Predicción de la tasa de deserción de Clientes de Claro Colombia

Taller 2 Grupal

María Fernanda Izquierdo Aparicio & Juan Pablo Cuellar Solano

# BUSINESS UNDERSTANDING

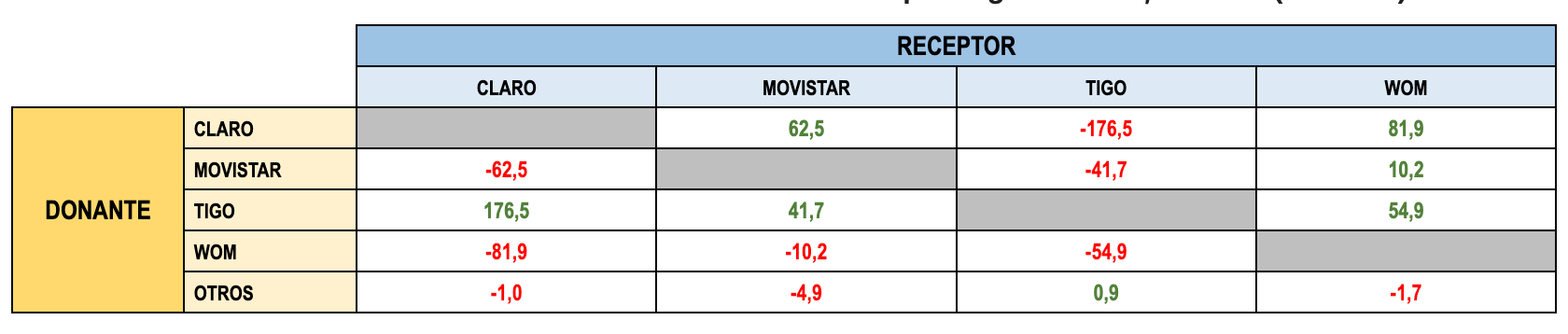
## Background

La telefonía móvil en Colombia aumentó en el segundo trimestre en 1,40% versus el trimestre anterior, con una base de abonados a telefonía móvil de más de 82 millones. De acuerdo con los datos presentados por la Comisión de Regulación de Comunicaciones (CRC), en Colombia la penetración de la telefonía para el segundo trimestre del 2023 fue de 159,57% (Comisión de Regulación de Comunicaciones, 2023a), indicador calculado como el número de abonados sobre el total de la población. El mismo reporte presentado por la CRC muestra la conformación del mercado de telefonía móvil, la cual está liderada por Claro (Comcel S.A), seguido por Movistar, Tigo y WOM.

La información telefónica presentada por la CRC a través de su portal postdata.gov.co, presenta el detalle de los principales rubros que reportan los proveedores de servicios móviles en el país. Éstos muestran cómo Claro por medio de Comcel S.A., se posiciona como el líder en ganancias por su prestación de servicios de voz móvil, tanto en planes de telefonía prepago como pospago, con ingresos que superan los 93 mil millones de pesos y una participación del 46% del *market share* del mercado de telecomunicaciones. Al final del segundo trimestre del año 2023, el reporte muestra que Claro tiene el 66.18% del *market share* del mercado prepago y el 47.77% del *market share* del mercado pospago (Comisión de Regulación de Comunicaciones, 2023a).

Dadas las características del mercado de telecomunicaciones, el cual tiende a ser muy competitivo, y la medida de Portabilidad Numérica Móvil implementada en 2011[[1]](#footnote-2),se ha registrado un total de 45,7 millones de operaciones (ingresos y abandonos de clientes con diversos operadores) desde la fecha de implementación de la medida hasta Junio del 2023 (Comisión de Regulación de Comunicaciones, 2023b). Aun así, en el reporte *Data Flash* de la CRC, se reporta que en el segundo trimestre del 2023, las operaciones de portación llegaron a los 2.06 millones disminuyendo en un 0.85% frente al trimestre anterior (Comisión de Regulación de Comunicaciones, 2023c).

Teniendo en cuenta las características de dicho mercado, Claro, líder en Colombia en servicios de telefonía móvil con una participación de mercado del 46% (Ministerio de Tecnologías de la Información y las Comunicaciones de Colombia, 2023) tiene interés en poder mantenerse en dicha posición, por lo cual es de gran relevancia mantener o disminuir la cantidad de cancelación de sus clientes (Churn Rate). Según el reporte Data Flash 2023-006 - Portabilidad Numérica Móvil de la CRC (2023), Claro realizó 692.2 mil operaciones como donante (perdió líneas), y recibió 726.3mil operaciones (ganó líneas), dejándolo con un saldo neto positivo de 34.1 mil portaciones. En las operaciones netas de Claro, este es el principal donante de 2 de sus tres principales competidores (Movistar y WOM).



**Tabla 1.** **Resultados netos de Portabilidad Numérica Móvil por origen-destino, en miles (2023-1T)** tomada de Data Flash 2023-006 - Portabilidad Numérica Móvil. Creada por la Comisión de Regulación de Comunicaciones (Mayo 29, 2023).

Teniendo en cuenta el contexto del mercado de telefonía móvil en Colombia se realiza un análisis PESTLE.

***Factores económicos***

Entre los factores económicos que tiene el sector de telecomunicaciones en Colombia, se encuentran la inflación, el desempleo y el crecimiento económico. En septiembre el cambio la inflación fue de 10,99 % según el informe del DANE, disminuyendo frente al mismo mes del año pasado, y la categoría de información y comunicación fue la de menor variación mensual con un 0,07 %. En esta categoría, la subclase de servicios de comunicación fija y móvil no tuvo variación, mostrando que hasta el momento es un mercado que se mantiene constante ante las fluctuaciones de la economía (Suárez, 2023). Además, el desempleo y crecimiento económico, afectan de igual forma a las empresas de telecomunicaciones, dado que si el desempleo es muy elevado, este puede hacer que se pierdan clientes a estos no tener con que costear los servicios, y por su parte, el crecimiento y desarrollo de la economía tienen una relación estrecha con las comunicaciones móviles, como lo refleja el estudio realizado por las Naciones Unidas, ya que si aumenta el número de personas que usan el servicio, el efecto de redes, infraestructura y desarrollo de comunicaciones también lo hace, trayendo mejoras y efectos positivos tanto en la economía como en las compañías de telefonía móvil (Fernández-Ardèvol, 2023)

***Factores Políticos***

Las regulaciones sobre telefonía móvil en Colombia se pueden ver en la reglamentación de la Ley 37 de 1993, "por la cual se regula la prestación del servicio de telefonía móvil celular, la celebración de contratos de sociedad y de asociación en el ámbito de las telecomunicaciones y se dictan otras disposiciones”, tal como lo muestra la compilación jurídica del MINTIC (Ministerio de Tecnologías de la Información y Comunicaciones, 2019). Así, decretos y medidas implementadas, como la de portación mencionada, son importantes para el manejo de las telefonías móviles y cómo funciona el mercado de estas en Colombia, reguladas por el MINTIC, la CRC y la Superintendencia de Industria y Comercio.

***Factores Sociales***

La demanda de servicios de está influenciada por los cambios en los patrones de consumo y las tendencias culturales. Por telecomunicaciones ejemplo, el aumento en el uso de las redes sociales y las aplicaciones de mensajería instantánea ha llevado a un aumento en la demanda de planes de datos móviles (Perez Becerra & Yamal Matos, 2011). Además, la adopción de tecnologías de telecomunicaciones, como el acceso a Internet a través de dispositivos móviles, ha tenido un impacto significativo en la vida diaria de las personas en desarrollados y en crecimiento. Los usuarios pueden utilizar sus dispositivos para acceder a información, comunicarse, realizar transacciones financieras y participar en actividades educativas y de ocio (Ciuró et al., 2020).

Ahora, la importancia en el cubrimiento de telefonía móvil e internet en zonas rurales, junto con el crecimiento de la población, el aumento en la urbanización y la búsqueda de una disminución de la brecha digital, son parte importante de las acciones que deben ejecutar las empresas de telefonía móvil. por un lado para aumentar su cobertura, como con el fin de promover el desarrollo de la sociedad, que de acuerdo con lo encontrado en el estudio realizado por las naciones unidas, la difusión de las comunicaciones móviles favorece en la mejora de la situación de pobreza, ya que por medio de la telefonía móvil, se pueden fortalecer las relaciones sociales, las cuales a su vez pueden fortalecer las redes económicas, incorporándose en la producción rural y actividades comerciales como un instrumento de comunicación (Fernández-Ardèvol, 2023)

***Factores Tecnológicos***

El desarrollo de la tecnología 5G representa una oportunidad y un potencial en la revolución del sector de telecomunicaciones, coincidiendo con la posibilidad de que dado un ecosistema digital propicio como el que tiene Colombia actualmente, el país se encuentra preparado para recibir esta tecnología., la cual a su vez traerá avances tanto en infraestructura como en conectividad del país (Portafolio, 2023b). Como ésta, las nuevas tendencias tecnológicas, pueden cambiar la forma en que se prestan los servicios de telecomunicaciones y afectar la demanda de los consumidores (Uanataca, 2021).

**Factores Legales**

Las regulaciones que tienen las empresas de telecomunicaciones son vitales para el ejercicio de dichas empresas, ya reglas buscan proteger al consumidor y promover una competencia justa. Tal como sucede con la normatividad de portabilidad de la Ley 1245 de 2008, por la que la SIC puede sancionar aquellos operadores que incumplan lo pasos para hacer efectivo el trámite de portación (Ávila Forero, 2023). Otro aspecto de importancia legal es el manejo y protección de los datos, que es supervisado a su vez por la SIC, la cual busca el beneficio y protección de los usuarios e invita a los operadores a cumplir con dichos regímenes (SIC, 2023).

**Factores Ecologicos**

Los factores ecológicos por su parte tienen una normatividad para la industria de las telecomunicaciones, teniendo en cuenta la Ley 2294 de 2023 en el *Plan Nacional de Desarrollo-Colombia Potencia Mundial de la Vida*, por medio del cual se menciona el incremento que deben hacer las empresas de telecomunicaciones por el uso del espectro radioeléctrico (Universidad Externado de Colombia, 2023). Igualmente, la dependencia de la telefonía móvil al uso de celulares genera un impacto medioambiental negativo en términos de contaminación, por lo que es muy importante que las empresas tecnológicas y de telecomunicación cambien, y aporten en la disminución de la contaminación, a la vez que pueden jugar un papel importante en la lucha contra el cambio climático empleando tecnología renovable dentro de su infraestructura.

En cuanto al análisis PESTLE, se observa que el sector de telecomunicaciones en Colombia tiene un entorno óptimo para seguir creciente, que aunque presenta desafíos en algunos factores, la estructura y organización del país, favorecen el desarrollo del sector, esto impacta positivamente a Claro, al ser el líder del mercado en Colombia, que a su vez tiene mayor cobertura y red en el país.

A continuación, se revisan 5 Fuerzas de Porter para Claro:

**Rivalidad entre competidores**

El mercado tiene una rivalidad entre competidores alta, donde Movistar, Tigo y WOM, son los principales competidores de Claro, que de acuerdo al informe de EConcept, Claro no domina el mercado, pero si concentra gran parte de él (Portafolio, 2023a). De acuerdo con el mismo informe, el churn del mercado es cercano al 15% trimestral, mientras que la rivalidad entre competidores no sólo es por lucha de precios, sino por calidad del servicio, experiencia de usuario, cobertura y variedad en el servicio.

**Poder de clientes**

La posibilidad de los clientes de cambiar de proveedor hace que el poder de negociación de Claro sea moderado. Varios proveedores con servicios similares hacen que los clientes decidan si el precio o el servicio ofrecido es de su agrado o no, y en caso poder decidir cambiar de telefonía móvil cuando deseen.

**Proveedores:**

Los proveedores tienen un poder de negociación bajo, ya que son los operadores móviles, los que toman la decisión de las tarifas de precios y condiciones que se deben manejar con los equipos y las tarifas de telefonía móvil.

**Productos Sustitutos:**

La amenaza de productos sustitutos es alta, ya que existen distintas alternativas para las personas a la hora de elegir el medio y el canal de comunicación que las personas quieran emplear, como lo es internet o el uso de WhatsApp. Adicionalmente el cambio social generado por la pandemia como lo es el tener la posibilidad de trabajar en casa puede inducir a que las personas busquen reducir el gasto o uso en telefonía móvil al no tener necesidad de salir de casa.

**Nuevos Competidores:**

Dado las características del mercado, donde existe una alta competencia entre los diferentes prestadores de servicios móviles, el alto costo de infraestructura y las leyes que deben cumplir de acuerdo con las normativas colombiana, hace que la entrada de nuevos competidores sea baja, ya que aquellos que desean ingresar, deben contar con el suficiente presupuesto para poder costear su ingreso a este mercado.

A continuación, se presenta un análisis DOFA de la entidad:

**Fortalezas**

La principal fortaleza de Claro se tiene en su posicionamiento de marca donde tiene una porción del mercado importante, una gran cobertura telefónica y una infraestructura que le permite cubrir la mayor cantidad del territorio colombiano. Su posicionamiento no solo se evidencia en la cantidad de clientes abonados, sino en su su posición en el Top Of Mind, siendo la principal marca de telecomunicaciones en recordación en Colombia, con un 50% (Semana, 2023).

**Debilidades**

La calidad del servicio postventa de Claro es una de sus principales debilidades, ya que su fuerza de trabajo y venta tiene muchas fallas y es así donde las quejas recibidas por los clientes no se solucionan adecuadamente, haciendo que los clientes busquen cambiar y migrar de servicio, y aquí donde se pueden entrenar y entrenar con atención constante, para mantener una buena experiencia. Igualmente, la calidad del servicio móvil y la falencia en la señal de comunicación, por lo que el soporte debe estar siempre disponible

**Oportunidades**

Teniendo en cuenta el desarrollo tecnológico de los últimos años, se puede apalancar del uso de nuevas tecnologías para ampliar su red móvil y llegar a nuevos clientes en aquellas zonas rurales y con poco acceso a las telecomunicaciones, como lo es el uso de tecnología 5G satelital (López, 2022)

**Amenazas**

La alta competencia es una amenaza siempre latente, teniendo en cuenta que competidores como Movistar, Tigo y WOM, están en constante cambio, ofreciendo nuevos planes y servicios, lo que obliga a Claro a mantenerse a la actualizado, desarrollo de alianzas, y diferentes planes de servicio para captar y retener su pool de clientes; de igual forma, es importante que con el cambio tecnológico que se enfrena, Claro realice inversiones que le permitan mantenerse actualizado y que sus equipos, instalaciones e infraestructura no queden obsoletos.

## Business goal

El objetivo de negocio es incrementar la tasa de retención de los clientes (*Retention Rate*) a partir del pronóstico de deserción de clientes basado en la información de la compañía.

## Business success criteria

Incrementar la tasa de retención de clientes entre 5 y 10 puntos porcentuales en un semestre.

## Determine Data mining goals

Desarrollar un modelo de regresión logística adecuado que permita pronosticar la deserción de clientes de la compañía de telefonía móvil que cumpla con los estándares de capacidad predictiva.

## Data mining success criteria

El criterio de éxito para la regresión logística es que tenga la *mayor* área (AUC) bajo la curva ROC (Receiver Operating Characteristics curve).

# DATA UNDERSTANDING

## Describe data

Se tiene una base de datos con información de 8.243 clientes con contratos de suscripción (pospago) de la compañía de telefonía móvil, la cual presenta información sobre el “consumo (de) los últimos seis meses de (cada) cliente con el operador” (Beltrán et al, 2023) con corte a primero de enero de 2019. Esta información corresponde tanto a clientes que han entrado en churn como clientes que siguen con el operador, razón por la cual los datos de aquellos clientes que han entrado en churn, “corresponden a los últimos seis meses con el operador, y el churn pudo haber ocurrido en cualquier momento dentro de la ventana de tiempo del segundo semestre de 2018” (Beltrán et al, 2023). Teniendo este contexto, el dataset muestra información relacionada con cada cliente en específico (fecha de nacimiento del cliente, si el cliente es hombre, mujer o empresa); tipo y cantidad de facturación (facturación online, tenencia de plan premium de datos, facturación total); consumo (minutos consumidos), antigüedad del equipo en el contrato, antigüedad del contrato y días de mora acumulados en caso de que hubiere.

## Explore data

Se realizó tanto análisis univariado, como multivariado sobre la base de datos entera. De éste análisis se debe mencionar que la data contiene 11 variables en total, 2 variables tipo fecha, 1 variable categórica, 3 variables binarias, y 5 variables escalares. Solamente se analizan 10 variables escalares ya que la variable *id* es un identificador único para cada cliente.

A partir de un análisis univariado de cada una de las variables se puede hacer el siguiente resumen:

* **Fecha de nacimiento:** Esta variable indica la fecha de nacimiento de cada cliente. Para la muestra se identifica que hay clientes nacidos entre los años de 1940 hasta 1992. Revisando la distribución de clientes por año de nacimiento se identifica que el rango comprendido entre 1940 a 1948 tiene el 1.35% del total de clientes; el rango entre 1949 a 1954 tiene el 19.47%; el rango entre 1955 a 1960 comprende el 49.52% de la totalidad de clientes;el rango entre 1961 a 1967 comprende un 27.38% del total de clientes y el rango entre 1968 y 1992 comprende el restante 2,28% de la totalidad de clientes.
* **Tipo cliente:** Esta variable categórica indica si el cliente es hombre, mujer o es un cliente empresarial. Se tiene que de los 8243 clientes, 3740 (45.37%) son hombres; 3605 (43.73%) son mujeres y 898 (10.89%) son clientes empresariales.
* **Factura online:** Es una variable binaria que indica si el cliente recibe únicamente su factura online. En el dataset hay 4032 clientes (48.91%) que reciben únicamente su factura online mientras que los restante 4211 (51.09%) reciben sus facturas de otras formas aparte de una factura online.
* **Antigüedad equipo:** Variable escalar que indica el número de meses de antigüedad del equipo. Esta variable muestra que los clientes pospago tiene equipos de un mes hasta equipos de 48 meses. El promedio de los clientes tienen equipos con 24.27 meses de antigüedad. Esta variable tiene una desviación estándar de 15.15 meses y un coeficiente de variación de 0.62, métricas que muestran que la variabilidad de la antigüedad de los equipos de los clientes es alta en relación con la media. Además, esta distribución tiene un coeficiente de asimetría de 0.2 y una curtosis de -1.63, lo que indica que tiene una asimetría ligeramente positiva pero tiene colas pronunciadas, lo que implica la presencia de varios valores en los extremos y pocos valores alrededor de la media.
* **Plan de datos:** Esta variable binaria indica si el cliente tiene plan premium de datos. En la base de datos hay 6933 clientes (84.11%) que tienen plan premium de datos y 1310 (15.89%) que no tienen plan premium de datos.
* **Facturación:** Es una variable escalar que indica la suma total de facturación de los últimos seis meses de cada cliente. La variable tiene 23 registros negativos y 8220 registros positivos y su rango se encuentra entre - $112,588 hasta $626,111. El promedio de facturación de los clientes es $267,765, la distribución tiene una desviación estándar de $95,596.02 y un coeficiente de variación de 0.36 , medidas que muestra que la variabilidad de la nómina es relativamente baja en relación con su media. Esta distribución también tiene una asimetría ligeramente pues su coeficiente de asimetría es de -0.2, lo que indica que es una distribución muy parecida a la distribución normal y tiene pocos valores en los extremos. Además, tiene una curtosis de 0.1 medida que evidencia una alta concentración de valores alrededor de la media y tanto colas derecha como izquierda relativamente igual de pesadas en comparación con una distribución normal.
* **Mora:** Es una variable escalar que indica el número de días de mora acumulados por los clientes. Esta variable tiene 3169 clientes con mora igual a cero, lo que indica que el 38.45% de los clientes en la base de datos no han caído en mora. Los restantes 5074 clientes que han caído en mora presentan un rango de días de mora entre 1 y 118, con promedio de 53.5 días de mora. La desviación estándar de la distribución completa es de 21.03 días de mora con un coeficiente de variación de 1.16, lo que indica una variabilidad alta de mora. También tiene un coeficiente de asimetría de 0.91 y una curtosis de 0.10, medidas que indican que la distribución tiene una asimetría positiva con una cola izquierda pronunciada, reflejan que hay una mayor cantidad de clientes con mora cero o mora de pocos días y que la distribución es relativamente plana con una menor concentración de valores entorno a la media en comparación de una distribución norma
* **Fecha inicio contrato:** Esta variable indica la fecha de inicio del primer contrato de plan pospago con la compañíade cada cliente. Para la muestra se identifica que los clientes iniciaron su contrato pospago entre el 2011 y el 2019. Revisando la distribución, se identifica que en el periodo comprendido entre el 2011 al 2014, 280 clientes (3.4%) iniciaron su primer contrato pospago. En el año 2015 fueron 914 clientes (11.1%), en el 2016 fueron 2515 clientes (30.5%), en el 2017 fueron 3736 clientes (45.3%), en el 2018 fueron 797 clientes (9.7%) y en el 2019 fue 1 cliente (0.01%) que respectivamente iniciaron su primer contrato pospago con la compañía.
* **Minutos:** Esta es una variable escalar que indica el total de minutos consumidos en los últimos seis meses por cada cliente. Se identifica que los clientes consumieron entre 100 y 14,103 minutos, siendo 1,183.82 minutos consumidos en promedio por cliente. Esta distribución tiene una desviación estándar de 932.84 minutos y un coeficiente de variación de 0.79. También tiene un coeficiente de asimetría de 6.01 y una curtosis de 54.81. En general, estas medidas indican que la distribución que minutos consumidos tiene es altamente sesgada hacia la derecha (asimetría positiva) con valores atípicos en el extremo superior, tiene además colas pesadas y picos agudos, lo que sugiere una alta concentración de valores alrededor de la media, pero con algunos valores extremadamente altos, por lo que no es similar a una distribución normal.
* **Resultado:** Esta es una variable binaria que indica si hubo o no deserción por parte del cliente en los últimos 6 meses. En la base de datos se tiene que 6882 clientes (83.49%) no desertaron a corte del primero de enero de 2019, pero hay 1361 clientes (16.51%) que si desertaron en algún momento del segundo semestre del año de 2018.

Del análisis multivariado cabe mencionar que:

* Existen correlaciones positivas importantes (coeficientes de correlación mayores que 0.4) entre: 1)el plan de datos y la antigüedad del equipo y 2) la facturación con el consumo de minutos.
* Hay correlaciones negativas importantes (coeficientes de correlación menores que -0,4) entre: 1) el la antigüedad del equipo y la factura online y 2) el plan de datos y la factura online.

## Verify Data Quality

Al revisar la data no se detectaron valores nulos ni información faltante tanto para observaciones como para variables, por lo que se puede concluir que la calidad de los datos es sólida. La revisión de las variables escalares indica que hay 23 observaciones con facturación negativa comprendidos entre -112588 y -202, los cuales corresponden a clientes que no han desertado, estos valores tienen sentido de negocio, ya que en algunos casos las personas tienen saldos a favor en su factura, que son deudas de la compañía con el usuario (Comunidad Claro, 2016). También hay 21 clientes con consumo de minutos total mayor a 10,000 minutos. Finalmente es necesario hacer hincapié en que la variable “Resultado”, la cual será usada como variable a predecir, esta desbalanceada ya que el 83.49% de los clientes no desertaron a corte del primero de enero de 2019, pero hay el 16.51% si desertaron.

# DATA PREPARATION

## Clean and Construct data

Ya que la revisión de la data evidencia que no hay que no hay problemas con la calidad de datos, se decide que se trabajará con la totalidad de datos entregada. Aun así, debido a que se usará con un algoritmo de regresión logística, hemos decidido realizar ciertas trasformaciones a las variables con el objetivo de poder usarlas todas, para obtener mejores *insights* en la predicción. Además, se crean datasets de entrenamiento, validación y evaluación con el fin de entrenar el modelo y validar la calidad de los resultados del mismo. A continuación, se detallan las transformaciones realizadas y la creación de datasets para entrenamiento, validación y evaluación:

1. ***Transformaciones:***
   * **Fechas:** Teniendo el cuenta que en la data original se tienen las variables “Fecha de Nacimiento” y “Fecha inicio del Contrato”, las cuales están en formato fecha (día, mes, año), se procede a calcular la edad del cliente y la antigüedad del contrato respectivamente. Esto se realiza para tener variables escalares que trabajan mejor en el tipo de modelo seleccionado.

|  |  |
| --- | --- |
| **Nueva Variable** | **Cálculo** |
| Edad | 2019 menos el año correspondiente a la Fecha de Nacimiento del cliente |
| Antigüedad\_contrato\_años | Diferencia en días entre el 1 de Enero de 2019 y la fecha de inicio del contrato. El resultado es dividido en 365 para tener los años del contrato. |

* + **Binarias para tipo de cliente:** Teniendo en cuenta que hay tres categorías para tipo de cliente, se crean tres variables binarias, una para identificar si el cliente es un hombre, una segunda para representar si el cliente es una mujer y una tercera para identificar si el cliente es empresarial. Esto para evitar suposiciones de orden, una interpretación más clara al evaluar el modelo y evitar multicolinealidad.

|  |  |
| --- | --- |
| **Nueva Variable** | **Cálculo** |
| Hombres | data$hombres=ifelse(data$tipo\_cliente==1,1,0) |
| Mujeres | data$mujeres=ifelse(data$tipo\_cliente==2,1,0) |
| Empresas | data$empresas=ifelse(data$tipo\_cliente==3,1,0) |

* + **Transformaciones lineales:** Para poder capturar varias relaciones que no pueden capturarse por una relación lineal, se crean las siguientes variables para evaluarlas y definir si su uso en el modelo mejora su calidad de la predicción.

|  |  |  |
| --- | --- | --- |
| **Tipo de transformación**  **(se presenta el sufijo utilizado para identificar la nueva variable)** | **Variables a las que se le aplica la transformación** | **Cálculo**  **(X representa la variable)** |
| Cuadrática  ( \_sqr ) | * Antigüedad del Equipo * Facturación * Mora * Minutos * Edad * Antigüedad\_contrato\_años | data$X^2 |
| Logaritmo Natural  ( \_ln ) | * Antigüedad del Equipo * Minutos * Edad | Log(data$X) |
| Inversa  ( \_inversa ) | * Antigüedad del Equipo * Facturación * Minutos * Edad * Antigüedad\_contrato\_años | 1/(data$X) |

* + **Relaciones:** También se crean relaciones entre la variable de mora con las variables de minutos, facturación y edad; y entre facturación con minutos, con el fin de identificar posibles interacciones no lineales entre la variable de mora y variables de consumo/demográficas, así como la variable de facturación y minutos.

|  |  |
| --- | --- |
| **Nueva Variable** | **Cálculo** |
| mora\_minutos | data$mora/data$minutos |
| mora\_facturación | data$mora/data$facturación |
| mora\_edad | data$mora/data$Edad |
| facturación\_minutos | data$facturación/ data$minutos |

* + **Grupos de clientes por minutos consumidos**: Se crea una variable jerárquica que representa 4 grados de contacto que puede tener un cliente con otras personas o entidades usando como bace la variable de minutos. Según la lógica que presentan Meza Rodríguez y Chue Gallardo (2020), se puede clasificar a los clientes dependiendo del uso de minutos que tienen y se agrupan en cuatro grupos (identificados por cuartiles: comunidad baja; comunidad media; comunidad media alta y comunidad alta) dependiendo de la ubicación en la que están dentro de la distribución de la variable minutos.

|  |
| --- |
| **Cálculo** |
| cuartiles quantile(data$minutos, probs = c(0.25, 0.5, 0.75))    Q1 <- quartiles[1]  Q2 <- quartiles[2]  Q3 <- quartiles[3]    data$comunidad\_baja = ifelse(data$minutos<= Q1,1,0)  data$comunidad\_media = ifelse(data$minutos>Q1 & data$minutos<= Q2,1,0)  data$comunidad\_media\_alta = ifelse(data$minutos>Q2 & data$minutos<= Q3,1,0)  data$comunidad\_alta = ifelse(data$minutos> Q3,1,0) |

* + **Marginales:** Según Manďák y Hančlová (2019), el efecto del consumo de minutos y de la duración del contrato sobre la probabilidad de abandono no es el mismo dependiendo del año de nacimiento del cliente. En su estudio “” probaron el efecto del consumo de minutos y la duración del contrato es mayor para los clientes mayores. Con esto se decide replicar la lógica de Manďák y Hančlová para este estudio con la creación de las siguientes variables:

|  |  |
| --- | --- |
| **Nueva Variable** | **Cálculo** |
| ME\_Minutos\_BirthYear | (data$Edad + 2019)\*data$minutos |
| ME\_duracion\_BirthYear | (data$Edad + 2018)\*data$Antigüedad\_contrato\_años |

1. ***Eliminación de variables***

Una vez realizadas las transformaciones mencionadas, se procede a eliminar las variables que no serán parte del análisis y la evaluación de los modelos.

|  |  |
| --- | --- |
| **Variables eliminadas** | **Justificación** |
| ID | Es un identificador numérico único para cada cliente que si se llega a utilizar puede distorsionar el análisis. |
| grupo\_de\_cliente | Es una variable categórica de la cual ya se crearon variables binarias para cada categoría por lo cual si se conserva puede generar correlaciones. |
| Fecha de Nacimiento  Antigüedad del contrato | Son variables con en formato fecha y de las cuales se crearon variables escalares que representan la misma información. Es por esto que se eliminan del dataset final. |

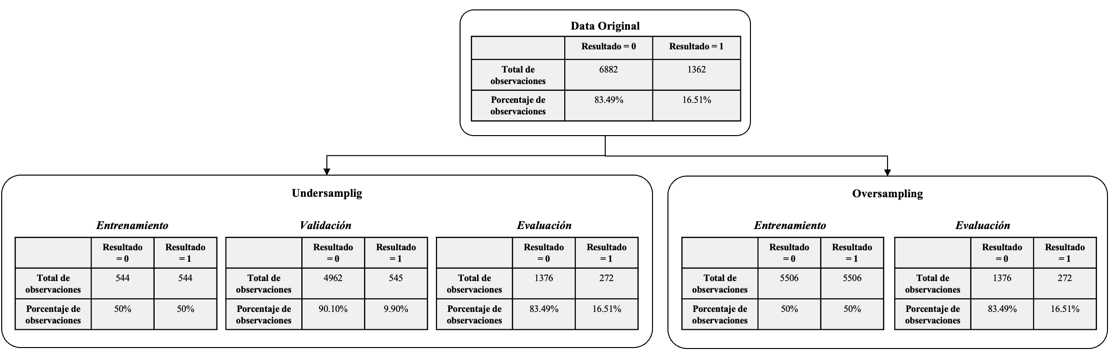
1. ***Creación datasets entrenamiento, validación y evaluación:***

Una vez creadas nuevas variables y eliminado variables que no se van a emplear, se procede a dividir la base de datos en una porción para entrenamiento y otra porción para la evaluación. Dada la cantidad de datos, se decide que la separación será aleatoria, usando un 80% de las observaciones para entrenamiento y el restante 20% para la evaluación del modelo en una primera instancia. Pero, teniendo en cuenta que la variable “Resultado”, la cual será usada como variable a predecir, esta desbalanceada, se realiza un proceso de balanceo para tener un dataset de entrenamiento equilibrado usando los porcentajes para *entrenamiento* y *evaluación* mencionados anteriormente. Se realizan dos tipos de balanceo para evaluar los modelos propuestos.

* + **Undersampling:** En el primer balanceo se reduce la cantidad de observaciones de la clase mayoritaria (Resultado igual a 0), ya que en lugar de utilizar todos las observaciones de la clase mayoritaria correspondientes al 80% para evaluación, se selecciona un subconjunto aleatorio de observaciones de esa clase para que se equipare en tamaño con la clase minoritaria (Resultado igual a 1) que se obtiene de la selección aleatoria del 80% del total de observaciones para evaluación. De este proceso queda un restante de observaciones que se usan para una tercer dataset de validación.
  + **Oversampling:** En el segundo balanceo se aumenta la cantidad de registros de la clase minoritaria (Resultado igual a 1) a partir de la replicación de observaciones de la clase minoritaria hasta que alcance una proporción igual que la clase mayoritaria seleccionada para evaluación (80% del total de observaciones con Resultado igual a 0).

## Dataset description

De acuerdo a lo presentado se tendrán dos conjuntos de datasets. El primer conjunto contine los datasets de evaluación, validación y evaluación resultantes del proceso de *undersampling*, y el segundo conjunto tiene los datasets de evaluación y entrenamiento resultantes del proceso de *oversampling*. Todos los datasets presentados tienen 36 variables, entre escalares y binarias. En el siguiente gráfico se muestra el total de observaciones por clases en cada conjunto y correspondientes datasets.



# MODELING

## Select modeling techniques:

Para este proyecto se elige usar una regresión logística porque este tipo de modelo supervisado permite modelar probabilidades ya que, a diferencia de la regresión lineal, la regresión logística modela la probabilidad de que una observación pertenezca a una categoría específica, lo cual es útil en un problema como el de este proyecto donde se pretende predecir una clasificación binaria. Además, este tipo de regresión maneja tanto variables categóricas como numéricas como predictores. En cuanto a sus supuestos no requiere supuestos de normalidad de los errores y no asume homocedasticidad de los errores, lo que significa que la varianza de los errores no necesita ser constante en todos los niveles de las variables predictoras.

Con el fin obtener las variables independientes que van a ir en las diversas iteraciones de la regresión, y teniendo en cuenta que se realizaron diversas transformaciones las variables originales, se aplicará el método stepwise. Este método, usando la heurística del vecino más cercano, indica qué variables independientes deberán incluirse en el modelo de regresión y cuáles se deberán excluir. Esto es importante porque automatiza el proceso de selección de características, con las cuales se puede mejorar el rendimiento de la regresión y puede mitigar problemas de multicolinealidad que puedan existir en nuestros datasets.

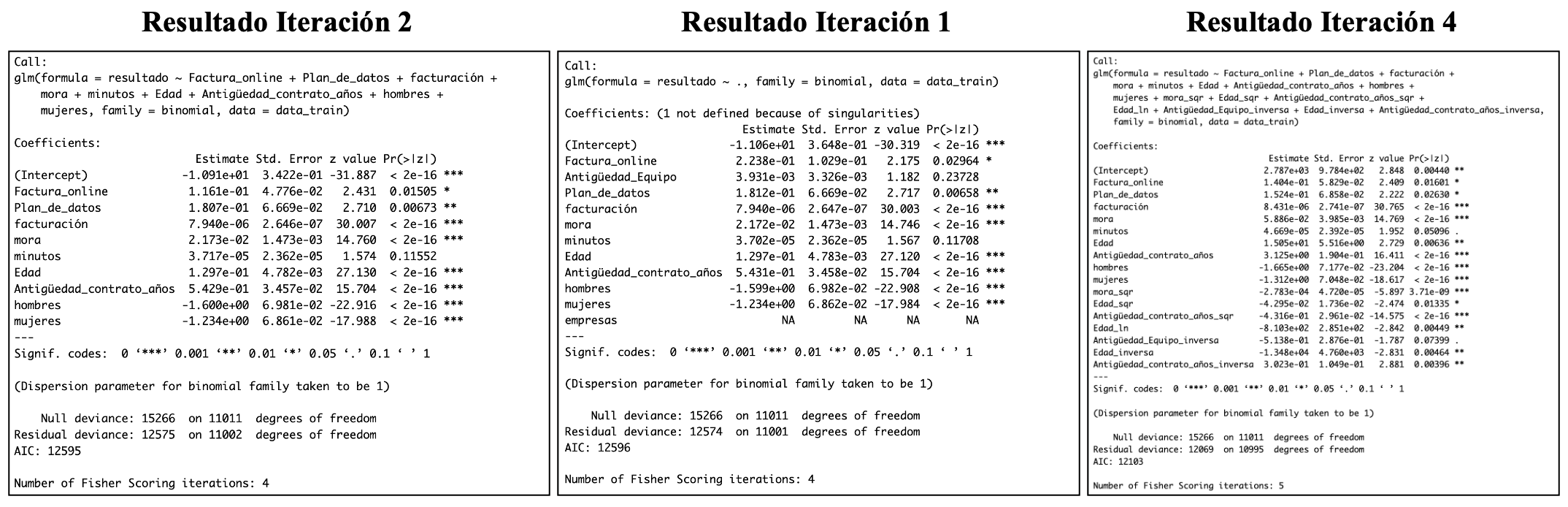
## Generate test design:

Para el diseño y evaluación de las regresiones de datos, se comenzó por revisar las variables originales y se propusieron modelos de regresión logística utilizando solamente estas variables, considerando el conocimiento del negocio. Luego, se llevaron a cabo transformaciones graduales en las variables originales, como se muestra en el punto anterior, generando iteraciones adicionales de regresiones logísticas. Para seleccionar las variables más relevantes, se aplicó el método de stepwise. Además, se compararon los modelos de regresión en función del área bajo la curva (AUC) medida en la curva de *Receiver Operating Characteristics* (ROC). Este proceso permitió optimizar y refinar el modelo de regresión logística para mejorar su capacidad predictiva. En el *Anexo 1* se pueden ver cada una de las 14 iteraciones realizadas con sus métricas correspondientes tanto los datasets producidos con *undersampling* como en los datasets con *oversampling*.

## Build model:

De acuerdo al proceso mencionado se evaluaron 14 iteraciones de tipo regresión logística usando las variables explicativas originales así como transformaciones creadas y las interacciones propuestas. Del proceso se seleccionaron los siguientes tres modelos conforme al criterio de selección de mayor área bajo la curva – AUC y evaluando con un punto de corte de 0.5:

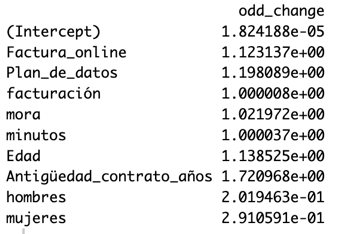
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Iteración** | **AIC** | **Precisión**  **Train** | **Exhaustividad**  **Train** | **F\_Score**  **Train** | **AUC**  **Train** | **Precisión**  **Test** | **Exhaustividad**  **Test** | **F\_Score**  **Test** | **AUC**  **Test** | **Tipo Balanceo** |
| 2 | 12595.40 | 0.7041178 | 0.6863422 | 0.6951163 | 0.6989648 | 0.3427562 | 0.7132353 | 0.4630072 | 0.8000251 | Oversampling |
| 1 | 12596.00 | 0.7045752 | 0.6852525 | 0.6947795 | 0.6989648 | 0.3421517 | 0.7132353 | 0.4624553 | 0.7994533 | Oversampling |
| 4 | 12102.59 | 0.7105066 | 0.6877951 | 0.6989664 | 0.7037777 | 0.3375681 | 0.6838235 | 0.4520049 | 0.7933962 | Oversampling |



# EVALUATION:

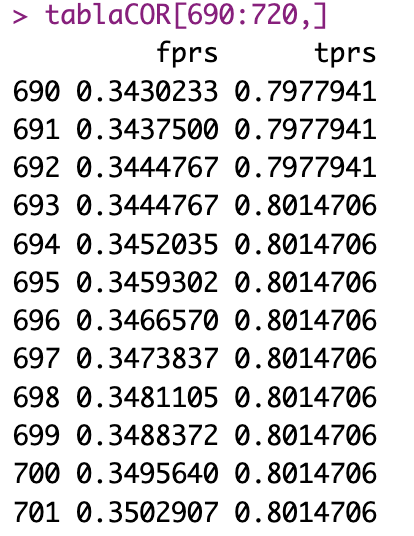
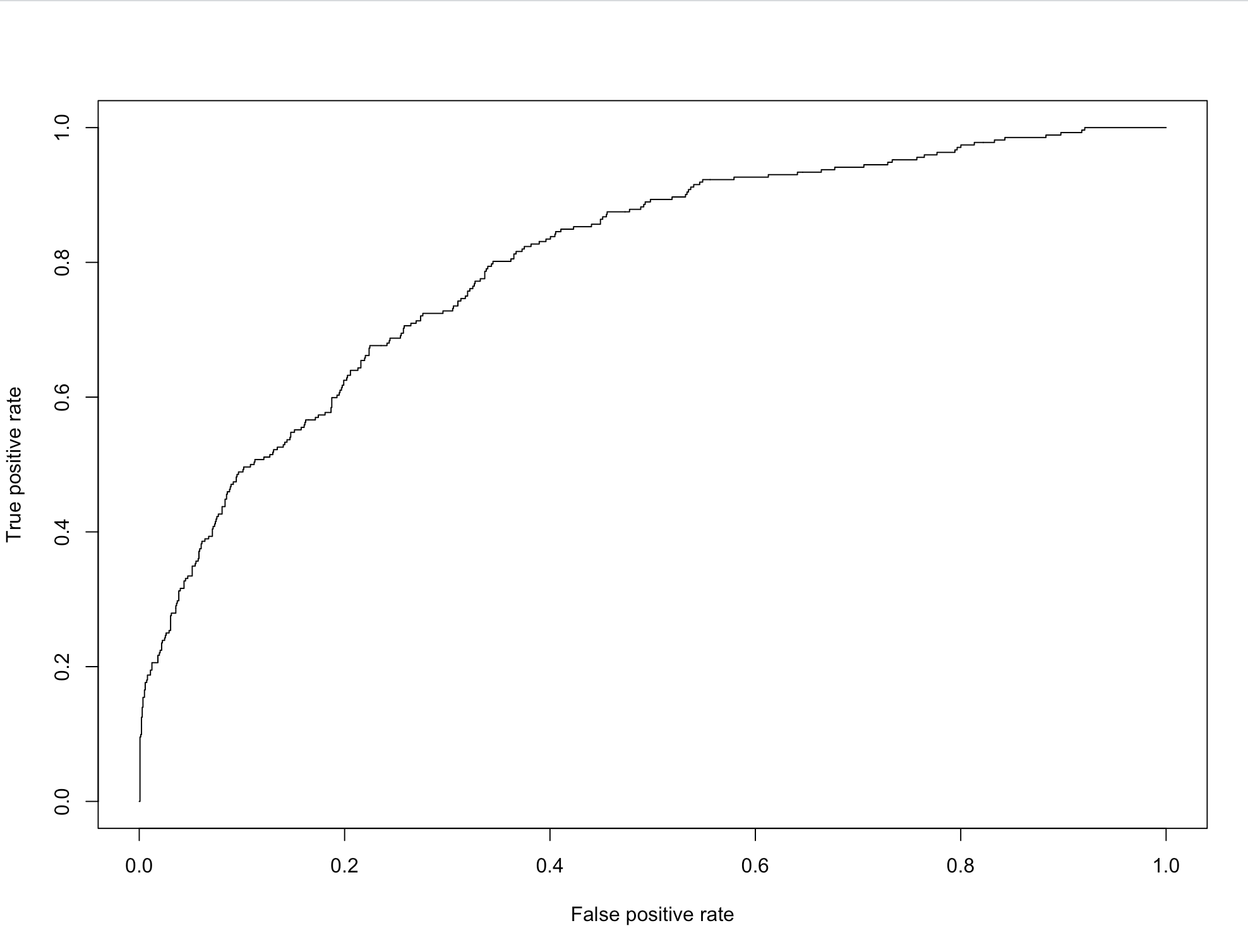
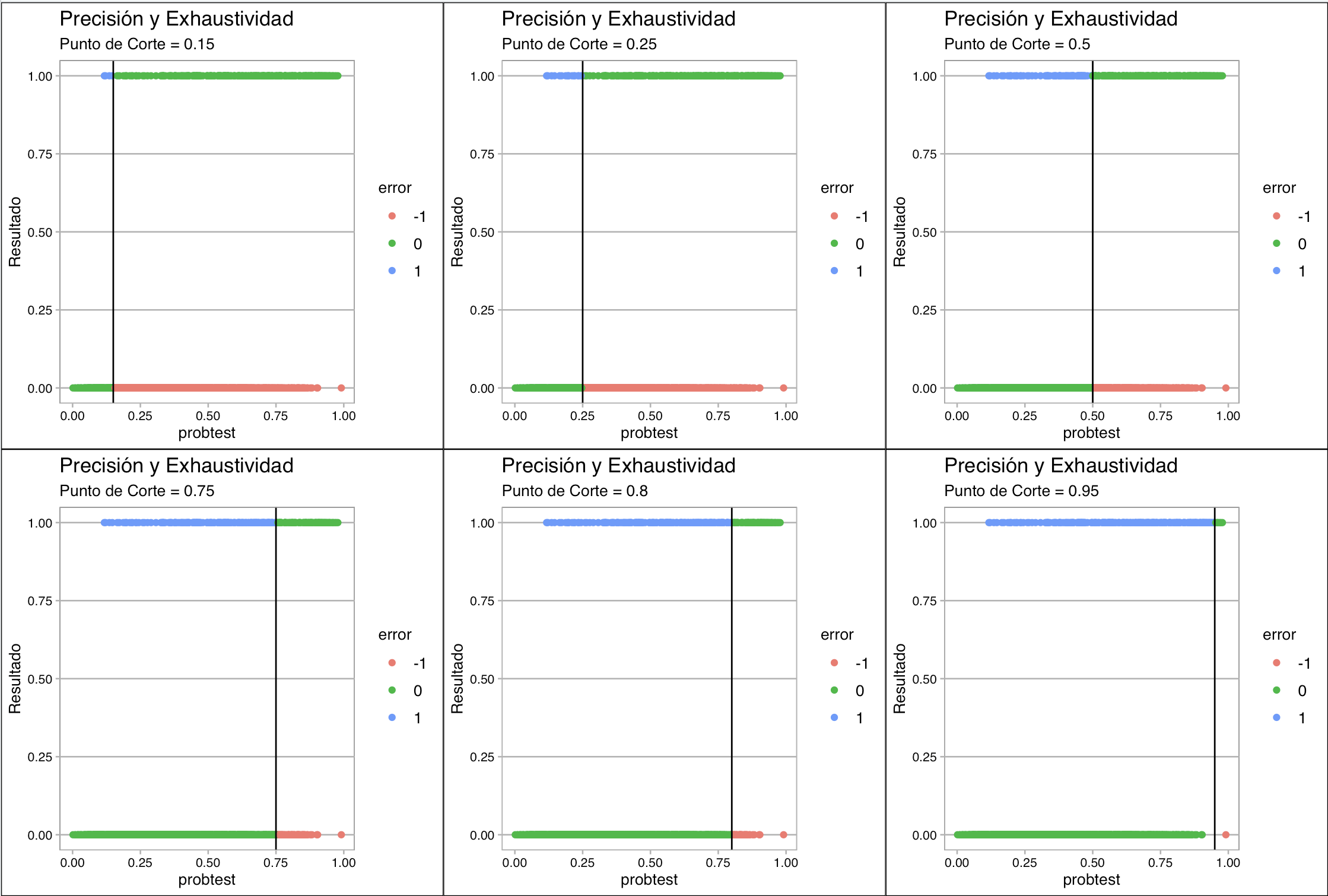
## Assess model:

Teniendo en cuenta el objetivo de data mining del proyecto el cual es escoger aquella regresión logística con la mayor área bajo la curva – AUC, el modelo elegido es el modelo de la iteración 2. Este modelo tiene la forma de:

Los coeficientes resultantes representan el logaritmo del odds ratio para cada variable predictiva usada. El odds ratio es una medida de cuánta más o menos probabilidad hay de que ocurra un evento para un grupo en comparación con otro. Un coeficiente positivo indica que es más probable que ocurra el evento, mientras que un coeficiente negativo indica que es menos probable que ocurra. De los resultados, y calculando el odd change con base en los coeficientes se puede concluir:

* **Factura\_online:** Recibir solamente una factura online, dejando las demás variables constantes, aumenta en 12.31% la probabilidad de que un cliente se retire de la compañía.
* **Plan\_de\_datos:** Tener plan de datos premium, dejando las demás variables constantes, aumenta en 19.81% la probabilidad de que un cliente se retire de la compañía.
* **Facturación:** El aumento en la facturación de una unidad monetaria, dejando las demás variables constantes, aumenta en un 0.0008% la probabilidad de que un cliente se retire de la compañía. Es decir, es una variable que no cambia la probabilidad de que el cliente se retire de la compañía.
* **Mora:** El aumento de un día de mora,dejando las demás variables constantes, aumenta en un 2.20% la probabilidad de que un cliente se retire de la compañía.
* **Minutos:** El aumento de un minuto de consumo, dejando las demás variables constantes, aumenta en un 0.0037% la probabilidad de que un cliente se retire de la compañía. Esta es una variable similar a Facturación, cuya afectación es muy baja sobre la probabilidad de que le cliente se retire.
* **Edad:** El aumento de un año más de edad del cliente, dejando las demás variables constantes, aumenta en un 13.85% la probabilidad de que un cliente se retire de la compañía.
* **Antigüedad\_contrato\_años:** El aumento de un año más en el contrato, dejando las demás variables constantes, aumenta en un 72.1% la probabilidad de que un cliente se retire de la compañía. Esta variable nos puede interesar estratégicamente, al ser una de las que más influencia tiene sobre la probabilidad de retirarse.
* **Hombres:** Clientes que son hombres, dejando las demás variables constantes, por un cliente nuevo que sea hombre este disminuye en 20.19% la probabilidad de retirarse de la compañía.
* **Mujeres:** Clientes que son mujeres, dejando las demás variables constantes, por un cliente nuevo que sea mujer esta disminuye en 29.11% la probabilidad de retirarse de la compañía.

Del modelo también se puede evaluar cómo cambia la precisión y exhaustividad a diversos puntos de corte:



El área bajo la curva es de 80%, lo cual se encuentra dentro de un rango adecuado ya que está por encima del 50% de una predicción al azar. Además, La curva COR muestra que para lograr un 80% de exhaustividad (tomar acciones sobre el 80% de los clientes desertores), se tendría que aceptar casi un 35% de falsos positivos, y en el caso de la Claro, se tendría que incomodar al 35% de clientes que van a seguir con la compañía.

## Produce final report:

Ya que de acuerdo con la cantidad de clientes perdidos por Claro el año anterior de 62.9 mil clientes.6mil millones de pesos (plan pospago promedio es de $65.900), , que resultó en una perdida aproximada de 45Claro desea abordar este problema de retiro de usuarios, buscando tener una tasa de retención de por lo menos 10%, teniendo en cuenta la competencia agresiva que se tiene en el mercado. Por esto, se impactarán las 6.882 de la base que aún continúan con el servicio, buscando retener aquellas personas que es probable que deseen retirarse, proyectando una ganancia al retenerlas de 45.3 millones de pesos.

La primera estrategia es una campaña de SMS y email marketing. Se hará a los clientes de base trabajada en este proyecto, llegando a la totalidad de la base, sin importar que se llegue a usuarios que no estén interesados en retirarse. El costo por mensaje oscila entre los $7 y $40[[2]](#footnote-3) por mensaje y el costo por mil mensajes de email se encuentra aproximadamente entre los $15.000 y $20.000, haciendo costo efectiva la campaña. El mensaje de la campaña está enfocado en incentivar la renovación y actualización de los contratos, comunicando a aquellas personas que lleven más de 2 años, beneficios adicionales al hacer la renovación y actualización del plan pospago con la compañía, ofreciendo descuentos con marcas aliadas (restaurantes, tiendas de ropa, planes turísticos), acceso exclusivo de Claro Video, y descuentos en la renovación de celular, esto con el fin de disminuir la probabilidad de que clientes con mayor tiempo con la compañía tengan incentivos a retirarse, teniendo en cuenta que entre más tiempo de contrato tienen con nosotros su probabilidad de retirarse aumenta en un 79%, al renovar su equipo y entregar nuevos productos de valor, se busca desincentivar esta deserción.

Si la campaña tiene éxito y se logra una retención del 10% (688), la ganancia sería de 45.3 millones versus el costo de la campaña de 415.280 ($40\*6.882 + $20.000 \* 7), generando un ROI de 109.08. Esto es costo eficiente dado que la precisión del modelo es de 0.34 y la tasa costo/ganancia es de 0.009.

Teniendo en cuenta lo hallado, donde la edad es un factor que afecta la toma de decisión de si continuar o no con la compañía, se desea impactar a las generaciones *Z* y *millenials*, ya que ambas generaciones tienen preferencia por la tecnología y las redes sociales (Indeed, 2022). La estrategia busca la creación de planes más modernos y convenios tecnológicos, para apalancar ambas generaciones por medio de innovación, con promociones que lleguen a los intereses, creando planes tipo “el influencer” que no solo cuenta con acceso a todas las redes sociales, sino que cuenta con convenio de cursos online como Platzi o Coursera a un costo total de 99.900 mensuales, o “el gammer” el cual tiene beneficios de poder contar con suscripción a Xbox game pass o Play Plus, para poder jugar sus juegos favoritos junto con la factura del plan, a un costo de 89.900 mil pesos mensuales, servicios con la más alta velocidad y cobertura del mercado, ambos beneficios son llamativos, dado que los servicios solo de gaming y de educación online mensual están aproximadamente entre los 35mil y 80mil mensuales respectivamente.

### Potenciales siguientes pasos

Sería importante poder tener más información del comportamiento e intereses de los usuarios, con el fin de poder realizar Clústeres por intereses de clientes, o grupos de productos, para realizar promociones que beneficien tanto al usuario como a la compañía y la puedan hacer más atractiva y competitiva de lo que actualmente es, no solo disminuyendo su pérdida de usuarios y aumentando la retención de estos, si no aumentando la captación de nuevos usuarios. Finalmente, también es importante que en futuras iteraciones de este ejercicio se agreguen variables adicionales como el consumo de plan de datos, SMS y consumo de apps populares, ubicación del cliente, y servicios adicionales que el cliente puede tener con la compañía, para así realizar un análisis más exhaustivo de la retención.

**Anexo 1**

**Iteraciones de modelos usando datasets de train y test (80%-20%) y balanceo con oversampling**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **#** | **Modelo\*** | **Punto**  **corte** | **AIC** | **Precisión**  **Train** | **Exhaustividad**  **Train** | **F\_Score**  **Train** | **AUC**  **Train** | **Presicion Test** | **Exhaustividad Test** | **F\_Score Test** | **AUC Test** |
| 1 | β + Factura\_online + Antigüedad\_Equipo + Plan\_de\_datos + facturación + mora + minutos + Edad + Antiguedad\_contrato\_años + hombres + mujeres | 0.5 | 12596.00 | 0.7045752 | 0.6852525 | 0.6947795 | 0.6989648 | 0.3421517 | 0.7132353 | 0.4624553 | 0.7994533 |
| 2 | β + Factura\_online + Plan\_de\_datos + facturación + mora + minutos + Edad + Antiguedad\_contrato\_años + hombres + mujeres | 0.5 | 12595.40 | 0.7041178 | 0.6863422 | 0.6951163 | 0.6989648 | 0.3427562 | 0.7132353 | 0.4630072 | 0.8000251 |
| 4 | β + Factura\_online + Plan\_de\_datos + facturación + mora + minutos + Edad + Antigüedad\_contrato\_años + hombres + mujeres + mora\_sqr + Edad\_sqr + Antigüedad\_contrato\_años\_sqr + Edad\_ln + Antigüedad\_Equipo\_inversa + Edad\_inversa + Antigüedad\_contrato\_años\_inversa | 0.5 | 12102.59 | 0.7105066 | 0.6877951 | 0.6989664 | 0.7037777 | 0.3375681 | 0.6838235 | 0.4520049 | 0.7933962 |
| 5 | β + Factura\_online + Plan\_de\_datos + facturación + mora + minutos + Edad + Antigüedad\_contrato\_años + hombres + mujeres + mora\_sqr + Edad\_sqr + Antigüedad\_contrato\_años\_sqr + Edad\_ln + Antigüedad\_Equipo\_inversa + Edad\_inversa + Antigüedad\_contrato\_años\_inversa + mora\_edad | 0.5 | 12102.59 | 0.7131519 | 0.6854341 | 0.6990183 | 0.7048674 | 0.3363309 | 0.6875000 | 0.4516908 | 0.7930168 |
| 6 | β + Factura\_online + Antigüedad\_Equipo + Plan\_de\_datos + facturación + mora + minutos + Edad + Antigüedad\_contrato\_años + hombres + mujeres + empresas + Antigüedad\_Equipo\_sqr + facturación\_sqr + mora\_sqr + minutos\_sqr + Edad\_sqr + Antigüedad\_contrato\_años\_sqr + Antigüedad\_Equipo\_ln + minutos\_ln + Edad\_ln + Antigüedad\_Equipo\_inversa + facturación\_inversa + minutos\_inversa + Edad\_inversa + Antigüedad\_contrato\_años\_inversa + mora\_minutos + mora\_facturacion + mora\_edad + comunidad\_baja + comunidad\_media + comunidad\_media\_alta | 0.5 | 12096.66 | 0.7133962 | 0.6867054 | 0.6997964 | 0.7054123 | 0.3339318 | 0.6838235 | 0.4487334 | 0.7930382 |
| 7 | β + Factura\_online + Plan\_de\_datos + facturación + mora + Edad + Antigüedad\_contrato\_años + hombres + mujeres + mora\_sqr + Edad\_sqr + Antigüedad\_contrato\_años\_sqr + minutos\_ln + Edad\_ln + Antigüedad\_Equipo\_inversa + Edad\_inversa + Antigüedad\_contrato\_años\_inversa + mora\_edad + comunidad\_baja + comunidad\_media | 0.5 | 12079.62 | 0.7154426 | 0.6899746 | 0.7024778 | 0.7077733 | 0.3333333 | 0.6727941 | 0.4457978 | 0.7929848 |
| 8 | β + Factura\_online + Plan\_de\_datos + facturación + mora + Edad + Antigüedad\_contrato\_años + hombres + mujeres + mora\_sqr + Edad\_sqr + Antigüedad\_contrato\_años\_sqr + minutos\_ln + Edad\_ln + Antigüedad\_Equipo\_inversa + Edad\_inversa + Antigüedad\_contrato\_años\_inversa + mora\_edad + comunidad\_baja + comunidad\_media + minutos\_facturacion | 0.5 | 12081.46 | 0.7154426 | 0.6899746 | 0.7024778 | 0.7077733 | 0.3333333 | 0.6727941 | 0.4457978 | 0.7929741 |
| 9 | β + Factura\_online + Plan\_de\_datos + facturación + mora + minutos + Edad + hombres + mujeres + mora\_sqr + Edad\_sqr + Antigüedad\_contrato\_años\_sqr + minutos\_ln + Edad\_ln + Antigüedad\_Equipo\_inversa + minutos\_inversa + Edad\_inversa + Antigüedad\_contrato\_años\_inversa + mora\_edad + comunidad\_baja + comunidad\_media + ME\_Minutos\_BirthYear + ME\_duracion\_BirthYear | 0.5 | 12071.24 | 0.7154426 | 0.6899746 | 0.7024778 | 0.7077733 | 0.3382084 | 0.6801471 | 0.4517705 | 0.7926775 |

**Iteraciones de modelos usando datasets de train y test (80%-20%) y balanceo con undersampling**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **#** | **Modelo\*** | **Punto**  **corte** | **AIC** | **Precisión**  **Train** | **Exhaustividad**  **Train** | **F\_Score**  **Train** | **AUC**  **Train** | **Precisión Test** | **Exhaustividad Test** | **F\_Score Test** | **AUC Test** |
| 1 | β + Factura\_online + Antigüedad\_Equipo + Plan\_de\_datos + facturación + mora + minutos + Edad + Antiguedad\_contrato\_años + hombres + mujeres | 0.5 | 1267.939 | 0.6958955 | 0.6856618 | 0.6907407 | 0.6930147 | 0.3173844 | 0.7316176 | 0.4427141 | 0.7807370 |
| 2 | β + facturación + mora + Edad + Antiguedad\_contrato\_años + hombres + mujeres | 0.5 | 1261.883 | 0.7033582 | 0.6930147 | 0.6981481 | 0.7003676 | 0.3165584 | 0.7169118 | 0.4391892 | 0.7797885 |
| 4 | β + facturación + mora + Edad + Antiguedad\_contrato\_años + hombres + muejeres + mora\_sqr + Antigüedad\_contrato\_años\_sqr + Edad\_ln | 0.5 | 1218.569 | 0.7099812 | 0.6930147 | 0.7013953 | 0.7049632 | 0.3227953 | 0.7132353 | 0.4444444 | 0.7847902 |
| 5 | β + facturación + mora + Edad + Antiguedad\_contrato\_años + hombres + muejres + Antigüedad\_contrato\_años\_sqr + Edad\_ln + mora\_edad | 0.5 | 1218.569 | 0.7080979 | 0.6911765 | 0.6995349 | 0.7031250 | 0.3259386 | 0.7022059 | 0.4452214 | 0.7787491 |
| 6 | β + Factura\_online + Antigüedad\_Equipo + Plan\_de\_datos + facturación + mora + minutos + Edad + Antigüedad\_contrato\_años + hombres + mujeres + empresas + Antigüedad\_Equipo\_sqr + facturación\_sqr + mora\_sqr + minutos\_sqr + Edad\_sqr + Antigüedad\_contrato\_años\_sqr + Antigüedad\_Equipo\_ln + minutos\_ln + Edad\_ln + Antigüedad\_Equipo\_inversa + facturación\_inversa + minutos\_inversa + Edad\_inversa + Antigüedad\_contrato\_años\_inversa + mora\_minutos + mora\_facturacion + mora\_edad + comunidad\_baja + comunidad\_media + comunidad\_media\_alta | 0.5 | 1244.384 | 0.7161905 | 0.6911765 | 0.7034612 | 0.7086397 | 0.3216667 | 0.7095588 | 0.4426606 | 0.7782415 |
| 7 | β + Plan\_de\_datos + mora + Antigüedad\_contrato\_años + hombres + mujeres + facturación\_sqr + Edad\_sqr + Antigüedad\_contrato\_años\_sqr + mora\_minutos + comunidad\_alta | 0.5 | 1239.152 | 0.7123810 | 0.6875000 | 0.6997194 | 0.7049632 | 0.3260870 | 0.7169118 | 0.4482759 | 0.7814557 |
| 8 | β + facturación + mora + Antigüedad\_contrato\_años + hombres + mujeres + Antigüedad\_contrato\_años\_sqr + mora\_edad + minutos\_facturacion | 0.5 | 1262.297 | 0.6996198 | 0.6764706 | 0.6878505 | 0.6930147 | 0.2801303 | 0.6323529 | 0.3882619 | 0.7440631 |
| 9 | β + facturación + mora + Edad + hombres + mujeres + Antigüedad\_contrato\_años\_sqr + Edad\_ln + mora\_edad + ME\_duracion\_BirthYear | 0.5 | 1211.328 | 0.6996198 | 0.6764706 | 0.6878505 | 0.6930147 | 0.3259386 | 0.7022059 | 0.4452214 | 0.7787839 |

\***Nota**: los modelos presentados son de la forma:

# REFERENCIAS

Ávila Forero, R. (2023, 20 febrero). Portabilidad numérica: más allá de la norma. *Semana.com*. Recuperado 4 de noviembre de 2023, de <https://www.semana.com/economia/opinion/articulo/portabilidad-numerica-mas-alla-de-la-norma/202300/>

Beltrán, A. M., Alvarado, J., García, J. C., González Sanmiguel, J., & Cuervo, O. A. (2023). *Retención en telefonía móvil 2023-30*. Kaggle. Recuperado 13 de noviembre de 2023, de <https://www.kaggle.com/competitions/retencin-en-telefona-mvil-2023-30/overview>

Ciuró, J., Jiménez, J. F., Moura, L., & Camargos, G. (2020). *Impacto económico y social de Android en México*. Bain & Company. Recuperado 10 de noviembre de 2023, de <https://www.bain.com/contentassets/a9200a057a0241b8963c05a9b09e33fe/impacto-economico-y-social-de-android-en-mexico-1.pdf>

Comisión de Regulación de Comunicaciones. (2023a). *Data Flash 2023-006 - Portabilidad numérica móvil | PostData*. Postdata - CRC. Recuperado 30 de octubre de 2023, de <https://postdata.gov.co/dataflash/data-flash-2023-006-portabilidad-numerica-movil#:~:text=La%20Comisi%C3%B3n%20de%20Regulaci%C3%B3n%20de,de%20la%20Resoluci%C3%B3n%20CRC%205050>

Comisión de Regulación de Comunicaciones. (2023b). *Información Telefonía móvil*. Postdata. Recuperado 30 de octubre de 2023, de <https://postdata.gov.co/informaci%C3%B3n-telefon%C3%ADa-m%C3%B3vil>

Comisión de Regulación de Comunicaciones. (2023c). *La CRC simplifica el proceso de portabilidad numérica móvil en beneficio de los usuarios*. Recuperado 30 de octubre de 2023, de <https://www.crcom.gov.co/es/noticias/proyectos-regulatorios/crc-simplifica-proceso-portabilidad-numerica-movil-en-beneficio#:~:text=La%20Portabilidad%20Num%C3%A9rica%20M%C3%B3vil%20es,se%20ajuste%20a%20sus%20necesidades>

Comisión de Regulación de Comunicaciones. (2023d). *Usuarios realizaron 1,78 millones de operaciones de portabilidad numérica móvil en el primer trimestre de 2023*. Recuperado 30 de octubre de 2023, de <https://www.crcom.gov.co/es/noticias/comunicado-prensa/usuarios-realizaron-178-millones-operaciones-portabilidad-numerica-movil>

Comunidad Claro (2016). ¿Por qué mi facturación viene en negativo?. Claro. Recuperado de: https://comunidad.claro.com.pe/t5/M%C3%B3vil/Por-qu%C3%A9-mi-facturaci%C3%B3n-viene-en-negativo/td-p/1304#:~:text=Re%3A%20%C2%BFPor%20qu%C3%A9%20mi%20facturaci%C3%B3n%20viene%20en%20negativo%3F,-Hola%20%40JorgeBellido&text=%C2%A1Significa%20que%20tienes%20un%20saldo,reclamo%20pendiente%20o%20alg%C3%BAn%20bloqueo.

Fernández-Ardèvol, M. (2023). *Las comunicaciones móviles y el desarrollo socioeconómico:Una perspectiva latinoamericana*. United Nations. Recuperado 4 de noviembre de 2023, de <https://www.un.org/es/chronicle/article/las-comunicaciones-moviles-y-el-desarrollo-socioeconomicouna-perspectiva-latinoamericana>

Indeed. (2022). Gen Z vs. Millennials: Definitions and Key Differences. Indeed. Recuperado 12 de noviembre de 2023, de <https://ca.indeed.com/career-advice/career-development/gen-z-vs-millennials>

López, J. M. (2022, 24 marzo). *El 5G y la conexión por satélite, un trabajo conjunto*. ThinkBig. Recuperado 1 de noviembre de 2023, de <https://blogthinkbig.com/internet-satelite-5g>

Manďák, J., & Hančlová, J. (2019). Use of Logistic Regression for Understanding and Prediction of Customer Churn in Telecommunications. *Statistika: Statistics & Economy Journal,* *99*(2), 131-141. <https://web.s.ebscohost.com/abstract?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=0322788X&AN=137158738&h=OxgXtU8mdZNvJ6%2bwneWuG3%2fndOPyNWnq93KM3I3SXJRnB8Wn0KGXbFy8SN3cEcNJmISMN8iESvZEvb7ZGmJYbw%3d%3d&crl=c&resultNs=AdminWebAuth&resultLocal=ErrCrlNotAuth&crlhashurl=login.aspx%3fdirect%3dtrue%26profile%3dehost%26scope%3dsite%26authtype%3dcrawler%26jrnl%3d0322788X%26AN%3d137158738>

Meza Rodríguez, A. R., & Chue Gallardo, J. (2020). Uso del algoritmo Adaboost y la regresión logística para la predicción de fuga de clientes en una empresa de telefonía móvil. *Natura@economía*, *5*(2), 2226-9479. <https://dialnet.unirioja.es/servlet/articulo?codigo=9095465>

Ministerio de Tecnologías de la Información y Comunicaciones. (2019). *TELEFONIA MOVIL CELULAR - Definición legal / SERVICIO DE LARGA DISTANCIA NACIONAL E INTERNACIONAL - Definición legal / TELEFONIA BASICA CONMUTADA - Concepto*. Normograma Ministerio de Tecnologías de la Información y Comunicaciones. Recuperado 3 de noviembre de 2023, de <https://normograma.mintic.gov.co/mintic/docs/pdf/11001-03-24-000-2001-0212-01(7199).pdf>

Ministerio de Tecnologías de la Información y las Comunicaciones de Colombia. (2023). *Telefonía móvil*. Recuperado 30 de octubre de 2023, de <https://colombiatic.mintic.gov.co/679/w3-propertyvalue-47274.html>

Perez Becerra, G., & Yamal Matos, S. (2011). *Diagnóstico de los factores generales del entorno de las telecomunicaciones en Colombia en el periodo 2005-2011* [Tesis]. Universidad Tecnológica de Bolívar. Recuperado 1 de noviembre de 2023, de <https://repositorio.utb.edu.co/bitstream/handle/20.500.12585/977/0061366.pdf?sequence=1>

Portafolio. (2023a, agosto 9). *Pulso entre los operadores móviles seguirá al rojo vivo*. Portafolio.co. Recuperado 2 de noviembre de 2023, de <https://www.portafolio.co/negocios/empresas/aumenta-competencia-entre-empresas-de-telefonia-movil-587091>

Portafolio. (2023b, septiembre 17). *¿Qué tan preparada está Colombia para la llegada de la tecnología 5G?* Portafolio.co. Recuperado 1 de noviembre de 2023, de <https://www.portafolio.co/innovacion/expertos-senalan-que-colombia-esta-lista-para-la-llegada-del-5g-589193>

Semana. (2023, 29 abril). Ranking Top of Mind 2023. *Semana.com*. Recuperado 1 de noviembre de 2023, de <https://www.semana.com/especiales-editoriales/articulo/ranking-top-of-mind-2023/202337/>

Suárez, M. C. (2023, 6 octubre). Inflación en septiembre llegó a 10,99% y baja respecto a agosto y al año pasado. *Bloomberg Línea*. Recuperado 1 de noviembre de 2023, de <https://www.bloomberglinea.com/latinoamerica/colombia/inflacion-en-septiembre-llego-a-1099-y-baja-respecto-a-agosto-y-al-ano-pasado/#:~:text=La%20categor%C3%ADa%20Informaci%C3%B3n%20y%20comunicaci%C3%B3n,reparaci%C3%B3n%20(0%2C92%25)>

Uanataca. (2021, 9 diciembre). *5 tendencias tecnológicas en telecomunicaciones para 2022*. Uanataca. Recuperado 13 de noviembre de 2023, de <https://web.uanataca.com/es/blog/telecomunicaciones/tendencias-tecnologicas-para-2022>

Universidad Externado de Colombia. (2021). Normativa sobre las tecnologías de la información y las comunicaciones. *BOLETÍN e-Telecomunicaciones*, *44*, 8-12. Recuperado 13 de noviembre de 2023, de <https://www.uexternado.edu.co/wp-content/uploads/2023/07/Boletín-44.pdf>

1. Medida que permite a los usuarios cambiar de operador manteniendo su número, con el fin de que este pueda elegir qué servicio le beneficia más y decidir cuál le genera más satisfacción. [↑](#footnote-ref-2)
2. Costos aproximados tomando información de: <https://clickpanda.com/sms>, <https://www.hablame.co/sms/>, <https://www.smsmasivo.com.co> y <https://www.wausms.com/envio-masivo-sms/sms-empresas/Colombia/> [↑](#footnote-ref-3)