

CARACTERIZACIÓN DE LOS DOCENTES DE LA SECRETARÍA DE EDUCACIÓN
DE MEDELLÍN PARA EL PERFILAMIENTO EN TELETRABAJO

MONICA PATRICIA ESCUDERO CAÑAS

UNIVERSIDAD PONTIFICIA BOLIVARIANA
ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA EN TECNOLOGÍAS DE INFORMACIÓN Y
COMUNICACIÓN
MAESTRÍA EN TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN
MEDELLÍN-ANTIOQUIA

2016

CARACTERIZACIÓN DE LOS DOCENTES DE LA SECRETARÍA DE EDUCACIÓN
DE MEDELLÍN PARA EL PERFILAMIENTO EN TELETRABAJO

MONICA PATRICIA ESCUDERO CAÑAS

Trabajo de grado para optar al título de MAGÍSTER EN TECNOLOGÍAS DE LA
INFORMACIÓN Y LA COMUNICACIÓN

Asesor

CRISTINA GÓMEZ SANTAMARÍA

PhD en Ingeniería Área Telecomunicaciones

UNIVERSIDAD PONTIFICIA BOLIVARIANA

ESCUELA INGENIERÍA

FACULTAD DE INGENIERÍA EN TECNOLOGÍAS DE INFORMACIÓN Y
COMUNICACIÓN

MAESTRÍA EN TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN

MEDELLÍN

2016

NOTA DE ACEPTACIÓN

Firma
Nombre
Presidente del jurado

Firma
Nombre
Presidente del jurado

Firma
Nombre
Presidente del jurado

Medellín, Diciembre de 2016

A los docentes de Medellín, con respeto y admiración.

AGRADECIMIENTOS

A mi familia por ser mi inspiración y fuerza, en especial a mi esposo Gustavo Rodríguez por el acompañamiento, apoyo y comprensión durante el desarrollo de este trabajo.

A mis mascotas Sirio y Luna por su compañía, siempre bajo la mesa de mi escritorio.

A Cristina Gómez por su amabilidad y guía en el proceso.

A la Universidad Pontificia Bolivariana por disponer sus recursos para el desarrollo de este trabajo.

DECLARACIÓN ORIGINALIDAD

"Declaro que esta tesis (o trabajo de grado) no ha sido presentada para optar a un título, ya sea en igual forma o con variaciones, en esta o cualquier otra universidad". Art. 82 Régimen Discente de Formación Avanzada, Universidad Pontificia Bolivariana.

FIRMA AUTOR (ES)

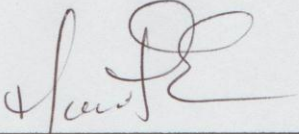

21 853701.

TABLA DE CONTENIDO

GLOSARIO.....	XIII
RESUMEN	XV
ABSTRACT	XVII
INTRODUCCIÓN.....	19
OBJETIVOS	23
OBJETIVO GENERAL	23
OBJETIVOS ESPECÍFICOS	23
FUNDAMENTOS TEÓRICOS.....	25
TELETRABAJO.....	25
<i>Innovación social.....</i>	<i>27</i>
<i>Características del teletrabajador.....</i>	<i>28</i>
<i>Rasgos de personalidad del teletrabajador.....</i>	<i>31</i>
<i>Competencias del teletrabajador.....</i>	<i>31</i>
INTELIGENCIA DE NEGOCIOS	34
SISTEMAS DE INFORMACIÓN	35
SISTEMA DSS	36
<i>Características de los sistemas DSS.....</i>	<i>37</i>
<i>Sistemas OLAP.</i>	<i>37</i>
BODEGAS O ALMACENES DE DATOS	39
ETL (EXTRACTION, TRANSFORMATION AND LOADING).....	40
PROCESO KDD	41
MINERÍA DE DATOS.....	43
<i>Aplicaciones de la minería de datos.....</i>	<i>44</i>
MÉTODO DE SOLUCIÓN	47
SISTEMA KDD Y METODOLOGÍA CRISP-DM.....	47
POBLACIÓN Y MUESTRA.....	52

VARIABLES	52
<i>Descripción de Variables.</i>	52
INSTRUMENTOS.....	55
HERRAMIENTAS	56
DESARROLLO DEL PROCESO.....	57
CAPÍTULO I: DESCRIPCIÓN DE VARIABLES, ANÁLISIS E IDENTIFICACIÓN	57
CAPÍTULO II: CALIDAD DE DATOS.....	58
<i>Primer análisis de perfil con herramienta DQ analyzer.</i>	59
<i>Segundo análisis de perfil con herramienta DQ Analyzer.</i>	66
<i>Tercer análisis de perfil con herramienta DQ Analyzer.</i>	86
CAPÍTULO III: MINERÍA DE DATOS	97
<i>Reglas de Asociación.</i>	100
<i>Clustering, algoritmo EM.</i>	109
IMPORTANCIA DEL TELETRABAJO PARA DOCENTES	114
CONCLUSIONES	118
TRABAJOS FUTUROS.....	120
REFERENCIAS.....	121
ANEXOS.....	125
ANEXO 1: ENCUESTA DE CARACTERIZACIÓN Y FORMACIÓN DOCENTE	125
ANEXO 2: INFORME DIRIGIDO A LA SECRETARÍA DE EDUCACIÓN DE MEDELLÍN .	126
ANEXO 3: ALGUNAS VARIABLES EN EL PRIMER ANÁLISIS CON HERRAMIENTA DQ	
ANALYSER.....	134

LISTA DE TABLAS

<i>Tabla 1: Descripción de variables.....</i>	<i>52</i>
<i>Tabla 2: Perfil #1 de calidad de datos.....</i>	<i>60</i>
<i>Tabla 3: Perfil #2 de calidad de datos.....</i>	<i>66</i>
<i>Tabla 4: Análisis para imputaciones de variables con datos nulos.....</i>	<i>72</i>
<i>Tabla 5: Relación porcentual de CODCARGOEMPRESA con Nulos en las variables NIVELDICTA, AREAEDUCATIVECNICA, AREAEDUCATIVA, NIVELEDUCACION.....</i>	<i>75</i>
<i>Tabla 6: Relación CODCARGOEMPRESA con nulos en variables FECHANACIMIENTO, SEXO, CONDICIONDEAMENAZADO.....</i>	<i>76</i>
<i>Tabla 7: Perfil # 3 de calidad de datos.....</i>	<i>87</i>
<i>Tabla 8: Perfil #3 de variable CODEMPLEADO</i>	<i>89</i>
<i>Tabla 9: Perfil #3 de variable FECHAINGRESONOMINA.....</i>	<i>89</i>
<i>Tabla 10: Perfil #3 de variable VINCULACIONCATEGORIA y VINCULACIONESTADO</i>	<i>90</i>
<i>Tabla 11: Perfil #3 de variable VACACIONESMETODO y CODCALENDARIO.....</i>	<i>90</i>
<i>Tabla 12: Perfil #3 de variable CODCARGOEMPRESA.....</i>	<i>91</i>
<i>Tabla 13: Perfil #3 de variable CENTROCOSTO.....</i>	<i>91</i>
<i>Tabla 14: Perfil #3 de variable NIVELCONTRATACION y CODEDICACIONLABORAL.....</i>	<i>92</i>
<i>Tabla 15: Perfil #3 de variable BASICO.....</i>	<i>92</i>
<i>Tabla 16: Perfil #3 de variable CODTURNO y SITUACIONLABORAL</i>	<i>93</i>
<i>Tabla 17: Perfil #3 de variable VINCULACIONNOVEDAD.....</i>	<i>93</i>
<i>Tabla 18: Perfil #3 de variable CIUDAD y AREAEDUCATIVA</i>	<i>94</i>
<i>Tabla 19: Perfil #3 de variable NIVELEDUCACION.....</i>	<i>94</i>
<i>Tabla 20: Perfil #3 de variable NIVELDICTA y REGIMENDEVINCULACION</i>	<i>95</i>

Tabla 21: Perfil #3 de variable HORASTRAB y AREAEDUCATIVECNICA	95
Tabla 22: Perfil #3 de variable GRADO	96
Tabla 23: Perfil #3 de variable FECHANACIMIENTO y SEXO	96
Tabla 24: Perfil #3 de variable CONDICIONDEAMENAZADO.....	96
Tabla 25: Configuración del algoritmo.....	104
Tabla 26: Configuración del algoritmo.....	105
Tabla 27: Configuración del algoritmo.....	106
Tabla 28: Tabla de caracterización docente según criterios de selección de teletrabajadores con el uso de Reglas de Asociación	108
Tabla 29: Información relevante de la segmentación de clustering.....	110
Tabla 30: Caracterización docente según criterios de selección de teletrabajadores con el uso de Clustering.....	112
Tabla 31: Perfil #1 de variables CODEMPLEADO y FECHAINGRESONOMINA	134
Tabla 32: Perfil #1 de variables FECHAINGRESONOMINA y CODVINCULACIONCATEGORIA	134
Tabla 33: Perfil #1 de variables VINCULACIONCATEGORIA y CODVINCULACIONESTADO	135
Tabla 34: Perfil #1 de variables VINCULACIONESTADO y CODVACACIONESMETODO.....	135
Tabla 35: Perfil #1 de variables VACACIONESMETODO y CODCALENDARIO	136
Tabla 36: Perfil #1 de variables NOMBRAMIENTOCODACTOADMIN y NOMBRAMIENTOFecha.....	136
Tabla 37: Perfil #1 de variables NOMBRAMIENTONUMERO y POSESIONCODACTOADMIN	137
Tabla 38: Perfil #1 de variables POSESIONFECHA y POSESIONNUMERO	137
Tabla 39: Perfil #1 de variables CONTINUIDAD y RETIROCODACTOADMIN.....	138
Tabla 40: Perfil #1 de variables RETIROFECHAACTOADMIN y RETIRONUMEROACTOADMIN	138

Tabla 41: Perfil #1 de variables CODCAUSARETIRO y CAUSARETIRO.....139

Tabla 42: Perfil #1 de variables FECHA CREACION y USUARIO139

LISTA DE FIGURAS

<i>Figura 1: Proceso de caracterización docente para teletrabajo del municipio de Medellín.</i>	<i>20</i>
<i>Figura 2: Proceso de extracción de conocimiento, desde un sistema transaccional</i>	<i>41</i>
<i>Figura 3: Etapas del KDD.....</i>	<i>42</i>
<i>Figura 4: Proceso que da como resultado la obtención de datos operacionales.....</i>	<i>47</i>
<i>Figura 5: Modelo Crisp_DM.....</i>	<i>48</i>
<i>Figura 6: Proceso para unir las bases de datos</i>	<i>58</i>
<i>Figura 7: Incremento del número de Teletrabajadores en los diferentes sectores (Molano, 2015) .</i>	<i>97</i>
<i>Figura 8: Planeación de un piloto de teletrabajo, según el Libro Blanco del Teletrabajo (MinTic, 2012)</i>	<i>99</i>
<i>Figura 9: Visualización del set de datos final para evaluar Reglas de Asociación</i>	<i>103</i>
<i>Figura 10: Configuración por defecto de algoritmo EM para Minería de datos con un número de 8 clustering o grupos.....</i>	<i>109</i>

GLOSARIO

TELETRABAJO: El teletrabajo es una alternativa nueva, atrevida, que propone soluciones a problemáticas de la sociedad moderna; introduce un cambio de paradigma laboral frente a las organizaciones, la economía, la educación, la ética, las estructuras y ordenes laborales tradicionales y la persona como sujeto social.

MINERÍA DE DATOS: Es entendida como el proceso de extraer información o patrones no triviales de una base de datos con medios automatizados, y dar sentido a la información para convertirla en un activo útil para la organización. Un proceso como la minería de datos, con distintos procesos o métodos cualitativos, convierte en conocimiento todos los detalles que a simple vista de la experiencia humana pasan inadvertidos en un universo masivo de información. Al final nos encontraremos con una bodega de información organizada, estructurada, y dispuesta para ser manipulada y entendida con parámetros ya automatizados.

INTELIGENCIA DE NEGOCIOS: Las grandes cantidades de información generada por las organizaciones pueden convertirse en un activo que apoye la toma de decisiones, por medio de técnicas, que sustentan todo ese conocimiento generado al interior de estas; pues bien, la inteligencia de negocios es la posibilidad de convertir grandes cantidades de datos en una fuente de soluciones al servicio de toda una estructura empresarial que más que reinventarse, necesita mejorarse así misma

cada día. Toda esta estrategia tiene un soporte en la tecnología, factor clave para encontrar las relaciones ocultas y formar escenarios posibles de la información, explorar los datos de la misma forma como pensamos, ser intuitivos. En conclusión serán los datos transformados en información e información transformada en conocimiento. (IN-FORMA-CION = no es más que la acción de dar forma a las cosas).

RESUMEN

TÍTULO: CARACTERIZACIÓN DE LOS DOCENTES DE LA SECRETARÍA DE EDUCACIÓN DE MEDELLÍN PARA EL PERFILAMIENTO EN TELETRABAJO

AUTOR: MONICA PATRICIA ESCUDERO CAÑAS

PALABRAS CLAVE: teletrabajo; inteligencia de negocios; minería de datos; docentes, calidad de datos; perfilamiento.

El presente proyecto desarrolla cuatro componentes: el teletrabajo, la calidad de datos, la inteligencia de negocios y la minería de datos, con el objetivo de buscar información que nos acerque a un perfilamiento de docentes en la Secretaría de Educación del municipio de Medellín; que permita implementar el teletrabajo como una innovación laboral para el sector educativo.

La Secretaría de Educación de Medellín y entidades afines poseen importantes bases de datos con información detallada de sus docentes, lo que permitiría con ayuda de herramientas BI (*Bussiness Intelligent*) identificar reglas y grupos de docentes para aplicar a proyectos pilotos en teletrabajo.

Los procesos de minería de datos pueden convertirse en una metodología aplicada en áreas como la educación que se implementa a la infraestructura de información como las bases de datos de docentes.

El resultado final será un documento que apoye los procesos de selección y clasificación de personal para el teletrabajo.

ABSTRACT

TITLE: Teacher Characterization from Secretary of Education in Medellin City for
Profiling the Telecommuting Worker

AUTHOR: MONICA PATRICIA ESCUDERO CAÑAS

KEYWORDS: Telecommuting; Business Intelligence; Data Mining; Teacher; Data Quality; Profiling.

The project below develops four components: telecommuting, data quality, business intelligence and data mining, with the purpose of searching information to develop a teacher profiling in Secretary of Education from Medellin city; in order to allow the implementation of telecommuting as innovation for educative sector.

Secretary of Education of Medellín and the related entities have important databases with detailed information about a big number of their teachers. It would allow by using Business Intelligent tools (BI) a clear identification of rules and groups of teachers who could apply for telecommuting pilot projects.

Data Mining Processes could become a methodology applied in areas such as education to the information infrastructure such as teacher databases.

The final result of this project will be a document to support processes of selecting and classifying of teacher staff for flexible working conditions inside educative sector.

INTRODUCCIÓN

El teletrabajo es una manera de flexibilización laboral que ofrece alternativas de solución a las situaciones particulares de las empresas y los trabajadores. Boiarov y Poliszuk (2011) explican que:

Durante la década del 70, en los EEUU, y en plena crisis del petróleo, el físico Jack Nilles comenzó a pensar formas de optimización de recursos no renovables. Su primera idea fue "llevar el trabajo al trabajador y no el trabajador al trabajo", tras lo cual creó el concepto de "telecommuting".

En la actualidad, la idea anterior ha cobrado mayor importancia gracias a la gran cantidad de herramientas que favorecen su implementación en todos los sectores, entre ellos, la educación, que requiere una transformación desde las tecnologías de la información y el diseño de nuevas propuestas que permitan realizar innovaciones de carácter social acordes con las necesidades actuales de los actores del sector educativo.

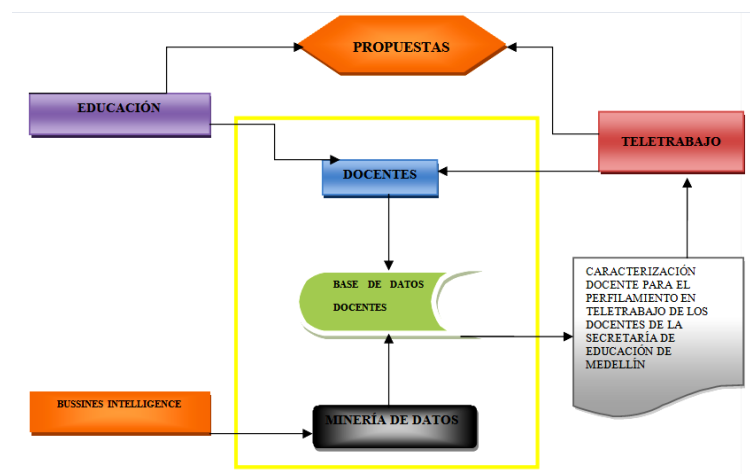
Es importante tener en cuenta que la educación aún no logra modificaciones ni transformaciones significativas pensadas para la era de la información, quedándose en estrategias pensadas para una era industrial que resultan ya obsoletas. Esto sugiere un conflicto entre las necesidades actuales de la población y los currículos tradicionales que las instituciones educativas ofrecen.

Los docentes son actores fundamentales de este proceso educativo y, por lo tanto, es necesario analizar qué transformaciones pueden desarrollar si se tiene en cuenta su entorno

laboral, familiar, y social. En consecuencia, un estudio de perfiles permitiría la identificación y observación de los docentes para llevar a cabo propuestas y estrategias de mejoramiento pensadas desde sus condiciones específicas que beneficien la transformación del sector educativo en general.

Por esta razón, se ha recurrido a técnicas de Bussiness Intelligence, como la minería de datos, que permite “explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto” (Sinnexus, 2007). De esta manera, se puede analizar y estudiar las características de los docentes para clasificarlos e identificar quiénes pueden involucrarse en actividades de teletrabajo. Esta es una incursión en un sector donde aún no se evidencian estudios orientados a una transformación de vida laboral.

Figura 1: Proceso de caracterización docente para teletrabajo del municipio de Medellín.



PLANTEAMIENTO DEL PROBLEMA

El teletrabajo es una alternativa que se ha implementado en diferentes países desde hace tiempo. Documentos como la Ley *Bassannini* en Italia, el Código de conducta en Irlanda o el Acuerdo macroeconómico de teletrabajo que se firmó el 16 de julio de 2002, en Bruselas, dan cuenta de cómo se implementa esta modalidad laboral en Europa.

Sin embargo, como propuesta laboral, el teletrabajo es desconocido para la sociedad colombiana en su forma y en su aplicación, y por lo tanto, se requiere de una identidad en su metodología, como lo evidencia Segura Jiménez (2014) en su blog:

A diferencia del trabajo tradicional, de sus estructuras y relaciones, los sujetos teletrabajadores, parecen tener que hacer un esfuerzo mayor por el reconocimiento de los otros como sujetos productivos, en la medida en que son mirados a partir de unos estereotipos en los que no encajan.

Es decir, la cultura laboral en Colombia aún se visualiza desde lo tradicional. Sin embargo, el teletrabajo es una innovación organizacional que el Estado implementó desde el año 2012. Esto supone la puesta en marcha de la estrategia en todas las organizaciones a partir de la fecha, brindando a todos los sectores la oportunidad de involucrar propuestas innovadoras en la modalidad laboral, a la luz de las nuevas exigencias mundiales.

Este es el caso del sector educativo, que se halla hoy frente a la oportunidad de introducir nuevas alternativas que cambien las dinámicas laborales de los docentes teniendo en cuenta sus necesidades personales, familiares y sociales.

Para ello es necesario determinar las características y rasgos apropiados de los docentes, a través de herramientas de análisis que permiten tomar decisiones y generar estrategias como la minería de datos. Este ejercicio sugiere una novedad para el sector ya que la minería de datos es muy utilizada en sectores como el comercial pero no en el sector educativo.

Una vez se acepte de forma cultural el teletrabajo, se creará para Colombia un marco con perspectivas en la vida laboral insospechadas hasta el momento.

.

OBJETIVOS

Objetivo general

Desarrollar un perfilamiento para la caracterización docente en la Secretaría de Educación de Medellín, mediante la utilización de herramientas de Inteligencia de negocios que favorezca la implementación del teletrabajo como una alternativa laboral para el gremio.

Objetivos específicos

- Realizar un levantamiento de información que permita hacer una caracterización de las herramientas y metodologías de inteligencia de negocios para teletrabajo de los docentes del municipio de Medellín.
- Identificar variables determinantes en el perfilamiento de los docentes para identificar bajo qué condiciones y qué profesor puede tele trabajar.
- Consolidar información buscada para la perfilación de los docentes de la Secretaría de Educación de Medellín mediante procesos de extracción, transformación y carga de la información de inteligencia de negocios.

- Implementar el proceso de minería de datos para el perfilamiento de los docentes según las variables seleccionadas de manera previa, y que permiten obtener información sobre la posible aplicación de teletrabajo en los docentes del municipio de Medellín.

FUNDAMENTOS TEÓRICOS

En el marco de las demandas de la sociedad actual se reúnen infinidad de necesidades que apuntan a establecer nuevas conexiones a nivel mundial; esta dinámica no es un cúmulo de individualidades sino que plantea procesos eficientes que se integren para la solución de los problemas a los que hoy se enfrenta la sociedad.

En este sentido, las tecnologías de la información parecieran ser el centro de la integración y convergencia hacia una nueva sociedad, donde la responsabilidad social como ente articulador de todas las organizaciones es un propósito.

Por esta razón, se presenta una investigación que tiene un carácter social, tecnológico, sistémico, innovador y práctico. Así pues, se presenta una propuesta de flexibilidad laboral, a través del estudio de perfilamiento para el sector educativo que se une a una propuesta de optimización de medios y recursos tal como lo ofrece DATA en el manejo de la información, dirigido a los docentes como los actores principales.

En consecuencia, es importante tener en cuenta los conceptos y criterios que al respecto se hallan en la literatura y están relacionados de forma directa con el tema.

Teletrabajo

El *teletrabajo* es una propuesta que flexibiliza la vida laboral de la persona, lo que permite crear una dinámica de soluciones viables frente a un entorno laboral, económico, personal, social, de constante cambio. Es importante establecer un estudio riguroso para

determinar las características y rasgos propios del teletrabajador de acuerdo con las necesidades de la empresa o el sector laboral específico. En este sentido, se requiere de profundos cambios de perspectiva que permitan establecer una plataforma estructurada desde los diferentes niveles organizacionales, que permitan el trabajo conjunto.

Para ampliar este tema se pueden citar algunas definiciones que fundamentan esta investigación:

“El *teletrabajo* se entiende como una modalidad laboral a distancia que usa las tecnologías de la información y las comunicaciones para alcanzar sus objetivos”
(Institución Universitaria CESMAG, 2014).

De acuerdo con la Ley 1221 de 2008, en el libro Blanco del Ministerio de las Tecnologías de la Información y las Comunicaciones (2012) el *teletrabajo* se define como:

Una forma de organización laboral, que consiste en el desempeño de actividades remuneradas o prestación de servicios a terceros utilizando como soporte las tecnologías de la información y la comunicación -TIC- para el contacto entre el trabajador y la empresa, sin requerirse la presencia física del trabajador en un sitio específico de trabajo. (p.11)

En otras palabras, el *teletrabajo* es una manera nueva de resolver una necesidad básica, es entender el trabajo como expresión de supervivencia, que permite al trabajador distribuir sus tiempos de una manera diferente y enriquecer su vida gracias al aprovechamiento del tiempo ya que no hay una presencia física en una instalación o empresa específica.

Como se mencionó, la implementación del teletrabajo es una innovación porque converge con las posibilidades actuales que ofrecen las tecnologías de la información y la comunicación y tiene un carácter social porque genera un impacto y un sinnúmero de posibilidades en la vida de las personas que llevan a cabo sus actividades laborales de esta manera. Se puede hablar entonces de *innovación social*.

Innovación social

La *innovación social* se refiere a la necesidad de insertar nuevas propuestas que mejoren lo que ya existe, que se introduzcan nuevas formas de hacer las cosas, que surjan soluciones a las problemáticas que el progreso genera. Una propuesta de *innovación social* debe estar enmarcada, desde la tecnología hacia los centros sociales estructurados y modelados, para fines de progresos económicos y de poder. Sugerir nuevos ordenes de carácter mundial, con implicaciones urbanas y de aceptación social, no implica otra cosa que pensar sobre lo que ya se tiene y reinventarlo con una consciencia social despierta y activa.

Como lo expresa Morales Gutiérrez *et al*, (2012):

Una generación relativamente reciente de emprendedores sociales ha comprendido, de manera notable, que la aplicación de las nuevas tecnologías genera innovación y multiplica el impacto social. La vinculación del emprendimiento social con las nuevas tecnologías explica por qué los últimos avances en apoyo al emprendimiento social se han centrado en la creación de sitios virtuales y físicos dedicados a la creación de valor social a través del desarrollo exhaustivo de herramientas TIC inteligentes y de una red competente de participantes civiles y públicos vinculados internacionalmente. (p.2)

Características del teletrabajador.

De acuerdo con Jaramillo y Bustamante (2011), en una publicación de la revista virtual de la Universidad Católica del Norte:

El éxito de su labor, eficiente y efectiva, depende de una excelente integración de la tecnología, su perfil (personal y laboral), las competencias comunicativas (esencialmente de lector-escritura, entre otras) y las funciones que tiene a cargo... Por ello, se habla de la necesidad de diagnosticar si las personas poseen el perfil del teletrabajador, con unas características que, si bien no son una camisa de fuerza, harían mucho más probable que una persona pueda ser exitosa en esta modalidad laboral. (p.1)

Por lo tanto, a continuación se enumeran algunas características que debe poseer una persona para teletrabajar tal y como las enuncian Jaramillo y Bustamante (2011):

Se orienta a obtener resultados: para el teletrabajo es esencial estar enfocado más en las metas que en el proceso. Esto no quiere decir que el proceso se descuide, sino que se debe tener siempre en mente la meta a la cual se quiere llegar.

Tiene excelentes habilidades para la comunicación: trabajar a distancia implica que no es posible obtener información directamente por observación, como se podría tener en el lugar de trabajo físico.

No requiere mucha supervisión: no necesita sentirse vigilado y supervisado para motivarse a realizar las funciones que se le han consignado y obtener los resultados previstos.

Es adaptable: el teletrabajador debe ser capaz de adaptarse, entre otras cosas, a su nuevo entorno de trabajo, en el cual no tendrá colegas a su alrededor, un jefe que lo supervise, ni clientes físicos en persona a quienes deberá atender. Además, debe poder adaptarse a las implicaciones del trabajo por medios tecnológicos, así como a las diversas innovaciones en materia de comunicación. Finalmente debe poder adaptarse a un entorno laboral en el que probablemente parte del tiempo esté rodeado de su familia, y de las tentaciones que podría tener en el hogar, como pueden ser la posibilidad de la televisión, la nevera, realizar llamadas personales extensas o frecuentes, entre otras.

Es muy organizado: si tener un trabajo directivo presencial requiere de organización, esto es aún más riguroso en un trabajo a distancia. La organización no solamente se refiere a un espacio físico que sea apropiado, cómodo, ergonómico, y tranquilo, sino también al adecuado manejo del tiempo, planeación, seguimiento y evaluación de las actividades que se deben realizar, así como otros aspectos que por estar a distancia requieren de una buena dosis de autogestión y autocontrol.

Tiene un sólido conocimiento sobre su trabajo.

Entiende profundamente los objetivos de la organización: al estar por fuera de la empresa física, el teletrabajador puede perder algunas formas de comunicación que usualmente se practican en la empresa como son las carteleras, boletines físicos, algunas

reuniones, y por ende olvidar, o no tener muy presente, cuáles son la misión, visión, y objetivos institucionales.

Puede establecer prioridades y manejar bien el tiempo: como se indicó en un punto anterior, establecer prioridades y manejar bien el tiempo son habilidades esenciales cuando se está en teletrabajo

Ser exitoso en las posiciones actuales y previas

El entorno hogareño es favorable para el teletrabajo: este punto es básico. Si el entorno hogareño no se presta para el teletrabajo entonces esto afectará negativamente su desempeño.

Su trabajo tiene claramente definidos sus objetivos: para que el teletrabajador pueda tener éxito, debe tener unas tareas con objetivos bien delimitados y medibles. Además, se debe tener en cuenta que se deben priorizar los resultados más que los procesos, como ya se indicó anteriormente.

En conclusión, ser teletrabajador no solo es responder a un piloto de viaje preestablecido de normas y exigencias, es también entender que el trabajo reflejará los rasgos propios de la personalidad, y si se es consciente de ello, el teletrabajador podrá ofrecer al mundo comportamientos que correspondan con los claros propósitos de un ideal definido.

Rasgos de personalidad del teletrabajador.

1. Ser auto motivado: al no tener un supervisor o jefe inmediato cerca, y al tener unos grandes espacios de tiempo antes de ser contactado para una supervisión, es necesario que la motivación sea interna y no externa.
2. Tener tendencia al orden y al control de las situaciones: existen tipos de personalidad que usualmente son bastante organizados, controladores y perfeccionistas. Estas son cualidades que resultan positivas para el teletrabajo puesto que, como se indicó antes, requiere de bastante de organización, control de los detalles, y prevención de posibles riesgos.
3. Ser perfeccionista: dado que no existe un supervisor que esté prestando atención a los detalles del trabajo y los resultados esperados constantemente, es necesario que el control de calidad se haga desde la misma tele trabajadora.
4. Ser alegre y optimista: tener un buen estado de ánimo es vital en la tele trabajadora, puesto que no tendrá la presión o compañía de otros colegas o jefes si se siente pesimista o aburrido en el trabajo.

Competencias del teletrabajador.

- Clarificar el pensamiento: Los usuarios comprenden cómo se conectan las ideas y se dan cuenta de cómo se puede organizar o agrupar la información. Con el aprendizaje visual, los nuevos conceptos son más profundos y asimilables.

- Reforzar la comprensión: Los lectores reproducen en sus propias palabras (parafrasear) lo que han aprendido. Esto les ayuda a absorber e interiorizar nueva información, dándoles posesión sobre sus propias ideas.
- Integrar nuevo conocimiento: Los esquemas incitan a los lectores a construir sobre su conocimiento previo y a integrar la nueva información.
- Detectar conceptos erróneos: Al tiempo que un mapa conceptual, un esquema u otra herramienta de aprendizaje visual refuerza el mensaje y los contenidos, los enlaces mal dirigidos o conexiones erradas dejan al descubierto vacío de información.

De acuerdo con las características, rasgos de la personalidad y competencias del teletrabajador, en el sector educativo es importante que los docentes tengan alternativas acorde con sus capacidades en aras de la diversificación. El docente puede flexibilizar su rol como agente acompañante de forma externa con el uso de las nuevas plataformas que ofrece la tecnología de la información y la comunicación, puede salir de la estructura formal que representa un “encierro” para el estudiante y el mismo docente. Permitirle al docente que haga uso de otras propuestas como el teletrabajo, le permite hacer un acompañamiento desde la libertad y flexibilidad del estudiante y de él mismo, razón por la cual es necesario que la Secretaría de Educación tenga un perfil de docentes aptos para teletrabajo: de esta manera, se optimiza el personal a través de prácticas de teletrabajo.

Entre las ventajas que puede ofrecer este sistema de trabajo se halla la optimización del tiempo a través de una planeación orientada hacia los objetivos, la reducción del

desplazamiento laboral de los docentes a los lugares de trabajo, la conciliación laboral entre el docente y su vida familiar, la potenciación de la autonomía y la creación de rutinas dinámicas en la escuela que permitan la mejora de la educación y la calidad de vida del docente.

Por otra parte, el teletrabajo puede representar una oportunidad para atender la población con discapacidad pues, como se mencionó, no requiere la presencia física en una institución. Al respecto, se presenta un fragmento de un artículo publicado por el Ministerio de Educación Nacional –MEN- (2007).

Más de 2 millones 600 mil colombianos padecen algún tipo de discapacidad, es decir, el 6.4% de la población nacional. Frente a este hecho, el uso de las Tecnologías de Información y Comunicación se está convirtiendo en un aliado eficaz para hacer realidad una educación inclusiva. No obstante que la brecha digital, aún entre quienes tienen acceso o no a esas tecnologías es todavía muy grande, para aquellos con discapacidad, Internet representa una posibilidad única de reinserción social, mediante opciones como el teletrabajo y la participación en comunidades virtuales y de aprendizaje.

Además, nuevas alternativas laborales para los docentes generan la posibilidad de progreso, por ello la adopción del teletrabajo es una construcción social más que tecnológica y una oportunidad de cara al futuro, tal y como lo proclama el CAF (Banco de Desarrollo de América Latina) la Organización Iberoamericana de Juventud, la Secretaría General Iberoamericana, entre otras organizaciones.

La educación hoy más que en otras épocas necesita del abordaje y modificación de sus escenarios para que sean compatibles con todos los contextos actuales, con la educación a distancia, con los programas e-learning, con las comunidades de educación especial, los discapacitados, la geografía y las necesidades de los docentes y la dignificación de su labor.

Pero esta transformación educativa implica un proceso a largo plazo en el que es necesario involucrar conocimientos relacionados con las tecnologías y el manejo de las bases de datos, con el fin de identificar aquellos docentes que pueden teletrabajar atendiendo a sus necesidades y capacidades específicas.

En el caso concreto de Medellín, un análisis de las bases de datos de la Secretaría de Educación permitiría generar un perfil de docentes para tal fin. Sin embargo, esta acción implica la intervención de una serie de procesos en los que se involucran varios aspectos que se mencionan a continuación.

Inteligencia de negocios

Entendida como la capacidad para transformar los datos en información a través de aplicaciones y tecnologías que permitan predecir acciones y tomar decisiones al respecto.

En la literatura se hallan definiciones concretas que permiten ampliar esta definición:

Habilidad para entender la interconexión de los hechos presentes como medio para descubrir guías que nos permitan llegar a metas futuras, lo que significa mejorar nuestra capacidad de hacer juicios predictivos para el acierto en la toma de decisiones. Gómez y Vargas (2012)

Business Intelligence is a broad category of applications and technologies used to collect, archive, analyze and access data, which helps users in decision.(Vasile & Mirela, 2008, p.1364)

Habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios. (Sinnexus, 2007)

Sistemas de información

Hace referencia a la optimización del desarrollo de las actividades de una organización al organizar la información para poder crear escenarios de conocimiento concretos; no todos los datos se transforman en información a menos que sean procesados e interpretados, los datos pueden ser clasificados, y formar una síntesis; no tiene que ser computarizado, pero hoy debido al apoyo se hace en computadoras. En esta era de la información, el enfoque de las compañías ha cambiado de la orientación hacia el producto a la orientación hacia el conocimiento.

Los *sistemas de información* responden al campo de conocimiento al que se aplican y su clasificación puede hacerse de forma genérica en sistemas competitivos, sistemas cooperativos y sistemas que pueden modificar el estilo de operación del negocio de acuerdo al entorno de aplicación transaccional o decisiones como son la informática, la teoría de sistemas, la seguridad informática, las matemáticas, la geografía y cartografía, entre otros.

En el National Information Systems Security Glossary (2014) se halla la siguiente definición sobre *sistemas de información*: “Sistema que se utiliza para obtener, almacenar,

manipular, administrar, mover, controlar, desplegar, intercambiar, transmitir o recibir voz y/o datos, e incluye tanto los programas de computación como el equipo de cómputo”.

En informática por ejemplo, los *sistemas de información* son un sistema computacional, se puede clasificar en EIS (sistemas estratégicos o sistema de apoyo a los ejecutivos, sistemas de información ejecutiva), TPS (servicio a nivel operativo o sistema de procesamiento de transacciones), DSS (sistema de nivel de conocimiento, Sistemas de Apoyo a las Decisiones o soporte a las decisiones), MIS (herramientas de uso gerencial o sistemas de información gerencial).

Sistema DSS

El sistema DSS hace referencia a los sistemas que soportan la toma de decisiones y apoyan el proceso. En otras palabras, el análisis de información por sí sola no genera un conocimiento, requiere una técnica en la que se utilizan herramientas tecnológicas.

En este sentido, el sistema DSS es un sistema de tipo informático que apoya las decisiones, de tipo estructurada como las también llamadas no estructuradas. Esto conlleva un proceso inicial cuya metodología es la etapa de inteligencia o búsqueda de problemas, luego sigue con la realización de modelos o diseño, solución al modelo y desarrollo de la solución. Esto no garantiza un resultado eficiente pero puede dar luces para retomar el proceso desde el inicio con directrices diferentes que se acerquen en mayor medida a los propósitos establecidos inicialmente. La Universidad Nacional Abierta y a Distancia (2014), retoma la definición de Turban (1995) para indicar que es un sistema que “Utiliza

datos, proporciona una interfaz amigable y permite la toma de decisiones en el propio análisis de la situación"

Características de los sistemas DSS.

- Estos sistemas permiten profundizar en los datos, navegar entre ellos, manejarlos desde distintas perspectivas, en otras palabras, son informes dinámicos, flexibles e interactivos.
- No requiere conocimientos técnicos. Un usuario no técnico puede crear nuevos gráficos e informes y navegar entre ellos. No es imprescindible la ayuda del personal de tecnologías de información.
- Rapidez en el tiempo de respuesta. Utilizan estructuras optimizadas para la consulta y el análisis de grandes volúmenes de información (*DataWarehouse/Datamart*).
- Integración de datos provenientes de diferentes fuentes que pueden integrarse a datos provenientes de Hojas de cálculo, Bases de datos internas/externas
- Cada usuario dispone de información adecuada a su perfil, no se trata de que todo el mundo tenga acceso a toda la información, sino de que tenga acceso a la que necesita para hacer su trabajo lo más eficiente posible.

Sistemas OLAP.

Proviene del término en inglés “*On-line Analytical Processing*”. Es un sistema en línea que permite el planteamiento de soluciones para el análisis de grandes cantidades de

información. De esta manera se agilizan las consultas de grandes cantidades de datos y permite a los usuarios analizar diferentes dimensiones.

Esta categoría de herramientas de software que permite organizar, presentar y analizar, es útil para reportes rápidos, permite al usuario un acercamiento ágil ya que no necesita de conocimientos muy técnicos en la estructura en la que se organiza la información.

Por otra parte, OLAP tiene una relación directa con las bodegas de datos. Al respecto Orlando *et al* (2007) señalan que:

El análisis OLAP tiene un mejor desempeño si la información reposa consolidada en una sola fuente, y más aún si esa fuente es una base de datos multidimensional que se comporte como una única bodega de datos. Esta característica ha llevado a pensar que OLAP depende de las bodegas de datos o que OLAP y la bodega de datos son la misma tecnología, pero aunque ambas usan el concepto de modelo multidimensional, su origen fue independiente y se han integrado para facilidad de los análisis.

De cualquier forma, la idea es centralizar la información de la organización, no dejar en manos de la parte operativa detalles de la información que pueden ser importantes en la toma de decisiones.

Sobre cubos OLAP se puede hacer minería de datos; los datos se estructuran según las áreas del negocio y los formatos se alimentan de los sistemas operacionales, así su implementación se haya hecho de forma independiente. Esta es una tecnología muy usada

en base de datos, que se caracteriza por ser amigable y atender problemas o requerimientos específicos.

Bodegas o almacenes de datos

Las decisiones actuales se toman basadas en información, cuyos grandes volúmenes necesitan de una estructura optimizada para facilitar consultas muy complejas que contengan históricos de datos. Las bodegas de datos cumplen con esta labor porque son estables y tienen un tipo de acceso de solo lectura que rara vez se borra; sus datos no tienen frecuencia de modificación y sus usuarios son la gerencia, direcciones y ejecutivos, es decir, los tomadores de decisiones o los que apoyan la toma de decisiones, lo que hace que el análisis sea de características multidimensionales y sirva para hacer análisis de tendencias.

Es importante tener en cuenta que el número de registros por acceso es millonario, pero un consolidado único de bases de datos, alimentado de múltiples fuentes, tanto internas como externas, mejora la recolección e integración de datos, mejora el almacenamiento masivo de la información, acelera la recuperación y soporte de nuevos tipos de datos. Por otra parte, las bodegas de datos se consideran un punto de partida para la minería de datos o también la base de datos de un sistema de inteligencia de negocios y su primera necesidad es facilitar los análisis o consolidar la información OLAP.

De acuerdo con William H. Inmon, una bodega de datos o *DataWarehouse* es “una colección de datos, orientados a hechos relevantes del negocio, integrados, que incluyen el

tiempo como característica importante de referencia y no volátiles para el proceso de toma de decisiones” (Orlando *et al*, 2007).

ETL (Extraction, Transformation and Loading)

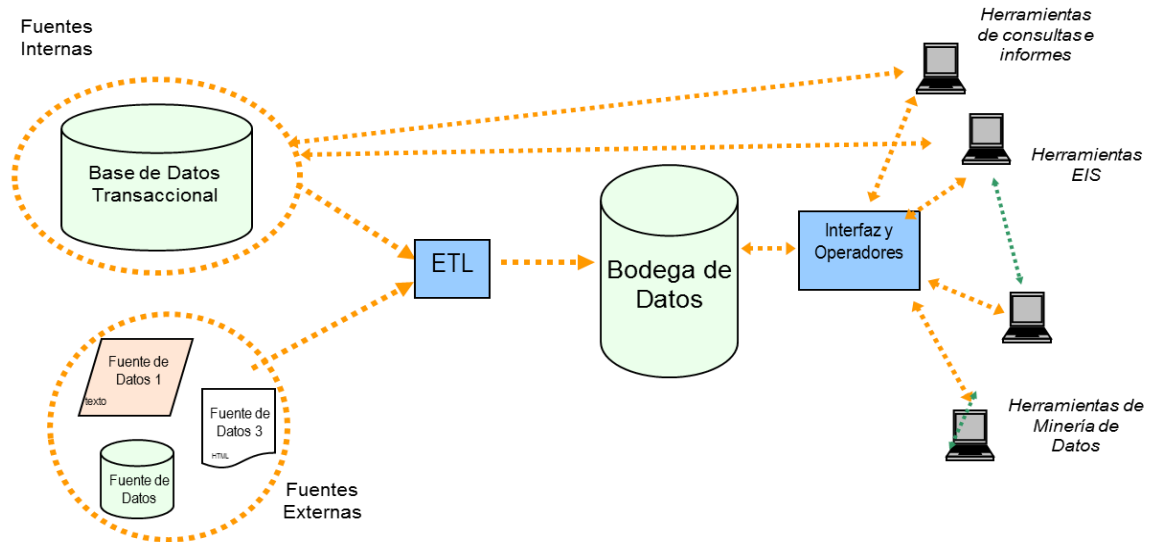
Es una aplicación compuesta de un proceso de extracción, un proceso de transformación y un último proceso de carga, cuyo objetivo es establecer conexiones que faciliten de mejor forma el viaje de datos, ya transformados, sin dificultad a los propósitos iniciales de calidad que se pretenden. Los tres procesos se describen a continuación:

- La *extracción* tiene su inicio en las fuentes de información, con la responsabilidad de tener el mínimo o nada de impacto sobre los datos, y trabajar con el tamaño para que no afecte la calidad en el servicio.
- La *transformación* es un proceso sujeto a cambios y diseños constantes en las funciones que afectarán las decisiones determinantes que llevarán al cambio; tienen la responsabilidad de ofrecer al siguiente proceso los recursos adecuados para la generación de un resultado final óptimo e ideal.
- La *carga* es un proceso que tiene unas especificaciones de acuerdo con la organización y garantiza la calidad de los datos.

Es importante anotar que las ETL son parte esencial del ciclo de vida de una implementación de inteligencia de negocios, que su intervención y destino se da en las bases de datos de las cuales se dispone.

En palabras de Amón Uribe (2010) es “un proceso de ETL es un conjunto de programas utilizado para extraer datos, transformarlos de acuerdo con una serie de criterios de calidad, integrarlos y cargarlos en un repositorio de información”.

Figura 2: Proceso de extracción de conocimiento, desde un sistema transaccional



Tomado de la presentación del docente Amón Uribe en el año 2014

Proceso KDD

La sigla se traduce como Proceso de Extracción de Conocimiento (*Knowledge Discovery in Databases*, por sus siglas en inglés). Alcanzar un nuevo conocimiento completo y productivo requiere de la implementación de un sistema que permita desarrollar la esencia del problema, que se integre a un ambiente de trabajo preestablecido y provea soluciones en tiempos reales. El KDD es un proceso más general que la minería de datos, pero la incluye, se centra en el usuario y depende de las decisiones que éste tome. Este

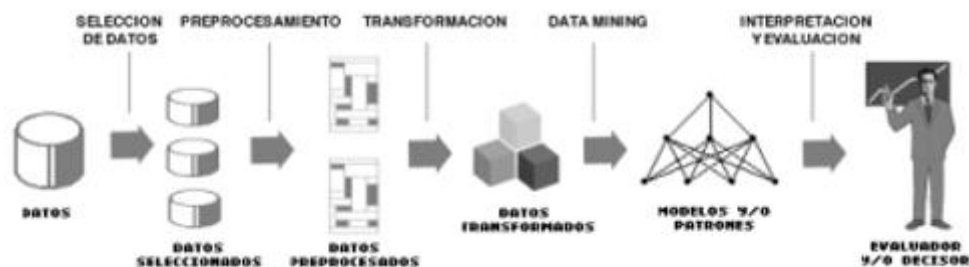
proceso va desde la limpieza de datos hasta la extracción de conocimiento. Es un campo de la inteligencia artificial y ha cobrado importancia con el crecimiento actual de las bases de datos; no es un producto de software sino un proceso compuesto de varias etapas en las que las librerías son los pilares fundamentales.

Fayyad *et al*, (1996) lo definen como un “proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia comprensibles a partir de los datos”. (Pag.40).

El KDD nace como interfaz y se nutre de diferentes disciplinas como Estadística, Sistemas de información o bases de datos, aprendizaje automático o IA, visualización de datos, computación paralela o distribuida e interfaces de lenguaje natural a bases de datos.

Como el proceso explora grandes volúmenes de datos, se divide en cinco etapas que se describen a continuación en la Figura 3

Figura 3: Etapas del KDD.



Las etapas del proceso son descritas por *Webmining Consultores* (2011) en el artículo *KDD: Proceso de Extracción de conocimiento*.

Minería de datos

La llamada sociedad de la información requiere de procesos automatizados que generen innovación y competitividad. En este sentido la minería de datos ofrece estas posibilidades ya que es un soporte para extraer conocimiento de una base de datos en forma automatizada o semiautomatizada. En la literatura se hallan algunas definiciones que permiten ampliar este concepto y se menciona a continuación:

La minería de datos es el arte o ciencia de descubrir, de manera “automática o “semi automática”, patrones o información de valor, en grandes bases de datos para soportar la toma de decisiones del negocio, fase que integra los métodos de aprendizaje y estadísticas para obtener hipótesis de patrones y modelos (Amón Uribe, 2010).

La minería de datos es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto (Sinnexus, 2007).

En otras palabras, la minería de datos surge para ayudar a comprender el contenido de un repositorio de datos. Con este fin, hace uso de prácticas estadísticas y, en algunos casos, de algoritmos de búsqueda próximos a la Inteligencia Artificial y a las redes neuronales.

Es importante mencionar que los datos son la materia prima. Cuando el usuario les atribuye un significado especial, estos pasan a convertirse en información; y cuando surge una interpretación entre la información y el modelo que elaboran o encuentran los especialistas, entonces se habla de conocimiento.

Aunque cada caso concreto puede ser distinto al anterior en la minería de datos, el proceso es común a todos ellos y suele componerse de cuatro etapas principales como se explica en Sinnexus (2007):

- Determinación de los objetivos. Trata de la delimitación de los objetivos que el cliente desea bajo la orientación del especialista en minería de datos.
- Pre-procesamiento de los datos. Se refiere a la selección, la limpieza, el enriquecimiento, la reducción y la transformación de las bases de datos. Esta etapa consume, por lo general, alrededor del 70% del tiempo total de un proyecto de minería de datos.
- Determinación del modelo. Comienza con un análisis estadístico de los datos, y después se lleva a cabo una visualización gráfica de los mismos para tener una primera aproximación. Según los objetivos planteados y la tarea, pueden utilizarse algoritmos desarrollados en diferentes áreas de la Inteligencia Artificial.
- Análisis de los resultados. Se verifica si los resultados obtenidos son coherentes y se cotejan con los obtenidos por los análisis estadísticos y de visualización gráfica. El cliente determina si son novedosos y si le aportan un nuevo conocimiento que le permita considerar sus decisiones.

Aplicaciones de la minería de datos.

En el sector de la salud ofrece a los usuarios el análisis de reclamos, de proveedores, de costos, de fraudes, de enfermedades, cuidado preventivo, diagnósticos mejorados, gestión de enfermedades, evaluaciones de salud, e investigación.

Al respecto, Vega *et al* (2012), explican que:

En la actualidad son muchas las áreas del sector salud en las que ha incursionado la minería de datos desde aplicaciones muy complejas como el reconocimiento de imágenes en el cerebro, hasta los procesos para la gestión de los pacientes en hospitales, como por ejemplo, en la predicción y el tratamiento de enfermedades de especialidad analizando la sintomatología, las enfermedades y los resultados de estos tratamientos, siendo un factor importante en la atención de enfermedades tales como: cáncer, problemas del corazón, tratamiento y procesamiento de imágenes, entre otras. (p.1)

También tiene aplicaciones en comercio para marketing y análisis de clientes, segmentación de clientes, Market Basket Analysis, planeación de promociones, mercado multicanal; y aplicaciones en operación de tiendas tales como planeación de espacios, pronósticos de ventas, análisis de persona x tienda, análisis de fraude, análisis de mercados, planeación y presupuesto de inventarios, planeación de distribución de productos, análisis de precios y gestión de marcas.

Otro elemento que se involucra en el análisis de datos está relacionado con el análisis predictivo que Herschel *et al* (2010), pertenecientes al grupo Gartner, explican “ayuda a conectar los datos con la toma de decisiones al extraer conclusiones fiables acerca de las circunstancias actuales y los eventos futuros”.

Como se ha expuesto hasta ahora, en la revisión de la literatura no se hallan registros o evidencias de investigaciones acerca del perfilamiento de docentes para teletrabajo que utilicen herramientas de Inteligencia de negocios, ni está sugerido en la documentación encontrada de aplicación que se hace a la minería de datos en los diferentes campos. Esto supone una novedad en la aplicación y utilización de las herramientas de BI a un campo

social y tan importante como lo es la educación. Por lo tanto, se presenta a continuación una propuesta de caracterización para el perfilamiento de docentes, aplicado a las bases de datos de la Secretaría de Educación de Medellín.

MÉTODO DE SOLUCIÓN

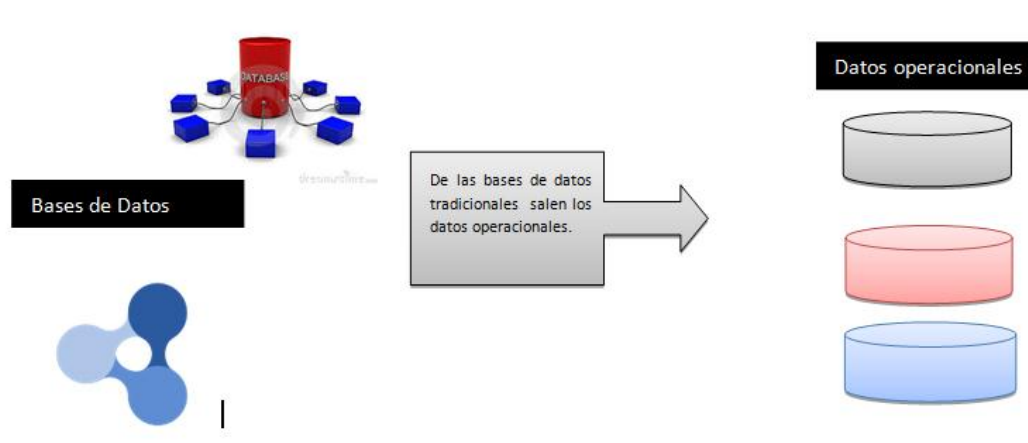
Sistema KDD y metodología CRISP-DM

Todo el proceso parte de una necesidad del sector de la educación, donde se reconoce que existen unos actores importantes que son los docentes y que dentro de sus múltiples necesidades está la de hallar propuestas laborales flexibles. Por este motivo, se propone a continuación un procedimiento que permita obtener información laboral de las bases de datos de Secretaría de Educación de Medellín, analizar y tomar decisiones respecto a las posibilidades de teletrabajo que pueden ofrecerse a los docentes.

Se inicia con la extracción de las bases de datos que se hallan en diferentes áreas de la Secretaría de Educación como son: nómina, recursos humanos, Casa del maestro, vivero del software y entidades externas a la Secretaría como las universidades y sindicatos. La extracción de esta información entrega los datos operacionales del proyecto.

De forma gráfica sería así:

Figura 4: Proceso que da como resultado la obtención de datos operacionales

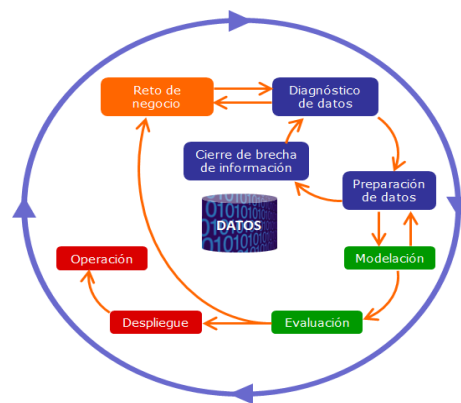


Para transformar la información, en primer lugar, se aplica la minería de datos como un proceso técnico a través del KDD que facilita la reunión de sistemas de información, su análisis y la generación de conocimiento

En segundo lugar, se aplica la metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*) para el desarrollo del proceso. La metodología CRISP-DM describe una serie de pasos a seguir y se sugiere porque responde al estudio que se hace sobre las reglas del negocio previo, hecho que permite conocer un entorno, la posibilidad de retroalimentación en fases anteriores y la confianza de ser una de las más utilizadas.

Tener una metodología de minería de datos ofrece orden y guía a través de todo el proceso; es importante subrayar que durante el desarrollo se hace una mezcla de diferentes procesos, metodologías y técnicas, descritas en líneas anteriores. La metodología se enmarca dentro de la resolución de problemas que se apoyan en el análisis predictivo, que ayuda a conectar los datos con la toma de decisiones.

Figura 5: Modelo Crisp_DM



Fuente: Logyt Discovery con base en el Modelo de CRISP. 2010

Figura obtenida de <http://images.frompo.com/8b9556da604f551126c595ecbcfdea22>

Fases de la metodología CRISP-DM:

Comprensión del negocio: (Objetivos y requerimientos desde una perspectiva no técnica) Establecimiento de los objetivos del negocio: Contexto inicial, objetivos, criterios de éxito, Evaluación de la situación (Inventario de recursos, requerimientos, supuestos, terminologías propias del negocio, entre otros)

Establecimiento de los objetivos de la minería de datos: (Objetivos y criterios de éxito) Generación del plan del proyecto (plan, herramientas, equipo y técnicas). La búsqueda de información debe iniciar en la entidad que tiene almacenada toda la base de datos de los docentes públicos del municipio Medellín y luego en otras entidades.

Gestión de la información: Hacer una gestión para tener el acceso a la información, a través de cartas sustentadas y apoyadas por la universidad, visitas y contactos en las organizaciones.

Comprensión de los datos: (Familiarizarse con los datos sin descuidar los objetivos del negocio) Integrar la información que está distribuida en diferentes bases de datos para convertirlo en una sola:

- ✓ Descripción de los datos
- ✓ Exploración de los datos
- ✓ Verificación de calidad de datos

Preparación de los datos: (Obtener la vista minable o *dataset*) Aplicar herramientas de software que permitan evaluar la calidad de los datos y garantice una información

objetiva. Se sugieren las herramientas *DQ Analyzer* y *Fril*, además de hacer correlación de variables. El resultado esperado serían unos datos aptos para el análisis sin variables complicadas con las que se corre el riesgo de tener datos duplicados o que una variable determina la otra, en conclusión información poco confiable.

Modificación: Transformación y filtración de variables para adecuarse a los requisitos del problema o cuestión que se pretende analizar

- ✓ Selección de los datos
- ✓ Limpieza de datos
- ✓ Construcción de datos
- ✓ Integración de datos
- ✓ Formateo de datos

Modelado: (Aplicar las técnicas de minería de datos a los *Dataset*) Determinar el tipo de tarea de minería de datos que aplica para el desarrollo del proyecto. Para este el estudio en la presente investigación, se determinó la tarea de asociación, porque busca similitud entre los datos, pero a diferencia del *Clustering*, lo hace en forma de reglas, no requiere de un campo de entrenamiento o historial de la información.

Diseño de patrones: Diseñar los patrones que definirán la información para hacer la minería de datos, solucionar los problemas apoyados en datos, descubrir patrones, reglas ocultas en los datos y predecir un comportamiento a través de la elección de unas variables significativas, y llegar a patrones de asociación.

- ✓ Diseño de la evaluación

- ✓ Construcción del modelo

- ✓ Evaluación del modelo

Aplicar herramienta de minería de datos: Se sugiere la herramienta Weka por ser un software gratuito, de fácil acceso e instalación, ofrece múltiples algoritmos para hacer minería y permite aplicar, analizar y evaluar técnicas de análisis de datos. Weka está escrito en lenguaje Java y fue desarrollada en la universidad de Waikato.

Evaluación: Respecto a los modelos de las fases anteriores para determinar si son útiles a las necesidades del negocio

Conclusiones y resultados: Información nueva:

- ✓ Evaluación de resultados

- ✓ Revisar el proceso

- ✓ Establecimiento de los siguientes pasos o acciones.

Despliegue: Explotar utilidad de los modelos, integrándolos en las tareas de toma de decisiones de la organización.

- ✓ Planificación de despliegue

- ✓ Planificación de la monitorización y del mantenimiento

- ✓ Generación de informe final

- ✓ Revisión del proyecto

Esta última fase no estaba dentro del alcance del proyecto, por tal motivo no se aborda para tal caso.

Población y muestra

El proyecto está dirigido a los docentes que se hallan vinculados a la Secretaría de Educación de Medellín, a través de cualquier tipo de contrato tradicional.

Variables

De acuerdo con la literatura acerca del teletrabajo, no todas las personas pueden teletrabajar. Al respecto, Jaramillo y Bustamante (2011) exponen:

Sin embargo, no todos los trabajos son susceptibles actualmente de ser realizados a distancia, e igualmente, no todos los trabajadores tendrían éxito o mejorarían su calidad de vida tanto personal como profesional si hicieran el tránsito hacia el teletrabajo. Por ello, se habla actualmente de la necesidad de diagnosticar si las personas poseen el perfil del teletrabajador, con unas características que, si bien no son una camisa de fuerza, harían mucho más probable que una persona pueda ser exitosa en esta modalidad laboral...”

En consecuencia, se debe analizar una serie de rasgos y características que pueden hacer parte del perfil del teletrabajador. Dichos aspectos surgen de la información que se halla en las bases de datos y se han denominado, para esta investigación, como variables que se evalúan para el cumplimiento de los objetivos.

A continuación se indica el nombre de las variables tal y como surgen de la base de datos y se describen sus características.

Descripción de Variables.

Tabla 1: Descripción de variables

EXPRESIÓN O VARIABLE	DESCRIPCIÓN DE LA VARIABLE
----------------------	----------------------------

EXPRESIÓN O VARIABLE	DESCRIPCIÓN DE LA VARIABLE
CODEMPLEADO	Código único para cada empleado.
FECHAINGRESONOMINA	Fecha desde la cual se realizan pagos de nómina al funcionario.
FECHAINGRESOEMPRESA	Fecha real desde la que se posesionó en el cargo.
CODVINCULACIONCATEGORIA	Código de la categoría de vinculación del funcionario.
VINCULACIONCATEGORIA	Categoría de vinculación del funcionario.
CODVINCULACIONESTADO	Estado de la actividad. Ejemplo A=Abiertas Inscripciones, C=Cerradas Inscripciones, E=Ejecutando
VINCULACIONESTADO	Código del estado de la vinculación del funcionario.
CODVACACIONESMETODO	Código del método de vacaciones que aplica para el funcionario.
VACACIONESMETODO	Método de vacaciones para el funcionario.
CODCALENDARIO	Código del calendario al que pertenece la vinculación del funcionario
NOMBRAMIENTOCODACTOADMIN	Código del tipo de acto administrativo de nombramiento.
NOMBRAMIENTOFecha	Fecha del acto administrativo de nombramiento.
NOMBRAMIENTONUMERO	Número del acto administrativo de nombramiento.
POSESIONCODACTOADMINISTRATIVO	Código del tipo de acto administrativo de posesión.
POSESIONFecha	Fecha del acto administrativo de posesión.
POSESIONNUMERO	Número del acto administrativo de posesión.
CONTINUIDAD	Almacena un carácter (S o N) si el funcionario inicia la vinculación con continuidad o la pierde.
RETIROCODACTOADMINISTRATIVO	Código del tipo de acto administrativo de retiro.
RETIROFechaAACTOADMINISTRATIVO	Fecha del acto administrativo de retiro.
RETIRONUMEROAACTOADMINISTRATIVO	Número del acto administrativo de retiro.
CODCAUSARETIRO	Código de la causa de retiro del funcionario.
CAUSARETIRO	Descripción de la causa del retiro del funcionario.
FECHACREACION	Fecha de creación del registro dentro del sistema.
USUARIO	Nombre de usuario que realizó la actualización del registro dentro del sistema.
DETORIGEN	Detalle del origen de inserción de la novedad.
DETORIGEN2	Detalle 2 de la inserción de la novedad.
ORIGEN	Origen de inserción de la novedad de vinculación.
RETIROCONFIRMADO	Define si el retiro está o no confirmado.
CODCONTINUIDADMOTIVO	Código del esquema de motivos de continuidad del docente.
CODEMPRESACONTINUIDAD	Código de continuidad en empresa.
CODEMPRESATRASLADO	Código de traslado en empresa.
CODESQUEMA	Código del esquema de pago al cual pertenece el funcionario durante cada detalle dentro de la vinculación.
CODCARGOEMPRESA	Código del cargo que desempeña el funcionario durante cada detalle dentro de la vinculación.
CENTROCOSTO	Nombre del centro de costo.
CODNIVELCONTRATACION	Código del nivel de contratación del funcionario durante cada detalle dentro de la vinculación.
NIVELCONTRATACION	Nombre del nivel de contratación del funcionario durante cada detalle dentro de la vinculación.
CODDEDICACIONLABORAL	Código de la dedicación laboral del funcionario durante cada detalle dentro de la vinculación.
CODCENTROCOSTO	Código del centro de costo al cual pertenece el

EXPRESIÓN O VARIABLE	DESCRIPCIÓN DE LA VARIABLE
	funcionario durante cada detalle dentro de la vinculación.
BASICO	Valor del sueldo básico de acuerdo al cargo del funcionario durante cada detalle dentro de la vinculación.
CODTURNO	Código del turno o jornada laboral del funcionario durante cada detalle dentro de la vinculación.
CODSITUACIONLABORAL	Código de la situación laboral del funcionario durante cada detalle dentro de la vinculación.
SITUACIONLABORAL	Nombre de la situación laboral del funcionario durante cada detalle dentro de la vinculación.
BASICOENCARGO	Valor del básico del encargo.
CODSUCURSAL	Código de la sucursal a la cual pertenece el funcionario durante cada detalle dentro de la vinculación.
CODDEPENDENCIA	Código de la dependencia a la cual pertenece el funcionario durante cada detalle dentro de la vinculación.
CODVINCULACIONNOVEDAD	Código de la novedad durante cada detalle dentro de la vinculación.
VINCULACIONNOVEDAD	Nombre de la novedad durante cada detalle dentro de la vinculación.
CODAREA	Código del área de la entidad de la que depende el funcionario durante cada detalle dentro de la vinculación.
FUENTESDERECURSOSPARAPAGO	Para los docentes los recursos con los cuales se pagan pueden provenir de la nación, la gobernación, el municipio o de un fondo común.
CODCIUDADUBICACIONLABORAL	Código de la ciudad donde está ubicado el funcionario durante cada detalle dentro de la vinculación.
CIUDAD	Ciudad donde se realizara la actividad.
TIPOREG	Almacena un número (1, 2 ó 3) si el registro es histórico, actual o proyectado.
REFNUMVINCULACION	Describe la vinculación de los funcionarios que está realizando la referencia.
REALVINCULACIONDET	Detalle de vinculación.
REFVINCULACIONDET	Número real del detalle de vinculación.
CODAREAEDUCATIVA	Código del área educativa del funcionario durante cada situación laboral.
AREAEDUCATIVA	Área educativa del funcionario durante cada situación laboral.
CODNIVELEDUCACION	Código del nivel de educación del funcionario durante cada situación laboral.
NIVELEDUCACION	Nombre del nivel de educación del funcionario durante cada situación laboral.
CODNIVELDICTA	Código del nivel en el cual dicta cátedra el docente durante cada situación laboral.
NIVELDICTA	Nombre del nivel en el cual dicta cátedra el docente durante cada situación laboral.
CODVINCULACIONTIPO	Código del tipo de vinculación del funcionario durante cada detalle dentro de la vinculación.
REGIMEN_DE_VINCULACION	Nombre del régimen de vinculación.
HORASTRAB	Valor de horas trabajadas por los funcionarios al mes.

EXPRESIÓN O VARIABLE	DESCRIPCIÓN DE LA VARIABLE
CODAREAEDUCATIVECNICA	Código del área educativa técnica del funcionario durante cada situación laboral.
AREAEDUCATIVECNICA	Nombre del área educativa técnica del funcionario durante cada situación laboral.
GRADO	Establece el grado del funcionario.
GRADOENCARGO	Establece el grado de encargo.
OTRAAREAEDUCATIVECNICA	Se definen otras áreas educativas técnicas no comunes dentro de las descritas en otra variable.
FECHA_NACIMIENTO	Fecha de nacimiento del empleado.
SEXO	Almacena un carácter (F o M) dependiendo del sexo del empleado.
CONDICION_DE_AMENAZADO	Determina la condición del docente para ejercer su labor en el lugar de trabajo, teniendo en cuenta su seguridad.
ESCALAFON	El escalafón consiste en la lista de rangos en que se agrupan los docentes integrados a la Secretaría de Educación.

Instrumentos

Se han determinado dos instrumentos fundamentales para la recolección de la información:

- ✓ Las base de datos de la Secretaría de Educación de Medellín.
- ✓ Las encuestas: Si bien existen varios mecanismos para obtener información sobre los docentes, la encuesta más completa es la elaborada por la Secretaría de Educación de Medellín. Ésta llega a través de los correos de los núcleos a las instituciones y solicita al docente llenar toda su información (ver anexo). Es importante aclarar que este instrumento se usó como punto de partida en este trabajo, sin embargo, durante el proceso se determinó que no arroja resultados satisfactorios porque los docentes se abstienen de llenar la encuesta.

Herramientas

Se usan herramientas de Minería de Datos (DM) que consisten en la extracción no trivial de la información que reside de manera implícita en los datos, se desconoce y puede resultar útil para algún proceso. En otras palabras, las herramientas de minería de datos preparan, sondean y exploran los datos para extraer la información oculta. Se utilizará la herramienta Weka y herramientas de limpieza de datos como *DQ Analyzer* y *Microsoft Excel*.

DESARROLLO DEL PROCESO

Este capítulo describe el proceso de perfilamiento para la caracterización de los docentes en la Secretaría de Educación de Medellín a partir de la información existente en sus bases de datos. Por lo tanto, evidencia cómo se implementaron las herramientas de Inteligencia de negocios, se determinaron las variables y se analizaron los datos con el fin de hallar patrones para identificar posibles docentes teletrabajadores.

El proceso inicia con la consecución del instrumento, es decir, con la gestión para el uso de las bases de datos. La Secretaría de Educación de Medellín, a través del departamento de personal autoriza su entrega a la estudiante Mónica Patricia Escudero Cañas para el desarrollo de su trabajo de grado en maestría.

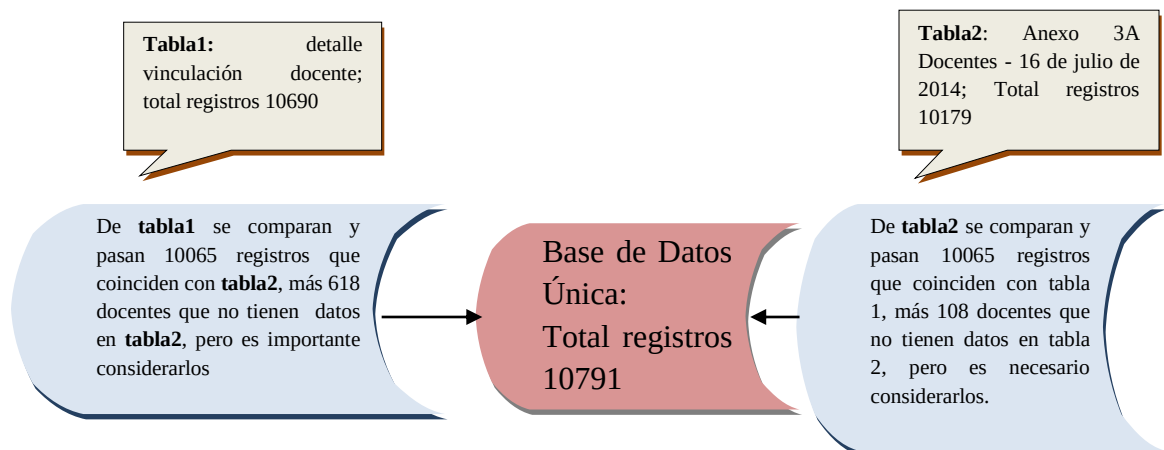
Una vez se obtienen las bases de datos, se verifica la información para su estudio, como se describe a continuación.

Capítulo I: Descripción de variables, análisis e identificación

Con la función **CONTAR.SI** se procede a hacer la verificación e integración de variables. De su análisis se concluye que existen dificultades de consistencia: En **Tabla1** hay 618 registros que no tienen información en **Tabla2** y, por otra parte, en **Tabla2** existen 108 registros sin que no tienen información en **Tabla1**.

Sin embargo, en este primer análisis, es necesario unir los registros en una sola Tabla de Datos debido a la relevancia de los mismos, arrojando un total de 10791 registros. Aunque surgen problemas de completitud porque algunos registros no coinciden entre las fuentes, se conservan para evitar que se deseche información importante. El proceso se grafica a continuación en la *figura 6*.

Figura 6: Proceso para unir las bases de datos



Todas las variables de la Base de Datos se analizan con la herramienta *DQ Analyzer* para buscar problemas de calidad en la información que no se detectan de manera manual y para limpiar los datos antes del proceso de minería al que se somete la información.

Capítulo II: Calidad de datos

En este capítulo se describe la identificación de variables importantes en el perfilamiento de los docentes y la consolidación de la información buscada mediante procesos de calidad de datos que permiten obtener información minable.

Primer análisis de perfil con herramienta *DQ analyzer*.

Se utilizó la herramienta de perfilamiento de datos *DQ Analyzer* para analizar el de los datos en diferentes etapas del proceso de depuración.

La herramienta, mediante la presentación visual de análisis de frecuencia, máscara, valores extremos y otros, arroja resultados que permiten desarrollar hipótesis y sacar conclusiones preliminares acerca del estado real de la fuente de datos a explotar. El estudio se realizó por medio de revisiones detalladas sobre los atributos disponibles en el reporte extraído del sistema transaccional.

A continuación se describe cada etapa del proceso desde la generación del reporte inicial hasta la consolidación del set de datos depurado que se usará en las tareas de minería de datos. Los atributos iniciales disponibles se presentan en la *Tabla 2*

Tabla 2: Perfil #1 de calidad de datos

Expression	Type	Nulls	Not nulls	Distinct	Unique	Min	Max	Mediana
CODEMPLEADO	INTEGER	0	10,791	10,787	10,783	129,921	1,216,715,181	43,252,246
FECHAINGRESONOMINA	STRING	108	10,683	1,646	1,073	01/13/2003 00:00:00	2014-12-08 00:00:00	2003-01-01 00:00:00
FECHAINGRESOEMPRESA	STRING	108	10,683	1,689	1,076	01/13/2003 00:00:00	2014-12-08 00:00:00	2003-01-01 00:00:00
CODVINCULACIONCATEGORIA	INTEGER	108	10,683	2	0	1	3	1
VINCULACIONCATEGORIA	STRING	108	10,683	2	0	Contratista	Planta	Planta
CODVINCULACIONESTADO	INTEGER	108	10,683	1	0	1	1	1
VINCULACIONESTADO	STRING	108	10,683	1	0	Activa	Activa	Activa
CODVACACIONESMETODO	INTEGER	108	10,683	3	1	0	2	2
VACACIONESMETODO	STRING	108	10,683	3	1	Docentes	No Aplica	Docentes
CODCALENDARIO	INTEGER	108	10,683	2	0	1	2	1
NOMBRAMIENTOCODACTOADMIN	INTEGER	108	10,683	3	0	2	6	2
NOMBRAMIENTOFECHA	STRING	109	10,682	1,339	794	01/13/1994 00:00:00	2014-12-03 00:00:00	08/19/2004 00:00:00
NOMBRAMIENTONUMERO	INTEGER	118	10,673	3,131	2,374	0	22,681	321
POSESIONCODACTOADMINISTRATIVO	INTEGER	108	10,683	5	2	1	7	1
POSESIONFECHA	STRING	110	10,681	1,379	803	01/13/1994 00:00:00	2014-12-08 00:00:00	08/23/2011 00:00:00
POSESIONNUMERO	INTEGER	118	10,673	3,402	2,673	0	22,681	312
CONTINUIDAD	STRING	108	10,683	2	0	N	S	N
RETIROCODACTOADMINISTRATIVO	INTEGER	709	10,082	2	0	0	3	0
RETIROFECHAACTOADMINISTRATIVO	STRING	10,766	25	16	10	05/26/2014 00:00:00	2014-11-08 00:00:00	07/28/2014 00:00:00
RETIRONUMEROACTOADMINISTRATIVO	STRING	10,764	27	27	27	111	9680	8898
CODCAUSARETIRO	STRING	10,753	38	5	2	1	7	1
CAUSARETIRO	STRING	10,753	38	5	2	Destitución por Sanción	Voluntario	Voluntario
FECHACREACION	STRING	108	10,683	2,176	1,647	01/13/2003 00:00:00	2014-12-08 16:27:17	2003-01-01 00:00:00
USUARIO	STRING	108	10,683	13	2	beatriz.lopez@medellin.gov.co	paula.correa@medellin.gov.co	Migración
DETORIGEN	STRING	10,791	0	0	0			
DETORIGEN2	STRING	10,791	0	0	0			
ORIGEN	STRING	10,200	591	1	0	1	1	1
RETIROCONFIRMADO	STRING	10,764	27	1	0	S	S	S
CODCONTINUIDADMOTIVO	STRING	10,768	23	2	0	1	2	1

Expression	Type	Nulls	Not nulls	Distinct	Unique	Min	Max	Mediana
CODEMPRESACONTINUIDAD	STRING	10,787	4	3	2	2	5631	5001
CODEMPRESATRASLADO	STRING	10,791	0	0	0			
CODESQUEMA	INTEGER	108	10,683	5	0	1	8	4
CODCARGOEMPRESA	INTEGER	108	10,683	16	2	219	9,032	9,001
CENTROCOSTO	STRING	108	10,683	297	35	ANEXO ESQUINA AL FRENTE	SEDE LA SUSANITA	INST EDUC JOSE ASUNCION SILVA
CODNIVELCONTRATACION	INTEGER	108	10,683	7	0	1	99	1
NIVELCONTRATACION	STRING	108	10,683	7	0	Emp Público Sin A.P	Provisional Vacante Temporal	Propiedad
CODDEDICACIONLABORAL	INTEGER	108	10,683	3	1	0	5	1
CODCENTROCOSTO	LONG	108	10,683	466	80	5,001	50,012,909,250,801	50,012,079,220,201
BASICO	INTEGER	108	10,683	57	4	616,000	4,009,527	1,844,811
CODTURNO	INTEGER	108	10,683	1	0	1	1	1
CODSITUACIONLABORAL	INTEGER	108	10,683	12	1	1	31	1
SITUACIONLABORAL	STRING	108	10,683	12	1	Comisión	Suspensión	Normal
BASICOENCARGO	INTEGER	1,985	8,806	19	6	0	4,009,527	0
CODSUCURSAL	INTEGER	108	10,683	2	0	1	2	2
CODDEPENDENCIA	LONG	108	10,683	475	87	5,001	50,012,909,250,801	50,012,079,220,105
CODVINCULACIONNOVEDAD	INTEGER	108	10,683	22	4	1	96	2
VINCULACIONNOVEDAD	STRING	108	10,683	22	4	Aclaratoria	Traslados	Cambios de Sueldo
CODAREA	INTEGER	108	10,683	2	0	1	3	1
FUENTESDERECURSOSPARAPAGO	STRING	108	10,683	2	0	Donaciones	Sistema General Participación	Sistema General Participación
CODCIUDADUBICACIONLABORAL	INTEGER	108	10,683	1	0	5,001	5,001	5,001
CIUDAD	STRING	108	10,683	1	0	Medellín (Ant)	Medellín (Ant)	Medellín (Ant)
TIPOREG	INTEGER	108	10,683	1	0	2	2	2
REFNUMVINCULACION	INTEGER	10,754	37	36	35	3,518,188	1,077,420,717	43,086,105
REALVINCULACIONDET	INTEGER	108	10,683	35	4	1	43	1
REFVINCULACIONDET	STRING	10,791	0	0	0			
CODAREAEDUCATIVA	INTEGER	527	10,264	20	1	1	22	4
AREAEDUCATIVA	STRING	527	10,264	20	1	Áreas de Apoyo para Educación Especial	Tecnología de Informática	No Aplica
CODNIVELEDUCACION	INTEGER	375	10,416	20	2	0	26	11

Expression	Type	Nulls	Not nulls	Distinct	Unique	Min	Max	Mediana
NIVELEDUCACION	STRING	375	10,416	20	2	Bachiller	Tecnólogo	Licenciado
CODNIVELDICTA	INTEGER	608	10,183	6	0	1	7	3
NIVELDICTA	STRING	608	10,183	6	0	Ciclo complementario normal	Sin Asignación Directa	Primaria
CODVINCULACIONTIPO	INTEGER	108	10,683	7	0	0	9	2
REGIMENDEVINCULACION	STRING	108	10,683	6	0	Cofinanciado Departamental	Nacionalizados	Nacionalizados
HORASTRAB	INTEGER	108	10,683	2	1	15	240	240
CODAREAEDUCATIVECINICA	INTEGER	242	10,549	11	2	1	15	11
AREAEDUCATIVECINICA	STRING	242	10,549	11	2	Agropecuaria(NA)	Salud	No aplica
GRADO	STRING	108	10,683	63	6	1	BC	2ª
GRADOENCARGO	STRING	10,678	113	18	6	10	3DM	14
OTRAAREAEDUCATIVECINICA	STRING	10,643	148	52	38	AGRICOLA	TURISMO	Tecnología informática
FECHANACIMIENTO	DATETIME	601	10,190	7,056	4,688	1900-12-25 00:00:00.000	1995-03-14 00:00:00.000	1968-02-14 00:00:00.000
SEXO	STRING	619	10,172	2	0	F	M	F
CONDICIONDEAMENAZADO	STRING	619	10,172	2	0	N	S	N
ESCALAFON	STRING	619	10,172	34	0	1	BC	2ª

Luego de generar el perfil de calidad sobre el set de datos original se determina que los procedimientos que se aplicaron, a su vez, generaron modificaciones sobre los campos presentes en el reporte. A través de cada etapa en la preparación de datos, se presentan tablas similares a la anterior con el fin de describir la inclusión de nuevos atributos, así como la eliminación o modificación de algunos ya presentes. Las transformaciones culminan al consolidar la vista minable que contiene los campos seleccionados al final para ser modelados.

Como se mencionó en párrafos anteriores, el análisis se realiza con un total de 10791 registros que se clasifican de acuerdo con el tipo de variable así: entero 29, tipo cadena de caracteres 38, 1 dato tiempo, 1 longitud.

Decisiones tomadas sobre la calidad de datos.

Se toman las siguientes decisiones:

- A. Eliminar de la base de datos las variables que no son necesarias para el objeto de estudio y 108 registros con un alto porcentaje de nulos.
- B. Imputar datos de registros con valor nulo en la base de datos.

De acuerdo con estas decisiones, se realizan las siguientes acciones:

- Eliminar las variables que representan códigos, es decir, se van a utilizar los nombres para una mayor comprensión por parte del usuario ya que representan la misma información para el objeto estudio. Las variables son: CODVINCULACIONESTADO, CODVINCULACIONCATEGORIA, CODNIVELCONTRATACION,

CODSITUACIONLABORAL, CODVINCULACIONNOVEDAD,
 CODVACACIONESMETODO, CODCIUDADUBICACIONLABORAL,
 CODAREAEDUCATIVA, CODNIVELEDUCACION, CODNIVELDICTA,
 CODAREAEDUCATIVATECNICA, CODVINCULACIONTIPO, CODCAUSARETIRO.

- Eliminar las variables que no contienen datos en la tabla: DETORIGEN, DETORIGEN2, CODEMPRESATRASLADO.

- Eliminar 108 registros que contienen solo el 5.1% de la información y la posibilidad de que generen “ruido” en la tabla, las variables que se afectan de forma directa y que en el análisis con la herramienta representan el 1% de nulos son:

- FECHAINGRESONOMINA, FECHAINGRESOEMPRESA, VINCULACIONCATEGORIA, VINCULACIONESTADO, VACACIONESMETODO, CODCALENDARIO, NOMBRAMIENTOCODACTOADMIN, NOMBRAMIENTONUMERO, CODAREA, POSESIONCODACTOADMINISTRATIVO, NOMBRAMIENTOFECHA, POSESIONFECHA, POSESIONNUMERO, CONTINUIDAD, FECHACREACION, USUARIO, CODESQUEMA, CODTURNO, CODCARGOEMPRESA, CENTRO COSTO, NIVELCONTRATACION, CODDEDICACIONLABORAL, BASICO, CODTURNO, CODSUCURSAL, SITUACIONLABORAL, VINCULACIONNOVEDAD, REJVINCULACIONDET, FUENTESDERECURSOSPARAPAGO, CIUDAD, TIPOREG, CODVINCULACIONTIPO, REGIMENDEVINCULACION, HORASTRAB, GRADO, CODDEPENDENCIA.

- Eliminar la variable ESCALAFON ya que posee la misma información que la variable GRADO y esta última variable está en la tabla de nómina que se abordará en este estudio.

- Eliminar la variable FECHAINGRESOEMPRESA ya que contienen en un 95% la misma información FECHAINGRESONOMINA, solo varía en los momentos en los cuales

el docente tarda un tiempo en posesionarse en el cargo, momento donde hace su ingreso real a la nómina.

- Eliminar las variables que no ofrecen información importante para la presente investigación: NOMBRAMIENTOCODACTOADMIN, NOMBRAMIENTOFecha, NOMBRAMIENTONUMERO, POSESIONCODACTOADMINISTRATIVO, POSESIONFECHA, POSESIONNUMERO, CONTINUIDAD, RETIROCODACTOADMINISTRATIVO, RETIROFECHAACTOADMINISTRATIVO, RETIRONUMEROACTOADMINISTRATIVO, CAUSARETIRO, FECHACREACION, USUARIO, ORIGEN, CODCONTINUIDADMOTIVO, CODEMPRESACONTINUIDAD, CODESQUEMA, CODCENTROCOSTO, CODSUCURSAL, CODDEPENDENCIA, CODAREA, TIPOREG, REFNUMVINCULACION, REALVINCULACIONDET, REFVINCULACIONDET.

Luego de hacer modificaciones de calidad a la Base de Datos, el set de datos se somete a un segundo perfil de calidad para evaluar su estado como se observa en la *Tabla 3*.

Segundo análisis de perfil con herramienta DQ Analyzer.

Tabla 3: Perfil #2 de calidad de datos

Expression	Type	Nulls	Not nulls	Distinct	Unique	Min	Max	Median
CODEMPLEADO	INTEGER	0	10,683	10,679	10,675	129,921	1,216,715,181	43,258,606
FECHAINGRESONOMINA	STRING	0	10,683	1,646	1,073	01/13/2003 00:00:00	2014-12-08 00:00:00	2003-01-01 00:00:00
VINCULACIONCATEGORIA	STRING	0	10,683	2	0	Contratista	Planta	Planta
VINCULACIONESTADO	STRING	0	10,683	1	0	Activa	Activa	Activa
VACACIONESMETODO	STRING	0	10,683	3	1	Docentes	No Aplica	Docentes
CODCALENDARIO	INTEGER	0	10,683	2	0	1	2	1
RETIROCONFIRMADO	STRING	10,656	27	1	0	S	S	S
CODCARGOEMPRESA	INTEGER	0	10,683	16	2	219	9,032	9,001
CENTROCOSTO	STRING	0	10,683	297	35	ANEXO ESQUINA AL FRENTE	SEDE LA SUSANITA	INST EDUC JOSE ASUNCION SILVA
NIVELCONTRATACION	STRING	0	10,683	7	0	Emp Público Sin A.P	Provisional Vacante Temporal	Propiedad
CODDEDICACIONLABORAL	INTEGER	0	10,683	3	1	0	5	1
BASICO	INTEGER	0	10,683	57	4	616,000	4,009,527	1,844,811
CODTURNO	INTEGER	0	10,683	1	0	1	1	1
SITUACIONLABORAL	STRING	0	10,683	12	1	Comisión	Suspensión	Normal
BASICOENCARGO	INTEGER	1,877	8,806	19	6	0	4,009,527	0
VINCULACIONNOVEDAD	STRING	0	10,683	22	4	Aclaratoria	Traslados	Cambios de Sueldo
AREAFUENTESDERECURSOSPARAPAGO	STRING	0	10,683	2	0	Donaciones	Sistema General Participación	Sistema General Participación
CIUDAD	STRING	0	10,683	1	0	Medellín (Ant)	Medellín (Ant)	Medellín (Ant)
AREAEDUCATIVA	STRING	419	10,264	20	1	Áreas de Apoyo para Educación Especial	Tecnología de Informática	No Aplica
NIVELEDUCACION	STRING	267	10,416	20	2	Bachiller	Tecnólogo	Licenciado
NIVELDICTA	STRING	500	10,183	6	0	Ciclo complementario normal	Sin Asignación Directa	Primaria
REGIMENDEVINCULACION	STRING	0	10,683	6	0	Cofinanciado Departamental	Nacionalizados	Nacionalizados

Expression	Type	Nulls	Not nulls	Distinct	Unique	Min	Max	Median
HORASTRAB	INTEGER	0	10,683	2	1	15	240	240
AREAEDUCATIVECNICA	STRING	134	10,549	11	2	Agropecuaria(NA)	Salud	No aplica
GRADO	STRING	0	10,683	63	6	1	BC	2A
GRADOENCARGO	STRING	10,570	113	18	6	10	3DM	14
OTRAAREAEDUCATIVECNICA	STRING	10,535	148	52	38	AGRICOLA	TURISMO	Tecnología informática
FECHANACIMIENTO	DATETIME	601	10,082	6,977	4,627	1900-12-25 00:00:00.000	1995-03-14 00:00:00.000	1968-03-04 00:00:00.000
SEXO	STRING	619	10,064	2	0	F	M	F
CONDICIONDEAMENAZADO	STRING	619	10,064	2	0	N	S	N

Nuevos hallazgos y decisiones tomadas.

- Hay valores nulos en las variables RETIROCONFIRMADO, BASICOENCARGO, AREAEDUCATIVA, NIVELEDUCACION, NIVELDICTA, AREAEDUCATIVATECNICA, GRADOENCARGO, OTRAAREAEDUCATIVATECNICA, FECHANACIMIENTO, SEXO, CONDICIONDEAMENAZADO.
- Hay una inconsistencia en la variable CODEMPLEADO en 4 registros, si tenemos en cuenta que una clave primaria debe ser única.
- Hay cambios en los valores para la Mediana en las variables: CODEMPLEADO (43252246 por 43258606), FECHANACIMIENTO (1968-02-14 00:00:00.000 por 1968-03-04 00:00:00.000) para el máximo y en la mediana no se observa alteraciones.
- VINCULACIONESTADO tiene el 0.01% d, es decir 1 dato no único, es necesario verificarlo.
- VACACIONESMETODO tiene un valor con el atributo *No aplica*; es necesario verificarlo.
- Se toma la decisión de imputar valores en las variables que se describen a continuación y reemplazarlos por ceros u otro valor de acuerdo a la necesidad ya que no se trata de una información nula, sino de una información que no tiene aplicación para el docente en esa variable: RETIROCONFIRMADO, BASICOENCARGO,

AREAEDUCATIVA, NIVELEDUCACION, NIVELDICTA,
 AREAEDUCATIVECNICA, GRADOENCARGO,
 OTRAAREAEDUCATIVECNICA, FECHA_NACIMIENTO, SEXO,
 CONDICION_DE_AMENAZADO.

Imputación de Datos.

- RETIROCONFIRMADO: En esta variable no se debe reemplazar por la mediana ya que de hacerlo se cambiaría el sentido de la información, por ejemplo: si a un docente adicionamos S (que es la mediana), confirmaríamos que se retiró del cargo. Es recomendable un valor como *No aplica* o un valor parecido que sustente la información.
- BASICOENCARGO: Para esta variable la mediana es Cero (0), es necesario imputar los datos con otra variable diferente a cero ya que puede indicar que el docente desempeña unas funciones por encargo, pero que su salario no sufre modificación, lo que puede generar cambios significativos en la información. Se sugiere *No aplica*
- AREAEDUCATIVA: Esta variable puede ser reemplazada por valores como *No asignada*, ya que si se reemplaza por Cero (0) no es coherente con el tipo de valor que ésta contiene o es posible que la mediana (*No aplica*) sea de gran ayuda para este caso.
- NIVELEDUCACION: Para la imputación de los datos es necesario descartar la media por tratarse de una información que mide el nivel de educación de un docente y se da por hecho que este dato cambia el sentido de la información. Así mismo, existe un dato (sin

definir) que indica que no se le asignó nivel educativo al docente. Es necesario verificar si se utiliza éste u otro que no altere la información; puede ser *Sin información*.

- **NIVELDICTA:** Para la imputación de los datos es necesario descartar la media por tratarse de una información que mide el nivel de educación de un docente y se da por hecho que este dato cambia el sentido de la información. Así mismo, existe un dato (Sin asignación directa) que indica que no se le asignó el nivel en el que dicta clase. Es necesario verificar si se utiliza este u otro que no altere la información; puede ser *Sin información* u *Otro*.

- **AREAEDUCATIVECNICA:** Para la imputación de los datos nulos es necesario considerar la media (*No aplica*) ya que corresponde a una opción viable si se compara con la información suministrada y corresponde a una lógica de esta variable que tiene relación con los docentes que solo tienen una asignación para la media.

- **GRADOENCARGO:** Para la imputación no se recomienda usar la media catorce (14), ya que distorsiona la información y clasifica a los docentes en la categoría 14. Se debe aclarar que esta clasificación es temporal y no corresponde al grado definitivo que tiene el docente. Se puede recomendar usar un valor como *No aplica*.

- **OTRAAREAEDUCATIVECNICA:** la imputación de los datos nulos deben estar hechos para evidenciar que no todos los docentes pertenecen a *OTRAAREATECNICA*, sino unos cuantos que no están en la categoría área técnica y que por necesidad de la Secretaría de Educación es necesario clasificarlos. Puede ser un dato como *No aplica*.

- FECHANACIMIENTO: La imputación de datos en esta variable debe estar dirigida a los valores nulos y a eliminar de la variable una información basura que fue introducida en el cruce de variables y que es evidenciable de forma clara. Se sugiere que sea un dato que evidencia que no hay registros para ese docente, como *No registra*; no se recomienda que sea la mediana.
- SEXO: Se sugiere que la imputación de datos evidencie que no hay registro sobre este ya que para el objeto de estudio es importante resaltar el hecho de ser de género femenino y es importante no alterar de forma significativa la información. Se recomienda el valor *No registra*, *No hay dato*, entre otros.
- CONDICIONDEAMENAZADO: La imputación de datos en esta variable debe ser reemplazada por la mediana ya que corresponde a la lógica de la información y no alteraría su sentido.

Luego del análisis, se concluye que se debe desarrollar un tercer perfil de calidad para evaluar el estado de los datos, e identificar las variables que contienen valores faltantes. Se hace una selección de las variables, se identifica a qué tipo corresponden, el porcentaje de nulos, el patrón de ausencia (MCAR, MNAR, MAR), la técnica recomendada, si es necesario imputar y algunas observaciones.

En la *Tabla 4* se exponen los análisis realizados en los párrafos anteriores con respecto a las imputaciones de variables con datos nulos:

Tabla 4: Análisis para imputaciones de variables con datos nulos

VARIABLE	TIPO DE VARIABLE	% EN NULOS	PATRÓN DE AUSENCIA (PA)	TÉCNICA DE IMPUTACIÓN RECOMENDADA	OBSERVACIONES
RETIROCONFIRMADO	STRING	99,75	MNAR	No se recomienda imputar	Variable y no se recomienda usarla por el porcentaje alto de nulos que contiene. Se concentra solo en una parte de la población
BASICOENCARGO	INTEGER	17,57	MNAR	N/A	Se identifica que la variable contiene la misma información, contenida en el BASICO y se decide no incluirla en el grupo de variables necesarias para el análisis del trabajo en estudio.
AREAEDUCATIVA	STRING	3,92	MAR	Hot Deck	No es posible utilizar la media, la mediana para imputar los datos porque son variables categóricas. La mediana no afecta los valores extremos
NIVELEDUCACION	STRING	2,50	MAR	Hot Deck	La mediana no afecta los valores extremos
NIVELICTA	STRING	4,68	MAR	Moda	La mediana no afecta los valores extremos
AREAEDUCATIVECATECNICA	STRING	1,25	MAR	Hot Deck	La mediana no afecta los valores extremos
GRADOENCARGO	STRING	98,94	MNAR	No se recomienda imputar.	Variable y no se recomienda usarla por el porcentaje alto de nulos que contiene.
OTRAAREAEDUCATIVECATECNICA	STRING	98,61	MNAR	No se recomienda imputar	Variable y no se recomienda usarla por el porcentaje alto de nulos que contiene.
FECHANACIMIENTO	DATETIME	5,63	MAR	Hot Deck	
SEXO	STRING	5,79	MAR	Moda	Es el valor que más se repite y se pueden calcular variables que no son numéricas.
CONDICIONDEAMENAZADO	STRING	5,79	MAR	Mediana o moda	La mediana no afecta los valores extremos

Luego de un análisis de relaciones entre las variables y de acuerdo con la metodología sugerida en el diagrama para la selección de técnicas para *Valores Faltantes*, sugerida en Amón Uribe (2010), se toman las siguientes decisiones:

- Se realiza una nueva verificación de la cantidad de valores faltantes con la información de la tabla de nómina original para verificar y evidenciar un posible error inicial. Sin embargo, se logra evidenciar que los valores faltantes no sufren cambios diferentes a los que ya se presentaron.
- No realizar imputación en las variables RETIROCONFIRMADO, GRADOENCARGO, OTRAAREAEDUCATIVECNICA por su alto porcentaje de valores faltantes, declarado por la herramienta. Se toma la decisión de retirarlos del objeto de estudio.
- En la variable SEXO es conveniente imputar los datos dado que es una variable con un porcentaje de valores faltantes equivalentes al 5.79%. Para variables categóricas “como género o estado civil, se recomienda la imputación usando la moda.” (Amón Uribe, 2010)
- En la variable CONDICIONDEAMENAZADO es conveniente imputar los datos porque es una variable con un porcentaje de valores faltantes equivalentes al 5.79%.
- Se debe eliminar valores artificios que fueron utilizados en la unión de las dos tablas en la variable FECHANACIMIENTO y verificar de nuevo nivel de calidad con la herramienta. De igual manera esta pasa de tener un 5.63% a un 5.79% que no representa

una cifra significativa de acuerdo al objeto de estudio, ni cambia de forma sustancial los límites de manejo de valores faltantes.

- Se elimina la variable BASICOENCARGO del grupo preseleccionado para el estudio por contener la misma información de la variable BÁSICO.
- Se descarta el patrón MCAR en todas las variables ya que los valores nulos no se encuentran distribuidos de forma aleatoria en la base de datos.
- La relación de las variables CODCARGOEMPRESA con NIVELDICTA, AREAEDUCATIVECNICA, AREAEDUCATIVA, NIVELEDUCACION se hace evidente al identificar que los valores nulos tienen una concentración de acuerdo al código del CARGOEMPRESA.
- Se logra identificar que los nulos están concentrados en los CÓDIGOS DE CARGO EN LA EMPRESA, como se indica en la Tabla 5 y 6:

Tabla 5: Relación porcentual de CODCARGOEMPRESA con Nulos en las variables NIVELDICTA, AREAEDUCATIVECNICA, AREAEDUCATIVA, NIVELEDUCACION

RELACIÓN CODCARGOEMPRESA CON NULOS EN VARIABLES: NIVELDICTA, AREAEDUCATIVECNICA, AREAEDUCATIVA, NIVELEDUCACION, MEDIDOS EN %											
AREAEDUCATIVA CANTIDAD TOTAL NULOS=419			NIVELEDUCACION CANTIDAD TOTAL NULOS=267			NIVELDICTA CANTIDAD TOTAL NULOS=500			AREAEDUCATIVECNICA CANTIDAD TOTAL NULOS=134		
CODCARGOEMPRESA	Cantidad Nulos	%	CODCARGOEMPRESA	Cantidad Nulos	%	CODCARGOEMPRESA	Cantidad de Nulos	%	CODCARGOEMPRESA	Cantidad Nulos	%
219	8	1,9	219	7	2,6	219	9	1,8	219	4	2,9
314	7	1,6	314	6	2,2	314	11	2,2	314	1	0,74
407	223	53	407	127	47,5	407	263	52,6	407	80	5,9
440	239	9	440	19	7,1	440	36	7,2	440	14	10,4
470	83	19	470	58	21,7	470	94	18,8	470	14	10,4
477	72	17	477	47	17,6	477	83	16,6	477	14	10,4
480	3	0.7	480	3	1,12	480	3	0,6	480	3	2,2
						425	1	0,2	902	3	2,2
									9032	1	0,7
Total	419	100%	Total	267	100	Total	500	100%	Total	134	100%

Tabla 6: Relación CODCARGOEMPRESA con nulos en variables FECHANACIMIENTO, SEXO, CONDICIONDEAMENAZADO

RELACIÓN CODCARGOEMPRESA CON NULOS EN VARIABLES: FECHANACIMIENTO, SEXO, CONDICIONDEAMENAZADO, MEDIDOS EN %								
FECHANACIMIENTO CANTIDAD TOTAL NULOS=619			SEXO CANTIDAD TOTAL NULOS=619			CONDICIONDEAMENAZADO CANTIDAD TOTAL NULOS=619		
CODCARGOEMPRESA	Cantidad Nulos	%	CODCARGOEMPRESA	Cantidad Nulos	%	CODCARGOEMPRESA	Cantidad Nulos	%
219	9	1,45	219	9	1,45	219	9	1,45
314	11	1,77	314	11	1,77	314	11	1,77
407	263	42,4	407	263	42,4	407	263	42,4
440	1	0,16	440	1	0,16	440	1	0,16
470	36	5,8	470	36	5,8	470	36	5,8
477	94	15,18	477	94	15,18	477	94	15,18
480	83	13,4	480	83	13,4	480	83	13,4
425	3	0,48	425	3	0,48	425	3	0,48
9001	119	19,2	9001	119	19,2	9001	119	19,2
Total	619	100%	Total	619	100%	Total	619	100%

Se observa que en las variables analizadas, la mayor cantidad de nulos están concentrados en el código 407 del CODCARGOEMPRESA.

Grupos de datos a imputar de acuerdo con Variable AREAEDUCATIVA.

- Se imputa dato en variable AREAEDUCATIVA, por medio de técnica *Hot Deck*.

Se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESONOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

21404806	01/01/2003 0:00	Planta	Activa	Estándar 1	314	INST EDUC INEM JOSE FELIX DE
RESTREPO	Propiedad	1	2458869 1	Normal	Cambios de Sueldo Sistema	General
Participación	Medellín (Ant)	No Aplica		Nacional	240	No aplica 1T3

- Se imputa dato en variable AREAEDUCATIVA, por medio de técnica *Hot Deck*.

Se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESONOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

32015164	01/01/2003 0:00	Planta	Activa	Estándar 1	219	INST EDUC ESCUELA NORMAL
SUPERIOR DE MEDELLIN	Emp Público Sin A.P		1	3096165 1	Normal	Cambios de Sueldo
Sistema General Participación	Medellín (Ant)		Áreas de Apoyo para Educación Especial			Posgrado en
Educación	Departamental	240	No aplica 1P6			

- Se imputa dato en variable AREAEDUCATIVA, por medio de técnica *Hot Deck*.

Se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes

que comparten características, como CODCARGOEMPRESA, FECHAINGRESONOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

3348423	05/23/2011 00:00:00	Planta	Activa	Estándar 1	407	II.E. MARCO FIDEL SUAREZ -
SEDE PRINCIPAL	Propiedad	1	1515515 1	Normal	Cambios de Sueldo Sistema General	Participación Medellín (Ant) Áreas de Apoyo para Educación Especial
					Posgrado en Educación	Nacional
	240	No aplica 2A1				

- Se imputa dato en variable AREAEDUCATIVA, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESONOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

22200937	07/05/2011 0:00	Planta	Activa	Estándar 1	440	INST EDUC PEDRO LUIS VILLA
Propiedad	1	1515515 1	Normal	Cambios de Sueldo Sistema General	Participación	Medellín (Ant) Áreas de Apoyo para Educación Especial
					Bachiller Académico	Nacional
	240	No aplica 2A2				

- Se imputa dato en variable AREAEDUCATIVA, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESONOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

21871008	03/23/2003 00:00:00	Planta	Activa	Estándar 1	470	INST EDUC CRISTOBAL COLON
Propiedad	1	1358243 1	Normal	Cambios de Sueldo Sistema General	Participación	Medellín (Ant) No Aplica
				Nacional	240	No aplica 1A3

- Se imputa dato en variable AREAEDUCATIVA, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESONOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

3622303	06/13/2011 00:00:00	Planta	Activa	Estándar 1	477	INST EDUC OCTAVIO HARRY
Propiedad	1	1255371	1	Normal	Cambios de Sueldo Sistema	General Participación
Medellín (Ant)	Áreas de Apoyo para Educación Especial				Bachiller Académico	
Nacional	240	No aplica 1A2				

- Se imputa dato en variable AREAEDUCATIVA, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESONOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

3352337	01/01/2003 0:00	Planta	Activa	Estándar 1	480	INST EDUC ESCUELA NORMAL SUPERIOR
DE MEDELLIN	Propiedad	1	1515515	1	Normal	Cambios de Sueldo Sistema General
Participación	Medellín (Ant)	Áreas de Apoyo para Educación Especial				Nacional
240	2A3					

Grupos de datos a imputar de acuerdo a Variable NIVELEDUCACION.

- Se imputa dato en variable NIVELEDUCACION, por medio de técnica *Hot Deck*. Se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA,

FECHAINGRESOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

32015164	01/01/2003 0:00	Planta	Activa	Estándar	1	219	INST EDUC ESCUELA NORMAL
SUPERIOR DE MEDELLIN	Emp Público Sin A.P			1	3096165	1	Normal Cambios de Sueldo
	Sistema General Participación		Medellín (Ant)				Áreas de Apoyo para Educación Especial
	Posgrado en Educación		Departamental		240		No aplica 1P6

- Se imputa dato en variable NIVELEDUCACION, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

3716049	06/01/2012 0:00	Planta	Activa	Estándar	1	314	INST EDUC INEM JOSE FELIX DE
RESTREPO	Propiedad	1	2013621	1	Normal	Cambios de Sueldo Sistema General	
	Participación Medellín (Ant)					Bachiller	Nacional 240 No
aplica	1T1						

- Se imputa dato en variable NIVELEDUCACION, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

21495776	06/01/2011 0:00	Planta	Activa	Estándar	1	407	INST EDUC REPUBLICA DE
URUGUAY	Propiedad	1	1515515	1	Normal	Cambios de Sueldo Sistema General	
	Participación Medellín (Ant)					Bachiller Académico	
	Nacional 240					No aplica 2A1	

- Se imputa dato en variable NIVELEDUCACION, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESONOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

43003397	07/05/2011 0:00	Planta	Activa	Estándar	1	440	INST EDUC HORACIO MUÑOZ
SUESCUN	Propiedad	1	1515515	1	Normal	Cambios de Sueldo Sistema General	
Participación Medellín (Ant)	Áreas de Apoyo para Educación Especial					Bachiller Académico	
Nacional	240	No aplica 2A2					

- Se imputa dato en variable NIVELEDUCACION, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESONOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

32524023	06/01/2011 0:00	Planta	Activa	Estándar	1	470	INST EDUC ASAMBLEA
DEPARTAMENTAL	Propiedad	1	1255371	1	Normal	Cambios de Sueldo Sistema General	
Participación Medellín (Ant)	Áreas de Apoyo para Educación Especial					Bachiller Académico	
Nacional	240	No aplica 1A1					

- Se imputa dato en variable NIVELEDUCACION, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESONOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

8351651	06/13/2011 00:00:00	Planta	Activa	Estándar	1	477	INST EDUC LA MILAGROSA
---------	---------------------	--------	--------	----------	---	-----	------------------------

Propiedad	1	1255371	1	Normal	Cambios de Sueldo Sistema	General	Participación
Medellín (Ant)	Áreas de Apoyo para Educación Especial				Bachiller Académico		
Nacional	240	No aplica	1A2				

- Se imputa dato en variable NIVELEDUCACION, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESONOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

8150524	01/01/2003 0:00	Planta	Activa	Estándar	1	440	INST EDUC	JUAN DE LA CRUZ POSADA
Propiedad	1	1758304	1	Normal	Cambios de Sueldo Sistema	General	Participación	
Medellín (Ant)	Áreas de Apoyo para Educación Especial				Bachiller Académico	Departamental	240	No aplica
	2AB							

Nota: en esta variable no hay grupo de datos de acuerdo al parámetro que se viene desarrollando, se toma la decisión de imputar eligiendo el código 440, ya que hay datos que coinciden en otras variables del grupo

Grupos de datos a imputar de acuerdo a Variable AREAEDUCATIVECENICA.

- Se imputa dato en variable AREAEDUCATIVECENICA, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESONOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

32015164	01/01/2003 0:00	Planta	Activa	Estándar	1	219,	INST EDUC	ESCUELA NORMAL
SUPERIOR DE MEDELLIN	Emp Público Sin A.P			1	3096165	1	Normal	Cambios de Sueldo
	Sistema General Participación			Medellín (Ant)	Áreas de Apoyo para Educación Especial			

Posgrado en Educación	Departamental	240	No aplica 1P6
-----------------------	---------------	-----	---------------

- Se imputa dato en variable AREAEDUCATIVECNICA, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESONOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

70320719	01/01/2003 0:00	Planta	Activa	Estándar 1	314	INST EDUC CASD JOSE MARIA
ESPINOSA PRIETO	Propiedad	1	2458869 1	Normal	Cambios de Sueldo Sistema General	
Participación Medellín (Ant)	No Aplica	Técnico o tecnólogo en Educación			Nacional 240	No aplica
	1T3					

- Se imputa dato en variable AREAEDUCATIVECNICA, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESONOMINA y VINCULACIONNOVEDAD ; las otras variables coinciden en sus valores con los datos, así:

3348423	05/23/2011 00:00:00	Planta	Activa	Estándar 1	407	II.E. MARCO FIDEL SUAREZ -
SEDE PRINCIPAL	Propiedad	1	1515515 1	Normal	Cambios de Sueldo Sistema General	
Participación Medellín (Ant)	Áreas de Apoyo para Educación Especial	Posgrado en Educación			Nacional	240
		No aplica 2A1				

- Se imputa dato en variable AREAEDUCATIVECNICA, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA₂.

FECHAINGRESOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

22200937	07/05/2011 0:00	Planta	Activa	Estándar 1	440	INST EDUC PEDRO LUIS VILLA
Propiedad	1	1515515	1	Normal	Cambios de Sueldo Sistema General Participación	
Medellín (Ant)	Áreas de Apoyo para Educación Especial				Bachiller Académico	
Nacional	240	No aplica 2A2				

- Se imputa dato en variable AREAEDUCATIVECINICA, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

32524023	06/01/2011 0:00	Planta	Activa	Estándar 1	470	INST EDUC ASAMBLEA
DEPARTAMENTAL Propiedad	1	1255371	1	Normal	Cambios de Sueldo Sistema General Participación	
Medellín (Ant)	Áreas de Apoyo para Educación Especial				Bachiller Académico	
Nacional	240	No aplica 1A1				

- Se imputa dato en variable AREAEDUCATIVECINICA, por medio de *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

8351651	06/13/2011 00:00:00	Planta	Activa	Estándar 1	477	INST EDUC LA MILAGROSA
Propiedad	1	1255371	1	Normal	Cambios de Sueldo Sistema General Participación	
Medellín (Ant)	Áreas de Apoyo para Educación Especial				Bachiller Académico	
Nacional	240	No aplica 1A2				

- Se imputa dato en variable AREAEDUCATIVECNICA, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESONOMINA y VINCULACIONNOVEDAD; las otras variables coinciden en sus valores con los datos, así:

3362714	01/01/2003 0:00	Planta	Activa	Docentes 1	902	SECRETARÍA DE EDUCACION DE			
MEDELLIN	Propiedad	1	2711939 1	Normal	Cambios de Sueldo	Sistema General Participación			
	Medellín (Ant)	No Aplica	Licenciado	Sin Asignación Directa		Nacionalizados	240	No	
aplica	14	25/11/1956	M	N					

- Se imputa dato en variable AREAEDUCATIVECNICA, por medio de técnica *Hot Deck*, se hace la elección de grupos de datos así: se tiene en cuenta datos de variables donantes que comparten características, como CODCARGOEMPRESA, FECHAINGRESONOMINA y VINCULACIONNOVEDAD ; las otras variables coinciden en sus valores con los datos, así:

15362962	03/12/2003 0:00	Planta	Activa	Docentes 1	9032	INST EDUC LA ESPERANZA			
	Propiedad	1	2711939 1	Normal	Continuidad	Sistema General Participación			
	Medellín (Ant)	No Aplica	Licenciado	Sin Asignación Directa		Nacionalizados	240	No	
aplica	14	18/01/1954	M	N					

Eliminación de datos duplicados.

Se identifican los siguientes registros duplicados:

39169650	02/01/2014 0:00	Planta	Activa	Docentes 1	9001	INST	EDUC	JOSE	HORACIO
BETANCUR	Provisional Vacante Definitiva			1	1411890 1	Normal	Cambios	de	Sueldo
	Sistema General Participación		Medellín (Ant)	Primaria	Licenciado		Primaria	Nacional	
240	No aplica 2A	20/08/1965	F	N					

39170358	03/18/2012 00:00:00	Planta	Activa	Docentes 1	9001	INST EDUC RAMON
GIRALDO CEBALLOS	Provisional Vacante Definitiva			1	1411890 1	Normal Cambios
de Sueldo	Sistema General Participación		Medellín (Ant)		Ciencias Sociales	Profesional
Secundaria	Nacionalizados	240	No aplica 2A	26/01/1961	F	N

43623866	10/10/2002 0:00	Planta	Activa	Estándar 1	407	INST EDUC TULIO OSPINA
Propiedad	1	1682814 1		Normal	Cambios de Sueldo Sistema	General
Participación	Medellín (Ant)	Áreas de Apoyo para Educación Especial			Bachiller	Académico
Secundaria	Departamental	240	No aplica 2AA		F	N

88144665	01/09/2007 0:00	Planta	Activa	Docentes 1	9001	INST EDUC JOSE ACEVEDO Y
GOMEZ	Propiedad	1	2797931 1	Normal	Cambios de Sueldo Sistema	General
Participación	Medellín (Ant)	Tecnología de Informática			Postgrado Magister	Secundaria
Nacional	240	No aplica 3BM	04/12/1970	M	N	

Tercer análisis de perfil con herramienta DQ Analyzer.

Para finalizar la etapa de preparación de datos, se presenta el perfil final del conjunto de datos minables como se describe la *Tabla 7*:

Tabla 7: Perfil # 3 de calidad de datos

Expression	Type	Nulls	Not nulls	Distinct	Unique	Min	Max	Mediana
CODEMPLEADO	INTEGER	0	10,679	10,679	10,679	129,921	1,216,715,181	43,258,606
FECHAINGRESONOMINA	STRING	0	10,679	1,646	1,075	01/13/2003 00:00:00	2014-12-08 00:00:00	2003-01-01 00:00:00
VINCULACIONCATEGORIA	STRING	0	10,679	2	0	Contratista	Planta	Planta
VINCULACIONESTADO	STRING	0	10,679	1	0	Activa	Activa	Activa
VACACIONESMETODO	STRING	0	10,679	3	1	Docentes	No Aplica	Docentes
CODCALENDARIO	INTEGER	0	10,679	2	0	1	2	1
CODCARGOEMPRESA	INTEGER	0	10,679	16	2	219	9,032	9,001
CENTROCOSTO	STRING	0	10,679	297	35	ANEXO ESQUINA AL FRENTE	SEDE LA SUSANITA	I.E. JOSE ASUNCIÓN SILVA
NIVELCONTRATACION	STRING	0	10,679	7	0	Emp. Público sin A.P	Provisional Vacante Temporal	Propiedad
CODDEDICACIONLABORAL	INTEGER	0	10,679	3	1	0	5	1
BASICO	INTEGER	0	10,679	57	4	616,000	4,009,527	1,844,811
CODTURNO	INTEGER	0	10,679	1	0	1	1	1
SITUACIONLABORAL	STRING	0	10,679	12	1	Comisión	Suspensión	Normal
VINCULACIONNOVEDAD	STRING	0	10,679	22	4	Aclaratoria	Traslados	Cambios de Sueldo
FUENTESDERECURSOSPARAPAGO	STRING	0	10,679	2	0	Donaciones	Sistema General Participación	Sistema General Participación
CIUDAD	STRING	0	10,679	1	0	Medellín (Ant.)	Medellín (Ant.)	Medellín (Ant.)
AREAEDUCATIVA	STRING	0	10,679	20	1	Áreas de Apoyo para Educación Especial	Tecnología de Informática	No Aplica

NIVELEDUCACION	STRING	0	10,679	20	2	Bachiller	Tecnólogo	Licenciado
NIVELICTA	STRING	0	10,679	6	0	Ciclo Compl. Normales	Sin Asignación Directa	Primaria
REGIMENDEVINCULACION	STRING	0	10,679	6	0	Cofinanciado Departamental	Nacionalizados	Nacionalizados
HORASTRAB	INTEGER	0	10,679	2	1	15	240	240
AREAEDUCATIVECNICA	STRING	0	10,679	11	2	Agropecuaria (NA)	Salud	No aplica
GRADO	STRING	0	10,679	63	6	1	BC	2A
FECHANACIMIENTO	DATETIME	0	10,679	6,959	4,610	1946-01-20 00:00:00.000	1995-03-14 00:00:00.000	1968-03-19 00:00:00.000
SEXO	STRING	0	10,679	2	0	F	M	F
CONDICIONDEAMENAZADO	STRING	0	10,679	2	0	N	S	N

A continuación, se presentan la información que se ha extraído del perfil y evidencia el proceso llevado a cabo con los datos.

Tabla 8: Perfil #3 de variable CODEMPLEADO

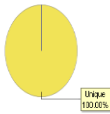
General		Estadísticas		Frecuencia		Máscara/Extremos		Observaciones		
Expression:	CODEMPLEADO									
Datatype:	INTEGER									
Rows:	10,679									
 Unique 100.00%		Counts		Group Frequency Analysis		N/A		El campo de tipo entero presenta los siguientes aspectos: - Es un campo obligatorio. - No presenta duplicados, ni valores nulos. - Existen 10679 CODEMPLEADO con valores únicos de un total de 10679.		
		Type	Count	Group Size	Group Count					Percentage
		Non-null values:	10,679	1	10679					100.00%
		Null values:	0							
		Distinct values:	10,679							
		Duplicate values:	0							
		Unique values:	10,679							
		Non-unique	0							
		Statistics								
		Type	value							
		Minimum value:	129,921							
		Median value:	43,258,606							
		Maximum value:	1,216,715,181							
		Sum:	874,880,373,823							
		Average:	81,925,308.91							
Variance:	37,360,270,437,314,692.69									
Standard deviation:	193,288,050.43									
		Group sizes								

Tabla 9: Perfil #3 de variable FECHAINGRESO NOMINA

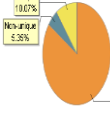
General		Estadísticas		Frecuencia		Máscara/Extremos		Observaciones	
Expression:	FECHAINGRESO NOMINA	Counts		Group Frequency Analysis				- El 84,59% de los valores son declarados como duplicados, es decir 9033 datos, 10,07% son valores únicos equivalentes a 1075 y 5,31% valores no únicos equivalentes a 571, situación considerada normal si se tiene en cuenta que uno o varios docentes pueden ingresar a nómina en una misma fecha o en fechas diferentes y esta dinámica hace parte del proceso.	
Datatype:	STRING								
Rows:	10,679								
 Unique 10.07% Non-unique 84.59% Duplicate 5.31%		Type	Count	Group Size	Group Count	Percentage			
		Non-null values:	10,679	1	1075	65.31%			
		Null values:	0	2	249	15.13%			
		Distinct values:	1,646	3	101	6.14%			
		Duplicate values:	9,033	4	47	2.86%			
		Unique values:	1,075	5	32	1.94%			
		Non-unique values:	571	6	24	1.46%			
		Statistics							
		Type	value						
		Minimum value:	01/13/2003 00:00:00						
Median value:	01/01/2003 00:00:00								
Maximum value:	08/12/2014 00:00:00								
Minimum length:	19								
Median length:	19								
Average length:	19								
Maximum length:	19								
		13	4	0.24%					
		17	3	0.18%					
		15	3	0.18%					
		67	2	0.12%					
		41	2	0.12%					
		23	2	0.12%					
		22	2	0.12%					
		21	2	0.12%					
		20	2	0.12%					
		19	2	0.12%					
		16	2	0.12%					
		2,82	1	0.06%					
		1,771	1	0.06%					
		505	1	0.06%					
		284	1	0.06%					
		232	1	0.06%					
		199	1	0.06%					
		179	1	0.06%					
		128	1	0.06%					
		109	1	0.06%					
		106	1	0.06%					
		88	1	0.06%					
		76	1	0.06%					
		75	1	0.06%					
		64	1	0.06%					
		61	1	0.06%					
		60	1	0.06%					
		57	1	0.06%					
		39	1	0.06%					
		38	1	0.06%					
		28	1	0.06%					
		24	1	0.06%					

Tabla 10: Perfil #3 de variable VINCULACIONCATEGORIA y VINCULACIONESTADO

General		Estadísticas		Frecuencia		Máscara/Extremos		Observaciones		
Expression:	VINCULACIONCATEGORIA	Counts		Frequency Analysis		Mask Analysis		El número de registros evaluados es de 10679 con un total de 98,98% registros duplicados que equivalen a 10677; situación normal para el campo, ya que ofrece la posibilidad de que el docente tenga la misma categoría de vinculación igual a la de otro docente. El 0,02% equivalente 2 como valor no único es un campo de tipo string. La frecuencia de los datos está representada por el 99,96%, equivalente a 10675 docentes vinculados con categoría de planta; 0,04% docentes vinculados a la categoría contratista.		
Datatype:	STRING	Type	Count	Value	Count	Percentage	Value		Count	Percent age
Rows:	10,679	Non-null values:	10,679	Planta	10675	99.96%	LLLLL		10675	99.96%
		Null values:	0	Contratista	4	0.04%	LLLLLLLLL		4	0.04%
		Distinct values:	2							
		Duplicate values:	10,677							
		Unique values:	0							
		Non-unique values:	2							
		Statistics								
		Type	value							
		Minimum value:	Contratista							
		Median value:	Planta							
		Maximum value:	Planta							
		Minimum length:	6							
		Median length:	6							
		Average length:	6							
		Maximum length:	11							
Expression:	VINCULACIONESTADO	Counts		Frequency Analysis		Mask Analysis		Es un campo de tipo string, con un total de 10679 registros analizados que presentan las siguientes características: - 99,99% datos duplicados equivalentes a 10678. - Existe una única categoría asignada a la variable VINCULACIONESTADO = Activa. - La máscara de entrada asignada está conformada por letras y es constante en todo el set de datos.		
Datatype:	STRING	Type	Count	Value	Count	Percentage	Value		Count	Percent age
Rows:	10,679	Non-null values:	10,679	Activa	10679	100.00%	LLLLL		10679	100.00%
		Null values:	0							
		Distinct values:	1							
		Duplicate values:	10,678							
		Unique values:	0							
		Non-unique values:	1							
		Statistics								
		Type	value							
		Minimum value:	Activa							
		Median value:	Activa							
		Maximum value:	Activa							
		Minimum length:	6							
		Median length:	6							
		Average length:	6							
		Maximum length:	6							

Tabla 11: Perfil #3 de variable VACACIONESMETODO y CODCALENDARIO

General		Estadísticas		Frecuencia		Máscara/Extremos		Observaciones		
Expression:	VACACIONESMETODO							Es un campo de tipo string, con un total de 10679 registros analizados que presentan las siguientes características: - 99,97% datos duplicados equivalentes a 10679, 3 valores distintos y como valores únicos el 0,02 % correspondiente a 2 registros. - En el análisis de frecuencia se categorizan el método vacaciones con tres valores: Docentes con un 95,33%, Estándar con un 4,66% y No aplica con un 0,01%. El tipo de máscara de entrada está determinado por letras en dos formatos diferentes.		
Datatype:	STRING									
Rows:	10,679									
 Docentes 95.33% No aplica 0.01%		Counts		Frequency Analysis		Mask Analysis				
		Type	Count	Value	Count	Percentage	Value		Count	Percent age
		Non-null values:	10,679	Docentes	10180	95.33%	LLLLLLLL		10678	99.99%
		Null values:	0	Estándar	498	4.66%	LL LLLLLL		1	0.01%
		Distinct values:	3	No Aplica	1	0.01%				
		Duplicate values:	10,676							
		Unique values:	1							
		Non-unique values:	2							
		Statistics								
		Type	value							
Minimum value:	Docentes									
Median value:	Docentes									
Maximum value:	No Aplica									
Minimum length:	8									
Median length:	8									
Average length:	8									
Maximum length:	9									
Expression:	CODCALENDARIO							Es un campo de tipo entero, con un total de 10679 registros analizados que presentan las siguientes características: - 99,98% datos duplicados equivalentes a 10677, y 2 valores distintos y como valores únicos el 0,02% correspondiente a 2 registros. - No hay una máscara de entrada definida para este campo. En su frecuencia de análisis presenta como valores definidos el uno(1) y dos (2).		
Datatype:	INTEGER									
Rows:	10,679									
 1 99.84% 2 0.16%		Counts		Frequency Analysis		N/A				
		Type	Count	Value	Count	Percentage				
		Non-null values:	10,679	1	10662	99.84%				
		Null values:	0	2	17	0.16%				
		Distinct values:	2							
		Duplicate values:	10,677							
		Unique values:	0							
		Non-unique values:	2							
		Statistics								
		Type	value							
Minimum value:	1									
Median value:	1									
Maximum value:	2									
Sum:	10,696									
Average:	1									
Variance:	0									
Standard deviation:	0.04									

Tabla 12: Perfil #3 de variable CODCARGOEMPRESA

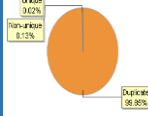
General		Estadísticas		Frecuencia			Máscara/Extremos	Observaciones
Expression:	CODCARGOEMPRESA						N/A	El 99,85% de la información está diagnosticada como valores duplicados, es decir 10663 valores; el 0,13% = 14 valores no únicos, y, 16 valores distintos; 0,02% valores únicos. - El campo no tiene una máscara de entrada definida. - En el análisis de frecuencia el valor con mayor representación es 9001 con un 89,44%.
Datatype:	INTEGER							
Rows:	10,679							
		Counts		Frequency Analysis				
		Type	Count	Value	Count	Percentage		
		Non-null values:	10,679	9,001	9562	89.54%		
		Null values:	0	907	385	3.61%		
		Distinct values:	16	407	262	2.45%		
		Duplicate values:	10,663	9,032	146	1.37%		
		Unique values:	2	470	94	0.88%		
		Non-unique values:	14	477	83	0.78%		
		Statistics		9,005	58	0.54%		
		Type	value	440	36	0.34%		
		Minimum value:	219	902	14	0.13%		
		Median value:	9,001	314	11	0.10%		
		Maximum value:	9,032	219	9	0.08%		
		Sum:	88,546,100	908	9	0.08%		
		Average:	8,291.61	9,003	5	0.05%		
Variance:	5,434,278.25	480	3	0.03%				
Standard deviation:	2,331.15	425	1	0.01%				
		9,031	1	0.01%				

Tabla 13: Perfil #3 de variable CENTROCOSTO

General		Estadísticas		Frecuencia			Máscara/Extremos	Observaciones	
Expression:	CENTROCOSTO			Tiene una máscara de entrada definida para cada decena por					Es un campo de tipo dato string, con un total de 10679 registros analizados que presentan las siguientes características: - El 97,32% de la información está diagnosticada como valores duplicados, es decir 10382 valores; el 2,40% = 262 valores no únicos, 297 valores distintos; 0,33% = 35 valores únicos. - La información corresponde a las instituciones educativas de la Secretaría de Educación a la que se afilia cada docente, en el análisis de frecuencia aparecen, en primer lugar la institución Concejo de Medellín con 1,48% y en su orden la siguen la institución Finca la Mesa con 1,35% y así en su orden. - Tiene una máscara de entrada definida en letras para cada institución educativa de la Secretaría de Educación de Medellín.
Datatype:	STRING								
Rows:	10,679								
11.78% 6.40% 4.25% 3.70% 2.69% 2.36% 2.36% 1.68% 1.68% 1.68% 1.68% 1.68% 1.35% 1.35% 1.35% 1.35% 1.35% 1.35% 1.01									

Tabla 14: Perfil #3 de variable NIVELCONTRATACION y CODEDICACIONLABORAL

General		Estadísticas		Frecuencia		Máscara/Extremos		Observaciones																																																																					
Expression:	NIVELCONTRATACION	Counts		Frequency Analysis		Mask Analysis		Es un campo de tipo dato string, con un total de 10679 registros analizados que presentan características: - El 99,93% la información está diagnosticada como valores duplicados, es decir 10672 valores; el 0,07% = 7 valores no únicos, y 7 valores distintos; 0valores únicos. - En el análisis de frecuencia el nivel de contratación tiene como clasificación en valores lo siguiente: en propiedad un 81,64%, provisional vacante definitiva un 17,54%, provisional vacante temporal un 0,57%, periodo de prueba un 0,15%, emp público sin A.P un 0,07%, no definido un 0,02%, planta temporal 0,02%. - El campo tiene 7 máscaras de entrada definidas para el set de datos.																																																																					
Datatype:	STRING																																																																												
Rows:	10,679																																																																												
Non-unique 0,07% Unique 99,93% Duplicate 10,672		<table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>10,679</td></tr><tr><td>Null values:</td><td>0</td></tr><tr><td>Distinct values:</td><td>7</td></tr><tr><td>Duplicate values:</td><td>10,672</td></tr><tr><td>Unique values:</td><td>0</td></tr><tr><td>Non-unique values:</td><td>7</td></tr></tbody></table> <table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td>Propiedad</td><td>8718</td><td>81.64%</td></tr><tr><td>Provisional Vacante Definitiva</td><td>1873</td><td>17.54%</td></tr><tr><td>Provisional Vacante Temporal</td><td>61</td><td>0.57%</td></tr><tr><td>Periodo de Prueba</td><td>16</td><td>0.15%</td></tr><tr><td>Emp Público Sin A.P</td><td>7</td><td>0.07%</td></tr><tr><td>No Definido</td><td>2</td><td>0.02%</td></tr><tr><td>Planta Temporal</td><td>2</td><td>0.02%</td></tr></tbody></table> <table><thead><tr><th>Type</th><th>value</th></tr></thead><tbody><tr><td>Minimum</td><td>Emp Público Sin A.P</td></tr><tr><td>Median value:</td><td>Propiedad</td></tr><tr><td>Maximum</td><td>Provisional Vacante Temporal</td></tr><tr><td>Minimum</td><td>9</td></tr><tr><td>Maximum</td><td>30</td></tr></tbody></table>	Type	Count	Non-null values:	10,679	Null values:	0	Distinct values:	7	Duplicate values:	10,672	Unique values:	0	Non-unique values:	7	Value	Count	Percentage	Propiedad	8718	81.64%	Provisional Vacante Definitiva	1873	17.54%	Provisional Vacante Temporal	61	0.57%	Periodo de Prueba	16	0.15%	Emp Público Sin A.P	7	0.07%	No Definido	2	0.02%	Planta Temporal	2	0.02%	Type	value	Minimum	Emp Público Sin A.P	Median value:	Propiedad	Maximum	Provisional Vacante Temporal	Minimum	9	Maximum	30	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td>LLLLLLLL</td><td>8718</td><td>81.64%</td></tr><tr><td>LLLLLLLL LLLLLL</td><td>1873</td><td>17.54%</td></tr><tr><td>LLLLLLLL LLLLLL</td><td>61</td><td>0.57%</td></tr><tr><td>LLLLLLLL L LLLLL</td><td>16</td><td>0.15%</td></tr><tr><td>LL LLLLLL L L L</td><td>7</td><td>0.07%</td></tr><tr><td>L LLLLLLL</td><td>2</td><td>0.02%</td></tr><tr><td>LLLLL LLLLLL</td><td>2</td><td>0.02%</td></tr></tbody></table>	Value	Count	Percentage	LLLLLLLL	8718	81.64%	LLLLLLLL LLLLLL	1873	17.54%	LLLLLLLL LLLLLL	61	0.57%	LLLLLLLL L LLLLL	16	0.15%	LL LLLLLL L L L	7	0.07%	L LLLLLLL	2	0.02%	LLLLL LLLLLL	2	0.02%
Type	Count																																																																												
Non-null values:	10,679																																																																												
Null values:	0																																																																												
Distinct values:	7																																																																												
Duplicate values:	10,672																																																																												
Unique values:	0																																																																												
Non-unique values:	7																																																																												
Value	Count	Percentage																																																																											
Propiedad	8718	81.64%																																																																											
Provisional Vacante Definitiva	1873	17.54%																																																																											
Provisional Vacante Temporal	61	0.57%																																																																											
Periodo de Prueba	16	0.15%																																																																											
Emp Público Sin A.P	7	0.07%																																																																											
No Definido	2	0.02%																																																																											
Planta Temporal	2	0.02%																																																																											
Type	value																																																																												
Minimum	Emp Público Sin A.P																																																																												
Median value:	Propiedad																																																																												
Maximum	Provisional Vacante Temporal																																																																												
Minimum	9																																																																												
Maximum	30																																																																												
Value	Count	Percentage																																																																											
LLLLLLLL	8718	81.64%																																																																											
LLLLLLLL LLLLLL	1873	17.54%																																																																											
LLLLLLLL LLLLLL	61	0.57%																																																																											
LLLLLLLL L LLLLL	16	0.15%																																																																											
LL LLLLLL L L L	7	0.07%																																																																											
L LLLLLLL	2	0.02%																																																																											
LLLLL LLLLLL	2	0.02%																																																																											
Expression:	CODIFICACIONLABORAL	Counts		Frequency Analysis		N/A		Es un campo de tipo dato entero, con un total de 10679 registros analizados que presentan las siguientes características: - El 99,97% la información está diagnosticada como valores duplicados, es decir 10679 valores; el 0,02% = 2 valores no únicos, y el 3 valores distintos; 1valores únicos. - En el análisis de frecuencia, los valores tienen la siguiente clasificación el 1 con el 99,96%, valor 0 con el 0,03% y valor 5 con el 0,01%. - El campo no tiene máscara entrada definida.																																																																					
Datatype:	INTEGER																																																																												
Rows:	10,679																																																																												
Unique 0,01% Non-unique 99,97% Duplicate 10,676		<table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>10,679</td></tr><tr><td>Null values:</td><td>0</td></tr><tr><td>Distinct values:</td><td>3</td></tr><tr><td>Duplicate values:</td><td>10,676</td></tr><tr><td>Unique values:</td><td>1</td></tr><tr><td>Non-unique values:</td><td>2</td></tr></tbody></table> <table><thead><tr><th>Type</th><th>value</th></tr></thead><tbody><tr><td>Minimum</td><td>0</td></tr><tr><td>Median value:</td><td>1</td></tr><tr><td>Maximum</td><td>5</td></tr><tr><td>Sum:</td><td>10,68</td></tr><tr><td>Average:</td><td>1</td></tr><tr><td>Variance:</td><td>0</td></tr><tr><td>Standard deviation:</td><td>0.04</td></tr></tbody></table>	Type	Count	Non-null values:	10,679	Null values:	0	Distinct values:	3	Duplicate values:	10,676	Unique values:	1	Non-unique values:	2	Type	value	Minimum	0	Median value:	1	Maximum	5	Sum:	10,68	Average:	1	Variance:	0	Standard deviation:	0.04	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td>1</td><td>10679</td><td>99.96%</td></tr><tr><td>0</td><td>3</td><td>0.03%</td></tr><tr><td>5</td><td>1</td><td>0.01%</td></tr></tbody></table>	Value	Count	Percentage	1	10679	99.96%	0	3	0.03%	5	1	0.01%																																
Type	Count																																																																												
Non-null values:	10,679																																																																												
Null values:	0																																																																												
Distinct values:	3																																																																												
Duplicate values:	10,676																																																																												
Unique values:	1																																																																												
Non-unique values:	2																																																																												
Type	value																																																																												
Minimum	0																																																																												
Median value:	1																																																																												
Maximum	5																																																																												
Sum:	10,68																																																																												
Average:	1																																																																												
Variance:	0																																																																												
Standard deviation:	0.04																																																																												
Value	Count	Percentage																																																																											
1	10679	99.96%																																																																											
0	3	0.03%																																																																											
5	1	0.01%																																																																											

Tabla 15: Perfil #3 de variable BASICO

General		Estadísticas		Máscara/Extremos		Observaciones																																																																																																																																																																																																													
Expression:	BASICO	Counts		Frequency Analysis		N/A	Es un campo de tipo dato entero, con un total de 10,679 registros analizados que presentan las siguientes características: - El 99,47% la información está diagnosticada como valores duplicados, es decir 10622 valores; el 0,49% = 53 valores no únicos, y 5Valores distintos; 0,4% =4 valores únicos. - El análisis de frecuencia arroja como datos de mayor relevancia en su orden : valor 1411890 con el 27,34%, 2711939 el 26,88%, 2381197 con el 7,97% y así.... - El campo no tiene definida máscara de entrada para el set de datos.																																																																																																																																																																																																												
Datatype:	INTEGER																																																																																																																																																																																																																		
Rows:	10,679	<table><tr><th>Type</th><th>Count</th></tr><tr><td>Non-null values:</td><td>10,679</td></tr><tr><td>Null values:</td><td>0</td></tr><tr><td>Distinct values:</td><td>57</td></tr><tr><td>Duplicate values:</td><td>10,622</td></tr><tr><td>Unique values:</td><td>4</td></tr><tr><td>Non-unique values:</td><td>53</td></tr></table>		Type	Count			Non-null values:	10,679	Null values:	0	Distinct values:	57	Duplicate values:	10,622	Unique values:	4	Non-unique values:	53	<table><tr><th>Value</th><th>Count</th><th>Percentage</th></tr><tr><td>1,411,890</td><td>2920</td><td>27.34%</td></tr><tr><td>2,711,939</td><td>2871</td><td>26.88%</td></tr><tr><td>2,381,197</td><td>851</td><td>7.97%</td></tr><tr><td>1,534,628</td><td>738</td><td>6.91%</td></tr><tr><td>1,844,811</td><td>646</td><td>6.05%</td></tr><tr><td>1,960,718</td><td>381</td><td>3.57%</td></tr><tr><td>2,151,184</td><td>328</td><td>3.07%</td></tr><tr><td>1,121,819</td><td>279</td><td>2.61%</td></tr><tr><td>2,154,714</td><td>198</td><td>1.85%</td></tr><tr><td>1,515,515</td><td>166</td><td>1.55%</td></tr><tr><td>1,623,673</td><td>140</td><td>1.31%</td></tr><tr><td>2,429,075</td><td>122</td><td>1.14%</td></tr><tr><td>1,305,676</td><td>101</td><td>0.95%</td></tr><tr><td>2,363,041</td><td>97</td><td>0.91%</td></tr><tr><td>1,255,371</td><td>91</td><td>0.85%</td></tr><tr><td>1,808,388</td><td>91</td><td>0.85%</td></tr><tr><td>1,583,721</td><td>81</td><td>0.76%</td></tr><tr><td>2,121,532</td><td>51</td><td>0.48%</td></tr><tr><td>2,797,931</td><td>48</td><td>0.45%</td></tr><tr><td>1,682,814</td><td>45</td><td>0.42%</td></tr><tr><td>3,460,354</td><td>37</td><td>0.35%</td></tr><tr><td>1,329,674</td><td>34</td><td>0.32%</td></tr><tr><td>2,574,881</td><td>31</td><td>0.29%</td></tr><tr><td>1,607,320</td><td>30</td><td>0.28%</td></tr><tr><td>1,358,243</td><td>29</td><td>0.24%</td></tr><tr><td>1,271,702</td><td>29</td><td>0.23%</td></tr><tr><td>1,430,005</td><td>23</td><td>0.22%</td></tr><tr><td>2,477,921</td><td>23</td><td>0.22%</td></tr><tr><td>2,874,648</td><td>21</td><td>0.20%</td></tr><tr><td>1,062,143</td><td>17</td><td>0.16%</td></tr><tr><td>1,758,304</td><td>16</td><td>0.15%</td></tr><tr><td>4,009,527</td><td>16</td><td>0.15%</td></tr><tr><td>1,718,968</td><td>15</td><td>0.14%</td></tr><tr><td>1,598,605</td><td>14</td><td>0.13%</td></tr><tr><td>1,641,853</td><td>14</td><td>0.13%</td></tr><tr><td>1,446,417</td><td>13</td><td>0.12%</td></tr><tr><td>1,843,384</td><td>11</td><td>0.10%</td></tr><tr><td>856,266</td><td>10</td><td>0.09%</td></tr><tr><td>944,535</td><td>9</td><td>0.08%</td></tr><tr><td>1,108,666</td><td>7</td><td>0.07%</td></tr><tr><td>2,458,869</td><td>7</td><td>0.07%</td></tr><tr><td>826,06</td><td>5</td><td>0.05%</td></tr><tr><td>2,285,199</td><td>4</td><td>0.04%</td></tr><tr><td>2,961,113</td><td>4</td><td>0.04%</td></tr><tr><td>616</td><td>2</td><td>0.02%</td></tr><tr><td>1,004,109</td><td>2</td><td>0.02%</td></tr><tr><td>2,013,621</td><td>2</td><td>0.02%</td></tr><tr><td>2,060,252</td><td>2</td><td>0.02%</td></tr><tr><td>2,208,620</td><td>2</td><td>0.02%</td></tr><tr><td>2,708,610</td><td>2</td><td>0.02%</td></tr><tr><td>2,744,883</td><td>2</td><td>0.02%</td></tr><tr><td>3,134,755</td><td>2</td><td>0.02%</td></tr><tr><td>3,162,686</td><td>2</td><td>0.02%</td></tr><tr><td>2,104,515</td><td>1</td><td>0.01%</td></tr><tr><td>2,651,640</td><td>1</td><td>0.01%</td></tr><tr><td>2,673,903</td><td>1</td><td>0.01%</td></tr><tr><td>3,096,165</td><td>1</td><td>0.01%</td></tr></table>		Value	Count	Percentage	1,411,890	2920	27.34%	2,711,939	2871	26.88%	2,381,197	851	7.97%	1,534,628	738	6.91%	1,844,811	646	6.05%	1,960,718	381	3.57%	2,151,184	328	3.07%	1,121,819	279	2.61%	2,154,714	198	1.85%	1,515,515	166	1.55%	1,623,673	140	1.31%	2,429,075	122	1.14%	1,305,676	101	0.95%	2,363,041	97	0.91%	1,255,371	91	0.85%	1,808,388	91	0.85%	1,583,721	81	0.76%	2,121,532	51	0.48%	2,797,931	48	0.45%	1,682,814	45	0.42%	3,460,354	37	0.35%	1,329,674	34	0.32%	2,574,881	31	0.29%	1,607,320	30	0.28%	1,358,243	29	0.24%	1,271,702	29	0.23%	1,430,005	23	0.22%	2,477,921	23	0.22%	2,874,648	21	0.20%	1,062,143	17	0.16%	1,758,304	16	0.15%	4,009,527	16	0.15%	1,718,968	15	0.14%	1,598,605	14	0.13%	1,641,853	14	0.13%	1,446,417	13	0.12%	1,843,384	11	0.10%	856,266	10	0.09%	944,535	9	0.08%	1,108,666	7	0.07%	2,458,869	7	0.07%	826,06	5	0.05%	2,285,199	4	0.04%	2,961,113	4	0.04%	616	2	0.02%	1,004,109	2	0.02%	2,013,621	2	0.02%	2,060,252	2	0.02%	2,208,620	2	0.02%	2,708,610	2	0.02%	2,744,883	2	0.02%	3,134,755	2	0.02%	3,162,686	2	0.02%	2,104,515	1	0.01%	2,651,640	1	0.01%	2,673,903	1	0.01%	3,096,165	1	0.01%	<table><tr><th>Type</th><th>value</th></tr><tr><td>Minimum value:</td><td>0</td></tr><tr><td>Median value:</td><td>1,844,811</td></tr><tr><td>Maximum value:</td><td>4,009,527</td></tr><tr><td>Sum:</td><td>21,170,232,568</td></tr><tr><td>Average:</td><td>1,982,417.13</td></tr><tr><td>Variance:</td><td>331,761,739,753.54</td></tr><tr><td>Standard deviation:</td><td>575,987.62</td></tr></table>		Type	value	Minimum value:	0	Median value:	1,844,811	Maximum value:	4,009,527	Sum:	21,170,232,568	Average:	1,982,417.13	Variance:	331,761,739,753.54
Type	Count																																																																																																																																																																																																																		
Non-null values:	10,679																																																																																																																																																																																																																		
Null values:	0																																																																																																																																																																																																																		
Distinct values:	57																																																																																																																																																																																																																		
Duplicate values:	10,622																																																																																																																																																																																																																		
Unique values:	4																																																																																																																																																																																																																		
Non-unique values:	53																																																																																																																																																																																																																		
Value	Count	Percentage																																																																																																																																																																																																																	
1,411,890	2920	27.34%																																																																																																																																																																																																																	
2,711,939	2871	26.88%																																																																																																																																																																																																																	
2,381,197	851	7.97%																																																																																																																																																																																																																	
1,534,628	738	6.91%																																																																																																																																																																																																																	
1,844,811	646	6.05%																																																																																																																																																																																																																	
1,960,718	381	3.57%																																																																																																																																																																																																																	
2,151,184	328	3.07%																																																																																																																																																																																																																	
1,121,819	279	2.61%																																																																																																																																																																																																																	
2,154,714	198	1.85%																																																																																																																																																																																																																	
1,515,515	166	1.55%																																																																																																																																																																																																																	
1,623,673	140	1.31%																																																																																																																																																																																																																	
2,429,075	122	1.14%																																																																																																																																																																																																																	
1,305,676	101	0.95%																																																																																																																																																																																																																	
2,363,041	97	0.91%																																																																																																																																																																																																																	
1,255,371	91	0.85%																																																																																																																																																																																																																	
1,808,388	91	0.85%																																																																																																																																																																																																																	
1,583,721	81	0.76%																																																																																																																																																																																																																	
2,121,532	51	0.48%																																																																																																																																																																																																																	
2,797,931	48	0.45%																																																																																																																																																																																																																	
1,682,814	45	0.42%																																																																																																																																																																																																																	
3,460,354	37	0.35%																																																																																																																																																																																																																	
1,329,674	34	0.32%																																																																																																																																																																																																																	
2,574,881	31	0.29%																																																																																																																																																																																																																	
1,607,320	30	0.28%																																																																																																																																																																																																																	
1,358,243	29	0.24%																																																																																																																																																																																																																	
1,271,702	29	0.23%																																																																																																																																																																																																																	
1,430,005	23	0.22%																																																																																																																																																																																																																	
2,477,921	23	0.22%																																																																																																																																																																																																																	
2,874,648	21	0.20%																																																																																																																																																																																																																	
1,062,143	17	0.16%																																																																																																																																																																																																																	
1,758,304	16	0.15%																																																																																																																																																																																																																	
4,009,527	16	0.15%																																																																																																																																																																																																																	
1,718,968	15	0.14%																																																																																																																																																																																																																	
1,598,605	14	0.13%																																																																																																																																																																																																																	
1,641,853	14	0.13%																																																																																																																																																																																																																	
1,446,417	13	0.12%																																																																																																																																																																																																																	
1,843,384	11	0.10%																																																																																																																																																																																																																	
856,266	10	0.09%																																																																																																																																																																																																																	
944,535	9	0.08%																																																																																																																																																																																																																	
1,108,666	7	0.07%																																																																																																																																																																																																																	
2,458,869	7	0.07%																																																																																																																																																																																																																	
826,06	5	0.05%																																																																																																																																																																																																																	
2,285,199	4	0.04%																																																																																																																																																																																																																	
2,961,113	4	0.04%																																																																																																																																																																																																																	
616	2	0.02%																																																																																																																																																																																																																	
1,004,109	2	0.02%																																																																																																																																																																																																																	
2,013,621	2	0.02%																																																																																																																																																																																																																	
2,060,252	2	0.02%																																																																																																																																																																																																																	
2,208,620	2	0.02%																																																																																																																																																																																																																	
2,708,610	2	0.02%																																																																																																																																																																																																																	
2,744,883	2	0.02%																																																																																																																																																																																																																	
3,134,755	2	0.02%																																																																																																																																																																																																																	
3,162,686	2	0.02%																																																																																																																																																																																																																	
2,104,515	1	0.01%																																																																																																																																																																																																																	
2,651,640	1	0.01%																																																																																																																																																																																																																	
2,673,903	1	0.01%																																																																																																																																																																																																																	
3,096,165	1	0.01%																																																																																																																																																																																																																	
Type	value																																																																																																																																																																																																																		
Minimum value:	0																																																																																																																																																																																																																		
Median value:	1,844,811																																																																																																																																																																																																																		
Maximum value:	4,009,527																																																																																																																																																																																																																		
Sum:	21,170,232,568																																																																																																																																																																																																																		
Average:	1,982,417.13																																																																																																																																																																																																																		
Variance:	331,761,739,753.54																																																																																																																																																																																																																		
Standard deviation:	575,987.62																																																																																																																																																																																																																		

100%

0.94%

Non-unique
0.97%

2920 value
0.947%

Tabla 16: Perfil #3 de variable CODTURNO y SITUACIONLABORAL

General

Datatype: INTEGER

Rows: 10,679

Non-null values: 100%

Duplicate values: 99,99%

Estadísticas

Counts	
Type	Count
Non-null values:	10,679
Null values:	0
Distinct values:	1
Duplicate values:	10,678
Unique values:	0
Non-unique values:	1

Type	value
Minimum value:	1
Median value:	1
Maximum value:	1
Sum:	10,679
Average:	1
Variance:	0
Standard deviation:	0

Frecuencia

Frequency Analysis		
Value	Count	Percentage
1	10679	100.00%

Máscara/Extremos

N/A

Observaciones

Es un campo de tipo dato entero, con un total de 10679 registros analizados que presentan las siguientes características:

- El 99,99% la información está diagnosticada como valores duplicados, es decir 1078 valores; el 0,01% = 1 valores no únicos, y 1 valores distintos; O valores únicos.
- Para el análisis de frecuencia tenemos como único valor relevante el 1 con un 99,99%.
- El campo no tiene definida máscara de entrada para el set de datos.

Expression: SITUACION LABORAL

Datatype: STRING

Rows: 10,679

Non-null values: 100%

Duplicate values: 99,99%

Estadísticas

Counts	
Type	Count
Non-null values:	10,679
Null values:	0
Distinct values:	12
Duplicate values:	10,667
Unique values:	1
Non-unique values:	11

Type	value
Minimum value:	Comisión
Median value:	Normal
Maximum value:	Suspensión
Minimum length:	6
Median length:	6
Average length:	6.39
Maximum length:	28

Frecuencia

Frequency Analysis		
Value	Count	Percentage
Normal	10407	97.45%
Encargo Vacante Definitiva	91	0.85%
Comisión	65	0.61%
Situación Laboral Remunerada	58	0.54%
Encargo Vacante Temporal	20	0.19%
Licencia No Remunerada	15	0.14%
Reemplazo de Enfermedad	7	0.07%
Sanción	5	0.05%
Reemplazo Encargos	4	0.04%
Reemplazo - Lic. Ordinaria	3	0.03%
Reemplazo de Maternidad	3	0.03%
Suspension	1	0.01%

Máscara/Extremos

Mask Analysis		
Value	Count	Percent age
	10407	97.45%
	91	0.85%
	65	0.61%
	58	0.54%
	20	0.19%
	15	0.14%
	10	0.09%
	5	0.05%
	4	0.04%
	3	0.03%
	1	0.01%

Observaciones

Es un campo de tipo dato string, con un total de 10679 registros analizados que presentan las siguientes características:

- El 99,89% la información está diagnosticada como valores duplicados, es decir 10667 valores; el 0,10% = 11 valores no únicos y 12 valores distintos; 0,01% = 1 valores únicos.
- En el análisis de frecuencia un valor relevante es normal con un 97,45%, en segundo nivel estaría encargo vacante definitiva con 0,85%, comisión con un 0,61%, situación laboral remunerada con un 0,54%, encargo vacante temporal con un 0,19%, licencia no remunerada con un 0,14%, reemplazo de enfermedad con un 0,07%, sanción con un 0,05%, reemplazo encargos con un 0,04%, reemplazo _licencia ordinaria con un 0,03%, reemplazo maternidad con un 0,03%, suspensión con un 0,01%.
- Para el campo están definidas 11 máscaras de entrada.

Tabla 17: Perfil #3 de variable VINCULACIONNOVEDAD

[illegible]

Tabla 18: Perfil #3 de variable CIUDAD y AREAEDUCATIVA

General		Estadísticas		Frecuencia		Máscara/Extremos			Observaciones																																																																																																																																														
Expression: CIUDAD									Es un campo de tipo string, con un total de 10679 registros analizados que presentan las siguientes características:																																																																																																																																														
Datatype: STRING																																																																																																																																																							
Rows:	10,679								- El 99,99% la información está diagnosticada como valores duplicados, es decir 10678 valores; el 0,01% = 1 valor no únicos, 1 valores distintos; 0valores únicos. - Este campo tiene definido un formato de máscara de entrada en letras. - El valor relevante en todo el set de datos es la frecuencia correspondiente a Medellín(Ant.) con un 100% .																																																																																																																																														
Non-unique 0.01% Duplicate 99.99%		<table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>10,679</td></tr><tr><td>Null values:</td><td>0</td></tr><tr><td>Distinct values:</td><td>1</td></tr><tr><td>Duplicate values:</td><td>10,678</td></tr><tr><td>Unique values:</td><td>0</td></tr><tr><td>Non-unique values:</td><td>1</td></tr></tbody></table>		Type	Count	Non-null values:	10,679	Null values:		0	Distinct values:	1	Duplicate values:	10,678	Unique values:	0	Non-unique values:	1	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td>Medellin (Ant)</td><td>10679</td><td>100.00%</td></tr></tbody></table>		Value	Count	Percentage	Medellin (Ant)	10679	100.00%	<table><thead><tr><th>Value</th><th>Count</th><th>Percent age</th></tr></thead><tbody><tr><td>LLLLLLLL (LLL)</td><td>10679</td><td>100.00%</td></tr></tbody></table>			Value	Count	Percent age	LLLLLLLL (LLL)	10679	100.00%																																																																																																																				
Type	Count																																																																																																																																																						
Non-null values:	10,679																																																																																																																																																						
Null values:	0																																																																																																																																																						
Distinct values:	1																																																																																																																																																						
Duplicate values:	10,678																																																																																																																																																						
Unique values:	0																																																																																																																																																						
Non-unique values:	1																																																																																																																																																						
Value	Count	Percentage																																																																																																																																																					
Medellin (Ant)	10679	100.00%																																																																																																																																																					
Value	Count	Percent age																																																																																																																																																					
LLLLLLLL (LLL)	10679	100.00%																																																																																																																																																					
Expression: AREAEDUCATIVA									Es un campo de tipo string, con un total de 10179 registros analizados que presentan las siguientes características:																																																																																																																																														
Datatype: STRING																																																																																																																																																							
Rows:	10,679								- El 99,81% la información está diagnosticada como valores duplicados, es decir 10659 valores; el 0,18% = 19 valor no únicos, y 20 valores distintos; 1valores únicos. - Para el análisis de frecuencia los valores tiene una frecuencia así: primaria con un 32,64%, no aplicación un 9,05%, matemáticas con un 7,84%, humanidades y lengua castellana con un 7,16%, ciencias sociales con un 6,35%, preescolar con un 5,99%, idioma extranjero inglés con un 4,72%, tecnología e informática con un 4,21%, ciencias naturales y educación ambiental con un 3,70%, ed. física, recreación y deporte con un 3,47%, ciencias naturales química con un 2,63%, áreas de apoyo para la educación especial con un 2,23%, educación artística-artes plásticas con un 2,23%, educación religiosa con un 1,58%, edu. ética y en valores con un 1,06%, ciencias naturales físicas con un 1,05%, filosofía con un 0,85%, educ. artística - música 0,07%, educ. artística - artes escénicas con un 0,03%, ciencias económicas y políticas con un 0,01% . - La máscara de entrada para este campo está definida por 19 formatos en letras para el set de datos.																																																																																																																																														
Unique 0.01% Non-unique 99.81%		<table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>10,679</td></tr><tr><td>Null values:</td><td>0</td></tr><tr><td>Distinct values:</td><td>20</td></tr><tr><td>Duplicate values:</td><td>10,659</td></tr><tr><td>Unique values:</td><td>1</td></tr><tr><td>Non-unique values:</td><td>19</td></tr></tbody></table>		Type	Count	Non-null values:	10,679	Null values:		0	Distinct values:	20	Duplicate values:	10,659	Unique values:	1	Non-unique values:	19	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td>Primaria</td><td>3486</td><td>32.64%</td></tr><tr><td>No Aplica</td><td>966</td><td>9.05%</td></tr><tr><td>Matemáticas</td><td>837</td><td>7.84%</td></tr><tr><td>Humanidades y Lengua Castellana</td><td>765</td><td>7.16%</td></tr><tr><td>Ciencias Sociales</td><td>678</td><td>6.35%</td></tr><tr><td>Preescolar</td><td>640</td><td>5.99%</td></tr><tr><td>Áreas de Apoyo para Educación Especial</td><td>569</td><td>5.33%</td></tr><tr><td>Idioma Extranjero Ingles</td><td>504</td><td>4.72%</td></tr><tr><td>Tecnología de Informática</td><td>450</td><td>4.21%</td></tr><tr><td>Ciencias Naturales y Edu. Ambiental</td><td>395</td><td>3.70%</td></tr><tr><td>Educ. Física, Recreación y Deporte</td><td>371</td><td>3.47%</td></tr><tr><td>Ciencias Naturales</td><td>281</td><td>2.63%</td></tr><tr><td>Educ. Artística - Artes Plásticas</td><td>238</td><td>2.23%</td></tr><tr><td>Educ. Religiosa</td><td>169</td><td>1.58%</td></tr><tr><td>Educ. Ética y en Valores</td><td>114</td><td>1.07%</td></tr><tr><td>Ciencias Naturales Física</td><td>113</td><td>1.06%</td></tr><tr><td>Filosofía</td><td>91</td><td>0.85%</td></tr><tr><td>Educ. Artística - Música</td><td>8</td><td>0.07%</td></tr><tr><td>Educ. Artística - Artes Escénicas</td><td>3</td><td>0.03%</td></tr><tr><td>Ciencias Económicas y Política</td><td>1</td><td>0.01%</td></tr><tr><td>Ciencias Económicas y Política</td><td>1</td><td>0.01%</td></tr></tbody></table>		Value	Count	Percentage	Primaria	3486	32.64%	No Aplica	966	9.05%	Matemáticas	837	7.84%	Humanidades y Lengua Castellana	765	7.16%	Ciencias Sociales	678	6.35%	Preescolar	640	5.99%	Áreas de Apoyo para Educación Especial	569	5.33%	Idioma Extranjero Ingles	504	4.72%	Tecnología de Informática	450	4.21%	Ciencias Naturales y Edu. Ambiental	395	3.70%	Educ. Física, Recreación y Deporte	371	3.47%	Ciencias Naturales	281	2.63%	Educ. Artística - Artes Plásticas	238	2.23%	Educ. Religiosa	169	1.58%	Educ. Ética y en Valores	114	1.07%	Ciencias Naturales Física	113	1.06%	Filosofía	91	0.85%	Educ. Artística - Música	8	0.07%	Educ. Artística - Artes Escénicas	3	0.03%	Ciencias Económicas y Política	1	0.01%	Ciencias Económicas y Política	1	0.01%	<table><thead><tr><th>Value</th><th>Count</th><th>Percent age</th></tr></thead><tbody><tr><td>LLLLLLLL</td><td>3486</td><td>32.64%</td></tr><tr><td>LL LLLLLL</td><td>966</td><td>9.05%</td></tr><tr><td>LLLLLLLLLL</td><td>837</td><td>7.84%</td></tr><tr><td>LLLLLLLLLL L LLLLL</td><td>765</td><td>7.16%</td></tr><tr><td>LLLLLLLL LLLLLLL</td><td>678</td><td>6.35%</td></tr><tr><td>LLLLLLLLLL</td><td>640</td><td>5.99%</td></tr><tr><td>LLLL LL LLLL LLLL</td><td>569</td><td>5.33%</td></tr><tr><td>LLLLLLLL LLLLLLL LLLLL</td><td>504</td><td>4.72%</td></tr><tr><td>LLLLLLLLLL LL</td><td>450</td><td>4.21%</td></tr><tr><td>LLLLLLLL LLLLLLLLL L</td><td>395</td><td>3.70%</td></tr><tr><td>LLLL LLLLLLL LLLLLLLLL</td><td>371</td><td>3.47%</td></tr><tr><td>LLLLLLLL LLLLLLL</td><td>281</td><td>2.63%</td></tr><tr><td>LLLL LLLLLLL - LLLL LLLLLLL</td><td>241</td><td>2.26%</td></tr><tr><td>LLLL LLLLLLL</td><td>169</td><td>1.58%</td></tr><tr><td>LLLL LLLL L L LLLLLL</td><td>114</td><td>1.07%</td></tr><tr><td>LLLLLLLL LLLLLLL</td><td>113</td><td>1.06%</td></tr><tr><td>LLLL LLLLLLL - LLLLL</td><td>91</td><td>0.85%</td></tr><tr><td>LLLLLLLL LLLLLLL L</td><td>8</td><td>0.07%</td></tr><tr><td>LLLLLLLL LLLLLLL L</td><td>3</td><td>0.01%</td></tr><tr><td>LLLLLLLL LLLLLLL L</td><td>1</td><td>0.01%</td></tr></tbody></table>			Value	Count	Percent age	LLLLLLLL	3486	32.64%	LL LLLLLL	966	9.05%	LLLLLLLLLL	837	7.84%	LLLLLLLLLL L LLLLL	765	7.16%	LLLLLLLL LLLLLLL	678	6.35%	LLLLLLLLLL	640	5.99%	LLLL LL LLLL LLLL	569	5.33%	LLLLLLLL LLLLLLL LLLLL	504	4.72%	LLLLLLLLLL LL	450	4.21%	LLLLLLLL LLLLLLLLL L	395	3.70%	LLLL LLLLLLL LLLLLLLLL	371	3.47%	LLLLLLLL LLLLLLL	281	2.63%	LLLL LLLLLLL - LLLL LLLLLLL	241	2.26%	LLLL LLLLLLL	169	1.58%	LLLL LLLL L L LLLLLL	114	1.07%	LLLLLLLL LLLLLLL	113	1.06%	LLLL LLLLLLL - LLLLL	91	0.85%	LLLLLLLL LLLLLLL L	8	0.07%	LLLLLLLL LLLLLLL L	3	0.01%	LLLLLLLL LLLLLLL L	1
Type	Count																																																																																																																																																						
Non-null values:	10,679																																																																																																																																																						
Null values:	0																																																																																																																																																						
Distinct values:	20																																																																																																																																																						
Duplicate values:	10,659																																																																																																																																																						
Unique values:	1																																																																																																																																																						
Non-unique values:	19																																																																																																																																																						
Value	Count	Percentage																																																																																																																																																					
Primaria	3486	32.64%																																																																																																																																																					
No Aplica	966	9.05%																																																																																																																																																					
Matemáticas	837	7.84%																																																																																																																																																					
Humanidades y Lengua Castellana	765	7.16%																																																																																																																																																					
Ciencias Sociales	678	6.35%																																																																																																																																																					
Preescolar	640	5.99%																																																																																																																																																					
Áreas de Apoyo para Educación Especial	569	5.33%																																																																																																																																																					
Idioma Extranjero Ingles	504	4.72%																																																																																																																																																					
Tecnología de Informática	450	4.21%																																																																																																																																																					
Ciencias Naturales y Edu. Ambiental	395	3.70%																																																																																																																																																					
Educ. Física, Recreación y Deporte	371	3.47%																																																																																																																																																					
Ciencias Naturales	281	2.63%																																																																																																																																																					
Educ. Artística - Artes Plásticas	238	2.23%																																																																																																																																																					
Educ. Religiosa	169	1.58%																																																																																																																																																					
Educ. Ética y en Valores	114	1.07%																																																																																																																																																					
Ciencias Naturales Física	113	1.06%																																																																																																																																																					
Filosofía	91	0.85%																																																																																																																																																					
Educ. Artística - Música	8	0.07%																																																																																																																																																					
Educ. Artística - Artes Escénicas	3	0.03%																																																																																																																																																					
Ciencias Económicas y Política	1	0.01%																																																																																																																																																					
Ciencias Económicas y Política	1	0.01%																																																																																																																																																					
Value	Count	Percent age																																																																																																																																																					
LLLLLLLL	3486	32.64%																																																																																																																																																					
LL LLLLLL	966	9.05%																																																																																																																																																					
LLLLLLLLLL	837	7.84%																																																																																																																																																					
LLLLLLLLLL L LLLLL	765	7.16%																																																																																																																																																					
LLLLLLLL LLLLLLL	678	6.35%																																																																																																																																																					
LLLLLLLLLL	640	5.99%																																																																																																																																																					
LLLL LL LLLL LLLL	569	5.33%																																																																																																																																																					
LLLLLLLL LLLLLLL LLLLL	504	4.72%																																																																																																																																																					
LLLLLLLLLL LL	450	4.21%																																																																																																																																																					
LLLLLLLL LLLLLLLLL L	395	3.70%																																																																																																																																																					
LLLL LLLLLLL LLLLLLLLL	371	3.47%																																																																																																																																																					
LLLLLLLL LLLLLLL	281	2.63%																																																																																																																																																					
LLLL LLLLLLL - LLLL LLLLLLL	241	2.26%																																																																																																																																																					
LLLL LLLLLLL	169	1.58%																																																																																																																																																					
LLLL LLLL L L LLLLLL	114	1.07%																																																																																																																																																					
LLLLLLLL LLLLLLL	113	1.06%																																																																																																																																																					
LLLL LLLLLLL - LLLLL	91	0.85%																																																																																																																																																					
LLLLLLLL LLLLLLL L	8	0.07%																																																																																																																																																					
LLLLLLLL LLLLLLL L	3	0.01%																																																																																																																																																					
LLLLLLLL LLLLLLL L	1	0.01%																																																																																																																																																					

Tabla 19: Perfil #3 de variable NIVELEDUCACION

General		Estadísticas		Frecuencia		Máscara/Extremos		Observaciones	
Expression:	NIVELEDUCACION							Es un campo de tipo string, con un total de 10679 registros analizados que presentan las siguientes características: - El 99,81% la información está diagnosticada como valores duplicados, es decir 10659 valores; el 0,17% = 18 valor no únicos, y el 0,02% = 2 valores nulos, 20 valores distintos; 2 valores únicos.	
Datatype:	STRING								
Rows:	10,679							- De acuerdo al análisis de frecuencias los valores de mayor relevancia son licenciado en un 49,36%, posgrado especialización en un 28,72%, profesional en un 7,91%, posgrado magister en un 3,50%, normalista superior en un 2,58%, bachiller académico en un 1,16%, bachiller pedagógico en un 1,16%, posgrado en educación con un 0,96%, técnico o tecnólogo en educación en un 0,44%, básica primaria un 0,35%, posgrado en educación 0,28%, maestro con un 0,19%, bachiller con un 0,19%, tecnología con un 0,06%, posgrado en otras áreas con un 0,04%, técnico o intermedia profesional con un 0,04%, básica secundaria con un 0,02%, posgrado doctorado 0,02%, no definido un 0,02. La máscara de entrada para este campo está definida por 17 formatos diferentes con letras para todo el set de datos. - La máscara de entrada para este campo está definida por 17 formatos diferentes con letras para todo el set de datos.	
									
		Counts		Frequency Analysis		Mask Analysis			
		Type	Count	Value	Count	Percentage	Value	Count	Percent age
		Non-null values:	10,679	Licenciado	5271	49.36%	LLLLLLLLLL	5271	49.36%
		Null values:	0	Postgrado Especializacion	3067	28.72%	LLLLLLLLLL	3067	28.72%
		Distinct values:	20	Profesional	845	7.91%	LLLLLLLLLL	845	7.91%
		Duplicate values:	10,659	Bachiller Académico	445	4.17%	LLLLLLLLLL LLLLLLL	447	4.19%
		Unique values:	2	Postgrado Magister	374	3.50%	LLLLLLLLLL LLLLLLL	374	3.50%
		Non-unique values:	18	Normalista Superior	276	2.58%	LLLLLLLLLL LLLLLLL	276	2.58%
				Bachiller Pedagógico	124	1.16%	LLLLLLLL LL LLLLLLL	133	1.25%
		Statistics		Posgrado en Educación	103	0.96%	LLLLLLLL LLLLLLL L	124	1.16%
		Type	value	Técnico o tecnólogo en Educación	47	0.44%	LLLLLL L LLLLLLL LL	47	0.44%
		Minimum value:	Bachiller	Básica Primaria	37	0.35%	LLLL LLLLLLL	37	0.35%
		Median value:	Licenciado	Posgrado en Educación	30	0.28%	LLLLLLLL	26	0.24%
		Maximum value:	Tecnólogo	Bachiller	20	0.19%	LLLLLLL	20	0.19%
		Minimum value:		Maestro	20	0.19%	LLLLLLLL L LLLLLLL	4	0.04%
		Minimum length:	7	Tecnólogo	6	0.06%	LLLLLLLL LL LLLL LLLL	4	0.04%
		Median length:	11	Posgrado en otras áreas	4	0.04%	LLLLLL LLLLLLL	2	0.02%
		Average length:	15.65	Técnico o Intermedia Profesional	4	0.04%	LL LLLLLLL	1	0.01%
		Maximum length:	32	Básica Secundaria	2	0.02%	LLLLLL L LLLLLLL LL	1	0.01%
				Postgrado doctorado	2	0.02%	LLLLLL L LLLLLLL LL	1	0.01%
				No definido	1	0.01%			
				Perito o Experto en Educación	1	0.01%			
				Perito o Experto en Educación	1	0.01%			

Tabla 20: Perfil #3 de variable NIVELDICTA y REGIMENDEVINCULACION

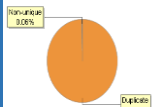
General		Estadísticas		Frecuencia		Máscara/Extremos			Observaciones
Expresión:	NIVELDICTA								Es un campo de tipo string, con un total de 10679 registros analizados que presentan las siguientes características: - El 99,94% la información está diagnosticada como valores duplicados, es decir 10673 valores; el 0,06% = 6 valor no únicos, y, 6 valores distintos, 0 valores únicos. - De acuerdo al análisis de frecuencia los valores de mayor relevancia son secundaria con un 41,94%, primaria con un 34,36%, media con un 11,76%, preescolar con un 5,99%, sin asignación directa con un 5,91%, ciclo complementario normal con un 0,04%. - El campo tiene definido 5 formatos en letras para la máscara de entrada en todo el set de datos.
Datatype:	STRING								
Rows:	10,679								
									
		Counts		Frequency Analysis		Mask Analysis			
		Type	Count	Value	Count	Percentage	Value	Count	Percent age
		Non-null values:	10,679	Secundaria	4479	41.94%	LLLLLLLLLL	5119	47.94%
		Null values:	0	Primaria	3669	34.36%	LLLLLLLL	3669	34.36%
		Distinct values:	6	Media	1256	11.76%	LLLLL	1256	11.76%
		Duplicate values:	10,673	Preescolar	640	5.99%	LLLL	631	5.91%
		Unique values:	0	Sin Asignación Directa	631	5.91%	LL LLLLLLLL LLLLLL	631	5.91%
		Non-unique values:	6	Ciclo complementario normal	4	0.04%	LLL LLLLLLLLLLLLLL	4	0.04%
		Type	value						
		Minimum value:	Ciclo complementario normal						
		Median value:	Primaria						
		Maximum value:	Sin Asignación Directa						
		Minimum length:	5						
		Median length:	10						
		Average length:	9.44						
		Maximum length:	27						
Expresión:		REGIMEN DE VINCULACION							Es un campo de tipo STRING, con un total de 10791 registros analizados que presentan las siguientes características: - El 98,94% la información está diagnosticada como valores duplicados, es decir 10677 valores; el 0,06% = 6 valor no únicos, y el 1,00% = 108 valores nulos, 6 valores distintos, 0 valores únicos. - De acuerdo al análisis de frecuencias, los valores que más relevancia obtuvieron son: nacionalizados con un 56,21%, nacional con un 20,52%, municipal con un 18,50%, departamental con un 3,06%, cofinanciado departamental con un 0,20% y cofinanciado municipal con un 0,10%. - La máscara de entrada para el campo está definida en 6 formatos con letras.
Datatype:		STRING							
Rows:		10,679							
									
		Counts		Frequency Analysis		Mask Analysis			
		Type	Count	Value	Count	Percentage	Value	Count	Percent age
		Non-null values:	10,679	Nacionalizados	6965	56.79%	LLLLLLLLLLLLLL	6965	56.79%
		Null values:	0	Nacional	2212	20.71%	LLLLLLLL	2212	20.71%
		Distinct values:	6	Municipal	2040	19.10%	LLLLLLLLL	2040	19.10%
		Duplicate values:	10,673	Departamental	329	3.08%	LLLLLLLLLLLLLL	329	3.08%
		Unique values:	0	Cofinanciado	22	0.21%	LLLLLLLLLLLLLL	22	0.21%
		Non-unique values:	6	Cofinanciado Municipal	11	0.10%	LLLLLLLLLLLLLL LLLLLL	11	0.10%
		Statistics							
		Type	value						
		Minimum value:	Cofinanciado Departamental						
		Median value:	Nacionalizados						
		Maximum value:	Nacionalizados						
		Minimum length:	8						
		Median length:	14						
		Average length:	11.81						
		Maximum length:	27						

Tabla 21: Perfil #3 de variable HORASTRAB y AREAEDUCATIVATECNICA

General		Estadísticas		Frecuencia		Máscara/Extremos		Observaciones																																																																																																		
Expression:	HORASTRAB																																																																																																									
Datatype:	INTEGER																																																																																																									
Rows:	10,679																																																																																																									
Unique 10.01% Non-unique 0.02% Duplicate 99.98%		<table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>10,679</td></tr><tr><td>Null values:</td><td>0</td></tr><tr><td>Distinct values:</td><td>2</td></tr><tr><td>Duplicate values:</td><td>10,677</td></tr><tr><td>Unique values:</td><td>1</td></tr><tr><td>Non-unique values:</td><td>1</td></tr></tbody></table> <table><thead><tr><th>Type</th><th>value</th></tr></thead><tbody><tr><td>Minimum value:</td><td>15</td></tr><tr><td>Median value:</td><td>240</td></tr><tr><td>Maximum value:</td><td>240</td></tr><tr><td>Sum:</td><td>2,562,735</td></tr><tr><td>Average:</td><td>239.98</td></tr><tr><td>Variance:</td><td>4.74</td></tr><tr><td>Standard deviation:</td><td>2.18</td></tr></tbody></table>	Type	Count	Non-null values:	10,679	Null values:	0	Distinct values:	2	Duplicate values:	10,677	Unique values:	1	Non-unique values:	1	Type	value	Minimum value:	15	Median value:	240	Maximum value:	240	Sum:	2,562,735	Average:	239.98	Variance:	4.74	Standard deviation:	2.18	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td>240</td><td>10678</td><td>99.99%</td></tr><tr><td>15</td><td>1</td><td>0.01%</td></tr></tbody></table>	Value	Count	Percentage	240	10678	99.99%	15	1	0.01%	N/A	Es un campo de tipo ENTERO, con un total de 10679 registros analizados que presentan las siguientes características: - El 99,98% la información está diagnosticada como valores duplicados, es decir 10677 valores; el 0,01% = 1 valor no únicos, y 2 valores distintos, 0,01% = 2 valores únicos. - De acuerdo al análisis de frecuencia los valores más relevantes son 240 con un 99,99% y 15 con un 0,01%. - El campo no tiene definida una máscara de entrada para todo el set de datos.																																																														
Type	Count																																																																																																									
Non-null values:	10,679																																																																																																									
Null values:	0																																																																																																									
Distinct values:	2																																																																																																									
Duplicate values:	10,677																																																																																																									
Unique values:	1																																																																																																									
Non-unique values:	1																																																																																																									
Type	value																																																																																																									
Minimum value:	15																																																																																																									
Median value:	240																																																																																																									
Maximum value:	240																																																																																																									
Sum:	2,562,735																																																																																																									
Average:	239.98																																																																																																									
Variance:	4.74																																																																																																									
Standard deviation:	2.18																																																																																																									
Value	Count	Percentage																																																																																																								
240	10678	99.99%																																																																																																								
15	1	0.01%																																																																																																								
Expression:	AREAEDUCATIVATECNICA																																																																																																									
Datatype:	INTEGER																																																																																																									
Rows:	10,679																																																																																																									
Unique 10.01% Non-unique 0.02% Duplicate 99.98%		<table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>10,679</td></tr><tr><td>Null values:</td><td>0</td></tr><tr><td>Distinct values:</td><td>11</td></tr><tr><td>Duplicate values:</td><td>10,668</td></tr><tr><td>Unique values:</td><td>2</td></tr><tr><td>Non-unique values:</td><td>9</td></tr></tbody></table> <table><thead><tr><th>Type</th><th>value</th></tr></thead><tbody><tr><td>Minimum value:</td><td>Agropecuaria(NA)</td></tr><tr><td>Median value:</td><td>No aplica</td></tr><tr><td>Maximum value:</td><td>No aplica</td></tr><tr><td>Median length:</td><td>9</td></tr><tr><td>Average:</td><td>9.2</td></tr><tr><td>Maximum length:</td><td>60</td></tr></tbody></table>	Type	Count	Non-null values:	10,679	Null values:	0	Distinct values:	11	Duplicate values:	10,668	Unique values:	2	Non-unique values:	9	Type	value	Minimum value:	Agropecuaria(NA)	Median value:	No aplica	Maximum value:	No aplica	Median length:	9	Average:	9.2	Maximum length:	60	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td>No aplica</td><td>10415</td><td>97.53%</td></tr><tr><td>Otras</td><td>147</td><td>1.38%</td></tr><tr><td>Finanzas - Administración y Seguros</td><td>70</td><td>0.66%</td></tr><tr><td>Ciencias Naturales y Aplicadas</td><td>12</td><td>0.11%</td></tr><tr><td>Salud</td><td>12</td><td>0.11%</td></tr><tr><td>Operadores del Equipo y Transporte/Instalación y</td><td>10</td><td>0.09%</td></tr><tr><td>Agropecuaria(NA)</td><td>7</td><td>0.07%</td></tr><tr><td>Cultura, Arte, Procesamiento y Deportes</td><td>2</td><td>0.02%</td></tr><tr><td>Procesamiento y Ensamble</td><td>2</td><td>0.02%</td></tr><tr><td>Comercial(NA)</td><td>1</td><td>0.01%</td></tr><tr><td>Comunicación y Medios(NA)</td><td>1</td><td>0.01%</td></tr></tbody></table>	Value	Count	Percentage	No aplica	10415	97.53%	Otras	147	1.38%	Finanzas - Administración y Seguros	70	0.66%	Ciencias Naturales y Aplicadas	12	0.11%	Salud	12	0.11%	Operadores del Equipo y Transporte/Instalación y	10	0.09%	Agropecuaria(NA)	7	0.07%	Cultura, Arte, Procesamiento y Deportes	2	0.02%	Procesamiento y Ensamble	2	0.02%	Comercial(NA)	1	0.01%	Comunicación y Medios(NA)	1	0.01%	<table><thead><tr><th>Mask Analysis</th><th>Value</th><th>Count</th><th>Percent</th></tr></thead><tbody><tr><td>LL LLLLLL</td><td>10415</td><td>97.53%</td></tr><tr><td>LLLLL</td><td>159</td><td>1.49%</td></tr><tr><td>LLLLL</td><td>70</td><td>0.66%</td></tr><tr><td>LLLLL LLLLLL L</td><td>12</td><td>0.11%</td></tr><tr><td>LLLLL LLLLLL L</td><td>10</td><td>0.09%</td></tr><tr><td>LLLLL L</td><td>7</td><td>0.07%</td></tr><tr><td>LLLLL LLL</td><td>2</td><td>0.02%</td></tr><tr><td>LLLLL LLLLLL L</td><td>2</td><td>0.02%</td></tr><tr><td>LLLLL L</td><td>2</td><td>0.02%</td></tr><tr><td>LLLLL L</td><td>1</td><td>0.01%</td></tr><tr><td>LLLLL L</td><td>1</td><td>0.01%</td></tr></tbody></table>	Mask Analysis	Value	Count	Percent	LL LLLLLL	10415	97.53%	LLLLL	159	1.49%	LLLLL	70	0.66%	LLLLL LLLLLL L	12	0.11%	LLLLL LLLLLL L	10	0.09%	LLLLL L	7	0.07%	LLLLL LLL	2	0.02%	LLLLL LLLLLL L	2	0.02%	LLLLL L	2	0.02%	LLLLL L	1	0.01%	LLLLL L	1	0.01%	A. Es un campo de tipo STRING, con un total de 10679 registros analizados que presentan las siguientes características: - El 99,90% la información está diagnosticada como valores duplicados, es decir 10668 valores; el 0,08% = 9 valor no únicos, y, 11 valores distintos, 0,02% = 2 valores únicos. - De acuerdo al análisis de frecuencia los valores con mayor relevancia son: no aplica con un 97,53%, otras con un 1,38%, finanzas, administración y seguros con un 0,66%, ciencias naturales y aplicadas con un 0,11%, salud con un 0,11%, operadores del equipo y transporte instalación y mantenimiento con un 0,09%, agropecuaria con un 0,07%, cultura, arte, esparcimiento, deporte con un 0,02%, procesamiento, fabricación y ensamblaje con un 0,02%, comercial (NA), comunicación y medios con (NA) con un 0,01%. - El campo tiene definido 11 formatos de máscara de entrada diferentes para el set de datos. - El campo tiene definido 11 formatos de máscara de entrada diferentes para el set de datos.
Type	Count																																																																																																									
Non-null values:	10,679																																																																																																									
Null values:	0																																																																																																									
Distinct values:	11																																																																																																									
Duplicate values:	10,668																																																																																																									
Unique values:	2																																																																																																									
Non-unique values:	9																																																																																																									
Type	value																																																																																																									
Minimum value:	Agropecuaria(NA)																																																																																																									
Median value:	No aplica																																																																																																									
Maximum value:	No aplica																																																																																																									
Median length:	9																																																																																																									
Average:	9.2																																																																																																									
Maximum length:	60																																																																																																									
Value	Count	Percentage																																																																																																								
No aplica	10415	97.53%																																																																																																								
Otras	147	1.38%																																																																																																								
Finanzas - Administración y Seguros	70	0.66%																																																																																																								
Ciencias Naturales y Aplicadas	12	0.11%																																																																																																								
Salud	12	0.11%																																																																																																								
Operadores del Equipo y Transporte/Instalación y	10	0.09%																																																																																																								
Agropecuaria(NA)	7	0.07%																																																																																																								
Cultura, Arte, Procesamiento y Deportes	2	0.02%																																																																																																								
Procesamiento y Ensamble	2	0.02%																																																																																																								
Comercial(NA)	1	0.01%																																																																																																								
Comunicación y Medios(NA)	1	0.01%																																																																																																								
Mask Analysis	Value	Count	Percent																																																																																																							
LL LLLLLL	10415	97.53%																																																																																																								
LLLLL	159	1.49%																																																																																																								
LLLLL	70	0.66%																																																																																																								
LLLLL LLLLLL L	12	0.11%																																																																																																								
LLLLL LLLLLL L	10	0.09%																																																																																																								
LLLLL L	7	0.07%																																																																																																								
LLLLL LLL	2	0.02%																																																																																																								
LLLLL LLLLLL L	2	0.02%																																																																																																								
LLLLL L	2	0.02%																																																																																																								
LLLLL L	1	0.01%																																																																																																								
LLLLL L	1	0.01%																																																																																																								

Capítulo III: Minería de Datos

Es necesario aclarar que el concepto de teletrabajo en Colombia y las características de un teletrabajador aún sigue siendo un tema limitado, ya que es un proceso emergente y no puede dimensionarse debido a la falta de estadísticas, de instituciones y experiencia, aspectos que son necesarios para que se establezca un proceso de maduración y consolidación. La *figura 7* confirma que para el caso del sector educativo, como se ha mencionado en párrafos anteriores, no existen datos relevantes respecto al número de teletrabajadores.

Figura 7: Incremento del número de Teletrabajadores en los diferentes sectores (Molano, 2015)



Los datos corresponden al “*Estudio de penetración y percepción del teletrabajo en Colombia – 2014*” adelantado por la alianza entre los Ministerios TIC y del Trabajo, la Corporación Colombia Digital y el centro Nacional de Consultoría.

Al respecto, se puede afirmar que es fundamental medir y analizar las dinámicas organizacionales teniendo en cuenta el contexto del país, y de forma específica, el contexto del municipio de Medellín, y por consiguiente, sus políticas administrativas, aspectos, sociales, económicos, de movilidad, y aquellos relacionados con las circunstancias personales de los docentes.

Para aportar a esta concepción laboral, se halla un procedimiento para la selección de teletrabajadores que se direcciona según las sugerencias del Libro Blanco del Teletrabajo (MinTic, 2012) que indica:

Puede elegirse en dos vías. Por una parte se puede elegir personal de una única área y poner a prueba el modelo. Por la otra, la más aconsejable, se puede elegir personal de distintas áreas de la organización para así medir impactos de una forma más amplia. En esta decisión también será importante tener en cuenta los costos que se deberán asumir y la disponibilidad de presupuesto definida en el proceso de autoevaluación. (p.49)

Es decir, una organización debe determinar quién coordinará el proyecto piloto teniendo en cuenta los objetivos e indicadores para realizar la selección del personal teletrabajador.

Figura 8: Planeación de un piloto de teletrabajo, según el Libro Blanco del Teletrabajo (MinTic, 2012)



Siguiendo las indicaciones anteriores, este capítulo presenta la guía donde se organizan los grupos, es decir, se sectoriza la información extraída de las bases de datos con el objetivo de caracterizar la población docente como posibles teletrabajadores.

El procedimiento seguido con las herramientas de minería de datos permitió explorar variables en la búsqueda de información luego de aplicar un proceso de calidad de datos (Data Quality) que llevó a la conformación de un set con 26 variables, evaluadas en consideración con el cumplimiento de los objetivos del trabajo propuesto. “La organización podrá definir distintos procedimientos para evaluar si un empleado es apto o no para el teletrabajo e incluirlos dentro de los procesos de selección de personal ya establecidos.” (MinTic, 2012, p.52).

Se usó la herramienta de minería Weka mediante el método de Reglas de Asociación y *Clustering* para la construcción del modelo con base en las 26 variables y 10679 registros.

Tanto las Reglas de Asociación como los *Clustering* exigen que todas las variables o atributos numéricos sean categóricos para garantizar que la herramienta procese esos campos y se amplíen las posibilidades de extracción máxima de información para el modelo.

Reglas de Asociación.

El objetivo de las Reglas de Asociación es encontrar asociaciones o correlaciones entre los elementos u objetos de bases de datos transaccionales, relacionales o *Datawarehouse*.

Las Reglas de Asociación pertenecen a la rama de la minería de datos no supervisada y son utilizadas para este análisis con la necesidad de conservar el mismo estatus entre las variables, mantener su rol y evitar que algunas de ellas sean discriminadas; es así como esta técnica se convierte en una herramienta que permite unir la experiencia laboral, las competencias, la formación académica y la estabilidad laboral, entre otras, en la búsqueda del objetivo.

Por este motivo, luego del análisis, se decide que los atributos FECHAINGRESOMINA, BASICO, FECHANACIMIENTO no deben eliminarse ya que podrían estar presentes en el análisis al ser categorizados por rangos, de esta manera:

FECHAINGRESOMINA

- Antes de 1959: Docentes que ingresan a laborar antes de 1959, es decir, docentes con 56 años o más en servicio.

- Entre 1960 y 1969: Docentes que ingresan a nómina después del año 1960 y antes de 1969, es decir, docentes en servicio de 55 a 46 años.
- Entre 1970 y 1979: Docentes que ingresan a nómina después del año 1970 y antes de 1979, es decir, docentes en servicio de 36 a 45 años.
- Entre 1980 y 1989: Docentes que ingresan a nómina después del año 1980 y antes de 1989, es decir, docentes en servicio de 26 a 35 años.
- Entre 1990 y 1999: Docentes que ingresan a nómina después del año 1990 y antes de 1999, es decir, docentes en servicio de 16 a 25 años.
- Entre el 2000 y 2009: Docentes que ingresan a nómina después del año 2000 y antes de 2009, es decir, docentes en servicio de 6 a 15 años.
- Después del 2010: Docentes que ingresan a nómina después del año 2010, es decir, docentes en servicio de 6 años máximo.

BASICO

La información sobre el salario que devenga cada docente se reporta así:

- Hasta 1.000.000
- Entre 1.000.000 y 1.500.000
- Entre 1.500.000 y 2.000.000
- Entre 2.000.000 y 2.500.000
- Entre 2.500.000 y 3.000.000
- Entre 3.000.000 y 3.500.000
- Más de 3.500.000

FECHANACIMIENTO

Si bien el campo es una fecha, la edad es el dato relevante para el análisis. Por lo tanto se hacen dos procesos. En primer lugar, se transforma la fecha en el número de años que pasaron y, en segundo lugar, se categoriza así:

- Mayor a 56 años: Docentes con edades superiores a 56 años de edad
- Entre 46 y 55 años de edad: Docentes con edades entre 46 y 55 años de edad.
- Entre 36 y 45 años de edad: Docentes con edades entre 36 y 45 años de edad.
- Entre los 26 y 35 años de edad: Docentes con edades entre 26 y 35 años de edad.
- Inferior a 25 años de edad: Docentes con edades inferiores a 25 años de edad.

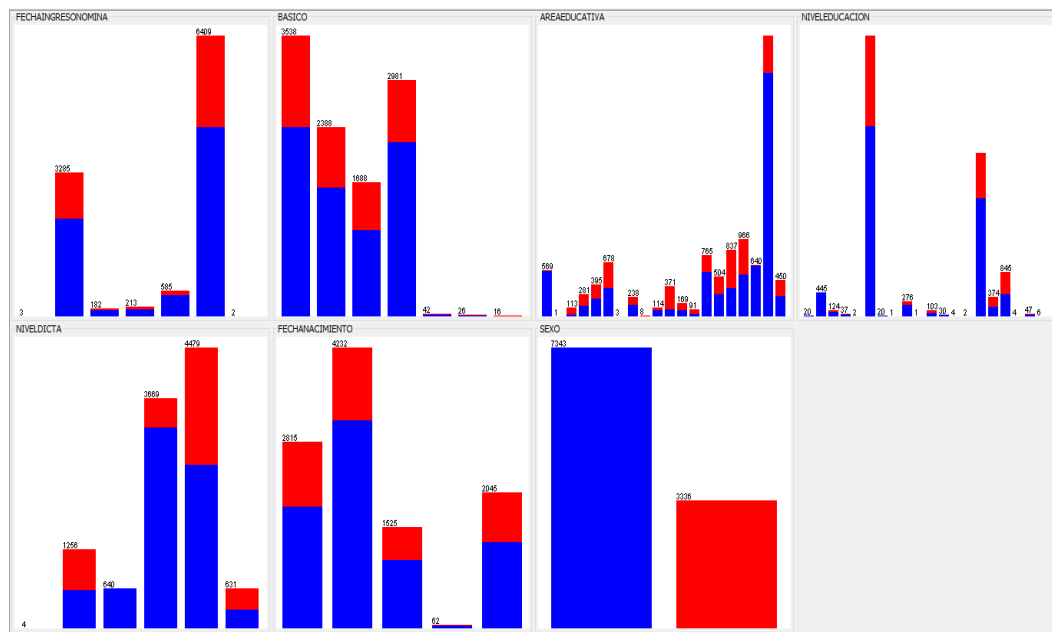
En forma simultánea, se hace un reconocimiento de variables que no aportan al estudio en curso por contener un único valor o porcentajes en algunos datos que no superan el 1% de la información total de la variable. Se toma la decisión de eliminarlos del set de datos minable:

VINCULACIONCATEGORIA, VINCULACIONESTADO, CENTRO COSTO,
 CODTURNO, FUENTESDERECURSOSPARAPAGO, CIUDAD, HORASTRAB,
 GRADO, CODCALENDARIO, CODCARGOEMPRESA,
 CODDEDICACIONLABORAL, REGIMENDEVINCULACION,
 SITUACIONLABORAL, AREAEDUCATIVATECNICA,

CONDICIONAMENAZADO, VACIONESMETODO, NIVELCONTRATACION,
VINCULACION NOVEDAD, CODEMPLEADO.

Para el logro de los objetivos se consideran siete variables, y se observa qué Reglas de Asociación extraer de ellas:

Figura 9: Visualización del set de datos final para evaluar Reglas de Asociación

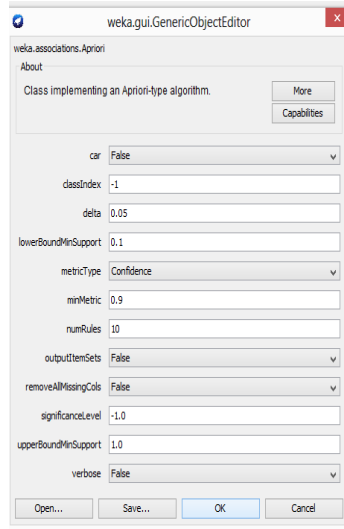


La decisión para elegir estas variables se basa en el análisis de calidad de la información, ninguna variable fue descartada por fuera del proceso sin justificación en dificultades de calidad o que su contenido no superara el 1% de la variable.

A continuación se presentan las Reglas de Asociación que este algoritmo extrae. En minería de datos el soporte y la confianza son las que evalúan las Reglas; como punto de partida en la configuración del algoritmo a priori, se dejan por defecto los valores predeterminados de la aplicación en este proceso de análisis.

Algoritmo a priori 1.

Tabla 25: Configuración del algoritmo

Algoritmo a priori 1	Configuración por defecto de Algoritmo a priori para Minería de datos, con la métrica Confianza.	Reglas de Asociación obtenidas al utilizar como número de regla 10 y la métrica Confianza
Run information === Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Relation: Setdatosfinal- weka.filters.unsupervised.attribute.Remove -R1,3-4,6-7,11-12,15 Instances: 10679 Attributes: 7 === Associator model (full training set) ===		<pre> 1 AREAEDUCATIVA=Primaria 3486 ==> NIVELDICTA=Primaria 3486 conf(1) 2 AREAEDUCATIVA=Primaria SEXO=F 3024 ==> NIVELDICTA=Primaria 3024 conf(1) 3 FECHAINGRESONOMINA=Entree2000y2009 AREAEDUCATIVA=Primaria 2187 ==> NIVELDICTA=Primaria 2187 conf(1) 4 FECHAINGRESONOMINA=Entree2000y2009 AREAEDUCATIVA=Primaria SEXO=F 1894 ==> NIVELDICTA=Primaria 1894 conf(1) 5 AREAEDUCATIVA=Primaria NIVELDICTA=Licenciado 1798 ==> NIVELDICTA=Primaria 1798 conf(1) 6 FECHAINGRESONOMINA=Entree2000y2009 NIVELDICTA=Primaria 2257 ==> AREAEDUCATIVA=Primaria 2187 conf(0.97) 7 FECHAINGRESONOMINA=Entree2000y2009 NIVELDICTA=Primaria SEXO=F 1962 ==> AREAEDUCATIVA=Primaria 1894 conf(0.97) 8 NIVELDICTA=Primaria 3609 ==> AREAEDUCATIVA=Primaria 3486 conf(0.95) 9 NIVELDICTA=Primaria SEXO=F 3201 ==> AREAEDUCATIVA=Primaria 3024 conf(0.94) 10 NIVELDICTA=Licenciado NIVELDICTA=Primaria 1927 ==> AREAEDUCATIVA=Primaria 1798 conf(0.95) </pre>

De las reglas 1 y 8; 2 y 9 se extrae que el total de docentes que pertenecen al AREAEDUCATIVA primaria y que dictan clases en este mismo nivel corresponde al 32.64% del total de docentes. El 86.7% corresponde a género femenino.

El total de docentes que ingresan a nómina en el grado y nivel primaria entre el año 2000 y 2009 equivalen al 62.73% del grupo de docentes en área y nivel primaria. (Regla 3 y 6).

La regla 4 y 7 indica que el total de docentes del área y nivel primaria que ingresaron entre el año 2000 y el 2009 a nómina, equivalen al 54.33% de los docentes de primaria, pertenecen al género femenino y llevan trabajando entre 6 y 15 años.

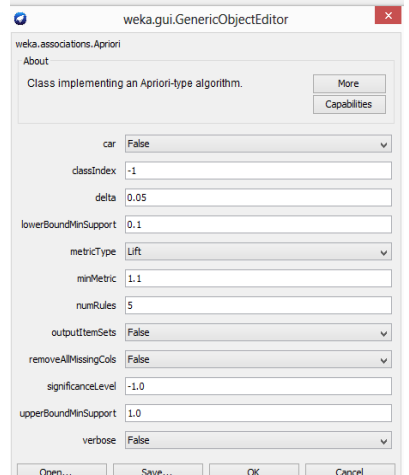
La regla 5 y 10 indica que el 51.57% de los docentes que pertenecen al área y nivel primaria tienen un nivel de educación de pregrado en licenciatura.

Es importante mencionar que la regla obedece a una confianza de 1% para las reglas 1, 2, 3, 4, para la regla 5 del 97%, para la regla 6 del 97%, para la regla 7 del 97%, para la regla 8 del 95%, para la regla 9 del 94% y para la regla 10 del 93%, en sus respectivos análisis.

El algoritmo descarta las variables básicas y fecha de nacimiento en el análisis.

Algoritmo a priori 2.

Tabla 26: Configuración del algoritmo

Algoritmo a priori 2	Configuración por defecto de algoritmo a priori para minería de datos, con la métrica Lift.	Reglas de Asociación obtenidas utilizando como número de reglas 5 y la métrica Lift
<pre> ==== Run information ==== Scheme: weka.associations.Apriori -N 5 -T 1 -C 1.1 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Relation: Setdatosfinal- weka.filters.unsupervised.attribute. Remove-R1,3-4,6-7,11-12,15 Instances: 10679 Attributes: 7 ==== Associator model (full training set) ==== </pre>		<pre> 1. AREAEDUCATIVA=Primaria 3486 ==> NIVELDICTA=Primaria 3486 conf:(1) < lift:(2.91)> lev:(0.21) [2288] conv:(2288.31) 2. NIVELDICTA=Primaria 3669 ==> AREAEDUCATIVA=Primaria 3486 conf:(0.95) < lift:(2.91)> lev:(0.21) [2288] conv:(13.43) 3. NIVELDICTA=Primaria 3669 ==> AREAEDUCATIVA=Primaria SEXO=F 3024 conf:(0.82) < lift:(2.91)> lev:(0.19) [1985] conv:(4.07) 4. AREAEDUCATIVA=Primaria SEXO=F 3024 ==> NIVELDICTA=Primaria 3024 conf:(1) < lift:(2.91)> lev:(0.19) [1985] conv:(1985.04) 5. AREAEDUCATIVA=Primaria 3486 ==> NIVELDICTA=Primaria SEXO=F 3024 conf:(0.87) < lift:(2.89)> lev:(0.19) [1979] conv:(5.27) </pre>

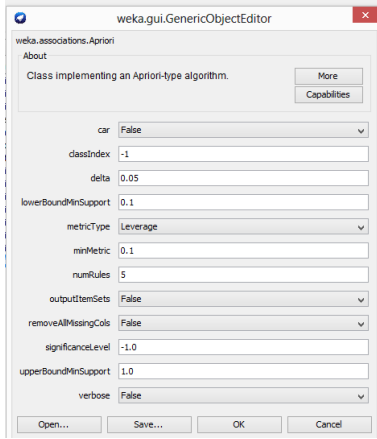
Al evaluar el set de datos con el uso de métrica Elevación o *Lift* se verifica que su medida sea superior a 1 y luego se evalúa al tomar los valores por defecto y permitir que el número de reglas sea de 5.

De las reglas 1, 2 y 3 se extrae que el total de docentes que pertenecen al AREAEDUCATIVA primaria es del 32.64% y que quienes dictan clases en este mismo nivel corresponden al 34.35% del total de docentes. El 86.7% corresponden al género femenino, y son evaluados con una confianza por encima del 82%. Las reglas 4 y 5 resumen las 3 primeras reglas.

Los resultados son similares a los obtenidos con la métrica de Confianza configurada en el ejercicio anterior, donde se destaca la participación de los docentes en el nivel y área de primaria pertenecientes al género femenino. Se evidencia que la tendencia se mantiene sobre las reglas iniciales.

Algoritmo a priori 3.

Tabla 27: Configuración del algoritmo

Algoritmo a priori 3	Configuración por defecto de algoritmo A priori para minería de datos, con la métrica <i>Leverage</i>	Reglas de Asociación obtenidas utilizando como número de reglas 5 y la métrica <i>Leverage</i>
<pre> === Run information === Scheme: weka.associations.Apriori -N 5 -T 2 -C 0.1 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Relation: Setdatosfinal- weka.filters.unsupervised.attrib ute.Remove-R1,3-4,6-7,11- 12,15 Instances: 10679 Attributes: 7 === Associator model (full training set) === Apriori </pre>		<pre> 1. AREAEDUCATIVA=Primaria 3486 ==> NIVELDICTA=Primaria 3486 conf:(1) lift:(2.91) < lev:(0.21) [2288]> conv:(2288.31) 2. NIVELDICTA=Primaria 3669 ==> AREAEDUCATIVA=Primaria 3486 conf:(0.95) lift:(2.91) < lev:(0.21) [2288]> conv:(13.43) 3. NIVELDICTA=Primaria 3669 ==> AREAEDUCATIVA=Primaria SEXO=F 3024 conf:(0.82) lift:(2.91) < lev:(0.19) [1985]> conv:(4.07) 4. AREAEDUCATIVA=Primaria SEXO=F 3024 ==> NIVELDICTA=Primaria 3024 conf:(1) lift:(2.91) < lev:(0.19) [1985]> conv:(1985.04) 5. AREAEDUCATIVA=Primaria 3486 ==> NIVELDICTA=Primaria SEXO=F 3024 conf:(0.87) lift:(2.89) < lev:(0.19) [1979]> conv:(5.27) </pre>

Al evaluar el set de datos que utiliza la métrica *Leverage* (cobertura, apalancamiento) se toman los valores por defecto, lo que permite que el número de reglas sea de 5.

La tendencia se mantiene en las Reglas de Asociación y muestra una evidente relación entre los docentes del AREAEDUCATIVA primaria y nivel de primaria que pertenecen en su mayoría al género femenino.

De las reglas 1, 2 y 3 se extrae que el total de docentes que pertenecen al AREAEDUCATIVA primaria es del 32.64%, quienes dictan clases en este mismo nivel corresponden al 34.35% del total de docentes y de este grupo el 86.7% es de género femenino, evaluados con una confianza por encima del 82%. Las reglas 4 y 5 resumen las 3 primeras reglas.

En concordancia con la información presentada, en la *Tabla 28* se presenta la síntesis de las Reglas de Asociación para la caracterización docente teniendo en cuenta los criterios de selección.

Tabla 28: Tabla de caracterización docente según criterios de selección de teletrabajadores con el uso de Reglas de Asociación

MÉTRICA	CONDICIONES EXCEPCIONALES	ANTIGÜEDAD EN LA ORGANIZACIÓN	POBLACIÓN OBJETIVO	GRADO DE GESTIÓN	ÁREAS SUSCEPTIBLES DE SER TRABAJADAS DESDE LA CASA
Reglas de Asociación obtenidas utilizando como número de regla 10 y la métrica <i>Confianza</i>	Del 32.64% de AREAEDUCATIVA primaria, El 86.7% corresponde a género femenino.	El total de docentes que ingresan a nómina en el grado y nivel primaria entre el año 2000 y 2009 equivalen al 62.73%. (Regla 3 y 6).	De las reglas 1 y 8; 2 y 9 se extrae que el total de docentes que pertenecen al AREAEDUCATIVA primaria y que dictan clases en este mismo nivel corresponde al 32.64% del total de docentes.	La regla 5 y 10 indica que el 51.57% de los docentes que pertenecen al área y nivel primaria tienen un nivel de educación de pregrado en licenciatura.	N/A
	Del total de docentes que ingresaron en 2009 el 54.33% de los docentes de primaria, pertenecen al género femenino.	N/A	N/A	N/A	N/A
Reglas de asociación obtenidas utilizando como número de reglas 5 y la métrica <i>Lift</i>	El 86.7% del 34.35% que pertenecen al AREAEDUCATIVA primaria corresponden al género femenino.	N/A	De las reglas 1, 2 y 3 se extrae que el total de docentes que pertenecen al AREAEDUCATIVA primaria es del 32.64% y que quienes dictan clases en este mismo nivel corresponden al 34.35% del total de docentes.	N/A	N/A
Reglas de Asociación obtenidas utilizando como número de reglas 5 y la métrica <i>Leverage</i>	El 86.7% del 34.35% que pertenecen al AREAEDUCATIVA primaria es de género femenino.	De las regla 1, 2 y 3 podemos extraer que el total de docentes que pertenecen al AREAEDUCATIVA primaria es del 32.64%, quienes dictan clases en este mismo nivel corresponden al 34.35% del total de docentes.	N/A	N/A	N/A

Clustering, algoritmo EM.

Los resultados limitados en el número de Reglas de Asociación implican una segunda evaluación con minería de datos *Clustering* que permite hacer un modelo de agrupamiento para partir la población en pequeños grupos y hallar conocimiento útil.

El análisis se desarrolla con 8 clústeres, luego de hacer la evaluación con 4 números diferentes de *Clustering* considerados idóneos para el negocio.

Figura 10: Configuración por defecto de algoritmo EM para Minería de datos con un número de 8 clustering o grupos

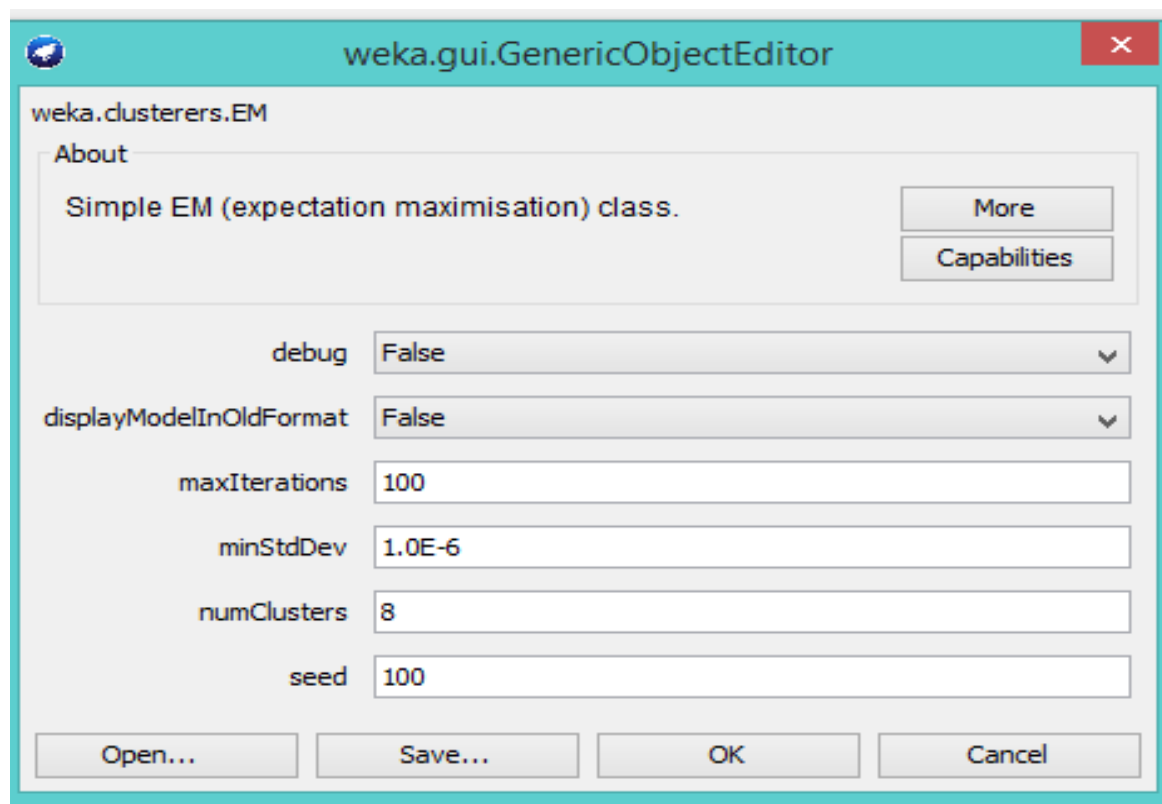


Tabla 29: Información relevante de la segmentación de clustering

Attribute	Cluster				
	0 (0.12)	1 (0.2)	2 (0.19)	4 (0.18)	7 (0.12)
FECHAINGRESOMINA					
Despuesdel2010		1529.5752			804.3641
Entreel2000y2009	954.1765		1711.997	1239.1927	
[total]	1257.8415	2126.8243	2041.0011	1893.9764	1257.7428
BASICO					
Entre1.000.000y1.500.000		1571.6109			1211.0475
Entre2.500.000y3.000.000	1156.3427		998.3937	626.0122	
[total]	1257.8415	2126.8243	2041.0011	1893.9764	1257.7428
AREAEDUCATIVA					
CienciasNaturalesyEdu.Ambiental				347.0335	
CienciasSociales			414.8544		
Matematicas		449.2989			
Primaria	1010.2631				1117.1102
[total]	1270.8415	2126.8243	2041.0011	1906.9764	1257.7428
NIVELEDCACION					
Licenciado		1381.5595	967.4036	793.6909	828.5588
PostgradoEspecializacion	957.1122				
[total]	1252.8415	2126.8243	2041.0011	1906.9764	1257.7428
NIVELICTA					
Media				1256.4007	
Primaria	1024.2732				1249.7136
Secundaria		1943.2784	2034.8075		
[total]	1252.8415	2126.8243	2041.0011	1892.9764	1256.7428
FECHANACIMIENTO					
Entre36y45añosdeedad		787.3709			407.4716
Entre46y55añosdeedad	723.0125		910.3951	697.2678	
[total]	1252.8415	2126.8243	2041.0011	1891.9764	1255.7428
SEXO					
F	1152.4962	1206.5032	1072.6416		1053.6335
M				987.9463	
[total]	1252.8415	2126.8243	2041.0011	1888.9764	1252.7428

De los 8 clústeres generados por el algoritmo, se seleccionan los clústeres 0, 1, 2, 4 y 7 que representan el 82 % de la información.

Dicha información, según los objetivos de ésta investigación, puede convertirse en una herramienta de apoyo para el equipo coordinador del Proyecto Piloto de Teletrabajo (se denomina “piloto”, ya que en la Secretaría de Educación aún no se implementa el teletrabajo). Por lo tanto, se generan unos grupos susceptibles de la realización del proyecto de teletrabajo que se caracterizan de acuerdo a los clústeres, así:

- Clúster #0= De 1257 docentes, 954 ingresaron a nómina entre el 2000 y el 2009, 1156 reciben un salario básico que oscila entre 2.500.000 y 3.000.000, 1010 pertenecen al área educativa primaria, 957 tienen un nivel de educación de posgrado con especialización, 1024 pertenecen al nivel de primaria, 723 oscilan entre las edades de 46 y 55 años de edad y 1152 son del género femenino.
- Clúster #1 = De 2126 docentes, 1529 ingresaron a nómina después del 2010, 1571 tienen un salario básico que oscila entre 1.000.000 y 1.500.000, 449, es decir, el 21% de los docentes pertenecen al área educativa matemáticas, 1381 tienen un nivel de educación en licenciatura, 1943 pertenecen al nivel de secundaria, 787 oscilan entre las edades de 36 y 45 años de edad y 1206 son del género femenino.
- Clúster #2 = De 2041 docentes, 1711 ingresaron a nómina entre los años 2000 y 2009, 998 tienen un salario básico que oscila entre 2.500.000 y 3.000.000, 414 pertenecen al área educativa ciencias sociales, 967 tienen un nivel de educación de pregrado en licenciatura, 2034 pertenecen al nivel de secundaria, 910 oscilan entre las edades de 46 y 55 años de edad y 10722 son del género femenino.
- Clúster #4 = De 1893 docentes, 1239 ingresaron a nómina entre el 2000 y el 2009, 626 tienen un salario básico que oscila entre 2.500.000 y 3.000.000, es decir, el 33%. De 1906 docentes, 347 pertenecen al área educativa ciencias naturales y educación ambiental, 793 tienen un nivel de educación de pregrado en licenciatura, 1256 pertenecen al nivel de media académica, 697 oscilan entre las edades de 46 y 55 años de edad y 987 son del género masculino, es decir el 52%.

- Clúster #7 = De 1257 docentes, 804 ingresaron a nómina después del año 2010, 1211 tienen un salario básico que oscila entre 1.000.000 y 1.500.000, 1117 pertenecen al área educativa primaria, 828 tienen un nivel de educación de pregrado en licenciatura, 1249 pertenecen al nivel de primaria, 787 oscilan entre las edades de 36 y 45 años de edad y 1053 son del género femenino.

En concordancia con los datos que se presentaron, en la *Tabla 30* se presenta la síntesis del *Clustering* para la caracterización docente teniendo en cuenta los criterios de selección

Tabla 30: Caracterización docente según criterios de selección de teletrabajadores con el uso de Clustering

CLUSTERS	CONDICIONES EXCEPCIONALES	ANTIGÜEDAD EN LA ORGANIZACIÓN	POBLACIÓN OBJETIVO	GRADO DE GESTIÓN	EXISTENCIA DE ÁREAS SUSCEPTIBLES DE SER TRABAJADAS DESDE LA CASA
Clúster #0= De 1257	1152 son del género femenino.	954 ingresaron a nómina entre el 2000 y el 2009	723 oscilan entre las edades de 46 y 55 años de edad.	957 tienen un nivel de educación de posgrado con especialización.	N/A
			1156 reciben un salario básico que oscila entre 2.500.000 y 3.000.000		
			1010 pertenecen al área educativa primaria.		
			1024 pertenecen al nivel de primaria.		
Clúster #1 = De 2126 docentes	1206 son del género femenino.	1529 ingresaron a nómina después del 2010.	1.571 tienen un salario básico que oscila entre 1.000.000 y 1.500.000	1381 tienen un nivel de educación de pregrado en licenciatura.	449, es decir, el 21% de los docentes pertenecen al área educativa

			787 oscilan entre las edades de 36 y 45 años de edad.		Matemáticas.
			1943 pertenecen al nivel de secundaria.		
Clúster #2 = De 2041 docentes.	10722 son del género femenino.	1711 ingresaron a nómina entre los años 2000 y 2009.	998 tienen un salario básico que oscila entre 2.500.000 y 3.000.000	967 tienen un nivel de educación de pregrado en licenciatura.	414 pertenecen al área educativa Ciencias Sociales.
			2034 pertenecen al nivel de secundaria.		
			910 oscilan entre las edades de 46 y 55 años de edad.		
Clúster #4 = De 1893 docentes	987 son del género masculino, es decir el 52%.	1239 ingresaron a nómina entre el 2000 y el 2009.	626 tienen un salario básico que oscila entre 2.500.000 y 3.000.000 es decir, el 33%	793 tienen un nivel de educación de pregrado en licenciatura.	De 1906 docentes, 347 pertenecen al área educativa Ciencias Naturales y Educación Ambiental.
			697 oscilan entre las edades de 46 y 55 años de edad.		1256 pertenecen al nivel de media académica.
Clúster #7 = De 1257 docentes.	1053 son del género femenino.	804 ingresaron a nómina después del año 2010	1211 tienen un salario básico que oscila entre 1.000.000 y 1.500.000	828 tienen un nivel de educación de pregrado en licenciatura.	N/A
			1117 pertenecen al área educativa primaria.		
			1249 pertenecen al nivel de primaria.		
			787 oscilan entre las edades de 36 y 45 años de edad.		

Importancia del teletrabajo para docentes

En el año 2006, la Ministra de Educación Cecilia María Vélez expuso en carta de apertura a la presentación de Estándares Básicos de Competencias en Lenguaje, Matemáticas, Ciencias y Ciudadanas, desarrollados por el Ministerio de Educación, lo siguiente:

El Gobierno nacional se propuso la tarea de adelantar una Revolución Educativa y la fijó como la primera de sus herramientas de equidad social, con el pleno convencimiento de que la educación es el camino para garantizar la paz, asegurar la igualdad de oportunidades y contribuir al desarrollo del país. De esta manera, incluyó como ejes de su plan de acción, el aumento de cobertura, el mejoramiento de la calidad de la educación y mayores niveles de eficiencia en el sector (Ministerio de Educación Nacional -MEN-, 2006, p.6).

En consecuencia, los retos que el Gobierno propone frente a la educación requieren de cambios en la infraestructura, aumento del servicio educativo, incremento en el número de docentes, vinculación a propuestas laborales en el marco de la ciencia y la tecnología y la satisfacción de las necesidades que tienen los jóvenes y niños del país en términos calidad.

Bajo este panorama, es evidente que el sector económico es el primero que debe generar recursos para alcanzar dicho propósito, además el Gobierno tiene una responsabilidad social y cultural respecto a la creación de un ambiente adecuado y de inclusión que avance al ritmo de los desarrollos actuales.

Por ello, hoy es innegable la necesidad de incluir propuestas que contribuyan a la construcción de ciudades inteligentes. Una de las propuestas de innovación social es el teletrabajo que, una vez implementado, tiene el potencial para aportar a la Revolución Educativa que plantea el Ministerio como un reto educativo. Como lo manifiesta el consultor educativo Curtis Johnson (2011) en la entrevista realizada por Eduardo Punset, la educación está viviendo un proceso de innovación disruptiva:

....De alguna manera el modelo vigente en el sector educativo está casi completamente desvinculado de la realidad del siglo 21; el futuro es digital, y los que pueblan ahora las escuelas pertenecen a esa esfera, el sistema educativo está en crisis por muchas razones y su modificación paulatina es demasiado lenta para la disrupción que las nuevas tecnologías está introduciendo en las aulas. Es difícil seguir adoptando el modelo de una clase estática con un profesor al frente de un grupo de alumnos que lo escuchan atentos anotando en sus libretas. El cambio está rompiendo la jerarquía profesor alumno por un sistema más cooperativo entre los estudiantes que trabajan en equipo y se supervisan entre sí, bajo la mirada del maestro, las nuevas tecnologías nos están desvinculando progresivamente de la escuela como centro de conocimiento, por un sistema más abierto, el del internet en el que una información muy dispar es accesible desde cualquier punto del planeta. El aprendizaje de la nueva generación también debe ser fluido y multidimensional, el uso de las nuevas tecnologías ha acostumbrado a los alumnos a mezclar conocimientos y no responden bien a nuestra manera lineal de presentarles las cosas; la tendencia en educación es hacia la personalización, el alumno mismo elegirá la información con la que desee trabajar y administrara su ritmo de aprendizaje bajo la supervisión de un tutor, quien habrá de seguir también el desarrollo emocional de sus estudiantes..... Por último, el aula perderá su entidad física, ahora podemos aprender *on-line* química desde cualquier lugar del mundo y en el mejor de los laboratorios,

hacer una excursión al museo de historia natural de Londres o aprender a tocar un instrumento musical desde casa, aún más, podemos compartir nuestra experiencia con los demás gracias a internet.

Como lo plantea el autor, la existencia física del aula desaparecerá de forma paulatina porque la información y, en consecuencia, quien enseña puede estar en cualquier lugar aprovechando las múltiples herramientas *on-line*, servir de tutor con una cobertura mayor, optimizar los centros de estudio y la infraestructura, ahorrando tiempo y dinero. De esta forma se establece un puente entre la vida laboral y personal del docente al hacer uso de plataformas tecnológicas que permiten atender diferentes tipos de estudiantes y necesidades y sobretodo, permite hacer accesible la educación por medio de la flexibilidad, una característica propia del ritmo de vida de las sociedades actuales.

Por lo tanto, se necesita una política que ordene y de forma a este tipo de necesidades. Al respecto,

El pasado 26 de julio de 2013, en el jardín botánico de Medellín, se firmó el primer Pacto Regional por el Teletrabajo, alianza entre el sector público y privado de Antioquia para promover la implementación de modelos laborales flexibles en tiempo y espacio, punto de partida para generar un cambio organizacional y modernización de las organizaciones de la región. (Molano, 2013)

En conclusión, la educación tiene grandes retos desde el punto de vista económico, cultural y personal; el docente se convierte en uno de los ejes principales de transformación con respecto a la reducción de costos, la reorganización de grupos, la optimización del

tiempo, la reducción de problemas como el ausentismo, la rotación, entre otros aspectos.

Como resultado, será posible obtener mayor eficiencia y productividad para el cumplimiento de objetivos.

CONCLUSIONES

- De acuerdo con los resultados obtenidos al usar el Algoritmo a priori y el análisis con la técnica de *Clustering*, se indica que las Reglas de Asociación y los grupos resultantes sirven como apoyo para la toma de decisiones y formulación de estrategias por parte del equipo coordinador del Proyecto de teletrabajo.
- El análisis realizado con ambas herramientas (Reglas de Asociación y *Clustering*) demuestra que existe una cantidad considerable de población femenina. Por lo tanto, es un grupo susceptible para teletrabajar si se tienen en cuenta características enmarcadas dentro de las condiciones excepcionales, la antigüedad en la empresa, su grado de gestión o si es una población objetivo.
- Los análisis con el Algoritmo EM de *clustering* arrojan importantes grupos de posibles teletrabajadores que pueden ser atendidos por parte del equipo coordinador del Proyecto de teletrabajo, según los perfiles extraídos de las variables edad, salario, nivel de educación, nivel que dicta, sexo, área educativa y la fecha de ingreso a la nómina.
- Se obtiene un documento para perfilar a un docente como posible teletrabajador. Sin embargo, es necesario ampliar la información con datos que evidencien relaciones entre competencias laborales y personales, rasgos de personalidad, valores, hobbies, entorno familiar, entorno laboral, trayectoria laboral, historial de formación, entre otros que den

cuenta de un espectro más amplio sobre el maestro; por otro lado, se generan unas necesidades latentes en la consecución de nueva información para el gremio.

- Las herramientas de inteligencia de negocios corresponden a un proceso de perfilación de docentes para teletrabajo que permite el uso de procesos de selección, exploración, modificación, modelización del comportamiento, evaluación y presentación gráfica de forma eficiente para el gremio. La información recolectada se convierte en una herramienta vital en el proceso de creación de un piloto en teletrabajo para la Secretaría de Educación de Medellín.
- La minería de datos se convierte en una herramienta de carácter útil en el análisis de áreas de corte social, y trasciende su carácter de aplicabilidad.
- La información obtenida fortalece el proceso de flexibilización laboral a largo plazo porque sin un antecedente como éste no es posible justificar la necesidad de adquirir nueva información en poblaciones como el gremio docente. Es necesario tener en cuenta que el teletrabajo como propuesta innovadora es muy reciente en Colombia; pero se solo se decretó desde el 2012 y en el municipio de Medellín se firmó el Plan Maestro de Teletrabajo desde el 28 de abril del 2015.

TRABAJOS FUTUROS

- Es necesario realizar una nueva evaluación de procesos para la selección de posibles teletrabajadores en una investigación futura, por medio de otras técnicas de minería de datos que permitan implementar o crear nuevos modelos para el aprovechamiento máximo de la información por parte del equipo coordinador de teletrabajo en la Secretaría de Educación de Medellín.
- Como resultado de este trabajo, se evidencia la necesidad de construir una base de datos de docentes con información suficiente que muestre las relaciones entre competencias laborales y personales, rasgos de personalidad, valores, hobbies, entorno familiar, entorno laboral, trayectoria laboral, historial de formación, entre otros; y que, en consecuencia, permita hacer un perfilamiento de docentes para teletrabajo ya que sin este tipo de base de datos no es posible obtener un avance significativo frente a dicho propósito.
- Con el presente trabajo queda abierta la propuesta para implementar pilotos de teletrabajo con el gremio de los docentes del municipio de Medellín.

REFERENCIAS

- Amón Uribe, Iván (2010) *Guía metodológica para la selección de técnicas de depuración de datos*. Universidad Nacional de Colombia, Medellín. Recuperado en abril de 2014, de <http://www.bdigital.unal.edu.co/2033/1/71644758.20101.pdf>
- Boiarov, Sonia; Naum, Poliszuk (2011). *¿Qué es el teletrabajo?* Argentina, caminandoutopias.org. Recuperado el 10 de enero de 2016, de http://www.caminandoutopias.org.ar/institucional/que_es_el_teletrabajo.php
- Fayyad, U.; Piatetsky Shapiro, G.; Smyth, P. (1996). *From Data Mining to Knowledge Discovery in Databases*. Recuperado el 14 de mayo de 2015, de <http://www.csd.uwo.ca/faculty/ling/cs435/fayyad.pdf>
- Franco Jaramillo, A; Restrepo Bustamante, F. (2011) *El perfil del teletrabajador y su incidencia en el éxito laboral*. En: *Revista Virtual Universidad Católica Del Norte*, No. 33, (mayo-agosto de 2011, Colombia). ISSN 0124-5821. Recuperado en mayo de 2014, de <http://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/1/2>
- Gómez Medina, H.; Vargas Jiménez, T. (2012). *Diseño en Excel mediante herramientas de Business Intelligence de un modelo de diagnóstico financiero para una pequeña empresa*. Recuperado el 4 de abril de 2014, de [http://repository.udem.edu.co/bitstream/handle/11407/227/Dise%C3%B1o en Excel mediante herramientas de Business Intelligence de un modelo de diagn%C3%B3stico financiero para una peque%C3%B1a empresa.pdf?sequence=1&isAllowed=y](http://repository.udem.edu.co/bitstream/handle/11407/227/Dise%C3%B1o%20en%20Excel%20mediante%20herramientas%20de%20Business%20Intelligence%20de%20un%20modelo%20de%20diagn%C3%B3stico%20financiero%20para%20una%20peque%C3%B1a%20empresa.pdf?sequence=1&isAllowed=y)
- Herschel, Gareth; Sarner, Adam; Thompson, Ed. (2010) *Gartner CRM Summit 2010, March 16-17, London, Features Industry Leading Analysts*. Recuperado el 4 de mayo de 2014, de <https://www.youtube.com/watch?v=qxKdOFkJOCY>

- Institución Universitaria CESMAG (2014). *Teletrabajo*. Recuperado el 20 de enero de 2016, de <https://guswillsan.files.wordpress.com/2014/10/teletrabajo.pdf>
- Johnson, Curtis (2011) *La manera disruptiva de aprender*. En: Redes (Nº 102). Recuperado el 01 de septiembre de 2016, de <https://www.youtube.com/watch?v=NQFMArcQI5o&list=PL32834A97F99676DE>
- Ministerio de Educación Nacional -MEN- (2006) *Estándares Básicos de Competencias en Lenguaje, Matemáticas, Ciencias y Ciudadanas*. Colombia, MEN. Recuperado el 5 de agosto de 2016, de http://www.mineduacion.gov.co/1621/articles-116042_archivo_pdf.pdf
- Ministerio de Educación Nacional. (2007) *Con la tecnología, respuestas para múltiples necesidades*. Colombia. En: *Altablero* No. 43 (septiembre-diciembre 2007) Recuperado el 28 de abril de 2014, de <http://www.mineduacion.gov.co/1621/article-150544.html>
- Ministerio de las Tecnologías de la Información y las Comunicaciones -MinTic- (2012). *Libro Blanco: El ABC del teletrabajo en Colombia*. Recuperado el 28 de abril de 2014, de http://www.teletrabajo.gov.co/622/articles-8228_archivo_pdf_libro_blanco.pdf
- Molano, Adriana (2013) *Memorias: Firma Pacto por el Teletrabajo Antioquia*. En: Colombia Digital, agosto 13 de 2013. Recuperado el 27 de julio de 2016, <https://colombiadigital.net/teletrabajo/item/5503-memorias-firma-pacto-por-el-teletrabajo-antioquia.html>
- Molano, Adriana (2015) *Así va el teletrabajo en Colombia (2): Las empresas y el teletrabajo*. En Colombia Digital, mayo 11 de 2015. Recuperado el 20 de mayo de 2016, de <https://colombiadigital.net/actualidad/articulos-informativos/item/8301-asi-va-el-teletrabajo-en-colombia-2-las-empresas-y-el-teletrabajo.html>

Morales Gutiérrez, A.; Ariza Montes, J.; Muñiz Rodríguez, N. (2012). *El emprendedor social y el E-empowerment de las redes virtuales*. (Spanish). Social Entrepreneurs and E-Empowerment through Virtual Networks. (English), (75), 153–177. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=zbh&AN=89234211&lang=es&site=ehost-live>

National Information Systems Security Glossary (2014) *Sistemas de Información*.

Recuperado el 23 de diciembre de 2014, de

<http://america11.bligoo.com.mx/content/view/1363411/sistemas-de-informacion.html#.VJm2sV4A8>

Orlando, D., Frade, A., Pérez, N. (2007). *Estado actual de las tecnologías de bodega de datos y OLAP aplicadas a bases de datos espaciales*. Recuperado de <http://www.scielo.org.co/pdf/iei/v27n1/v27n1a08.pdf>

Segura Jiménez, Herlaynne (2014) *Teletrabajo: Consultoría en Colombia*. Recuperado el 02 de mayo de 2015, de <http://blog.torresdeteletrabajo.com/tag/herlaynne-segura-jimenez/>.

Sinnexus. (2007) *¿Qué es business_intelligence?* Coruña. Recuperado el 28 de abril de 2014, de http://www.sinnexus.com/business_intelligence/

Sinnexus. (2007). *Datamining (Minería de datos)*. Coruña. Recuperado el 28 de abril de 2014, de http://www.sinnexus.com/business_intelligence/datamining.aspx.

Universidad Nacional Abierta y a Distancia –UNAD– (2014). *Lección 7. Sistemas de apoyo a las decisiones*. Colombia. Recuperado en abril de 2015, de http://datateca.unad.edu.co/contenidos/210102/Modulo_del_Curso_SIGP/Modulo%20del%20Curso%20SIGP%20EXE/leccin_7_sistemas_de_apoyo_a_las_decisiones_12.html

Vasile, Gorgan; Mirela, Oancea (2008) *Data Quality in Business Intelligence Applications*.

Recuperado en agosto de 2015, de

<http://steconomiceuoradea.ro/anale/volume/2008/v4-management-marketing/250.pdf>

Vega, Carlos A. *et al* (2012) *Data Mining Aplicado a la Predicción y Tratamiento de Enfermedades*. Investigación, Universidad Popular Autónoma del Estado de Puebla, México.

Recuperado el 4 de mayo de 2014, de

http://www.iiis.org/CDs2012/CD2012SCI/CISCI_2012/PapersPdf/CA732OV.pdf

Webmining Consultores (2011) *KDD: Proceso de Extracción de conocimiento*. Recuperado

el 04 de mayo de 2014, de <http://www.webmining.cl/2011/01/proceso-de-extraccion-de-conocimiento/>

ANEXOS

ANEXO 1: Encuesta de caracterización y formación docente

Para el Municipio de Medellín es muy importante conocer un poco más las condiciones de vida de cada uno de los maestros y sus familias, por eso los invitamos a responder de forma objetiva el siguiente cuestionario, cuyos resultados serán utilizados para generar políticas y programas de bienestar y capacitación, orientados a su mejoramiento familiar, personal y profesional. Los datos suministrados son totalmente confidenciales y su uso está reglamentado por la norma legal vigente (Ley 1266 de 2008 y la sentencia C-748 de 2011).



Medellín
todos por la vida

CIRCULAR No. 050

Medellín, Agosto 26 de 2013

Para: Directivos Docentes y Maestros y Maestras de las Instituciones Educativas Oficiales del Municipio de Medellín

De: Secretaría de Educación

Asunto: Realización de encuesta de caracterización docente en el marco del Programa Maestros para la Vida

El Programa MAESTROS PARA LA VIDA del Plan de Desarrollo Municipal 2012-2015 busca implementar y diseñar políticas y estrategias integrales que posibiliten el buen desempeño social y laboral de los educadores de nuestras instituciones educativas. Con esto queremos destacar y agradecer la calidad y trascendencia de la labor docente que realizan día a día en la ciudad nuestros maestros y maestras en pro del mejoramiento de la calidad educativa.

En este marco del Programa MAESTROS PARA LA VIDA, se continuará con la realización de la encuesta censal de caracterización personal, profesional y familiar de los educadores oficiales del municipio.

Por lo anterior estamos invitando a todos los docentes y directivos docentes que no hayan diligenciado la encuesta, para que ingresen desde cualquier computador con acceso a internet al enlace <http://aplicaciones.medellin.gov.co/Docentes/EncuestaV1.14/>, siendo esta la manera ideal de diligenciamiento de la encuesta.

Si lo desea, también pueden diligenciarla en formato word desde cualquier computador. Para acceder a este, debe ingresar a la suite de recursos de la Secretaría de Educación, en el link de maestros para la vida <http://secremed.aprender.com.co/aprender/> y remitirla completamente diligenciada al correo electrónico laboratorio.educacion@medellin.gov.co, o imprimirla, diligenciarla completamente y remitirla de manera física a la Secretaría de Educación, Edificio Carré, oficina 204 Carrera 52 44 B 17.

Se recuerda que toda la información diligenciada es totalmente confidencial y que no hay respuestas correctas o incorrectas.

Cordialmente,


LUZ ELENA GAVIRIA LÓPEZ
Secretaría de Educación


Aprobó: Alexandra Páez Botero
Subsecretaría de Calidad

Proyectó: Tatiana Mora



Centro Administrativo Municipal - C.A.M. - Calle 44 No. 50 - 105
Edificio de la Secretaría de Educación 14 44 105 Correo Electrónico: secremed@medellin.gov.co Medellín - Colombia

La entidad más innovadora



Alcalde de Medellín

ANEXO 2: Informe dirigido a la Secretaría de Educación de Medellín

CARACTERIZACIÓN DE LOS DOCENTES DE LA SECRETARÍA DE EDUCACIÓN DE MEDELLÍN PARA EL PERFILAMIENTO EN TELETRABAJO

Mónica Patricia Escudero Cañas,
Maestría en Tecnologías de Información y Comunicación
Escuela de Ingeniería - Universidad Pontificia Bolivariana
monicapescudero@gmail.com

Resumen: En esta investigación se realizó un análisis de la información de la base de datos de docentes de la Secretaría de Educación de Medellín, a través del uso de herramientas de Inteligencia de Negocios como la minería de datos, con el objetivo de extraer reglas y grupos de docentes que se pudieran perfilar para la realización de proyectos piloto en teletrabajo.

Palabras Clave: teletrabajo; inteligencia de negocios; minería de datos; docentes; perfilamiento.

INTRODUCCIÓN

Esta propuesta buscó el desarrollo de un documento que apoye decisiones para ofrecer alternativas laborales a los docentes de la Secretaría de Educación y se sustenta en

iniciativas políticas y jurídicas como la celebrada el 26 de julio de 2013, en el jardín botánico:

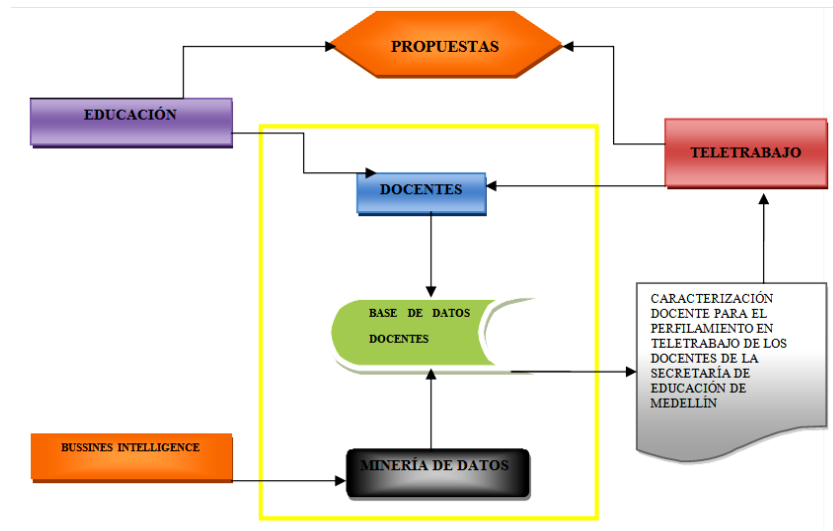
... Se firmó el primer Pacto Regional por el Teletrabajo, alianza entre el sector público y privado de Antioquia para promover la implementación de modelos laborales flexibles en tiempo y espacio, punto de partida para generar un cambio organizacional y modernización de las organizaciones de la región (Molano, 2013).

En este sentido, la investigación tiene un carácter social, tecnológico, sistémico, innovador y práctico. Así pues, se presenta una propuesta de flexibilidad laboral dirigida a los docentes como actores principales, a través del estudio de perfilamiento para el sector educativo que se une a una propuesta de optimización de medios y recursos como lo ofrece *Data Mining* en el manejo de la información.

¿QUÉ SE HIZO?

Se desarrolló un documento que muestra grupos y reglas de perfiles de docentes de la Secretaría de Educación de Medellín, caracterizados por sus condiciones excepcionales, la antigüedad en la empresa, la población objetivo, el grado de gestión y las áreas susceptibles de ser trabajadas desde la casa, mediante la utilización de herramientas de Inteligencia de Negocios que favorecen la implementación del teletrabajo como una alternativa laboral para el gremio.

Proceso de caracterización docente para teletrabajo del municipio de Medellín.



¿CÓMO SE HIZO?

La investigación contó con el aval de la Universidad Pontificia Bolivariana en asocio con la Secretaría de Educación de Medellín y se desarrolló en varias etapas:

1. Levantamiento de información que permita hacer una caracterización de las herramientas y metodologías de inteligencia de negocios para teletrabajo de los docentes del Municipio de Medellín.
2. Identificación de variables determinantes en el perfilamiento de los docentes para identificar bajo qué condiciones y qué profesor puede teletrabajar.
3. Consolidación de la información para la perfilación de los docentes de la Secretaría de Educación de Medellín mediante procesos de extracción, transformación y carga de la información de Inteligencia de Negocios.
4. Implementación del proceso de minería de datos para el perfilamiento de los docentes según las variables seleccionadas de manera previa, y que permiten obtener

información sobre la posible aplicación de teletrabajo en los docentes del Municipio de Medellín.

RESULTADOS:

Según los objetivos de esta investigación, la información puede convertirse en una herramienta de apoyo para un equipo coordinador de un Proyecto Piloto de Teletrabajo (se denomina “piloto”, ya que en la Secretaría de Educación aún no se implementa el teletrabajo). Por lo tanto, se generan unos grupos susceptibles de la realización del proyecto de teletrabajo que se caracterizan de acuerdo a unas métricas así:

Tabla de caracterización docente según criterios de selección de teletrabajadores con el uso de Reglas de Asociación

MÉTRICA	CONDICIONES EXCEPCIONALES	ANTIGÜEDAD EN LA ORGANIZACIÓN	POBLACIÓN OBJETIVO	GRADO DE GESTIÓN	ÁREAS SUSCEPTIBLES DE SER TRABAJADAS DESDE LA CASA
Reglas de Asociación obtenidas utilizando como número de regla 10 y la métrica Confianza	Del 32.64% de AREAEDUCATIVA primaria, El 86.7% corresponde a género femenino.	El total de docentes que ingresan a nómina en el grado y nivel primaria entre el año 2000 y 2009 equivalen al 62.73%.	El total de docentes que pertenecen al AREAEDUCATIVA primaria y que dictan clases en este mismo nivel corresponde al 32.64% del total de docentes.	El 51.57% de los docentes que pertenecen al área y nivel primaria tienen un nivel de educación de pregrado en licenciatura.	N/A
	Del total de docentes que ingresaron en 2009 el 54.33% de los docentes de primaria, pertenecen al género femenino.	N/A	N/A	N/A	N/A

Reglas de asociación obtenidas utilizando como número de reglas 5 y la métrica <i>Lift</i>	El 86.7% del 34.35% que pertenecen al AREAEDUCATIVA primaria corresponden al género femenino.	N/A	El total de docentes que pertenecen al AREAEDUCATIVA primaria es del 32.64% y que quienes dictan clases en este mismo nivel corresponden al 34.35% del total de docentes.	N/A	N/A
Reglas de Asociación obtenidas utilizando como número de reglas 5 y la métrica <i>Leverage</i>	El 86.7% del 34.35% que pertenecen al AREAEDUCATIVA primaria es de género femenino.	El total de docentes que pertenecen al AREAEDUCATIVA primaria es del 32.64%, quienes dictan clases en este mismo nivel corresponden al 34.35% del total de docentes.	N/A	N/A	N/A

Caracterización docente según criterios de selección de teletrabajadores con el uso de Clustering

CLUSTERS	CONDICIONES EXCEPCIONALES	ANTIGÜEDAD EN LA ORGANIZACIÓN	POBLACIÓN OBJETIVO	GRADO DE GESTIÓN	EXISTENCIA DE ÁREAS SUSCEPTIBLES DE SER TRABAJADAS DESDE LA CASA
Clúster #0 = De 1257	1152 son del género femenino.	954 ingresaron a nómina entre el 2000 y el 2009	723 oscilan entre las edades de 46 y 55 años de edad.	957 tienen un nivel de educación de posgrado con especialización.	N/A
			1156 reciben un salario básico que oscila entre 2.500.000 y 3.000.000		
			1010 pertenecen al área educativa primaria.		
			1024 pertenecen al nivel de primaria.		
Clúster #1 = De 2126 docentes	1206 son del género femenino.	1529 ingresaron a nómina después del 2010.	1.571 tienen un salario básico que oscila entre 1.000.000 y 1.500.000	1381 tienen un nivel de educación de pregrado en licenciatura.	449, es decir, el 21% de los docentes pertenecen al área educativa Matemáticas.
			787 oscilan entre las edades de 36 y 45 años de edad.		
			1943 pertenecen al nivel de secundaria.		
Clúster #2 = De 2041 docentes.	10722 son del género femenino.	1711 ingresaron a nómina entre los años 2000 y 2009.	998 tienen un salario básico que oscila entre 2.500.000 y 3.000.000	967 tienen un nivel de educación de pregrado en licenciatura.	414 pertenecen al área educativa Ciencias Sociales.
			2034 pertenecen al nivel de secundaria.		

			910 oscilan entre las edades de 46 y 55 años de edad.		
Clúster #4 = De 1893 docentes	987 son del género masculino, es decir el 52%.	1239 ingresaron a nómina entre el 2000 y el 2009.	626 tienen un salario básico que oscila entre 2.500.000 y 3.000.000 es decir, el 33%	793 tienen un nivel de educación de pregrado en licenciatura.	De 1906 docentes, 347 pertenecen al área educativa Ciencias Naturales y Educación Ambiental.
			697 oscilan entre las edades de 46 y 55 años de edad.		1256 pertenecen al nivel de media académica.
Clúster #7 = De 1257 docentes.	1053 son del género femenino.	804 ingresaron a nómina después del año 2010	1211 tienen un salario básico que oscila entre 1.000.000 y 1.500.000	828 tienen un nivel de educación de pregrado en licenciatura.	N/A
			1117 pertenecen al área educativa primaria.		
			1249 pertenecen al nivel de primaria.		
			787 oscilan entre las edades de 36 y 45 años de edad.		

SUGERENCIAS FUTURAS:

- Realización de una investigación futura por parte del equipo coordinador de un Proyecto de Teletrabajo en la Secretaría de Educación de Medellín, para la evaluación de procesos de selección de posibles teletrabajadores, por medio de otras técnicas de minería de datos que permitan implementar o crear nuevos modelos para el aprovechamiento máximo de la información.

- Como resultado de este trabajo, se evidencia la necesidad de construir una base de datos de docentes con información suficiente que muestre las relaciones entre competencias laborales y personales, rasgos de personalidad, valores, hobbies, entorno familiar, entorno laboral, trayectoria laboral, historial de formación, entre otros; y que, en consecuencia, permita hacer un perfilamiento de docentes para teletrabajo ya que sin este tipo de base de datos no es posible obtener un avance significativo frente a dicho propósito.
- Con el presente trabajo queda abierta la propuesta para implementar pilotos de teletrabajo con el gremio de los docentes del municipio de Medellín.

ANEXO 3: Algunas variables en el primer análisis con herramienta *DQ analyser*

Tabla 31: Perfil #1 de variables CODEMPLEADO y FECHAINGRESONOMINA

General		Estadísticas		Frecuencia		Máscara/Extremos		Observaciones																																																													
Expression:	CODEMPLEADO	Counts		Frequency Analysis		N/A		El campo de tipo entero presenta los siguientes aspectos: - Aunque es un campo obligatorio, su máxima frecuencia se halla en los datos: (39,169,650;39,170,358;43,623,866;88,144,665) - Existen 10783 CODEMPLEADO con valores únicos y 10787 valores distintos de un total de 10791; teniendo en cuenta que es un campo de valor único presenta una inconsistencia con 4 valores duplicados.																																																													
Datatype:	INTEGER	Type		Value																																																																	
Rows:	10,791	Count		Count Percentage																																																																	
Unique 99.00% Duplicate 0.60% Null 0.40%		Non-null values:		10,791		39,169,650		2		0.02%																																																											
		Null values:		0		39,170,358		2		0.02%																																																											
		Distinct values:		10,787		43,623,866		2		0.02%																																																											
		Duplicate values:		4		88,144,665		2		0.02%																																																											
		Unique values:		10,783		129,921		1		0.01%																																																											
		Non-unique values:		4		174,516		1		0.01%																																																											
						227,803		1		0.01%																																																											
						251,469		1		0.01%																																																											
						282,764		1		0.01%																																																											
						797,821		1		0.01%																																																											
						2,762,874		1		0.01%																																																											
						2,773,729		1		0.01%																																																											
						2,773,846		1		0.01%																																																											
						2,774,065		1		0.01%																																																											
						2,774,336		1		0.01%																																																											
						2,775,568		1		0.01%																																																											
						3,068,754		1		0.01%																																																											
Expression:	FECHAINGRESONOMINA	Counts		Frequency Analysis		Count Percentage		<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td>01/01/2003 00:00:00</td><td>108</td><td>1.00%</td></tr><tr><td>01/09/2007 00:00:00</td><td>2820</td><td>26.13%</td></tr><tr><td>02/28/2011 00:00:00</td><td>1772</td><td>16.42%</td></tr><tr><td>02/28/2009 00:00:00</td><td>526</td><td>4.87%</td></tr><tr><td>03/25/2011 00:00:00</td><td>284</td><td>2.63%</td></tr><tr><td>03/25/2011 00:00:00</td><td>232</td><td>2.15%</td></tr><tr><td>02/01/2014 00:00:00</td><td>200</td><td>1.85%</td></tr><tr><td>09/01/2003 00:00:00</td><td>179</td><td>1.66%</td></tr><tr><td>01/15/2007 00:00:00</td><td>128</td><td>1.19%</td></tr><tr><td>02/12/2008 00:00:00</td><td>109</td><td>1.01%</td></tr><tr><td>03/13/2012 00:00:00</td><td>106</td><td>0.98%</td></tr><tr><td>12/23/1997 00:00:00</td><td>88</td><td>0.82%</td></tr><tr><td>03/18/2011 00:00:00</td><td>87</td><td>0.81%</td></tr><tr><td>03/04/2010 00:00:00</td><td>87</td><td>0.81%</td></tr><tr><td>01/25/2013 00:00:00</td><td>76</td><td>0.70%</td></tr><tr><td>03/01/2014 00:00:00</td><td>75</td><td>0.70%</td></tr><tr><td>03/12/2003 00:00:00</td><td>64</td><td>0.59%</td></tr><tr><td>08/27/1996 00:00:00</td><td>61</td><td>0.57%</td></tr><tr><td>03/27/2003 00:00:00</td><td>60</td><td>0.56%</td></tr></tbody></table>		Value	Count	Percentage	01/01/2003 00:00:00	108	1.00%	01/09/2007 00:00:00	2820	26.13%	02/28/2011 00:00:00	1772	16.42%	02/28/2009 00:00:00	526	4.87%	03/25/2011 00:00:00	284	2.63%	03/25/2011 00:00:00	232	2.15%	02/01/2014 00:00:00	200	1.85%	09/01/2003 00:00:00	179	1.66%	01/15/2007 00:00:00	128	1.19%	02/12/2008 00:00:00	109	1.01%	03/13/2012 00:00:00	106	0.98%	12/23/1997 00:00:00	88	0.82%	03/18/2011 00:00:00	87	0.81%	03/04/2010 00:00:00	87	0.81%	01/25/2013 00:00:00	76	0.70%	03/01/2014 00:00:00	75	0.70%	03/12/2003 00:00:00	64	0.59%	08/27/1996 00:00:00	61	0.57%	03/27/2003 00:00:00	60	0.56%
Value	Count	Percentage																																																																			
01/01/2003 00:00:00	108	1.00%																																																																			
01/09/2007 00:00:00	2820	26.13%																																																																			
02/28/2011 00:00:00	1772	16.42%																																																																			
02/28/2009 00:00:00	526	4.87%																																																																			
03/25/2011 00:00:00	284	2.63%																																																																			
03/25/2011 00:00:00	232	2.15%																																																																			
02/01/2014 00:00:00	200	1.85%																																																																			
09/01/2003 00:00:00	179	1.66%																																																																			
01/15/2007 00:00:00	128	1.19%																																																																			
02/12/2008 00:00:00	109	1.01%																																																																			
03/13/2012 00:00:00	106	0.98%																																																																			
12/23/1997 00:00:00	88	0.82%																																																																			
03/18/2011 00:00:00	87	0.81%																																																																			
03/04/2010 00:00:00	87	0.81%																																																																			
01/25/2013 00:00:00	76	0.70%																																																																			
03/01/2014 00:00:00	75	0.70%																																																																			
03/12/2003 00:00:00	64	0.59%																																																																			
08/27/1996 00:00:00	61	0.57%																																																																			
03/27/2003 00:00:00	60	0.56%																																																																			
Datatype:	STRING	Type		Value																																																																	
Rows:	10,791	Count		Count Percentage																																																																	
Unique 9.94% Null 0.00% Duplicate 89.94%		Non-null values:		10,683		108		1.00%																																																													
		Null values:		108		01/09/2007 00:00:00		2820		26.13%																																																											
		Distinct values:		1,646		02/28/2011 00:00:00		1772		16.42%																																																											
		Duplicate values:		9,037		02/28/2009 00:00:00		526		4.87%																																																											
		Unique values:		1,073		03/25/2011 00:00:00		284		2.63%																																																											
		Non-unique values:		573		03/25/2011 00:00:00		232		2.15%																																																											
						02/01/2014 00:00:00		200		1.85%																																																											
						09/01/2003 00:00:00		179		1.66%																																																											
						01/15/2007 00:00:00		128		1.19%																																																											
						02/12/2008 00:00:00		109		1.01%																																																											
						03/13/2012 00:00:00		106		0.98%																																																											
						12/23/1997 00:00:00		88		0.82%																																																											
						03/18/2011 00:00:00		87		0.81%																																																											
						03/04/2010 00:00:00		87		0.81%																																																											
						01/25/2013 00:00:00		76		0.70%																																																											
						03/01/2014 00:00:00		75		0.70%																																																											
						03/12/2003 00:00:00		64		0.59%																																																											
						08/27/1996 00:00:00		61		0.57%																																																											
						03/27/2003 00:00:00		60		0.56%																																																											
Expression:	CODVINCULACIONCATEGORIA	Counts		Frequency Analysis		Count Percentage		<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td>DDDDDDDDDD DDDDDDD</td><td>7823</td><td>64.17%</td></tr><tr><td>DDDDDDDDDD DDDDDDD</td><td>3650</td><td>31.82%</td></tr></tbody></table>		Value	Count	Percentage	DDDDDDDDDD DDDDDDD	7823	64.17%	DDDDDDDDDD DDDDDDD	3650	31.82%																																																			
Value	Count	Percentage																																																																			
DDDDDDDDDD DDDDDDD	7823	64.17%																																																																			
DDDDDDDDDD DDDDDDD	3650	31.82%																																																																			
Datatype:	STRING	Type		Value																																																																	
Rows:	10,791	Count		Count Percentage																																																																	
Unique 5.31% Null 0.00% Duplicate 94.69%		Non-null values:		10,683		108		1.00%																																																													
		Null values:		108		01/09/2007 00:00:00		2820		26.13%																																																											
		Distinct values:		1,646		02/28/2011 00:00:00		1772		16.42%																																																											
		Duplicate values:		9,037		02/28/2009 00:00:00		526		4.87%																																																											
		Unique values:		1,073		03/25/2011 00:00:00		284		2.63%																																																											
		Non-unique values:		573		03/25/2011 00:00:00		232		2.15%																																																											
						02/01/2014 00:00:00		200		1.85%																																																											
						09/01/2003 00:00:00		179		1.66%																																																											
						01/15/2007 00:00:00		128		1.19%																																																											
						02/12/2008 00:00:00		109		1.01%																																																											
						03/13/2012 00:00:00		106		0.98%																																																											
						12/23/1997 00:00:00		88		0.82%																																																											
						03/18/2011 00:00:00		87		0.81%																																																											
						03/04/2010 00:00:00		87		0.81%																																																											
						01/25/2013 00:00:00		76		0.70%																																																											
						03/01/2014 00:00:00		75		0.70%																																																											
						03/12/2003 00:00:00		64		0.59%																																																											
						08/27/1996 00:00:00		61		0.57%																																																											
						03/27/2003 00:00:00		60		0.56%																																																											
Expression:	CODVINCULACIONCATEGORIA	Counts		Frequency Analysis		Count Percentage		<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td>DDDDDDDDDD DDDDDDD</td><td>7823</td><td>64.17%</td></tr><tr><td>DDDDDDDDDD DDDDDDD</td><td>3650</td><td>31.82%</td></tr></tbody></table>		Value	Count	Percentage	DDDDDDDDDD DDDDDDD	7823	64.17%	DDDDDDDDDD DDDDDDD	3650	31.82%																																																			
Value	Count	Percentage																																																																			
DDDDDDDDDD DDDDDDD	7823	64.17%																																																																			
DDDDDDDDDD DDDDDDD	3650	31.82%																																																																			
Datatype:	STRING	Type		Value																																																																	
Rows:	10,791	Count		Count Percentage																																																																	
Unique 5.31% Null 0.00% Duplicate 94.69%		Non-null values:		10,683		108		1.00%																																																													
		Null values:		108		01/09/2007 00:00:00		2820		26.13%																																																											
		Distinct values:		1,646		02/28/2011 00:00:00		1772		16.42%																																																											
		Duplicate values:		9,037		02/28/2009 00:00:00		526		4.87%																																																											
		Unique values:		1,073		03/25/2011 00:00:00		284		2.63%																																																											
		Non-unique values:		573		03/25/2011 00:00:00		232		2.15%																																																											
						02/01/2014 00:00:00		200		1.85%																																																											
						09/01/2003 00:00:00		179		1.66%																																																											
						01/15/2007 00:00:00		128		1.19%																																																											
						02/12/2008 00:00:00		109		1.01%																																																											
						03/13/2012 00:00:00		106		0.98%																																																											
						12/23/1997 00:00:00		88		0.82%																																																											
						03/18/2011 00:00:00		87		0.81%																																																											
						03/04/2010 00:00:00		87		0.81%																																																											
						01/25/2013 00:00:00		76		0.70%																																																											
						03/01/2014 00:00:00		75		0.70%																																																											
						03/12/2003 00:00:00		64		0.59%																																																											
						08/27/1996 00:00:00		61		0.57%																																																											
						03/27/2003 00:00:00		60		0.56%																																																											
Expression:	CODVINCULACIONCATEGORIA	Counts		Frequency Analysis		Count Percentage		<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td>DDDDDDDDDD DDDDDDD</td><td>7823</td><td>64.17%</td></tr><tr><td>DDDDDDDDDD DDDDDDD</td><td>3650</td><td>31.82%</td></tr></tbody></table>		Value	Count	Percentage	DDDDDDDDDD DDDDDDD	7823	64.17%	DDDDDDDDDD DDDDDDD	3650	31.82%																																																			
Value	Count	Percentage																																																																			
DDDDDDDDDD DDDDDDD	7823	64.17%																																																																			
DDDDDDDDDD DDDDDDD	3650	31.82%																																																																			
Datatype:	STRING	Type		Value																																																																	
Rows:	10,791	Count		Count Percentage																																																																	
Unique 5.31% Null 0.00% Duplicate 94.69%		Non-null values:		10,683		108		1.00%																																																													
		Null values:		108		01/09/2007 00:00:00		2820		26.13%																																																											
		Distinct values:		1,646		02/28/2011 00:00:00		1772		16.42%																																																											
		Duplicate values:		9,037		02/28/2009 00:00:00		526		4.87%																																																											
		Unique values:		1,073		03/25/2011 00:00:00		284		2.63%																																																											
		Non-unique values:		573		03/25/2011 00:00:00		232		2.15%																																																											
						02/01/2014 00:00:00		200		1.85%																																																											
						09/01/2003 00:00:00		179		1.66%																																																											
						01/15/2007 00:00:00		128		1.19%																																																											
						02/12/2008 00:00:00		109		1.01%																																																											
						03/13/2012 00:00:00		106		0.98%																																																											
						12/23/1997 00:00:00		88		0.82%																																																											
						03/18/2011 00:00:00		87		0.81%																																																											
						03/04/2010 00:00:00		87		0.81%																																																											
						01/25/2013 00:00:00		76		0.70%																																																											
						03/01/2014 00:00:00		75		0.70%																																																											
						03/12/2003 00:00:00		64		0.59%																																																											
						08/27/1996 00:00:00		61		0.57%																																																											
						03/27/2003 00:00:00		60		0.56%																																																											
Expression:	CODVINCULACIONCATEGORIA	Counts		Frequency Analysis		Count Percentage		<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td>DDDDDDDDDD DDDDDDD</td><td>7823</td><td>64.17%</td></tr><tr><td>DDDDDDDDDD DDDDDDD</td><td>3650</td><td>31.82%</td></tr></tbody></table>		Value	Count	Percentage	DDDDDDDDDD DDDDDDD	7823	64.17%	DDDDDDDDDD DDDDDDD	3650	31.82%																																																			
Value	Count	Percentage																																																																			
DDDDDDDDDD DDDDDDD	7823	64.17%																																																																			
DDDDDDDDDD DDDDDDD	3650	31.82%																																																																			
Datatype:	STRING	Type		Value																																																																	
Rows:	10,791	Count		Count Percentage																																																																	
Unique 5.31% Null 0.00% Duplicate 94.69%		Non-null values:		10,683		108		1.00%																																																													
		Null values:		108		01/09/2007 00:00:00		2820		26.13%																																																											
		Distinct values:		1,646		02/28/2011 00:00:00		1772		16.42%																																																											
		Duplicate values:		9,037		02/28/2009 00:00:00		526		4.87%																																																											
		Unique values:		1,073		03/25/2011 00:00:00		284		2.63%																																																											
		Non-unique values:		573		03/25/2011 00:00:00		232		2.15%																																																											
						02/01/2014 00:00:00		200		1.85%																																																											
						09/01/2003 00:00:00		179		1.66%																																																											
						01/15/2007 00:00:00		128		1.19%																																																											
						02/12/2008 00:00:00		109		1.01%																																																											
						03/13/2012 00:00:00		106		0.98%																																																											
						12/23/1997 00:00:00		88		0.82%																																																											
						03/18/2011 00:00:00		87		0.81%																																																											
						03/04/2010 00:00:00		87		0.81%																																																											
						01/25/2013 00:00:00		76		0.70%																																																											
						03/01/2014 00:00:00		75		0.70%																																																											
						03/12/2003 00:00:00		64		0.59%																																																											
						08/27/1996 00:00:00		61		0.57%																																																											
						03/27/2003 00:00:00		60		0.56%																																																											
Expression:	CODVINCULACIONCATEGORIA	Counts		Frequency Analysis		Count Percentage		<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td>DDDDDDDDDD DDDDDDD</td><td>7823</td><td>64.17%</td></tr><tr><td>DDDDDDDDDD DDDDDDD</td><td>3650</td><td>31.82%</td></tr></tbody></table>		Value	Count	Percentage	DDDDDDDDDD DDDDDDD	7823	64.17%	DDDDDDDDDD DDDDDDD	3650	31.82%																																																			
Value	Count	Percentage																																																																			
DDDDDDDDDD DDDDDDD	7823	64.17%																																																																			
DDDDDDDDDD DDDDDDD	3650	31.82%																																																																			
Datatype:	STRING	Type		Value																																																																	
Rows:	10,791	Count		Count Percentage																																																																	
Unique 5.31% Null 0.00% Duplicate 94.69%		Non-null values:		10,683		108		1.00%																																																													
		Null values:		108		01/09/2007 00:00:00		2820		26.13%																																																											
		Distinct values:		1,646		02/28/2011 00:00:00		1772		16.42%																																																											
		Duplicate values:		9,037		02/28/2009 00:00:00		526		4.87%																																																											
		Unique values:		1,073		03/25/2011 00:00:00		284		2.63%																																																											
		Non-unique values:		573		03/25/2011 00:00:00		232		2.15%																																																											
						02/01/2014 00:00:00		200		1.85%																																																											
						09/01/2003 00:00:00		179		1.66%																																																											
						01/15/2007 00:00:00		128		1.19%																																																											
						02/12/2008 00:00:00		109		1.01%																																																											
						03/13/2012 00:00:00																																																															

Tabla 33: Perfil #1 de variables VINCULACIONCATEGORIA y CODVINCULACIONESTADO

General		Estadísticas		Frecuencia		Máscara/Extremos		Observaciones
Datatype:	STRING	Type	Count	Value	Count	Percentage	Value	Count
Rows:	10,791	Non-null values:	10,683		108	1.00%		108
		Null values:	108	Planta	106,79	98.96%	LLLLLL	106,79
		Distinct values:	2	Contratista	4	0.04%	LLLLLLLLLLLL	4
		Duplicate values:	10,681					
		Unique values:	0					
		Non-unique values:	2					
		Statistics						
		Type	value					
		Minimum value:	Contratista					
		Median value:	Planta					
		Maximum value:	Planta					
		Minimum length:	0					
		Median length:	6					
		Average length:	6					
		Maximum length:	11					
Expression:	CODVINCULACIONESTADO	Counts		Frequency Analysis		Mask Analysis		El número de registros evaluados es de 10791 con un total de 98,98% registros duplicados que equivalen a 10681; situación normal para el campo, ya que se da la posibilidad de que el docente tenga la misma categoría de vinculación igual a la de otro docente, se presentan como dato diferenciador el 1% equivalente a 108 como valor nulo y el 0,02% equivalente a 2 como valor no único. Es un campo de tipo string. La frecuencia de los datos es esta representada por el 98,96%, equivalente a 10679 docentes vinculados con categoría de planta; 0,04% docentes vinculados a la categoría contratista y sin categoría serían 108 docentes correspondientes al 1% de los datos.
Datatype:	INTEGER	Type	Count	Value	Count	Percentage		
Rows:	10,791	Non-null values:	10,683		108	1.00%		Es un dato de tipo entero y presenta los siguientes aspectos: - Están catalogados como valores nulos 108, equivalentes al 1% - El 98,99% de los valores son duplicados y la categoría asignada a CODVINCULACIONESTADO = 1. - No existe máscara de entrada asignada para este campo.
		Null values:	108	1	106,83	99.00%		
		Distinct values:	1					
		Duplicate values:	10,682					
		Unique values:	0					
		Non-unique values:	1					
		Statistics						
		Type	value					
		Minimum value:	1					
		Median value:	1					
		Maximum value:	1					
		Sum:	10,683					
		Average:	1					
		Variance:	0					
		Standard deviation:	0					

Tabla 34: Perfil #1 de variables VINCULACIONESTADO y CODVACACIONESMETODO

General		Estadísticas		Frecuencia		Máscara/Extremos		Observaciones
Datatype:	STRING	Type	Count	Value	Count	Percentage	Value	Count
Rows:	10,791	Non-null values:	10,683		108	1.00%		108
		Null values:	108	Activa	106,83	99.00%	LLLLLL	106,83
		Distinct values:	1					
		Duplicate values:	10,682					
		Unique values:	0					
		Non-unique values:	1					
		Statistics						
		Type	value					
		Minimum value:	Activa					
		Median value:	Activa					
		Maximum value:	Activa					
		Minimum length:	6					
		Median length:	6					
		Average length:	6					
		Maximum length:	6					
Expression:	CODVACACIONESMETODO	Counts		Frequency Analysis		Mask Analysis		Es un campo de tipo string, con un total de 10791 registros analizados. Presentan las siguientes características: - 98,99% datos duplicados equivalentes a 10682 - Hay un 1% de datos nulos equivalentes a 108 con un único valor distinto. - Existe una única categoría asignada a la variable VINCULACIONESTADO = Activa. - La máscara de entrada asignada está conformada por letras y es constante en todo el set de datos.
Datatype:	INTEGER	Type	Count	Value	Count	Percentage		
Rows:	10,791	Non-null values:	10,683		108	1.00%		Es un campo de tipo entero, con un total de 10791 registros analizados. Presentan las siguientes características: - Hay un 98,97% datos duplicados equivalentes a 10680, el 1% de datos nulos equivalentes a 108 y 3 valores distintos y como valores únicos el 0,02% correspondiente a 2 registros. - Existen 3 grupos de análisis de frecuencia con los valores 0, 1, 2. No hay máscara definida para el campo. - En el momento de tomar una decisión al reemplazar los valores nulos, es necesario descartar la categoría cero(0) ya que ya existe un valor cero asignado para esta variable.
		Null values:	108	2	101,83	94.37%		
		Distinct values:	3	1	499	4.62%		
		Duplicate values:	10,680	0	1	0.01%		
		Unique values:	1					
		Non-unique values:	2					
		Statistics						
		Type	value					
		Minimum value:	0					
		Median value:	2					
		Maximum value:	2					
		Sum:	20,865					
		Average:	0.95					
		Variance:	0.04					
		Standard deviation:	0.21					

Tabla 35: Perfil #1 de variables VACACIONESMETODO y CODCALENDARIO

General		Estadísticas		Frecuencia		Máscara/Extremos		Observaciones																																																										
Expression:	VACACIONESMETODO																																																																	
Datatype:	STRING																																																																	
Rows:	10,791																																																																	
Unique 0.01% Null 1.00% Non-unique 99.97% Duplicate 98.97%		<table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>10,683</td></tr><tr><td>Null values:</td><td>108</td></tr><tr><td>Distinct values:</td><td>2</td></tr><tr><td>Duplicate values:</td><td>10,680</td></tr><tr><td>Unique values:</td><td>1</td></tr><tr><td>Non-unique values:</td><td>2</td></tr></tbody></table> <table><thead><tr><th>Type</th><th>value</th></tr></thead><tbody><tr><td>Minimum value:</td><td>Docentes</td></tr><tr><td>Median value:</td><td>Docentes</td></tr><tr><td>Maximum value:</td><td>No Aplica</td></tr><tr><td>Minimum length:</td><td>8</td></tr><tr><td>Median length:</td><td>8</td></tr><tr><td>Average length:</td><td>8</td></tr><tr><td>Maximum length:</td><td>9</td></tr></tbody></table>		Type	Count	Non-null values:	10,683	Null values:	108	Distinct values:	2	Duplicate values:	10,680	Unique values:	1	Non-unique values:	2	Type	value	Minimum value:	Docentes	Median value:	Docentes	Maximum value:	No Aplica	Minimum length:	8	Median length:	8	Average length:	8	Maximum length:	9	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td></td><td>108</td><td>1.00%</td></tr><tr><td>Docentes</td><td>10183</td><td>94.37%</td></tr><tr><td>Exceder</td><td>499</td><td>4.62%</td></tr><tr><td>No Aplica</td><td>1</td><td>0.01%</td></tr></tbody></table>		Value	Count	Percentage		108	1.00%	Docentes	10183	94.37%	Exceder	499	4.62%	No Aplica	1	0.01%	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td></td><td>108</td><td>1.00%</td></tr><tr><td>LLLLLLLLL</td><td>10682</td><td>98.99%</td></tr><tr><td>L.L.L.L.L.L.L</td><td>1</td><td>0.01%</td></tr></tbody></table>		Value	Count	Percentage		108	1.00%	LLLLLLLLL	10682	98.99%	L.L.L.L.L.L.L	1	0.01%	Es un campo de tipo string, con un total de 10791 registros analizados. Presentan las siguientes características: - 98,97% datos duplicados equivalentes a 10680, el 1% de datos nulos equivalentes a 108 y 3 valores distintos y como valores únicos el 0,02 % correspondiente a 2 registros. - En el análisis de frecuencia se categorizan el método vacaciones con tres valores: Docentes con un 94,37%, estándar con un 4,62% y No aplica con un 0,01%. El tipo de máscara de entrada está determinado por letras en dos formatos diferentes.	
Type	Count																																																																	
Non-null values:	10,683																																																																	
Null values:	108																																																																	
Distinct values:	2																																																																	
Duplicate values:	10,680																																																																	
Unique values:	1																																																																	
Non-unique values:	2																																																																	
Type	value																																																																	
Minimum value:	Docentes																																																																	
Median value:	Docentes																																																																	
Maximum value:	No Aplica																																																																	
Minimum length:	8																																																																	
Median length:	8																																																																	
Average length:	8																																																																	
Maximum length:	9																																																																	
Value	Count	Percentage																																																																
	108	1.00%																																																																
Docentes	10183	94.37%																																																																
Exceder	499	4.62%																																																																
No Aplica	1	0.01%																																																																
Value	Count	Percentage																																																																
	108	1.00%																																																																
LLLLLLLLL	10682	98.99%																																																																
L.L.L.L.L.L.L	1	0.01%																																																																
Expression:	CODCALENDARIO																																																																	
Datatype:	INTEGER																																																																	
Rows:	10,791																																																																	
Non-unique 0.02% Null 1.00% Duplicate 98.98%		<table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>10,683</td></tr><tr><td>Null values:</td><td>108</td></tr><tr><td>Distinct values:</td><td>2</td></tr><tr><td>Duplicate values:</td><td>10,681</td></tr><tr><td>Unique values:</td><td>0</td></tr><tr><td>Non-unique values:</td><td>2</td></tr></tbody></table> <table><thead><tr><th>Type</th><th>value</th></tr></thead><tbody><tr><td>Minimum value:</td><td>1</td></tr><tr><td>Median value:</td><td>1</td></tr><tr><td>Maximum value:</td><td>2</td></tr><tr><td>Sum:</td><td>10,7</td></tr><tr><td>Average:</td><td>1</td></tr><tr><td>Variance:</td><td>0</td></tr><tr><td>Standard deviation:</td><td>0.04</td></tr></tbody></table>		Type	Count	Non-null values:	10,683	Null values:	108	Distinct values:	2	Duplicate values:	10,681	Unique values:	0	Non-unique values:	2	Type	value	Minimum value:	1	Median value:	1	Maximum value:	2	Sum:	10,7	Average:	1	Variance:	0	Standard deviation:	0.04	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td></td><td>108</td><td>1.00%</td></tr><tr><td>1</td><td>10666</td><td>98.84%</td></tr><tr><td>2</td><td>17</td><td>0.16%</td></tr></tbody></table>		Value	Count	Percentage		108	1.00%	1	10666	98.84%	2	17	0.16%	N/A		Es un campo de tipo entero, con un total de 10791 registros analizados que presentan las siguientes características: - 98,98% datos duplicados equivalentes a 10681, el 1% de datos nulos equivalentes a 108 valores nulos y 2-valores distintos y como valores únicos el 0,02% correspondiente a 2 registros. - No hay una máscara de entrada definida para este campo. En su frecuencia de análisis presenta como valores definidos el uno(1) y dos (2).																
Type	Count																																																																	
Non-null values:	10,683																																																																	
Null values:	108																																																																	
Distinct values:	2																																																																	
Duplicate values:	10,681																																																																	
Unique values:	0																																																																	
Non-unique values:	2																																																																	
Type	value																																																																	
Minimum value:	1																																																																	
Median value:	1																																																																	
Maximum value:	2																																																																	
Sum:	10,7																																																																	
Average:	1																																																																	
Variance:	0																																																																	
Standard deviation:	0.04																																																																	
Value	Count	Percentage																																																																
	108	1.00%																																																																
1	10666	98.84%																																																																
2	17	0.16%																																																																

Tabla 36: Perfil #1 de variables NOMBRAMIENTOCODACTOADMIN y NOMBRAMIENTOFECHA

General		Estadísticas	Frecuencia		Máscara/Extremos		Observaciones																																																																								
Expression: NOMBRAMIENTO CODACTO ADMIN		Counts <table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>10,683</td></tr><tr><td>Null values:</td><td>108</td></tr><tr><td>Distinct values:</td><td>3</td></tr><tr><td>Duplicate values:</td><td>10,680</td></tr><tr><td>Unique values:</td><td>0</td></tr><tr><td>Non-unique values:</td><td>3</td></tr></tbody></table> Statistics <table><thead><tr><th>Type</th><th>value</th></tr></thead><tbody><tr><td>Minimum value:</td><td>-2</td></tr><tr><td>Median value:</td><td>-2</td></tr><tr><td>Maximum value:</td><td>6</td></tr><tr><td>Sum:</td><td>21,569</td></tr><tr><td>Average:</td><td>2.06</td></tr><tr><td>Variance:</td><td>0.06</td></tr><tr><td>Standard deviation:</td><td>0.24</td></tr></tbody></table>	Type	Count	Non-null values:	10,683	Null values:	108	Distinct values:	3	Duplicate values:	10,680	Unique values:	0	Non-unique values:	3	Type	value	Minimum value:	-2	Median value:	-2	Maximum value:	6	Sum:	21,569	Average:	2.06	Variance:	0.06	Standard deviation:	0.24	Frequency Analysis <table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td></td><td>108</td><td>1.00%</td></tr><tr><td>2</td><td>10092</td><td>93.52%</td></tr><tr><td>3</td><td>588</td><td>5.45%</td></tr><tr><td>6</td><td>3</td><td>0.03%</td></tr></tbody></table>		Value	Count	Percentage		108	1.00%	2	10092	93.52%	3	588	5.45%	6	3	0.03%	N/A		Es un campo de tipo entero, con un total de 10791 registros analizados que presentan las siguientes características: - Hay 98,97% datos duplicados equivalentes a 10680, el 1% de datos nulos equivalentes a 108 valores nulos y 3 valores distintos y como valores únicos el 0,03% correspondiente a 3 registros. - Los valores clasificados por el análisis de frecuencia corresponden a dos (2) con un 93,52%, tres (3) con un 5,45% y seis (6) con un 0,03%. - No hay una máscara de entrada definida para este campo.																											
Type	Count																																																																														
Non-null values:	10,683																																																																														
Null values:	108																																																																														
Distinct values:	3																																																																														
Duplicate values:	10,680																																																																														
Unique values:	0																																																																														
Non-unique values:	3																																																																														
Type	value																																																																														
Minimum value:	-2																																																																														
Median value:	-2																																																																														
Maximum value:	6																																																																														
Sum:	21,569																																																																														
Average:	2.06																																																																														
Variance:	0.06																																																																														
Standard deviation:	0.24																																																																														
Value	Count	Percentage																																																																													
	108	1.00%																																																																													
2	10092	93.52%																																																																													
3	588	5.45%																																																																													
6	3	0.03%																																																																													
Datatype: INTEGER																																																																															
Rows: 10,791																																																																															
Non-unique 98.97% Null 1.00% Duplicate 08.97%																																																																															
Expression: NOMBRAMIENTO FECHA		Counts <table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>10,682</td></tr><tr><td>Null values:</td><td>109</td></tr><tr><td>Distinct values:</td><td>1,339</td></tr><tr><td>Duplicate values:</td><td>9,343</td></tr><tr><td>Unique values:</td><td>794</td></tr><tr><td>Non-unique values:</td><td>845</td></tr></tbody></table> Statistics <table><thead><tr><th>Type</th><th>value</th></tr></thead><tbody><tr><td>Minimum value:</td><td>01/13/1994 00:00:00</td></tr><tr><td>Median value:</td><td>08/19/2004 00:00:00</td></tr><tr><td>Maximum value:</td><td>03/12/2014 00:00:00</td></tr><tr><td>Minimum length:</td><td>19</td></tr><tr><td>Median length:</td><td>19</td></tr><tr><td>Average length:</td><td>19</td></tr><tr><td>Maximum length:</td><td>19</td></tr></tbody></table>	Type	Count	Non-null values:	10,682	Null values:	109	Distinct values:	1,339	Duplicate values:	9,343	Unique values:	794	Non-unique values:	845	Type	value	Minimum value:	01/13/1994 00:00:00	Median value:	08/19/2004 00:00:00	Maximum value:	03/12/2014 00:00:00	Minimum length:	19	Median length:	19	Average length:	19	Maximum length:	19	Frequency Analysis <table><thead><tr><th></th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td>01/06/2004 00:00</td><td>2926</td><td>27.12%</td></tr><tr><td>01/15/2007 00:00:00</td><td>1820</td><td>16.87%</td></tr><tr><td>02/28/2011 00:00:00</td><td>524</td><td>4.86%</td></tr><tr><td>02/20/2009 00:00:00</td><td>385</td><td>3.54%</td></tr><tr><td>03/25/2011 00:00:00</td><td>230</td><td>2.13%</td></tr><tr><td>03/27/2003 00:00:00</td><td>138</td><td>1.28%</td></tr><tr><td>01/26/2013 00:00:00</td><td>113</td><td>1.07%</td></tr><tr><td>02/12/2008 00:00</td><td>119</td><td>1.02%</td></tr><tr><td>03/13/2012 00:00:00</td><td>106</td><td>0.98%</td></tr></tbody></table>			Count	Percentage	01/06/2004 00:00	2926	27.12%	01/15/2007 00:00:00	1820	16.87%	02/28/2011 00:00:00	524	4.86%	02/20/2009 00:00:00	385	3.54%	03/25/2011 00:00:00	230	2.13%	03/27/2003 00:00:00	138	1.28%	01/26/2013 00:00:00	113	1.07%	02/12/2008 00:00	119	1.02%	03/13/2012 00:00:00	106	0.98%	Mask Analysis <table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td></td><td>109</td><td>1.01%</td></tr><tr><td>DDDDDDDD DD-DD-DD</td><td>5992</td><td>55.53%</td></tr><tr><td>DDDD-DD-DD DD-DD-DD</td><td>4690</td><td>43.46%</td></tr></tbody></table>		Value	Count	Percentage		109	1.01%	DDDDDDDD DD-DD-DD	5992	55.53%	DDDD-DD-DD DD-DD-DD	4690	43.46%	Es un campo de tipo string, con un total de 10791 registros analizados y los cuales presentan las siguientes características: - 86,58% datos duplicados equivalentes a 9343, el 1,01% de datos nulos equivalentes a 109 valores nulos y 1339 valores distintos y como valores únicos el 7,36% correspondiente a 794 registros. - Existe una máscara de entrada con 2 tipos de formato definida para el campo - Es importante resaltar que el mayor número de datos está registrado el 01 de junio de 2004 y agrupe un grupo de docentes de 2926, equivalente al 27,12%, seguido del 16,87% equivalente a 1820 docentes.
Type	Count																																																																														
Non-null values:	10,682																																																																														
Null values:	109																																																																														
Distinct values:	1,339																																																																														
Duplicate values:	9,343																																																																														
Unique values:	794																																																																														
Non-unique values:	845																																																																														
Type	value																																																																														
Minimum value:	01/13/1994 00:00:00																																																																														
Median value:	08/19/2004 00:00:00																																																																														
Maximum value:	03/12/2014 00:00:00																																																																														
Minimum length:	19																																																																														
Median length:	19																																																																														
Average length:	19																																																																														
Maximum length:	19																																																																														
	Count	Percentage																																																																													
01/06/2004 00:00	2926	27.12%																																																																													
01/15/2007 00:00:00	1820	16.87%																																																																													
02/28/2011 00:00:00	524	4.86%																																																																													
02/20/2009 00:00:00	385	3.54%																																																																													
03/25/2011 00:00:00	230	2.13%																																																																													
03/27/2003 00:00:00	138	1.28%																																																																													
01/26/2013 00:00:00	113	1.07%																																																																													
02/12/2008 00:00	119	1.02%																																																																													
03/13/2012 00:00:00	106	0.98%																																																																													
Value	Count	Percentage																																																																													
	109	1.01%																																																																													
DDDDDDDD DD-DD-DD	5992	55.53%																																																																													
DDDD-DD-DD DD-DD-DD	4690	43.46%																																																																													
Datatype: STRING																																																																															
Rows: 10,791																																																																															
Non-unique 86.58% Null 1.01% Duplicate 96.58% Unique 7.36%																																																																															

Tabla 37: Perfil #1 de variables NOMBRAMIENTONUMERO y POSESIONCODACTOADMIN

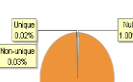
General		Estadísticas		Frecuencia		Máscara/Extremos		Observaciones																																																																			
Expression: NOMBRAMIENTONUMERO																																																																											
Datatype: INTEGER																																																																											
Rows: 10,791																																																																											
		<table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>10,673</td></tr><tr><td>Null values:</td><td>118</td></tr><tr><td>Distinct values:</td><td>3,131</td></tr><tr><td>Duplicate values:</td><td>7,542</td></tr><tr><td>Unique values:</td><td>2,374</td></tr><tr><td>Non-unique values:</td><td>727</td></tr></tbody></table> <table><thead><tr><th>Type</th><th>value</th></tr></thead><tbody><tr><td>Minimum value:</td><td>0</td></tr><tr><td>Median value:</td><td>321</td></tr><tr><td>Maximum value:</td><td>22,681</td></tr><tr><td>Sum:</td><td>19,964,231</td></tr><tr><td>Average:</td><td>1,870.54</td></tr><tr><td>Variance:</td><td>7,926,830.32</td></tr><tr><td>Standard deviation:</td><td>2,815.46</td></tr></tbody></table>		Type	Count	Non-null values:	10,673	Null values:	118	Distinct values:	3,131	Duplicate values:	7,542	Unique values:	2,374	Non-unique values:	727	Type	value	Minimum value:	0	Median value:	321	Maximum value:	22,681	Sum:	19,964,231	Average:	1,870.54	Variance:	7,926,830.32	Standard deviation:	2,815.46	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td></td><td>118</td><td>1.09%</td></tr><tr><td>13</td><td>2931</td><td>27.36%</td></tr><tr><td>152</td><td>1453</td><td>13.46%</td></tr><tr><td>154</td><td>257</td><td>3.31%</td></tr><tr><td>1,647</td><td>109</td><td>1.57%</td></tr><tr><td>312</td><td>138</td><td>1.28%</td></tr><tr><td>1,643</td><td>103</td><td>0.95%</td></tr><tr><td>3,81</td><td>54</td><td>0.57%</td></tr><tr><td>242</td><td>64</td><td>0.59%</td></tr><tr><td>4,311</td><td>61</td><td>0.57%</td></tr><tr><td>1,365</td><td>53</td><td>0.49%</td></tr></tbody></table>		Value	Count	Percentage		118	1.09%	13	2931	27.36%	152	1453	13.46%	154	257	3.31%	1,647	109	1.57%	312	138	1.28%	1,643	103	0.95%	3,81	54	0.57%	242	64	0.59%	4,311	61	0.57%	1,365	53	0.49%	N/A		Es un campo de tipo entero, con un total de 10791 registros analizados, que presentan las siguientes características: - El 69.89% datos duplicados equivalentes a 7542, el 1.09% de datos nulos equivalentes a 118 valores nulos y 3131 valores distintos y como valores únicos el 22.00% correspondiente a 2374 registros. - No hay una máscara de entrada definida para este campo. - El código de acto administrativo presenta muchos valores duplicados, lo que supone que los diferentes nombramientos están	
Type	Count																																																																										
Non-null values:	10,673																																																																										
Null values:	118																																																																										
Distinct values:	3,131																																																																										
Duplicate values:	7,542																																																																										
Unique values:	2,374																																																																										
Non-unique values:	727																																																																										
Type	value																																																																										
Minimum value:	0																																																																										
Median value:	321																																																																										
Maximum value:	22,681																																																																										
Sum:	19,964,231																																																																										
Average:	1,870.54																																																																										
Variance:	7,926,830.32																																																																										
Standard deviation:	2,815.46																																																																										
Value	Count	Percentage																																																																									
	118	1.09%																																																																									
13	2931	27.36%																																																																									
152	1453	13.46%																																																																									
154	257	3.31%																																																																									
1,647	109	1.57%																																																																									
312	138	1.28%																																																																									
1,643	103	0.95%																																																																									
3,81	54	0.57%																																																																									
242	64	0.59%																																																																									
4,311	61	0.57%																																																																									
1,365	53	0.49%																																																																									
Expression: POSESIONCODACTOADMIN																																																																											
Datatype: INTEGER																																																																											
		<table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>10,683</td></tr><tr><td>Null values:</td><td>108</td></tr><tr><td>Distinct values:</td><td>5</td></tr><tr><td>Duplicate values:</td><td>10,678</td></tr><tr><td>Unique values:</td><td>2</td></tr><tr><td>Non-unique values:</td><td>3</td></tr></tbody></table> <table><thead><tr><th>Type</th><th>value</th></tr></thead><tbody><tr><td>Minimum value:</td><td>1</td></tr><tr><td>Median value:</td><td>1</td></tr><tr><td>Maximum value:</td><td>7</td></tr><tr><td>Sum:</td><td>13,56</td></tr><tr><td>Average:</td><td>1.27</td></tr><tr><td>Variance:</td><td>1.26</td></tr><tr><td>Standard deviation:</td><td>1.12</td></tr></tbody></table>		Type	Count	Non-null values:	10,683	Null values:	108	Distinct values:	5	Duplicate values:	10,678	Unique values:	2	Non-unique values:	3	Type	value	Minimum value:	1	Median value:	1	Maximum value:	7	Sum:	13,56	Average:	1.27	Variance:	1.26	Standard deviation:	1.12	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td>1</td><td>10092</td><td>93.52%</td></tr><tr><td>6</td><td>563</td><td>5.22%</td></tr><tr><td>3</td><td>26</td><td>0.24%</td></tr><tr><td>5</td><td>1</td><td>0.01%</td></tr><tr><td>7</td><td>1</td><td>0.01%</td></tr></tbody></table>		Value	Count	Percentage	1	10092	93.52%	6	563	5.22%	3	26	0.24%	5	1	0.01%	7	1	0.01%	N/A		Es un campo de tipo entero, con un total de 10791 registros analizados. Presentan las siguientes características: - El 98.95% datos duplicados equivalentes a 10678, el 1.00% de datos nulos equivalentes a 108 valores nulos y 3131 valores distintos y como valores únicos el 0.02% correspondiente a 2 registros y el 0.03% equivalentes a 3 como valores no únicos. - Los valores 1 son relevantes con mayor número de frecuencia correspondiente a un total de 10092 registros equivalente al 93.52%, y tiene 5 valores son 5 diferentes (1, 6, 3, 5, 7). - No hay una máscara de entrada definida para este campo.																			
Type	Count																																																																										
Non-null values:	10,683																																																																										
Null values:	108																																																																										
Distinct values:	5																																																																										
Duplicate values:	10,678																																																																										
Unique values:	2																																																																										
Non-unique values:	3																																																																										
Type	value																																																																										
Minimum value:	1																																																																										
Median value:	1																																																																										
Maximum value:	7																																																																										
Sum:	13,56																																																																										
Average:	1.27																																																																										
Variance:	1.26																																																																										
Standard deviation:	1.12																																																																										
Value	Count	Percentage																																																																									
1	10092	93.52%																																																																									
6	563	5.22%																																																																									
3	26	0.24%																																																																									
5	1	0.01%																																																																									
7	1	0.01%																																																																									

Tabla 38: Perfil #1 de variables POSESIONFECHA y POSESIONNUMERO

General		Estadísticas	Frecuencia	Máscara/Extremos		Observaciones																																							
Expression:	POSESIONFECHA						Es un campo de tipo string, con un total de 10791 registros analizados. Presentan las siguientes características: - El 86,20% datos duplicados equivalentes a 9302, el 1,02% de datos nulos equivalentes a 110 valores nulos y 1379 valores distintos y como valores únicos el 7,44% correspondiente a 603 registros y el 5,34% equivalentes a 576 como valores no únicos. - El valor más relevante en la frecuencia es 2004-06-01 00:00:00, con un porcentaje de 27,12%, con una cantidad de 2926 registros. Lo que evidencia la posesión de docentes en masa en esta fecha determinada. - Existe una máscara de entrada con 2 tipos de formato definida para el campo.																																						
Datatype:	STRING																																												
Rows:	10,791																																												
		Counts <table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>10,681</td></tr><tr><td>Null values:</td><td>110</td></tr><tr><td>Distinct values:</td><td>1,379</td></tr><tr><td>Duplicate values:</td><td>9,302</td></tr><tr><td>Unique values:</td><td>603</td></tr><tr><td>Non-unique values:</td><td>576</td></tr></tbody></table> Statistics <table><thead><tr><th>Type</th><th>value</th></tr></thead><tbody><tr><td>Minimum value:</td><td>01/13/1994 00:00:00</td></tr><tr><td>Median value:</td><td>08/23/2011 00:00:00</td></tr><tr><td>Maximum value:</td><td>08/12/2014 00:00:00</td></tr><tr><td>Minimum length:</td><td>19</td></tr><tr><td>Median length:</td><td>19</td></tr><tr><td>Average length:</td><td>19</td></tr><tr><td>Maximum length:</td><td>19</td></tr></tbody></table>	Type	Count	Non-null values:	10,681	Null values:	110	Distinct values:	1,379	Duplicate values:	9,302	Unique values:	603	Non-unique values:	576	Type	value	Minimum value:	01/13/1994 00:00:00	Median value:	08/23/2011 00:00:00	Maximum value:	08/12/2014 00:00:00	Minimum length:	19	Median length:	19	Average length:	19	Maximum length:	19	Mask Analysis <table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td></td><td>110</td><td>1.02%</td></tr><tr><td>DDDDDDDD DD-DDDD</td><td>5955</td><td>55.18%</td></tr><tr><td>DDDD-DD-DD DD-DDDD</td><td>4726</td><td>43.80%</td></tr></tbody></table>	Value	Count	Percentage		110	1.02%	DDDDDDDD DD-DDDD	5955	55.18%	DDDD-DD-DD DD-DDDD	4726	43.80%
Type	Count																																												
Non-null values:	10,681																																												
Null values:	110																																												
Distinct values:	1,379																																												
Duplicate values:	9,302																																												
Unique values:	603																																												
Non-unique values:	576																																												
Type	value																																												
Minimum value:	01/13/1994 00:00:00																																												
Median value:	08/23/2011 00:00:00																																												
Maximum value:	08/12/2014 00:00:00																																												
Minimum length:	19																																												
Median length:	19																																												
Average length:	19																																												
Maximum length:	19																																												
Value	Count	Percentage																																											
	110	1.02%																																											
DDDDDDDD DD-DDDD	5955	55.18%																																											
DDDD-DD-DD DD-DDDD	4726	43.80%																																											
Expression:	POSESIONNUMERO						Es un campo de tipo entero, con un total de 10791 registros analizados y los cuales presentan las siguientes características: - 67,38% datos duplicados equivalentes a 7271, el 1,09% de datos nulos equivalentes a 118 y 3402 valores distintos y como valores únicos el 24,77% correspondientes a 2673 registros y el 6,76% equivalentes a 729 como valores no únicos. - No hay una máscara de entrada definida para este campo.																																						
Datatype:	INTEGER																																												
Rows:	10,791																																												
		Counts <table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>10,673</td></tr><tr><td>Null values:</td><td>118</td></tr><tr><td>Distinct values:</td><td>3,402</td></tr><tr><td>Duplicate values:</td><td>7,271</td></tr><tr><td>Unique values:</td><td>2,657</td></tr><tr><td>Non-unique values:</td><td>729</td></tr></tbody></table> Statistics <table><thead><tr><th>Type</th><th>value</th></tr></thead><tbody><tr><td>Minimum value:</td><td>0</td></tr><tr><td>Median value:</td><td>312</td></tr><tr><td>Maximum value:</td><td>22,681</td></tr><tr><td>Sum:</td><td>22,724,828</td></tr><tr><td>Average:</td><td>2,129.19</td></tr><tr><td>Variance:</td><td>9,470,236.31</td></tr><tr><td>Standard deviation:</td><td>3,077.37</td></tr></tbody></table>	Type	Count	Non-null values:	10,673	Null values:	118	Distinct values:	3,402	Duplicate values:	7,271	Unique values:	2,657	Non-unique values:	729	Type	value	Minimum value:	0	Median value:	312	Maximum value:	22,681	Sum:	22,724,828	Average:	2,129.19	Variance:	9,470,236.31	Standard deviation:	3,077.37	N/A												
Type	Count																																												
Non-null values:	10,673																																												
Null values:	118																																												
Distinct values:	3,402																																												
Duplicate values:	7,271																																												
Unique values:	2,657																																												
Non-unique values:	729																																												
Type	value																																												
Minimum value:	0																																												
Median value:	312																																												
Maximum value:	22,681																																												
Sum:	22,724,828																																												
Average:	2,129.19																																												
Variance:	9,470,236.31																																												
Standard deviation:	3,077.37																																												

Tabla 39: Perfil #1 de variables CONTINUIDAD y RETIROCODACTOADMIN

General		Estadísticas		Frecuencia		Máscara/Extremos		Observaciones																																																				
Expression:	CONTINUIDAD																																																											
Datatype:	STRING																																																											
Rows:	10,791																																																											
Non-unique 6.02% Null 1.39% Duplicate 98.80%		<table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>10,683</td></tr><tr><td>Null values:</td><td>108</td></tr><tr><td>Distinct values:</td><td>2</td></tr><tr><td>Duplicate values:</td><td>10,681</td></tr><tr><td>Unique values:</td><td>0</td></tr><tr><td>Non-unique values:</td><td>2</td></tr></tbody></table> <table><thead><tr><th>Type</th><th>value</th></tr></thead><tbody><tr><td>Minimum value:</td><td>N</td></tr><tr><td>Median value:</td><td>N</td></tr><tr><td>Maximum value:</td><td>S</td></tr><tr><td>Minimum length:</td><td>1</td></tr><tr><td>Median length:</td><td>1</td></tr><tr><td>Average length:</td><td>1</td></tr><tr><td>Maximum length:</td><td>1</td></tr></tbody></table>		Type	Count	Non-null values:	10,683	Null values:	108	Distinct values:	2	Duplicate values:	10,681	Unique values:	0	Non-unique values:	2	Type	value	Minimum value:	N	Median value:	N	Maximum value:	S	Minimum length:	1	Median length:	1	Average length:	1	Maximum length:	1	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td></td><td>108</td><td>1.00%</td></tr><tr><td>N</td><td>10660</td><td>98.79%</td></tr><tr><td>S</td><td>23</td><td>0.21%</td></tr></tbody></table>		Value	Count	Percentage		108	1.00%	N	10660	98.79%	S	23	0.21%	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td></td><td>108</td><td>1.00%</td></tr><tr><td>L</td><td>10683</td><td>99.00%</td></tr></tbody></table>		Value	Count	Percentage		108	1.00%	L	10683	99.00%	Es un campo de tipo string, con un total de 10791 registros analizados que presentan las siguientes características: - El 98,98% datos duplicados equivalentes a 10681, el 1,00% de datos nulos equivalentes a 108 y 2 valores distintos, 0 valores únicos y el 0,02% equivalentes a 2 como valores no únicos. - Existe una máscara de entrada con 1 tipo de formato en letras definida para el campo. - Existen 2 tipos de valores definidos para el campo y son N y S. - Aun no se posee información clara sobre el significado de los dos valores que aplican para la continuidad, ya que si se asume como negativo; se afirma que el 98,79% de los docentes no tienen continuidad en el cargo, situación que afectaría de forma sustancial el proceso de investigación.	
Type	Count																																																											
Non-null values:	10,683																																																											
Null values:	108																																																											
Distinct values:	2																																																											
Duplicate values:	10,681																																																											
Unique values:	0																																																											
Non-unique values:	2																																																											
Type	value																																																											
Minimum value:	N																																																											
Median value:	N																																																											
Maximum value:	S																																																											
Minimum length:	1																																																											
Median length:	1																																																											
Average length:	1																																																											
Maximum length:	1																																																											
Value	Count	Percentage																																																										
	108	1.00%																																																										
N	10660	98.79%																																																										
S	23	0.21%																																																										
Value	Count	Percentage																																																										
	108	1.00%																																																										
L	10683	99.00%																																																										
Expression:	RETIROCODACTOADMIN																																																											
Datatype:	INTEGER																																																											
Rows:	10,791																																																											
Non-unique 6.02% Null 6.67% Duplicate 93.41%		<table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>10,082</td></tr><tr><td>Null values:</td><td>709</td></tr><tr><td>Distinct values:</td><td>2</td></tr><tr><td>Duplicate values:</td><td>10,080</td></tr><tr><td>Unique values:</td><td>0</td></tr><tr><td>Non-unique values:</td><td>2</td></tr></tbody></table> <table><thead><tr><th>Type</th><th>value</th></tr></thead><tbody><tr><td>Minimum value:</td><td>0</td></tr><tr><td>Median value:</td><td>0</td></tr><tr><td>Maximum value:</td><td>3</td></tr><tr><td>Sum:</td><td>81</td></tr><tr><td>Average:</td><td>0.01</td></tr><tr><td>Variance:</td><td>0.02</td></tr><tr><td>Standard deviation:</td><td>0.16</td></tr></tbody></table>		Type	Count	Non-null values:	10,082	Null values:	709	Distinct values:	2	Duplicate values:	10,080	Unique values:	0	Non-unique values:	2	Type	value	Minimum value:	0	Median value:	0	Maximum value:	3	Sum:	81	Average:	0.01	Variance:	0.02	Standard deviation:	0.16	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td></td><td>709</td><td>6.57%</td></tr><tr><td>0</td><td>10055</td><td>93.18%</td></tr><tr><td>3</td><td>27</td><td>0.25%</td></tr></tbody></table>		Value	Count	Percentage		709	6.57%	0	10055	93.18%	3	27	0.25%	N/A		Es un campo de tipo entero, con un total de 10791 registros analizados y los cuales presentan las siguientes características: - El 93,41% datos duplicados equivalentes a 10080, el 6,57% de datos nulos equivalentes a 709 valores nulos y 2 valores distintos y 0 valores únicos y 2 valores no únicos equivalentes al 0,02%. - El campo con una frecuencia más alta corresponde es de 0 y equivale al 93,18%. - No hay una máscara de entrada definida para este campo.										
Type	Count																																																											
Non-null values:	10,082																																																											
Null values:	709																																																											
Distinct values:	2																																																											
Duplicate values:	10,080																																																											
Unique values:	0																																																											
Non-unique values:	2																																																											
Type	value																																																											
Minimum value:	0																																																											
Median value:	0																																																											
Maximum value:	3																																																											
Sum:	81																																																											
Average:	0.01																																																											
Variance:	0.02																																																											
Standard deviation:	0.16																																																											
Value	Count	Percentage																																																										
	709	6.57%																																																										
0	10055	93.18%																																																										
3	27	0.25%																																																										

Tabla 40: Perfil #1 de variables RETIROFECHAACTOADMIN y RETIRONUMEROACTOADMIN

General		Estadísticas	Frecuencia	Máscara/Extremos	Observaciones																																																																																																																																	
Expression:	RETIROFECHAACTOADMINISTRATIVO																																																																																																																																					
Datatype:	STRING																																																																																																																																					
Rows:	10,791																																																																																																																																					
		<table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>25</td></tr><tr><td>Null values:</td><td>10,766</td></tr><tr><td>Distinct values:</td><td>16</td></tr><tr><td>Duplicate values:</td><td>9</td></tr><tr><td>Unique values:</td><td>10</td></tr><tr><td>Non-unique values:</td><td>6</td></tr></tbody></table> <table><thead><tr><th>Type</th><th>value</th></tr></thead><tbody><tr><td>Minimum value:</td><td>05/26/2014 00:00:00</td></tr><tr><td>Median value:</td><td>07/23/2014 00:00:00</td></tr><tr><td>Maximum value:</td><td>08/31/2014 00:00:00</td></tr><tr><td>Minimum length:</td><td>19</td></tr><tr><td>Median length:</td><td>19</td></tr><tr><td>Average length:</td><td>19</td></tr><tr><td>Maximum length:</td><td>19</td></tr></tbody></table>	Type	Count	Non-null values:	25	Null values:	10,766	Distinct values:	16	Duplicate values:	9	Unique values:	10	Non-unique values:	6	Type	value	Minimum value:	05/26/2014 00:00:00	Median value:	07/23/2014 00:00:00	Maximum value:	08/31/2014 00:00:00	Minimum length:	19	Median length:	19	Average length:	19	Maximum length:	19	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td></td><td>10766</td><td>99.77%</td></tr><tr><td>07/18/2014 00:00:00</td><td>3</td><td>0.03%</td></tr><tr><td>07/21/2014 00:00:00</td><td>3</td><td>0.03%</td></tr><tr><td>07/28/2014 00:00:00</td><td>3</td><td>0.03%</td></tr><tr><td>07/22/2014 00:00:00</td><td>2</td><td>0.02%</td></tr><tr><td>08/06/2014 00:00:00</td><td>2</td><td>0.02%</td></tr><tr><td>08/11/2014 00:00:00</td><td>2</td><td>0.02%</td></tr><tr><td>05/26/2014 00:00:00</td><td>1</td><td>0.01%</td></tr><tr><td>06/24/2014 00:00:00</td><td>1</td><td>0.01%</td></tr><tr><td>07/23/2014 00:00:00</td><td>1</td><td>0.01%</td></tr><tr><td>07/24/2014 00:00:00</td><td>1</td><td>0.01%</td></tr><tr><td>07/29/2014 00:00:00</td><td>1</td><td>0.01%</td></tr><tr><td>08/13/2014 00:00:00</td><td>1</td><td>0.01%</td></tr><tr><td>08/15/2014 00:00:00</td><td>1</td><td>0.01%</td></tr><tr><td>08/01/2014 00:00:00</td><td>1</td><td>0.01%</td></tr><tr><td>08/04/2014 00:00:00</td><td>1</td><td>0.01%</td></tr><tr><td>05/05/2014 00:00:00</td><td>1</td><td>0.01%</td></tr></tbody></table>	Value	Count	Percentage		10766	99.77%	07/18/2014 00:00:00	3	0.03%	07/21/2014 00:00:00	3	0.03%	07/28/2014 00:00:00	3	0.03%	07/22/2014 00:00:00	2	0.02%	08/06/2014 00:00:00	2	0.02%	08/11/2014 00:00:00	2	0.02%	05/26/2014 00:00:00	1	0.01%	06/24/2014 00:00:00	1	0.01%	07/23/2014 00:00:00	1	0.01%	07/24/2014 00:00:00	1	0.01%	07/29/2014 00:00:00	1	0.01%	08/13/2014 00:00:00	1	0.01%	08/15/2014 00:00:00	1	0.01%	08/01/2014 00:00:00	1	0.01%	08/04/2014 00:00:00	1	0.01%	05/05/2014 00:00:00	1	0.01%	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td></td><td>10766</td><td>99.77%</td></tr><tr><td>DDDDDDDD DD-DD-DD</td><td>18</td><td>0.17%</td></tr><tr><td>DDDD-DD-DD DD-DD-DD</td><td>7</td><td>0.06%</td></tr></tbody></table>	Value	Count	Percentage		10766	99.77%	DDDDDDDD DD-DD-DD	18	0.17%	DDDD-DD-DD DD-DD-DD	7	0.06%	Este campo tiene en la mayoría de los campos vacíos y la herramienta lo determina como valor nulo en un 99,77%, es decir 10766 registros. - Es de aclarar que los docentes aquí descritos aparecen en su mayoría activos para el servicio de la Secretaría de Educación. - Las fechas de retiro del acto administrativo tiene una clasificación con 3 valores 1,2,3 y campo vacío. Existe una máscara de entrada con 2 tipos de formato definida para el campo Es un campo de tipo string, con un total de 10791 registros analizados que presentan las siguientes características: - 99,77% datos nulos 10766, 16 valores distintos, 10 valores únicos con el 0,06% , 0,08 valores duplicados equivalentes 9 y como valores no únicos estarían 6 registros.																																	
Type	Count																																																																																																																																					
Non-null values:	25																																																																																																																																					
Null values:	10,766																																																																																																																																					
Distinct values:	16																																																																																																																																					
Duplicate values:	9																																																																																																																																					
Unique values:	10																																																																																																																																					
Non-unique values:	6																																																																																																																																					
Type	value																																																																																																																																					
Minimum value:	05/26/2014 00:00:00																																																																																																																																					
Median value:	07/23/2014 00:00:00																																																																																																																																					
Maximum value:	08/31/2014 00:00:00																																																																																																																																					
Minimum length:	19																																																																																																																																					
Median length:	19																																																																																																																																					
Average length:	19																																																																																																																																					
Maximum length:	19																																																																																																																																					
Value	Count	Percentage																																																																																																																																				
	10766	99.77%																																																																																																																																				
07/18/2014 00:00:00	3	0.03%																																																																																																																																				
07/21/2014 00:00:00	3	0.03%																																																																																																																																				
07/28/2014 00:00:00	3	0.03%																																																																																																																																				
07/22/2014 00:00:00	2	0.02%																																																																																																																																				
08/06/2014 00:00:00	2	0.02%																																																																																																																																				
08/11/2014 00:00:00	2	0.02%																																																																																																																																				
05/26/2014 00:00:00	1	0.01%																																																																																																																																				
06/24/2014 00:00:00	1	0.01%																																																																																																																																				
07/23/2014 00:00:00	1	0.01%																																																																																																																																				
07/24/2014 00:00:00	1	0.01%																																																																																																																																				
07/29/2014 00:00:00	1	0.01%																																																																																																																																				
08/13/2014 00:00:00	1	0.01%																																																																																																																																				
08/15/2014 00:00:00	1	0.01%																																																																																																																																				
08/01/2014 00:00:00	1	0.01%																																																																																																																																				
08/04/2014 00:00:00	1	0.01%																																																																																																																																				
05/05/2014 00:00:00	1	0.01%																																																																																																																																				
Value	Count	Percentage																																																																																																																																				
	10766	99.77%																																																																																																																																				
DDDDDDDD DD-DD-DD	18	0.17%																																																																																																																																				
DDDD-DD-DD DD-DD-DD	7	0.06%																																																																																																																																				
	Free 0.05% Not-free 0.8% Typical 3.85% NA 99.77%																																																																																																																																					
Expression:	RETIRONUMEROACTOADMINISTRATIVO																																																																																																																																					
Datatype:	STRING																																																																																																																																					
Rows:	10,791																																																																																																																																					
		<table><thead><tr><th>Type</th><th>Count</th></tr></thead><tbody><tr><td>Non-null values:</td><td>27</td></tr><tr><td>Null values:</td><td>10,764</td></tr><tr><td>Distinct values:</td><td>27</td></tr><tr><td>Duplicate values:</td><td>0</td></tr><tr><td>Unique values:</td><td>27</td></tr><tr><td>Non-unique values:</td><td>0</td></tr></tbody></table> <table><thead><tr><th>Type</th><th>value</th></tr></thead><tbody><tr><td>Minimum value:</td><td>111</td></tr><tr><td>Median value:</td><td>8898</td></tr><tr><td>Maximum value:</td><td>9680</td></tr><tr><td>Minimum length:</td><td>3</td></tr><tr><td>Median length:</td><td>4</td></tr><tr><td>Average length:</td><td>3.36</td></tr><tr><td>Maximum length:</td><td>4</td></tr></tbody></table>	Type	Count	Non-null values:	27	Null values:	10,764	Distinct values:	27	Duplicate values:	0	Unique values:	27	Non-unique values:	0	Type	value	Minimum value:	111	Median value:	8898	Maximum value:	9680	Minimum length:	3	Median length:	4	Average length:	3.36	Maximum length:	4	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td></td><td>10764</td><td>99.75%</td></tr><tr><td>111</td><td>1</td><td>0.01%</td></tr><tr><td>1111</td><td>1</td><td>0.01%</td></tr><tr><td>5872</td><td>1</td><td>0.01%</td></tr><tr><td>6647</td><td>1</td><td>0.01%</td></tr><tr><td>7724</td><td>1</td><td>0.01%</td></tr><tr><td>8676</td><td>1</td><td>0.01%</td></tr><tr><td>8627</td><td>1</td><td>0.01%</td></tr><tr><td>8700</td><td>1</td><td>0.01%</td></tr><tr><td>8764</td><td>1</td><td>0.01%</td></tr><tr><td>8765</td><td>1</td><td>0.01%</td></tr><tr><td>8767</td><td>1</td><td>0.01%</td></tr><tr><td>8808</td><td>1</td><td>0.01%</td></tr><tr><td>8844</td><td>1</td><td>0.01%</td></tr><tr><td>8898</td><td>1</td><td>0.01%</td></tr><tr><td>8931</td><td>1</td><td>0.01%</td></tr><tr><td>8999</td><td>1</td><td>0.01%</td></tr><tr><td>9015</td><td>1</td><td>0.01%</td></tr><tr><td>9026</td><td>1</td><td>0.01%</td></tr><tr><td>9105</td><td>1</td><td>0.01%</td></tr><tr><td>9269</td><td>1</td><td>0.01%</td></tr><tr><td>9303</td><td>1</td><td>0.01%</td></tr><tr><td>9384</td><td>1</td><td>0.01%</td></tr><tr><td>9424</td><td>1</td><td>0.01%</td></tr><tr><td>9489</td><td>1</td><td>0.01%</td></tr><tr><td>9578</td><td>1</td><td>0.01%</td></tr><tr><td>9579</td><td>1</td><td>0.01%</td></tr><tr><td>9680</td><td>1</td><td>0.01%</td></tr></tbody></table>	Value	Count	Percentage		10764	99.75%	111	1	0.01%	1111	1	0.01%	5872	1	0.01%	6647	1	0.01%	7724	1	0.01%	8676	1	0.01%	8627	1	0.01%	8700	1	0.01%	8764	1	0.01%	8765	1	0.01%	8767	1	0.01%	8808	1	0.01%	8844	1	0.01%	8898	1	0.01%	8931	1	0.01%	8999	1	0.01%	9015	1	0.01%	9026	1	0.01%	9105	1	0.01%	9269	1	0.01%	9303	1	0.01%	9384	1	0.01%	9424	1	0.01%	9489	1	0.01%	9578	1	0.01%	9579	1	0.01%	9680	1	0.01%	<table><thead><tr><th>Value</th><th>Count</th><th>Percentage</th></tr></thead><tbody><tr><td></td><td>10764</td><td>99.75%</td></tr><tr><td>DDDD</td><td>25</td><td>0.24%</td></tr><tr><td>DDD</td><td>1</td><td>0.01%</td></tr></tbody></table>	Value	Count	Percentage		10764	99.75%	DDDD	25	0.24%	DDD	1	0.01%	Es un campo de tipo string, con un total de 10791 registros analizados que presentan las siguientes características: - El 99,75% datos nulos es decir 10764, 27 valores distintos, 27 valores únicos con el 0,25%, 0 valores duplicados y como valores no únicos estarían 0 registros. - Tiene dos máscaras de entrada determinada en letras, campos vacíos en un porcentaje del 99,75%; es un campo que tiene relevancia si el docente se retira de la planta.
Type	Count																																																																																																																																					
Non-null values:	27																																																																																																																																					
Null values:	10,764																																																																																																																																					
Distinct values:	27																																																																																																																																					
Duplicate values:	0																																																																																																																																					
Unique values:	27																																																																																																																																					
Non-unique values:	0																																																																																																																																					
Type	value																																																																																																																																					
Minimum value:	111																																																																																																																																					
Median value:	8898																																																																																																																																					
Maximum value:	9680																																																																																																																																					
Minimum length:	3																																																																																																																																					
Median length:	4																																																																																																																																					
Average length:	3.36																																																																																																																																					
Maximum length:	4																																																																																																																																					
Value	Count	Percentage																																																																																																																																				
	10764	99.75%																																																																																																																																				
111	1	0.01%																																																																																																																																				
1111	1	0.01%																																																																																																																																				
5872	1	0.01%																																																																																																																																				
6647	1	0.01%																																																																																																																																				
7724	1	0.01%																																																																																																																																				
8676	1	0.01%																																																																																																																																				
8627	1	0.01%																																																																																																																																				
8700	1	0.01%																																																																																																																																				
8764	1	0.01%																																																																																																																																				
8765	1	0.01%																																																																																																																																				
8767	1	0.01%																																																																																																																																				
8808	1	0.01%																																																																																																																																				
8844	1	0.01%																																																																																																																																				
8898	1	0.01%																																																																																																																																				
8931	1	0.01%																																																																																																																																				
8999	1	0.01%																																																																																																																																				
9015	1	0.01%																																																																																																																																				
9026	1	0.01%																																																																																																																																				
9105	1	0.01%																																																																																																																																				
9269	1	0.01%																																																																																																																																				
9303	1	0.01%																																																																																																																																				
9384	1	0.01%																																																																																																																																				
9424	1	0.01%																																																																																																																																				
9489	1	0.01%																																																																																																																																				
9578	1	0.01%																																																																																																																																				
9579	1	0.01%																																																																																																																																				
9680	1	0.01%																																																																																																																																				
Value	Count	Percentage																																																																																																																																				
	10764	99.75%																																																																																																																																				
DDDD	25	0.24%																																																																																																																																				
DDD	1	0.01%																																																																																																																																				
	Unfree 0.25% NA 99.75%																																																																																																																																					

Tabla 41: Perfil #1 de variables CODCAUSARETIRO y CAUSARETIRO

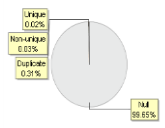
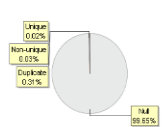
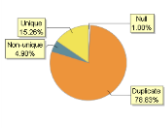
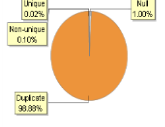
General		Estadísticas		Frecuencia		Máscara/Extremos		Observaciones																																																																					
Expression:	CODCAUSARETIRO							Es un campo de tipo string, con un total de 10791 registros analizados que presentan las siguientes características: - El 99,65% datos nulos es decir 10753, 5 valores distintos, 2 valores únicos con el 0,02%, 0,33% valores duplicados equivalentes a 33 y como valores no únicos estarían 3 registros equivalentes al 0,03%. - Su máscara de entrada está definida con una letra para todo el set de datos. - Es válido que los campos no contengan información ya que este campo solo se activa si el docente se retira de la planta. - En la escala de frecuencia los valores con mayor relevancia serían: valor 1= 0,20%, valor 7= 0,10%, valor 5= 0,03%, valor 4 =0,01%, valor 6= 0,01%																																																																					
Datatype:	STRING																																																																												
Rows:	10,791																																																																												
		<table><tr><th>Type</th><th>Count</th></tr><tr><td>Non-null values:</td><td>38</td></tr><tr><td>Null values:</td><td>10753</td></tr><tr><td>Distinct values:</td><td>5</td></tr><tr><td>Duplicate values:</td><td>33</td></tr><tr><td>Unique values:</td><td>2</td></tr><tr><td>Non-unique values:</td><td>3</td></tr></table> <table><tr><th>Type</th><th>value</th></tr><tr><td>Minimum value:</td><td>1</td></tr><tr><td>Median value:</td><td>1</td></tr><tr><td>Maximum value:</td><td>7</td></tr><tr><td>Minimum length:</td><td>1</td></tr><tr><td>Median length:</td><td>1</td></tr><tr><td>Average length:</td><td>1</td></tr><tr><td>Maximum length:</td><td>1</td></tr></table>		Type	Count	Non-null values:	38	Null values:	10753	Distinct values:	5	Duplicate values:	33	Unique values:	2	Non-unique values:	3	Type	value	Minimum value:	1	Median value:	1	Maximum value:	7	Minimum length:	1	Median length:	1	Average length:	1	Maximum length:	1	<table><tr><th>Value</th><th>Count</th><th>Percentage</th></tr><tr><td></td><td>10753</td><td>99.65%</td></tr><tr><td>1</td><td>22</td><td>0.20%</td></tr><tr><td>7</td><td>11</td><td>0.10%</td></tr><tr><td>5</td><td>3</td><td>0.03%</td></tr><tr><td>4</td><td>1</td><td>0.01%</td></tr><tr><td>6</td><td>1</td><td>0.01%</td></tr></table> <table><tr><th>Value</th><th>Count</th><th>Percentage</th></tr><tr><td></td><td>10753</td><td>99.65%</td></tr><tr><td>D</td><td>38</td><td>0.35%</td></tr></table>		Value	Count	Percentage		10753	99.65%	1	22	0.20%	7	11	0.10%	5	3	0.03%	4	1	0.01%	6	1	0.01%	Value	Count	Percentage		10753	99.65%	D	38	0.35%												
Type	Count																																																																												
Non-null values:	38																																																																												
Null values:	10753																																																																												
Distinct values:	5																																																																												
Duplicate values:	33																																																																												
Unique values:	2																																																																												
Non-unique values:	3																																																																												
Type	value																																																																												
Minimum value:	1																																																																												
Median value:	1																																																																												
Maximum value:	7																																																																												
Minimum length:	1																																																																												
Median length:	1																																																																												
Average length:	1																																																																												
Maximum length:	1																																																																												
Value	Count	Percentage																																																																											
	10753	99.65%																																																																											
1	22	0.20%																																																																											
7	11	0.10%																																																																											
5	3	0.03%																																																																											
4	1	0.01%																																																																											
6	1	0.01%																																																																											
Value	Count	Percentage																																																																											
	10753	99.65%																																																																											
D	38	0.35%																																																																											
Expression:	CAUSARETIRO							Es un campo de tipo string, con un total de 10791 registros analizados que presentan las siguientes características: - El 98,65% están diagnosticados por la herramienta como valores nulos es decir 10753 registros, un 0,31% de valores duplicados equivalentes a 33, 5 valores distintos y como valores únicos estarían 2 con el 0,02%. - En la escala de frecuencia de análisis los valores con mayor número estarían voluntario con un 0,20%, destitución por sanción 0,10%, fallecimiento 0,03%, retiro forzoso (edad) con el 0,01% y terminación provisionalidad 0,01%. - Para este campo están determinadas 5 máscaras de entrada en todo el set de datos con letras.																																																																					
Datatype:	STRING																																																																												
Rows:	10,791																																																																												
		<table><tr><th>Type</th><th>Count</th></tr><tr><td>Non-null values:</td><td>38</td></tr><tr><td>Null values:</td><td>10753</td></tr><tr><td>Distinct values:</td><td>5</td></tr><tr><td>Duplicate values:</td><td>33</td></tr><tr><td>Unique values:</td><td>2</td></tr><tr><td>Non-unique values:</td><td>3</td></tr></table> <table><tr><th>Type</th><th>value</th></tr><tr><td>Minimum value:</td><td>Destitucion por Sancion</td></tr><tr><td>Median value:</td><td>Voluntario</td></tr><tr><td>Maximum value:</td><td>Voluntario</td></tr><tr><td>Minimum length:</td><td>1</td></tr><tr><td>Median length:</td><td>10</td></tr><tr><td>Average length:</td><td>14.74</td></tr><tr><td>Maximum length:</td><td>27</td></tr></table>		Type	Count	Non-null values:	38	Null values:	10753	Distinct values:	5	Duplicate values:	33	Unique values:	2	Non-unique values:	3	Type	value	Minimum value:	Destitucion por Sancion	Median value:	Voluntario	Maximum value:	Voluntario	Minimum length:	1	Median length:	10	Average length:	14.74	Maximum length:	27	<table><tr><th>Value</th><th>Count</th><th>Percentage</th></tr><tr><td></td><td>10753</td><td>99.65%</td></tr><tr><td>Voluntario</td><td>22</td><td>0.20%</td></tr><tr><td>Destitucion por Sancion</td><td>11</td><td>0.10%</td></tr><tr><td>Fallecimiento</td><td>3</td><td>0.03%</td></tr><tr><td>Retiro Forzoso (Edad)</td><td>1</td><td>0.01%</td></tr><tr><td>Terminacion Provisionalidad</td><td>1</td><td>0.01%</td></tr></table> <table><tr><th>Value</th><th>Count</th><th>Percentage</th></tr><tr><td></td><td>10753</td><td>99.65%</td></tr><tr><td>LLLLLLLLLL</td><td>22</td><td>0.20%</td></tr><tr><td>LLLLLLLLLL LLL</td><td>11</td><td>0.10%</td></tr><tr><td>LLLLLL</td><td>3</td><td>0.03%</td></tr><tr><td>LLLLLLLLLLLLLL (LLLL)</td><td>1</td><td>0.01%</td></tr><tr><td>LLLLLLLLLLLLLL</td><td>1</td><td>0.01%</td></tr></table>		Value	Count	Percentage		10753	99.65%	Voluntario	22	0.20%	Destitucion por Sancion	11	0.10%	Fallecimiento	3	0.03%	Retiro Forzoso (Edad)	1	0.01%	Terminacion Provisionalidad	1	0.01%	Value	Count	Percentage		10753	99.65%	LLLLLLLLLL	22	0.20%	LLLLLLLLLL LLL	11	0.10%	LLLLLL	3	0.03%	LLLLLLLLLLLLLL (LLLL)	1	0.01%	LLLLLLLLLLLLLL	1	0.01%
Type	Count																																																																												
Non-null values:	38																																																																												
Null values:	10753																																																																												
Distinct values:	5																																																																												
Duplicate values:	33																																																																												
Unique values:	2																																																																												
Non-unique values:	3																																																																												
Type	value																																																																												
Minimum value:	Destitucion por Sancion																																																																												
Median value:	Voluntario																																																																												
Maximum value:	Voluntario																																																																												
Minimum length:	1																																																																												
Median length:	10																																																																												
Average length:	14.74																																																																												
Maximum length:	27																																																																												
Value	Count	Percentage																																																																											
	10753	99.65%																																																																											
Voluntario	22	0.20%																																																																											
Destitucion por Sancion	11	0.10%																																																																											
Fallecimiento	3	0.03%																																																																											
Retiro Forzoso (Edad)	1	0.01%																																																																											
Terminacion Provisionalidad	1	0.01%																																																																											
Value	Count	Percentage																																																																											
	10753	99.65%																																																																											
LLLLLLLLLL	22	0.20%																																																																											
LLLLLLLLLL LLL	11	0.10%																																																																											
LLLLLL	3	0.03%																																																																											
LLLLLLLLLLLLLL (LLLL)	1	0.01%																																																																											
LLLLLLLLLLLLLL	1	0.01%																																																																											

Tabla 42: Perfil #1 de variables FECHA CREACION y USUARIO

General		Estadísticas		Frecuencia		Máscara/Extremos			Observaciones																																																																																																													
Expression:	FECHA CREACION								Es un campo de tipo string, con un total de 10791 registros analizados que presentan las siguientes características: - El 78,83% de los registros son diagnosticados como duplicados, esto es un total de 8507 de 10791, un 1,00% como datos nulos es decir unos 108 registros, un 15,26% valores únicos o sea 1647 y un 4,90% como valores no únicos o sea 529 registros. - Existe una máscara de entrada con 2 tipos de formato definida para el campo.																																																																																																													
Datatype:	STRING																																																																																																																					
Rows:	10,791																																																																																																																					
		<table><tr><th>Type</th><th>Count</th></tr><tr><td>Non-null values:</td><td>10,683</td></tr><tr><td>Null values:</td><td>108</td></tr><tr><td>Distinct values:</td><td>13</td></tr><tr><td>Duplicate values:</td><td>2,176</td></tr><tr><td>Duplicate values:</td><td>8,507</td></tr><tr><td>Unique values:</td><td>1,647</td></tr><tr><td>Non-unique values:</td><td>529</td></tr></table> <table><tr><th>Type</th><th>value</th></tr><tr><td>Minimum value:</td><td>01/13/2003 00:00:00</td></tr><tr><td>Median value:</td><td>01/01/2003 00:00:00</td></tr><tr><td>Maximum value:</td><td>08/12/2014 16:27</td></tr><tr><td>Minimum length:</td><td>19</td></tr><tr><td>Median length:</td><td>19</td></tr><tr><td>Average length:</td><td>19</td></tr><tr><td>Maximum length:</td><td>19</td></tr></table>		Type	Count	Non-null values:	10,683	Null values:		108	Distinct values:	13	Duplicate values:	2,176	Duplicate values:	8,507	Unique values:	1,647	Non-unique values:	529	Type	value	Minimum value:	01/13/2003 00:00:00	Median value:	01/01/2003 00:00:00	Maximum value:	08/12/2014 16:27	Minimum length:	19	Median length:	19	Average length:	19	Maximum length:	19	<table><tr><th>Value</th><th>Count</th><th>Percentage</th></tr><tr><td></td><td>108</td><td>1.00%</td></tr><tr><td>DDDD-DD-DD DDDDD-DD</td><td>6807</td><td>63.08%</td></tr><tr><td>DDDDDDDD DD-DDDD</td><td>3871</td><td>35.92%</td></tr></table>		Value	Count	Percentage		108	1.00%	DDDD-DD-DD DDDDD-DD	6807	63.08%	DDDDDDDD DD-DDDD	3871	35.92%																																																																				
Type	Count																																																																																																																					
Non-null values:	10,683																																																																																																																					
Null values:	108																																																																																																																					
Distinct values:	13																																																																																																																					
Duplicate values:	2,176																																																																																																																					
Duplicate values:	8,507																																																																																																																					
Unique values:	1,647																																																																																																																					
Non-unique values:	529																																																																																																																					
Type	value																																																																																																																					
Minimum value:	01/13/2003 00:00:00																																																																																																																					
Median value:	01/01/2003 00:00:00																																																																																																																					
Maximum value:	08/12/2014 16:27																																																																																																																					
Minimum length:	19																																																																																																																					
Median length:	19																																																																																																																					
Average length:	19																																																																																																																					
Maximum length:	19																																																																																																																					
Value	Count	Percentage																																																																																																																				
	108	1.00%																																																																																																																				
DDDD-DD-DD DDDDD-DD	6807	63.08%																																																																																																																				
DDDDDDDD DD-DDDD	3871	35.92%																																																																																																																				
Expression:	USUARIO								Es un campo de tipo dato string, con un total de 10791 registros analizados que presentan las siguientes características: - El 98,88% de los registros son diagnosticados como duplicados, esto es un total de 10670 de 10791, un 1,00% como datos nulos es decir unos 108 registros, un 0,02% valores únicos o sea 2 y un 0,10% como valores no únicos o sea 11 registros. - El campo se refiere básicamente a las direcciones de correo electrónico de los docentes. - De acuerdo con el análisis de frecuencia existe un dato no correspondiente al formato de dirección electrónica con un 93,19%, es decir, 10056 registros con información no conforme o errónea. - El campo tiene 11 tipos de formato de finidos para todo el set de datos.																																																																																																													
Datatype:	STRING																																																																																																																					
Rows:	10,791																																																																																																																					
		<table><tr><th>Type</th><th>Count</th></tr><tr><td>Non-null values:</td><td>10,683</td></tr><tr><td>Null values:</td><td>108</td></tr><tr><td>Distinct values:</td><td>13</td></tr><tr><td>Duplicate values:</td><td>10,67</td></tr><tr><td>Duplicate values:</td><td>2</td></tr><tr><td>Unique values:</td><td>11</td></tr><tr><td>Non-unique values:</td><td>11</td></tr></table> <table><tr><th>Type</th><th>value</th></tr><tr><td>Minimum value:</td><td>beatriz.lopez@medellin.gov.co</td></tr><tr><td>Median value:</td><td>Migacion</td></tr><tr><td>Maximum value:</td><td>paula.correa@medellin.gov.co</td></tr><tr><td>Minimum length:</td><td>4</td></tr><tr><td>Median length:</td><td>9</td></tr><tr><td>Average length:</td><td>10.09</td></tr><tr><td>Maximum length:</td><td>33</td></tr></table>		Type	Count	Non-null values:	10,683	Null values:		108	Distinct values:	13	Duplicate values:	10,67	Duplicate values:	2	Unique values:	11	Non-unique values:	11	Type	value	Minimum value:	beatriz.lopez@medellin.gov.co	Median value:	Migacion	Maximum value:	paula.correa@medellin.gov.co	Minimum length:	4	Median length:	9	Average length:	10.09	Maximum length:	33	<table><tr><th>Value</th><th>Count</th><th>Percentage</th></tr><tr><td>Migacion</td><td>10056</td><td>93.19%</td></tr><tr><td>elkin.mejia@medellin.gov.co</td><td>270</td><td>2.58%</td></tr><tr><td>paula.correa@medellin.gov.co</td><td>168</td><td>1.56%</td></tr><tr><td>paula.cifuenteg@medellin.gov.co</td><td>49</td><td>0.45%</td></tr><tr><td>liliana.sanchez@medellin.gov.co</td><td>46</td><td>0.43%</td></tr><tr><td>beatriz.lopez@medellin.gov.co</td><td>38</td><td>0.35%</td></tr><tr><td>humano</td><td>18</td><td>0.17%</td></tr><tr><td>lina.montoya@medellin.gov.co</td><td>12</td><td>0.11%</td></tr><tr><td>enrique.zuluaga@medellin.gov.co</td><td>10</td><td>0.09%</td></tr><tr><td>liliana.mesa@medellin.gov.co</td><td>3</td><td>0.03%</td></tr><tr><td>jordi.carmona@medellin.gov.co</td><td>3</td><td>0.03%</td></tr><tr><td>carlos.albarracín@medellin.gov.co</td><td>1</td><td>0.01%</td></tr><tr><td>monica.macias@medellin.gov.co</td><td>1</td><td>0.01%</td></tr></table> <table><tr><th>Value</th><th>Count</th><th>Percentage</th></tr><tr><td></td><td>108</td><td>1.00%</td></tr><tr><td>LLLLLLLLLL</td><td>10056</td><td>93.19%</td></tr><tr><td>LLLLLLLLLL@LLLLLLLLLL</td><td>270</td><td>2.58%</td></tr><tr><td>LLLLLLLLLL@LLLLLLLLLL</td><td>168</td><td>1.56%</td></tr><tr><td>LLLLLLLLLLLL@LLLLLLLLLL</td><td>56</td><td>0.52%</td></tr><tr><td>LLLLLLLLLLLL@LLLLLLLLLL</td><td>49</td><td>0.45%</td></tr><tr><td>LLLLLLLLLLLL@LLLLLLLLLL</td><td>38</td><td>0.35%</td></tr><tr><td>LLLLLLLL</td><td>18</td><td>0.17%</td></tr><tr><td>LLLLLLLLLLLL@LLLLLLLLLL</td><td>15</td><td>0.14%</td></tr><tr><td>LLLLLLLLLLLL@LLLLLLLLLL</td><td>3</td><td>0.03%</td></tr><tr><td>LLLLLLLLLLLL@LLLLLLLLLL</td><td>1</td><td>0.01%</td></tr><tr><td>LLLLLLLLLLLL@LLLLLLLLLL@L</td><td>1</td><td>0.01%</td></tr></table>		Value	Count	Percentage	Migacion	10056	93.19%	elkin.mejia@medellin.gov.co	270	2.58%	paula.correa@medellin.gov.co	168	1.56%	paula.cifuenteg@medellin.gov.co	49	0.45%	liliana.sanchez@medellin.gov.co	46	0.43%	beatriz.lopez@medellin.gov.co	38	0.35%	humano	18	0.17%	lina.montoya@medellin.gov.co	12	0.11%	enrique.zuluaga@medellin.gov.co	10	0.09%	liliana.mesa@medellin.gov.co	3	0.03%	jordi.carmona@medellin.gov.co	3	0.03%	carlos.albarracín@medellin.gov.co	1	0.01%	monica.macias@medellin.gov.co	1	0.01%	Value	Count	Percentage		108	1.00%	LLLLLLLLLL	10056	93.19%	LLLLLLLLLL@LLLLLLLLLL	270	2.58%	LLLLLLLLLL@LLLLLLLLLL	168	1.56%	LLLLLLLLLLLL@LLLLLLLLLL	56	0.52%	LLLLLLLLLLLL@LLLLLLLLLL	49	0.45%	LLLLLLLLLLLL@LLLLLLLLLL	38	0.35%	LLLLLLLL	18	0.17%	LLLLLLLLLLLL@LLLLLLLLLL	15	0.14%	LLLLLLLLLLLL@LLLLLLLLLL	3	0.03%	LLLLLLLLLLLL@LLLLLLLLLL	1	0.01%	LLLLLLLLLLLL@LLLLLLLLLL@L	1
Type	Count																																																																																																																					
Non-null values:	10,683																																																																																																																					
Null values:	108																																																																																																																					
Distinct values:	13																																																																																																																					
Duplicate values:	10,67																																																																																																																					
Duplicate values:	2																																																																																																																					
Unique values:	11																																																																																																																					
Non-unique values:	11																																																																																																																					
Type	value																																																																																																																					
Minimum value:	beatriz.lopez@medellin.gov.co																																																																																																																					
Median value:	Migacion																																																																																																																					
Maximum value:	paula.correa@medellin.gov.co																																																																																																																					
Minimum length:	4																																																																																																																					
Median length:	9																																																																																																																					
Average length:	10.09																																																																																																																					
Maximum length:	33																																																																																																																					
Value	Count	Percentage																																																																																																																				
Migacion	10056	93.19%																																																																																																																				
elkin.mejia@medellin.gov.co	270	2.58%																																																																																																																				
paula.correa@medellin.gov.co	168	1.56%																																																																																																																				
paula.cifuenteg@medellin.gov.co	49	0.45%																																																																																																																				
liliana.sanchez@medellin.gov.co	46	0.43%																																																																																																																				
beatriz.lopez@medellin.gov.co	38	0.35%																																																																																																																				
humano	18	0.17%																																																																																																																				
lina.montoya@medellin.gov.co	12	0.11%																																																																																																																				
enrique.zuluaga@medellin.gov.co	10	0.09%																																																																																																																				
liliana.mesa@medellin.gov.co	3	0.03%																																																																																																																				
jordi.carmona@medellin.gov.co	3	0.03%																																																																																																																				
carlos.albarracín@medellin.gov.co	1	0.01%																																																																																																																				
monica.macias@medellin.gov.co	1	0.01%																																																																																																																				
Value	Count	Percentage																																																																																																																				
	108	1.00%																																																																																																																				
LLLLLLLLLL	10056	93.19%																																																																																																																				
LLLLLLLLLL@LLLLLLLLLL	270	2.58%																																																																																																																				
LLLLLLLLLL@LLLLLLLLLL	168	1.56%																																																																																																																				
LLLLLLLLLLLL@LLLLLLLLLL	56	0.52%																																																																																																																				
LLLLLLLLLLLL@LLLLLLLLLL	49	0.45%																																																																																																																				
LLLLLLLLLLLL@LLLLLLLLLL	38	0.35%																																																																																																																				
LLLLLLLL	18	0.17%																																																																																																																				
LLLLLLLLLLLL@LLLLLLLLLL	15	0.14%																																																																																																																				
LLLLLLLLLLLL@LLLLLLLLLL	3	0.03%																																																																																																																				
LLLLLLLLLLLL@LLLLLLLLLL	1	0.01%																																																																																																																				
LLLLLLLLLLLL@LLLLLLLLLL@L	1	0.01%																																																																																																																				