

## Método para la retención de clientes de telefonía móvil pospago

Víctor H. Aristizábal O.  
Máster Universitario en Análisis y Visualización de Datos Masivos  
Universidad Internacional de La Rioja  
Bogotá, Colombia

### Notas del autor

Víctor H. Aristizábal O, facultad de Ingeniería y Tecnología, Universidad Internacional de La Rioja – Bogotá

Víctor H. Aristizábal O se encuentra actualmente en la facultad de Ingeniería y Tecnología de la Universidad Internacional de La Rioja, Bogotá.

Este trabajo cuenta con la corrección de estilo del director Francisco José Soltero Domingo, facultad de Ingeniería y Tecnología de la Universidad Internacional de La Rioja.

## Resumen

La principal amenaza para las empresas de telefonía móvil es la pérdida de abonados, debido a que es más costoso para una compañía del sector atraer nuevos clientes que mantener clientes satisfechos.

Por tal razón, en este trabajo de fin de máster se tiene como objetivo principal ayudar a la disminución en la pérdida de abonados para la compañía, identificando los posibles clientes que estén pensando realizar el cambio de operador móvil.

En este trabajo de fin de máster se propone un método de gestión de datos para la retención de clientes de telefonía móvil pospago, lo cual incluye la identificación de las peticiones, quejas y reclamos (PQR's) generadas por los usuarios, para identificar los clientes potencialmente interesados en cambiar de operador de telefonía móvil, y mostrar visualizaciones de la data, para que el área de mercadeo de la compañía pueda definir los pasos a seguir para la gestión de retención de clientes, y otras áreas de la compañía pueden realizar acciones para disminuir los PQR's.

El sustento teórico del trabajo está directamente relacionado con el concepto de churn que se utiliza internacionalmente para la gestión de abandono de servicios contratados mediante suscripción, en consecuencia, la conclusión principal de este trabajo se debe enfocar en identificar los abonados que potencialmente pretenden realizar el cambio de operador móvil.

**Palabras Clave:** gestión de datos, retención, clientes, telefonía móvil, método

## Abstract

The main threat to mobile phone companies is the loss of subscribers, since it is more expensive for a company in the sector to attract new customers than to keep satisfied customers.

For this reason, the main objective of this master's thesis is to help reduce the loss of subscribers for the company, identifying potential customers who are thinking of making the change of mobile operator.

In this master's thesis, a data management method is proposed for the retention of post-paid mobile phone customers, which includes the identification of requests, complaints and claims (PQR's) generated by users, to identify the customers potentially interested in changing mobile phone operator, and show visualizations of the data, so that the company's marketing area can define the steps to follow for customer retention management, and other areas of the company can take actions to lower the PQR's.

The theoretical support of the work is directly related to the concept of churn that is used internationally for the management of abandonment of services contracted by subscription, consequently, the main conclusion of this work should focus on identifying the subscribers who potentially intend to perform the change of mobile operator.

**Keywords:** data management, retention, customers, mobile phone, method.

# Índice

Resumen .....	2
Abstract.....	3
1. Introducción.....	8
1.1. Justificación.....	8
1.2. Planteamiento del trabajo .....	10
1.3. Estructura del documento.....	11
2. Estado del arte .....	13
2.1. Contexto actual .....	17
2.2.1. Criterios de inclusión y exclusión .....	17
2.2.2. Estrategias de búsqueda .....	18
2.2.3. Revisión bibliográfica .....	18
2.2.4. Analítica visual.....	26
2.2.5. Tecnologías .....	28
2.2. Metodologías.....	31
2.2.1. KDD.....	31
2.2.2. CRISP-DM.....	32
2.2.3. SEMMA .....	35
3. Objetivos .....	37
3.1. Objetivo general .....	37
3.2. Objetivos específicos.....	37
4. Metodología .....	38
5. Desarrollo del método .....	41
5.1. Evaluación Inicial.....	41
5.2. Fuente de datos y análisis de variables .....	47
5.3. Diseño del modelo de datos .....	50
5.3.1. Extraer .....	52
5.3.2. Transformar .....	55
5.3.3. Cargar .....	56
5.4. Presentación de la herramienta de diseño y visualización .....	59
6. Resultados .....	62
a. Resumen.....	62
b. Motivos y causas .....	69
c. Cierres y solución.....	72
d. Mapa .....	78
7. Conclusiones.....	80
8. Trabajo futuro .....	82
9. Referencias bibliográficas.....	83

## Índice de figuras

	Pág.
Figura 1 Porcentaje de clientes retirados .....	13
Figura 2 Técnica de minería de datos .....	23
Figura 3 Método minería de datos.....	24
Figura 4 Herramientas minería de datos .....	25
Figura 5 Alcance de la analítica visual .....	27
Figura 6 Proceso de la analítica visual .....	28
Figura 7 Relación fases software utilizado .....	29
Figura 8 Pasos ejecución metodología KDD .....	32
Figura 9 Fases metodología CRISP-DM .....	33
Figura 10 Fases metodología SEMMA - GUI .....	36
Figura 11 Tablero de control .....	42
Figura 12 Tablero de control - Ingresos.....	43
Figura 13 Tablero de control - Efectividad.....	44
Figura 14 Tablero de control - Envíos .....	44
Figura 15 Tablero de control - Cierres.....	45
Figura 16 Tablero de control - Sin correspondencia .....	46
Figura 17 Tablero de control - Productividad.....	46
Figura 18 Procedimiento transformación.....	51
Figura 19 Fases ETL.....	51
Figura 20 Extraer información PQR's - Abonados.....	52
Figura 21 Informe ingresos por canal creador .....	53
Figura 22 Informe Abonados.....	53
Figura 23 Resultado información Abonados.....	54
Figura 24 Extraer información archivos adicionales .....	54
Figura 25 Transformación PQR's - Abonados .....	55
Figura 26 Transformaciones adicionales.....	56
Figura 27 Cargue información BD .....	57
Figura 28 Información en la BD.....	57
Figura 29 PQR's Mes.....	62
Figura 30 PQR's Clase .....	63
Figura 31 PQR's Tipo de cliente.....	64
Figura 32 PQR's Estado.....	64

Figura 33 Total valor ajustado.....	65
Figura 34 Resumen filtro departamento ciudad.....	66
Figura 35 Filtro mes .....	67
Figura 36 Filtro clase.....	67
Figura 37 Filtro tipo de cliente .....	68
Figura 38 Filtro estado .....	68
Figura 39 PQR's Motivos .....	69
Figura 40 PQR's Causal.....	70
Figura 41 Filtro Motivos y causas departamento ciudad.....	71
Figura 42 Filtro motivos.....	71
Figura 43 Filtro causal.....	72
Figura 44 PQR's tipo de cierre .....	73
Figura 45 PQR's días de solución .....	73
Figura 46 PQR's canal.....	74
Figura 47 PQR's área de solución.....	74
Figura 48 Filtro cierres y solución departamento ciudad .....	75
Figura 49 Filtro tipo de cierre .....	76
Figura 50 Filtro días solución .....	76
Figura 51 Filtro canal .....	77
Figura 52 Filtro área de solución .....	77
Figura 53 Mapa.....	78
Figura 54 Filtro mapa.....	79

## Índice de tablas

	Pág.
Tabla 1 PQR's.....	47
Tabla 2 Abonados.....	48
Tabla 3 Campos generados ETL .....	58

# 1. Introducción

## 1.1. Justificación

El Churn es un sistema de análisis de rotación o Churn que ayuda a una empresa a analizar, predecir y reducir la rotación de clientes, este sistema analiza las experiencias de los clientes para correlacionar esa experiencia con la rotación de clientes, por lo que, puede predecir con mayor precisión y reducir efectivamente la pérdida de clientes en el futuro. Como resultado de lo anterior, las empresas pueden reducir los costos de adquisición de clientes y mejorar las tasas de retención de clientes (Amin, y otros, 2017).

El objetivo del Churn es medir la infidelidad o la falta de lealtad de los clientes, se calcula como el porcentaje de clientes perdidos en un período frente al total de clientes vigentes al comienzo de ese mismo período (Brândușoiu, Todorean, & Beleiu, 2016):

$$\text{Churn rate} = \frac{\# \text{ de clientes perdidos en el periodo}}{\# \text{ de clientes al comienzo del periodo}} * 100$$

La cancelación se refiere a la desactivación de una cuenta de cliente activa. Las organizaciones que proporcionan o entregan información en un entorno de competencia significativa son particularmente propensas a la rotación, dichas organizaciones incluyen proveedores de telecomunicaciones celulares, proveedores de servicios de localización, proveedores de televisión por cable, proveedores de bases de datos en línea, y proveedores de servicios de telecomunicaciones de larga distancia y similares (Zhu, Baesens, & Broucke, 2017).

La situación descrita también se presenta en el negocio de telefonía móvil, en este mercado y particularmente en el segmento de los contratos de pospago, los clientes suscriben un contrato de adhesión que tiene una tarifa preestablecida mensual por la cual reciben a cambio un paquete que puede incluir llamadas telefónicas a diferentes operadores, acceso móvil a internet y a determinadas redes sociales, entre otros servicios; aunque esos elementos suelen ser los



más promocionados y los que usualmente se emplean para diferenciar los distintos planes y sus correspondientes tarifas, los clientes también reciben otros servicios que necesariamente hacen parte de todos los paquetes, como son los servicios de atención telefónica o digital para la generación de una petición, queja o reclamo o un recurso, conocido como PQR's. De acuerdo con la Comisión de Regulación de Comunicaciones (CRC), entidad pública encargada de la vigilancia y control de las empresas del sector de las comunicaciones en Colombia, con los servicios antes descritos se busca una libre competencia entre los operadores y en consecuencia se evita un monopolio del mercado por parte de alguna de estas empresas. Así las cosas, la entidad mencionada tiene como fin principal las búsquedas de altos niveles de calidad del servicio por parte de los operadores de telecomunicaciones.

Mediante la Resolución 5111 de 2017 “por la cual se establece el Régimen de Protección de los Derechos de los Usuarios de Servicios de Comunicaciones, se modifica el Capítulo 1 del Título II de la Resolución CRC 5050 de 2016 y se dictan otras disposiciones”, expedida por la CRC, la entidad aclara lo referente a las PQR: (CRC, 2017)

**Petición:** Solicitud de servicios o de información en relación con los servicios prestados por el operador, o cualquier manifestación del usuario en relación con sus derechos.

**Queja o reclamo:** Manifestación de inconformidad por parte del usuario al operador en relación con la prestación de sus servicios o el ejercicio de sus derechos.

**Recurso:** Manifestación de inconformidad del usuario en relación con la decisión tomada por el operador de telefonía y/o de internet frente a una queja presentada (relacionada con actos de negativa del contrato, suspensión del servicio, terminación del contrato, corte y facturación), y mediante la cual solicita la revisión por parte del operador (recurso de reposición) y en forma subsidiaria la revisión y decisión de la Superintendencia de Industria y Comercio (recurso en subsidio de apelación)

Esos factores mencionados que hacen parte del servicio de telefonía móvil inciden en que el cliente decida mantenerse vinculado contractualmente a un determinado operador, o que en determinado momento prefiera cambiarse a otro que le ofrezca mejores expectativas, ya sea respecto de las especificaciones del paquete (megabytes de internet al mes, minutos de comunicación telefónica y acceso a redes sociales frente a la tarifa) o respecto a los valores agregados (PQR, facturación, recaudo, cobertura, entre otros), en ese sentido, se habla entonces de la tasa de cancelación o abandono, conocida como Churn, que mide el porcentaje periódico de clientes que deciden cancelar los contratos de telefonía móvil pospago.

La situación descrita en el párrafo anterior conlleva al mejoramiento de la tasa Churn en las empresas de telefonía móvil que requieren el uso de un método de gestión masiva de datos, de manera que, se puedan identificar oportunamente algunas características o tendencias dentro de las bases de datos, con el propósito de adelantar acciones preventivas tendientes a mejorar la fidelidad de los clientes.

En consecuencia, y con relación al abandono de clientes, se identifica una oportunidad en la Empresa de Telecomunicaciones de Bogotá (ETB) de desarrollar un método práctico e innovador en el cual se utilicen técnicas para el análisis de datos y la visualización de la información para la gestión de los clientes potenciales a abandono.

## **1.2. Planteamiento del trabajo**

La razón fundamental de la existencia de las empresas en el mercado son los clientes, quiénes son el activo principal de cualquier compañía, por tal razón, las empresas se centran en mantener a sus clientes satisfechos y generar estrategias de marketing para obtener nuevos clientes, esto debido a que, es más costoso atraer un nuevo cliente que mantener un cliente satisfecho, es una proporción de seis a siete veces más. (Marketing directo, 2020)

En conclusión, retener clientes para las empresas es mucho más importante que captar nuevos, por tal razón, los esfuerzos se deben centrar en la retención de clientes, pero sin dejar de lado la necesidad de captación de nuevos clientes.

“Dentro de las principales razones para que un cliente deje de comprar productos de una compañía se destacan la disconformidad y la falta de políticas de retención efectivas expresadas en un mejor trato hacia ellos” (Barrientos & Ríos, Aplicación de Minería de Datos para Predecir Fuga de Clientes en la Industria de las Telecomunicaciones, 2013)

Por lo anterior, el TFM tiene como objetivo desarrollar un método para la gestión de los datos para la retención de clientes, con el fin de coadyuvar en la toma de decisiones con relación a la retención de los abonados, beneficiando de esta forma al desarrollo económico de la empresa.

En una etapa inicial del TFM, se realizará la recolección de datos de las PQR's generadas por los clientes, a fin de realizar el análisis de variables y la transformación de los datos, por consiguiente, una vez finalizada la primera fase, se procederá a diseñar e implementar la estructura del conjunto de los datos, lo cual estará enmarcado dentro de la fase dos; por lo que, en una última fase, se realizará la generación de visualizaciones para el análisis de la información.

### **1.3. Estructura del documento**

A continuación, se presenta la estructura y contenido definida para el presente documento, buscando cumplir el objetivo planteado.

En el capítulo 2, se encuentra definido el estado del arte, en el que se muestra la visualización como solución de análisis de datos que, si se desarrolla bajo unos conceptos teóricos, optimiza el análisis y la toma de decisiones empresariales.

En el capítulo 3, se definen los objetivos, tanto generales como específicos.

En el capítulo 4, se despliega el marco teórico, sustentando el porqué del desarrollo del método; cabe aclarar que en este capítulo se aplica el conocimiento adquirido en el máster, tales como, conceptos teóricos y manejo de herramientas relacionadas con el análisis y visualización de la información.

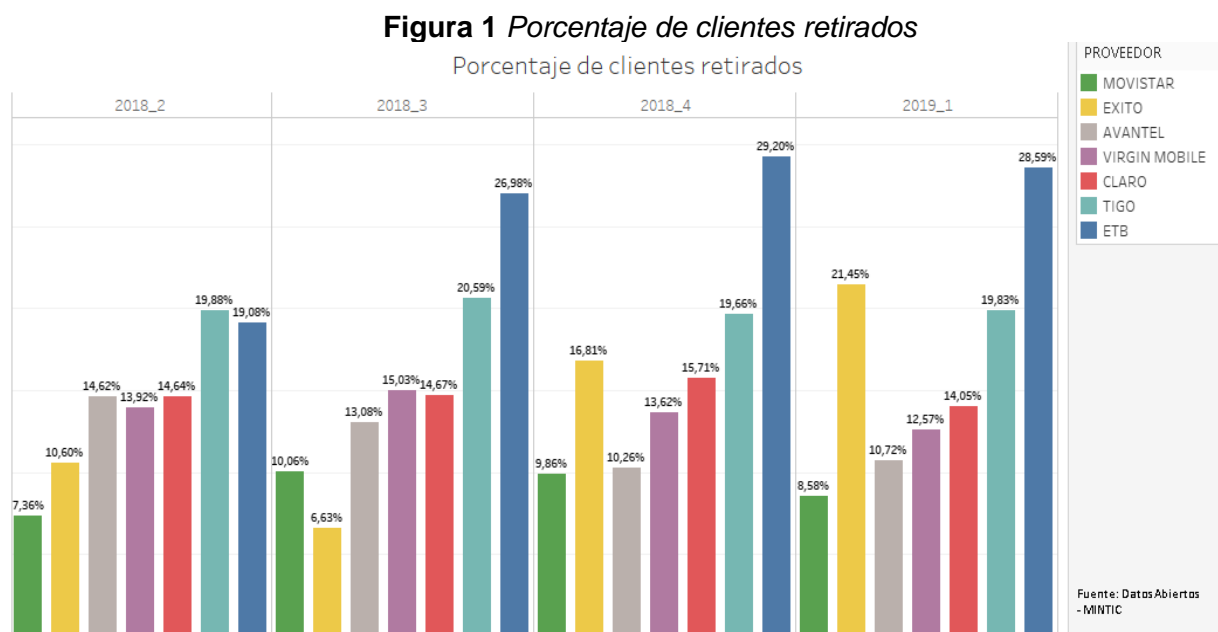
En el capítulo 5, se desarrolla el método para la retención de clientes, el cual está definido en tres fases, teniendo como fase inicial el análisis de datos, una fase Intermedia donde se define la estructura de la información, y por último, la fase donde se generan las visualizaciones de la información.

En el capítulo 6, se realiza un análisis de documento, en el que se presentan los resultados y conclusiones del método desarrollado.

## 2. Estado del arte

En la actualidad las empresas de telecomunicaciones presentan una alta cancelación de los servicios pospago contratados por los clientes, debido a que, en todo momento el abonado está tratando de buscar el mejor servicio ofrecido por las empresas del sector, y esto hace que los clientes cambien de operador buscando el mejor costo beneficio.

Al analizar la información relacionada con la tasa de cancelación de usuarios en el sector de las telecomunicaciones en Colombia, se tiene que, la información más reciente pertenece al primer semestre del año 2021, la cual se encuentra recopilada y publicada por el Ministerio de las Telecomunicaciones de Colombia, a través del portal de datos abiertos, en donde se tienen todos los datos publicados por todas la entidades públicas del país, los cuales pueden ser consultados y manipulados por los usuarios. (MINTIC, 2020). En la figura 1 se puede observar con detalle el porcentaje de retiro de clientes de los operadores móviles



Fuente: Elaboración propia con base a datos abiertos MINTIC

En la gráfica se observa los datos de los últimos tres trimestres del año 2018 hasta el primer trimestre del 2019, correspondiente a las siete compañías del sector de la telefonía móvil en Colombia que más abonados poseen.

Después de realizar un análisis de la representación gráfica, se puede concluir que la ETB en el rango de medición, es la empresa del sector de las telecomunicaciones que más tasa de cancelación de clientes tiene.

Por tal razón, se tiene la necesidad de analizar los datos para generar visualizaciones y de esta manera contar con información más precisa, oportuna y clara, para apoyar la toma de decisiones para mejora la retención de los clientes que poseen planes pospago con la compañía. Es por ello que, surgen métodos y herramientas de visualización para el análisis y la transformación de la data en información que sea útil para las áreas de mercadeo.

Con la utilización de las bases de datos en los sistemas informáticos que apoyan los procesos empresariales han surgido grandes volúmenes de datos como el resultado obtenido en la gestión ordinaria de las empresas; estos, en la mayoría de los casos, no son correlacionados para la toma de decisiones, situación que hace que los datos sean una cantera en bruto para diagnosticar e identificar variables que expliquen situaciones anómalas o típicas empresariales. (Gutiérrez & Molina, 2015)

Gracias a la cantidad de datos almacenados en las bases de datos de la empresas, se buscaron estrategias para el procesamiento de estos, por tal razón, se habla de la minería de datos (Data Mining) y sus técnicas de modelos predictivos aplicado en las empresas de telecomunicaciones, ayudando a predecir los posibles clientes que van a solicitar la cancelación de los servicios contratados, esto gracias a la generación de visualizaciones de la información de los modelos predictivos, ayudando a identificar los servicios que están generando impacto negativo

en la satisfacción de los clientes, por lo que, mejorando estos servicios se puede lograr una reducción en el churn.

A continuación, se aclara que es la minería de datos según (Gutiérrez & Molina, 2015), por lo que se tiene que es una “tecnología y estrategia de modelado matemático que intenta ayudar a comprender el contenido de una base de datos, teniendo en cuenta que los datos por sí sólo no dicen nada, estos se deben procesar para tener un significado y de esta manera se convierten en información”.

Otros autores también tienen un punto de vista acerca de la minería de datos:

“Los datos en bruto raramente son beneficiosos directamente. Su verdadero valor se basa en: (a) la habilidad para extraer información útil la toma de decisiones o la exploración, y (b) la comprensión del fenómeno gobernante en la fuente de datos” (Riquelme, Ruiz, & Gilbert, 2006).

“Minería de Datos o Data Mining es la extracción de información implícita, previamente desconocida y potencialmente útil. La idea es construir programas que examinen cuidadosa y automáticamente bases de datos, buscando regularidades, patrones. Patrones claros, si es que existen, generales y probables para realizar predicciones precisas en datos futuros” (Witten, Eibe, & Mark, 2011).

De acuerdo con la definición de minería de datos de los tres autores, se evidencia que la minería de datos es un procesamiento de datos, el cual se debe realizar a través de software especializado debido a las grandes cantidades de registros almacenadas en las bases de datos. En ese sentido, se debe tener presente que la minería de datos existe hace más de 50

años, pero esta era realizada por personas de forma manual, porque no se tenía una gran cantidad de información, y el análisis de esta, se podía realizar de una manera sencilla por un humano.

La minería de datos hoy en día es ampliamente utilizada por las compañías, para la gestión de la data almacenada en las bases de datos y procesarla para obtener información, y a partir de ella poder obtener el máximo provecho de esta, obteniendo visualizaciones de la información e indicadores para la toma de decisiones por parte de los directivos de la compañía; dentro de la información que se puede analizar se tiene, los productos más vendidos de una empresa, cuáles son los clientes que más llaman por PQR's, cuáles son los servicios que presentan más falla al mes, entre muchas más consultas que se pueden obtener, gracias al procesamiento de los datos.

La minería de datos se ha enfocado en el desarrollo de técnicas, para obtener el mayor provecho de la información, algunas de ellas son árboles de decisión, redes neuronales, algoritmos genéticos, entre otros. Sin embargo, la minería de datos no ha tenido un análisis profundo de las metodologías, las cuales ayudan a entender como ejecutar el proceso para obtener nuevo conocimiento.

“Las metodologías permiten llevar a cabo el proceso de minería de datos en forma sistemática y no trivial. Ayudan a las organizaciones a entender el proceso de descubrimiento de conocimiento y proveen una guía para la planificación y ejecución de los proyectos” (Moine, Haedo, & Gordillo, Estudio comparativo de metodologías para minería de datos, 2011).

Hoy en día existen tres metodologías en la minería de datos que predominan, las cuales son: KDD, CRISP-DM y SEMMA, entre tanto, al final de este capítulo se realizará una descripción de cada una de las metodologías anteriormente nombradas.



## **2.1. Contexto actual**

Para el desarrollo del estado del arte, se realizó una revisión de documentación, ejecutando las siguientes fases: inclusión, exclusión y estrategias de búsqueda, para ayudar a encontrar e identificar los trabajos de investigación relacionados con el trabajo de fin de máster, y también que apalanquen la consecución de los objetivos aquí planteados. Las fuentes de búsquedas utilizadas fueron los siguientes repositorios de información: Biblioteca virtual de UNIR, Repositorio Re-UNIR, Dialnet y Google académico.

### **2.2.1. Criterios de inclusión y exclusión**

Debido a que hoy en día el acceso a la información es mucho más fácil gracias a la internet, pero esto representa algunas dificultades puesto que no se encuentra documentación suficiente relacionada con el trabajo de investigación del presente trabajo de fin de máster, se hace necesario definir unos criterios de inclusión y exclusión de la documentación, por lo que, a continuación de definen estos:

#### **Inclusión:**

- ✓ La documentación se toma a partir del año 2010.
- ✓ La documentación debe hacer referencia a la utilización de la minería de datos.
- ✓ La documentación debe estar relacionado con modelos para el churn en empresas de telecomunicaciones.
- ✓ La documentación debe hacer referencia a técnicas de machine learning

#### **Exclusión:**

- ✓ Documentación que no estén relacionados con la minería de datos.
- ✓ Documentación que no estén relacionados con técnicas de machine learning.
- ✓ Documentación anterior al año 2010.

### **2.2.2. Estrategias de búsqueda**

Los siguientes son las fases ejecutadas para la estrategia de búsqueda de la información:

✓ **Revisión inicial**

En los repositorios de información anteriormente mencionados, se realizó una búsqueda inicial de documentación, para identificar los documentos relacionados con el propósito del trabajo del fin de máster. Dichos documentos son entrada para la siguiente fase.

✓ **Validación cruzada de estudios**

Validación de los documentos basado en los criterios de inclusión y exclusión para saber si los cumplen o no, y de esta manera depurar el listado inicial, siendo estos los insumos para la siguiente fase de la estrategia de búsqueda.

✓ **Integración del grupo de control**

Es el listado de documentos de investigación que cumplen con los criterios de inclusión y exclusión definidos, por tal razón, se realiza un análisis de los títulos de las investigaciones, resumen, introducción, objetivo general, método utilizado, conclusiones y palabras clave. A continuación, se describen los documentos de investigación seleccionados:

### **2.2.3. Revisión bibliográfica**

En la bibliografía podemos encontrar distintos trabajos analizados, como el titulado “Comparativa de técnicas Machine Learning sobre comportamiento de pago de clientes con cuentas por cobrar” (Pesantez Chuqui, 2019), en él se puede encontrar mejoras en la forma de hacer predicciones precisas sobre el comportamiento de pago de clientes con cuentas por cobrar, basada en datos históricos mediante una comparativa de soluciones, a efectos de encontrar el mejor modelo posible de predicciones.

Las técnicas que se utilizan son las siguientes: Árboles de decisión, Redes Neuronales, Clasificador bayesiano, Regresión logística, K vecino más cercano (KNN), Máquina de vectores de soporte (SVM) y la metodología utilizada es KDD. También cabe destacar el uso de las herramientas, Weka y Tableau. De igual forma, en este trabajo se observa como los algoritmos J48 y Random Forest han presentado los mejores resultados en las métricas de precisión, exactitud. Área ROC y F-measure con los valores más altos y con una baja tasa de error tanto en la predicción de falsos positivos y falsos negativos; y la técnica CorrelationAttributeEval ha presentado los mejores resultados tanto en el algoritmo J48 y Random Forest, y también que los árboles de decisión son una excelente herramienta de ayuda para la toma de decisiones para el usuario final, ya que son precisos, fáciles de entender y de aplicar.

De otra parte, se analizó el trabajo titulado “Técnicas de machine learning en el análisis del CHURN RATE” (Gutierrez González, 2019), en el cual se encuentra una explicación sobre en qué consiste el Churn Rate (tasa de abandono de clientes), el problema que plantea y las posibles maneras de detectarlo y reducirlo. Para ello, se logrará analizar en qué consiste y algunas de las técnicas de Data Mining existentes para su estudio, de igual forma, se realizará una revisión bibliográfica de varios estudios relacionados con el tema, lo que, nos permitirá ver la importancia real que este despierta y los métodos más comunes o efectivos para medirlo.

La metodología utilizada es KDD y las técnicas que se utilizan son las siguientes: Regresiones, Árboles de decisión, Redes Neuronales y Máquina de vectores de soporte (SVM). En el trabajo una vez analizadas todas las técnicas, se logra concluir que los árboles de decisión y las regresiones son las dos técnicas que superan al 80% de precisión media; así mismo, se logra concluir que, para la identificación del Churn Rate las técnicas más eficaces son las regresiones y los Árboles, debido a su precisión en cuanto al procesamiento de datos.

El siguiente trabajo analizado es el titulado “Modelo predictivo de Churn de clientes para el negocio de Telecomunicaciones” (Echeverri Giraldo, 2019). En el trabajo se logra desarrollar un modelo de Churn de clientes usando técnicas de Machine Learning que permita, de manera proactiva, predecir el abandono de clientes. Las técnicas que se utilizan son las siguientes: Árboles de decisión, Random Forest, XGBoost, Máquina de vectores de soporte (SVM) y la metodología utilizada es CRISP-DM. También cabe destacar el uso de la herramienta, Python para su desarrollo. En este trabajo se observa como la técnica seleccionada fue XGBoost gracias a su alta adaptabilidad, a su robustez y a su capacidad de trabajar con datos desbalanceados (esta afirmación está sustentada en la evaluación de las métricas de desempeño); también identifica que para que la implementación sea exitosa es preferible definir sprints, ya que estos ayudan a tener claras las etapas del desarrollo, en este sentido el modelo CRISP-DM es altamente recomendable de utilizar, ya que segmenta de manera concisa los pasos necesarios para un proyecto de este tipo.

Otro trabajo analizado es el titulado “Métodos para la mejora de predicciones en clases desbalanceadas en el estudio de bajas de clientes (CHURN)” (Arnejo Calviño, 2017). En el trabajo se identifica el estudio de las técnicas que permitan hacer estas predicciones de la forma más fiable posible atendiendo al problema del desbalanceo y obteniendo un número razonable de bajas predichas. Allí no se aclara la metodología utilizada y la técnica que se esgrime es la de Árboles de decisión. También cabe destacar el uso de las herramientas, KNIME, R y RStudio para su desarrollo. En este trabajo se logra identificar que los modelos híbridos utilizados en el trabajo no mejoraron los resultados por los modelos particulares. Se concluye que el algoritmo Random Forest obtuvo métricas favorables, pero es muy complejo de interpretar, por tal razón, no es ideal para la empresa, debido a que no se identifican con claridad las reglas por las cuales se asignan las probabilidades de baja de cada cliente.

El siguiente trabajo analizado es el titulado “Diseño de un modelo predictivo de fuga de clientes utilizando árboles de decisión” (Contreras Morales, Ferreira Correa, & Valle, 2017). En el que se logró establecer un modelo de clasificación que permita predecir las condiciones que se cumplen para el abandono voluntario de clientes, basado en antecedentes de comportamiento histórico. La técnica de minería de datos utilizada fue: Árboles de decisión, y la metodología utilizada es CRISP-DM. También cabe destacar el uso del software R para su desarrollo. En el trabajo las variables predictoras del modelo de clasificación permitieron establecer reglas de decisión desde las cuales es posible encontrar qué condiciones se debieran cumplir para que un cliente decida voluntariamente abandonar la empresa. También el modelo de árbol de decisión permite explicar en forma satisfactoria las condiciones que se deben cumplir para que ocurra un abandono voluntario de clientes. Por tal razón, los resultados del trabajo están cumpliendo el objetivo planteado.

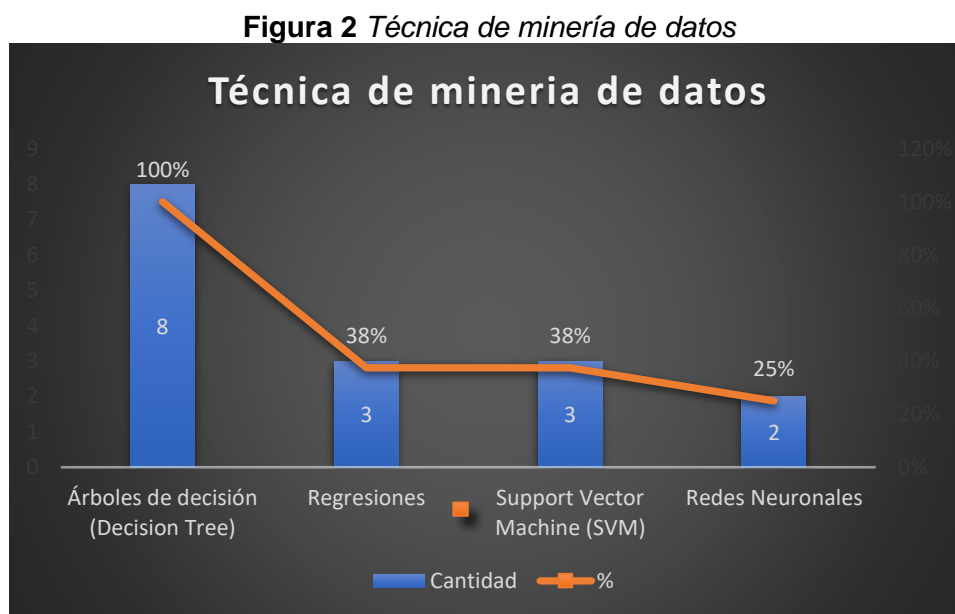
El trabajo titulado “Modelos predictivos del churn – abandono de clientes – para operadores de telecomunicaciones” (Lozano Núñez, 2015) también fue analizado. Allí se creó un modelo para la identificación temprana de bajas voluntarias de los clientes, basado en los datos demográficos, patrones de consumo, facturación, historial de incidentes, patrones de navegación, entre otros, de los que dispone el operador. Las técnicas utilizadas fueron: Árboles de decisión y Tree ensemble. La metodología utilizada es CRISP-DM. Las herramientas utilizadas fueron: RapidMiner, R, WEKA y KNIME. En el trabajo el modelo implementado cumplió con el objetivo planteando. El índice de acierto del modelo ha aumentado con el tiempo, debido a que se han incluido nuevas variables para el análisis del modelo. Con la ejecución del modelo, se logró un descenso en las bajas de clientes, pasando de un 45% a un 25% en telefonía fija y en telefonía móvil de un 80% a un 20%.

Otro trabajo analizado, es el titulado “Modelo de predicción de fuga de clientes de telefonía móvil post pago” (Pérez Villanueva, 2014). Construir un modelo de predicción de comportamiento de término de contrato que considere el efecto de la actividad de la red de contactos. Se logró gracias a la determinación de variables que afectan la fuga de clientes, las cuales fueron entradas para los modelos predictivos de regresión lineal y los árboles de decisión, donde fueron las técnicas utilizadas. No se aclara en el trabajo el método ni las herramientas de software utilizados. En el trabajo se realizó análisis de variables a través de dos modelos, donde el árbol de decisión posee un 18% menos de precisión que el de regresión lineal, pero este último se desempeña, pero con respecto a la variable de exactitud del modelo. Los datos de los fugados se ajustan mejor en el Árbol de decisión, por lo que éste puede aprender de mejor manera el fenómeno de fuga que la Regresión lineal. La variable que más peso en los dos modelos analizados, que tiene en la explicación de fuga es la variación en la cantidad de reclamos que realiza un cliente.

El último trabajo analizado, es el titulado “Aplicación de Minería de Datos para Predecir Fuga de Clientes en la Industria de las Telecomunicaciones” (Barrientos & Ríos, Aplicación de Minería de Datos para Predecir Fuga de Clientes en la Industria de las Telecomunicaciones, 2013) . En el trabajo se mostró una metodología para poder realizar predicción de fuga de clientes o Churn en un ambiente multiplataforma en la industria de las telecomunicaciones. La técnica utilizada fue: Árboles de decisión, y la metodología Knowledge Discovery in Databases (KDD). No se indica herramientas usadas. En el trabajo la integración fue muy importante para cada una de las etapas del modelo KDD. Se concluye que el preprocesamiento como la transformación se deben realizar de manera rigurosa, porque se puede tener pérdida de valores para muchas de las variables, y un valor perdido no se puede ignorar.

**2.1.1.1. Resumen de la bibliografía.** A continuación, se analizan las investigaciones seleccionadas que tienen relación con el churn en las empresas de telecomunicaciones y que como consecuencia generan aporte en el presente trabajo del fin de máster.

- ✓ La técnica de minería de datos más utilizada en los documentos es la de árboles de decisión, con un 100% de utilización de dicha técnica en cada una de las investigaciones analizadas, la segunda técnica más utilizada es la de regresiones y Support Vector Machine (SVM), donde un 38% de las investigaciones utilizaron dichas técnicas, y por último las redes neuronales, que fue utilizada en 25% de las investigaciones. En la figura 2 se observa más claramente la utilización de las técnicas.

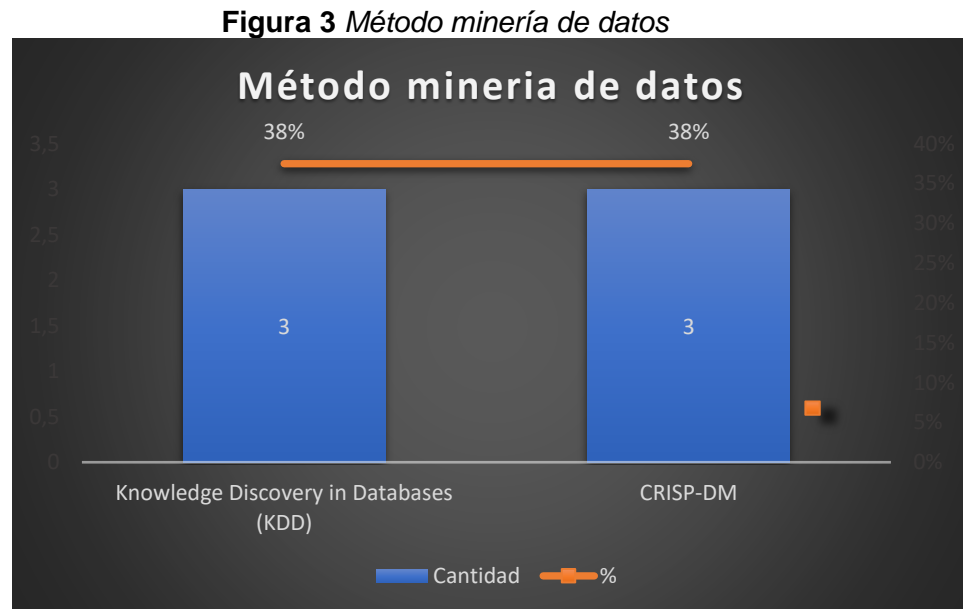


Fuente – Elaboración propia

El objetivo de las distintas técnicas de la minería de datos es comparar el desempeño de cada una de ellas con respecto al objetivo del problema planteado con base en las variables seleccionadas, para identificar cuál de ellas se comporta mejor con respecto al análisis de la información introducida en la técnica seleccionada.

Debido a que, con cada una de las técnicas mencionadas los resultados serán distintos, se considera que el desempeño de las técnicas de la minería de datos puede ser apropiado para la solución del problema, dependiendo del problema planteado y sus objetivos. Ejemplo: para algún caso lo más importante puede ser la variable de clasificación, mientras que en otro problema no.

- ✓ Los métodos de minería utilizados fueron el Knowledge Discovery in Databases (KDD) y el CRISP-DM, ambos con un 38% de utilización en las investigaciones analizadas. En la figura 3 se observa el detalle de los métodos de minería de datos.

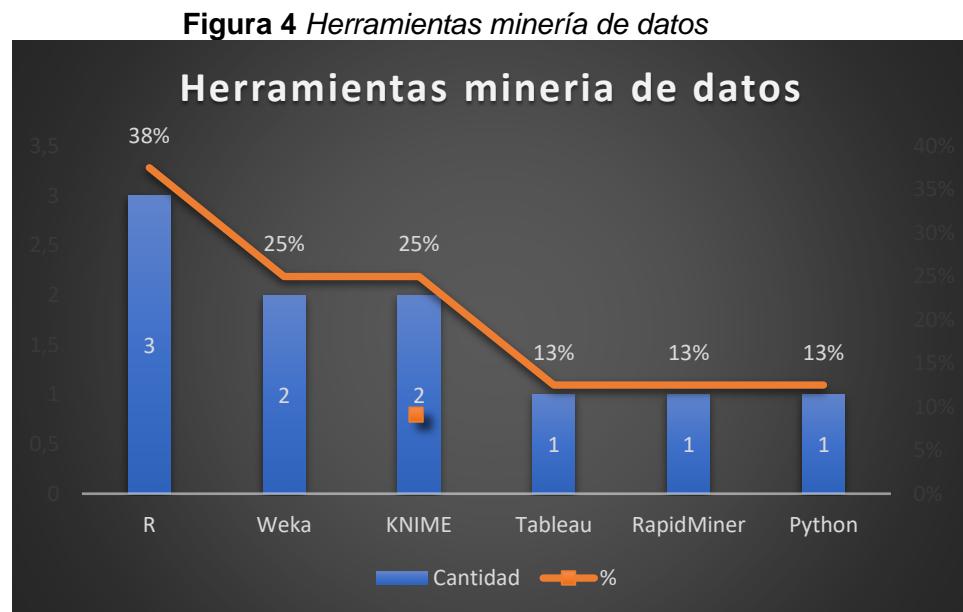


Fuente – Elaboración propia

Con el propósito de obtener más información de los métodos de minería de datos, estos fueron descritos en los apartados 2.1 y 2.2 respectivamente, del presente documento.



- ✓ Por último, el análisis de las investigaciones fue la variable de herramientas utilizadas. Con un 38% el software más empleado fue R, seguido con un 25% de las aplicaciones de Weka y KNIME. En la figura 4 podemos observar con detalle los softwares utilizados en las investigaciones.



Fuente – Elaboración propia

Con respecto al software, se puede concluir, que en la actualidad se tienen muchas opciones de aplicaciones, las cuales son apoyo fundamental para el procesamiento de grandes cantidades de información en tiempos cortos, ayudando a la toma de decisiones para la predicción de los posibles clientes que van a abandonar la compañía por algún motivo de inconformismo con el servicio prestado.

#### **2.2.4. Analítica visual**

Dentro de la minería de datos se tiene la necesidad del realizar un análisis y comprensión de datos, el cual facilite soluciones interactivas de visualizaciones de datos, por tal razón, surge la analítica visual como ciencia.

“La analítica visual es un proceso iterativo que implica la recopilación de información, datos preprocesamiento, representación del conocimiento, interacción y toma de decisiones”. (Keim, Mansmann, Schneidewind, Thomas, & Ziegler, 2008).

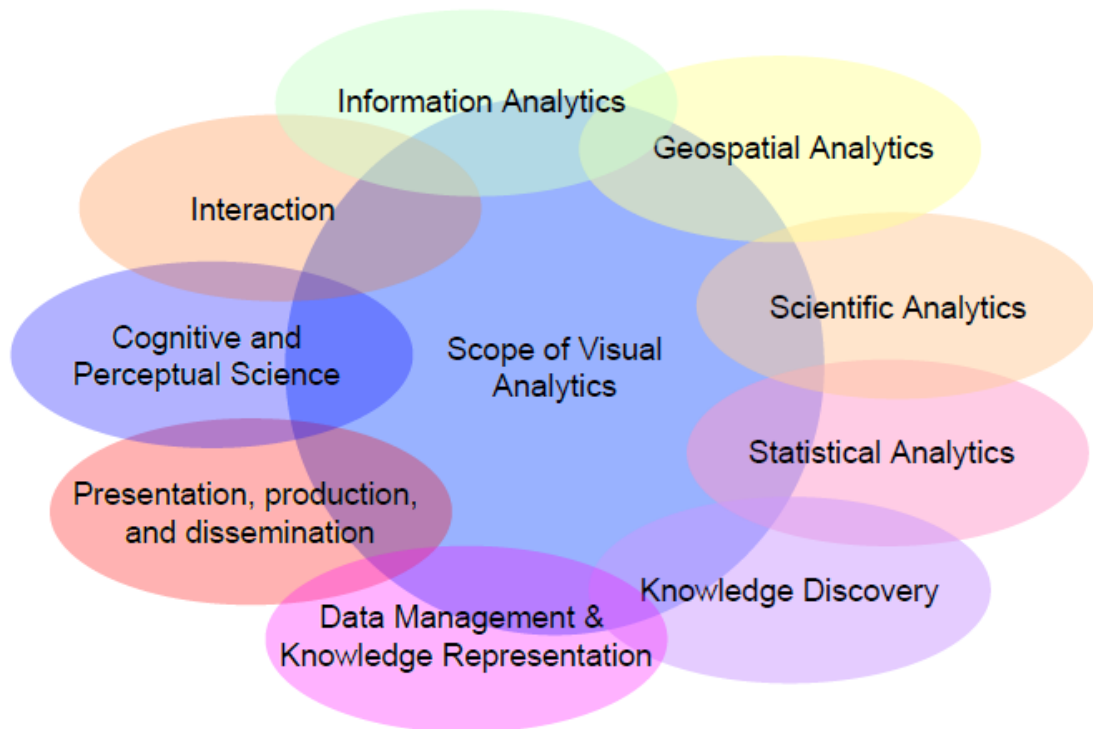
El objetivo de la analítica visual es la conexión entre la gestión de datos, el diseño y elaboración de visualizaciones de datos complejos, esto se logra gracias a que, el proceso de visualización representa los valores numéricos o de otro tipo de una forma más entendible, para que los usuarios puedan comprenderlos y poder coadyuvar a la toma de decisiones.

Con la ayuda de la analítica visual, se logra identificar inconsistencia, tendencias, patrones, entre otros análisis que nos muestra la información procesada. Estos resultados pueden ser analizados por los científicos de datos y por todos los expertos de las áreas de la compañía, logrando una transferencia y un intercambio de información entre las distintas divisiones de la empresa, y buscando ayudar a una mejor toma de decisiones empresariales, debido a que estas son más integrales.

En la actualidad se cuenta con una gran cantidad de datos en los distintos sistemas de las empresas, y por eso se hace tan importante la analítica visual para su procesamiento, análisis y posterior visualización de la información, esto se logra gracias a los sistemas de procesamiento de datos, pero también se debe tener presente que, en las primeras etapas de la analítica visual, se debe involucrar el talento humano para dicho análisis de los datos

debido a la complejidad de los mismos, por tal razón, se tiene una visión integral que combina el recurso humano, análisis y visualizaciones, como se observa en la figura 5, el alcance de la analítica visual.

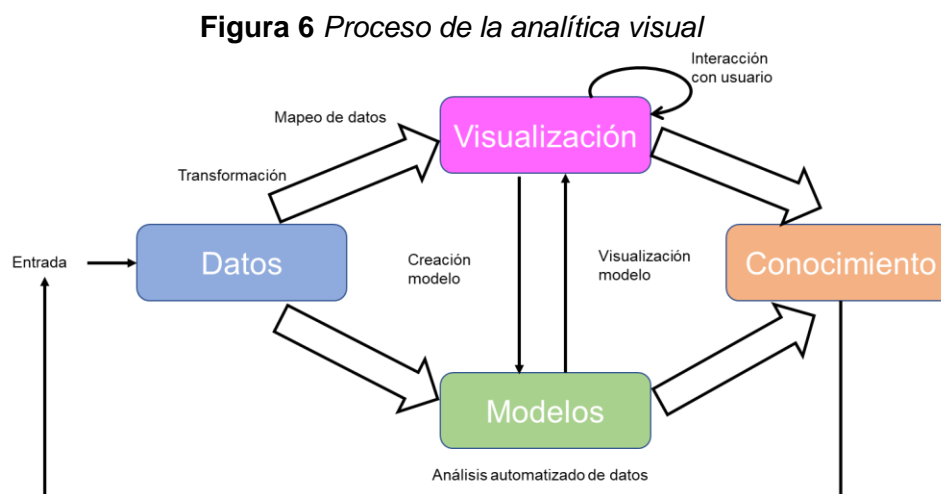
**Figura 5** Alcance de la analítica visual



Fuente: [http://kops.uni-konstanz.de/bitstream/handle/123456789/5631/Visual\\_Analytics\\_Scope\\_and\\_Challenges.pdf](http://kops.uni-konstanz.de/bitstream/handle/123456789/5631/Visual_Analytics_Scope_and_Challenges.pdf)

Se debe tener presente que, en la analítica visual la obtención de los datos no se puede lograr de forma automática para su análisis, por tal razón, se tiene la necesidad de realizar la extracción de los datos con métodos automáticos a partir de estadísticas y el método de la minería de datos de descubrimiento de conocimiento en bases de datos (KDD). También se debe tener claro, que el conocimiento del recurso humano tampoco se puede llegar a automatizar.

**2.2.4.1. Proceso de la analítica visual.** El proceso tiene como objetivo, obtener conocimiento de los datos a través de análisis automáticos y visualizaciones, logrando las interacciones con usuarios finales. En la figura 6 del proceso análisis visual se presentan las diferentes etapas del proceso y las relaciones.



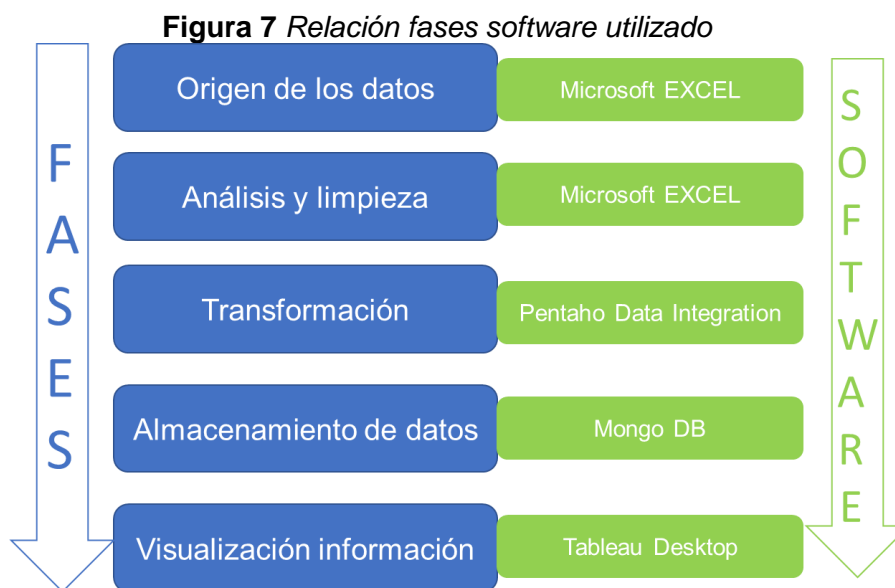
Fuente: Elaboración propia basado en [http://kops.uni-konstanz.de/bitstream/handle/123456789/5631/Visual\\_Analytics\\_Scope\\_and\\_Challenges.pdf](http://kops.uni-konstanz.de/bitstream/handle/123456789/5631/Visual_Analytics_Scope_and_Challenges.pdf)

En el contexto de Big data y por la diversidad que pueden tener los datos, antes de realizar algún tipo de visualización o aplicar algún método de análisis automático, se hace necesario la integración de los datos, por ello el primer paso en el proceso de análisis visual es preprocesamiento y transformación de los datos, que también incluye limpieza, normalización, agrupación o integración de fuentes de datos heterogénea (Thomas & Cook, 2005).

## 2.2.5. Tecnologías

Con el fin de lograr el objetivo planteado en el presente documento, se utilizó software tecnológico con bastante usabilidad en la industria de la minería de datos, debido a los buenos resultados obtenidos en la generación de metodologías para la retención de clientes.

De acuerdo con cada una de las fases de la metodología del trabajo, se detalla el software utilizado para cada una de las fases, con el propósito de cumplir con el objetivo de la metodología. En la figura 7 se observa la relación de las fases con el software utilizado.



Fuente: Elaboración propia

**2.2.5.1. Microsoft Excel.** Los archivos que se recibirán con la información relacionada con las PQR's, datos de los usuarios y la(s) líneas de telefonía móvil de los usuarios, están en formato de Excel. A partir de estos archivos se inicia con un análisis y limpieza de los datos, adicionalmente se puede identificar errores en los datos y validar la integridad de estos, debido a que un mismo campo con información tiene distinto nombre o tipo.

**2.2.5.2. Kettle Pentaho Data Integration.** "Pentaho vio la luz en el año 2004, y se trata de una plataforma de gratuita, es decir open source. Esta plataforma está compuesta por una serie de componentes de Big Data, como pueden ser informes, cuadros de mandos, dashboards o integración de datos, entre muchos otros." (Retos en Supply Chain, 2020)

Con Pentaho se pueden crear ETL, en otras palabras, se pueden realizar procedimientos de extracción, transformación y cargue de datos. Adicionalmente es una herramienta de que puede ser utilizada como BI, ya que ayuda a mejorar la capacidad de análisis y toma de decisiones.

Los archivos de Excel analizados se pueden cargar a Pentaho para realizar el proceso de ETL de la información.

**2.2.5.3.      Mongo DB.** Mongo es una BD gratuita, NoSQL y orientada documentos, por tal razón tiene escalabilidad y flexibilidad y modelo de consultas de indexación.

MongoDB es una base de datos distribuida en su núcleo, por lo que la alta disponibilidad, la escalabilidad horizontal y la distribución geográfica están integradas y son fáciles de usar. (MongoDB, Inc, 2021)

Es la BD definida para almacenar los datos que se le realizó el proceso de ETL con Pentaho, buscando obtener una escalabilidad de la BD a medida que la cantidad de datos aumentó, gracias a los beneficios anteriormente mencionados.

**2.2.5.4.      Tableau Desktop.** Tableau es una plataforma de análisis visual que transforma la manera en que usamos los datos para resolver problemas. Además, permite a las personas y las organizaciones sacar el máximo partido de los datos. (Tableau Software, LLC, 2003)

Es una de las herramientas líderes en el mercado, de las visualizaciones de datos interactivos, utilizada en el área de BI, ayudando a las empresas y personas a ver y comprender la información.

Adicionalmente Tableau ofrece una gran cantidad de conexiones de datos, donde se puede conectar a distintas fuentes de datos, sin necesidad de programar la conexión a dichas fuentes.

Debido a todos los beneficios que ofrece el software y el fácil uso para la generación de los informes, es la herramienta seleccionada para la generación de las visualizaciones de la información procesada.

## **2.2. Metodologías**

Las siguiente son las metodologías a describir:

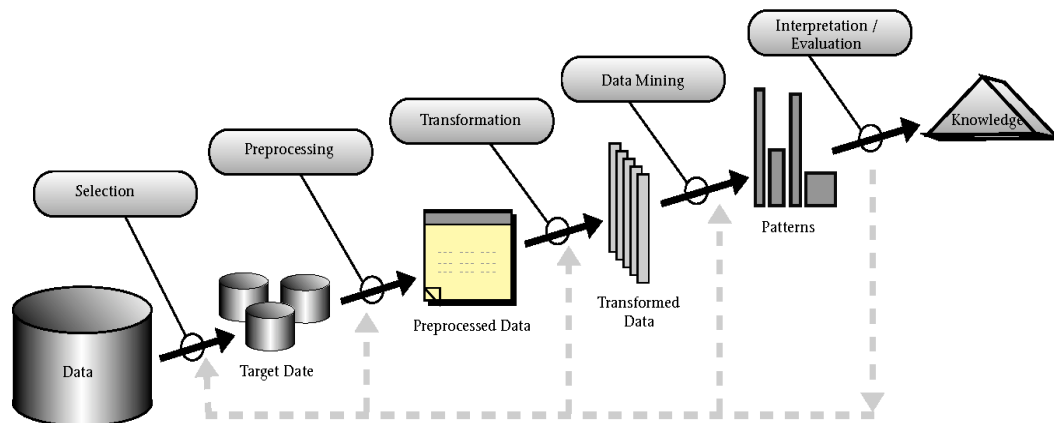
### **2.2.1. KDD**

El término Descubrimiento de Conocimiento en Bases de Datos (Knowledge Discovery in Databases, o KDD para abreviar) empezó a utilizarse en 1989 para referirse al amplio proceso de búsqueda de conocimiento en bases de datos, y para enfatizar la aplicación a "alto nivel" de métodos específicos de minería de datos. En general, el descubrimiento es un tipo de inducción de conocimiento, no supervisado" (Beltrán Martínez, s.f.).

Dentro del descubrimiento de información, se tiene como alcance obtener datos.

"El campo del Descubrimiento de Conocimiento en Bases de Datos, denominado Knowledge Discovery in Data Bases en inglés y usualmente abreviado KDD, es la convergencia del Aprendizaje Automático, la Estadística, el Reconocimiento de Patrones, la Inteligencia Artificial, las Bases de Datos, la Visualización de Datos, los Sistemas para el Apoyo a la Toma de Decisiones, la Recuperación de Información, y otros muchos campos" (Beltrán Martínez, s.f.).

La metodología KDD es una proceso iterativo e interactivo, y se definen los siguientes pasos para su ejecución. En la figura 8 se observan los pasos.

**Figura 8** Pasos ejecución metodología KDD

Fuente: <https://www.semanticscholar.org/paper/What-Data-Is-Necessary-to-Data-Mine-for-Knowledge-AI-par/fca8b47e29e75d0676af8ec1b36a1dd29c9a1e6e?p2df>

### 2.2.2. CRISP-DM

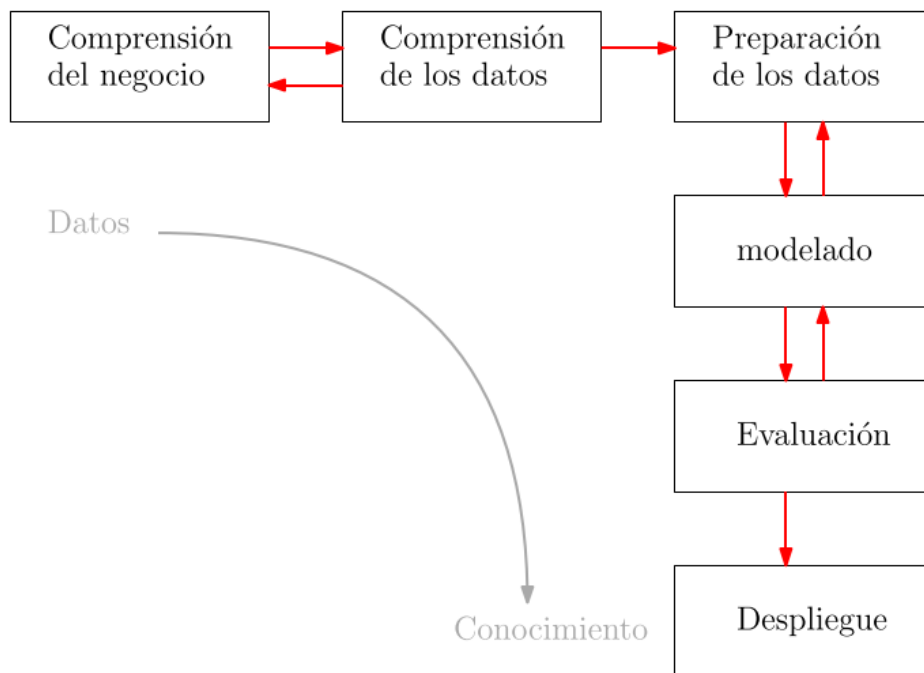
Por sus siglas en inglés Cross-Industry Standard Process for Data Mining, esta metodología fue creada por dos empresas SPSS y Daimler Chrysler en el año 2000, en la actualidad es la metodología más utilizada en proyectos relacionados con minería de datos.

“Su éxito se debe a que está basada en la práctica y experiencia real de analistas de minería de datos que han contribuido activamente al desarrollo de esta” (Gironés Roig, Casas Roma, Minguillón Alfonso, & Caihuelas Quiles, 2017).

Debemos tener claro que la metodología es una guía de trabajo, la cual ayuda a la entrega del trabajo final, pero el éxito del trabajo depende de la ejecución de cada una de las fases, sin tener prioridad en una de ellas, por eso se entiende que todas las fases son importantes y dependientes para lograr el objetivo del trabajo.

Las siguientes son las fases de la metodología CRISP-DM las cuales se observan en la figura 9



**Figura 9** Fases metodología CRISP-DM

Fuente: (Gironés Roig, Casas Roma, Minguillón Alfonso, & Caihuelas Quiles, 2017)

**2.2.2.1. Comprensión del negocio.** En esta fase se realiza la definición del problema de minería de datos, también se tiene la comprensión y definición de los objetivos y los requerimientos del proyecto. Para lograr cumplir con los objetivos, se hace necesario definir un plan de proyecto.

**2.2.2.2. Comprensión de los datos.** En esta fase debemos familiarizarnos con los datos, ya que, resulta ser una parte muy importante, debido a que se busca la calidad de los datos por ser parte fundamental de la minería de datos.

“Tener una buena calidad de los datos será siempre una condición necesaria, aunque no suficiente para tener éxito en el proyecto” (Gironés Roig, Casas Roma, Minguillón Alfonso, & Caihuelas Quiles, 2017).

En consonancia con lo anterior, debemos:

- ✓ Obtener el conjunto inicial de datos
- ✓ Explorar el conjunto de datos
- ✓ Identificar las características de calidad de los datos
- ✓ Identificar los resultados iniciales

**2.2.2.3. Preparación de los datos.** En esta fase se inicia con la selección de datos y posterior limpieza de los mismos, a fin de lograr el objetivo de la fase, el cual es obtener un conjunto de datos final para poder aplicar los modelos.

**2.2.2.4. Modelado.** Esta fase tiene relación directa con la fase anterior de la preparación de los datos y con la fase posterior de evaluación del modelo, por tal razón la hace importante.

“El objetivo último de esta fase será el de disponer de un modelo que nos ayude a alcanzar los objetivos de la minería de datos y los objetivos de negocio establecidos en el proyecto. Podemos entender el modelo como la habilidad de aplicar una técnica a un juego de datos con el fin de predecir una variable objetivo o encontrar un patrón desconocido” (Gironés Roig, Casas Roma, Minguillón Alfonso, & Caihuelas Quiles, 2017).

En esta fase se implementa herramientas de minería de datos, sin embargo, se debe realizar un análisis de las herramientas existentes para identificar cual es la más adecuada para el proyecto.

**2.2.2.5. Evaluación.** En esta fase se debe determinar si los resultados obtenidos coinciden con los objetivos planteados en la fase de comprensión del negocio. También se deben lograr identificar los temas del negocio que deberían haberse abordado, y en consecuencia con esta evaluación se logra definir los pasos a seguir, estos pueden ser, retornar a ejecutar una de las fases anteriores o redefinir cada una de ellas.

**2.2.2.6. Despliegue.** Como última fase, se realiza el despliegue de los modelos resultantes en la práctica, y se deben ejecutar mantenimientos de forma repetida y continua sobre estos modelos.

También se deben definir como los resultados obtenidos serán distribuidos con los usuarios objetivo del proyecto.

En esta fase deberemos estar preparados para:

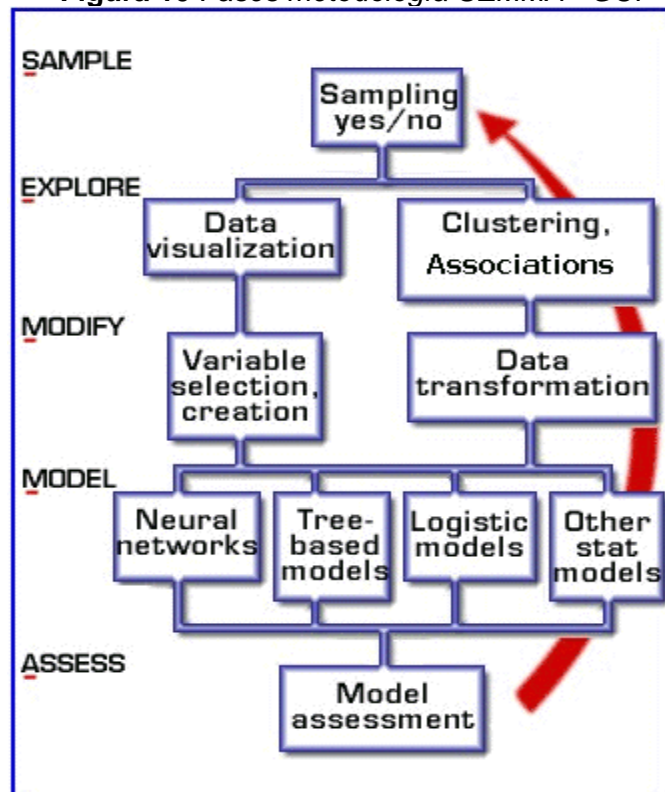
- ✓ “Diseñar un plan de despliegue de modelos y conocimiento sobre nuestra organización.
- ✓ Realizar seguimiento y mantenimiento de la parte más operativa del despliegue.
- ✓ Revisar el proyecto en su globalidad con el objetivo de identificar lecciones aprendidas” (Gironés Roig, Casas Roma, Minguillón Alfonso, & Caihuelas Quiles, 2017)

### **2.2.3. SEMMA**

Es el acrónimo de las cinco fases: Sample, Explore, Modify, Model, Assess.

“Metodología creada por el SAS Intitute, quien define la minería de datos como el proceso de muestreo, exploración, modificación, modelado y evaluación (SEMMA) de grandes cantidades de datos para descubrir patrones previamente desconocidos que pueden utilizarse como una ventaja comercial” (SAS Institute, 2017).

En figura 10, se logra ver con más claridad el funcionamiento de la metodología con el apoyo de la aplicación SAS.

**Figura 10** Fases metodología SEMMA - GUI

Fuente: <https://documentation.sas.com/doc/en/emref/14.3/n061bzurmej4j3n1jn8bbjm1a2.htm>

### **3. Objetivos**

A continuación, se presenta el objetivo general y los objetivos específicos del presente trabajo de fin de máster:

#### **3.1. Objetivo general**

Desarrollar un método para gestionar los datos relacionados con las PQR's generadas por los clientes, para identificar los posibles usuarios a finalizar el contrato con la empresa de telefonía móvil.

#### **3.2. Objetivos específicos**

Para lograr el objetivo general, se debe cumplir con cada uno de los siguientes objetivos específicos:

- ✓ Analizar la información proporcionada por las PQR's, para identificar variables relevantes y sus relaciones.
- ✓ Diseñar un modelo de datos que permita unificar todos los datos en una base de datos centralizada.
- ✓ Transformar y normalizar la data con el fin de permitir su cargue en una estructura estandarizada.
- ✓ Elaborar tableros de visualización de la información, los cuales puedan ser usados por las áreas encargadas para la retención de clientes.

## 4. Metodología

Una empresa está constituida por diferentes actores, como lo son, accionistas, los proveedores, los empleados, los clientes, entre otros. Sin embargo, uno de esos actores es el más difícil de sustituir y son los clientes, dado que mientras una empresa puede cambiar de socios que aporten capital, proveedores de insumos y de servicios, empleados y de bancos, cambiar de clientes resulta más difícil, debido a que se requiere invertir tiempo y esfuerzo para atraerlos, vincularlos y, especialmente, mantenerlos. Es precisamente debido a esa complejidad de acciones entre las áreas de marketing, ventas, servicio al cliente y operación que intervienen de manera coordinada, que se torna difícil sostener el volumen de clientes de las empresas, y que por ello éstas se ven en la necesidad de hacer todos los esfuerzos posibles para conservarlos, es decir, para evitar que se desvinculen de la empresa después de que han recibido sus productos y/o servicios. Al fin y al cabo, de todos los actores que intervienen en el ciclo del negocio de una empresa, los clientes son los únicos que aportan el dinero que las empresas necesitan para retribuirles a todos los demás, incluidos los accionistas que aportaron capital en el momento de fundar la empresa (Pérez, 2018).

La situación descrita también se presenta en el negocio de las telecomunicaciones, en este mercado y particularmente en el segmento de los contratos de pospago, los clientes suscriben un contrato de adhesión que tiene una tarifa preestablecida mensual por la cual reciben a cambio un paquete que puede incluir llamadas telefónicas a diferentes operadores, acceso móvil a internet y a determinadas redes sociales, entre otros servicios. Aunque esos elementos suelen ser los más promocionados y los que usualmente se emplean para diferenciar los diferentes planes y sus correspondientes tarifas, los clientes también reciben otros servicios que necesariamente hacen parte de todos los paquetes, como son los servicios de atención telefónica o digital para consulta o creación de peticiones, quejas y reclamos o recursos (conocidos como PQR), facturación y recaudo.

Esos factores mencionados que hacen parte del servicio de telefonía móvil inciden en que el cliente decida mantenerse vinculado contractualmente a un determinado operador, o que en determinado momento prefiera cambiarse a otro operador que le ofrezca mejores expectativas, ya sea respecto de las especificaciones del paquete (megabytes de internet al mes, minutos de comunicación telefónica y acceso a redes sociales frente a la tarifa) o respecto a los valores agregados (PQR, facturación, recaudo, cobertura, entre otros). Se habla entonces de la tasa de cancelación o abandono, conocida como Churn, que mide el porcentaje periódico de clientes que deciden cancelar los servicios contratados con una empresa.

También se debe tener presente que retener a un cliente resulta aproximadamente diez veces más económico que conseguir un nuevo, por ello, debe ser una prioridad saber aplicar estrategias de retención y fidelización que consigan mantener y desarrollar a los clientes rentables y fieles. (Domínguez & Hermo, 2007), por tal razón, los esfuerzos y estrategias de la compañía se deben centrar en mejorar la retención de clientes, el cual se puede ver reflejado en la mejora de ingresos económicos para la empresa.

Cuando se tiene una buena tasa de retención de cliente se genera una fidelización de los clientes con la compañía, como lo menciona Domínguez y Hermo, la fidelidad no implica necesariamente un compromiso, puede ser simplemente un hábito. Este índice formado por las métricas de satisfacción y retención ayudan a conocer qué porcentaje de clientes lo es por hábito y qué porcentaje lo es por satisfacción o por preferencia. Los estudios de mercado pueden colaborar, intentando diferenciar estas cuestiones (Domínguez & Hermo, 2007)

También se debe tener presente que la satisfacción del cliente es impulsada por la retención de los mismos. La satisfacción del cliente por sí misma no garantiza fidelidad. Se tiene un alto grado de fidelidad si la experiencia de compra tiene un grado muy alto de satisfacción (Domínguez & Hermo, 2007)

Por las razones anteriormente expuestas, el objetivo principal del documento es la creación de un método que ayude a mejorar la retención de clientes con planes pospago, ya que estos son los que generan la mayor cantidad de ingresos económicos a la empresa y de esta manera se está ayudando a que la compañía pueda realizar inversiones económicas en otros ámbitos, como en la infraestructura para mejorar el servicio a los clientes, y de esta forma obtener una mayor satisfacción de los clientes.

Para lograr el objetivo planteado en el TFM, se debe tener presente que, en la actualidad las aplicaciones generan una gran cantidad de información a cada segundo y se tiene la capacidad de ser almacenada, gracias a la evolución rápida que se ha tenido en el almacenamiento de la información en las bases de datos, ya sean relacionales o no relacionales. Sin embargo, en la capacidad de analizar la gran cantidad de datos producidos por las aplicaciones no se ha logrado tener una evolución tan acelerada, por tal razón, se tiene un desafío grande con respecto a los procesos de análisis de información, surgiendo la analítica visual de información, la cual se explicará a continuación.



## **5. Desarrollo del método**

En el desarrollo de este capítulo se encuentra el método propuesto para la retención de clientes de telefonía móvil pospago. A continuación, se explica con detalle el paso a paso de las actividades realizadas y cómo el resultado de cada una de estas encaja con el inicio de la siguiente fase del método, y de este modo se muestra el flujo de las fases del método propuesto.

### **5.1. Evaluación Inicial**

Con el objetivo de identificar las variables importantes para la retención de clientes de telefonía móvil pospago, el primer paso fue realizar acercamiento con el área encargada del análisis de las PQR's generados por los clientes e identificar los reportes que se generan de dichas solicitudes.

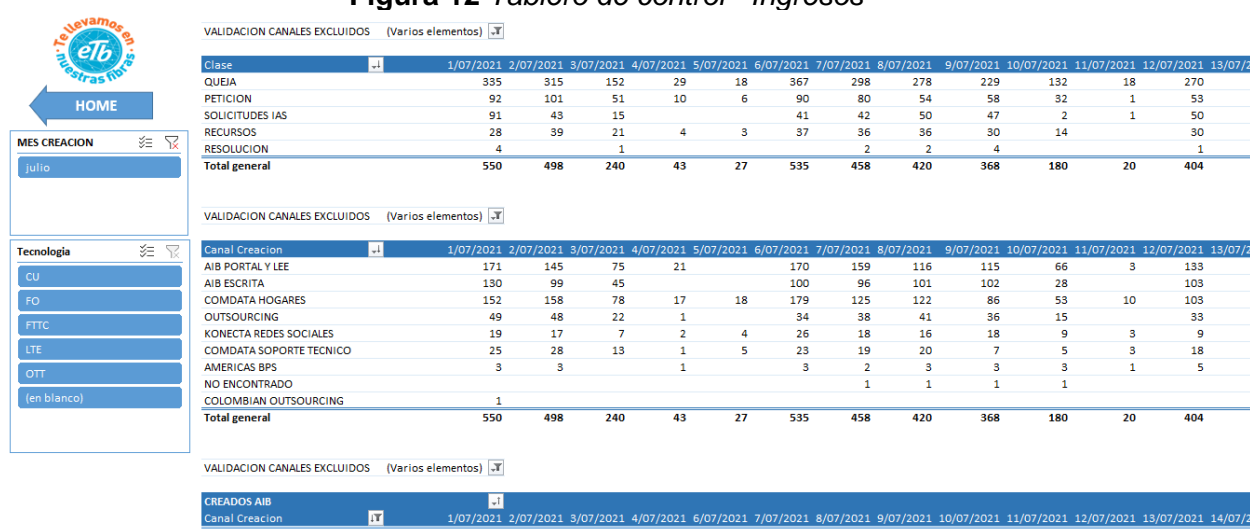
En la dirección de experiencia de primera instancia, se realiza un análisis cada 15 días de las PQR'S y se genera un tablero de control, obteniendo la información de los ingresos, efectividad, envíos, cierres, sin correspondencia y productividad de las PQR's, cómo se muestra en la figura 11. Así mismo, más adelante se explica brevemente cada una de las opciones del tablero de control.

**Figura 11** *Tablero de control*

Fuente: Tablero de control PQR's ETB

En la opción de “ingresos” se consigue visualizar la información de todas las PQR's con respecto al tipo de solicitud y los canales de creación, las cuales se tienen en dos tablas, así las cosas, en la primera tabla, se tiene la información detallada de la cantidad de PQR's generadas por cada día y la clase de solicitud, la cuales pueden ser de queja, petición, solicitudes IAS, recursos y resolución, esta información se observa de acuerdo con el mes filtrado. En la segunda tabla, se tiene la información detallada de la cantidad de PQR's generadas por cada día y el canal de creación de estas, estos pueden ser de cada uno de los Call Center contratados por ETB para la recepción de las PQR's: AIB PORTAL Y LEE, AIB ESCRITA, COMDATA HOGARES, OUTSOURCING, KONECTA REDES SOCIALES, COMDATA SOPORTE TECNICO, AMERICAS BPS Y NO ENCONTRADO. En la figura 12 se observa el tablero de control de ingresos.

Figura 12 Tablero de control - Ingresos




Fuente: Tablero de control PQR's ETB

En la opción de “efectividad” se puede visualizar la información de todas las PQR's que han sido solucionadas, esta solución se mide en rango de días con respecto al área que soluciona el requerimiento y el tipo de solicitud solucionando en los rangos de tiempos definidos para la medición, los cuales son:

- Entre 0 y 3 días
- Entre 4 y 5 días
- Entre 6 y 9 días
- Entre 10 y 12 días
- Entre 13 y 15 días
- Mayor a 15 días

En la figura 13 se observa el tablero de control de efectividad.

Figura 13 Tablero de control - Efectividad



HOME

Mes Solucion: julio

TIPO CLIENTE: PLATINO, ESPECIAL

Clase Pqr: PETICION, QUEJA, RECURSOS, RESOLUCION, SOLICITUDES IAS, Tecnologia: CU

Area Usuario/Rango	Entre 0 y 3 dias	Entre 4 y 5 dias	Entre 6 y 9 dias	Entre 10 y 12 dias	Entre 13 y 15 dias	Mayor a 15 dias	Total
PRIMERA INSTANCIA	763	949	2795	1578	745	2	6832
PORTAL ZEC	2081	279	95	8			2463
LINEA ESPECIAL	343	323	378	118	35		1197
SOPORTE DEFENSA	51	49	166	264	344	1	875
SEGUNDA INSTANCIA	56	108	244	210	224	8	850
LINEA ESPECIAL EMPLEADOS	291	45	13	3	4		356
CASOS TIENDAS	6						6
CORRESPONDENCIA	3						3
<b>Total</b>	<b>3594</b>	<b>1753</b>	<b>3691</b>	<b>2181</b>	<b>1352</b>	<b>11</b>	<b>12582</b>

Area Usuario/Rango	Entre 0 y 3 dias	Entre 4 y 5 dias	Entre 6 y 9 dias	Entre 10 y 12 dias	Entre 13 y 15 dias	Mayor a 15 dias	Total
PRIMERA INSTANCIA	11,2%						
PORTAL ZEC	84,5%						
LINEA ESPECIAL	28,7%						
SOPORTE DEFENSA	5,8%						
SEGUNDA INSTANCIA	6,6%						
LINEA ESPECIAL EMPLEADOS	81,7%						
CASOS TIENDAS	100,0%						
CORRESPONDENCIA	100,0%						
<b>Total</b>	<b>28,6%</b>						


Area Usuario/Rango	Entre 0 y 3 dias	Entre 4 y 5 dias	Entre 6 y 9 dias	Entre 10 y 12 dias	Entre 13 y 15 dias	Mayor a 15 dias	Total
PETICION	1872	431	459	192	102	1	3057
QUEJA	1619	1173	2825	1518	685	2	7822
RECURSOS	51	101	220	188	205	2	767
RESOLUCION	2	4	16	6	6	5	39
SOLICITUDES IAS	50	44	171	277	354	1	897
<b>Total general</b>	<b>3594</b>	<b>1753</b>	<b>3691</b>	<b>2181</b>	<b>1352</b>	<b>11</b>	<b>12582</b>

Area Usuario/Rango	Entre 0 y 3 dias	Entre 4 y 5 dias	Entre 6 y 9 dias	Entre 10 y 12 dias	Entre 13 y 15 dias	Mayor a 15 dias	Total
PETICION	61,2%						
QUEJA	20,7%						
RECURSOS	6,6%						
RESOLUCION	5,1%						
SOLICITUDES IAS	5,6%						
<b>Total general</b>	<b>28,6%</b>						

Fuente: Tablero de control PQR's ETB

En la opción de “envíos” se logra visualizar la información de todas las cartas enviadas a los usuarios que han generado PQR's, y se logra evidenciar si estas fueron enviadas de manera física o virtual. En la figura 14 se observa el tablero de control de envíos

Figura 14 Tablero de control - Envíos



HOME

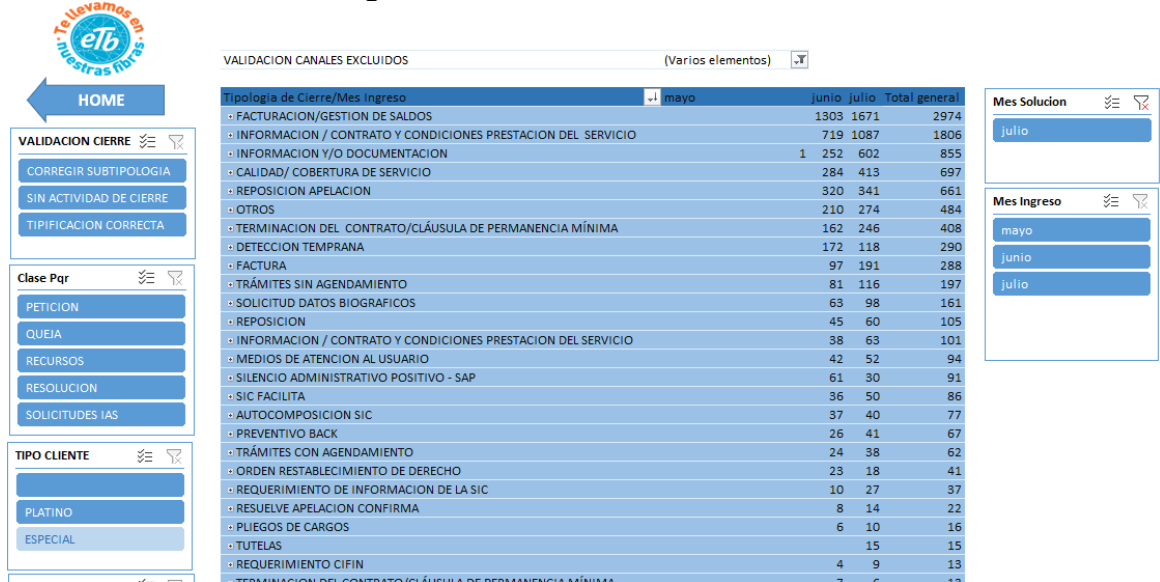
Nro día	ENVIOS FISICOS		
	Mayo	Junio	Julio
1		170	3
2		122	168
3		184	
4		84	
5	115		
6	65		88
7	50		
8		101	
9		116	
10	177	121	
11	105	164	
12	90		
13	128		
14	141		
15			
16		162	
17		326	
18	1	131	
19	265		
20	197		
21	99	126	
22		185	
23		3	
24	132	240	
25		137	
26			
27			
28	20	247	

Nro día	ENVIOS VIRTUALES		
	Mayo	Junio	Julio
1			
2			
3	22		
4	257		
5	642		
6	518		
7	852		
8	311		
9	0		
10	529		
11	661		
12	598		
13	589		
14	816		
15	635		
16	0		
17	176		
18	902		
19	696		
20	286		
21			
22			
23			
24			
25			
26			
27			
28			

Fuente: Tablero de control PQR's ETB

En la opción de “cierres” se puede visualizar la información de todas las PQR’s que han sido cerradas y su tipología de cierre, de acuerdo con el mes de solución y de ingreso de la solicitud generada por el usuario. En la figura 15 se observa el tablero de control de cierres

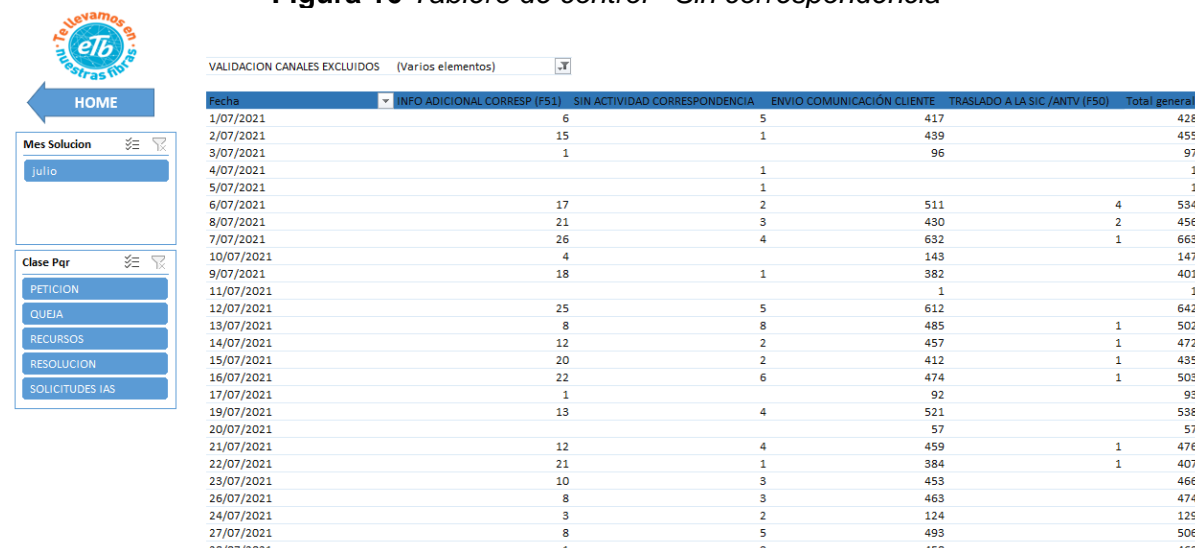
Figura 15 Tablero de control - Cierres



Fuente: Tablero de control PQR's ETB

En el enlace de “sin correspondencia”, se puede observar la información de las solicitudes que fueron cerradas en la aplicación SUMA, pero no tienen gestión con respecto a la correspondencia. En la figura 16 se observa el tablero de control de sin correspondencia.

Figura 16 Tablero de control - Sin correspondencia

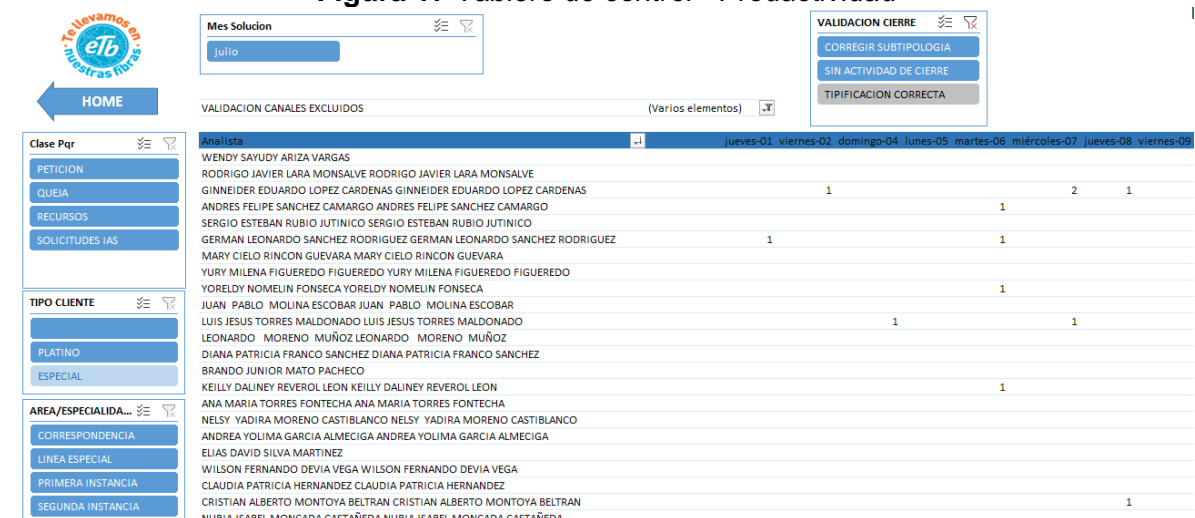


Fecha	INFO ADICIONAL CORRESP (F51)	SIN ACTIVIDAD CORRESPONDENCIA	ENVIO COMUNICACIÓN CLIENTE	TRASLADO A LA SIC /ANTV (F50)	Total general
1/07/2021	6	5	417		428
2/07/2021	15	1	439		455
3/07/2021	1		96		97
4/07/2021		1			1
5/07/2021		1			1
6/07/2021	17	2	511	4	534
8/07/2021	21	3	430	2	456
7/07/2021	26	4	632	1	663
10/07/2021	4		143		147
9/07/2021	18	1	382		401
11/07/2021			1		1
12/07/2021	25	5	612		642
13/07/2021	8	8	485	1	502
14/07/2021	12	2	457	1	472
15/07/2021	20	2	412	1	435
16/07/2021	22	6	474	1	503
17/07/2021	1		92		93
19/07/2021	13	4	521		538
20/07/2021			57		57
21/07/2021	12	4	459	1	476
22/07/2021	21	1	384	1	407
23/07/2021	10	3	453		466
26/07/2021	8	3	463		474
24/07/2021	3	2	124		129
27/07/2021	8	5	493		506
28/07/2021	1	0	450		450

Fuente: Tablero de control PQR's ETB

En el último enlace de “productividad”, se tiene la información de todas las PQR's solucionadas con respecto a cada uno de los agentes de los Call Center. En la figura 17 se observa el tablero de control de productividad

Figura 17 Tablero de control - Productividad



Analista	jueves-01	viernes-02	domingo-04	lunes-05	martes-06	miércoles-07	jueves-08	viernes-09
WENDY SAYUDY ARIZA VARGAS								
RODRIGO JAVIER LARA MONSALVE								
GINNEIDER EDUARDO LOPEZ CARDENAS								
ANDRES FELIPE SANCHEZ CAMARGO	1					1	2	1
SERGIO ESTEBAN RUBIO JUTINICO								
GERMAN LEONARDO SANCHEZ RODRIGUEZ	1					1		
MARY CIELO RINCON GUEVARA								
YURY MILENA FIGUEROA								
YORELDY NOME LIN FONSECA						1		
JUAN PABLO MOLINA ESCOBAR								
LUIS JESUS TORRES MALDONADO								
LEONARDO MORENO MUÑOZ				1			1	
DIANA PATRICIA FRANCO SANCHEZ								
BRANDO JUNIOR MATO PACHECO								
KEILLY DALINEY REVEROL LEON						1		
ANA MARIA TORRES FONTECHA								
NELSY YADIRA MORENO CASTIBLANCO								
ANDREA YOLIMA GARCIA ALMECIGA								
ELIAS DAVID SILVA MARTINEZ								
WILSON FERNANDO DEVIA WILSON								
CLAUDIA PATRICIA HERNANDEZ								
CRISTIAN ALBERTO MONTOYA BELTRAN								1
NUBIA YEABEL MONCADA CASTAÑEDA								

Fuente: Tablero de control PQR's ETB

Con este acercamiento se logró identificar que, se genera reporte diariamente de las PQR's, los cuales son analizados por personas de la dirección de experiencia de primera instancia para identificar la gestión realizada sobre los requerimientos generados por los usuarios y sobre todo para analizar los tiempos y la efectividad en el cierre de dichas solicitudes.

## 5.2. Fuente de datos y análisis de variables

Después de realizar el análisis inicial del tablero de control de la gestión de las PQR's, se logra evidenciar que dichas PQR's, no son analizadas para identificar cuáles son las solicitudes más recurrentes de los usuarios o cuales son los usuarios que tienen una visión de un mal servicio, y en consecuencia lograr tomar acciones para evitar el cambio de operador de telefonía móvil.

Se Concluye, que el operador móvil se concentra en la efectividad y solución de las PQR's creadas por los usuarios, pero no tienen un proceso definido en el que se logre analizar la causal de la generación de las solicitudes de los abonados para tomar acciones correctivas sobre los sistemas o la infraestructura, que permitan mejorar la percepción de buen servicio por parte de los usuarios.

La fuente de datos en donde se encuentra la data de las PQR's es "INFORME INGRESOS 2021 POR CANAL CREADOR.xlsx", allí se tiene toda la información relacionada con las PQR's reportadas por los abonados, las cuales son registradas en el CRM de la compañía de telefonía móvil. Se aclara que la data que se genera en el archivo son registros en los que se evidencia que las solicitudes ya fueron gestionadas. En la tabla 1 se muestran los campos definidos en el reporte:

Tabla 1 PQR's		
TABLA PQR'S		
NOMBRE COLUMNA	TIPO DATO	FORMATO/TAMAÑO
Numero Pqr	STRING	20

<b>Numero Identificación</b>	INTEGER	15
<b>Contacto Numero Documento</b>	INTEGER	15
<b>Numero conexión</b>	INTEGER	15
<b>Tecnología</b>	STRING	10
<b>Clase Pqr</b>	STRING	10
<b>Motivo Pqr</b>	STRING	50
<b>Causal Pqr</b>	STRING	50
<b>síntoma Pqr</b>	STRING	50
<b>Valor Ajustado</b>	INTEGER	20
<b>Fecha creación</b>	DATE	dd/mm/yyyy
<b>Día</b>	INTEGER	2
<b>mes</b>	INTEGER	2
<b>Fecha radicación</b>	DATE	dd/mm/yyyy
<b>Fecha solución</b>	DATE	dd/mm/yyyy
<b>Descripción Pqr</b>	STRING	200
<b>Estado Pqr</b>	STRING	10
<b>Tipo Cierre</b>	STRING	15
<b>Usuario Creador PQR</b>	STRING	8
<b>CANAL</b>	STRING	30
<b>Área</b>	STRING	50
<b>Usuario Asignado PQR</b>	STRING	8
<b>AREA ESCRITA</b>	STRING	30
<b>Días hábiles</b>	INTEGER	5

Otra fuente de datos importante es el reporte en el que se almacena la información de los abonados, ese reporte tiene como nombre “InfoDesarrolloCliente\_ProtegidaInfoSensibleyHabeasData.txt”. En la tabla 2 se muestran los campos definidos:

**Tabla 2 Abonados**

TABLA Abonados		
NOMBRE COLUMNA	TIPO DATO	FORMATO/TAMAÑO
TIPO_DOCUMENTO	INTEGER	20
NUMERO_IDENTIFICACION	INTEGER	15
ID_TIPO_CLIENTE	INTEGER	10
ID_SEGMENTO	INTEGER	10
ID_SEGMENTO_UEN	INTEGER	10



<b>ID_CIUADAD</b>	INTEGER	10
<b>ID_LOCALIDAD</b>	INTEGER	10
<b>ID_BARRIO</b>	INTEGER	10
<b>ID_DEPARTAMENTO</b>	INTEGER	10
<b>ID_ESTADO</b>	INTEGER	10
<b>UEN</b>	STRING	50
<b>PAIS</b>	STRING	50
<b>ESTRATO</b>	INTEGER	5
<b>GENERO</b>	STRING	10
<b>RANGO_EDAD</b>	STRING	15
<b>ESTADO_CIVIL</b>	STRING	10
<b>FECHA_CREACION</b>	DATE	dd/mm/yyyy
<b>FECHA_ACTUALIZACION</b>	DATE	dd/mm/yyyy
<b>CATEGORIA</b>	STRING	10

Las visualizaciones para presentar en el método deben organizarse de manera tal que ayuden al área de fidelización de clientes a tener un panorama amplio de los clientes que puedan ser potenciales a cambiar de operador móvil. En tales visualizaciones también se debe tener información de las posibles causas que pueda llevar a los abonados a tomar la decisión del cambio de operador.

A continuación, se establecen las variables claves para obtener los resultados del método:

- Clase Pqr
- Motivo Pqr
- Causal Pqr
- Descripción Pqr
- Numero Identificación
  - Tipo Documento
  - Tipo Cliente

- Segmento Uen
- Ciudad
- Localidad
- Barrio
- Departamento
- Estado
- Uen
- Estrato
- Genero
- Rango Edad
- Estado Civil
- Fecha creación
- Fecha actualización
- Categoría

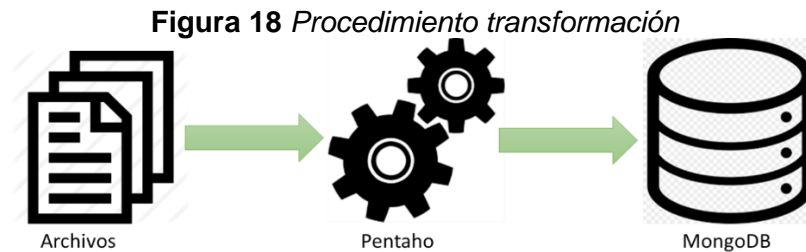
### 5.3. Diseño del modelo de datos

En la siguiente fase se desarrolla el procedimiento de la unificación de las fuentes definidas para el método, las cuales son proporcionadas por la empresa de telefonía móvil. Con base en la información obtenida se definen las variables a cargar en la BD (MongoDB), las cuales son importantes en la creación del modelo.

El procedimiento de cargue de información se realiza cada 15 días, debido a que es el tiempo en el que se cuenta con la data disponible, este proceso es llamado definición de ETL, el cual consiste en realizar una transformación de los archivos originales a una estructura que pueda ser leída por una BD, en este caso el almacenamiento se realiza en MongoDB.

La aplicación seleccionada para el diseño e implementación de la ETL es Pentaho Data Integration (PDI), software libre, que cuenta con una licencia GPL2, lo cual quiere decir que se puede usar, copiar, distribuir y modificar de forma gratuita, por tal razón, fue elegido este software para el desarrollo del método.

Estas transformaciones se han diseñado en Pentaho, en el cual se automatiza el cargue y transformación de la data, teniendo beneficios como tiempos cortos de análisis de información y evitando la manipulación de la información. En figura 18 se observa el proceso de transformación de forma general.



Fuente: Elaboración propia

En figura 19, podemos observar las fases que se realizan en la aplicación Pentaho, para el desarrollo de la ETL.



Fuente: Elaboración propia

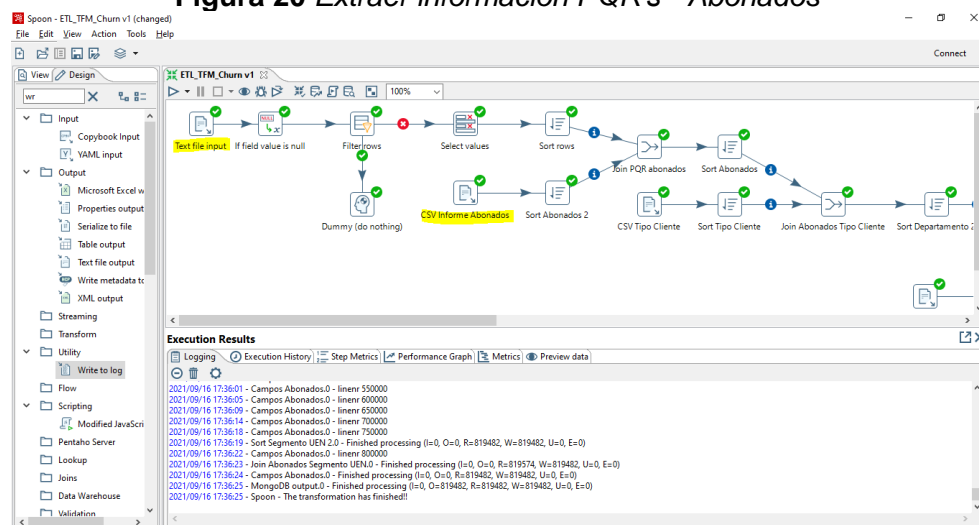
A continuación, se realiza una descripción de cada una de las fases, con las cuales se obtiene la generación de la ETL.

### 5.3.1. Extraer

En esta fase el objetivo es extraer los datos de las fuentes necesarias, y producir datos limpios y accesibles, los cuales puedan ser utilizados en la fase de transformación.

En Pentaho la ejecución de esta fase, se realiza a través la lectura de los archivos. En el área de trabajo de la aplicación y en el diseño se selecciona la entrada de datos CSV, como se observa en la figura 20.

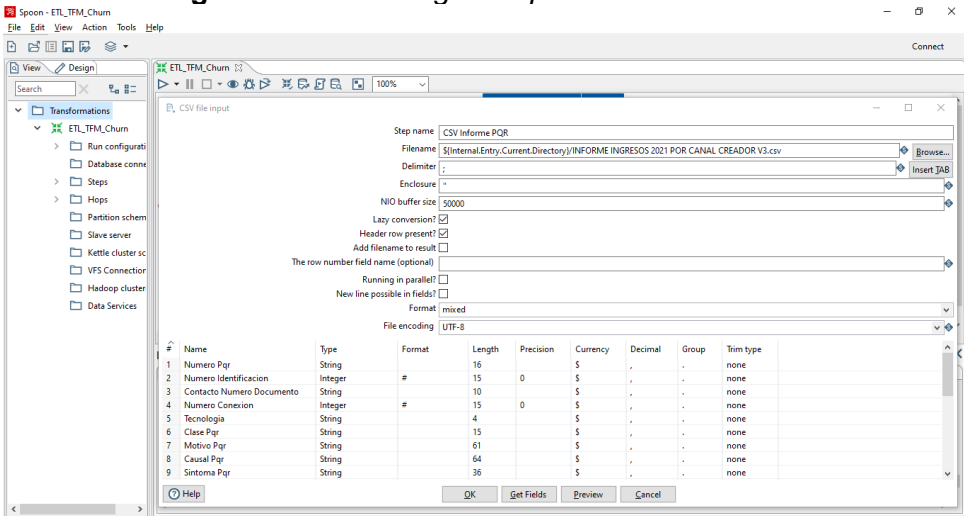
**Figura 20** Extraer información PQR's - Abonados



Fuente: Elaboración propia

En la figura 21 se observa el proceso de extracción de la información del archivo INFORME INGRESOS 2021 POR CANAL CREADOR V3.csv, donde se logra visualizar por cada tipo de registro las siguientes columnas: el nombre, tipo de dato, formato, tamaño, precisión, decimal y grupo. También se ajustaron los nombres de los campos que al realizar el proceso de extracción tenían caracteres especiales.

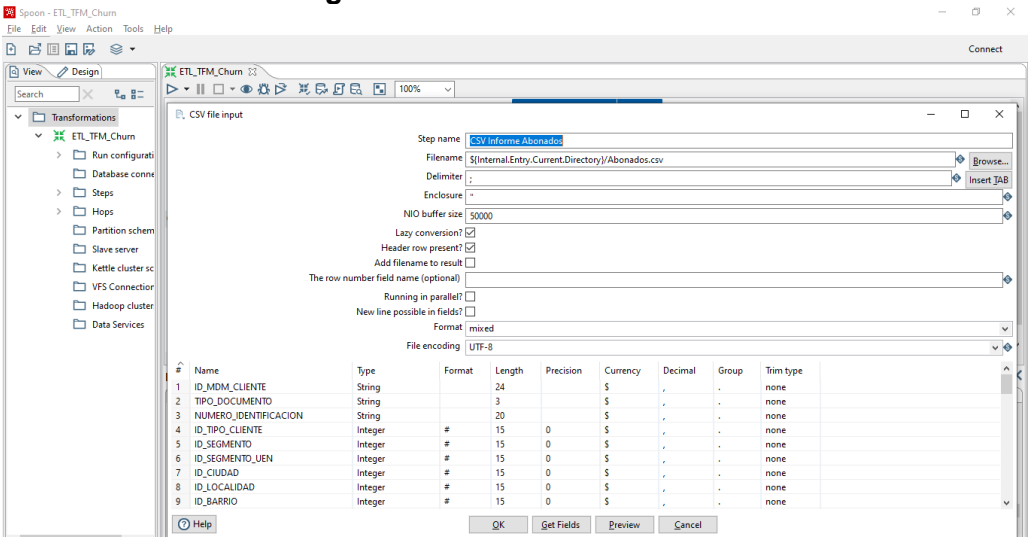
Figura 21 Informe ingresos por canal creador



Fuente: Elaboración propia

En la figura 22, se observa el proceso de extracción de la información del archivo InfoDesarrolloCliente\_ProtegidaInfoSensibleyHabeasData.txt, en el que, se logra visualizar por cada tipo de registro las siguientes columnas: el nombre, tipo de dato, formato, tamaño, precisión, decimal y grupo. También se ajustaron los nombres de los campos que al realizar el proceso de extracción tenían caracteres especiales.

Figura 22 Informe Abonados



Fuente: Elaboración propia

Como se observa en la figura 23, se puede visualizar los resultados de la extracción de la información de los abonados.

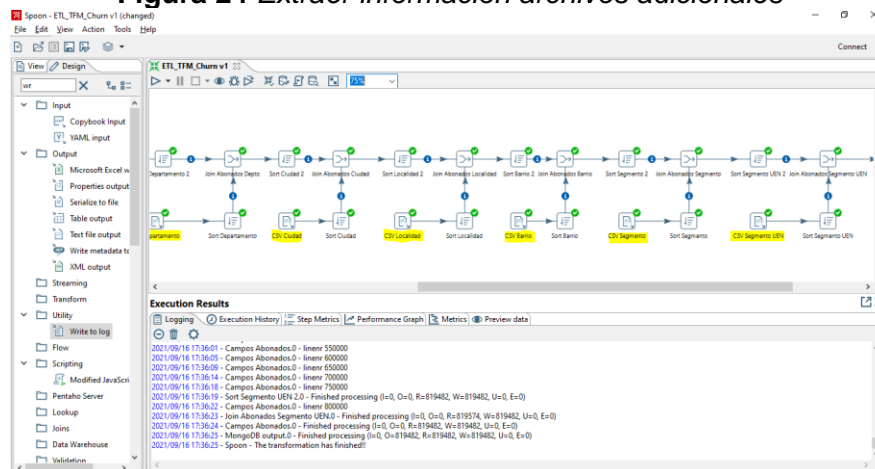
**Figura 23 Resultado información Abonados**

[illegible]

Fuente: Elaboración propia

También se realizó el cargue de información de los archivos de: tipo de cliente, departamento, ciudad, localidad, barrio, segmento y segmento UEN, ya que estos archivos son cruzados con el archivo de abonados, debido a que en este sólo están los ID de la información y de esta manera se obtiene la información de dichos campos. En la figura 24 se observa la extracción de la información.

### Figura 24 Extraer información archivos adicionales



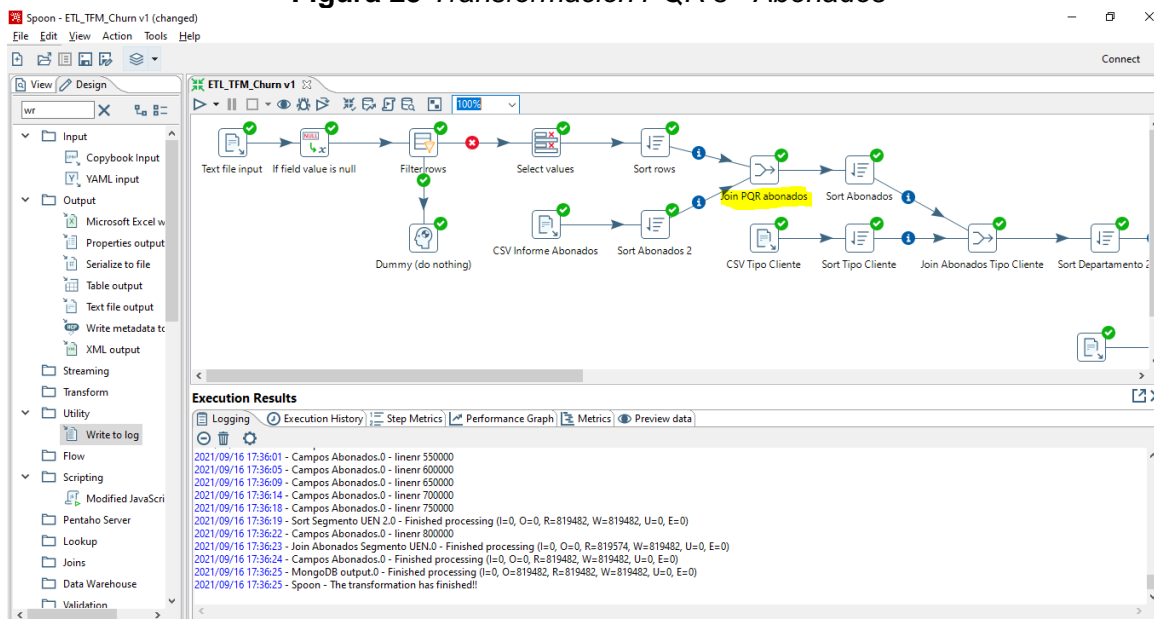
Fuente: Elaboración propia

### 5.3.2. Transformar

Esta fase es la más importante y crítica de todo el proceso de la generación de la ETL, debido a que se tiene como objetivo compilar, reformatear y limpiar los datos extraídos en la fase de extracción.

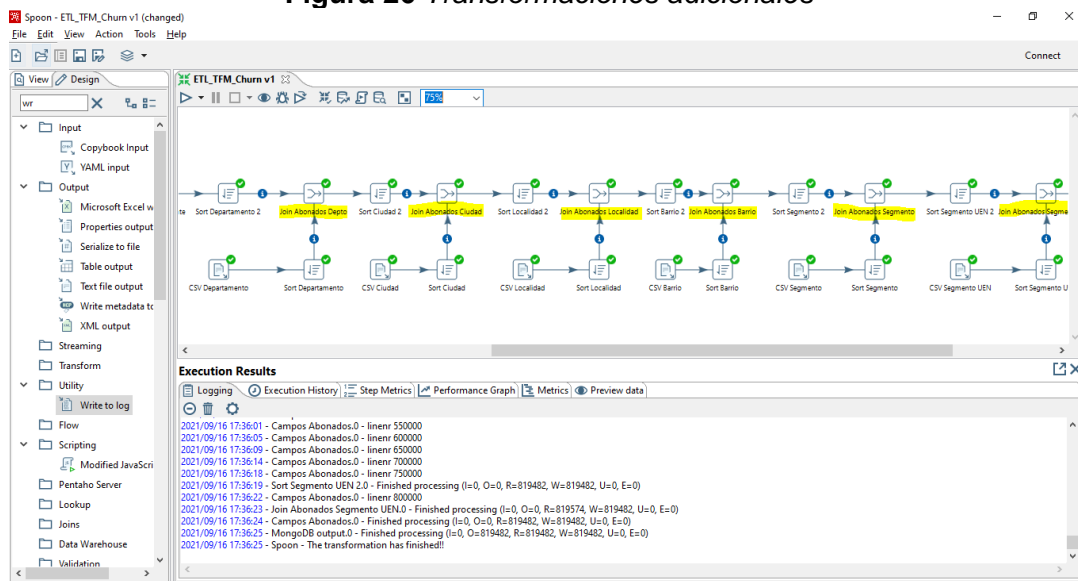
La ETL está compuesta de varias uniones de información, pero la más importante es la relacionada con la concatenación de las PQR's y los Abonados. En la figura 25 se observa la transformación.

**Figura 25 Transformación PQR's - Abonados**



Fuente: Elaboración propia

También se tiene otras concatenaciones de información, desde la data de abonados con cada uno de los campos donde se tienen los ID, estos son: tipo de cliente, departamento, ciudad, localidad, barrio, segmento y segmento UEN.

**Figura 26 Transformaciones adicionales**

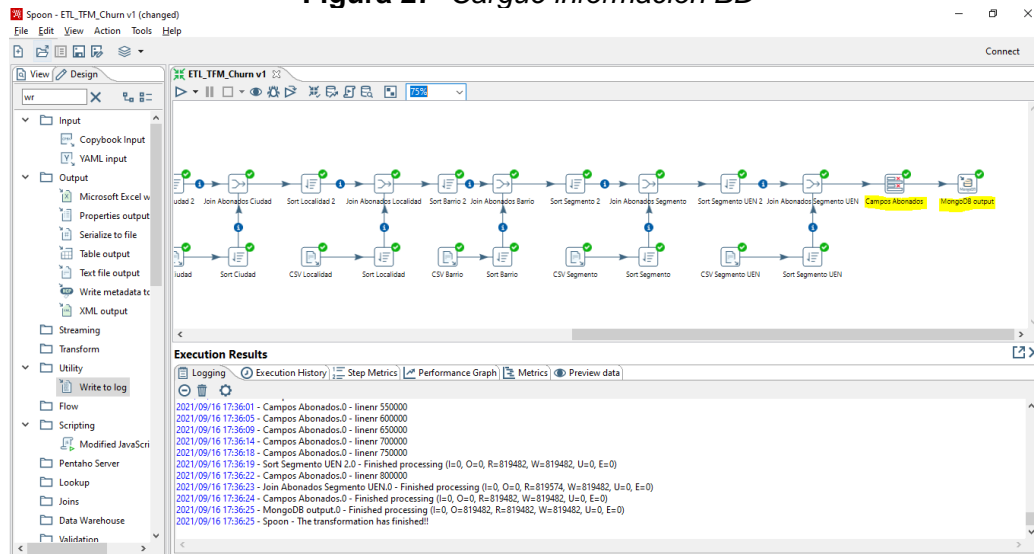
Fuente: Elaboración propia

### 5.3.3. Cargar

La última fase del proceso consiste en cargar los datos extraídos y transformados al nuevo destino. Este cargue se puede realizar de dos maneras, la primera es una inserción uno a uno de los registros a la base de datos destino, y la segunda opción es un cargue masivo de datos.

Para nuestro caso, se utilizó la forma de cargar la información directo a la BD de MongoDB. En figura 27 se observa el proceso de cargue.



**Figura 27** *Cargue información BD*

Fuente: Elaboración propia

En la figura 28 se logra visualizar la información almacenada en MongoDB, donde se creó la BD ET\_Churn y el esquema Churn, para guardar la información de las PQR's y sus abonados.

**Figura 28** *Información en la BD*

#	ID_BARRIO	BARRIO	ID_DEPARTAMENTO	DEPARTAMENTO	UEN
9	-1	"SIN DATA"	39	"VALLE DEL CAUCA"	No field
10	-1	"SIN DATA"	39	"VALLE DEL CAUCA"	No field
11	-1	"SIN DATA"	39	"VALLE DEL CAUCA"	No field
12	-1	"SIN DATA"	39	"VALLE DEL CAUCA"	No field
13	-1	"SIN DATA"	39	"VALLE DEL CAUCA"	No field
14	-1	"SIN DATA"	39	"VALLE DEL CAUCA"	No field
15	-1	"SIN DATA"	39	"VALLE DEL CAUCA"	No field
16	-1	"SIN DATA"	39	"VALLE DEL CAUCA"	No field
17	-1	"SIN DATA"	27	"MAGDALENA"	No field
18	-1	"SIN DATA"	9	"BOLIVAR"	No field
19	-1	"SIN DATA"	9	"BOLIVAR"	No field
20	-1	"SIN DATA"	9	"BOLIVAR"	No field

Fuente: Elaboración propia

En la tabla 3 se muestran los campos, el tipo y tamaño, de cada uno de los documentos almacenados en la BD de MongoDB, esto después de realizar la transformación de la información de las PQR's y sus abonados.

**Tabla 3** Campos generados ETL

TABLA PQR's		
NOMBRE COLUMNA	TIPO DATO	FORMATO/TAMAÑO
Numero Pqr	STRING	20
Contacto Numero Documento	INTEGER	15
tecnología	STRING	10
Clase Pqr	STRING	10
Motivo Pqr	STRING	50
Causal Pqr	STRING	50
síntoma Pqr	STRING	50
Valor Ajustado	INTEGER	20
Fecha creación	DATE	dd/mm/yyyy
día	INTEGER	2
mes	INTEGER	2
Fecha radicación	DATE	dd/mm/yyyy
Fecha solución	DATE	dd/mm/yyyy
descripción Pqr	STRING	200
Estado Pqr	STRING	10
Tipo Cierre	STRING	15
CANAL	STRING	30
AREA ESCRITA	STRING	30
Días hábiles	INTEGER	5
TIPO_DOCUMENTO	STRING	3
NUMERO_IDENTIFICACION	STRING	20
ID_TIPO_CLIENTE	INTEGER	15
TIPO_CLIENTE	INTEGER	15
ID_SEGMENTO	INTEGER	15
SEGMENTO	STRING	23
ID_SEGMENTO_UEN	INTEGER	15
SEGMENTO_UEN	STRING	23
ID_CIUADAD	INTEGER	15
CIUDAD	STRING	30
ID_LOCALIDAD	INTEGER	15
LOCALIDAD	STRING	25
ID_BARRIO	INTEGER	15

BARRIO	STRING	30
ID_DEPARTAMENTO	INTEGER	15
DEPARTAMENTO	STRING	15
UEN	STRING	22
CLIENTE_VALOR	STRING	6
COBERTURA	STRING	6
PAIS	STRING	8
ESTRATO	STRING	6
GENERO	STRING	6
RANGO_EDAD	STRING	6
ESTADO_CIVIL	STRING	6
FECHA_CREACION	DATE	dd/mm/yyyy
FECHA_ACTUALIZACION	DATE	dd/mm/yyyy
FECHA_INICIO	DATE	dd/mm/yyyy
FECHA_FIN	DATE	dd/mm/yyyy

#### 5.4. Presentación de la herramienta de diseño y visualización

Gracias al conocimiento adquirido en la maestría sobre herramientas de visualización de datos, se ha definido el software de Tableau desktop para la generación de las gráficas para el método de retención de clientes, ello debido a que un software líder en el mercado con respecto a las herramientas de visualización de datos se considera adicionalmente como una herramienta interactiva, donde el usuario puede interactuar con las visualizaciones, realizando filtros, conectando variables entre sí. Otro punto a favor es que las gráficas son muy visuales y esto hace que sean fáciles de comprender.

Debido al potencial de la aplicación Tableau se han desarrollado las visualizaciones, para que los usuarios puedan tener herramientas para la toma de decisiones con respecto a cuales son los usuarios que han generado más de una solicitud, también cuales son las causas más comunes de solicitud de PQR's, entre otros aspectos relevantes, los cuales ayudan a comprender a los usuarios por qué se presenta la generación de PQR's, y de esta manera contactar a otros abonados y dar una solución temprana con fallas en el servicio, previo a que estos

generen solicitudes relacionadas con los problemas y de esta manera dar una percepción de buen servicio y así evitar que los abonados cambien de operador.

Se creó un tablero de control para agrupar y comprender mejor las visualizaciones desarrolladas, como se muestra a continuación de manera general:

1. **Resumen:** En la primera visualización se muestran los resultados generales y de gran importancia para el negocio y para la toma de decisiones, entre las visualizaciones se tienen: PQR's generadas por mes, PQR's por clase, PQR's por tipo de cliente PQR's por estado, total de PQR's y el total del valor ajustado para la solución de las PQR's, de esta manera, se logra tener un acercamiento de los usuarios a la información y así se genera una necesidad de continuar con la navegación en el tablero de control.
2. **Motivos y Causas:** Las visualizaciones que se muestran en esta sección del tablero de control están relacionadas con la generación de las PQR's, donde se puede observar los gráficos sobre las solicitudes generadas por los abonados. Dentro de la información que se puede analizar se tiene, el motivo y la causal de la generación de las PQR's. Con las visualizaciones mostradas en esta sección, se logra identificar los motivos y las causas más frecuentes por las que, los usuarios generan PQR's, y así lograr gestionar las soluciones de fondo a los problemas más recurrentes para evitar que otros abonados generen PQR's relacionados con tales motivos, esto con el objetivo principal de satisfacción de los clientes respecto al servicio y buscar que declinen de la intención de cambiar de operador.
3. **Cierre y solución:** En este apartado del tablero de control se generaron visualizaciones relacionadas con la solución a las PQR's, la información que se observa en los gráficos y que se puede analizar, está relacionada con el tipo de cierre, días de solución, canal de solución y área de solución. Con estas visualizaciones se puede

obtener información importante de la gestión realizada de las solicitudes, en el que se logra definir qué tipo de solución se otorgó y si estuvo dentro del tiempo dispuesto para la atención, y el área que está gestionando la solución. Con base en esta información se logra realizar un análisis de los clientes que no están satisfechos con la solución y de cuales PQR's se gestionaron en tiempo fuera de los ANS, para que el área encargada de la relación de clientes realice un contacto con los usuarios a efectos de mejorar la percepción de servicio, gestionado de esta manera la retención de clientes y evitar un cambio de operador.

4. **Mapa:** Por último, se tienen visualizaciones relacionadas con la ubicación de los abonados que generaron las PQR's, esto para lograr identificar cuáles son los departamentos y las ciudades donde se registran más solicitudes y así establecer si se trata de un factor común con respecto a la ubicación para que los abonados generen estas solicitudes.

## 6. Resultados

El método diseñado tiene como propósito agrupar la información relacionada con la generación de las PQR's, para que esta se analizada para la toma de decisión por parte del área correspondiente de la retención de clientes, permitiendo una reducción en la deserción de los abonados y ayudando a tener un ahorro económico en la compañía.

Debido al análisis realizado del método de retención de clientes, se desarrolló un tablero de control, en el que se extrae la información más importante y relevante para el proceso de retención de clientes, pero esta debe ser analizada y gestionada por los empleados.

A continuación, se detallan cada una de las secciones del tablero de control:

### a. Resumen

En la visualización de resumen se muestran resultados generales y de gran importancia para el negocio y para la toma de decisiones, entre las visualizaciones se tienen:

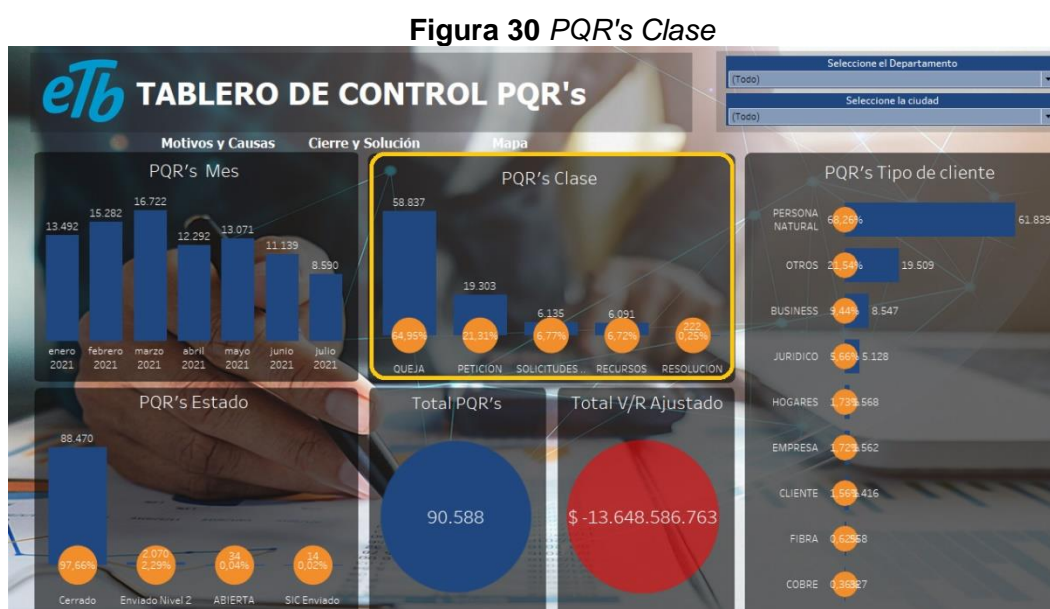
- PQR's generadas por mes: En la figura 29 se logra observar la cantidad de las solicitudes generadas por cada uno de los meses.

**Figura 29 PQR's Mes**



Fuente: Elaboración propia

- PQR's por clase: Con esta visualización se puede observar las clases o tipologías de las PQR's, ordenadas de mayor a menor, observando que las “quejas” son las solicitudes más comunes, el tipo de solicitud de “petición” ocupa un segundo lugar con mucho menos porcentaje con respecto a las quejas, así mismo, con menos cantidad en la generación de solicitudes se tienen “solicitudes de IAS” las cuales son generadas por entidades del estado, y otros tipos de solicitudes como “recursos” y “resolución”, con muy poco porcentaje de generación de PQR's. En la figura 30 se observa las PQR's por clase.



Fuente: Elaboración propia

- PQR's por tipo de cliente: En la figura 31 se observa los tipos de clientes que generan las solicitudes, se observa que en mayor porcentaje las personas naturales son las que más generan PQR's, también se tiene un elevado número de casos en los que no se tiene identificado el tipo de cliente, lo cual afecta el análisis de las PQR's relacionadas con los tipos de clientes.

**Figura 31 PQR's Tipo de cliente**

Fuente: Elaboración propia

- PQR's por estado: En la figura 32 se observa las PQR's por estado, en la cual un porcentaje muy alto las solicitudes fueron gestionadas, se evidencia que se tienen pocos casos donde dichas solicitudes siguen abiertas, otras escaladas a un segundo nivel o en su defecto fueron escaladas a la Superintendencia de Industria y Comercio (SIC) para la su solución.

**Figura 32 PQR's Estado**

Fuente: Elaboración propia



- Total del valor ajustado para la solución de las PQR's: Esta visualización es importante, debido a que nos muestra el valor en pesos colombianos de cuánto han sido los ajustes realizados a la facturación de los abonados para la solución de las PQR's. En la figura 33 se observa el valor ajustado en las PQR's.

**Figura 33** Total valor ajustado



Fuente: Elaboración propia

Adicionalmente todas las visualizaciones anteriormente mencionadas, se pueden filtrar por el departamento y la ciudad de ubicación del usuario que generó la PQR's, como se muestra en la figura 34:

**Figura 34** Resumen filtro departamento ciudad

Fuente: Elaboración propia

Al seleccionar un departamento y una ciudad, la información de las visualizaciones se actualiza de manera automática, ayudando a generar un análisis más sesgado de acuerdo con la ubicación.

También las visualizaciones se actualizan de manera automática, al permitir seleccionar un mes, una clase un tipo de cliente o un estado de las PQR's, haciendo mucho más interactivo el tablero de control. A continuación, se observan los gráficos:

- Filtro mes

Figura 35 *Filtro mes*

Fuente: Elaboración propia

- Filtro clase

Figura 36 *Filtro clase*

Fuente: Elaboración propia

- Filtro tipo cliente

Figura 37 Filtro tipo de cliente



Fuente: Elaboración propia

- Filtro estado

Figura 38 Filtro estado



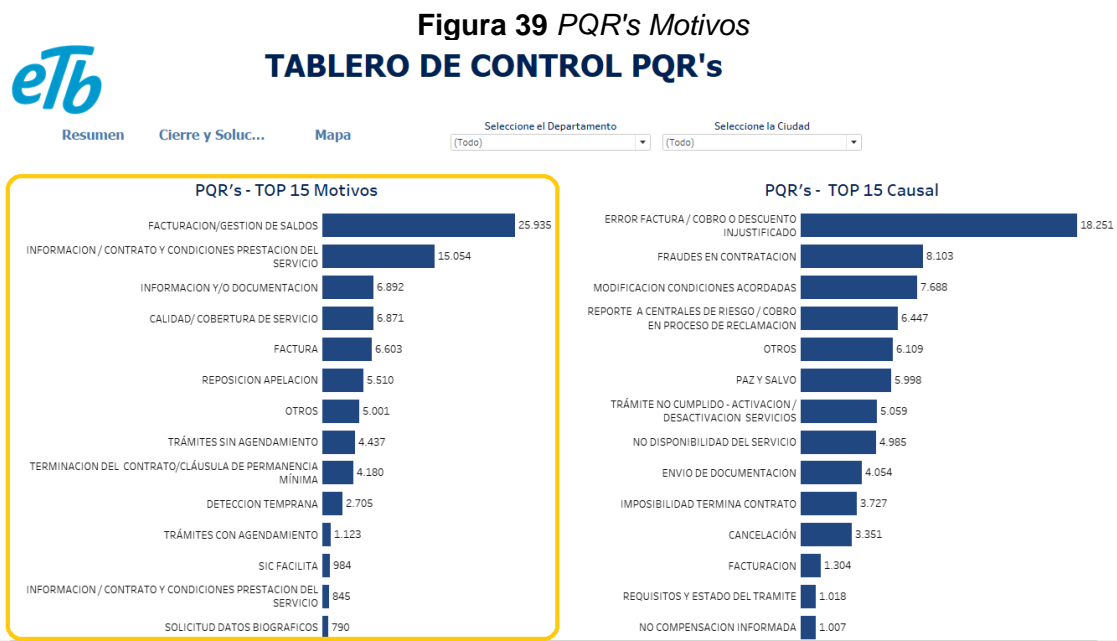
Fuente: Elaboración propia



**b. Motivos y causas**

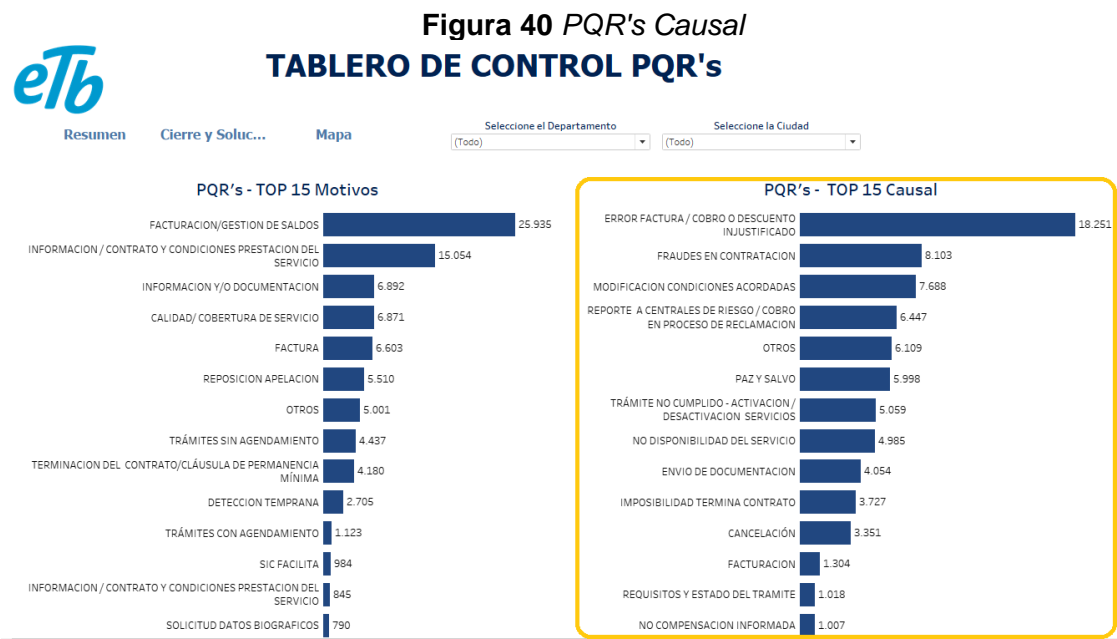
Las visualizaciones que se muestran en esta sección del tablero de control están relacionadas con la generación de las PQR's, lo que contribuye a identificar los motivos y causas principales de la generación de las solicitudes por los abonados.

- PQR's top 15 motivos: En dicha visualización se logra obtener los 15 motivos más comunes de generación de solicitudes de PQR's por los abonados. En la figura 39 se observan las PQR's por motivo



Fuente: Elaboración propia

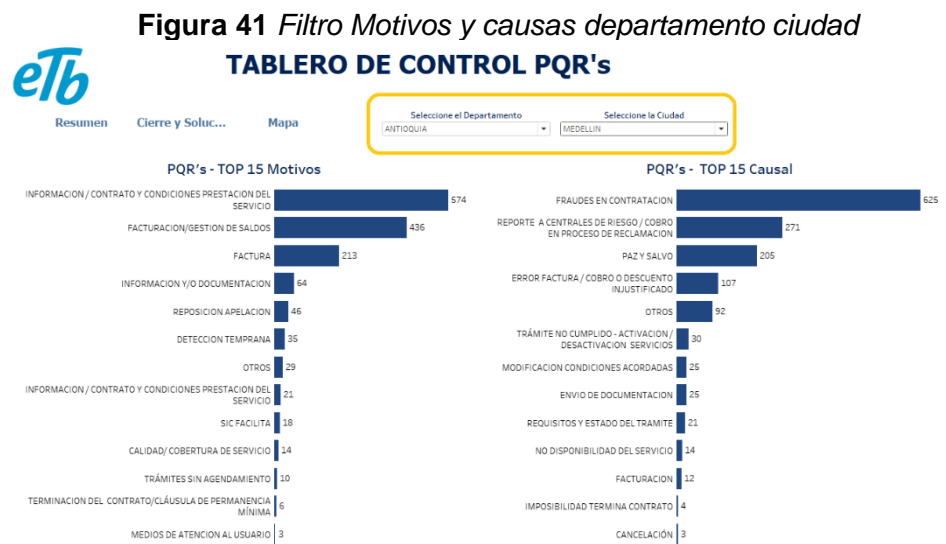
- PQR's top 15 causal: En dicha visualización se logra obtener las 15 causales más comunes de generación de solicitudes de PQR's por los abonados. En la figura 40 se observa las PQR's por causal



Fuente: Elaboración propia

Con esta información de motivo y causal de la generación de las PQR's, se puede realizar un análisis que conlleve a la creación de acciones de mejora en la solución de las fallas identificadas en el servicio, y de esta manera lograr disminuir la generación de PQR's, con lo cual se aporta a la retención de usuarios y a la disminución en el valor de la facturación que se debe que ajustar para la retención de los clientes.

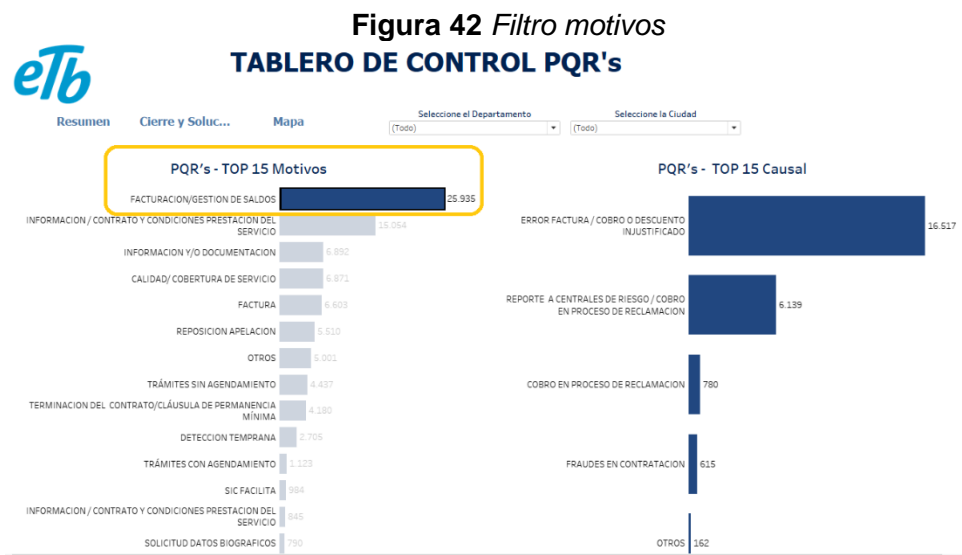
Adicionalmente los gráficos se pueden filtrar por el departamento y la ciudad de ubicación del usuario que generó la PQR's. En la figura 41 se observa las opciones de filtro de la visualización.



Fuente: Elaboración propia

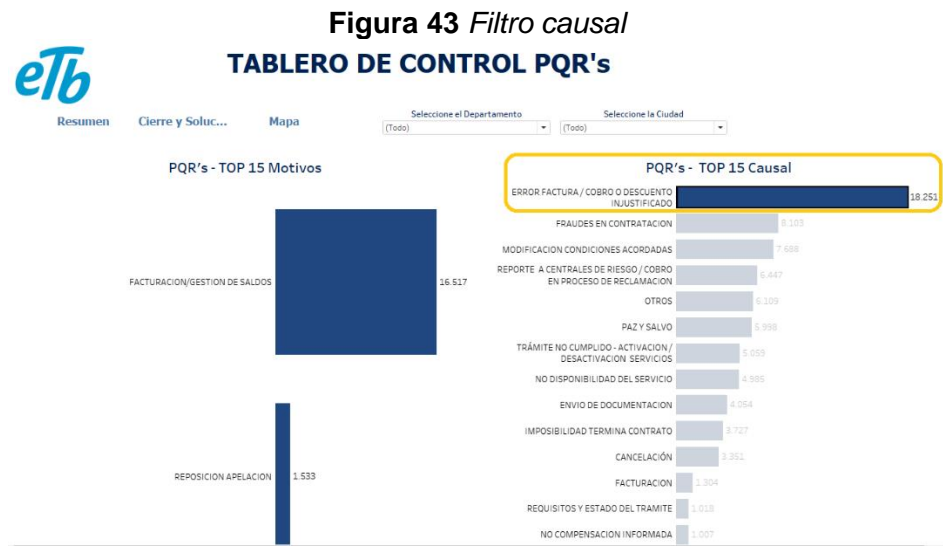
También las visualizaciones se actualizan de manera automática, al seleccionar un motivo, o una causal de las PQR's, haciendo mucho más interactivo el tablero de control. A continuación, se observan los gráficos:

- Filtro motivo



Fuente: Elaboración propia

- Filtro causal



Fuente: Elaboración propia

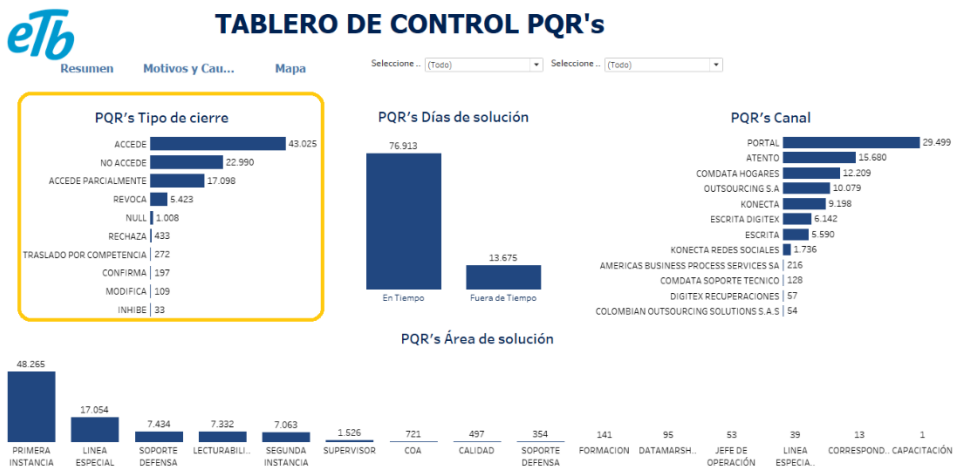
**c. Cierres y solución**

En esta sección del tablero de control se generaron gráficos relacionadas con la solución y cierre de las PQR's, por ellos la información que se observa en las visualizaciones está relacionada con:

- PQR's tipo de cierre: Con esta visualización se logra obtener los tipos de cierres de las PQR's creadas por los usuarios. En la figura 44 se observa las PQR's por tipo de cierre.



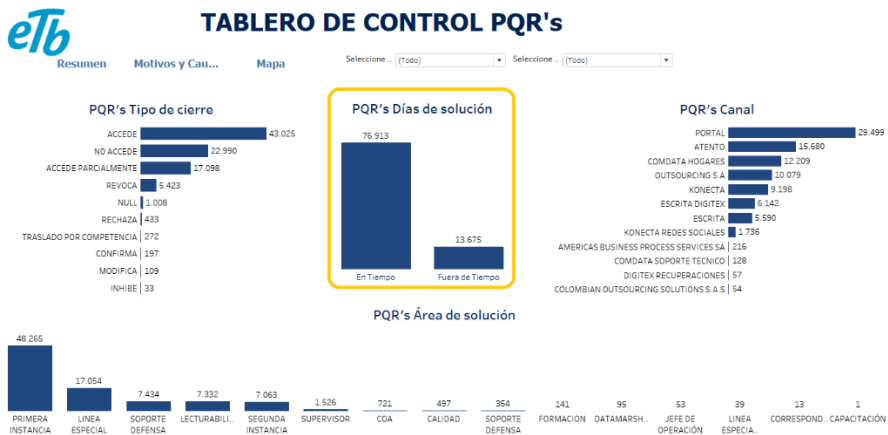
Figura 44 PQR's tipo de cierre



Fuente: Elaboración propia

- PQR's días de solución: Con esta visualización se logra identificar las PQR's solucionadas dentro del ANS de atención, el cual corresponde a 21 días hábiles. En la figura 45 se observa las PQR's por días de solución.

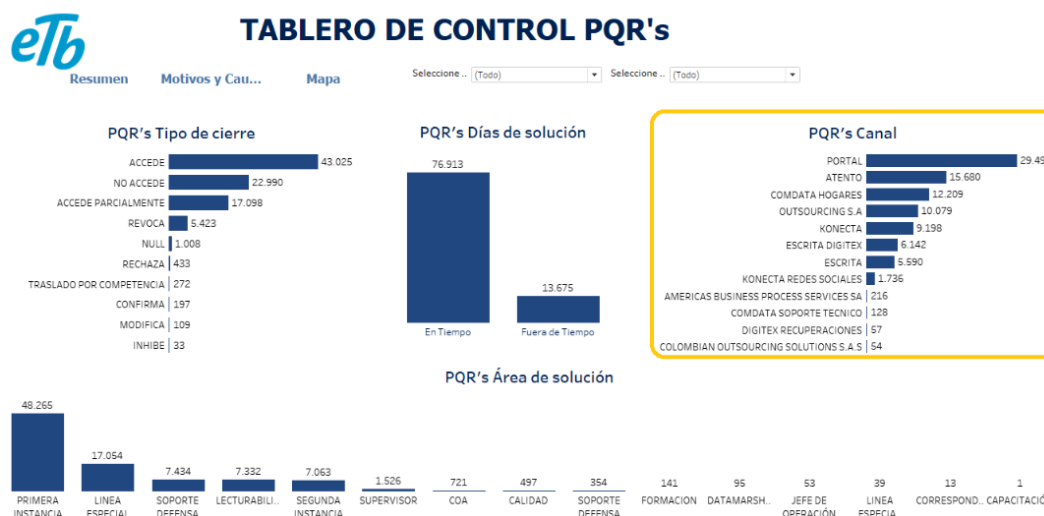
Figura 45 PQR's días de solución



Fuente: Elaboración propia

- PQR's canal de solución: Este gráfico se identifica las PQR's solucionadas por los distintos canales de atención, haciendo referencia a los call center. En la figura 46 se observa las PQR's por canal.

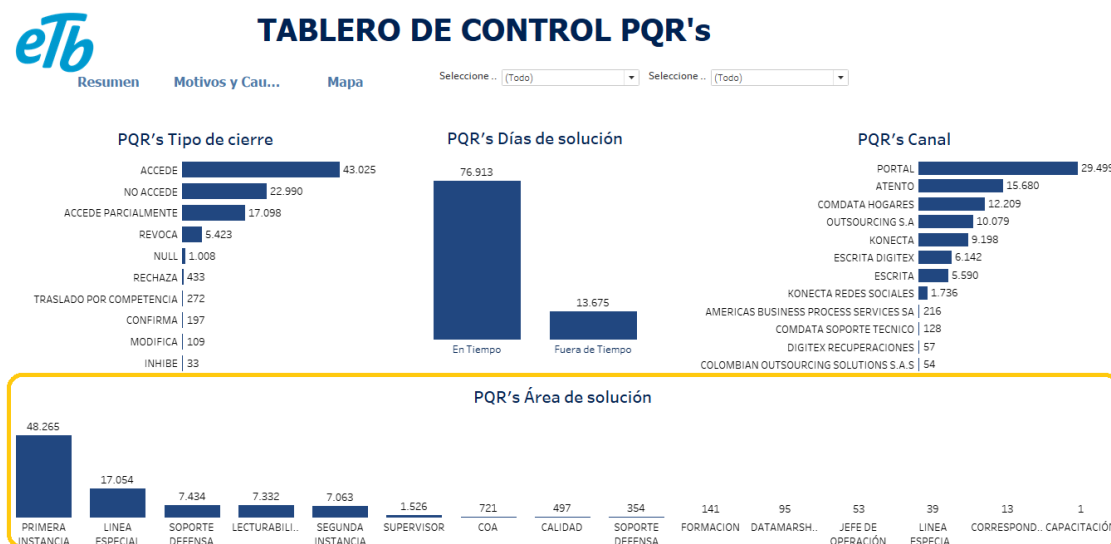
Figura 46 PQR's canal



Fuente: Elaboración propia

- PQR's área de solución: Este gráfico identifica las PQR's solucionadas por las distintas áreas en los canales de atención. En la figura 47 se observa las PQR's por área de solución.

Figura 47 PQR's área de solución



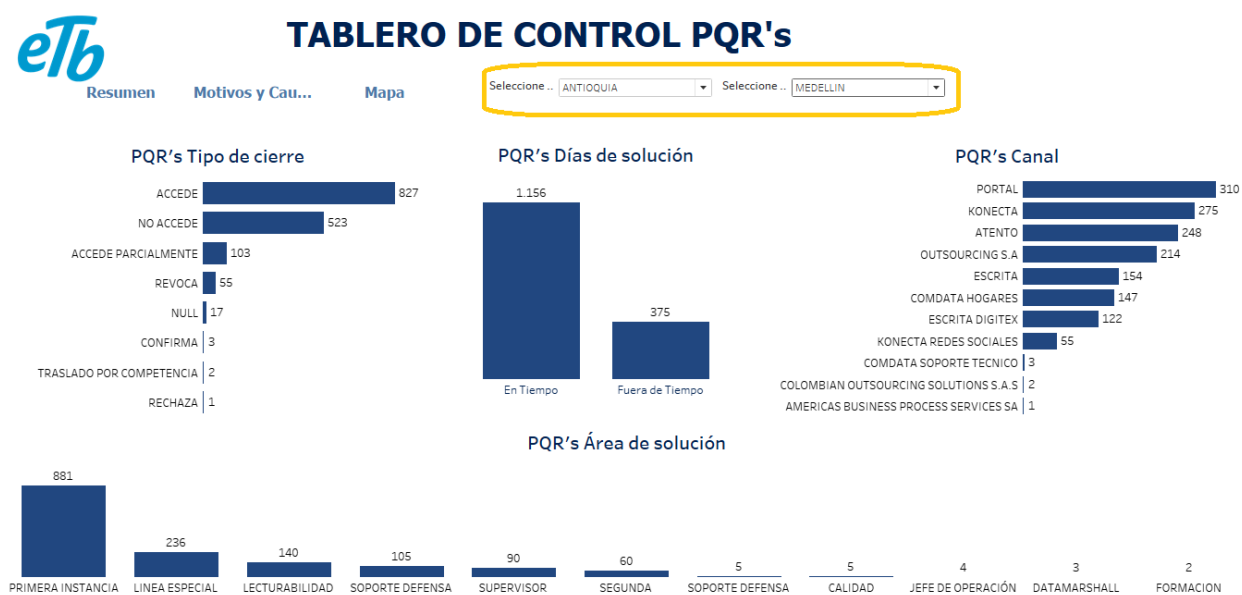
Fuente: Elaboración propia

Con estas visualizaciones se obtiene información importante de la gestión realizada sobre las solicitudes, así como, qué tipo de solución se otorgó y si la solución estuvo dentro del tiempo definido para la atención de estas, la cual es de 21 días hábiles, y el área y canal

que está gestionando la solución. Con base en esta información se consigue realizar un análisis de los clientes que no están satisfechos con la solución, así como, cuales PQR's se gestionaron en tiempo fuera de los ANS, para que el área encargada de la relación de clientes realice un contacto con los usuarios con el propósito de mejorar la percepción de servicio, gestionado de esta manera la retención de clientes y evitar un cambio de operador.

Todas las visualizaciones de la sección se pueden filtrar por el departamento y la ciudad de ubicación del usuario que generó la PQR's. En la figura 48 se observa las opciones de filtro.

**Figura 48** Filtro cierres y solución departamento ciudad

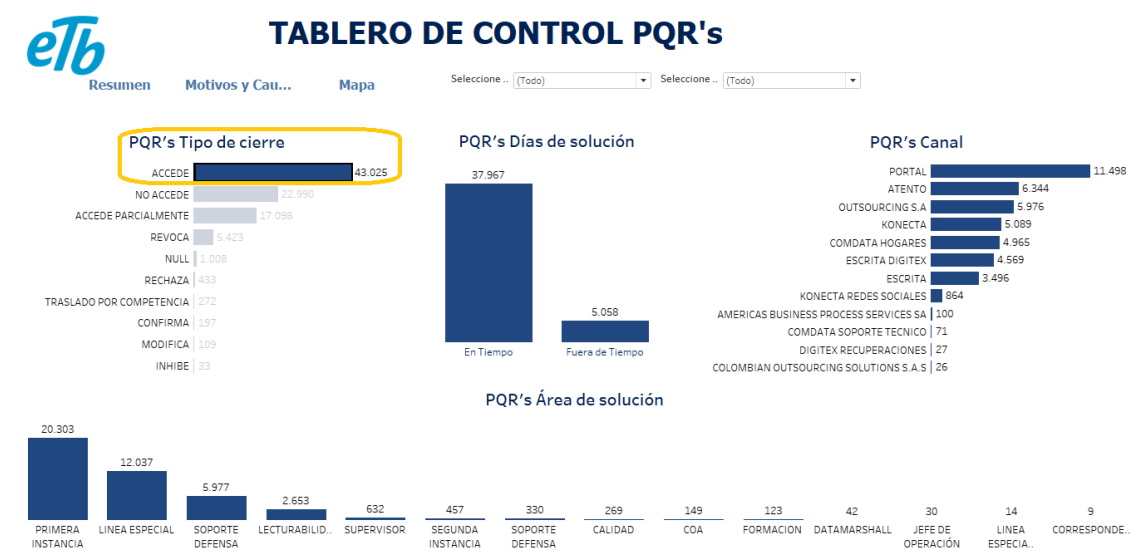


Fuente: Elaboración propia

También se logra actualizar las gráficas al seleccionar cualquiera de los valores de las visualizaciones de tipo de cierre, días de solución, canal o área de solución de las PQR's, haciendo aún más interactivo el tablero de control.

- Filtro tipo cierre

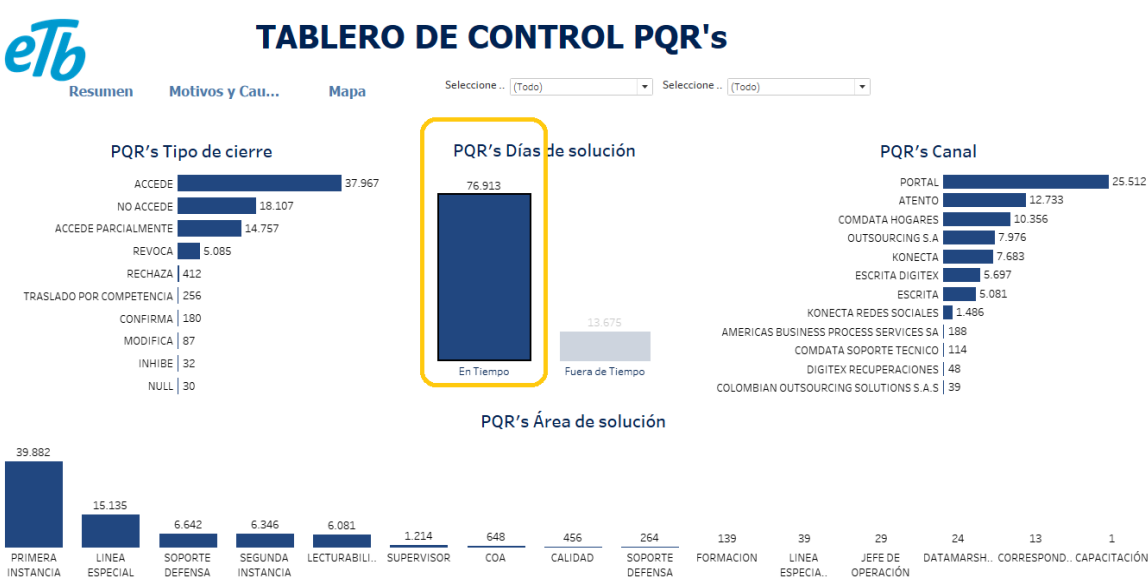
Figura 49 Filtro tipo de cierre



Fuente: Elaboración propia

- Filtro días solución

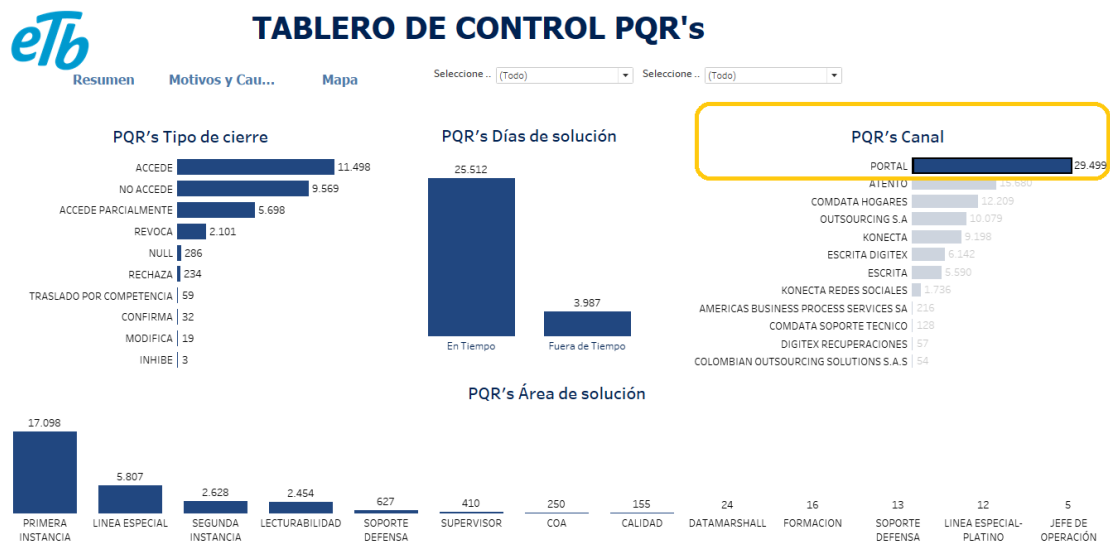
Figura 50 Filtro días solución



Fuente: Elaboración propia

- Filtro canal

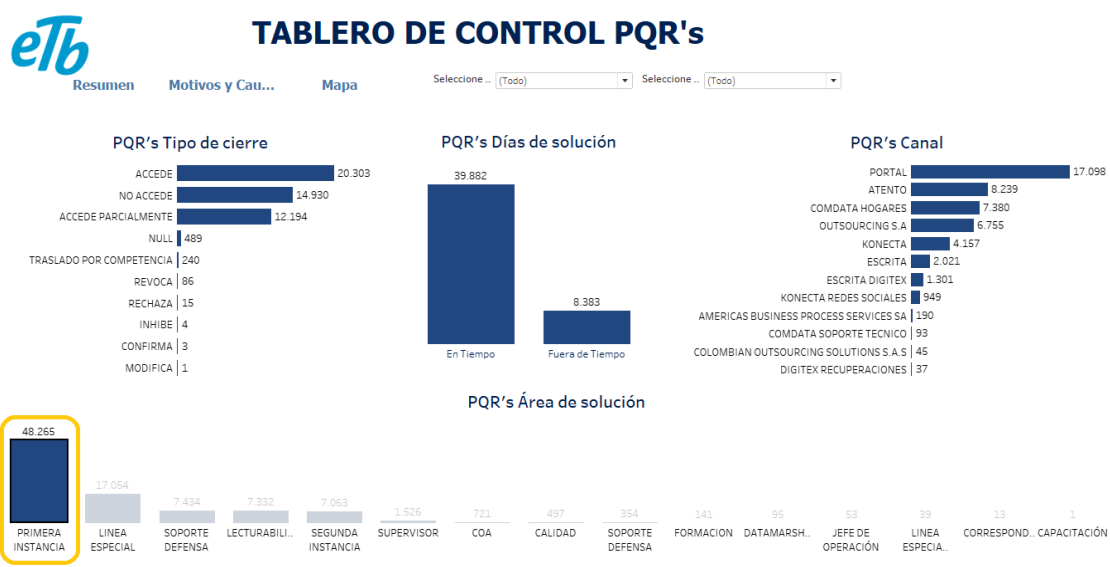
Figura 51 Filtro canal



Fuente: Elaboración propia

- Filtro área de solución

Figura 52 Filtro área de solución



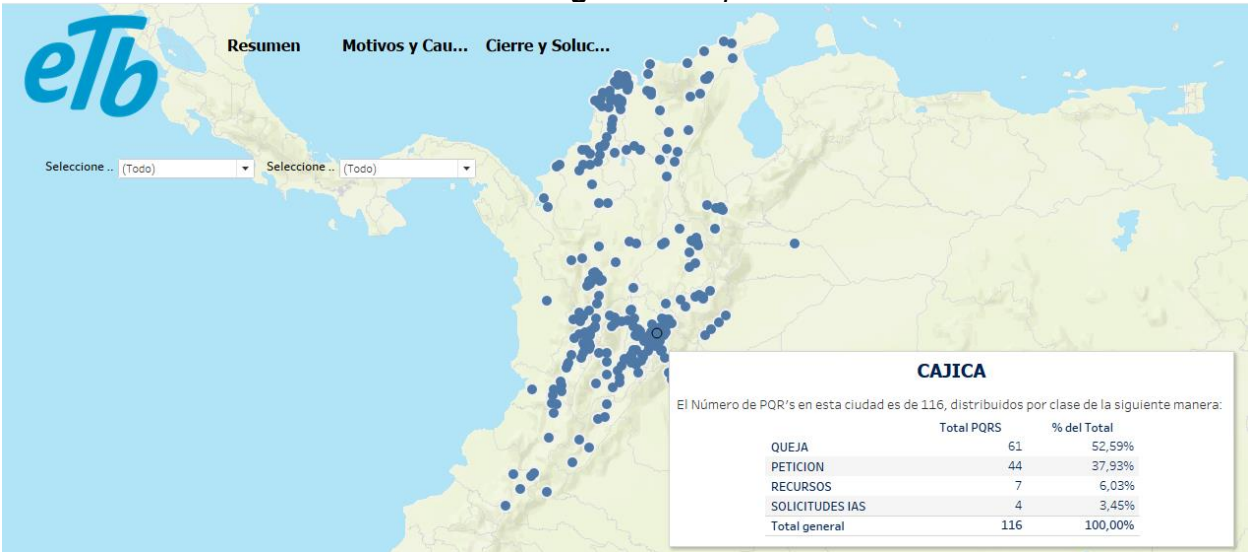
Fuente: Elaboración propia

d. Mapa

En esta sección, se tiene una visualización de mapa para identificar gráficamente la ubicación de los abonados que generaron las PQR's, basado en los parámetros de departamento y ciudad, y así lograr identificar las ubicaciones en las que, se registran más solicitudes de PQR's y de esta manera, conocer las causas comunes de estas solicitudes, con el objetivo de gestionar soluciones a posibles fallos de la infraestructura del servicio en dichas ubicaciones.

La información que se logra visualizar por ciudad corresponde al total de PQR's, tipo de solicitudes y el porcentaje de los tipos de solicitudes. En la figura 53 se observa las PQR's en mapa.

Figura 53 Mapa



Fuente: Elaboración propia

También se pueden filtrar por el departamento y la ciudad de ubicación del usuario que generó la PQR's. En la figura 54 se observa la opción de filtros.

Figura 54 Filtro mapa



Fuente: Elaboración propia

## 7. Conclusiones

En las visualizaciones presentadas en la sección anterior del documento, se logra identificar que el método para la retención de clientes de telefonía móvil pospago aporta una gran cantidad de información procesada, la cual puede ser analizada por el área de gestión de clientes, ellos con el propósito final de gestionar la retención de estos, así como, también coadyuvar a mejoría del servicio, para que los abonados perciban un servicio con buena calidad.

Como punto inicial en la ejecución del proyecto, se tuvo el análisis de la data, con el fin de obtener un prospecto de la información a manipular, y a partir de allí se definieron variables importantes para el método.

A partir de dichas variables se construyó un modelo de datos, y posteriormente se desarrollaron las fases de transformación a través de la construcción de una ETL, continuando con otra fase de almacenamiento de la data procesada en una BD no SQL, y como fase final se construyó un tablero de control, en la cual, se obtiene la información gráficamente por las variables o agrupamiento de estas, sumado a qué, en virtud del análisis que realicen los usuarios del tablero de control, se puede tomar decisiones relacionadas con la retención de clientes y contribuir así con la solución de los motivos por los cuales los usuarios generan una gran cantidad de PQR's.

Con el desarrollo de cada una de las fases del proyecto, se logra cumplir los objetivos específicos propuestos para el método y lo más importante, es que se obtiene un tablero de control el cual va a ser de gran utilidad en la empresa, por cuanto favorece y contribuye ampliamente al análisis y la toma de decisiones con respecto a la retención de clientes.



También se logra fomentar a la organización en la implementación de nuevos tableros de control con herramientas diseñadas para tal fin, esto debido a que en la actualidad en su mayoría la información no es analizada, porque esta se tiene en archivos de Excel o no se cuenta con dicha información de manera gráfica la cual es mucha más fácil de analizar por parte de los empleados.

Desde el aspecto económico, la organización logra una reducción de costos con respecto a la facturación de los abonados, debido a que se puede evitar el cambio de operador por parte de los clientes y la disminución de generación PQR's por parte de los usuarios, cuya solución de estas en algunos casos conlleva a la reducción del cobro en la facturación.

Adicionalmente a lo ya mencionado, se logra concluir que una de las falencias identificadas en el desarrollo del proyecto, es que la data suministrada por la organización no era de suficiente calidad, debido a que se tenían campos sin información, variables numéricas con datos alfanuméricos, entre otros casos.

## **8. Trabajo futuro**

Con el método desarrollado para la retención de clientes, se dio inicio a dicha gestión de usuarios en la compañía, y esta problemática fue abordada de una manera superficial, por tal razón, para que el método desarrollado continúe generando valor agregado a la compañía, se deben establecer controles de mejora continua para potenciarlo aún más y así obtener los resultados esperados.

Por tal razón, se propone a futuro, integrar más fuentes de información en donde adicionalmente a la información de las PQR's, se pueda obtener data de las redes sociales, para identificar la apreciación y percepción de los abonados.

De igual forma, se tiene como deber la implementación herramientas de machine learning, las cuales contribuyen a mejorar el procesamiento de dicha información, para identificar de una manera más precisa los usuarios que pueden ser potenciales al cambio de operador.

Todo lo anterior debe estar complementado con la implementación de una Bodega de datos, y de un proyecto que permita tener datos de calidad.

## 9. Referencias bibliográficas

- Álvarez Cabal, J. V., Rodríguez Montequín, M. M., Mesa Fernández, J. M., & González Valdés, A. (2003). METODOLOGÍAS PARA LA REALIZACIÓN DE PROYECTOS DE DATA MINING. *VII Congreso Internacional de Ingeniería de Proyectos*, (págs. 257-265). Pamplona.
- Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., & Huang, K. (2017). Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing*, 237, 242-254.
- Arnejo Calviño, H. A. (2017). *Métodos para la mejora de predicciones en clases desbalanceadas en el estudio de bajas de clientes (CHURN)*. [Tesis máster, Universidad de Santiago de Compostela - Universidade da Coruña - Universidad de Vigo]. Repositorio Universidad Santiago de Compostela. Obtenido de [http://eamo.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto\\_1469.pdf](http://eamo.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_1469.pdf)
- Barrientos, F., & Ríos, S. A. (2013). Aplicación de Minería de Datos para Predecir Fuga de Clientes en la Industria de las Telecomunicaciones. *Revista Ingeniería de Sistemas*, XXVII, 36.
- Barrientos, F., & Ríos, S. A. (2013). Aplicación de Minería de Datos para Predecir Fuga de Clientes en la Industria de las Telecomunicaciones. *Revista Ingeniería de Sistemas*, 27, 73-107. Obtenido de <http://www.dii.uchile.cl/~ris/RIS2013/rios.pdf>
- Beltrán Martínez, B. (s.f.). *MINERÍA DE DATOS*. Benemérita Universidad Autónoma de Puebla. Puebla: s.e. Obtenido de <https://www.cs.buap.mx/~bbeltran/NotasMD.pdf>

Brândușoiu, I., Todorean, G., & Beleiu, H. (2016). Methods for churn prediction in the pre-paid mobile telecommunications industry. *2016 International conference on communications (COMM)* , 97-100.

Contreras Morales, E. F., Ferreira Correa, F. M., & Valle, M. A. (2017). Diseño de un modelo predictivo de fuga de clientes utilizando árboles de decisión. *Revista Ingeniería Industrial*, 16(1), 7-23. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=6219421>

CRC, C. d. (24 de 02 de 2017). *Comisión de Regulación de Colombia*. Obtenido de [https://colombiatic.mintic.gov.co/679/articles-62266\\_doc\\_norma.pdf](https://colombiatic.mintic.gov.co/679/articles-62266_doc_norma.pdf)

Domínguez, A., & Hermo, S. (2007). *Métricas del marketing*. Madrid: ESIC. Obtenido de <https://docplayer.es/842849-Metricas-del-marketing.html>

Echeverri Giraldo, A. F. (2019). *Modelo predictivo de Churn de clientes para el negocio de Telecomunicaciones [Tesis pregado , Universidad de Antioquia ]*. Repositorio bibliotecadigital. Obtenido de <http://bibliotecadigital.udea.edu.co/handle/10495/15142>

*EcuRed*. (24 de 06 de 2021). Obtenido de EcuRed: [https://www.ecured.cu/Microsoft\\_Excel](https://www.ecured.cu/Microsoft_Excel)

Escobar, H. E., Alcivar, M., & Puris, A. (2016). Aplicaciones de Minería de Datos en Marketing. *Publicando*, 3(8), 503-512. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=5833425>

Gironés Roig, J., Casas Roma, J., Minguillón Alfonso, J., & Caihuelas Quiles, R. (2017). *Minería de datos: Modelos y algoritmos* (1 ed.). Barcelona: UOC.

Gutierrez González, D. (2019). *Técnicas de machine learning en el análisis del CHURN RATE [Trabajo máster, Universidad de Cantabria]*. Repositorio abierto de la Universidad de Cantabria. Obtenido de <https://repositorio.unican.es/xmlui/handle/10902/19075>

Gutiérrez, J. A., & Molina, B. (2015). Identificación de técnicas de minería de datos para apoyar la toma de decisiones en la solución de problemas empresariales. *Revista Ontare*, 33-51. doi:10.21158/23823399.v3.n2.2015.1440

Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J., & Ziegler, H. (2008). Visual Analytics: Scope and Challenges. En *Visual Data Mining: Theory, Techniques and Tools for Visual Analytics* (págs. 76-90). Springer. Obtenido de [https://link.springer.com/chapter/10.1007%2F978-3-540-71080-6\\_6](https://link.springer.com/chapter/10.1007%2F978-3-540-71080-6_6)

Lozano Núñez, D. (2015). *Modelos predictivos del churn – abandono de clientes – para operadores de telecomunicaciones. [Tesis máster, Universidad de Santiago de Compostela - Universidade da Coruña - Universidad de Vigo]*. Repositorio Universidad Santiago de Compostela. Obtenido de [http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto\\_1269.pdf](http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_1269.pdf)

Marketing directo. (28 de 04 de 2020). *Marketing Directo*. Obtenido de Marketing Directo: <https://www.marketingdirecto.com/marketing-general/marketing/porque-es-mas-barato-retener-un-cliente-que-conseguir-uno-nuevo>

MINTIC. (11 de 02 de 2020). *Datos Abiertos*. Obtenido de Datos Abiertos: <https://www.datos.gov.co/Ciencia-Tecnolog-a-e-Innovaci-n/Telefon-a-M-vil-abonados-por-categor-a/nrst-mwx4>

- Moine, J. M., Haedo, A. S., & Gordillo, S. E. (2011). Estudio comparativo de metodologías para minería de datos. *XIII Workshop de Investigadores en Ciencias de la Computación* (pág. 4). Buenos Aires: N/A. Obtenido de <http://sedici.unlp.edu.ar/handle/10915/20034>
- Moine, J. M., Haedo, A. S., & Gordillo, S. E. (2011). Estudio comparativo de metodologías para minería de datos. *XIII Workshop de Investigadores en Ciencias de la Computación*. Obtenido de <http://sedici.unlp.edu.ar/handle/10915/20034>
- MongoDB, Inc. (2021). *MongoDB*. Obtenido de <https://www.mongodb.com/es/what-is-mongodb>
- Pérez Villanueva, P. A. (2014). *Modelo de predicción de fuga de clientes de telefonía móvil post pago*. [Tesis pregrado, Universidad de Chile]. Repositorio académico Universidad de Chile. Obtenido de <http://repositorio.uchile.cl/handle/2250/115942>
- Pérez, S. (16 de 07 de 2018). *Un modelo de gestión empresarial: La responsabilidad social corporativa de las empresas del IBEX 35, actitudes y conductas de sus empleados y clientes*. Obtenido de <http://e-spacio.uned.es/fez/view/tesisuned:ED-Pg-PsiSal-Saperez>
- Pesantez Chuqui, M. C. (2019). *Comparativa de técnicas Machine Learning sobre comportamiento de pago de clientes con cuentas por cobrar* [Tesis máster, Universidad Internacional de la Rioja]. Repositorio digital de la UNIR. Obtenido de <https://reunir.unir.net/handle/123456789/9734>
- Retos en Supply Chain*. (9 de 04 de 2020). Obtenido de <https://retos-operaciones-logistica.eae.es/que-es-kettle-pentaho-y-cual-es-su-uso-en-la-empresa/>
- Riquelme, J., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias. *Inteligencia Artificial. Revista Iberoamericana*, 10(29), 11-18. Recuperado el 18 de 05 de 2021, de <https://www.redalyc.org/pdf/925/92502902.pdf>

SAS Institute. (30 de Agosto de 2017). *SAS® Help Center*. Obtenido de SAS® Help Center:

<https://documentation.sas.com/doc/en/emref/14.3/n061bzurmej4j3n1jnjb8bbj1a2.htm>

Tableau Software, LLC. (2003). *Tableau*. Obtenido de [https://www.tableau.com/es-es/why-](https://www.tableau.com/es-es/why-tableau/what-is-tableau)

[tableau/what-is-tableau](https://www.tableau.com/es-es/why-tableau/what-is-tableau)

Thomas, J., & Cook, K. (2005). *Illuminating the Path: Research and Development Agenda for Visual Analytics*. IEEE-Press.

Vargas, M. F., & Pineda, W. (2019). *Modelo Logístico con Datos Funcionales Aplicado al Churn en un Operador Movil. [Tesis pregrado, Universidad Santo Tomás]*. Centro de recursos para el aprendizaje y la investigación. Obtenido de <https://repository.usta.edu.co/handle/11634/21052>

Witten, I. H., Eibe, F., & Mark, A. (2011). *Data mining: practical machine learning tools and techniques*. Morgan Kaufmann.

Zhu, B., Baesens, B., & Broucke, S. (2017). An empirical comparison of techniques for the class imbalance problem in churn prediction. *Information sciences*, 408, 84-99.