



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

GENERACIÓN DE MODELO PARA INCREMENTAR LA MIGRACIÓN DE  
CLIENTES PREPAGO A POSPAGO EN UNA COMPAÑÍA DE  
TELECOMUNICACIONES

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

FELIPE ANDRÉS ALVARADO FUCHSLOCHER

PROFESOR GUÍA  
PABLO MARÍN VICUÑA

MIEMBROS DE LA COMISIÓN  
CAROLINA SEGOVIA RIQUELME  
DANIEL VARELA LÓPEZ

SANTIAGO DE CHILE  
2020

**RESUMEN DE LA MEMORIA PARA OPTAR AL  
TÍTULO DE:** Ingeniero Civil Industrial

**POR:** Felipe Andrés Alvarado Fuchslocher

**FECHA:** 07/05/2020

**PROFESOR GUÍA:** Pablo Marín Vicuña

**GENERACIÓN DE MODELO PARA INCREMENTAR LA MIGRACIÓN DE  
CLIENTES PREPAGO A POSPAGO EN UNA COMPAÑÍA DE  
TELECOMUNICACIONES**

El trabajo de memoria se desarrolla en una empresa de telecomunicaciones, particularmente en el sector de telefonía móvil, el cual se caracteriza por una tendencia en el aumento de los clientes pospago, usuarios que poseen un contrato con la compañía a cambio de un plan que les permite navegar por internet y/o realizar llamadas telefónicas. Dado el contexto de la industria, la compañía realiza una campaña mensual en la cual ofrece vía call center a un grupo de clientes prepago la opción de contratar un plan, siendo estos elegidos en base a reglas del negocio.

El objetivo del trabajo de memoria consiste en mejorar la tasa de conversión actual de la campaña mediante una metodología que defina a que clientes gestionar en esta, lo anterior, a través de la implementación de dos modelos analíticos. Para llevar a cabo el trabajo se utilizó la metodología CRISP-DM, bajo la cual se probaron tres algoritmos de clasificación diferentes correspondientes a Random Forest, Gradient Boosting Machine (GBM), y Extra Gradient Boosting Machine (XGBoost).

La metodología analítica desarrollada fue utilizada para perfilar los clientes a gestionar en la campaña del mes de diciembre 2019, obteniendo buenos resultados a partir de este. En base al trabajo de memoria desarrollado se duplica la utilidad promedio de un cliente migrado hacia pospago, pasando de una utilidad de \$1.753 a \$3.670.

Por otro lado, mediante la incorporación del trabajo realizado se aumenta la tasa de efectividad de la campaña, puesto que se deja fuera a clientes que normalmente se gestionaban y posee una baja tasa de conversión (3,9%), reemplazándolos por un nuevo segmento de clientes no considerados previamente, los cuales poseen una mayor tasa de efectividad (11,4%).

En conclusión, la incorporación de los modelos analíticos para definir la base de clientes a gestionar en la campaña genera valor para la compañía, puesto que, mediante la incorporación de estos, es posible obtener una mayor tasa de efectividad en relación a la gestión actual, mejorando la cantidad de clientes migrados, así como también la utilidad de la campaña.

## Agradecimientos

*A mis padres y hermanos, por apoyarme en cada una de mis decisiones.*

*A Carolina, por su apoyo y amor incondicional.*

*A todos mis amigos por hacer mi estadía en la universidad una experiencia increíble.*

*A Entel, por confiar en mí y darme la oportunidad de realizar mi memoria junto a ellos.*

## Tabla de contenido

1. Introducción .....	8
2. Antecedentes Generales .....	8
2.1. Industria de telefonía móvil.....	8
2.2. La Empresa.....	10
2.3. La Empresa en el mercado de prepago.....	11
3. Descripción y justificación del proyecto .....	14
3.1. Descripción del proyecto.....	14
3.2. Justificación del proyecto .....	15
3.3. Trabajo a realizar .....	19
4. Objetivos .....	20
4.1. Objetivo General .....	20
4.2. Objetivos Específicos .....	21
5. Alcances y resultados esperados .....	21
5.1. Alcances.....	21
5.2. Resultados esperados.....	21
6. Marco teórico.....	22
6.1. Metodología CRISP-DM .....	22
6.2. Algoritmos de clasificación .....	24
6.3. Métricas de evaluación.....	27
6.4. Técnicas de tratamiento de outliers .....	31
7. Desarrollo de la metodología .....	31
7.1. Conocimiento del negocio.....	32
7.1.1. Campaña de migración .....	32
7.1.2. Análisis de efectividad de Rut.....	34
7.1.3. Análisis de rentabilidad de campaña.....	36
7.1.4. Cambios en la gestión actual de la campaña de migraciones .....	41
7.1.5. Generación actual de base de clientes a cargar en campaña .....	43
7.1.6. Desarrollo del trabajo .....	43
7.2. Comprensión de los datos.....	44
7.3. Preparación de datos.....	49

7.3.2.	Creación de variables .....	53
7.3.3.	Transformación de variables .....	55
7.4.	Modelación .....	56
7.4.1.	Descripción de modelos .....	56
7.4.2.	Modelo de migración vía campaña .....	59
7.4.3.	Modelo de migración espontánea .....	77
7.5.	Despliegue .....	97
7.5.1.	Composición de base a cargar en campaña .....	98
7.5.2.	Distribución de scores de propensión .....	100
8.	Evaluación de campaña .....	102
8.1.	Evaluación general .....	102
8.2.	Evaluación según rango de recarga .....	103
8.3.	Análisis según deciles de propensión .....	104
8.4.	Análisis según score de propensión .....	107
8.5.	Análisis de cumplimiento del plan sugerido .....	107
8.6.	Análisis de utilidad de la campaña .....	108
9.	Conclusiones .....	110
10.	Recomendaciones y trabajo futuro .....	111
10.1.	Recomendaciones .....	111
10.2.	Trabajo futuro .....	112
11.	Bibliografía .....	114
12.	Anexos .....	115
12.1	Anexo 1: Detalle de aplicaciones utilizados en el modelo .....	115
12.2	Anexo 2: Importancia de variables modelo de migración vía campaña .....	116
12.3	Anexo 3: Importancia de variables modelo de migración espontánea .....	117
12.4	Anexo 4: Test de hipótesis para comparar diferencias de proporciones .....	118

## Índice de figuras

Figura 1: Evolución de clientes abonados por contrato.....	9
Figura 2: Participación de ingresos Entel por segmento de la Industria de Telecomunicaciones .....	11
Figura 3: Evolución de participación de mercado en telefonía móvil prepago.....	12
Figura 4: Ejemplo de desactivación por cobranza.....	17
Figura 5: Distribución de motivos de fuga para clientes migrados a suscripción vía campaña .....	18
Figura 6: Análisis de fuga acumulada para clientes migrados en enero y marzo del 2019 .....	19
Figura 7: Pasos metodología CRIPS-DM .....	24
Figura 8: Ejemplo del algoritmo Random Forest.....	26
Figura 9: Ejemplo de curvas ROC.....	30
Figura 10: Tasa de migración vía campaña según tipo de base.....	33
Figura 11: Distribución de clientes gestionados por campaña. ....	34
Figura 12: Análisis de Rut para clientes gestionados por campaña. ....	35
Figura 13: Análisis de clientes ingresados vía campaña con Rut asociado .....	35
Figura 14: Confiabilidad del Rut .....	36
Figura 15: Distribución de cargo fijo para migraciones outbound.....	37
Figura 16: Distribución de clientes según margen porcentual.....	38
Figura 17: Utilidad de clientes según margen porcentual .....	39
Figura 18: Distribución de clientes según rango de recarga .....	40
Figura 19: Utilidad según rango de recarga .....	41
Figura 20: Descripción de variables utilizadas en el modelo, parte 1. ....	44
Figura 21: Descripción de variables utilizadas en el modelo, parte 2. ....	47
Figura 22: Histograma de antigüedad de clientes prepago.....	49
Figura 23: Árbol de decisión para la antigüedad del cliente .....	50
Figura 24: Estratificación de antigüedad del cliente.....	51
Figura 25: Distribución de marca de equipos clientes prepago .....	52
Figura 26: Distribución de recargas de enero 2019.....	53
Figura 27: Esquema explicativo de construcción del modelo .....	57
Figura 28: Lift acumulado modelo GBM, migración vía campaña. ....	59
Figura 29: Curva ROC modelo GBM, migración vía campaña.....	61
Figura 30: Lift acumulado modelo Random Forest, migración vía campaña. ....	61
Figura 31: Curva ROC modelo Random Forest, migración vía campaña. ....	62
Figura 32: Lift acumulado modelo XGBoost, migración vía campaña. ....	63
Figura 33: Curva ROC modelo XGBoost, migración vía campaña.....	64
Figura 34: Comparación de lift acumulado para modelos de migración vía campaña .....	65
Figura 35: Comparación de curvas ROC modelos de migración vía campaña .....	66
Figura 36: Importancia de variables modelo XGBoost, migración vía campaña. ...	69
Figura 37: Tasa de migración con permanencia a suscripción vía campaña según Tipo de base.....	70

Figura 38: Tasa de migración con permanencia a suscripción vía campaña según llamadas recibidas en el mes previo (seg) .....	71
Figura 39: Tasa de migración con permanencia a suscripción vía campaña según llamadas emitidas en el mes previo (seg) .....	72
Figura 40: Tasa de migración con permanencia a suscripción vía campaña según antigüedad del cliente .....	73
Figura 41: Tasa de migración con permanencia a suscripción vía campaña según llamadas recibidas en el penúltimo mes (seg) .....	74
Figura 42: Tasa de migración con permanencia a suscripción vía campaña según llamadas emitidas en el penúltimo mes (seg).....	75
Figura 43: Tasa de migración con permanencia a suscripción vía campaña según tráfico en Facebook .....	76
Figura 44: Lift acumulado modelo GBM, migración espontánea. ....	78
Figura 45: Curva ROC modelo GBM, migración espontánea.....	79
Figura 46: Lift acumulado modelo Random Forest, migración espontánea. ....	80
Figura 47: Curva ROC modelo Random Forest, migración espontánea.....	81
Figura 48: Lift acumulado modelo XGBoost, migración espontánea.....	82
Figura 49: Curva ROC modelo XGBoost, migración espontánea. ....	83
Figura 50: Comparación de lift acumulado para modelos de migración espontánea .....	84
Figura 51: Comparación de curvas ROC para modelos de migración espontánea ..	86
Figura 52: Importancia de variables modelo XGBoost, modelo de migración espontánea.....	88
Figura 53: Tasa de migración espontánea con permanencia según cantidad de segundos recibidos el último mes .....	89
Figura 54: Tasa de migración espontánea con permanencia según cantidad de segundos recibidos el penúltimo mes. ....	90
Figura 55: Tasa de migración espontánea con permanencia según Tipo de cliente	91
Figura 56: Tasa de migración espontánea con permanencia según cantidad de segundos emitidos el último mes.....	92
Figura 57: Tasa de migración espontánea con permanencia según tráfico en Instagram. ....	93
Figura 58: Tasa de migración espontánea con permanencia según marca del equipo. ....	94
Figura 59: Tasa de migración espontánea con permanencia según gama del equipo .....	95
Figura 60: Esquema de despliegue .....	98
Figura 62: Distribución de clientes según score de propensión asignados por el modelo .....	101
Figura 63: Distribución de score de propensión según grupo de gestión.....	102
Figura 64: Distribución de densidad de clientes gestionados y migrados .....	107

## Índice de tablas

Tabla 1: Matriz de confusión .....	28
Tabla 2: Ejemplo de variables enviadas al call center para campaña de migración	32
Tabla 3: Cambios en la gestión actual de la campaña .....	42
Tabla 4: Distribución de bolsas más frecuentes para clientes prepago .....	46
Tabla 5: Matriz de confusión modelo GBM, migración vía campaña. ....	60
Tabla 6: Matriz de confusión modelo Random Forest, migración vía campaña ....	62
Tabla 7: Matriz de confusión modelo XGBoost, migración vía campaña .....	64
Tabla 8: Comparación de MCC para modelos de migración vía campaña .....	66
Tabla 9: Resumen de gestión según performance de mejor modelo vía campaña ..	68
Tabla 10: Matriz de confusión modelo GBM, migración espontánea .....	79
Tabla 11: Matriz de confusión modelo Random Forest, migración espontánea .....	81
Tabla 12: Matriz de confusión modelo XGBoost, migración espontánea .....	83
Tabla 13: Comparación de métrica MCC para modelo de migración espontánea ...	85
Tabla 14: Resumen de gestión según performance de mejor modelo de migración vía campaña .....	87
Tabla 15: Composición de base cargada a campaña en diciembre 2019 .....	99
Tabla 16: Detalle de composición de base cargada a campaña de migración en diciembre 2019 según grupo de gestión .....	100
Tabla 17: Tasas de migración según grupo de gestión.....	103
Tabla 18: Tasa de migración según rango de recarga.....	104
Tabla 19: Tasa de migración según deciles de propensión.....	105
Tabla 20: Tasa de migración según deciles de propensión y rango de recarga .....	106
Tabla 21: Comparación de utilidad para campañas de septiembre y diciembre 2019 .....	109

## 1. Introducción

El sector de telefonía móvil ha experimentado cambios significativos durante el último tiempo. El mayor uso de internet móvil junto a la reducción tarifaria del sector ha producido un incremento sostenible de los clientes pospago, usuarios que poseen un contrato con la compañía a cambio de un servicio que les permite llamar y navegar por internet.

El trabajo se desarrolla en una empresa de telecomunicaciones, la cual, dado el contexto de mercado, de forma mensual realiza una campaña para ofrecer a clientes prepago la opción de migrar hacia el mercado de pospago.

El objetivo del trabajo consiste en incrementar la tasa de migración de dicha campaña mediante la incorporación de modelos analíticos que permitan tomar la decisión respecto a que clientes debiesen ser gestionados en esta.

En el presente informe se desarrolla una metodología para mejorar la situación recién mencionada, para ello, se analiza la gestión actual de la campaña, y posteriormente se construyen diferentes algoritmos de clasificación.

Por último, a partir de los resultados obtenidos de los modelos, se realiza un piloto para evaluar el desempeño real de estos, concluyendo finalmente respecto a la mejora de la campaña en base a la incorporación del trabajo desarrollado.

## 2. Antecedentes Generales

En este capítulo se aborda el contexto de telefonía móvil en el mercado chileno, exponiendo los hitos más relevantes del último tiempo referente a la industria.

Adicionalmente, se describe la empresa de telecomunicaciones en la cual se trabajará junto con su participación de mercado en el rubro de prepago, explicando a grandes rasgos la caracterización de sus clientes.

### 2.1. Industria de telefonía móvil

La telefonía móvil corresponde a un sistema de comunicación para la transmisión de sonidos a larga distancia que permite hacer y recibir llamadas desde cualquier lugar, siempre que sea dentro del área de cobertura del servicio que lo facilita. En Chile, a

fin de marzo del 2019, este servicio cuenta con 25,7 millones de abonados, alcanzando una penetración de 135,3 abonados por cada 100 habitantes [1].

Uno de los cambios más significativos que ha experimentado el sector de telefonía móvil en el último tiempo corresponde a una constante reducción tarifaria de sus servicios. Por ejemplo, en el año 2019 el Ministerio de Transporte y Telecomunicaciones redujo en un 80% el costo de las tarifas que se pagan entre todas las compañías móviles y las llamadas desde teléfonos fijos a móviles, pasando de una tarifa promedio de \$8,7 por minuto a \$1,8.

Por otro lado, a partir del 2009 se ha evidenciado un aumento de la penetración de Internet Móvil, alcanzando en marzo del 2019 una penetración de 97,9 accesos por cada 100 habitantes, con un crecimiento anual de 7,2 puntos porcentuales.

De acuerdo al Ministerio de Transportes y Telecomunicaciones, la reducción tarifaria del sector junto a la alta penetración de Internet Móvil han provocado una tendencia creciente en la cantidad de clientes abonados por contrato, los cuales corresponden a clientes que pagan un cargo fijo constante mensualmente a cambio de minutos para llamar, mensajes y una cantidad de gigabytes para navegar por internet. A marzo del 2019, los clientes abonados por contrato representan el 48,6% del total de clientes que utilizan servicios móviles [1].

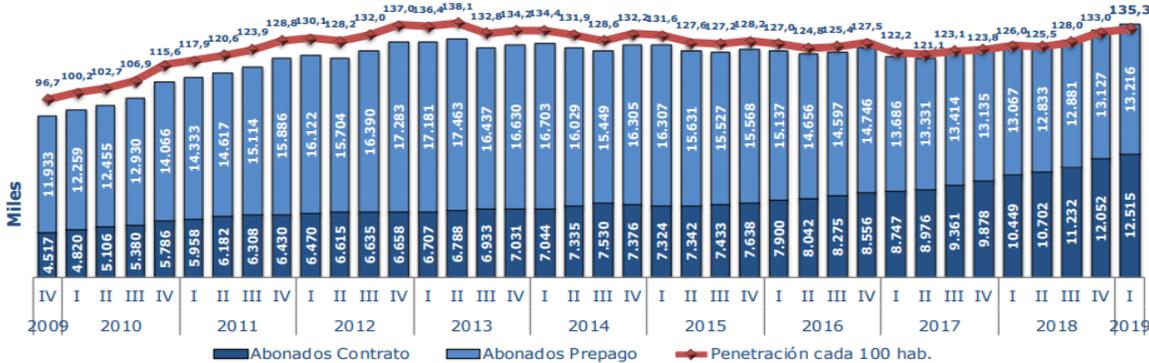


Figura 1: Evolución de clientes abonados por contrato

Fuente: Informe Trimestral del Sector Telecomunicaciones – Primer Trimestre 2019

## 2.2. La Empresa

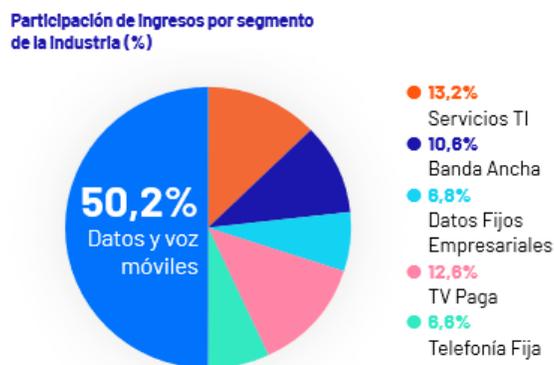
Entel Chile S.A es una empresa operadora líder en telecomunicaciones y tecnología en el país, con una creciente operación en Perú. La propuesta de valor de la compañía consiste en *“entregar experiencias de servicio simples y eficientes, a través de una infraestructura de primer nivel para comunicaciones móviles, fijas, y tecnologías de la información digitales, abriendo a nuestros clientes posibilidades infinitas”*[2].

En Chile, Entel ofrece una gama completa de servicios para sus clientes, los cuales se dividen en tres segmentos de mercado diferentes: Personas, Empresas y Corporaciones. Dichos servicios buscan atender desde necesidades cotidianas hasta procesos complejos, ya sea de negocios o instituciones a través de acompañamiento en transformación digital, contact center y outsourcing TI.

En particular, para los clientes personas, Entel ofrece los siguientes servicios:

- **Telefonía móvil:** El cual incluye servicio de llamado telefónico e internet móvil. El mercado de telefonía móvil posee dos grupos de clientes, clientes abonados por contrato o postpago, y clientes prepago, que corresponde a personas que no pagan un monto fijo a cambio de ciertos servicios, si no que recargan lo que estimen conveniente y utilizan este monto para poder llamar y/o comprar bolsas con el fin de navegar por internet.
- **Hogar:** Teléfono fijo, servicio de televisión e internet hogar, pudiendo ser este último internet de fibra óptica o banda ancha fija inalámbrica (BAFI).

La participación de mercado de Entel en la industria de telecomunicaciones a fines del 2018, en base a cálculos internos de la empresa, corresponde a un 24,1%. Teniendo una mayor participación (50,2%) en datos y voz móviles (Ver Figura 2).

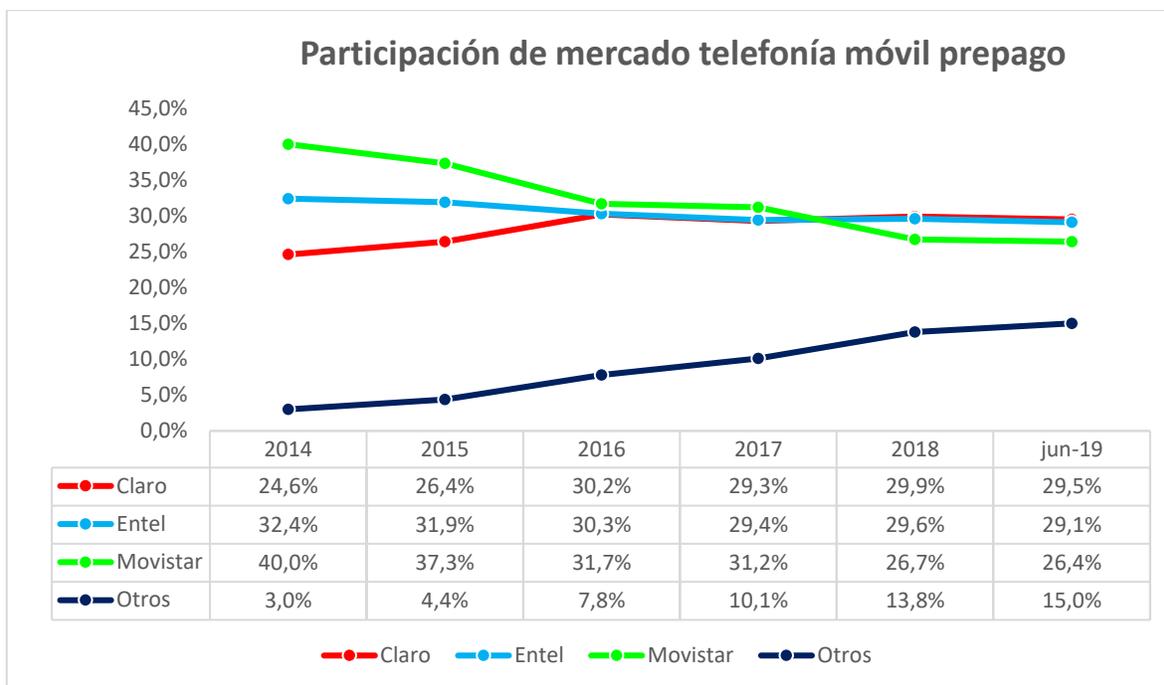


*Figura 2: Participación de ingresos Entel por segmento de la Industria de Telecomunicaciones*

*Fuente: Memoria corporativa Entel, 2018*

### 2.3. La Empresa en el mercado de prepago

A fines del 2018, Entel ocupa el segundo lugar en participación de mercado en telefonía móvil prepago respecto a la cantidad total de clientes existentes en el país [3], sin embargo, la compañía es líder en términos de ingresos recibidos por el sector [2]. A partir de la Figura 3, la cual muestra la evolución de la participación de mercado desde el 2014 en adelante, se evidencia un fuerte crecimiento de “Otras” compañías, compuesto principalmente por Wom, Virgin Mobile y VTR. La tendencia creciente de su participación se debe principalmente a la llegada de Wom al mercado de telefonía móvil a fines del 2014.



*Figura 3: Evolución de participación de mercado en telefonía móvil prepago*

*Fuente:* Elaboración propia en base a Informe Anual del Sector de Telecomunicaciones 2014, 2015, 2016, 2017, 2018 e Informe Semestral del Sector Telecomunicaciones 2019 Primer Semestre.

En la actualidad, Entel cuenta con más de 5 millones de móviles prepago, sin embargo, dicha cifra sobrestima la cantidad real de clientes que efectivamente utilizan su número, para ello, el mercado de prepago define los siguientes tipos de clientes:

- **Clientes activos:** Clientes que hayan realizado algún tipo de actividad durante los últimos 90 días, el cual puede ser realizar o recibir llamadas, o navegar por internet. Es importante mencionar que no se considera por sí sola la recarga como una actividad, es decir, un cliente que solo realiza recargas es considerado un cliente inactivo. Mensualmente, considerando los datos históricos del año 2019, la cantidad de móviles que cumplen dicha condición fluctúa entre 2 y 2.3 millones de móviles diferentes.
- **Clientes de calidad:** Corresponde a clientes que hayan realizado una recarga y algún tipo de actividad durante los últimos 30 días. La cantidad de móviles

que cumplen con la condición anterior mensualmente se encuentra entre 1.4 y 1.8 millones de teléfonos distintos.

Adicionalmente, la compañía clasifica a los clientes prepago en base al monto de recarga mensual que estos realizan, si este es mayor a \$8.000, corresponde a un cliente de alto valor, y en caso contrario, a uno de bajo valor.

Por último, también existe una clasificación de clientes prepago en relación al tipo de uso que le dan a su celular, para entender dicha definición, es necesario comprender los tipos de bolsas que los clientes pueden comprar:

- Bolsas de datos: Bolsas que entregan una cantidad de Gigabytes para navegar por internet por un período de tiempo determinado.
- Bolsas de voz: Otorgan una cantidad de minutos para realizar llamadas.
- Bolsas mixtas: Incluyen tantos minutos para llamar, como gigabytes para navegar.
- Bolsas roaming: Bolsas que permite realizar acciones estando en el extranjero.
- Promoción redes sociales: Al recargar un monto mayor o igual a \$3.000, es posible utilizar de forma gratuita las redes sociales por un determinado periodo de tiempo, el cual puede ser de 7 días si la recarga está entre \$3.000 y \$4.999, 15 días si la recarga se encuentra entre \$5.000 y \$9.999, y 30 días si el monto recargado es mayor o igual a \$10.000.

Teniendo en consideración lo anterior, existen tres segmentos de clientes prepago según el tipo de uso que estos realizan:

1. Clientes Redes Sociales (RRSS): Clientes que hacen uso de la promoción de redes sociales contra recargas, pero no compran bolsas mixtas ni de datos.
2. Clientes Datos: Personas que compran bolsas mixtas o solo de datos. Actualmente la única forma de navegar por internet, excluyendo las redes sociales, es mediante la compra de bolsas.
3. Clientes Voz: Clientes que no compran bolsas mixtas ni de datos, es decir, estas personas solamente compran bolsas de voz o utilizan directamente el monto de recarga para realizar llamadas.

Un cliente prepago puede transitar por 4 estados diferentes:

1. Cliente inactivo: corresponde a quienes no han realizado ningún tipo de actividad durante los últimos 3 meses, los cuales, posterior a este tiempo, dejan de ser considerados como clientes en la base final hasta una nueva reactivación.
2. Clientes activos.
3. Clientes portados a otra compañía (Port out): Corresponde a clientes que abandonan la compañía portándose a otro operador de la competencia.
4. Clientes migrados a postpago: Son clientes que optan por contratar un plan en el mercado de postpago. Un cliente puede migrar a postpago porque recibe una oferta de Entel mediante un llamado telefónico (vía call center), o porque lo decide por cuenta propia.

### 3. Descripción y justificación del proyecto

En el capítulo a continuación se detalla el trabajo que será realizado en la memoria, explicando la situación actual de las migraciones de clientes prepago hacia postpago y de que forma el trabajo genera valor para la empresa.

Se da énfasis en la importancia de incluir la fuga como un aspecto fundamental dentro del trabajo, considerando esta en la implementación de la solución del problema.

#### 3.1. Descripción del proyecto

El proyecto busca identificar los clientes prepago Entel más propensos a migrar hacia el mercado postpago en la misma compañía mediante la utilización de técnicas analíticas, con el objetivo de tomar acciones para facilitar su traspaso a este.

Junto con determinar los clientes propensos a migrar, se considerará que estos además posean una permanencia mínima en el mercado de postpago, con el fin de concentrar los esfuerzos en clientes que generarán un mayor valor a largo plazo para la empresa y que no se fugarán de forma temprana, período que se considera como dentro de los cuatro meses siguientes a la migración.

Mediante la identificación de los clientes con mayor probabilidad de migrar a suscripción, es posible enviar una base de datos al call center en la cual se priorice ofrecer el traspaso hacia pospago a clientes propensos tanto a migrar como a mantenerse en la compañía en dicho mercado, generando de esta forma un incremento sostenible en la base de pospago.

### 3.2. Justificación del proyecto

El trabajo resulta de gran relevancia para la empresa por diversas razones, en primer lugar, debido a la situación actual del mercado de telefonía móvil descrito en la sección anterior, el cual se caracteriza por el aumento de los abonados por contrato.

Dada la gran cantidad de clientes que posee Entel en el mercado de prepago, existe un gran potencial de estos que podrían migrar hacia pospago en el corto plazo, pudiendo ser dentro de la misma compañía o en otra empresa de telecomunicaciones. De acuerdo a un estudio de la consultora i-cuadrado [4], el 95% de los clientes prepago que se porta hacia otra compañía, opta por un plan en la empresa de destino.

Por otro lado, hoy en día Entel realiza una campaña mensual en la cual ofrece vía call center a un grupo de clientes prepago la opción de contratar un plan y así migrar hacia el mercado de pospago. Respecto a esta, actualmente no existe un modelo analítico para determinar a qué clientes dirigir la campaña, si no que la elección de estos se define en base a criterios comerciales, los cuales se describen a continuación:

- El cliente debe tener una antigüedad en la compañía de al menos 3 meses.
- El monto promedio de recarga durante los últimos 3 meses debe estar entre \$4.000 y \$30.000.
- Se excluyen clientes con deuda Entel o con riesgo financiero en base al Rut.
- No se envía la campaña a clientes que hayan migrado del mercado pospago a prepago durante los últimos 6 meses.
- Se excluyen clientes que hayan recibido la campaña en los últimos 2 meses.

En relación con lo anterior, existe un espacio de mejora con respecto a las campañas de migración, ya que, mediante la incorporación de inteligencia analítica, se podría mejorar la eficiencia de esta.

Otra forma de migrar de prepago a pospago es de forma espontánea, que corresponde a clientes que deciden por cuenta propia contratar un plan, ya sea visitando una tienda de Entel, mediante la página web o llamando directamente al

call center. En cuanto a este grupo de clientes, existe un desconocimiento respecto a qué características propias poseen qué gatillan finalmente en la decisión de migrar hacia pospago. Analizar este segmento de clientes, podría generar nuevos insights para facilitar y acelerar las migraciones dentro de la misma compañía, incluyendo estos en la campaña de migración.

Cuando un cliente de prepago migra hacia pospago, pueden ocurrir distintas situaciones en los meses posteriores:

1. Volver al mercado de prepago: El cliente tiene la opción de renunciar al plan y volver al mercado de prepago.
2. Portabilidad hacia otra compañía: Existe la opción de que el cliente decida irse de la compañía portándose a un operador de la competencia (port out).
3. Desactivación por renuncia: El cliente puede renunciar al plan y dar de baja el número.
4. Desactivación por Fraude: Consiste en una fuga involuntaria debido a un fraude por parte del cliente hacia la compañía.
5. Continuar en pospago: El mejor caso para la compañía es que el cliente se mantenga en el mercado de pospago.
6. Desactivación por cobranza: Si el cliente no paga el valor total de su plan durante dos meses, su número queda inhabilitado y este deja de ser considerado como un cliente pospago de la compañía, lo anterior ocurre al cuarto mes posterior a su ingreso. La Figura 4 muestra el momento desde el ingreso del cliente al mercado pospago hasta la desactivación por cobranza.

Cuando un cliente ingresa a pospago, el primer pago del monto total del plan elegido se factura en dos meses más. Por ejemplo, si un cliente ingresa a mediados de mayo del 2019, el primer pago lo realiza en el mes de junio, y el monto a pagar corresponde al valor del plan elegido proporcional a los días que utilizó este, siguiendo con el ejemplo descrito, si el cliente contrató el plan de \$9.990 y lo utilizó la mitad del mes de mayo, en junio del 2019 pagará la mitad del valor del plan, correspondiente a \$4.955.

Por lo tanto, recién en julio deberá pagar la primera cuota del monto real de su contrato. En caso de no pagar la boleta de julio ni agosto, el cliente será dado de baja del mercado pospago en el mes de septiembre.

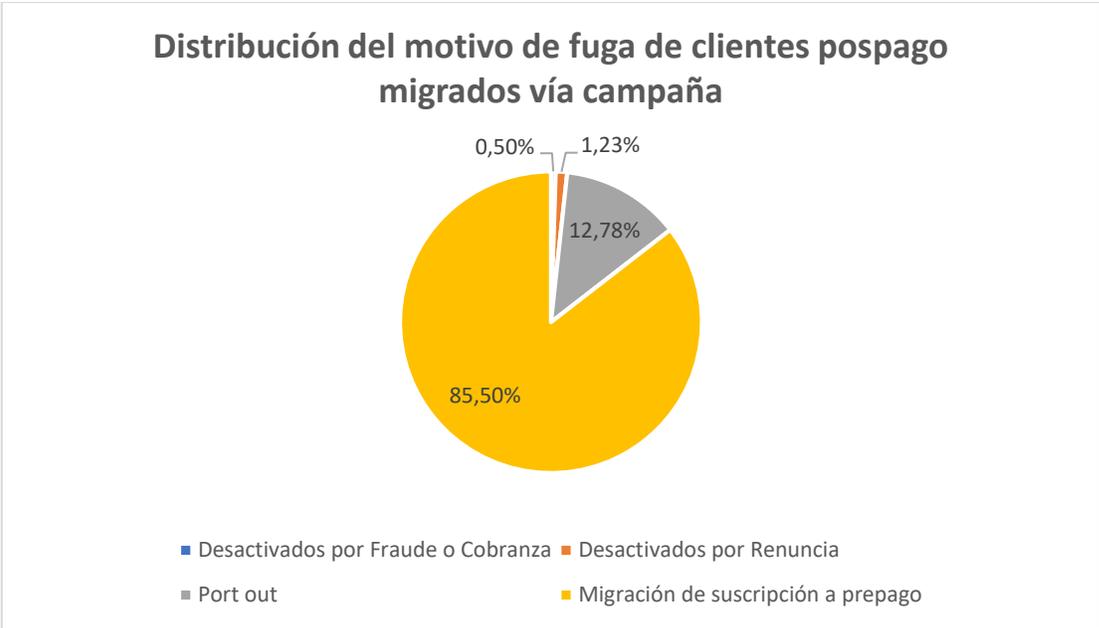
Si bien pasa un tiempo considerable desde el ingreso hasta que se elimina por completo al cliente, si este no paga no puede utilizar los servicios asociados al plan.



*Figura 4: Ejemplo de desactivación por cobranza*

*Fuente: Elaboración propia*

En la Figura 5, se aprecia que el 85,5% de los clientes que abandonan el mercado de suscripción vuelven hacia prepago, mientras que un 12,78% se porta hacia otra compañía durante los primeros cuatro meses. Las fugas asociadas a motivos de fraude y cobranza representan tan solo un 0,5% del total.



*Figura 5: Distribución de motivos de fuga para clientes migrados a suscripción vía campaña*

*Fuente:* Elaboración propia en base a clientes migrados vía campaña durante enero y abril del 2019

Un aspecto importante de las migraciones mediante la campaña de migración corresponde a la tasa de fuga que poseen estos usuarios, entendiendo esta como clientes que abandonan el mercado de pospago.

Para cuantificar la tasa de fuga al cuarto mes posterior al ingreso, se analizaron los clientes ingresados vía campaña desde enero a marzo del 2019. A partir de la Figura 6, se observa que al cuarto mes posterior a la migración (Mes 4), se alcanza una fuga acumulada superior al 16%. Por ejemplo, para los clientes ingresados en enero del 2019, la tasa de fuga acumulada en el Mes 4 (mayo del 2019) es del 16,6%, lo cual significa que, a fines de mayo del 2019, se mantienen en el mercado de pospago solamente el 83,4% de los clientes migrados inicialmente (enero del 2019).

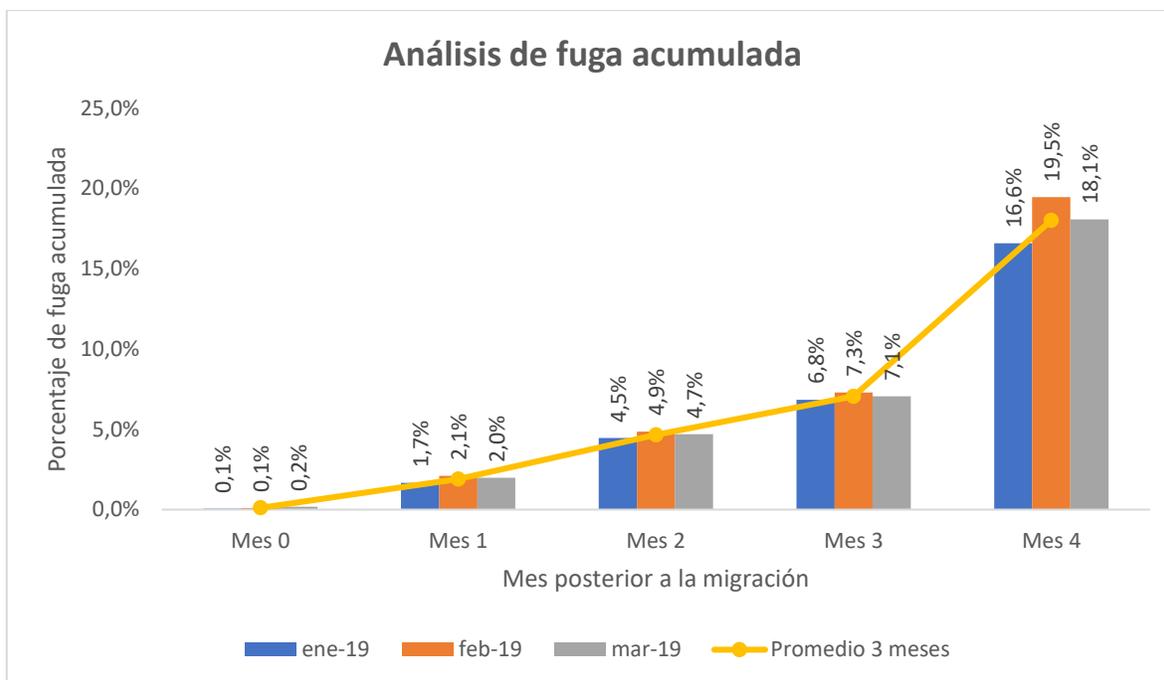


Figura 6: Análisis de fuga acumulada para clientes migrados en enero y marzo del 2019

Fuente: Elaboración propia en base a clientes migrados vía campaña durante enero y febrero del 2019

Es por esto que el trabajo de memoria busca migrar a clientes cuya permanencia esperada en el mercado de pospago sea de al menos 4 meses posterior a la migración, de esta forma, se evitar migrar a clientes que sean propensos a no pagar el cargo fijo asociado al plan, o a clientes probables de renunciar a su contrato tempranamente, ya sea portándose a otra compañía, renunciado a su contrato, o volviendo al mercado de prepago.

### 3.3. Trabajo a realizar

Para cumplir con la descripción del proyecto, se trabajará en la creación de dos modelos, uno para determinar los clientes más propensos a migrar vía campaña, y otro modelo para determinar los clientes con una mayor probabilidad de migrar hacia pospago de forma espontánea.

El modelo de migración vía campaña solo considera los clientes que son gestionados por esta, por lo tanto, no considera a todos los clientes del mercado prepago. A modo de ejemplo, no incluye clientes que lleven 1 o 2 meses en la compañía, ni tampoco a

clientes cuyo monto promedio de recarga en los últimos tres meses sea menor a \$4.000.

Por otro lado, el modelo de migración espontánea se encarga de incluir a los clientes que no cumplen con los filtros para ser gestionados y aun así migran hacia pospago, por lo que la unión de ambos considera el grupo total de clientes prepago que mensualmente migra hacia suscripción.

El objetivo del trabajo es analizar ambos modelos y en conjunto lograr determinar la base de clientes con mayor probabilidad de migrar hacia el mercado de pospago, de esta forma, se logrará enviar a campaña un grupo de clientes con una mayor tasa de aceptación esperada. Por ejemplo, para un cliente con dos meses de antigüedad y un monto promedio de \$3.500 en dicho período, el modelo de migración vía campaña podría entregar una baja propensión a migrar dado que no cumple con los filtros actuales de esta, sin embargo, el modelo de migración espontánea podría entregar que este cliente es altamente propenso a migrar, por lo que efectivamente si sería conveniente ofrecerle la campaña a este número.

Es decir, en caso de que un cliente posea una alta probabilidad de migrar de forma espontánea y este no cumpla con los filtros actuales de la campaña, también se incorporará en la gestión de esta con el objetivo de acelerar su migración hacia el mercado de suscripción. Por lo tanto, la base resultante de clientes a los cuales ofrecer la campaña de migración se obtendrá luego de analizar y combinar los dos modelos descritos anteriormente.

A partir del trabajo realizado será posible gestionar a un grupo de clientes prepago que genera mayor valor para la compañía, puesto que, junto con ser propensos a cambiarse al mercado de pospago, también serán propensos a no fugarse en el corto plazo de la compañía. Ambos puntos se alinean con los objetivos de la compañía dentro del mercado pospago, que corresponden al aumento de una base robusta y sostenible de los clientes pospago.

## 4. Objetivos

### 4.1. Objetivo General

Incrementar la migración de clientes prepago hacia el mercado de pospago mediante la incorporación de modelos analíticos para una empresa de telecomunicaciones.

## 4.2. Objetivos Específicos

- Determinar las variables más relevantes que identifican a los clientes propensos a migrar hacia un plan por suscripción.
- Construir un modelo para determinar clientes con mayor probabilidad de migrar vía campaña.
- Construir un modelo para identificar los clientes con mayor propensión a migrar de forma espontánea.
- Generar una metodología que defina mensualmente la base de clientes a los cuales dirigir la campaña.
- Evaluar el efecto en rentabilidad generado por la incorporación del modelo para perfilar la base de clientes a los cuales dirigir la campaña.
- Proponer cambios respecto al funcionamiento actual de las campañas de migración.

## 5. Alcances y resultados esperados

### 5.1. Alcances

Tal como se describió en los ítems anteriores, el trabajo de título busca migrar al mercado de pospago a clientes con una alta probabilidad de mantenerse en este al cuarto mes posterior a la migración, sin embargo, por temas de tiempo, no será posible verificar si la tasa de fuga de los clientes migrados y perfilados por el modelo poseen una menor tasa de fuga.

### 5.2. Resultados esperados

A partir del trabajo de título se esperan los siguientes resultados:

1. Extracción automática de base datos necesaria para correr modelos: En conjunto con el área de Big Data, se trabajará con el fin de generar todas las variables que utiliza el modelo de forma automática una vez al mes, lo anterior facilita el proceso considerando que las variables se obtienen de distintas plataformas.

2. Modelo analítico para predecir migración vía campaña: El cual incluye el script que permite pre procesar la información y ejecutar el modelo.
3. Modelo analítico para predecir migración espontánea: Al igual que en el caso anterior, este incluye el script que permite ejecutar el modelo.
4. Metodología para determinar la base óptima de clientes a los cuales perfilar la campaña mensualmente mediante la combinación de ambos modelos: A partir de todo el trabajo previo, el resultado esperado más importante consiste en identificar los clientes a los cuales dirigir la campaña con el objetivo de maximizar la cantidad de migraciones.
5. Realizar laboratorio para evaluar resultados del modelo: Una vez obtenido los resultados anteriores, se realizará un laboratorio con el fin de evaluar los resultados de los modelos desarrollados.
6. Propuesta de cambio en la gestión actual de campaña: En base a los resultados del laboratorio, se propondrá una nueva forma de gestionar y definir los clientes a los cuales ofrecerles la campaña de migración.

## 6. Marco teórico

En esta sección se describen los aspectos técnicos relevantes del trabajo, detallando y explicando la elección de la metodología a utilizar (CRISP-DM), los algoritmos de clasificación que serán desarrollados, las métricas utilizadas para evaluar la performance de los distintos modelos probados y la técnica utilizada para la detección y eliminación de outliers.

### 6.1. Metodología CRISP-DM

Sus siglas en inglés significan *Cross Industry Standard Process for Data Mining*, la metodología CRISP-DM se utiliza para trabajar con técnicas de Data Science o Data Analytics, siendo una de las más utilizadas actualmente para el desarrollo de proyectos analíticos [5].

Dado que Entel cuenta con una gran cantidad de clientes prepago y el objetivo principal del trabajo de memoria consiste en la creación de un modelo de clasificación, la metodología CRISP-DM se adecúa al trabajo de título.

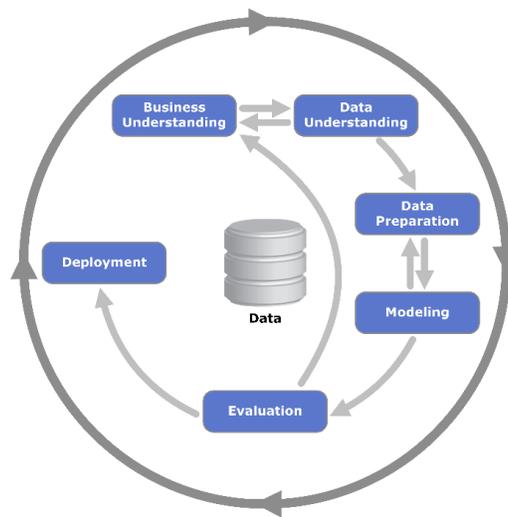
Al comparar esta metodología con otras utilizadas para proyecto de Minería de Datos como KDD o SEMMA, la principal diferencia consiste en que la metodología CRISP-DM posee mayor énfasis en analizar la solución de acuerdo al contexto de la situación, lo cual permite reevaluar y reestructurar el modelo en base al contexto del problema (Ver etapa de evaluación, Figura 7), por ejemplo, adecuarse a nuevas restricciones o reglas de gestión que no se tenían en consideración en primera instancia.

A continuación se describen los pasos que sigue esta metodología [6] [7]:

1. **Comprensión del negocio (Business Understanding):** Fase inicial que se enfoca en comprender los objetivos y requerimientos del proyecto.
2. **Comprensión de los datos (Data Understanding):** Incluye actividades que permiten familiarizarse con los datos, descubrir conocimiento preliminar sobre los datos, y/o encontrar subconjuntos interesantes para formar hipótesis en cuanto a la información oculta.
3. **Preparación de Datos (Data Preparation):** La fase de preparación de datos cubre todas las actividades necesarias para construir el conjunto final de datos (los datos que se utilizarán en las herramientas de modelado) a partir de los datos en bruto iniciales. Las tareas incluyen la selección de tablas, registros y atributos, así como la transformación y la limpieza de datos para las herramientas que modelan.
4. **Modelación (Modeling):** En esta fase, se seleccionan y aplican las técnicas de modelado que sean pertinentes al problema (cuantas más mejor), y se calibran sus parámetros a valores óptimos. Típicamente hay varias técnicas para el mismo tipo de problema de minería de datos y algunas técnicas tienen requerimientos específicos sobre la forma de los datos. Por lo tanto, casi siempre en cualquier proyecto se acaba volviendo a la fase de preparación de datos.
5. **Evaluación (Evaluation):** En esta etapa del proyecto, se han construido uno o varios modelos que parecen alcanzar una calidad suficiente desde una perspectiva de análisis de datos. Antes de proceder al despliegue final del modelo, es importante evaluarlo y revisar los pasos ejecutados para crearlo,

comparando el modelo obtenido con los objetivos del negocio. Un objetivo clave es determinar si hay algo importante del negocio que no haya sido considerado completamente. Al final de esta fase, se debería obtener una decisión sobre la aplicación de los resultados del proceso de análisis de datos.

6. Despliegue (Deployment): Generalmente, la creación del modelo no es el final del proyecto. Incluso si el objetivo del modelo es de aumentar el conocimiento de los datos, el conocimiento obtenido tendrá que organizarse y presentarse para que el cliente pueda usarlo. Dependiendo de los requisitos, la fase de desarrollo puede ser tan simple como la generación de un informe o tan compleja como la realización periódica y quizás automatizada de un proceso de análisis de datos en la organización.



*Figura 7: Pasos metodología CRIPS-DM*

*Fuente: Imagen obtenida de la página sngular.com*

## 6.2. Algoritmos de clasificación

Corresponden a técnicas utilizadas en minería de datos que provienen del área de Machine Learning, estas permiten clasificar los datos en dos o más clases en diversos contextos e industrias. A modo de ejemplo, los algoritmos se pueden usar para predecir si un cliente abandonará la compañía en el próximo periodo [8].

A continuación, se describirán los modelos de clasificación que serán utilizados en el trabajo de memoria:

- **Árbol de decisión:** Corresponde a un método de aprendizaje supervisado utilizado para clasificación y regresión. El objetivo del algoritmo es crear un modelo que prediga el valor de una variable objetivo mediante el aprendizaje de reglas de decisión simples inferidas de las características de los datos [9].

El árbol de decisión implica la partición de los datos en subconjuntos que contienen instancias con valores similares (homogéneos). Una métrica para determinar la homogeneidad de los datos es la entropía, que toma el valor 0 si la muestra es completamente homogénea y 1 si la muestra está igualmente dividida respecto a la variable de interés a predecir.

La construcción de un árbol de decisión consiste en ir encontrando los atributos que al dividir un nodo devuelvan la mayor ganancia de información, es decir, que los nodos resultantes sean los más homogéneos posibles en relación a su entropía [9] .

- **Random Forest:** Algoritmo que consiste en una gran cantidad de árboles de decisión que operan como un conjunto, es decir, cada árbol realiza una predicción para cada dato, y la clase con mayor cantidad de predicciones corresponde a la predicción final del modelo [10].

Para cada árbol de decisión probado, se escoge un subconjunto aleatorio de atributos del total disponible, evitando de esta forma una correlación alta entre los árboles probados, lo cual permite realizar predicciones más precisas que un árbol de decisión por sí solo.

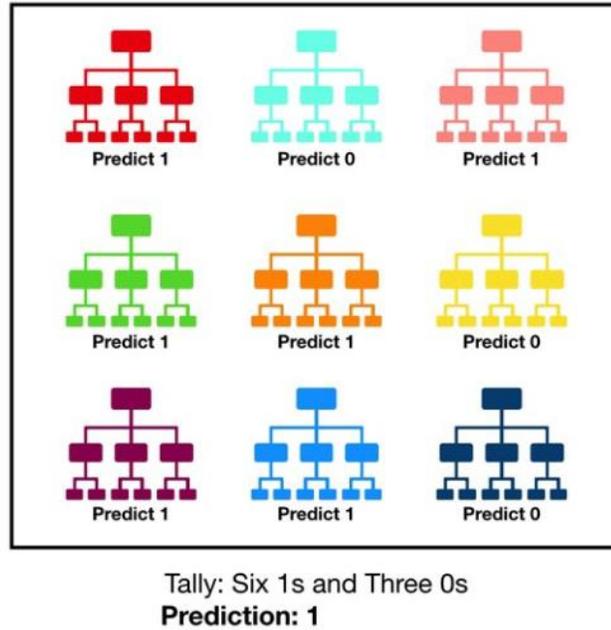


Figura 8: Ejemplo del algoritmo Random Forest

Fuente: Imagen obtenida de la página [towardsdatascience.com](https://towardsdatascience.com)

- **Gradient Boosting Machine (GBM):** Corresponde a un algoritmo que entrena árboles de decisión de manera gradual, aditiva y secuencial. Este modelo de clasificación posee la particularidad de ir aprendiendo de los errores previos de los datos mal clasificados, por lo que, en una siguiente iteración, pone particular atención en dichos datos (boosting).

Para ello, el algoritmo GBM identifica las deficiencias de un modelo mediante el uso de gradientes en la función de pérdida, siendo esta una medida que indica que tan buenos son los coeficientes del modelo para ajustar los datos subyacentes. [11].

- **Extra Gradient Boosting (XGBoost):** Algoritmo reciente publicado en el año 2016, desde su introducción, además de haber ganado numerosas competiciones de Kaggle, este algoritmo ha sido pionero en aplicaciones industriales de vanguardia.

El algoritmo XGBoost optimiza el algoritmo GBM estándar mediante la incorporación de procesamiento paralelo, poda de árboles (tree-pruning),

técnicas para lidiar de mejor forma con valores nulos y regularización para evitar el sobreajuste del modelo [12].

### 6.3. Métricas de evaluación

Existen diversas métricas que permiten evaluar y comparar la performance de distintos modelos, la elección de que métrica es mejor depende del problema en estudio y el objetivo que se busca con este. Para ello, es importante tener en consideración dos aspectos, los cuales se describen a continuación [13]:

1. Métrica robusta ante data no balanceada: En situaciones donde la clase positiva es significativamente menor a la clase negativa, es conveniente seleccionar una métrica que no incluya los casos donde el modelo predice correctamente la clase negativa.
2. Evaluación de métrica: Cada modelo entrega una probabilidad a cada registro de pertenecer a una cierta clase, teniendo esto en consideración, es importante determinar si se utilizarán las probabilidades obtenidas por el modelo, o se transformarán estas en asignar una clase a cada registro. Por ejemplo, si el objetivo del modelo es predecir si un cliente se fugará, resulta útil tomar la probabilidad predicha y transformarla en clase, dividiendo la población en clientes que se fugarán versus clientes que no lo harán, por otro lado, si se quiere predecir la pérdida esperada de ingresos, es mejor utilizar las probabilidades predichas.

A continuación, se definen métricas para la evaluación y comparación de modelos.

- Matriz de confusión [9]:

Herramienta que permite visualizar el desempeño de un algoritmo de clasificación. Esta matriz introduce cuatro conceptos que son utilizados en el cálculo de otras métricas.

- Verdaderos positivos: Casos en donde el modelo predice clase positiva y el valor real también corresponde a esta categoría.
- Falsos negativos: Casos donde el modelo predice clase negativa siendo esta positiva.

- Falsos positivos: Casos donde el modelo predice positivo pero el valor real de la clase es negativo.
- Verdaderos negativos: Casos donde el valor predicho y real es negativo.

		CLASE REAL	
		+	-
CLASE PREDICHA	+	Verdaderos positivos (VP)	Falsos positivos (FP)
	-	Falso Negativos (FN)	Verdaderos Negativos (VN)

*Tabla 1: Matriz de confusión*

- Coeficiente de correlación de Matthews (MCC):

El coeficiente MCC provee una medida para evaluar que tan bien un clasificador binario detecta verdaderos y falsos positivos, y verdaderos y falsos negativos. Este valor indica cuan correlacionado se encuentra el valor actual con el predicho por el modelo; 1 indica un clasificador perfecto, -1 un clasificador que predice la clase opuesta del valor real, y 0 significa que el clasificador no funciona mejor que asignar de forma aleatoria.

El coeficiente MCC es especialmente recomendado en casos donde existe desbalanceo de clases.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

*Ecuación 1: MCC Absoluto*

- Coeficiente Lift:

El coeficiente lift compara la probabilidad del fenómeno estudiado en un grupo en particular respecto al promedio de la población, por ejemplo, si el objetivo del modelo es predecir la tasa de fuga de una empresa que corresponde al 1% mensual, y el mejor modelo indica que en el decil más propenso a abandonar la compañía la tasa de fuga es de 7%, este grupo posee un lift de  $0,07/0,01 = 7$ , es decir, dicho grupo es 7 veces más probable a abandonar la compañía.

$$Lift = \frac{\text{Tasa predicha}}{\text{Tasa promedio}}$$

*Ecuación 2: Lift*

La curva de lift se grafica dividiendo a la población en percentiles en base a la propensión del fenómeno analizado, siendo los clientes del percentil uno los que poseen una mayor probabilidad de este, siguiendo con el ejemplo anterior, las personas pertenecientes al primer percentil son los clientes con una mayor probabilidad de fugarse de la empresa.

Los gráficos de lift utilizados corresponde al lift por percentil y lift acumulado hasta un cierto porcentaje de la población, ambos, debido a lo explicado anteriormente, poseen un comportamiento decreciente.

Esta métrica ha sido utilizada de forma exitosa en prácticas de marketing debido a que permite priorizar la gestión de los clientes poniendo esfuerzos solamente en un segmento de la población. Por ejemplo, si se desea entregar un descuento a los clientes para evitar su fuga, es óptimo ofrecer este solamente a los clientes con mayor probabilidad de abandono. Finalmente, el lift entrega buenos resultados cuando existe un desbalanceo de clases [14].

- Curva ROC (Receiver operating characteristic):

La curva ROC es una herramienta utilizada para clasificar la capacidad discriminante de una prueba diagnóstica dicotómica.

Esta capacidad depende del valor umbral elegido de entre todos los posibles resultados de la variable de decisión, es decir, la variable por cuyo resultado se clasifica a un individuo en un grupo u otro.

La curva representa, para cada valor umbral, las medidas de sensibilidad (tasa de verdaderos positivos) y especificidad de la prueba diagnóstica (1 - tasa de falsos positivos)[13].

Por un lado, la sensibilidad cuantifica la proporción de individuos que presenta el evento de interés y son clasificados por la prueba como portadores de dicho evento. Por otro lado, la especificidad cuantifica la proporción de individuos que no lo presentan y son clasificados por la prueba como tal.

El área bajo la curva ROC (AUC) se utiliza para evaluar la performance del modelo, mientras más cercana a 1 sea, mejor performance tendrá el modelo.

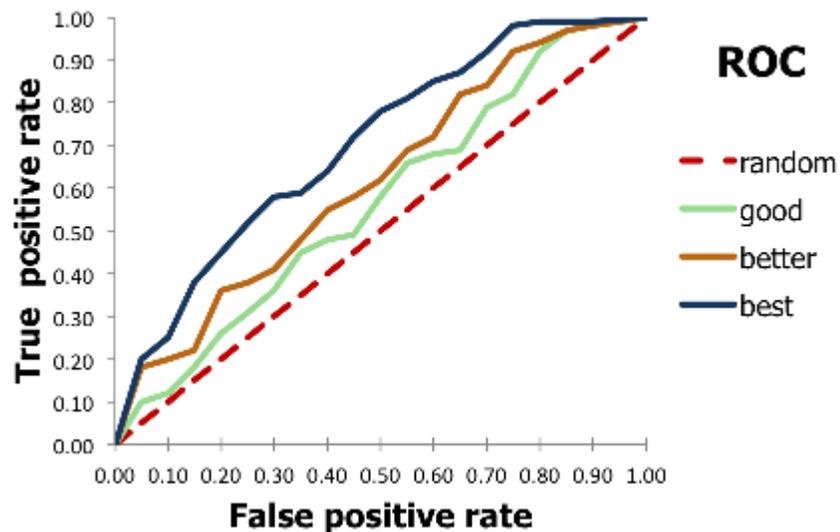


Figura 9: Ejemplo de curvas ROC

Fuente: Imagen obtenida de la página [towardsdatascience.com](https://towardsdatascience.com)

## 6.4 Técnicas de tratamiento de outliers

Los outliers corresponden a observaciones con características diferentes de las demás. Su principal problema radica en que son elementos que pueden no ser representativos de la población pudiendo distorsionar seriamente el comportamiento de los contrastes estadísticos.

Existen diversos métodos para tratar los datos atípicos dentro de la limpieza de los datos, siendo uno de los más conocidos y utilizados el rango intercuartílico. La detección de outliers de este método se basa en analizar los cuartiles de la variable en estudio.

Según dicho método es posible distinguir dos tipos de outliers, leves y extremos. Una observación es declarada como outlier extremo si esta cae fuera del intervalo  $(Q_1 - 3IQR, Q_3 + 3IQR)$ . Donde  $IQR = Q_3 - Q_1$  corresponde al rango intercuartílico. Así mismo, una observación es declarada como un outlier leve si cae fuera del intervalo  $(Q_1 - 1,5IQR, Q_3 + 1,5IQR)$ , donde los valores 3 y 1,5 son escogidos por comparaciones con una distribución normal.

## 7. Desarrollo de la metodología

El desarrollo metodológico describe todos los pasos de la metodología CRISP-DM, describiendo y entendiendo detalladamente el negocio, lo cual involucra el análisis de las campañas de migración vía call center, la comprensión y preparación de datos a utilizar para la resolución del problema, y la modelación y evaluación de los algoritmos utilizados.

En base a los resultados obtenidos, se explica el perfil de los clientes más propensos a migrar vía campaña y espontánea según las variables más importantes de los modelos.

Finalmente, se explica de qué forma serán utilizados los modelos para evaluar sus resultados en la gestión de la campaña de migración para el mes de diciembre.

## 7.1. Conocimiento del negocio

### 7.1.1. Campaña de migración

Dentro de la empresa de telecomunicaciones, se define como migración una transición de un cliente de un mercado a otro, pudiendo ser este prepago o pospago.

Para el contexto del trabajo de memoria, la migración de interés corresponde a la transición de un cliente prepago Entel hacia el mercado de pospago. Dicha migración puede ocurrir en la actualidad mediante dos formas, vía campaña de migración o de forma espontánea.

Tal como se explicó en la justificación del proyecto, mensualmente Entel ofrece vía call center a un grupo de clientes prepago la opción de contratar un plan, y la elección de estos clientes se define en base a reglas del negocio puesto que no existe un modelo analítico para ayudar en la toma de esta decisión. Las migraciones vía campaña del call center se denominan también como migraciones outbound.

Cada mes se envía una base de aproximadamente 177.000 clientes prepago al call center para ofrecerles la campaña, la cual, como se muestra en la Tabla 2, además de contener el móvil, posee información del cliente, tales como el Rut, el monto promedio de recarga durante los últimos 3 meses (PROM\_REC\_3), y la cantidad de veces que el cliente ha sido cargado en la campaña durante los últimos 6 meses (Q\_CARGAS).

En base a esto último, se divide el grupo en dos bases de clientes, los clientes B1, que corresponden a clientes que no han sido cargados en las campañas durante los últimos 6 meses (o nunca), y los clientes B2, quienes han estado en 1 o más campañas durante el periodo de tiempo descrito.

<b>MOVIL</b>	<b>RUT</b>	<b>PROM_REC_3</b>	<b>SUGERIDO 1</b>	<b>SUGERIDO 2</b>	<b>Q_CARGAS</b>	<b>B1 B2</b>
98432XXX	11692XXX	\$4.500	\$7.990	\$9.990	0	B1
83126XXX		\$16.000	\$19.990	\$17.990	3	B2

*Tabla 2: Ejemplo de variables enviadas al call center para campaña de migración*

Esta información se entrega al call center con el objetivo de que se priorice llamar primero a los clientes B1, dado que son clientes que presentan una mayor tasa de

aceptación de la campaña. Por otro lado, la base B1 se recorre una mayor cantidad de veces para los números que aún no han sido contactados.

Como se muestra en la Figura 10, la tasa de migración de clientes B1 es del 9,8% en comparación a un 5,4% para los clientes de base B2.

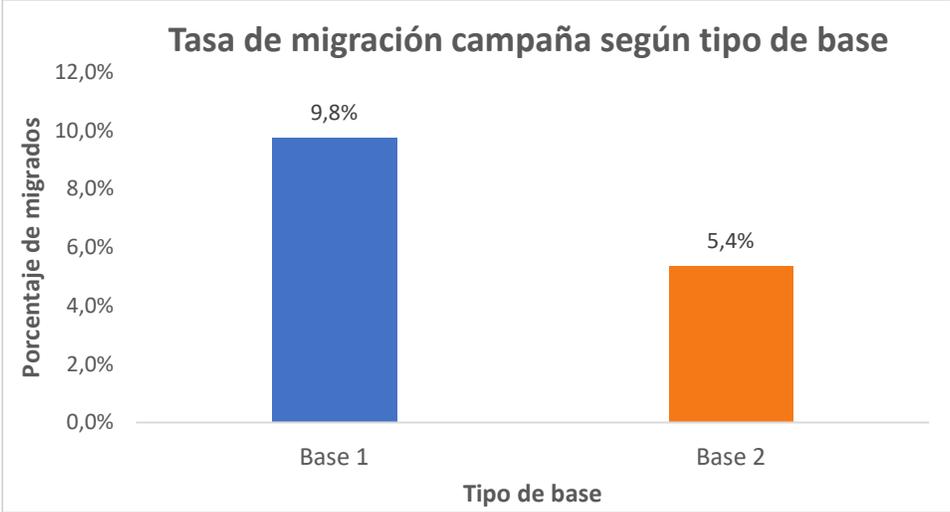
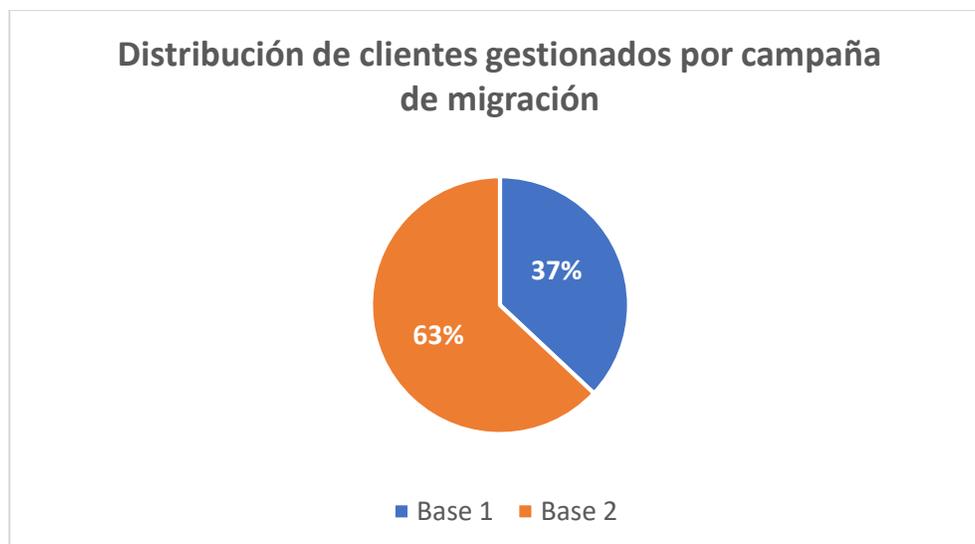


Figura 10: Tasa de migración vía campaña según tipo de base

Fuente: Elaboración propia en base a campañas de migración entre los meses de enero y abril del 2019.

A partir de la Figura 11, se observa que el 37% de los clientes que se envían a campaña corresponden a clientes B1, mientras que el 63% restante al segmento B2.



*Figura 11: Distribución de clientes gestionados por campaña.*

*Fuente: Elaboración propia en base a campañas de migración entre los meses de enero y abril del 2019.*

Por lo demás, es importante notar que en la base enviada al call center se sugieren dos planes para cada móvil, los cuales dependen del monto promedio recargado por el cliente en los últimos 3 meses, por ejemplo, de la Tabla 2 se observa que al cliente cuyo monto promedio de recarga es de \$4.500, se le sugiere el plan de \$7.990 y \$9.990. El plan de \$7.990 es exclusivo para la campaña de migración.

Los trabajadores del call center, encargados de contactar al cliente y ofrecerles un plan reciben \$4.500 por cada migración lograda, siendo este monto independiente del plan escogido por el cliente, es decir, para un ejecutivo es indiferente migrar a un cliente a un plan de \$7.990 que a uno de \$19.990. Lo anterior genera que no necesariamente el ejecutivo ofrezca el plan sugerido por la campaña, si no que este podría ofrecer uno más barato.

### 7.1.2. Análisis de efectividad de Rut

Dado que los clientes prepago no poseen un contrato con la compañía, los datos asociados al cliente que la empresa posee son menos y poco confiables, por ejemplo, no se posee información del Rut de todos los clientes prepago.

Del total de clientes que se gestionan por la campaña de migraciones solo se tiene un Rut asociado al 67% del total de clientes, y, de estos, existen Rut duplicados, por lo

que solo el 65% del total de registros posee Rut únicos, tal como se muestra en la Figura 12.

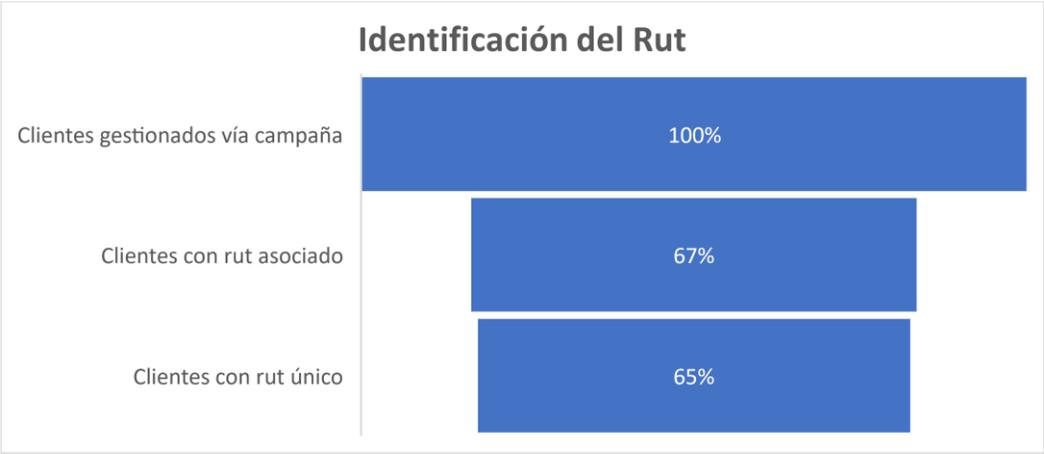


Figura 12: Análisis de Rut para clientes gestionados por campaña.

Fuente: Elaboración propia en base a clientes gestionados por campaña entre enero y abril del 2019

Si se analizan los clientes ingresados por campaña entre enero y abril del 2019, tan solo el 75% de los números tiene asociado un Rut, de estos, existen Rut asociados a más de un número, por lo que solo el 71% de los números posee un Rut único.

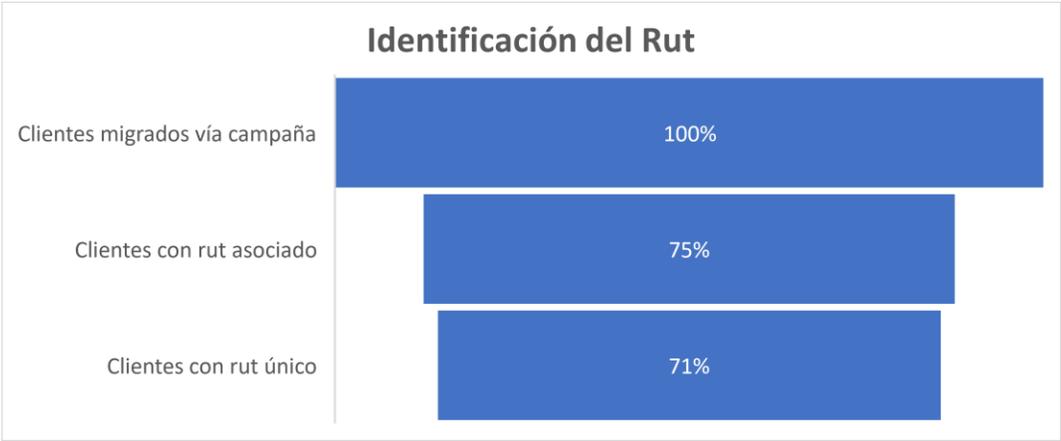


Figura 13: Análisis de clientes ingresados vía campaña con Rut asociado

Fuente: Elaboración propia en base a clientes ingresados a campaña entre enero y abril del 2019

Ahora bien, al comparar el Rut asociado en el mercado de prepago con el Rut del cliente que contrata el plan, este coincide solamente en un 53,4% del total de los casos.

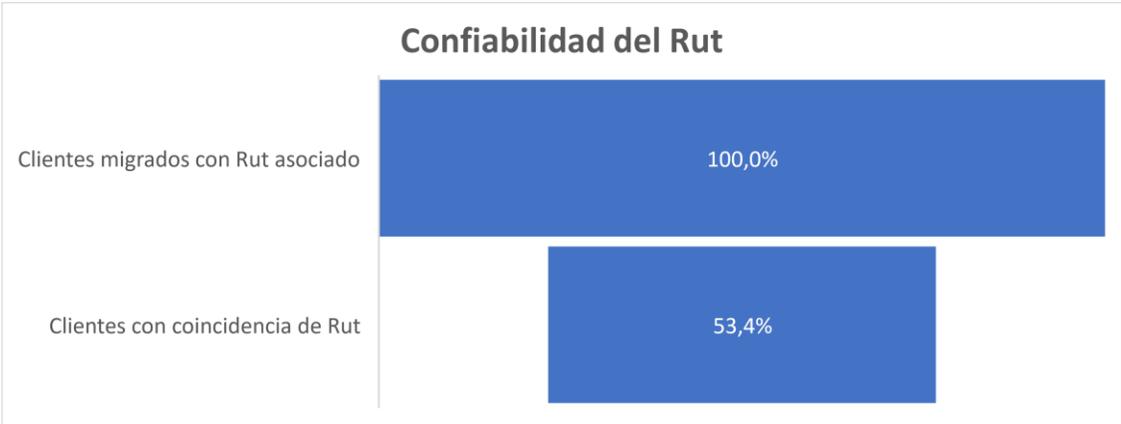


Figura 14: Confiabilidad del Rut

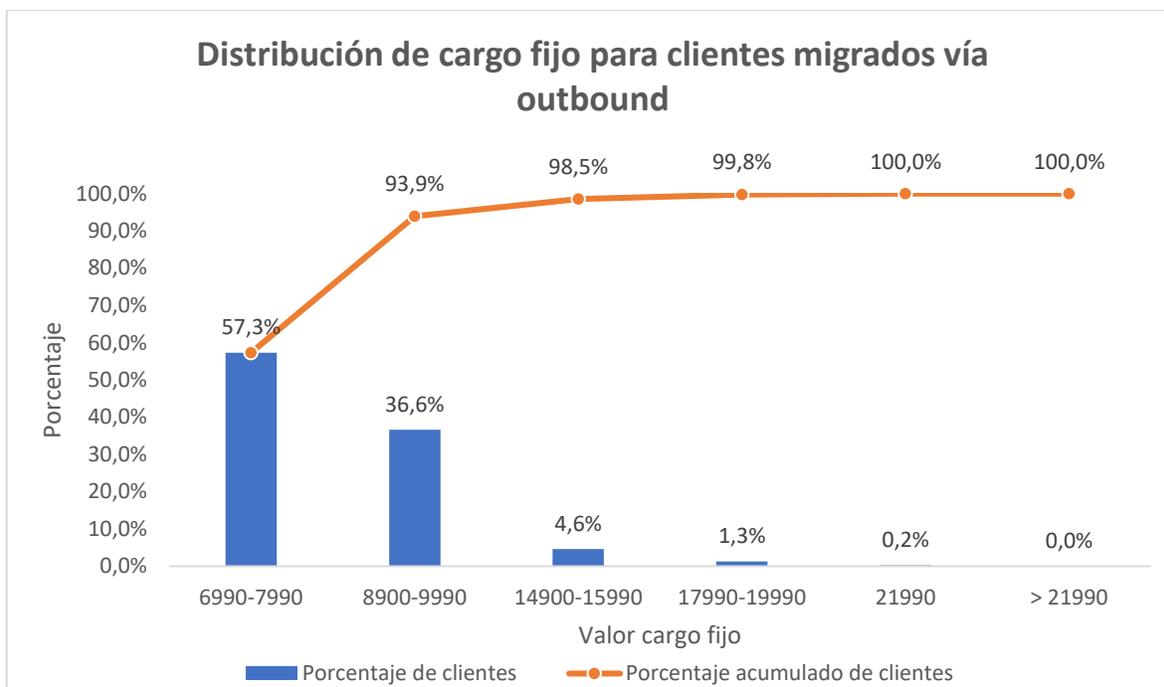
Fuente: Elaboración propia en base a clientes ingresados a campaña entre enero y abril del 2019

En base a los análisis descritos, los criterios de exclusión de la campaña que se basan en el Rut del cliente presentan un grado de error importante.

### 7.1.3. Análisis de rentabilidad de campaña

Tal como se comentó en la sección 5.1.1, los ejecutivos del call center reciben \$4.500 por cada migración lograda, siendo esto independiente de a qué cargo fijo se migre el cliente, es por esto que en reiteradas ocasiones no se respetan los planes que se sugieren en la campaña. Por ejemplo, para un cliente que tiene un monto promedio de recarga de \$16.000 en los últimos 3 meses, los planes sugeridos corresponden al de \$19.990 y \$17.990, sin embargo, un ejecutivo podría migrar a este cliente al plan de \$7.990, reduciendo a la mitad el valor que genera el cliente para la compañía.

A partir de la Figura 15, se observa que el 57,3% de los clientes migrados vía outbound migran hacia el plan cuyo valor se encuentra entre \$6.990 y \$7.990 (el valor oficial del plan es \$7.990, sin embargo, bajo ciertas condiciones de oferta especial, este queda a \$6.990). Además, del total de migraciones por este medio, el 93,9% de los clientes migran a un plan menor o igual a \$9.990.



*Figura 15: Distribución de cargo fijo para migraciones outbound*

*Fuente:* Elaboración propia en base a clientes migrados a pospago entre febrero y abril del 2019

Para cuantificar la rentabilidad de la campaña se realizaron dos análisis. El primero corresponde a analizar el margen porcentual que se genera de migrar a un cliente vía campaña hacia el mercado de pospago. Esta métrica representa la ganancia o pérdida de migrar al cliente hacia suscripción comparando el cargo fijo que elige, versus su monto promedio de recarga en los últimos 3 meses.

$$\text{Margen porcentual} = \frac{\text{Cargo Fijo} - \text{Promedio de recarga en los últimos 3 meses}}{\text{Cargo Fijo}}$$

*Ecuación 5: Margen porcentual*

Para realizar lo anterior, se ramificó el margen porcentual en siete categorías tal como se aprecia en la Figura 16. Tres categorías representan márgenes estrictamente negativos (menor a -30%, entre [-30%, -15%] y entre [-15%, -5%]), y análogamente, tres corresponden a márgenes positivos ([5,15%], [15%, 30%], > 30%). Por último, la categoría ± 5% incluye clientes cuyo margen se encuentra entre -5% y 5%, es decir,

corresponde a una categoría donde en promedio el margen porcentual es nulo en relación con la situación previa a la migración.

A partir de la Figura 16, se observa que en un 25% de los clientes se tiene un margen porcentual estrictamente negativo, siendo un 13% de estos clientes con un margen porcentual menor al 30%. Por otro lado, el 38% de los clientes migrados posee un margen porcentual mayor al 30%.

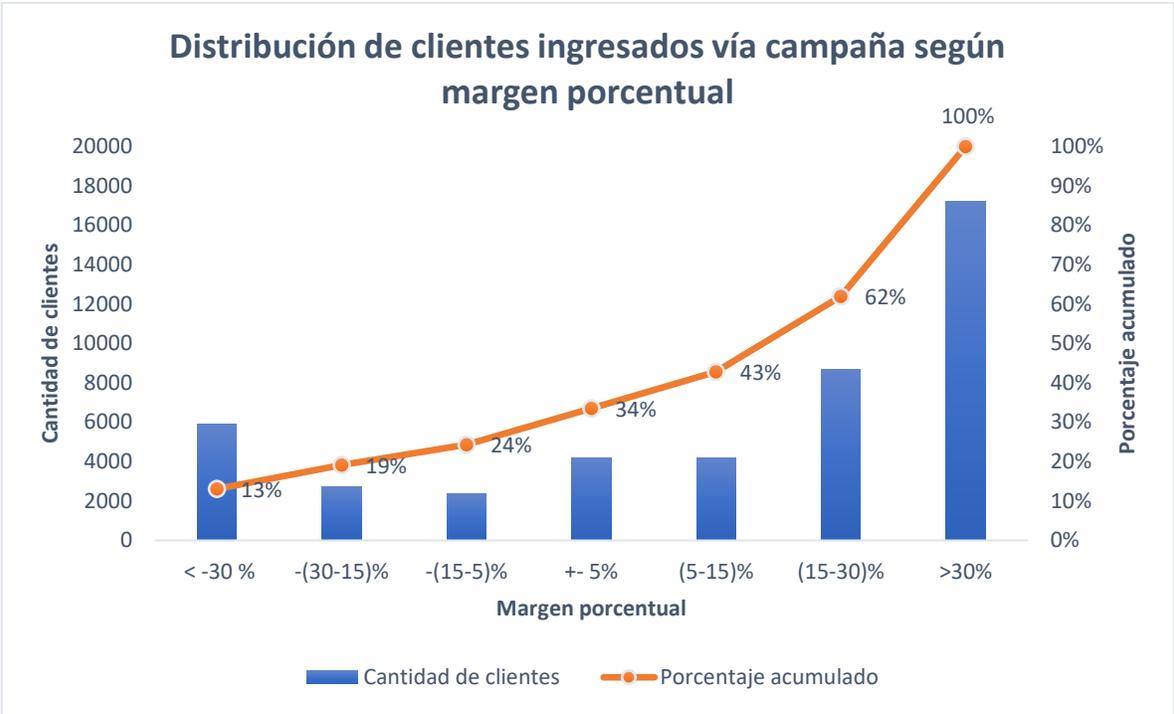


Figura 16: Distribución de clientes según margen porcentual

Fuente: Elaboración propia en base a clientes migrados a pospago entre febrero y abril del 2019

Para medir el beneficio de migrar un cliente prepago a pospago se calculó la utilidad de cada uno como la diferencia entre el cargo fijo y el promedio de recarga en los últimos tres meses.

$$Utilidad = \text{Cargo Fijo} - \text{Promedio de recarga en los últimos 3 meses}$$

Ecuación 6: Utilidad de la migración

De la Figura 17, se observa que los clientes migrados vía outbound (durante tres meses distintos) con un margen porcentual menor al 30%, generan una pérdida de

\$51,1 millones, esta pérdida acumulada aumenta hasta \$58,5 millones y luego comienza a disminuir en las categorías con margen positivo. Recién considerando el rango que entrega el mayor margen porcentual, la utilidad de la campaña se vuelve positiva, resultando una utilidad total de \$34,1 millones.

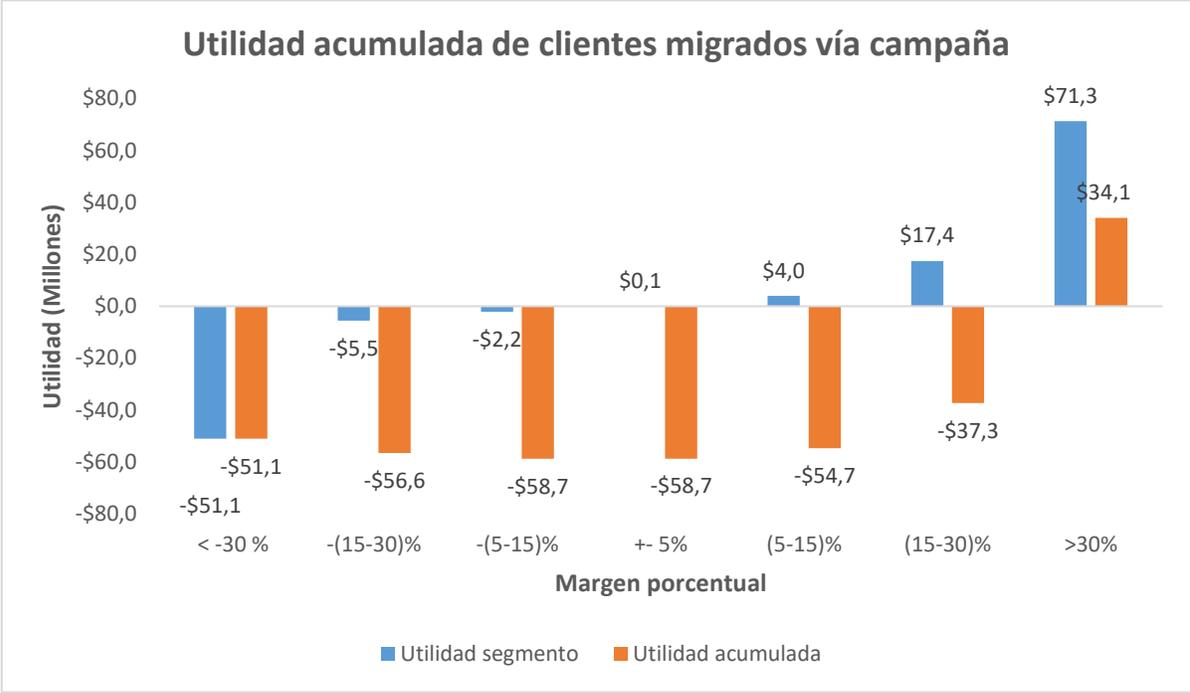
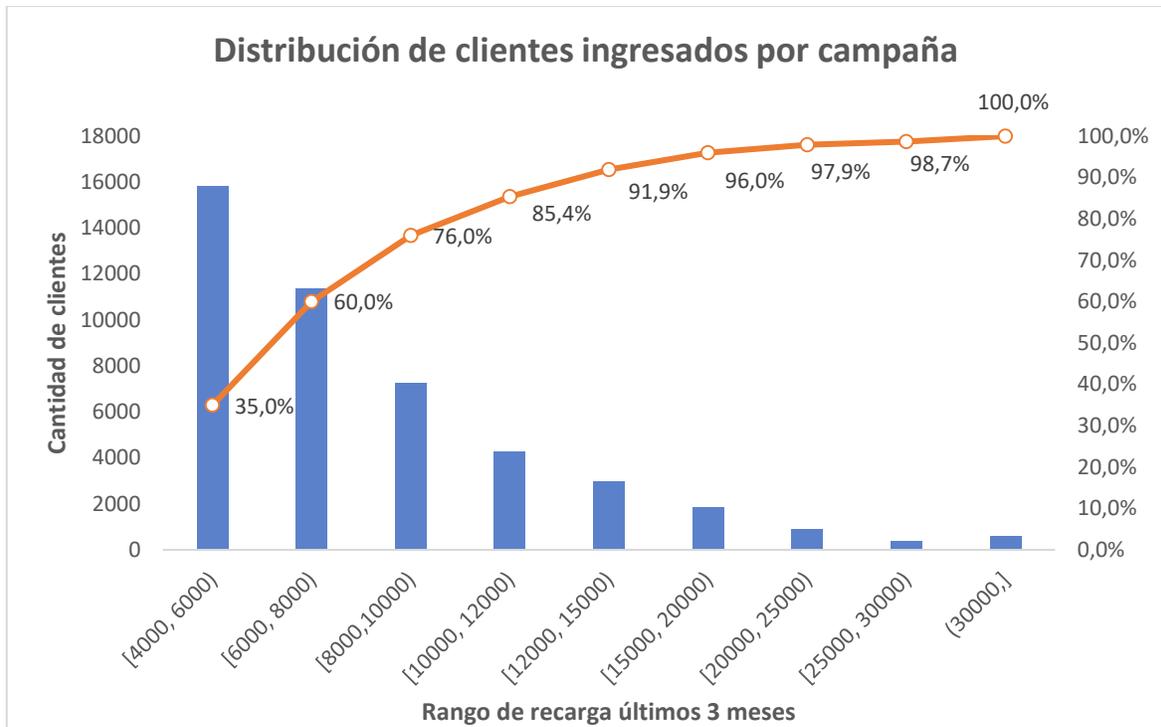


Figura 17: Utilidad de clientes según margen porcentual

Fuente: Elaboración propia en base a clientes migrados a pospago entre febrero y abril del 2019

Otra forma de analizar la rentabilidad de la campaña corresponde a comparar la utilidad según diferentes rangos de recarga que realizaban los clientes prepago previo a su migración. Para ello, en conjunto con el área de negocio, se definieron 9 tramos de recarga, tal como se muestra en la Figura 18. En esta, se aprecia que el 76% de los clientes migrados tiene un rango de recarga menor o igual a \$10.000.



*Figura 18: Distribución de clientes según rango de recarga*

*Fuente:* Elaboración propia en base a clientes migrados a pospago entre febrero y abril del 2019

A partir de la Figura 19, la cual muestra la utilidad según los distintos rangos de recarga descritos anteriormente, se observa que la utilidad es positiva para clientes cuyo rango de recarga es menor o igual a \$10.000. A partir de un monto de recarga superior, las utilidades son negativas.

En conclusión, en base a la gestión actual de la campaña, solo se genera utilidad para clientes cuyo rango de recarga es menor o igual a \$10.000, puesto que los clientes no migran hacia el plan sugerido, sino que lo hacen en su mayoría a cargos fijos menores o igual al de \$9.990.

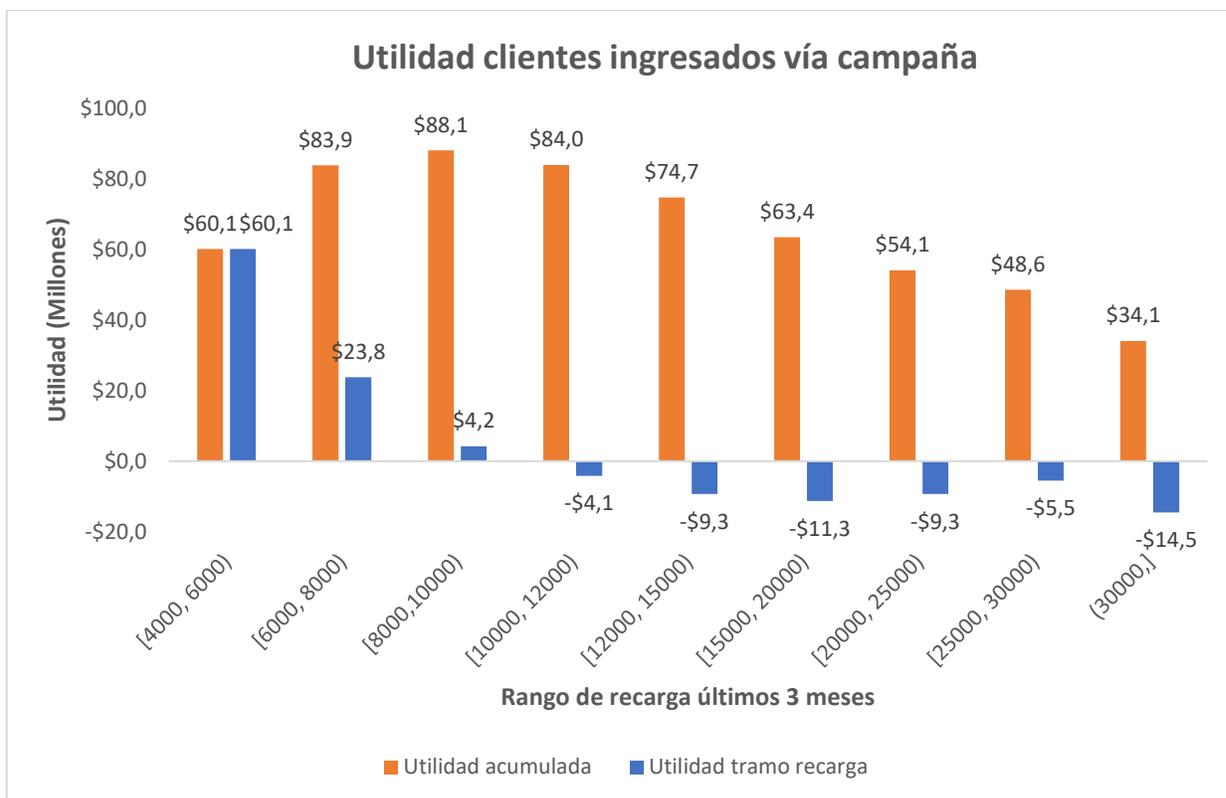


Figura 19: Utilidad según rango de recarga

Fuente: Elaboración propia en base a clientes migrados a pospago entre enero y marzo del 2019

#### 7.1.4. Cambios en la gestión actual de la campaña de migraciones

Tras presentar las conclusiones mostradas en la sección anterior a los encargados del negocio, se determinó hacer cambios en la gestión actual de la campaña con el objetivo de obtener una mayor rentabilidad a partir de esta.

Teniendo en consideración el valor actual de los planes a los cuales migran los clientes prepago, se decidió dejar fuera de la campaña a los clientes cuyo promedio de recarga en los últimos tres meses disponibles sea mayor a \$20.000, puesto que estos actualmente migran hacia planes de valor muy bajo en comparación a su comportamiento previo de recarga.

Adicionalmente, la exclusión de estos clientes será reemplazada por la incorporación a la campaña de clientes cuyo promedio de recarga en los últimos tres meses sea

menor a \$4.000. Clientes a los cuales anteriormente no se les hubiera ofrecido la campaña.

El potencial de ofrecer la opción de contratar un plan a los clientes cuyo rango de recarga es menor a \$4.000, es que necesariamente se incrementará el valor de este cliente para la compañía, puesto que el plan más barato que puede elegir corresponde al de \$7.990.

Finalmente, también se limitó el tipo de plan a ofrecer en base al rango de recarga previo del cliente, las nuevas reglas de gestión se resumen en la Tabla 3.

A partir de la Tabla 3, se observa que si el cliente posee un rango de recarga menor o igual a \$8.000, se mantiene la gestión actual en cuanto a la oferta del valor del plan que realiza el ejecutivo, esto puesto que cualquier plan generan un aumento en el valor del cliente.

Ahora bien, si el rango de recarga se encuentra entre \$8.001 y \$10.000, no se ofrecerá el plan de \$7.990. Mientras que si su rango está entre \$10.001 y \$12.000 se priorizará ofrecer en primera instancia el plan de \$14.990, y en caso de no aceptarlo, también se permitirá ofrecer el plan de \$9.990.

Por último, para clientes cuyo monto de recarga sea mayor a \$12.000, solo se ofrecerán planes superiores a \$14.990.

	<b>Valor del plan</b>				
<b>Rango de recarga</b>	\$7.990	\$9.990	\$14.990	\$19.990	\$21.990
< \$4.000	Incluir clientes de este segmento				
[\$4.000, \$6.000]	Mantener gestión actual				
[\$6.001, \$8.000]					
[\$8.001, \$10.000]	No ofrecer				
[\$10.001, \$12.000]	No ofrecer	Segunda opción	Primera opción		
[\$12.00, \$20.000]	No ofrecer	No ofrecer			
>\$20.000	Excluir clientes de campaña				

*Tabla 3: Cambios en la gestión actual de la campaña*

### 7.1.5. Generación actual de base de clientes a cargar en campaña

Hoy en día, la metodología para definir la base de clientes a los cuales ofrecer la campaña se realiza con la información de cierre de dos meses previos a la gestión de esta. Por ejemplo, en diciembre se elige la base de clientes a los cuales ofrecer la campaña en el mes de enero utilizando la información hasta el cierre del mes de noviembre.

Luego de los filtros explicados en la sección 5.1.1, se elige a los clientes cuyo rango de recarga en los últimos tres meses analizados sea superior a \$4.000. En caso de tener un conjunto de clientes mayor a lo que se debe cargar, se prioriza a los clientes que recargaron durante los tres meses analizados, luego a los que solo recargaron dos de los tres últimos meses, y finalmente a quienes recargaron solamente en 1 mes.

Durante los últimos meses de campaña, teniendo en consideración la exclusión de clientes cuyo rango de recarga es superior a \$20.000, la base de clientes es cada vez menor, por lo que, para las campañas de octubre y noviembre, simplemente se cargó en campaña todos los clientes con las características descritas anteriormente.

### 7.1.6. Desarrollo del trabajo

Dado lo explicado anteriormente, el primer punto a considerar corresponde a la inclusión de un nuevo segmento de clientes a la campaña de migración, los cuales son clientes cuyo rango de recarga es menor a \$4.000.

Este segmento normalmente no es cargado a campaña, por lo cual, no se tiene información histórica respecto a su efectividad en esta, sin embargo, existe un grupo de clientes con estas características que migra de forma espontánea hacia suscripción, por lo que se utilizará dicha información para construir un modelo que permita detectar a los clientes con una mayor probabilidad de migrar hacia pospago.

El objetivo de este modelo, denominado como modelo de migración espontánea, consiste en ofrecer la campaña a los clientes con rango de recarga bajo más propensos a migrar, acelerando de esta forma su traspaso hacia el mercado de pospago.

Por otro lado, teniendo en cuenta la disminución de la base total disponible a cargar en campaña con los filtros actuales, se creará un modelo para determinar la propensión de estos a migrar hacia pospago.

Mediante la combinación de ambos modelos, es posible determinar la base óptima de clientes a los cuales ofrecerles la campaña, dejando fuera a los clientes con una baja propensión a migrar que hubieran sido cargados, e incorporando así clientes de con un rango de recarga menor, pero con una alta probabilidad de migrar.

## 7.2. Comprensión de los datos

Para comenzar a trabajar en la creación de los modelos, es necesario seleccionar los datos que serán incluidos en el análisis, los cuales se dividen en 7 conjuntos de datos distintos. En primer lugar, se explicarán los datos mostrado en la Figura 20.

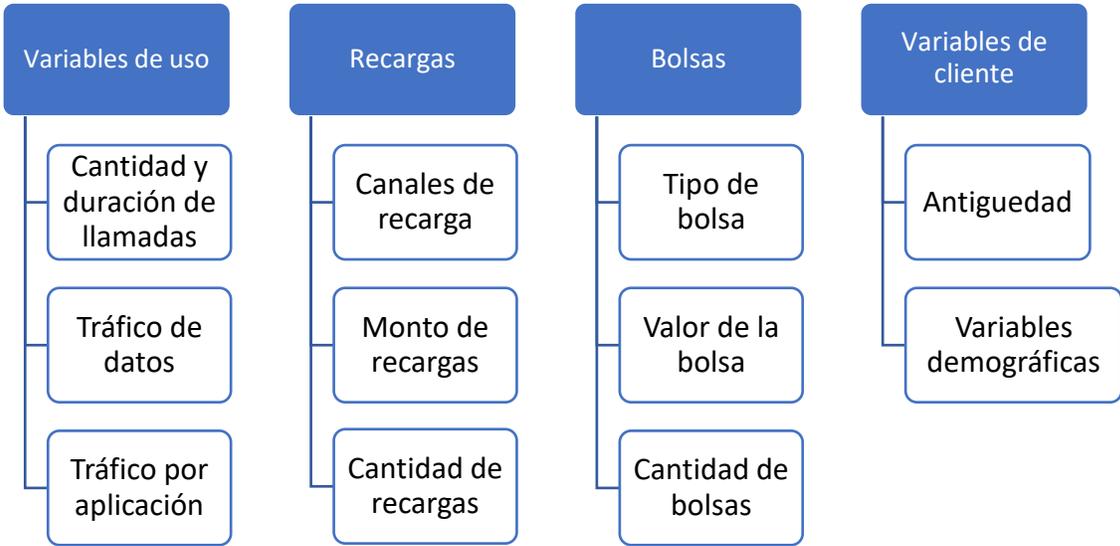


Figura 20: Descripción de variables utilizadas en el modelo, parte 1.

Fuente: Elaboración propia

- Variables de uso: Contiene información respecto a las llamadas, mensajes de texto, y tráfico de datos del cliente. Las variables disponibles son las siguientes:
  - Cantidad de llamadas recibidas
  - Cantidad de llamadas entrantes
  - Duración de las llamadas recibidas (segundos)

- Duración de las llamadas entrantes (segundos)
  - Cantidad de mensajes de texto enviados
  - Cantidad de mensajes de texto recibidos
  - Tráfico de datos (Bytes)
- Tráfico de datos por aplicación: Posee el desglose del tráfico de datos medido en bytes en más de 60 aplicaciones. Este reporte se entrega de forma quincenal. En el Anexo 1 se muestra el desglose de las aplicaciones utilizadas.
  - Recargas: Contiene todo el detalle de las recargas que realiza el cliente, tales como el monto, la cantidad de recargas y el canal mediante el cual realiza esta. Los canales disponibles para realizar recargas son 5:
    - Aplicación Entel
    - Página Web
    - Tiendas: Incluye negocios particulares, cajas vecinas, farmacias, supermercados, entre otros.
    - IVR: Llamada telefónica con respuesta de voz interactiva.
    - USSD: Línea de comandos, por ejemplo, \*103#.

Para cada canal, se utiliza la información de la cantidad de recargas realizadas, y el monto recargado en cada una de ellas.

- Bolsas: Al igual que en el caso de las recargas, se posee información respecto a la cantidad de bolsas compradas y el monto gastado en la compra de estas. Por lo demás, también se posee información específica del tipo de bolsa adquirida.

Actualmente es posible comprar más de 40 bolsas distintas, sin embargo, tan solo siete bolsas representan más del 80% de las bolsas adquiridas por los clientes. En base a lo anterior, para efectos de los análisis, se considerarán las bolsas que concentran más del 80% del total de compra, las cuales se muestran en la Tabla 4.

<b>Descripción de bolsa</b>	<b>Porcentaje de compra</b>	<b>Porcentaje acumulado</b>
Bolsa de datos por 3 días	43,8%	43,8%
Bolsa de datos por 7 días	12,7%	56,5%
Bolsa mixta por 7 días	7,1%	63,6%
Bolsa de datos por 2 días	7,1%	70,7%%
Bolsa mixta por 3 días	3,9%	74,6%
Bolsa mixta por 5 días	3,8%	78,5%
Bolsa de datos por 15 días	3,7%	82,2%

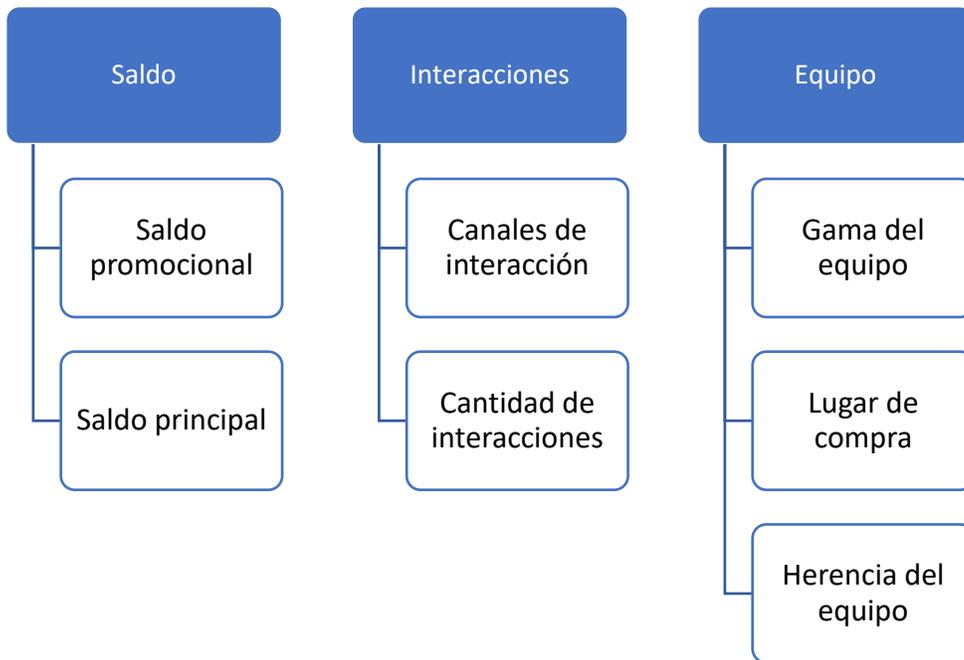
*Tabla 4: Distribución de bolsas más frecuentes para clientes prepago*

- Variables de cliente: Incluye la antigüedad del cliente en la compañía medida en meses, género del usuario y región donde habita.

Además, se registra el método bajo el cual ingresó el cliente a prepago, el cual puede ser mediante una habilitación de un nuevo número, portándose desde otra compañía (port in) u otro motivo desconocido.

Por último, se tiene registro del tipo de cliente según su comportamiento, el cual puede ser cliente de datos, redes sociales o de voz.

Adicionalmente existen variables asociadas al saldo, interacciones y equipos, las cuales se muestran en la Figura 21.



*Figura 21: Descripción de variables utilizadas en el modelo, parte 2.*

*Fuente: Elaboración propia*

- **Saldo:** Posee información diaria del saldo disponible para cada cliente, este se divide en dos, saldo principal y promocional. El saldo principal corresponde al saldo que considera las recargas realizadas por el cliente y el monto de este gastado en compra de bolsas y/o llamadas telefónicas, mientras que el saldo promocional solo incluye los montos que se le regalan al cliente producto de alguna oferta, por ejemplo, existen promociones en donde si el cliente prepago recarga \$5.000, se le regalan \$1.000 adicionales, este último monto correspondería al saldo promocional.
- **Interacciones:** Un cliente prepago puede interactuar con la compañía mediante diferentes plataformas, por ejemplo, llamando al call center, utilizando la aplicación para consultar el saldo disponible, o reclamando por algún inconveniente al SERNAC. Esta base almacena las interacciones que tiene el cliente con Entel a través de distintos canales. Para cada uno de ellos, se registran la cantidad de interacciones realizadas.

Los canales de interacción a utilizar son los siguientes:

- Agente en terreno: Interacción con agentes Entel en lugares públicos, como, por ejemplo, en una estación de metro.
  - Aplicación: Uso de la aplicación Entel.
  - Bot: Chat automatizado integrado en el sitio web de Entel.
  - Call center: Comunicación entre el cliente y el servicio de ayuda vía call center.
  - Franquicia: Visita a alguna tienda franquicia de Entel.
  - Tiendas propias: Visita en tiendas que pertenecen oficialmente a Entel.
  - Grandes Tiendas: Visita en grandes tiendas de retail, tales como Falabella, Ripley, Paris, entre otros.
  - Subtel: Interacciones del usuario con la Subsecretaría de Telecomunicaciones.
  - Sernac: Interacción con la compañía vía el Servicio Nacional del Consumidor, generalmente esta interacción hace alusión a cuando un cliente Entel se inscribe para no ser contactado por la compañía, por lo cual, no es posible ofrecerle ningún producto ni servicio.
- Equipo: Finalmente, se tiene información respecto al equipo que usa el cliente, las cuales se describen a continuación:
    - Gama del equipo: La cual puede ser baja, media o alta según las características del equipo.
    - Herencia: Indica si el equipo que usa el cliente lo había usado anteriormente otra persona.
    - Lugar de compra: Registra si el equipo fue comprado en Entel o no.
    - Marca del equipo
    - Antigüedad del modelo: Variable que se calcula como la diferencia entre la fecha actual y la fecha de lanzamiento del modelo del equipo.

Con respecto al almacenamiento de la información, los datos asociados a las variables de uso, recargas, interacciones y bolsas se tienen de forma diaria, mientras que las variables del cliente y equipos se almacenan mensualmente.

### 7.3. Preparación de datos

#### 7.3.1. Limpieza de datos

- Antigüedad del cliente

Tal como se mencionó anteriormente, la antigüedad del cliente se encuentra medida en meses. Esta va desde 0 hasta los 150 meses (12 años y medio), sin embargo, a medida que la antigüedad es mayor, menor es el número de clientes prepago tal como se muestra en la Figura 22. Sumado a lo anterior, también existe un grupo de clientes a los cuales se desconoce su antigüedad, el cual corresponde a un 3,1% del total de la base prepago.

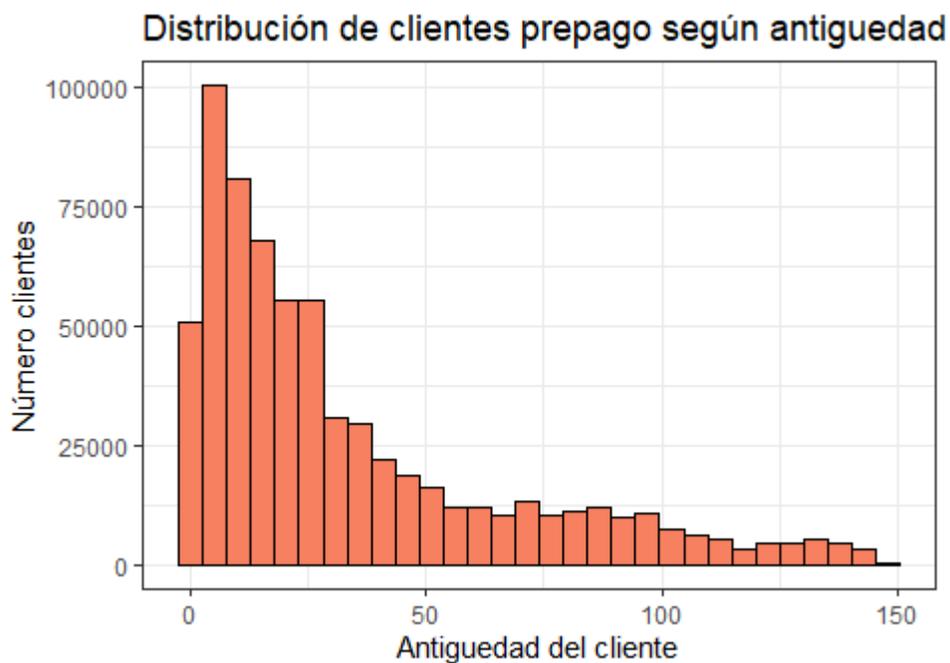


Figura 22: Histograma de antigüedad de clientes prepago

Fuente: Elaboración propia en R Studio

Por ambas razones, se utilizó la antigüedad del cliente agrupando en base a grupos que posean tasas similares de migración. Para definir la agrupación de la antigüedad, se utilizaron árboles de decisión que analizan la antigüedad en función de la tasa de efectividad de la campaña.

La Figura 23 corresponde a un árbol de decisión que divide a los clientes según su antigüedad en base a la tasa de migración con retención que estos tienen. El número superior dentro de cada nodo muestra la tasa de migración vía campaña del grupo respectivo, por ejemplo, en el primer nodo se observa la tasa de migración con retención promedio de la campaña entre febrero y mayo del 2019, la cual corresponde a 5,57%, mientras que en el nodo inferior derecho se observa la mayor tasa de migración, la cual se alcanza para clientes con una antigüedad menor a 2 meses, alcanzando una tasa de 10,5%. En función de la Figura 23 es posible definir distintos tramos para la antigüedad, entre 0 y 2 meses, entre 3 y 6 meses, entre 7 y 12 meses, entre 13 y 25 meses, entre 26 y 108 meses, superior a 108 meses y, por último, un segmento para los clientes que no poseen información de su antigüedad, los cuales no se incluyen en este árbol dado que no poseen un valor numérico consistente para este campo.

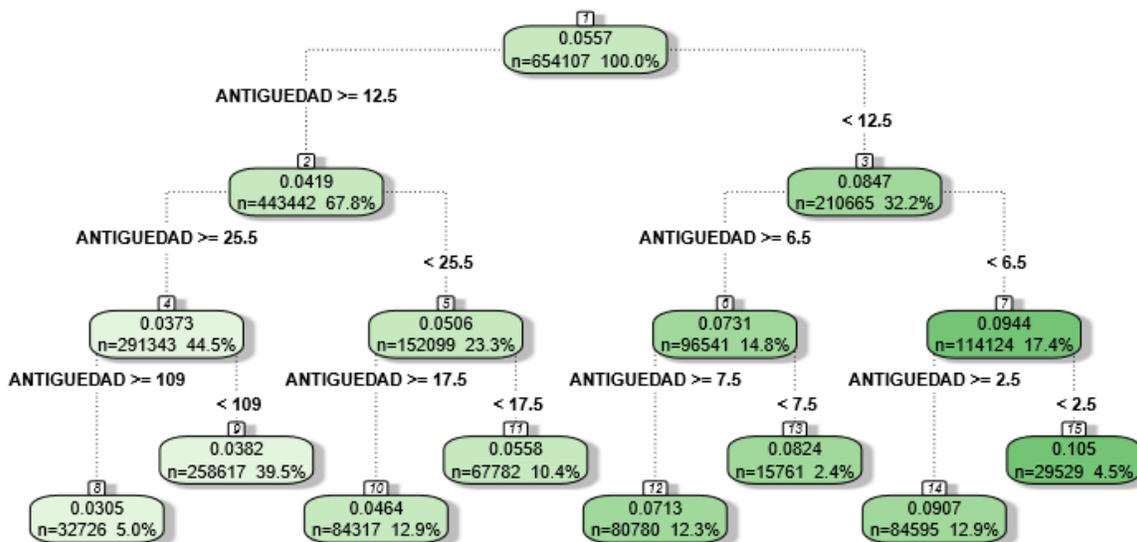


Figura 23: Árbol de decisión para la antigüedad del cliente

Fuente: Elaboración propia en R Studio

Estratificando de esta forma, quedan grupos con tasas de migración significativamente distintas tal como se aprecia en la Figura 24, en la cual se observa que mientras menor es la antigüedad, mayor es la tasa de migración.

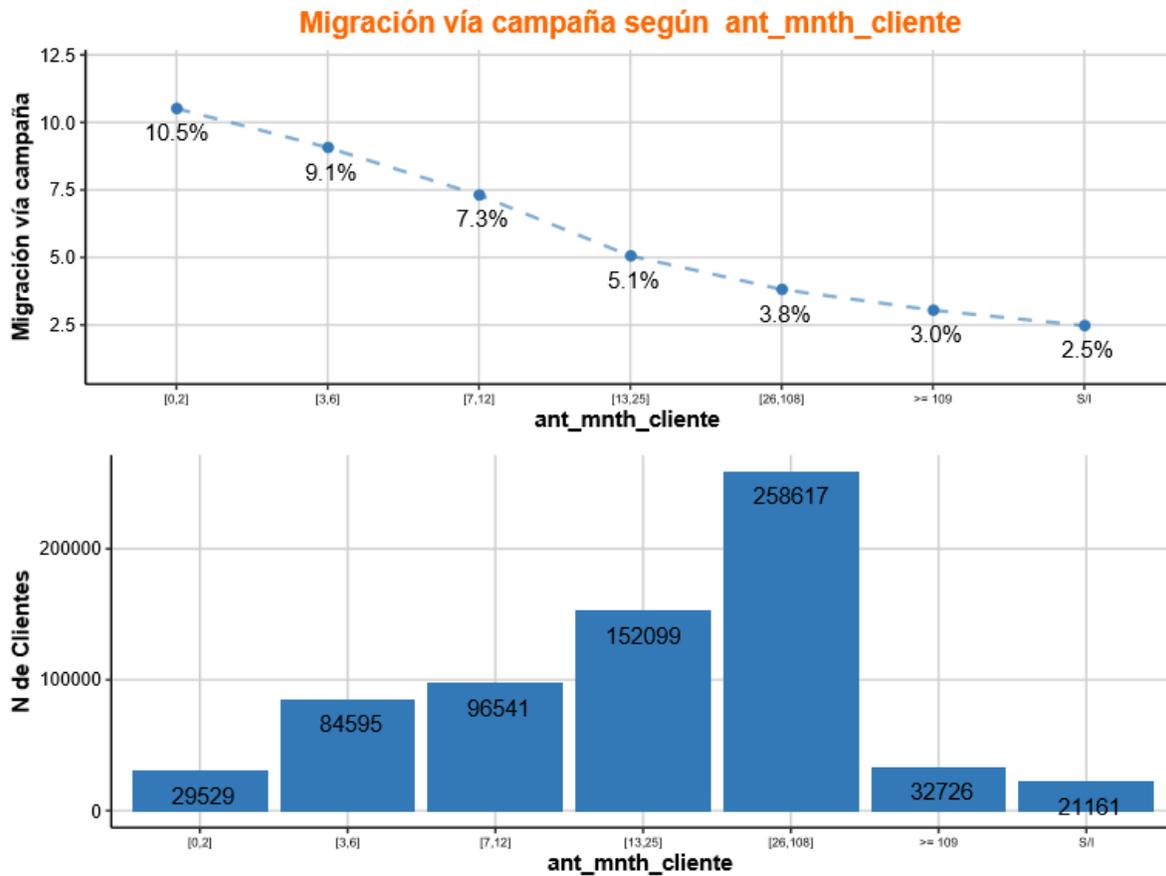


Figura 24: Estratificación de antigüedad del cliente

Fuente: Elaboración propia en R Studio

- Marca del equipo

Una marca puede venir escrita de distintas maneras, ya sea porque el nombre puede tener mayúsculas o minúsculas, o, en ocasiones, por errores de tipeo en lugar de la marca, también se encuentra el modelo del equipo, por ejemplo, en la marca para un dispositivo podría aparecer “Apple iPhone 6S”.

Para limpiar esta variable, en primer lugar, se limpió el campo marca viendo las palabras que más se repetían, para ello, se trabajó utilizando solamente caracteres en minúscula.

La Figura 25 muestra las marcas usadas por los clientes prepago de mayor a menor uso, a partir de esta, se aprecia que las tres marcas más usadas son Samsung, Own y Huawei, además, en segundo lugar, se encuentran los

clientes a los cuales se desconoce la marca del equipo. Por esta razón, para la construcción del modelo se decidió considerar las 8 marcas más usadas por los clientes correspondientes a Samsung, Own, Huawei, LG, Alcatel, Motorola, Nokia y Apple, y, adicionalmente, se agregó una categoría para incluir a los clientes de los cuales se desconoce la marca del equipo (Sin info). Por último, se agregó otra categoría que incluye a todas las demás marcas no consideradas, tales como Xiaomi, Sony, Azumi, etc.

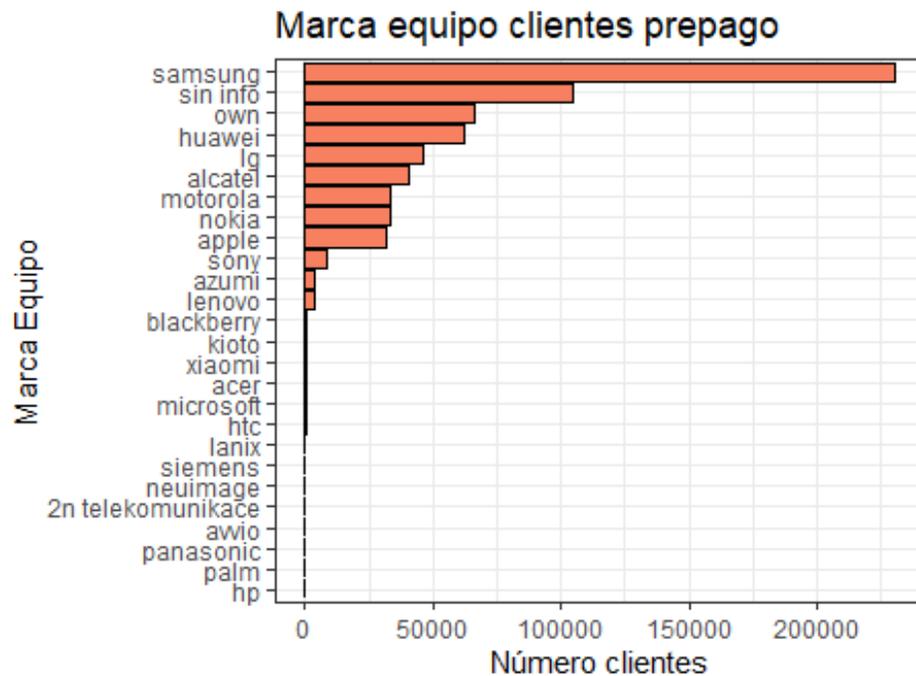


Figura 25: Distribución de marca de equipos clientes prepago

Fuente: Elaboración propia en R Studio

- Recargas y detección de outliers

Para las variables numéricas asociadas a recargas y monto gastado en la compra de bolsas se eliminaron los datos atípicos, lo anterior se realizó dejando fuera los outliers extremos según el método basado en el rango intercuartílico.

La Figura 26 muestra el boxplot de las recargas de enero 2019 antes de la eliminación de outliers junto con el boxplot de la variable luego de descartar

los datos atípicos. De los datos preprocesados es posible observar que el valor máximo se encuentra sobre \$1.000.000, mientras que luego de aplicar el método intercuartílico el valor máximo corresponde a \$31.700.

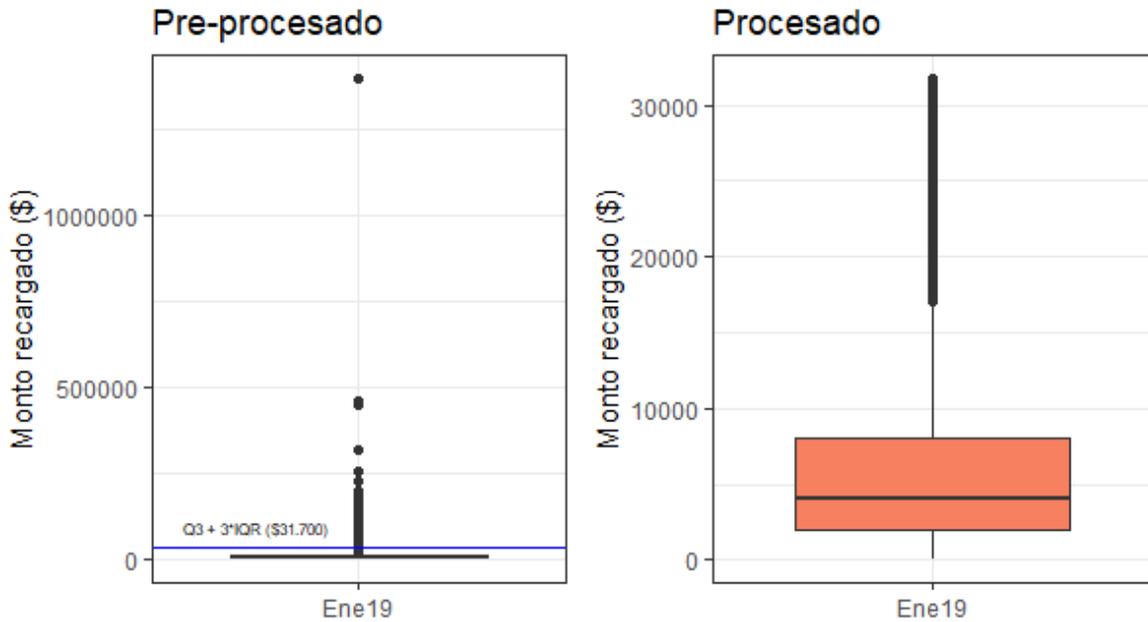


Figura 26: Distribución de recargas de enero 2019

Fuente: Elaboración propia en R Studio

### 7.3.2. Creación de variables

A partir de los datos descritos en el ítem anterior, se crearon variables de interés para entender de mejor manera el comportamiento de los clientes.

- Recargas

Con el fin de utilizar los datos de forma mensual, se calcularon variables como el mínimo y máximo monto de recarga durante el mes, junto con el monto promedio de recarga. Lo anterior tiene como objetivo poder diferenciar entre clientes que recargan frecuentemente un monto bajo, versus quienes recargan un menor número de veces, pero un monto mayor.

Por otro lado, se calcularon los meses y días desde la última recarga, el monto total recargado mensualmente, y el promedio y varianza del monto recargado en los últimos tres meses.

- Canales de recarga

Para incorporar los canales mediante los cuales recarga el cliente, se calculó el porcentaje de recargas realizadas por cada canal durante los últimos 3 meses. Por ejemplo, si durante tres meses el cliente realizó 6 recargas mediante la aplicación y 2 vía tiendas físicas, el porcentaje de recargas vía app será  $(6/8)*100 = 75\%$ , y el porcentaje mediante tiendas físicas será de 25%.

Además de incorporar el porcentaje de recargas por canal, se calculó también el porcentaje total del monto recargado para cada canal.

De esta forma es posible saber tanto el canal por donde realiza más recargas, como el canal donde recarga la mayor cantidad de dinero, agregando ambas variables al modelo.

- Canales de interacción

Para todos los canales de interacción descritos anteriormente, se creó una variable binaria que indica si el cliente interactuó o no con la compañía vía ese medio durante los últimos tres meses.

- Bolsas

Al igual que en el caso de recargas, se calculó el mínimo y máximo monto gastado en la compra de bolsas (no roaming) para cada mes analizado, además, se calculó el monto promedio que se gasta al adquirir una bolsa.

Para incorporar la compra de bolsas roaming, se agregó una variable binaria que indica si el cliente compró o no bolsas roaming durante los últimos tres meses.

Sumado a lo anterior, se incorporó una variable que calcula el porcentaje del monto disponible utilizado en la compra de bolsas, lo anterior, con el objetivo de ver qué porcentaje del monto recargado mensualmente se gasta en la compra de bolsa puesto que al utilizar bolsas los costos de uso son más baratos.

Finalmente, para incorporar el detalle de las bolsas más vendidas, se incluyó una variable binaria para cada bolsa que indica si el cliente la compró o no en los últimos tres meses.

- Frecuencia de recarga y bolsas

Por último, también se crearon variables para analizar la estabilidad de compra del cliente.

Respecto a las recargas, se calculó el porcentaje de meses en que realiza alguna recarga en los últimos tres y seis meses. Análogamente, se calcularon estas mismas variables para la compra de bolsas.

### 7.3.3. Transformación de variables

- Estratificación de variables
  - Tráfico de aplicaciones: El tráfico de datos en aplicaciones se encuentra en bytes, por lo que su valor, dependiendo de la aplicación, puede variar entre 0 hasta 2 MB o más (2.000.000 Bytes).

Para incorporar dichas variables al modelo, se procedió a agrupar estas en quintiles, siendo el quintil 1 el con mayor tráfico de datos y 5 el con menor. Adicionalmente, se agregó una categoría con valor 0 (decil 0) para clientes que no presentan tráfico en dicha aplicación, y otra para los clientes que no poseen información.

- Cantidad de segundos entrantes y salientes de llamadas: La cantidad de segundos hablados y recibidos es una de las variables con mayor varianza en los datos, por ello, se decidió incorporar esta variable agrupada por deciles, siendo el decil 1 los clientes con mayor cantidad de segundos emitidos/recibidos según sea el caso.
- Tráfico de datos: De manera análoga a la cantidad de segundos entrantes y salientes de llamadas, el tráfico de datos total de cada usuario también se utilizó en forma de decil.

## 7.4. Modelación

### 7.4.1. Descripción de modelos

- Variables independientes

Para modelar el problema, se utilizan los datos de los clientes prepago durante los últimos 3 meses disponibles. Por ejemplo, si se quiere predecir los clientes migrados en el mes de mayo del 2019, se tomarán en consideración los datos de los clientes en los meses de enero, febrero y marzo del mismo año.

El motivo de lo anterior se debe a que la base resultante se envía entre el 20 y 25 de cada mes, por lo que no es posible tener la información completa del último mes, sumado a esto, también existe un desfase de la disponibilización de información diaria de entre 3 a 9 días.

Sin embargo, si es posible utilizar el tráfico de aplicaciones del último mes puesto que este se entrega de forma quincenal, por ello, estos corresponden a los datos más actualizados que incorpora el modelo.

Con el fin de evaluar cambios en la tendencia del comportamiento del cliente, se incluirán variables de uso para cada mes analizado, así como también variables asociadas a la recarga y compra de bolsas.

La información asociada a las variables relacionadas a interacciones, distribución de recarga en los distintos canales disponibles y compra de bolsas más comunes, se utilizará de forma agregada durante los últimos 3 meses, el objetivo de esto es incorporar un comportamiento de los clientes más estables a través del tiempo.

- Variable dependiente

Tal como se explicó anteriormente, la variable dependiente del modelo corresponde a si el cliente migró (vía campaña/espontánea dependiendo del modelo) y tuvo una permanencia de al menos 4 meses en el mercado pospago posterior al mes de ingreso.

- Explicación de modelos

La Figura 27 muestra un esquema de cómo se construyó el modelo, en este se incluyen variables de uso, recargas y bolsas en cada uno de los últimos 3 meses disponibles. Además, también se incorporan datos asociados a los canales de

recarga, interacciones y bolsas más compradas considerando el total del período descrito.

Respecto al último mes disponible, también se incluyen las variables asociadas al equipo del cliente, tales como la gama y marca del equipo, adicionalmente, también se agregan variables demográficas de este, como, por ejemplo, sexo y región.

Finalmente, en el mes en el cual se debe definir el grupo de clientes a ofrecer la campaña, se incluye el tráfico de aplicaciones de la primera quincena.

La variable independiente modela si el cliente migró hacia suscripción y permaneció en el mercado de pospago al menos en los cuatro meses siguientes.



Figura 27: Esquema explicativo de construcción del modelo

Fuente: Elaboración propia

- Datos de entrenamiento

Para entrenar el modelo, se considerarán las migraciones de tres meses, los cuales corresponden a clientes migrados en febrero, marzo y abril del 2019.

Dado lo anterior, se utilizarán datos de los clientes prepago desde octubre del 2018, hasta febrero del 2019. La elección de estos meses se debe a que la variable de migración requiere una permanencia de al menos cuatro meses, por lo que, para los clientes migrados en abril del 2019, se debe verificar su estado en el mercado de pospago en el mes de agosto del mismo año.

- Modelo de migración vía campaña

Para entrenar el modelo de migración vía campaña, los datos de entrenamiento incluyen todos los clientes cargados en campaña los meses mencionados anteriormente.

- Modelo de migración espontánea

El modelo de migración espontánea considera todos los clientes activos con al menos una recarga en los últimos tres meses, es decir, corresponden a clientes que además de haber recargado al menos una vez, también han tenido tráfico de llamadas en los últimos tres meses, o bien han utilizado el servicio de datos en dicho período.

Dado que la tasa de migración espontánea con permanencia de al menos cuatro meses corresponde a tan solo el 0,74% de la base de clientes prepago descrita en el párrafo anterior, para entrenar el modelo se utilizó la técnica de undersampling.

La metodología para realizar undersampling consistió en tomar una muestra de 300.000 clientes mensualmente, la cual considera todos los clientes migrados hacia suscripción con permanencia, y una muestra aleatoria de los demás clientes hasta completar los 300.000 casos.

- Cross validation

Para entrenar los modelos descritos en el marco teórico, se utilizó cross-validation, la cual corresponde a una técnica que divide a la muestra en  $k$  conjuntos diferentes y utiliza cada uno de ellos para testear el modelo. A medida que mayor sea la partición escogida ( $k$ ), menor es el sesgo del modelo.

En base a lo anterior, para entrenar los modelos se utilizó cross-validation dividiendo la muestra en 5 conjuntos diferentes.

- Datos de validación

Posterior a entrenar los modelos, se validaron los resultados de este fuera de muestra, utilizando para ello los datos correspondientes a las migraciones del mes de mayo. Dado que se quiere encontrar a los clientes migrados con permanencia de al menos cuatro meses en pospago, se verificará que estos aun permanezcan en pospago en el mes de septiembre.

#### 7.4.2. Modelo de migración vía campaña

##### **GBM**

- Lift acumulado

La Figura 28 muestra el lift acumulado hasta el primer percentil (4,54), quinto (3,26), décimo (2,29) y veinteavo percentil (2,2) en los datos de validación.

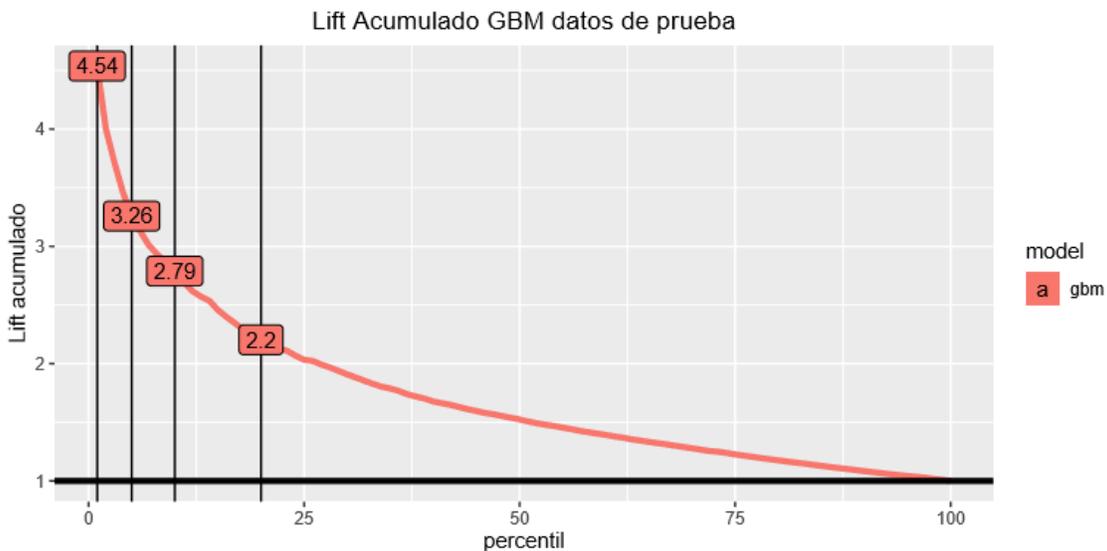


Figura 28: Lift acumulado modelo GBM, migración vía campaña.

Fuente: Elaboración propia en R Studio.

- Matriz de confusión

La matriz de confusión es posible obtenerla en base a distintos umbrales que permitan clasificar a un cliente en una categoría u otra. Para el modelo de migración vía campaña, el caso de éxito corresponde a un cliente migrado por esta vía con permanencia de al menos cuatro meses en el mercado de pospago.

La matriz mostrada para los modelos desarrollados es calculada para el umbral que maximiza el valor de MCC, variable descrita en el marco conceptual, y utilizada para evaluar la performance de los modelos en caso de data altamente desbalanceada. Para el modelo GBM, el umbral que maximiza la métrica de MCC es de 0,0879, es decir, este modelo clasifica como casos de éxito a todos los clientes que el modelo les asigna un score mayor o igual a 0,0879. En base a la matriz calculada, el MCC para los datos de entrenamiento corresponde a 0,123.

<b>GBM</b>	<b>CLASE REAL</b>		
		+	-
<b>CLASE PREDICHA</b>	+	2.168	20.321
	-	3.922	134.092

*Tabla 5: Matriz de confusión modelo GBM, migración vía campaña.*

- Curva ROC

El área bajo la curva ROC (AUC) para el modelo GBM corresponde a 0,698.

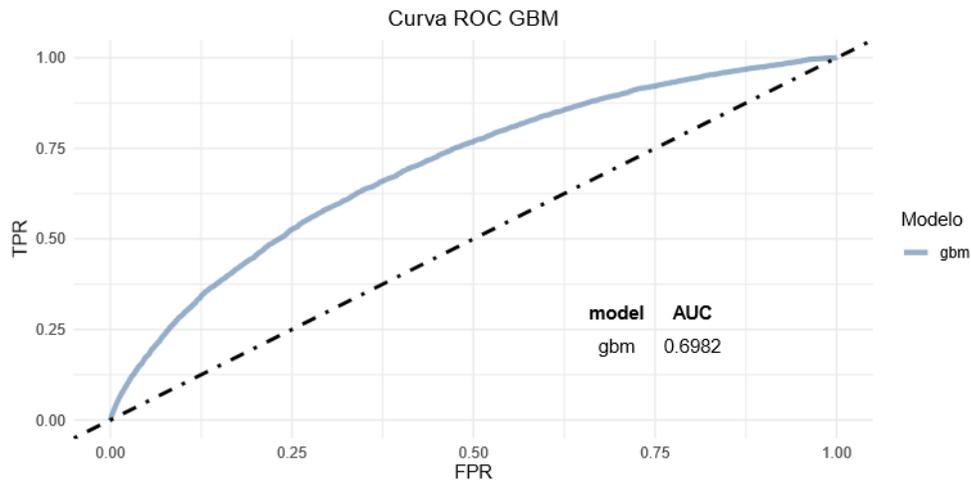


Figura 29: Curva ROC modelo GBM, migración vía campaña.

Fuente: Elaboración propia en R Studio.

## Random Forest

- Lift acumulado

La curva de lift muestra el lift acumulado hasta el primer percentil (3,7), quinto (2,9), décimo (2,4) y veinteavo percentil (1,93).

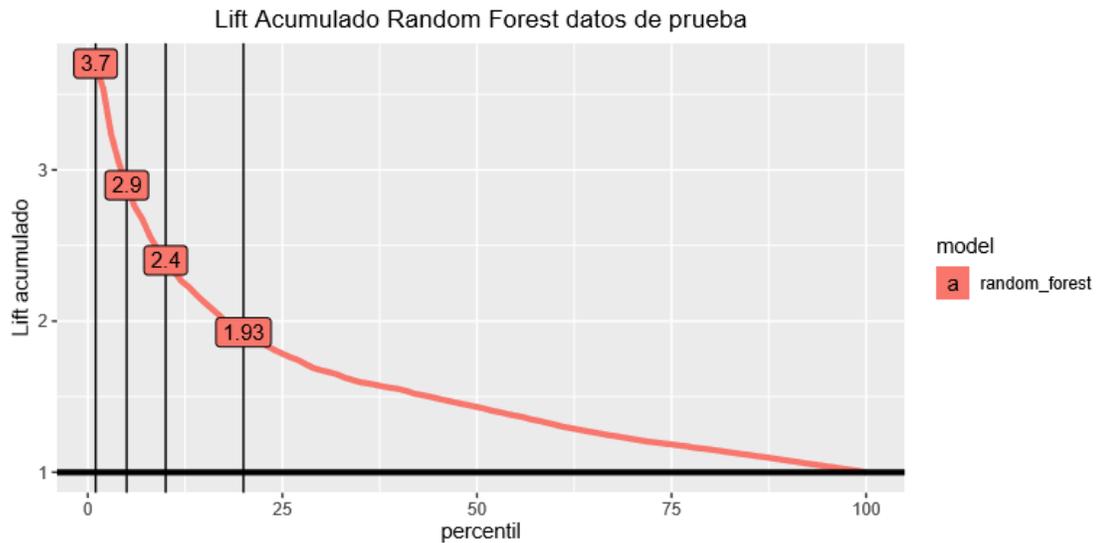


Figura 30: Lift acumulado modelo Random Forest, migración vía campaña.

Fuente: Elaboración propia en R Studio.

- Matriz de confusión

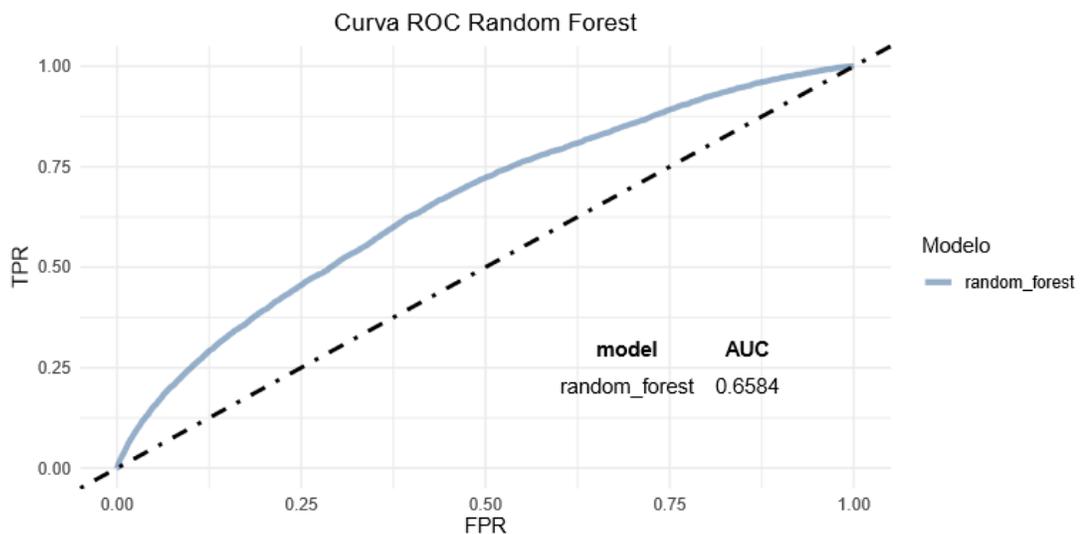
El umbral que maximiza el MCC para el modelo Random Forest corresponde a 0,045. En base a la matriz construida, el MCC para el modelo de Random Forest es de 0,094.

<b>Random Forest</b>	<b>CLASE REAL</b>		
		+	-
<b>CLASE PREDICHA</b>	+	1.755	18.897
	-	4.346	135.516

*Tabla 6: Matriz de confusión modelo Random Forest, migración vía campaña*

- Curva ROC

El área bajo la curva ROC para el modelo Random Forest corresponde a 0,658.



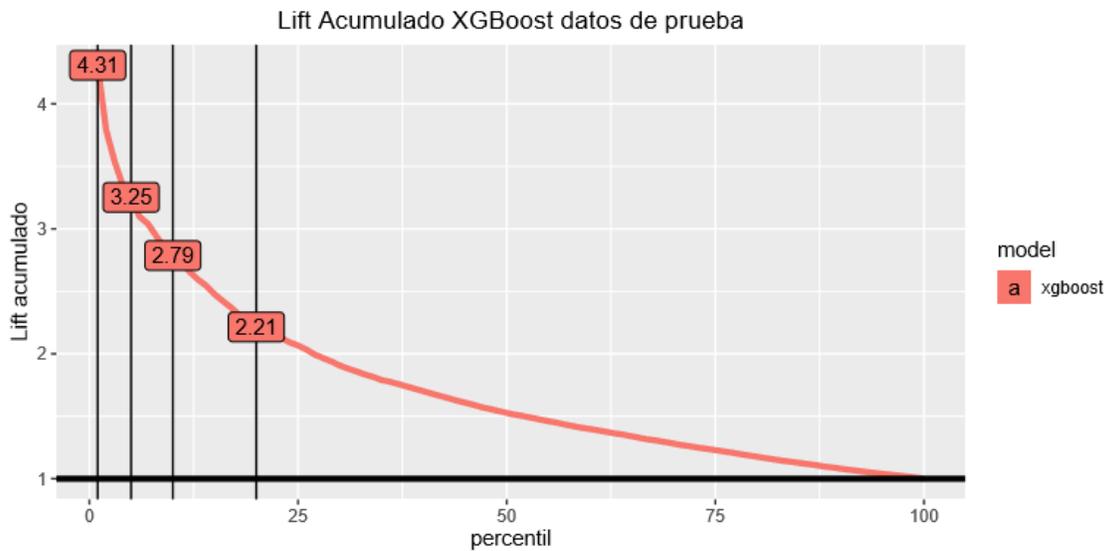
*Figura 31: Curva ROC modelo Random Forest, migración vía campaña.*

*Fuente: Elaboración propia en R Studio.*

## XGBoost

- Lift acumulado

La curva de lift muestra el lift acumulado hasta el primer percentil (4,31), quinto (3,25), décimo (2,79) y veinteavo percentil (2,21).



*Figura 32: Lift acumulado modelo XGBoost, migración vía campaña.*

*Fuente: Elaboración propia en R Studio.*

- Matriz de confusión

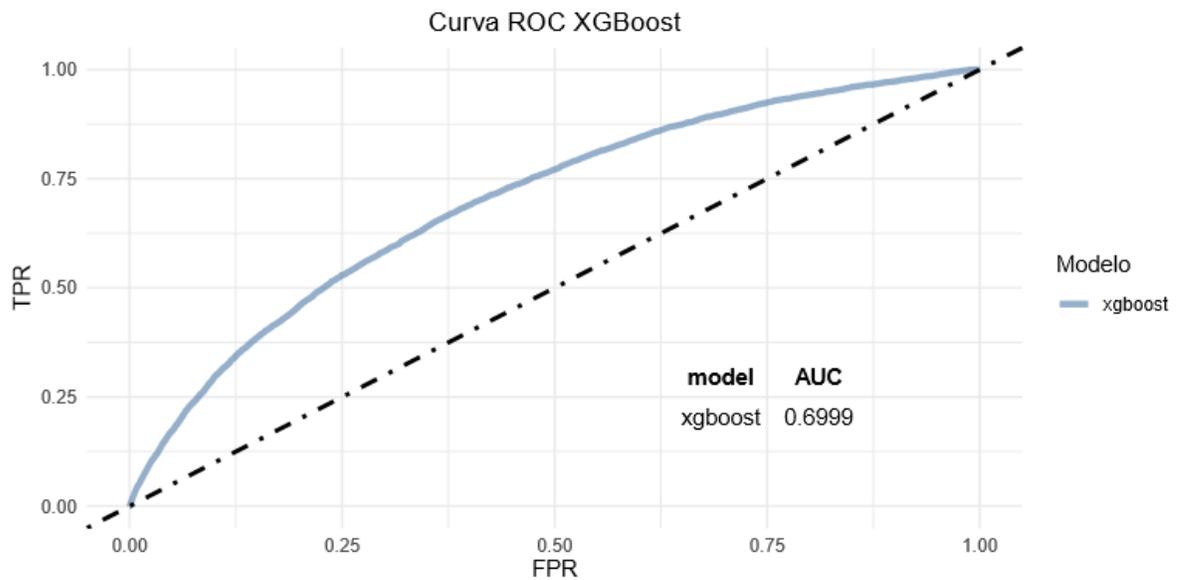
El umbral que maximiza el MCC para el modelo XGBoost corresponde a 0,093. Alcanzando un valor máximo de 0,124.

<b>XGBoost</b>	<b>CLASE REAL</b>		
		+	-
<b>CLASE PREDICHA</b>	+	2.123	19.522
	-	3.978	134.891

*Tabla 7: Matriz de confusión modelo XGBoost, migración vía campaña*

○ Curva ROC

El área bajo la curva ROC para el modelo XGBoost es de 0,699.



*Figura 33: Curva ROC modelo XGBoost, migración vía campaña.*

*Fuente: Elaboración propia en R Studio.*

#### 7.4.2.1. Evaluación del modelo de migración vía campaña

- Comparación de lift acumulado

A partir de la Figura 34, la cual muestra la curva de lift acumulado para los tres modelos desarrollados y el lift acumulado hasta el primer percentil, se observa que la curva asociada al modelo Random Forest se encuentra por debajo de los demás.

Por otro lado, al comparar las curvas de lift acumulado para los modelos GBM y XGBoost, se aprecia que el primero de estos posee un lift levemente mejor para los primeros percentiles, sin embargo, a partir del percentil 10, la curva del modelo XGBoost se encuentra ligeramente por encima del modelo Gradient Boosting Machine (GBM).

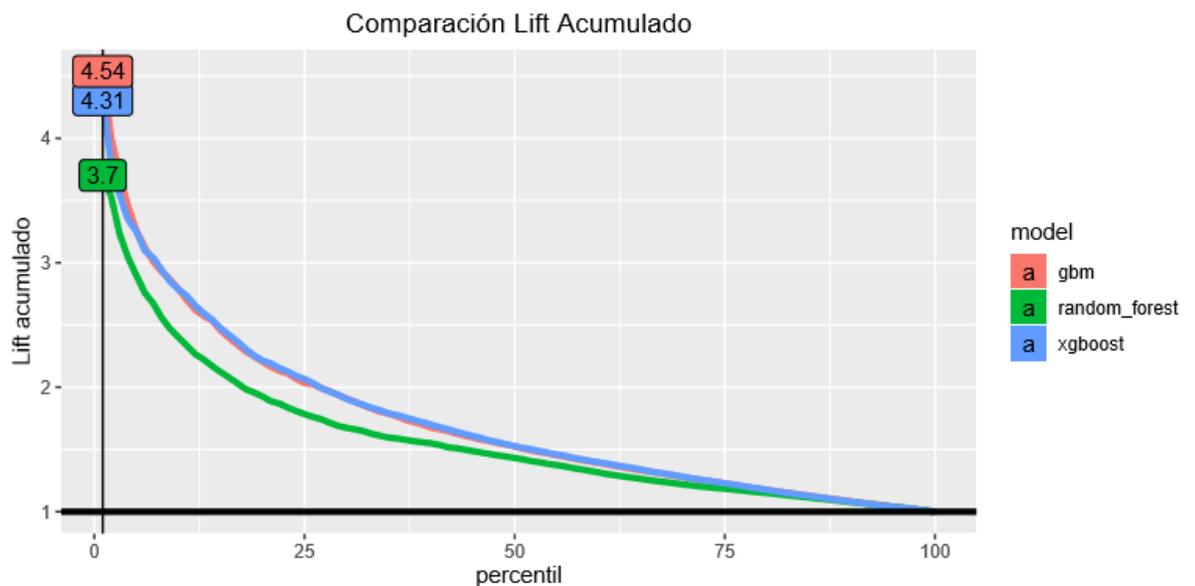


Figura 34: Comparación de lift acumulado para modelos de migración vía campaña

Fuente: Elaboración propia en R Studio.

- MCC

De forma análoga al gráfico asociado a la curva de lift acumulado, se observa que la métrica asociada al máximo coeficiente de correlación de Matthews (MCC) es menor para el modelo Random Forest, mientras que, si se compara

el coeficiente MCC para los modelos GBM y XGBoost, el valor es levemente superior para este último.

Métrica	GBM	Random Forest	XGBoost
MCC	0,123	0,094	0,124

Tabla 8: Comparación de MCC para modelos de migración vía campaña

- Curva ROC

Al comparar las curvas ROC, se observa que el modelo que presenta una peor performance nuevamente corresponde al algoritmo Random Forest. Por otro lado, las curvas de los modelos GBM y XGBoost se traslapan para la mayoría de los umbrales, resultando ser casi idénticas y difícil de distinguir las tal como se aprecia en la Figura 35, sin embargo, el modelo XGBoost posee un área bajo la curva superior.

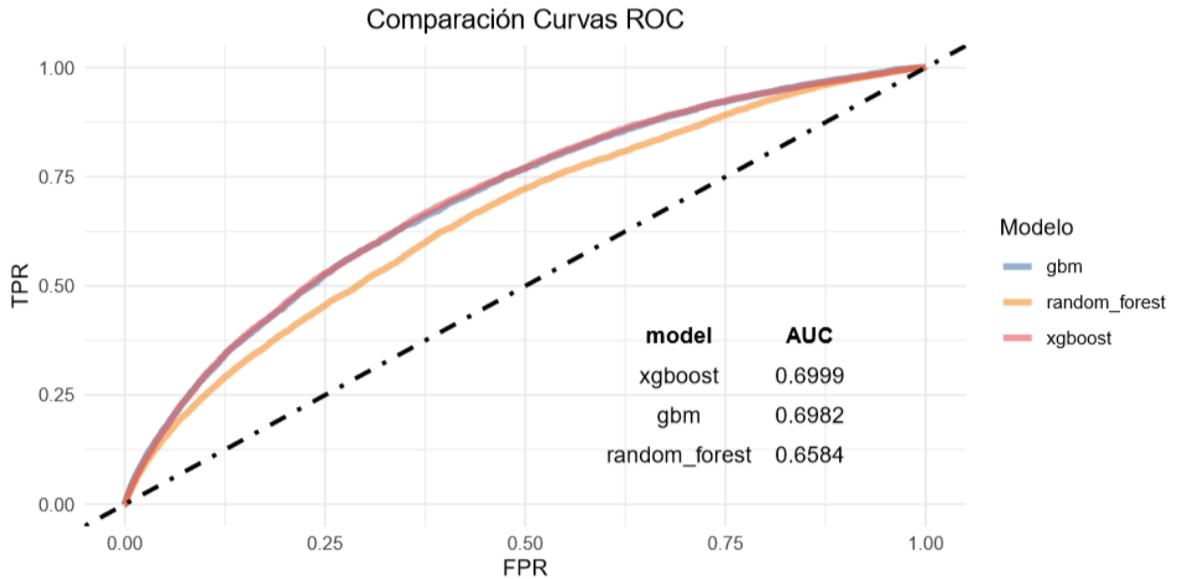


Figura 35: Comparación de curvas ROC modelos de migración vía campaña

Fuente: Elaboración propia en R Studio.

- **Mejor modelo**

En base a la evaluación de las distintas métricas utilizadas para comparar los modelos, el modelo XGBoost presenta mejores métricas de AUC (Area Under Curve) y MCC (Matthews Correlation Coefficient).

Con respecto a la curva de lift acumulado, si bien el modelo GBM presenta tasas sutilmente superiores en los primeros percentiles, el modelo XGBoost mantiene un lift acumulado mayor desde el percentil diez en adelante, lo anterior significa que, si se gestiona a más del 10% más propenso de la base, el modelo XGBoost entregaría una tasa de migración mayor. Teniendo en consideración que en la práctica se busca gestionar a un grupo de clientes superior al 10%, es óptimo utilizar entonces el modelo XGBoost.

Por lo tanto, el mejor modelo para identificar a los clientes con mayor probabilidad de migrar vía campaña hacia el mercado de pospago corresponde al modelo XGBoost.

### **Análisis del mejor modelo**

- **Resumen efectividad**

La Tabla 9 muestra los resultados obtenidos en los datos de validación del mejor modelo para predecir la migración vía campaña con permanencia de al menos cuatro meses en el mercado de pospago. A partir de esta, es posible observar que en el 1% más propenso la tasa de migración del grupo es 4,31 veces mayor a la tasa natural (3,8%) total de migración para el mes analizado, obteniendo una tasa de migración de 16,38% en dicho grupo. Lo anterior, permite identificar al 4,3% de los clientes migrados. De manera análoga, gestionando al 20% más propenso según el modelo, es posible identificar el 44,2% del total de clientes migrados dentro de un mes.

<b>Percentil</b>	<b>Lift Acumulado</b>	<b>Tasa acumulada</b>	<b>Porcentaje de migrados acumulados</b>
<b>1%</b>	4,31	16,38%	4,3%
<b>5%</b>	3,25	12,35%	16,3%
<b>10%</b>	2,79	10,6%	27,9%
<b>20%</b>	2,21	8,4%	44,2%
<b>100%</b>	1	3,8%	100%

*Tabla 9: Resumen de gestión según performance de mejor modelo vía campaña*

- **Importancia de variables**

En base al mejor modelo obtenido, el cual corresponde al modelo XGBoost, se analizarán las variables más importantes que entrega este, las cuales se muestran en la Figura 36. Lo anterior, con el fin de entender qué características poseen los clientes con mayor probabilidad de migrar a pospago mediante la campaña ofrecida vía call center.

Resulta importante mencionar que las variables más importantes obtenidas a partir del modelo XGBoost son similares a las de los modelos Random Forest y GBM (Ver Anexo 2).

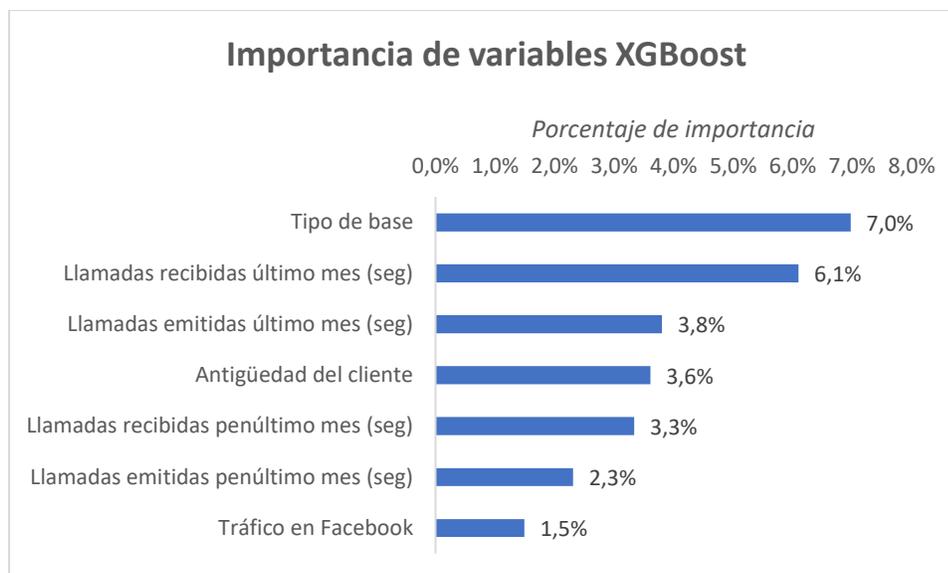


Figura 36: Importancia de variables modelo XGBoost, migración vía campaña.

Fuente: Elaboración propia

## 1. Tipo de base

El tipo de base es la variable más importante para el modelo, esta, como se explicó anteriormente, corresponde a la cantidad de veces que el cliente ha sido cargado en la campaña en los últimos seis meses.

La base B1 contiene clientes que nunca han sido cargados en la campaña, o bien no se les ha ofrecido esta en los últimos seis meses. Tal como se aprecia en la Figura 37, la tasa de migración para este grupo es significativamente superior (8,7%) a los clientes B2 (4,4%).

La línea punteada naranja indica la tasa de migración vía campaña con permanencia de la población total, con el objetivo de identificar que grupos se encuentra sobre y bajo este porcentaje.

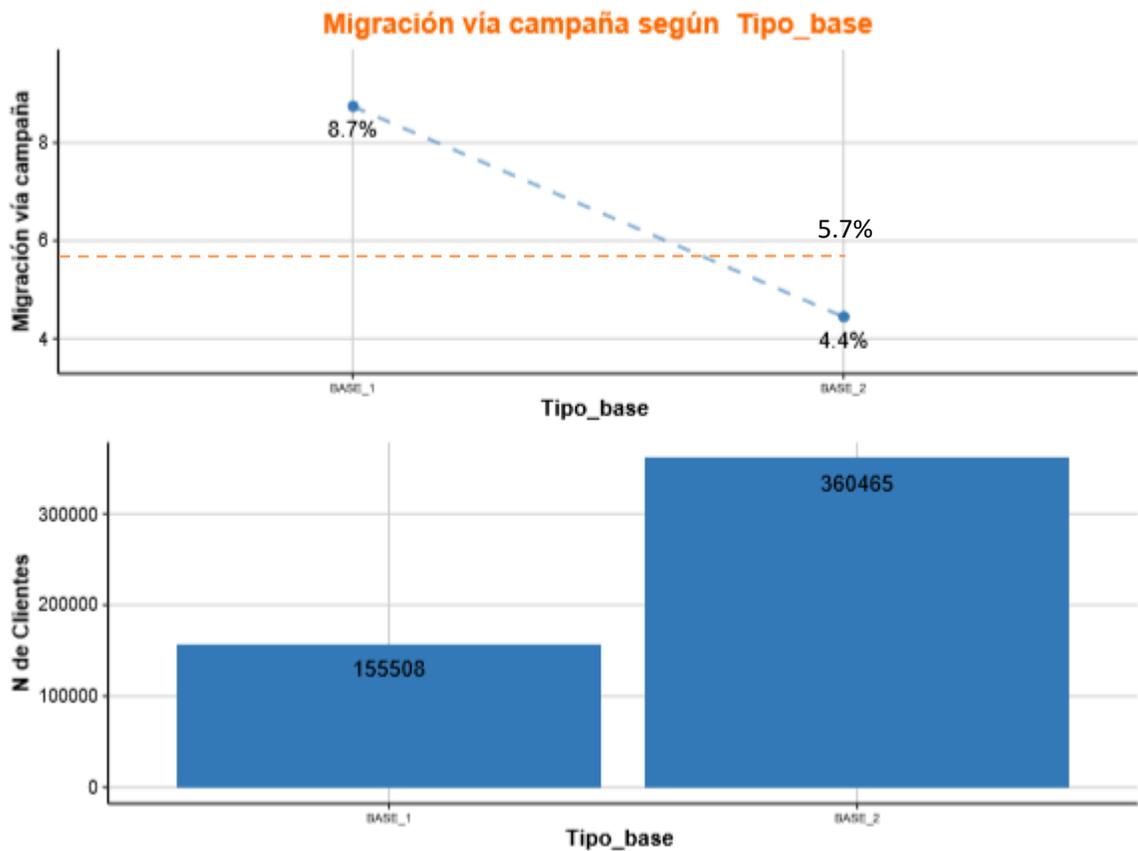


Figura 37: Tasa de migración con permanencia a suscripción vía campaña según Tipo de base

Fuente: Elaboración propia en R Studio

## 2. Cantidad de segundos recibidos el último mes (decil\_call\_in\_seg\_m1)

Esta variable se utilizó en deciles, donde en el decil 1 se encuentran los clientes con una mayor cantidad de segundos recibidos en el último mes disponible y en el decil 10 los clientes con la menor cantidad de segundos recibidos pero positivo. Por otro lado, en el decil 0 se encuentran todos los clientes prepago que no recibieron ninguna llamada el último mes (cantidad de segundos recibidos = 0).

A partir de la Figura 38, se observa que mientras mayor es la cantidad de segundos recibidos en el último mes disponible, mayor será la tasa de migración. En particular, para los primeros cinco deciles, la tasa de migración es superior a la tasa global de la campaña.

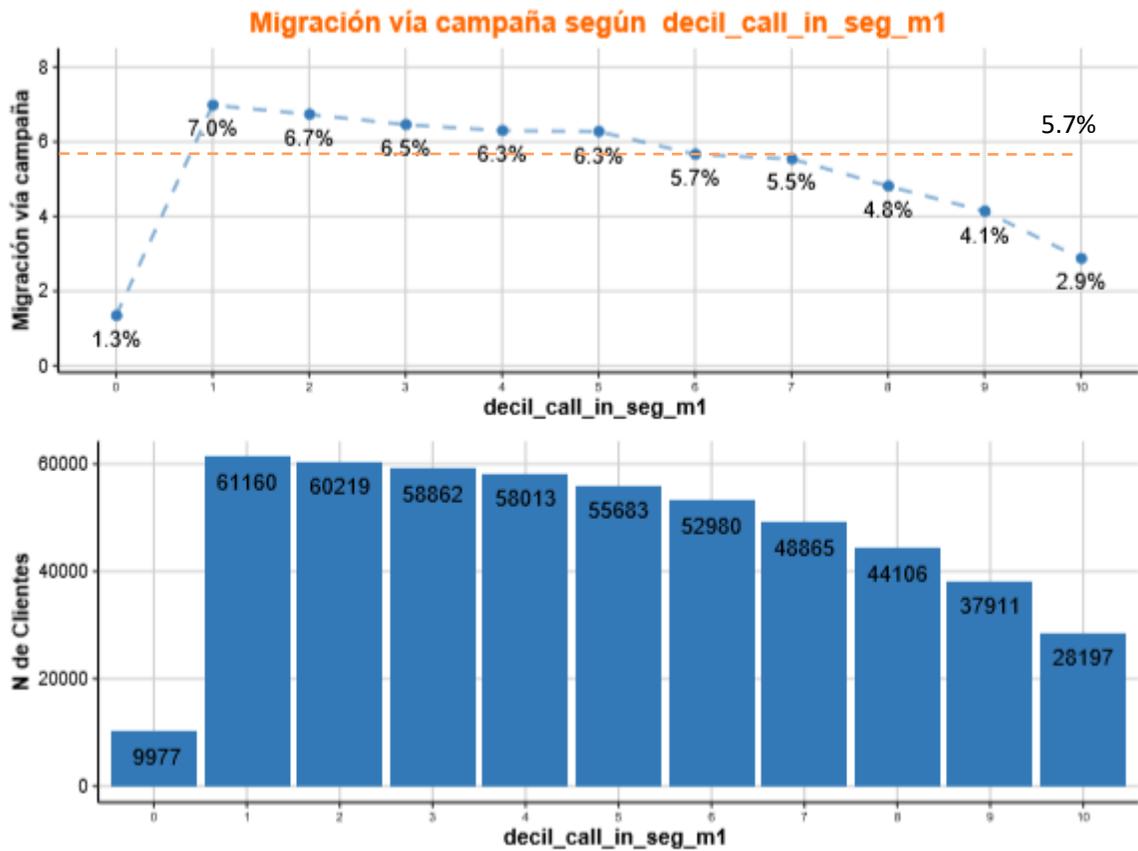


Figura 38: Tasa de migración con permanencia a suscripción vía campaña según llamadas recibidas en el mes previo (seg)

Fuente: Elaboración propia en R Studio

### 3. Cantidad de segundos emitidos el último mes (decil\_call\_out\_seg\_m1)

A partir de la Figura 39, es posible notar que mientras mayor es la cantidad de segundos emitidos en el último mes disponible, mayor es la tasa de migración en estos segmentos. A diferencia de los segundos recibidos, la tasa de migración decrece de forma más rápida para los deciles con menor cantidad de segundos realizados en llamadas, obteniendo tasas superiores a la tasa global solamente en los primeros cuatro deciles.

Adicionalmente, se observa que la tasa de migración en el primer decil es superior a todos los deciles asociados a los segundos recibidos.

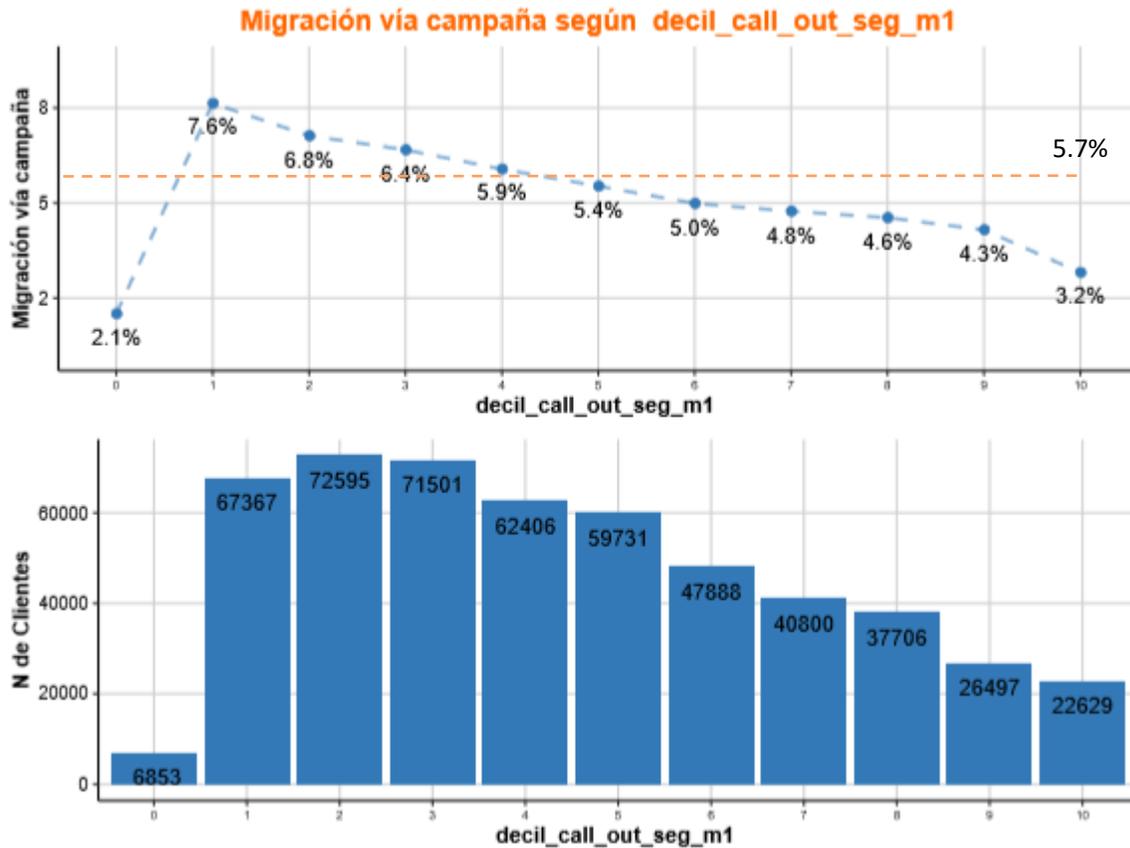


Figura 39: Tasa de migración con permanencia a suscripción vía campaña según llamadas emitidas en el mes previo (seg)

Fuente: Elaboración propia en R Studio

#### 4. Antigüedad del cliente (ant\_mnth\_cliente)

De la Figura 40 se aprecia que mientras mayor es la antigüedad del cliente, menor es la tasa de migración para dicho segmento. Por lo demás, las tasas de migración más altas se alcanzan para clientes con menos de seis meses de antigüedad en el mercado de prepago, experimentando una baja significativa para clientes cuya antigüedad es mayor a medio año.

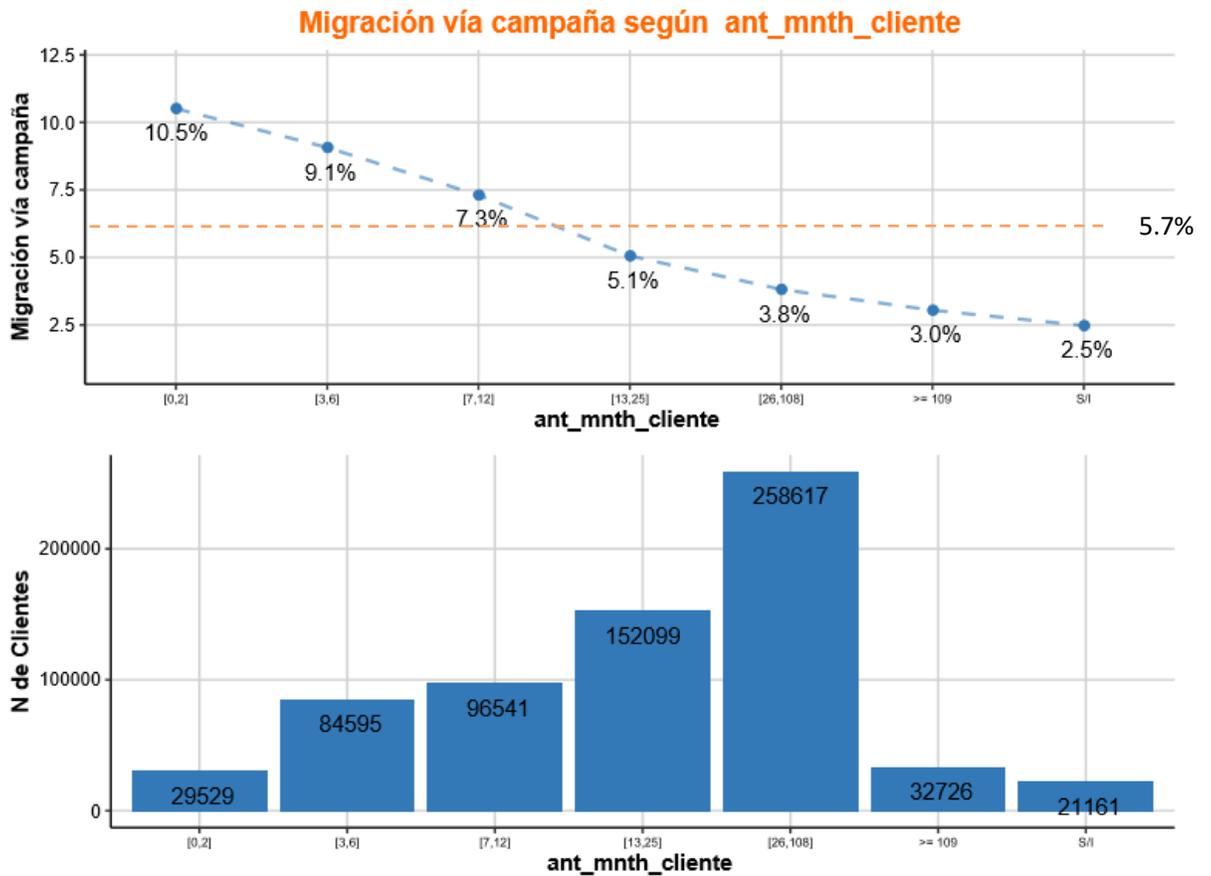


Figura 40: Tasa de migración con permanencia a suscripción vía campaña según antigüedad del cliente

Fuente: Elaboración propia en R Studio

## 5. Cantidad de segundos recibidos el penúltimo mes (decil\_call\_in\_seg\_m2)

Al igual que para la cantidad de segundos recibidos el último mes, a partir de la Figura 41 se observa que las mayores tasas de migración se alcanzan para los clientes con una mayor cantidad de segundos recibidos en el penúltimo mes.

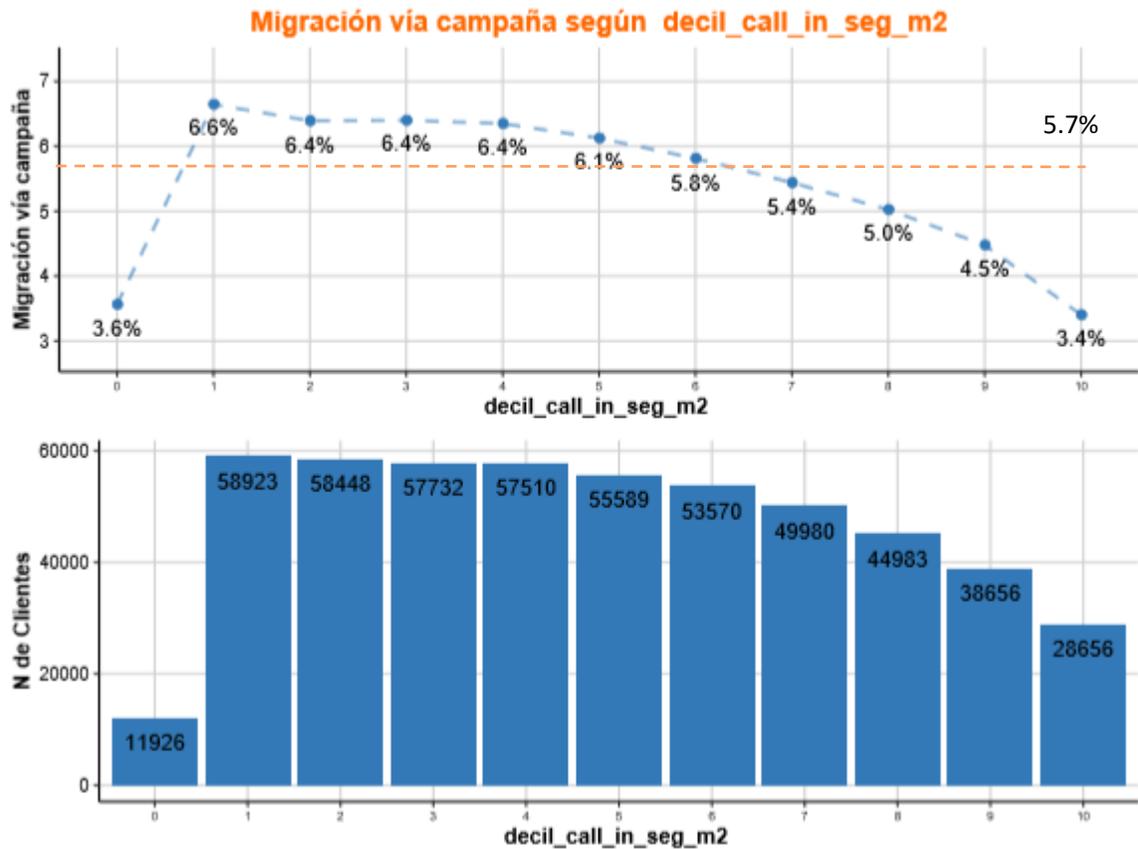


Figura 41: Tasa de migración con permanencia a suscripción vía campaña según llamadas recibidas en el penúltimo mes (seg)

Fuente: Elaboración propia en R Studio

## 6. Cantidad de segundos emitidos el penúltimo mes (decil\_call\_out\_seg\_m2)

Análogo a la variable anterior, los segundos asociados a las llamadas emitidas en el penúltimo mes disponible también resulta ser una variable importante.

A partir de la Figura 42, se observa que las tasas de migración son superiores a la tasa global para los tres deciles con mayor cantidad de segundos realizados vía llamadas.

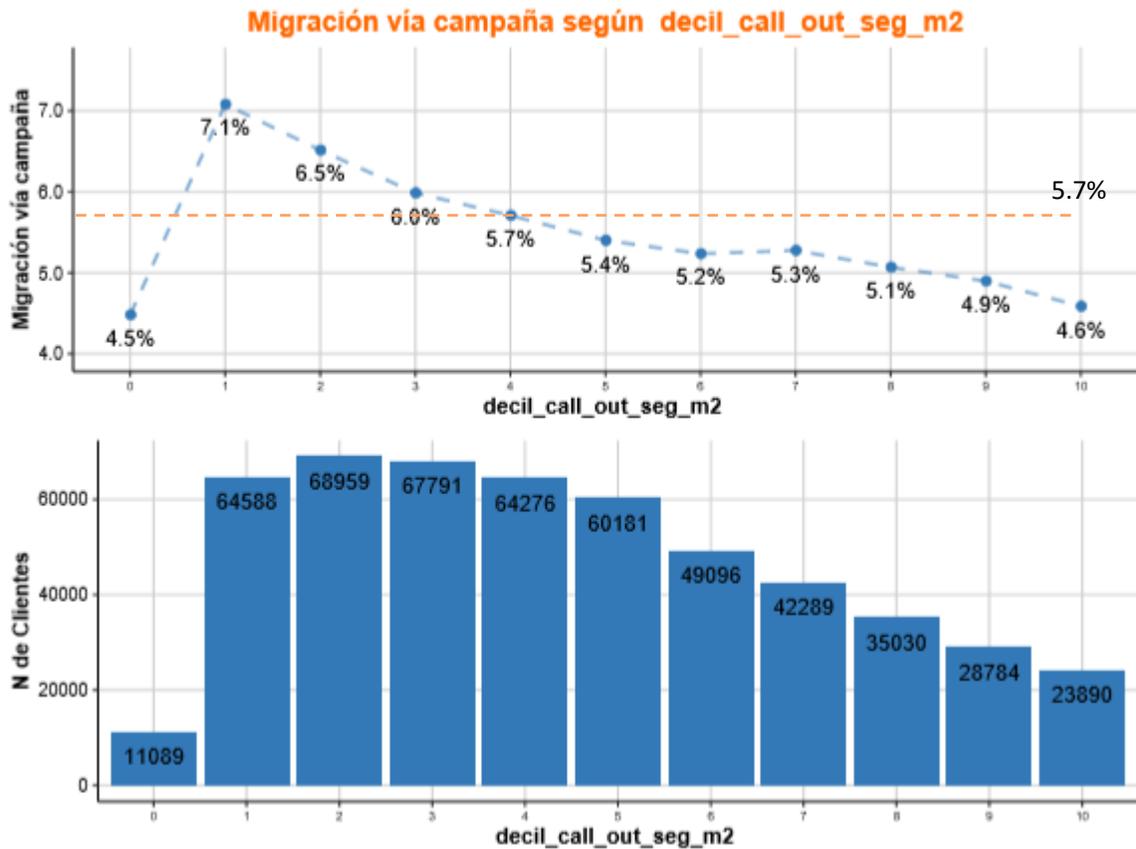


Figura 42: Tasa de migración con permanencia a suscripción vía campaña según llamadas emitidas en el penúltimo mes (seg)

Fuente: Elaboración propia en R Studio

## 7. Tráfico de datos en Facebook

El tráfico de datos en Facebook se utilizó en forma de quintiles, en el quintil 1 se encuentran los clientes con el mayor uso de esta red social.

A partir de la Figura 43, se observa que a medida que el tráfico en Facebook es mayor, la tasa de migración también incrementa. Por lo demás, es importante mencionar que en caso de que el cliente no use esta aplicación, o bien no posea información de uso, la tasa de migración disminuye significativamente.

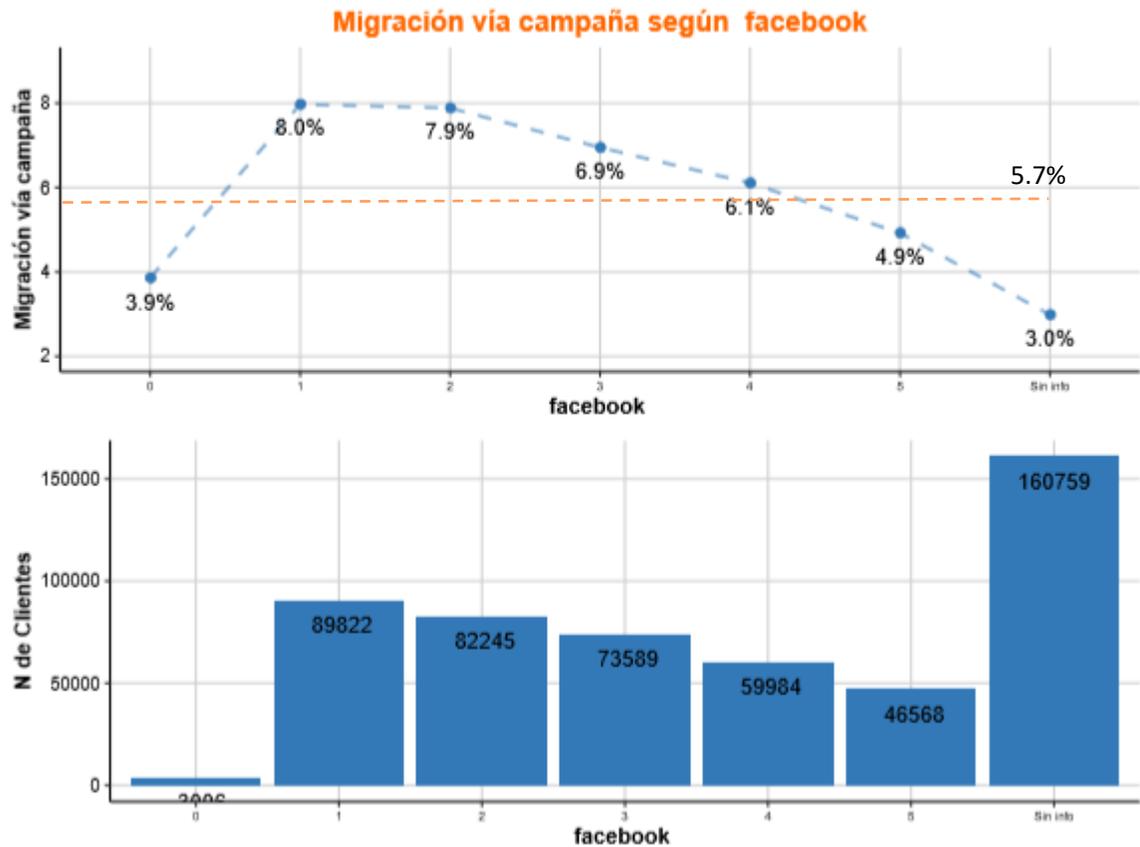


Figura 43: Tasa de migración con permanencia a suscripción vía campaña según tráfico en Facebook

Fuente: Elaboración propia en R Studio

### Caracterización de clientes propensos a migrar vía campaña

Si bien es importante mencionar que los clientes con una mayor propensión a migrar y permanecer en el mercado de pospago se obtienen a partir de la interacción de las variables utilizadas en los modelos, en base al análisis de las variables más importantes es posible tener una caracterización preliminar de este tipo de clientes.

De acuerdo con los análisis monovariantes presentados en la sección anterior, se observa que los clientes más propensos a migrar corresponden a clientes que no han sido contactados anteriormente por la campaña, esto teniendo en consideración que la base B1 posee una mayor conversión que B2, y los clientes con antigüedad menor a 6 meses son quienes tienen las tasas más altas de migración. Para clientes que

llevan mucho tiempo en la compañía siendo prepago, la probabilidad de migrar hacia suscripción es baja.

Por otro lado, estos clientes tanto realizan como reciben una gran cantidad de llamadas, lo cual se traduce en una alta cantidad de segundos emitidos y recibidos en el último mes disponible, que corresponde a dos meses previo a la migración. Además, es importante observar que este comportamiento lo poseen también en el penúltimo mes disponible, por lo que son clientes que tienen una alta actividad en cuanto al tráfico de llamadas en al menos los dos últimos meses utilizados en el modelo.

Por último, la variable asociada a Facebook entrega insights que este tipo de clientes no utilizan solamente el servicio de voz, sino que también son usuarios que utilizan el servicio de datos, siendo más propensos a migrar mientras mayor sea el uso de esta red social.

En contraste, los clientes menos propensos a migrar corresponden a clientes antiguos del mercado prepago, quienes en su mayoría ya han sido contactados anteriormente por la campaña y no han aceptado. Este tipo de clientes no se caracteriza por recibir ni emitir una gran cantidad de llamadas, ni tampoco de navegar en demasía mediante la red social Facebook.

### 7.4.3. Modelo de migración espontánea

#### **GBM**

- Lift acumulado

El gráfico muestra el lift acumulado hasta el primer percentil (5,1), quinto (3,47), décimo (2,9) y veinteavo percentil (2,42) en los datos de validación.

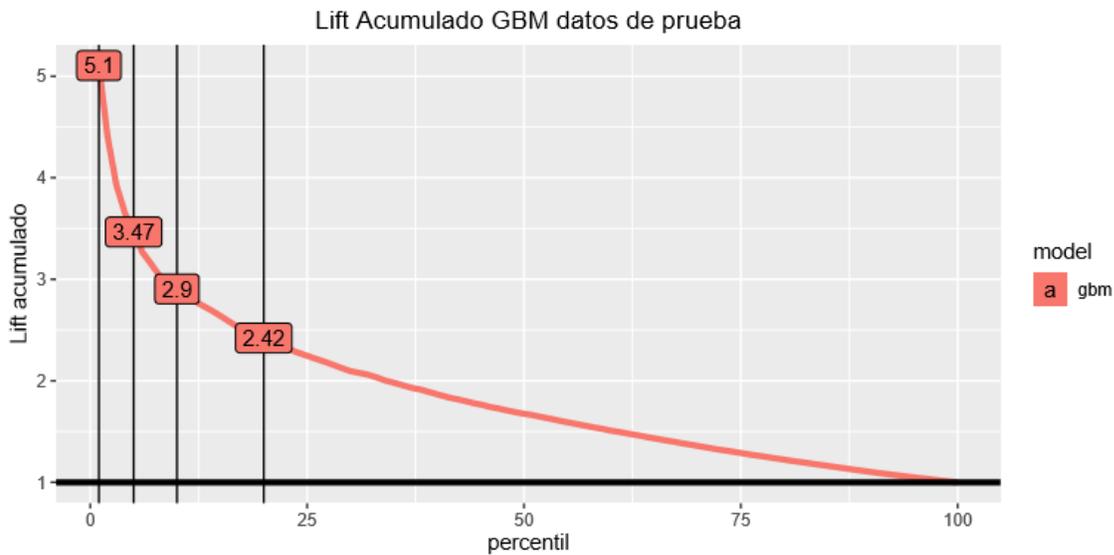


Figura 44: Lift acumulado modelo GBM, migración espontánea.

Fuente: Elaboración propia en R Studio

- Matriz de confusión

La matriz de confusión es posible calcularla para diferentes umbrales que separen a los clientes en una categoría u otra. En este caso, una clase corresponde a si el cliente migró vía espontánea y permaneció en el mercado de pospago en los próximos cuatro meses.

Al igual que para el modelo de migración vía campaña, para confeccionar la matriz de confusión para cada modelo, se escogió el umbral que maximiza el coeficiente MCC, variable utilizada para evaluar el desempeño de los modelos dada su utilidad en casos de problemas con desbalance de clases significativos.

En caso del modelo GBM, el umbral para clasificar a un cliente como migrado con permanencia en el mercado de pospago corresponde a 0,073, es decir, todos los clientes a los cuales el modelo les asigna un score superior a este valor son etiquetados como casos de éxito.

Con este umbral, el valor máximo del MCC corresponde a 0,062.

GBM	CLASE REAL		
		+	-
CLASE PREDICHA	+	7.074	462.881
	-	3.602	984.564

Tabla 10: Matriz de confusión modelo GBM, migración espontánea

- Curva ROC

El área bajo la curva ROC para el modelo GBM corresponde a 0,737.

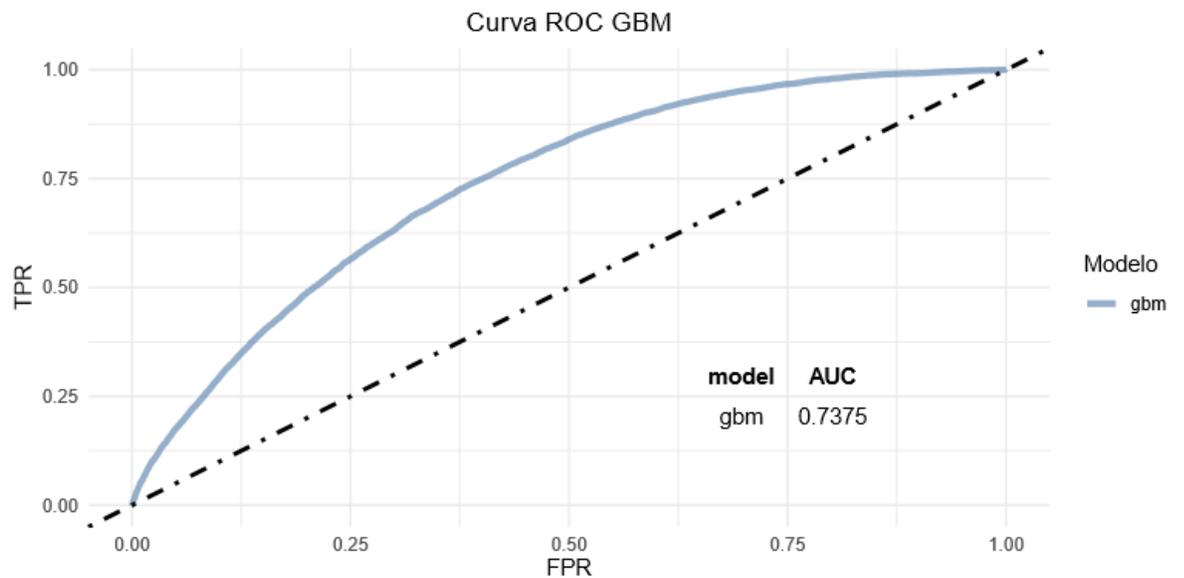


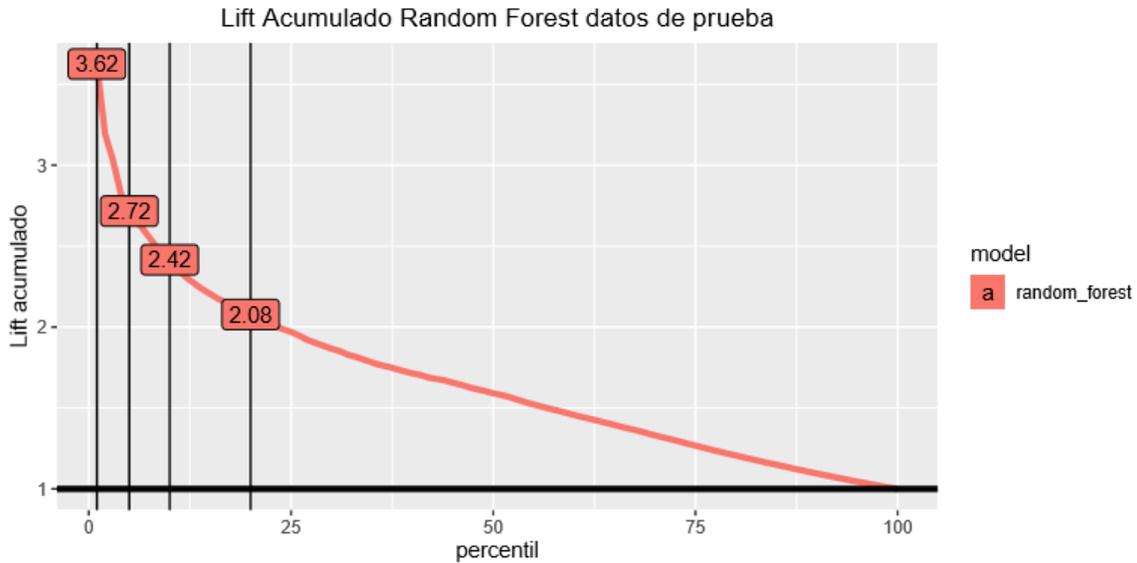
Figura 45: Curva ROC modelo GBM, migración espontánea.

Fuente: Elaboración propia en R Studio

## Random Forest

- Lift acumulado

El gráfico muestra el lift acumulado hasta el primer percentil (3,62), quinto percentil (2,72), décimo (2,42) y veinteavo percentil (2,08).



*Figura 46: Lift acumulado modelo Random Forest, migración espontánea.*

*Fuente: Elaboración propia en R Studio*

- Matriz de confusión

Para el modelo Random Forest, el umbral que maximiza la métrica MCC corresponde a 0,032. Utilizando este valor umbral, el MCC máximo es de 0,05.

Random Forest	CLASE REAL		
		+	-
CLASE PREDICHA	+	7.874	639.923
	-	2.802	810.522

Tabla 11: Matriz de confusión modelo Random Forest, migración espontánea

- Curva ROC

El área bajo la curva ROC para el modelo Random Forest es de 0,699.

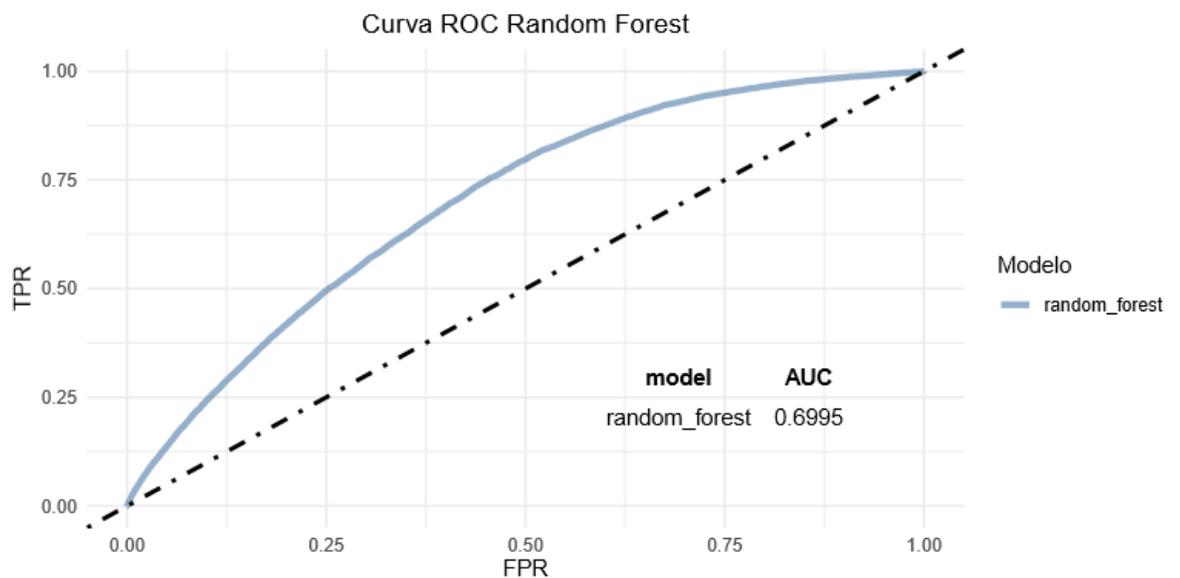


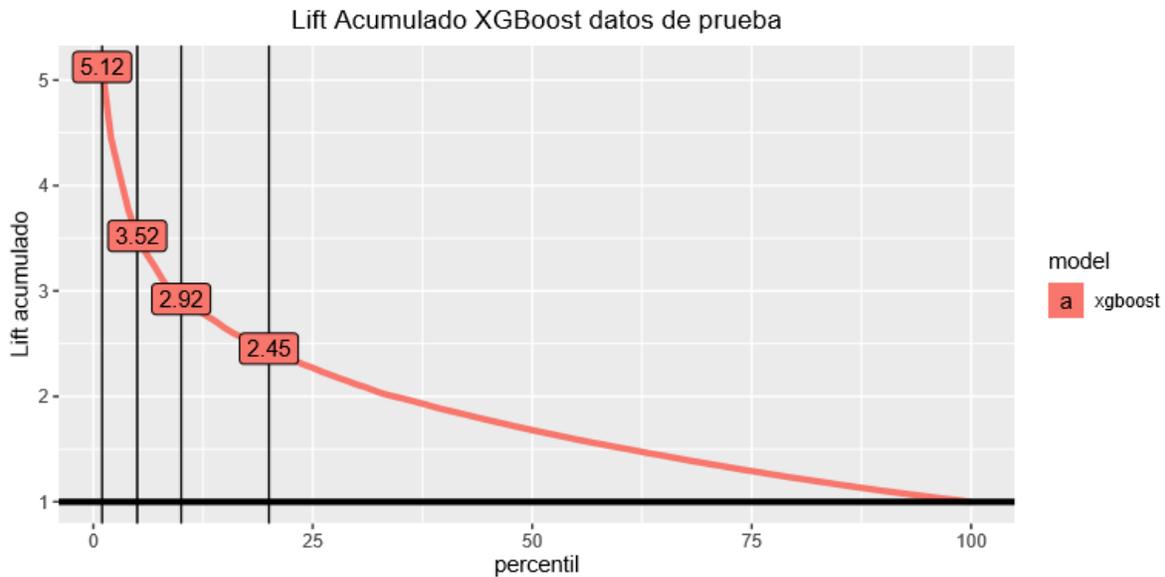
Figura 47: Curva ROC modelo Random Forest, migración espontánea.

Fuente: Elaboración propia en R Studio

## XGBoost

- Lift acumulado

El gráfico muestra el lift acumulado hasta el primer percentil (5,12), quinto (3,52), décimo (2,92) y veinteavo percentil (2,45).



*Figura 48: Lift acumulado modelo XGBoost, migración espontánea.*

*Fuente: Elaboración propia en R Studio*

- Matriz de confusión

El umbral que maximiza el valor MCC absoluto corresponde a 0,088, obteniendo un MCC máximo de 0,063.

<b>XGBoost</b>	<b>CLASE REAL</b>		
		+	-
<b>CLASE PREDICHA</b>	+	6.068	358.539
	-	4.608	1.088.906

Tabla 12: Matriz de confusión modelo XGBoost, migración espontánea

- Curva ROC

El área bajo la curva ROC para el modelo XGBoost es de 0,739.

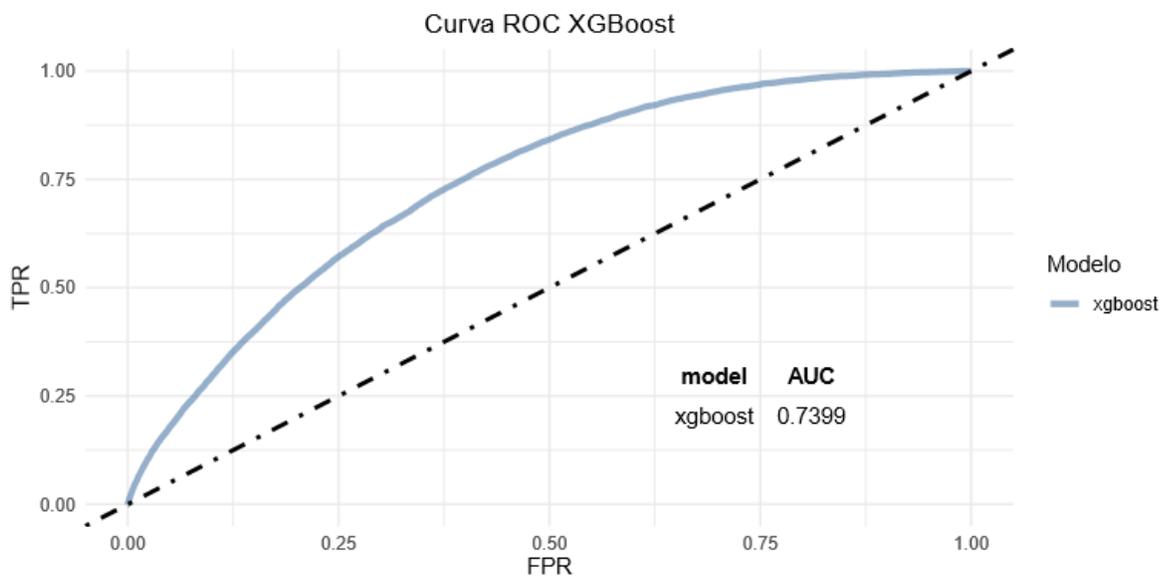


Figura 49: Curva ROC modelo XGBoost, migración espontánea.

Fuente: Elaboración propia en R Studio

### 7.4.3.1. Evaluación del modelo de migración espontánea

- Comparación de lift acumulado

A partir de la Figura 50, la cual muestra las curvas de lift acumulado para los tres modelos explicados anteriormente, se observa que el lift obtenido para el modelo Random Forest es peor en comparación a los otros dos modelos desarrollados.

Por lo demás, se aprecia que las curvas de lift para los modelos Gradient Boosting Machine y Extra Gradient Boosting Machine son similares. Pese a lo anterior, es posible observar que la curva asociada al modelo XGBoost se encuentra levemente por encima del modelo GBM, obteniendo un lift acumulado mayor para un mismo percentil.

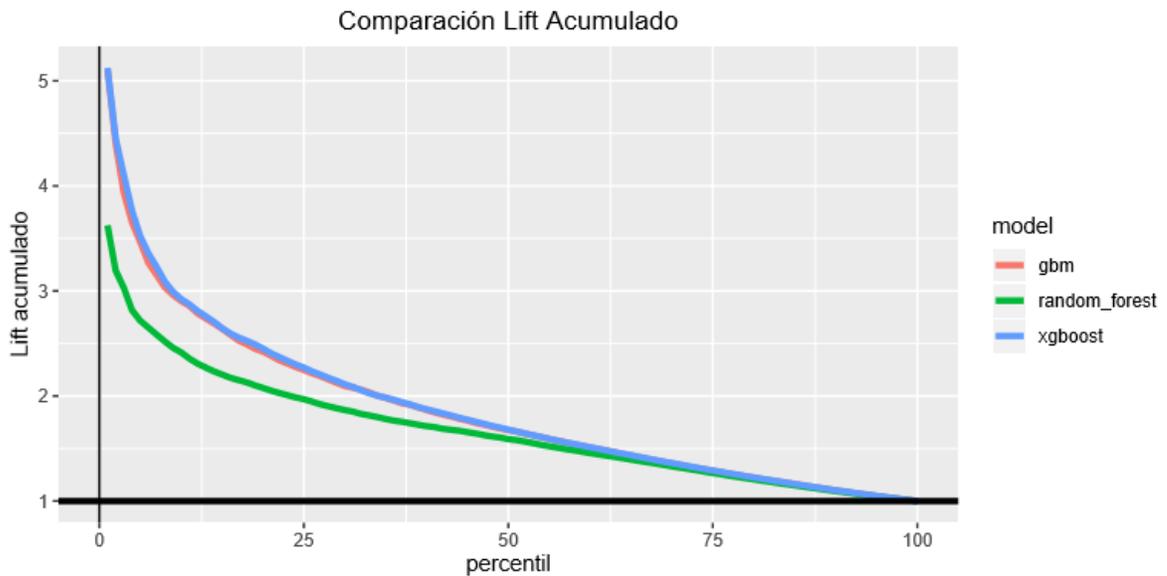


Figura 50: Comparación de lift acumulado para modelos de migración espontánea

Fuente: Elaboración propia en R Studio

- MCC

De forma análoga, la métrica asociada al máximo coeficiente de correlación de Matthews (MCC) es menor para el modelo Random Forest, mientras que este coeficiente es similar tanto para el modelo GBM como XGBoost, sin embargo, el valor es levemente superior para este último.

<b>Métrica</b>	<b>GBM</b>	<b>Random Forest</b>	<b>XGBoost</b>
MCC	0,062	0,05	0,063

*Tabla 13: Comparación de métrica MCC para modelo de migración espontánea*

- Comparación de curvas ROC

Al comparar las curvas ROC, se observa que la curva asociada al modelo Random Forest se encuentra por debajo de los demás, teniendo un AUC de 0,69.

Por otro lado, las curvas de los modelos GBM y XGBoost son similares, en efecto, en base a la Figura 51, se observa que estas se traslapan para la mayoría de los umbrales, resultando difícil visualizar ambas curvas en un mismo gráfico. Pese a lo anterior, el área bajo la curva del modelo XGBoost es levemente mejor que el modelo Gradient Boosting Machine.

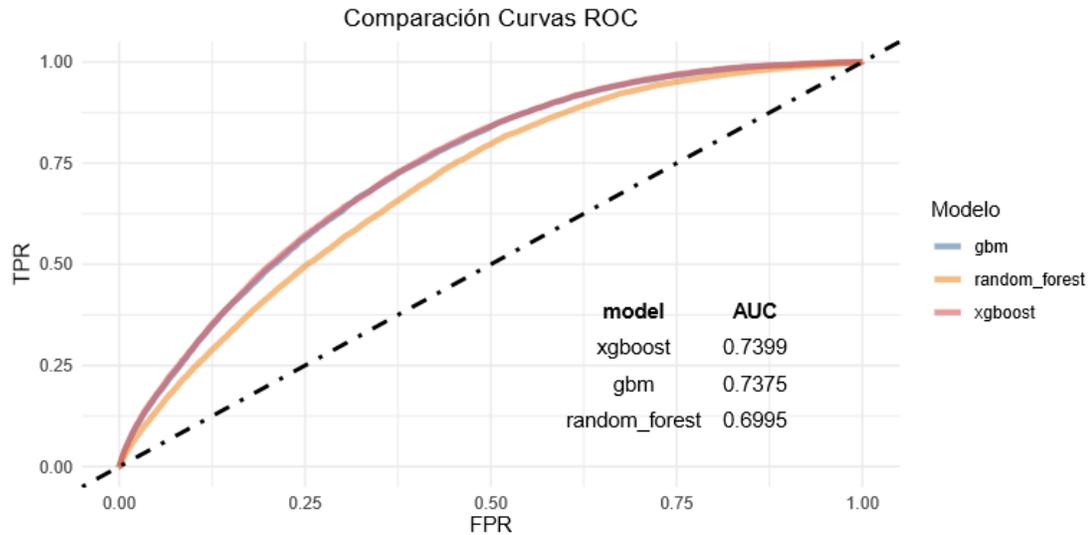


Figura 51: Comparación de curvas ROC para modelos de migración espontánea

Fuente: Elaboración propia en R Studio

- Mejor modelo

En base a los resultados obtenidos y la comparación de las métricas utilizadas para evaluar la performance de los distintos modelos planteados, es posible concluir que el mejor modelo para identificar a los clientes con mayor probabilidad de migrar hacia suscripción vía espontánea corresponde al algoritmo XGBoost, lo anterior se debe a que este posee las métricas más altas asociadas al área bajo la curva ROC (AUC), coeficiente de correlación de Matthews absoluto (MCC) y mayores tasas de lift acumulado.

## Análisis del mejor modelo

- Resumen efectividad

La Tabla 14 muestra los resultados obtenidos en los datos de validación del mejor modelo para predecir la migración espontánea con permanencia de al menos cuatro meses en el mercado de pospago. A partir de esta, es posible observar que en el 1% más propenso, la tasa de migración del grupo es 5,12 veces mayor a la tasa natural (0,73%) de migración espontánea del total de clientes prepago, logrando, por lo tanto, una tasa de migración de 3,75% en dicho grupo. Lo anterior, permite identificar al 5,1% de los clientes migrados con permanencia. De manera análoga, gestionando al 20% más propenso según el modelo, es posible identificar el 49% del total de clientes migrados dentro de un mes.

Dado que en la campaña se gestionan aproximadamente 177.000 clientes, esto corresponde a menos del 20% del total de clientes prepago activos dentro de un mes.

<b>Percentil</b>	<b>Lift Acumulado</b>	<b>Tasa acumulada</b>	<b>Porcentaje de migrados acumulados</b>
<b>1%</b>	5,12	3,75%	5,1%
<b>5%</b>	3,52	2,58%	17,6%
<b>10%</b>	2,92	2,14%	29,2%
<b>20%</b>	2,45	1,79%	49%
<b>100%</b>	1	0,73%	100%

*Tabla 14: Resumen de gestión según performance de mejor modelo de migración vía campaña*

- **Importancia de variables**

La Figura 52 muestra las siete variables más importantes según el modelo junto con su porcentaje de importancia, las cuales serán explicadas a continuación.

Al igual que para el modelo anterior, estas variables también resultan ser de las más importantes para los otros dos modelos desarrollados (Ver Anexo 3).

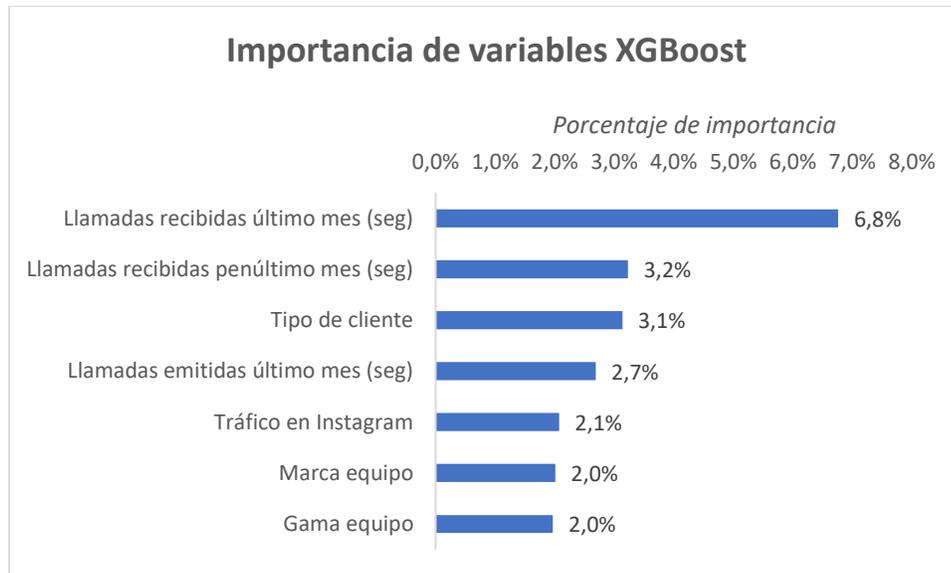


Figura 52: Importancia de variables modelo XGBoost, modelo de migración espontánea.

Fuente: Elaboración propia

### 1. Cantidad de segundos recibidos el último mes (decil\_call\_in\_seg\_m2)

La cantidad de segundos recibidos el último mes corresponde a la variable más importante del modelo. A partir de la Figura 53, se observa que mientras mayor es la cantidad de segundos recibidos, la tasa de migración para estos clientes también aumenta.

Particularmente, si se comparan las tasas de migración para cada decil en relación a la tasa de migración espontánea del total de la población (línea naranja), se aprecia que esta es mayor para los primeros cinco deciles con mayor cantidad de segundos recibidos.

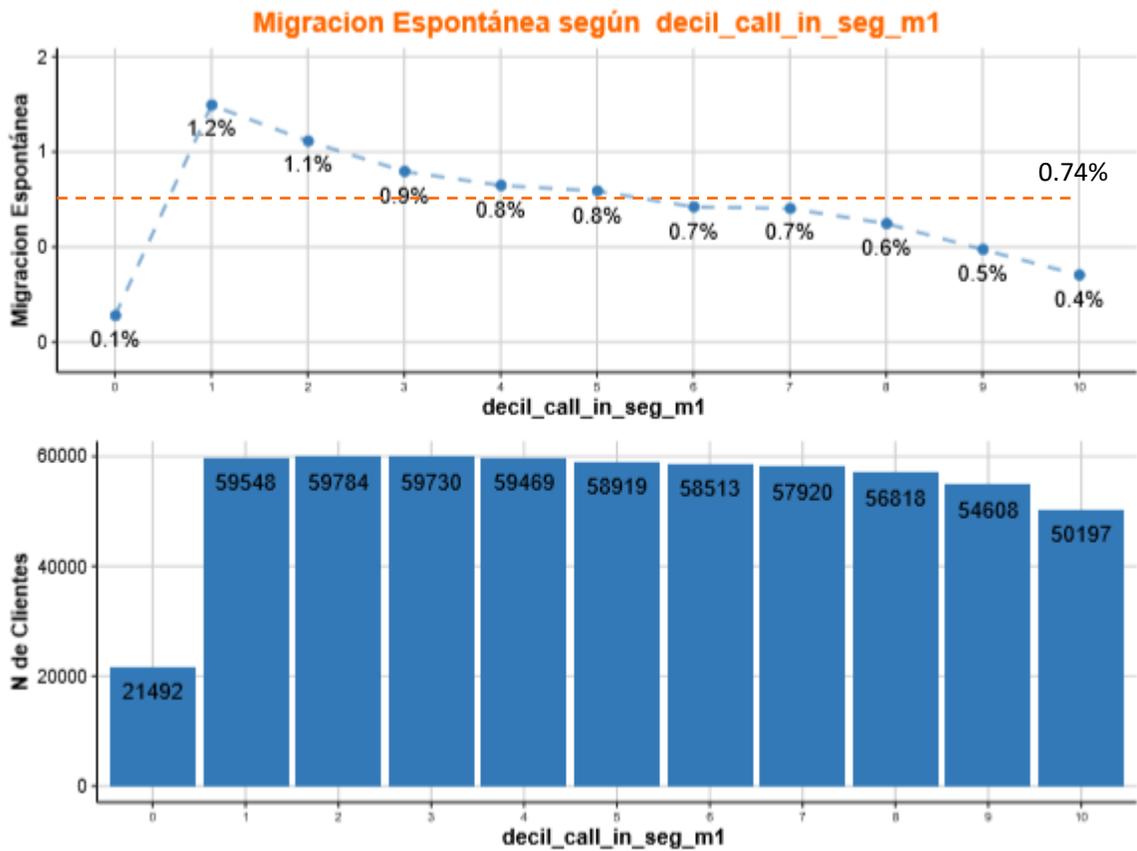


Figura 53: Tasa de migración espontánea con permanencia según cantidad de segundos recibidos el último mes

Fuente: Elaboración propia en R Studio

## 2. Cantidad de segundos recibidos el penúltimo mes (decil\_call\_in\_seg\_m2)

De manera análoga a la variable explicada anteriormente, se observa el mismo comportamiento para la cantidad de segundos recibidos el penúltimo mes. Es decir, mientras mayor sea la cantidad de segundos recibidos el penúltimo mes disponible, mayor es la tasa de migración hacia suscripción.

Al igual que en el caso anterior, las tasas de migración son superiores a la tasa total de migración espontánea para los primeros cinco deciles.

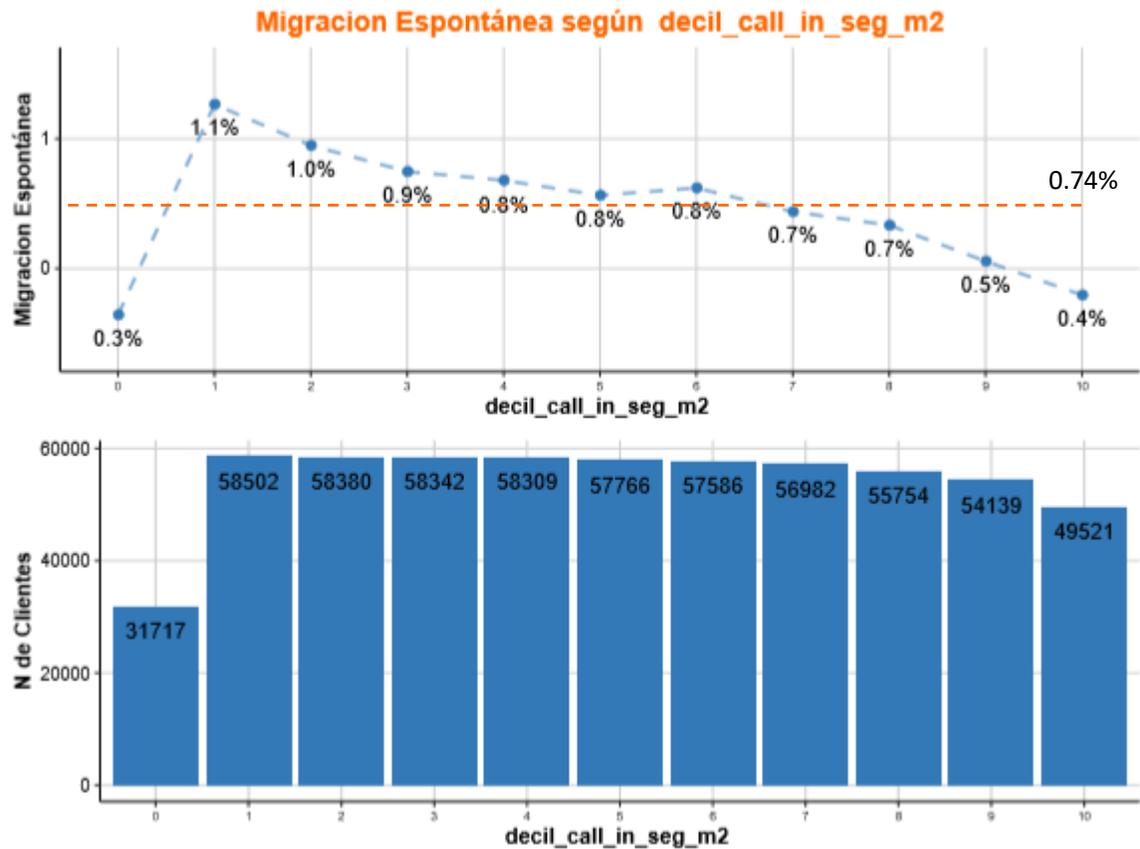


Figura 54: Tasa de migración espontánea con permanencia según cantidad de segundos recibidos el penúltimo mes.

Fuente: Elaboración propia en R Studio

### 3. Tipo de cliente

Tal como se explicó en la sección 1.3, la compañía clasifica a un cliente como de alto valor si en el último mes recarga más de \$8.000, y de bajo valor en caso contrario.

Adicionalmente, existen tres categorías que dependen del uso que estos tienen, los clientes de voz son usuarios que solamente utilizan su teléfono para realizar llamadas. Por otro lado, los clientes de datos usan su teléfono tanto para realizar llamadas como para navegar por internet, y los clientes de redes sociales realizan recargas para utilizar las promociones de redes sociales gratis.

A partir de la Figura 55, se observa que, si se compara los clientes según valor para una misma categoría de uso, los clientes de alto valor poseen una mayor tasa de migración en comparación a los de bajo valor. Por otro lado, los clientes de datos son quienes presentan una mayor tasa de migración, mientras que los clientes voz poseen las menores tasas de cambio hacia el mercado de pospago.

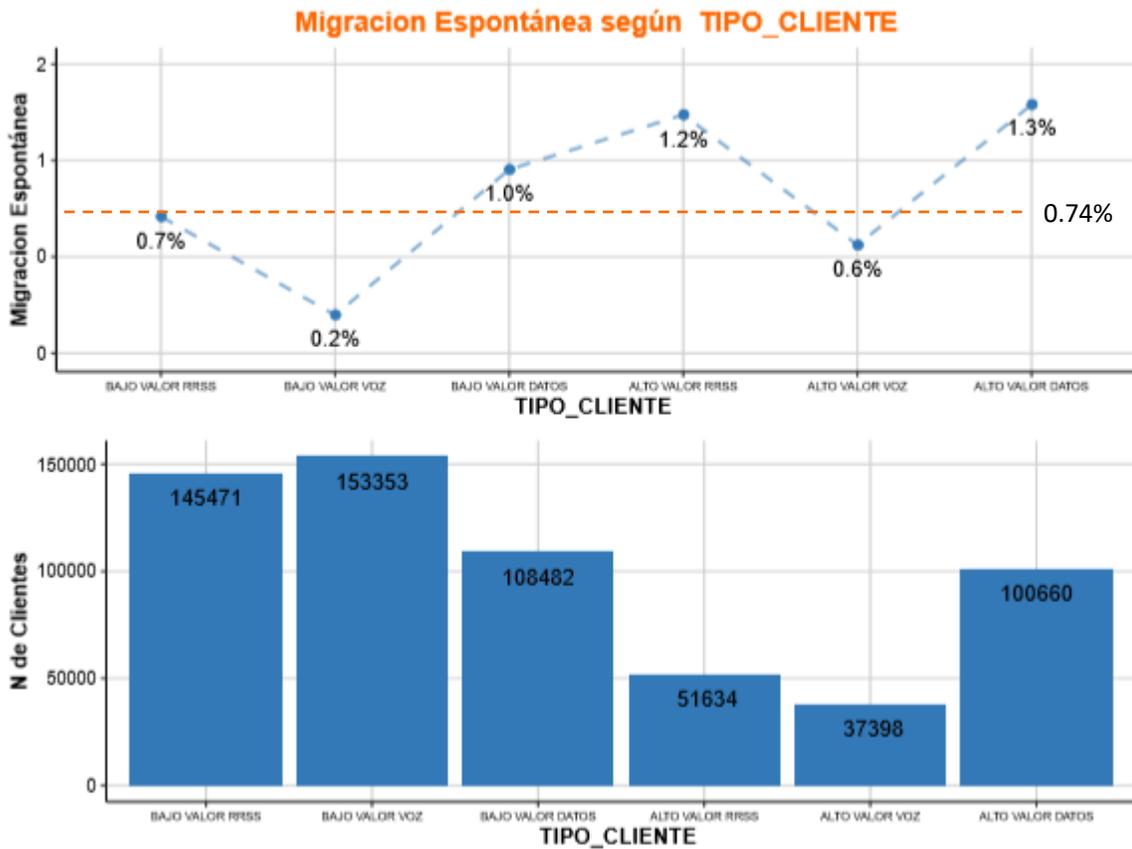


Figura 55: Tasa de migración espontánea con permanencia según Tipo de cliente

Fuente: Elaboración propia en R Studio

#### 4. Llamadas emitidas último mes (decil\_call\_out\_m1)

A partir de la Figura 56 se observa que las tasas de migración son mayores para los deciles que contienen los clientes con una mayor cantidad de segundos emitidos en el último mes disponible.

Al igual que para los segundos recibidos, mientras mayor es la cantidad de segundos emitidos, mayor es la tasa de migración.

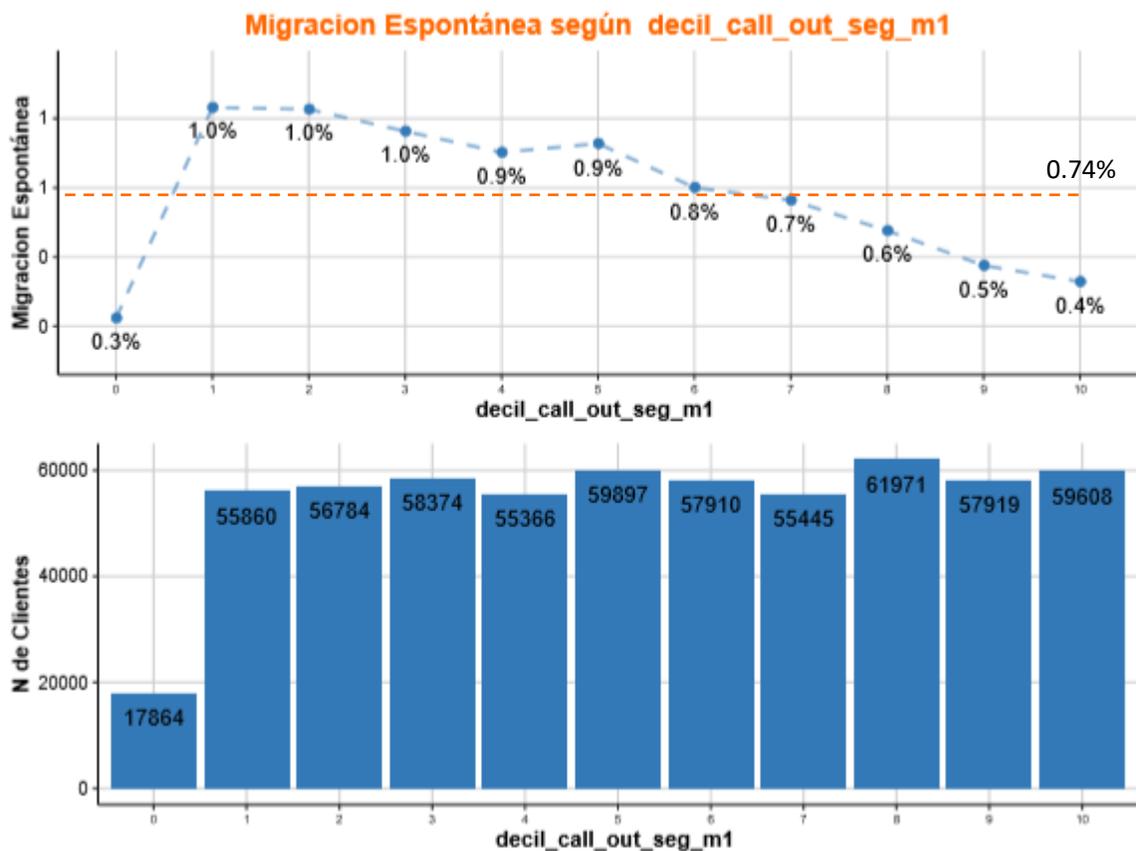


Figura 56: Tasa de migración espontánea con permanencia según cantidad de segundos emitidos el último mes.

Fuente: Elaboración propia en R Studio

## 5. Tráfico en Instagram

Mediante la Figura 57 se observa que las tasas de migración espontánea más altas se alcanzan para clientes con un mayor uso de la red social Instagram.

En particular, para el quintil con mayor tráfico de datos en esta aplicación, se observa una tasa de migración de más del doble que la tasa de migración espontánea total.

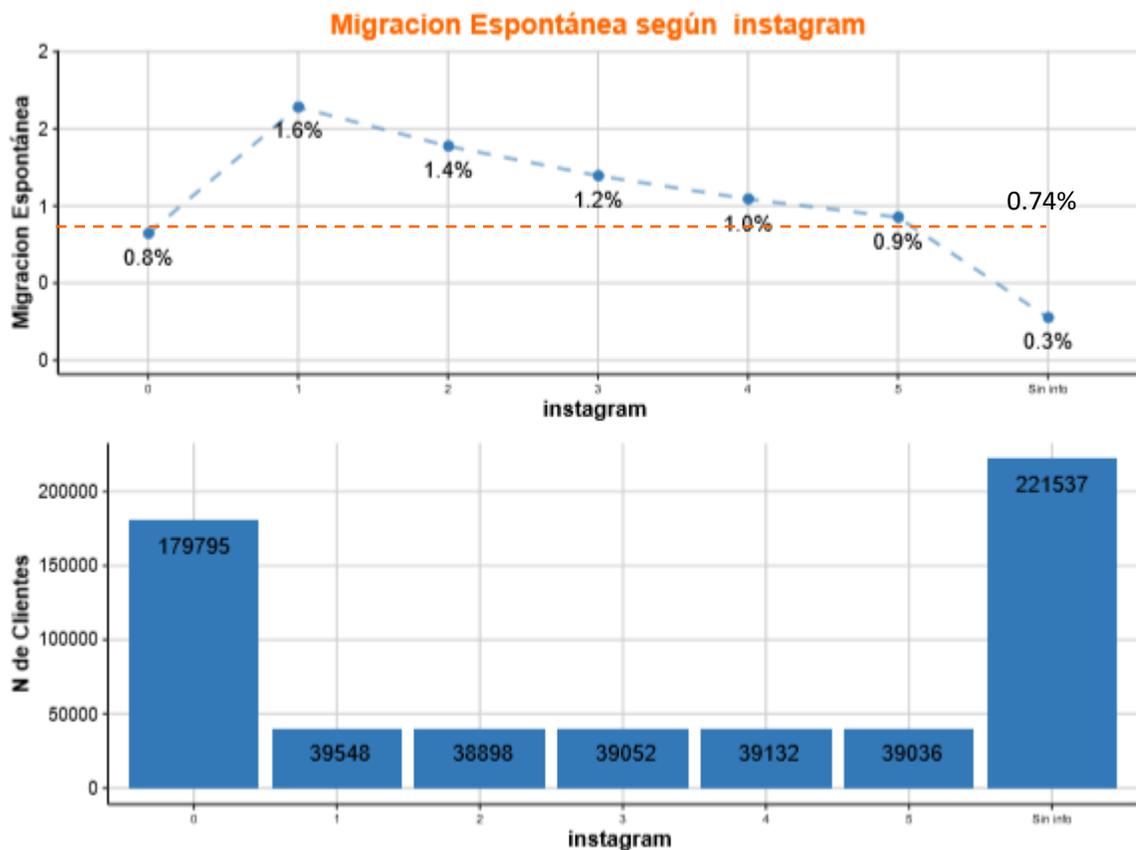


Figura 57: Tasa de migración espontánea con permanencia según tráfico en Instagram.

Fuente: Elaboración propia en R Studio

## 6. Marca del equipo

La Figura 58 muestra las tasas de migración de las ocho marcas de equipo más usadas por los clientes prepago, mientras que las demás marcas se agrupan en la categoría “Otro”. Por otro lado, también existe un grupo de clientes a los cuales se desconoce la información asociada al equipo, los cuales quedan en la categoría “Sin info”.

En base a la Figura 58 se observa que la marca con mayor tasa de migración corresponde a Apple (1,7%), siendo más del doble que la tasa de migración total. Además, existen otras marcas con tasas de migración altas, las cuales corresponden a Motorola, Samsung, LG y Huawei.

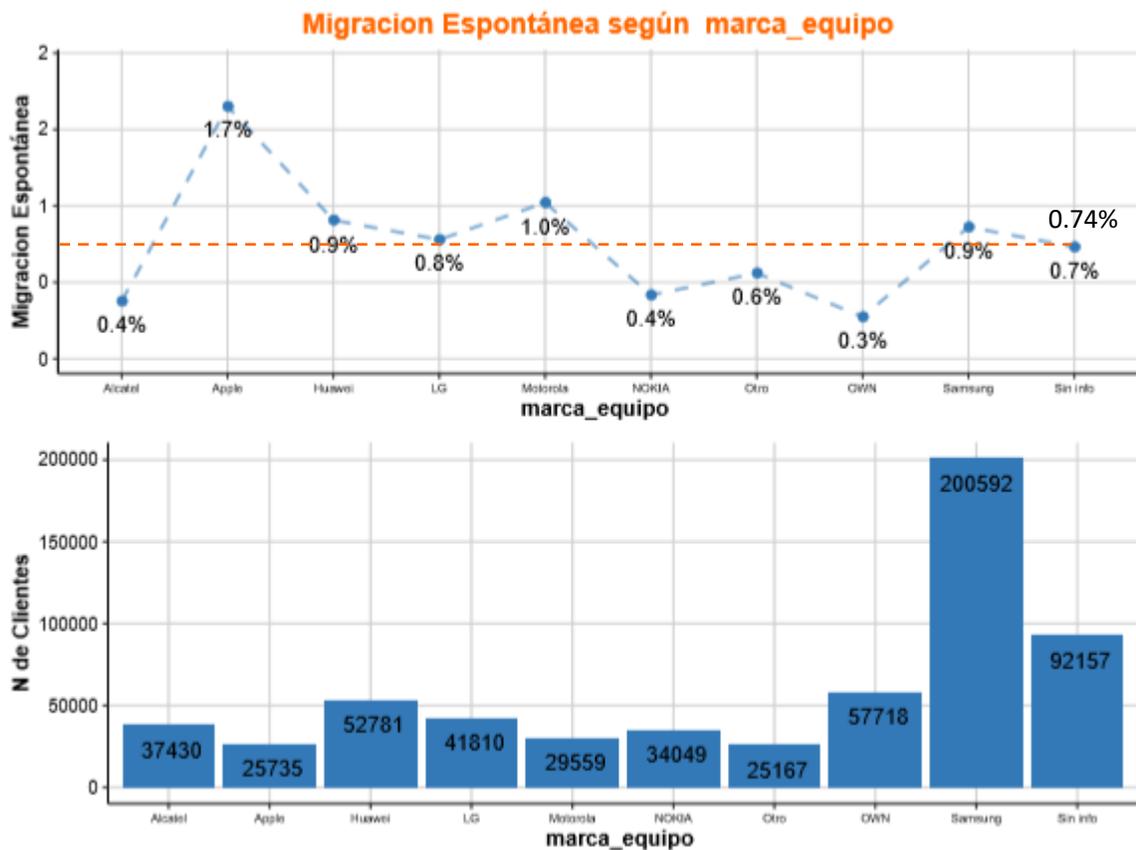


Figura 58: Tasa de migración espontánea con permanencia según marca del equipo.

Fuente: Elaboración propia en R Studio

## 7. Gama del equipo

Con respecto a la gama de los equipos, se aprecia que la tasa de migración es mayor para clientes con equipos de mejor gama, alcanzando la tasa de migración más alta en el conjunto de clientes que poseen equipos de gama alta. Sin embargo, la cantidad de clientes que poseen equipos de gama alta es menor en comparación a las otras categorías.

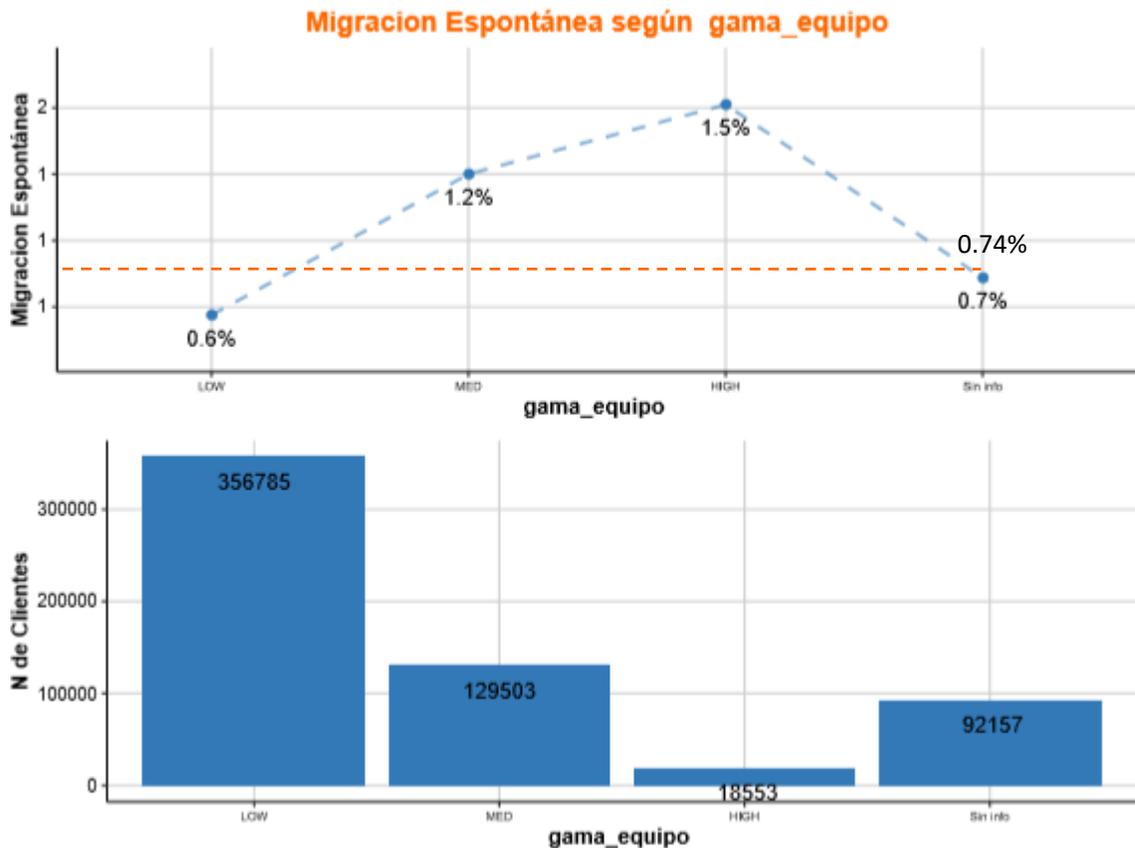


Figura 59: Tasa de migración espontánea con permanencia según gama del equipo

Fuente: Elaboración propia en R Studio

## **Caracterización de clientes propensos a migrar de forma espontánea**

Al igual que en la caracterización de los clientes propensos a migrar vía campaña se realizará una descripción de los clientes prepago con mayor probabilidad de migrar de forma espontánea hacia suscripción, lo anterior, con el fin de obtener un perfil para este tipo de clientes.

En base a los análisis de las variables más importantes obtenidas a partir del mejor modelo de migración espontánea, se observa que los clientes con mayor probabilidad de migrar tanto reciben como realizan una gran cantidad de llamadas, lo cual se traduce en una alta cantidad de segundos recibidos y emitidos el último mes disponible utilizado para la modelación del problema. Este comportamiento también se repite un mes antes, por lo cual, son clientes que en los últimos dos meses poseen un alto tráfico de llamadas.

Sumado a lo anterior, los usuarios prepago con mayor probabilidad de migrar se caracterizan por utilizar el servicio de datos, ya sea mediante la compra de bolsas para navegar por internet o la utilización de redes sociales, siendo Instagram una aplicación altamente utilizada en este grupo de clientes. Por otro lado, si el cliente solamente utiliza el servicio de voz, la probabilidad de migrar hacia un plan por contrato es baja.

El alto uso de datos y tráfico de llamadas en los últimos meses tiene relación a que los clientes recarguen un monto mayor al menos en el último mes, es por esto que los clientes de alto valor son más propensos a migrar que los de bajo valor.

Finalmente, los usuarios se caracterizan por poseer en su mayoría equipos de gama media o alta, teniendo como preferencia algunas marcas de equipo tales como Apple, Huawei, Samsung y LG.

En comparación, un cliente prepago poco propenso a migrar vía espontánea posee un bajo tráfico de llamadas en los últimos dos meses disponibles, además, tampoco se caracteriza por utilizar necesariamente el servicio de datos, si no que más bien corresponde a un perfil de clientes que principalmente realiza pocas llamadas telefónicas sin navegar en demasía por internet ni utilizar redes sociales (particularmente Instagram).

Dado su bajo tráfico de llamadas y de datos, estos clientes realizan un monto promedio de recarga menor en comparación a los más propensos. Por último, poseen equipos en su mayoría de gama baja, no inclinándose por marcas de equipos más conocidas como lo son Apple, Huawei, Samsung y LG.

## 7.5. Despliegue

Con el fin de evaluar los resultados obtenidos, en el mes de diciembre se utilizaron los modelos para perfilar la base enviada a campaña.

Para ello, se siguió la siguiente metodología que se resume en la Figura 60.

1. El primer paso consistió en tener el potencial de clientes a los cuales se les puede ofrecer la campaña en el mes de diciembre.
2. Posterior a lo anterior, en base a las reglas actuales de gestión, el área encargada de generar la campaña etiquetó a los clientes a los cuales les ofrecería la campaña normalmente. En este caso, el total de clientes inicialmente a cargar en la campaña corresponde a 144.410.
3. De manera análoga, utilizando los resultados de ambos modelos, se calculó la propensión de migrar y permanecer en pospago para cada cliente.

Para los clientes cuyo rango de recarga es mayor o igual a \$4.000, la propensión de migrar se obtiene aplicando a estos usuarios el mejor modelo de migración vía campaña.

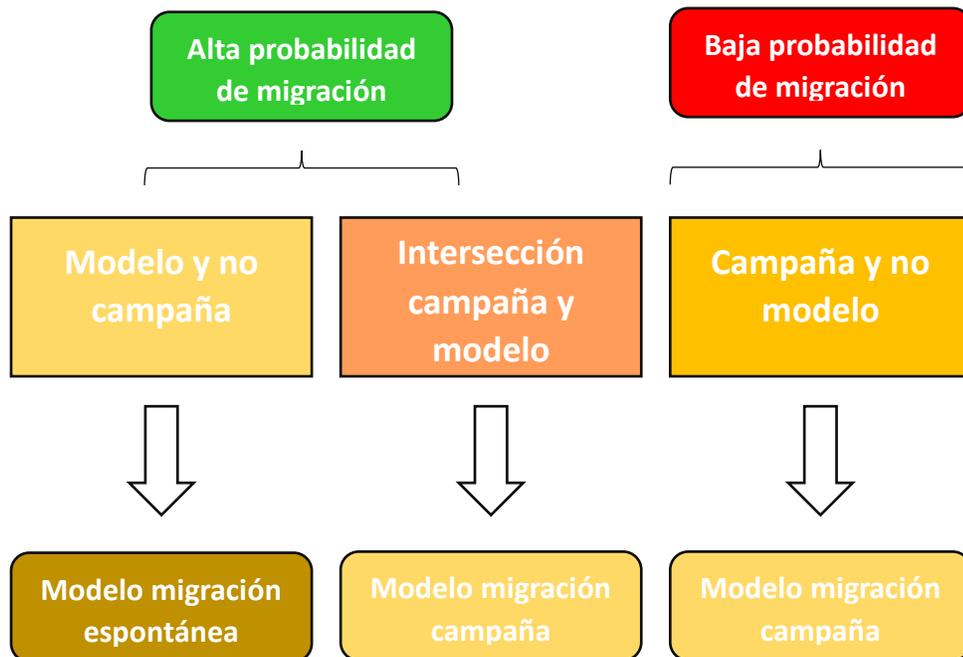
Por otro lado, para los clientes cuyo rango de recarga es menor a \$4.000, la propensión que se calcula es la probabilidad que tienen estos clientes para migrar a suscripción de forma espontánea, puesto que normalmente no serían gestionados por la campaña.

4. Una vez que se tiene calculada la propensión para cada cliente, se ordenaron estos de mayor a menor en base a su propensión.

Idealmente, si se utilizaran solamente los modelos para definir a que clientes ofrecer la campaña, la elección de la base de clientes a cargar correspondería a los 144.410 clientes con el mayor score entregado por los modelos. Sin embargo, dado que se quiere ver si el modelo posee una mejor performance que la gestión actual, se envió a campaña tres grupos diferentes descritos a continuación.

- 1) Intersección campaña y modelo: Corresponden a clientes a los cuales se les hubiera ofrecido la campaña originalmente, y el modelo también los predice como buenos candidatos a migrar.

- 2) Campaña y no modelo: Es un grupo de clientes que hubieran sido cargados según la gestión actual de la campaña, pero el modelo no los predice como buenos candidatos para contratar un plan.
- 3) Modelo y no campaña: Corresponde a clientes que no hubieran sido cargados en la campaña según las reglas actuales, pero el modelo los predice como buenos candidatos a migrar al mercado de pospago.



*Figura 60: Esquema de despliegue*

*Fuente: Elaboración propia*

### 7.5.1. Composición de base a cargar en campaña

Tal como se mencionó en las secciones anteriores, durante los últimos meses el número de clientes posibles a cargar en campaña con rango de recarga superior a \$4.000 ha disminuido considerablemente.

Particularmente, para el mes de diciembre, dado la reducción de la base potencial, la metodología consistió en cargar a todos los clientes cuyo rango de recarga es superior a \$4.000, y luego completar la base resultante con clientes con rango de recarga menor a dicho monto. Estos, normalmente son elegidos en base a reglas de negocio, priorizando a los clientes con promedio de recarga más alto que hayan recargado en los tres meses, luego a clientes que hayan recargado en dos de los últimos tres meses, y finalmente a quienes hayan recargado solo 1 mes.

Debido a lo anterior, los clientes pertenecientes al grupo Modelo y no campaña corresponden solamente a clientes cuyo rango de recarga es menor a \$4.000. La composición de la base se resume en la Tabla 15.

<b>Grupo</b>	<b>Cantidad de clientes</b>	<b>Porcentaje</b>
Intersección campaña y modelo	81.795	56,6%
Campaña y no modelo	28.878	20%
Modelo y no campaña	33.737	23,4%
<b>Total</b>	<b>144.408</b>	<b>100%</b>

*Tabla 15: Composición de base cargada a campaña en diciembre 2019*

La Tabla 16 muestra la composición de la base cargada a campaña según los rangos de recarga descritos anteriormente. A nivel global se observa que el grupo mayoritariamente gestionado corresponde a los clientes pertenecientes al rango de recarga menor a \$4.000, lo cual se debe a la inclusión del grupo Modelo y no campaña.

Si se analizan los grupos Intersección campaña y modelo, y Campaña y no modelo, se observa que el grupo de mayor porcentaje de clientes gestionados pertenece al rango de recarga entre \$4.001 y \$6.000.

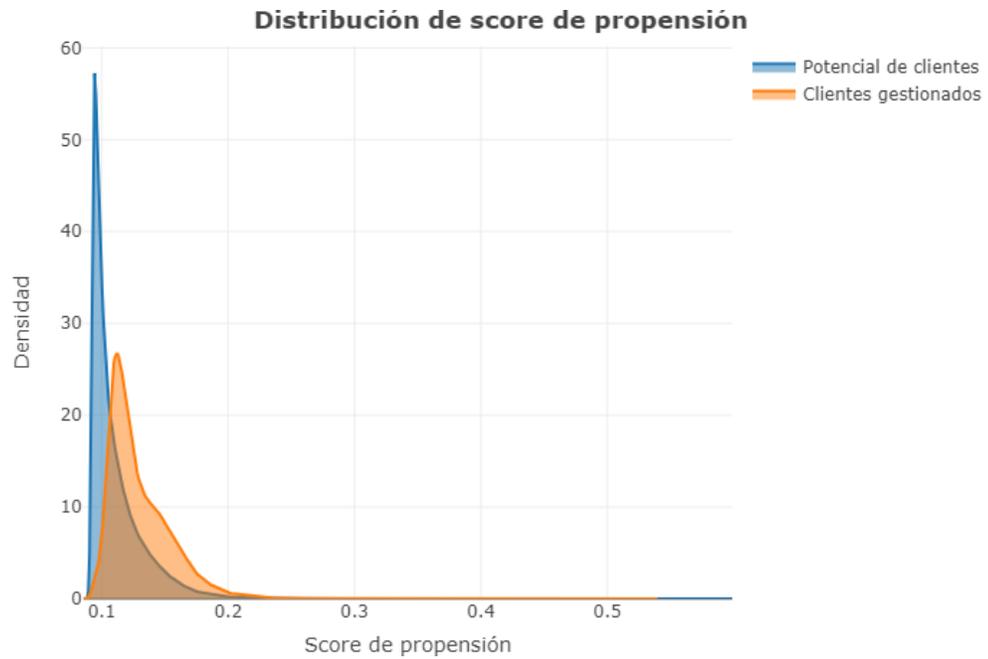
<b>Rango de recarga</b>	<b>Intersección campaña y modelo</b>	<b>Campaña y no modelo</b>	<b>Modelo y no campaña</b>	<b>Total</b>
< \$4.000	19%	17,3%	100%	37,6%
[\$4.001, \$6.000]	37,9%	43,9%		30,3%
[\$6.001, \$8.000]	19,4%	21%		15,2%
[\$8.001, \$10.000]	14%	11,6%		10,3%
[\$10.001, \$12.000]	7,4%	4,7%		5,2%
[\$12.001, \$15.000]	2,1%	1,5%		1,5%

*Tabla 16: Detalle de composición de base cargada a campaña de migración en diciembre 2019 según grupo de gestión*

### 7.5.2. Distribución de scores de propensión

La Figura 62 compara la distribución del score de propensión obtenido a partir de los modelos para el total de potenciales clientes a ofrecer la opción de contratar un plan, versus los clientes cargados finalmente en la campaña.

A partir de esta, se observa que la distribución de los clientes gestionados se encuentra desplazada hacia la derecha dada la mayor propensión que les asigna el modelo.

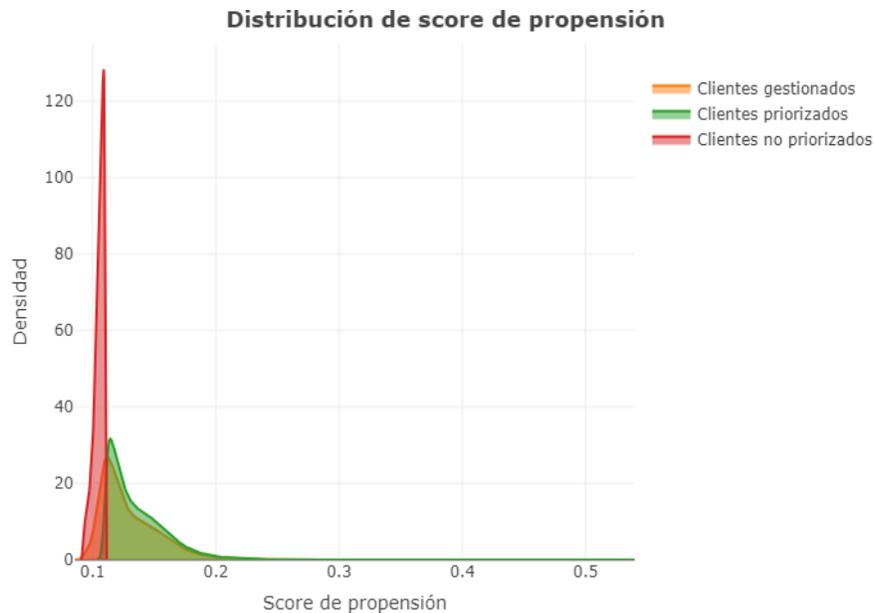


*Figura 61: Distribución de clientes según score de propensión asignados por el modelo*

*Fuente: Elaboración propia en R Studio*

La Figura 63 compara la distribución de los clientes priorizados por el modelo, pertenecientes a los grupos Intersección campaña y modelo, y Modelo y no campaña, con los clientes no priorizados, que corresponden al grupo Campaña y no modelo.

En base a la Figura 63, se observa que los clientes no priorizados son quienes presentan una distribución desplazada hacia la izquierda, teniendo como score máximo 0,11.



*Figura 62: Distribución de score de propensión según grupo de gestión*

*Fuente: Elaboración propia en R Studio*

## 8. Evaluación de campaña

En esta sección se evalúan y discuten los resultados obtenidos a partir de la campaña de migraciones del mes de diciembre, la cual incorpora los modelos desarrollados y explicados en la sección previa.

### 8.1. Evaluación general

Si se analiza el resultado de la campaña a nivel global, se observa que esta posee una tasa de migración del 9%. Sin embargo, se aprecian diferencias significativas entre los diferentes grupos explicados previamente.

En base a la Tabla 17, es posible identificar que el grupo con mejor tasa de conversión corresponde al grupo Modelo y no campaña, mientras que el de menor tasa es el grupo de Campaña y no modelo.

Este primer resultado es consistente a partir de los modelos obtenidos, puesto que el grupo con menor tasa de migración corresponde a un grupo de clientes que el

modelo indica que no debiesen ser cargados en la campaña dada su baja propensión a migrar.

<b>Grupo</b>	<b>Cientes gestionados</b>	<b>Cientes migrados</b>	<b>Tasa de migración</b>
Intersección campaña y modelo	81.795	8.003	9,8%
Campaña y no modelo	28.878	1.123	3,9%%
Modelo y no campaña	33.737	3.875	11,4%
<b>Total</b>	<b>144.408</b>	<b>12.983</b>	<b>9%</b>

*Tabla 17: Tasas de migración según grupo de gestión*

## 8.2. Evaluación según rango de recarga

Teniendo en consideración que existen clientes con distinto comportamiento previo en base a sus recargas, resulta interesante analizar las tasas de conversión según los tramos descritos.

En base a la Tabla 18, se observa a nivel global que las tasas de migración varían para cada segmento. El grupo con rango de recarga menor a \$4.000 presenta la mayor tasa de conversión, mientras que los clientes cuya recarga se encuentra entre \$8.001 y \$10.000 es el grupo con menor tasa de aceptación de la campaña.

Al comparar las tasas de conversión entre el grupo Campaña y no modelo e Intersección campaña y modelo, se observa que el primero de estos presenta menor tasa de conversión para todos los segmentos, siendo esta diferencia estadísticamente significativa con un 99% de confianza (Ver Anexo 4).

Por otro lado, al comparar el grupo Modelo y no campaña con los demás, es posible notar que este grupo presenta una mejor tasa de conversión que el grupo Campaña y no modelo. Ahora bien, al evaluar el grupo Modelo y no campaña con el de Intersección campaña y no modelo, su tasa de conversión, a priori, parece ser superior, sin embargo, en base al test de proporciones, no es posible afirmar que la

tasa del grupo Modelo y no campaña es estadísticamente superior al del grupo Intersección campaña y modelo.

Rango de recarga	Tasa de migración			
	Campaña y no modelo	Intersección campaña y modelo	Modelo y no campaña	Total
< \$4.000	5,1%	11,2%	11,4%	10,8%
[\$4.001, \$6.000]	3,8%	10,1%		8,3%
[\$6.001, \$8.000]	3,6%	9,1%		7,6%
[\$8.001, \$10.000]	3,7%	8,4%		7,3%
[\$10.001, \$12.000]	2,6%	8,9%		7,7%
[\$12.001, \$15.000]	3,8%	9,2%		8,2%
<b>Total</b>	<b>3,9%</b>	<b>9,8%</b>	<b>11,4%</b>	<b>9%</b>

*Tabla 18: Tasa de migración según rango de recarga*

Los resultados explicados recientemente son consistentes con lo esperado a partir de los modelos, puesto que las tasas de migración para el grupo de Intersección Campaña y Modelo son superiores en todos los tramos de recarga al grupo de Campaña y no modelo. De forma análoga, el grupo Modelo y no campaña también presenta mejor tasa de conversión en comparación al grupo Campaña y no modelo.

### 8.3. Análisis según deciles de propensión

Con el fin de evaluar la efectividad de predicción del modelo en base al score de propensión obtenido para cada cliente, se agrupó la base cargada en campaña según deciles de propensión.

En el decil 1 se encuentran los clientes prepago a los cuales el modelo les asigna una mayor probabilidad de contratar un plan en el mercado de suscripción, y en el decil

10 están los clientes que el modelo les asigna la probabilidad más baja de migrar en relación al total de clientes gestionados en la campaña.

Los clientes pertenecientes al grupo Campaña y no modelo corresponden a clientes que se ubican en los deciles de propensión 9 y 10, mientras que los clientes asignados a los grupos de Intersección modelo y campaña junto con Modelo y no campaña, se encuentran ubicados en los primeros ocho deciles.

A partir de la Tabla 19, se observa que las tasas más altas se alcanzan en los deciles con mayor probabilidad de migrar. En efecto, se aprecia un comportamiento estrictamente decreciente en las tasas de migración según los deciles de propensión, es decir, la tasa de migración disminuye a medida que el decil posee una menor probabilidad del fenómeno en estudio.

<b>Deciles de propensión</b>	<b>Cantidad de clientes</b>	<b>Cientes migrados</b>	<b>Tasa de migración</b>
1	14.442	2.010	13,9%
2	14.441	1.972	13,7%
3	14.442	1.822	12,6%
4	14.442	1.637	11,3%
5	14.441	1.435	9,9%
6	14.442	1.144	7,9%
7	14.441	1.009	7%
8	14.441	831	5,8%
9	14.441	656	4,5%
10	14.442	467	3,2%

*Tabla 19: Tasa de migración según deciles de propensión*

Dado que en los últimos meses se están incorporando clientes de bajo valor que antes no eran considerados en la campaña, resulta útil también analizar la composición de los deciles según el rango de recarga y su tasa de efectividad por separado.

La Tabla 20 muestra la tasa de migración por decil para dos grupos diferentes, el primero corresponde a clientes cuyo rango de recarga es menor a \$4.000, y el otro a clientes cuyo rango de recarga es superior a este monto.

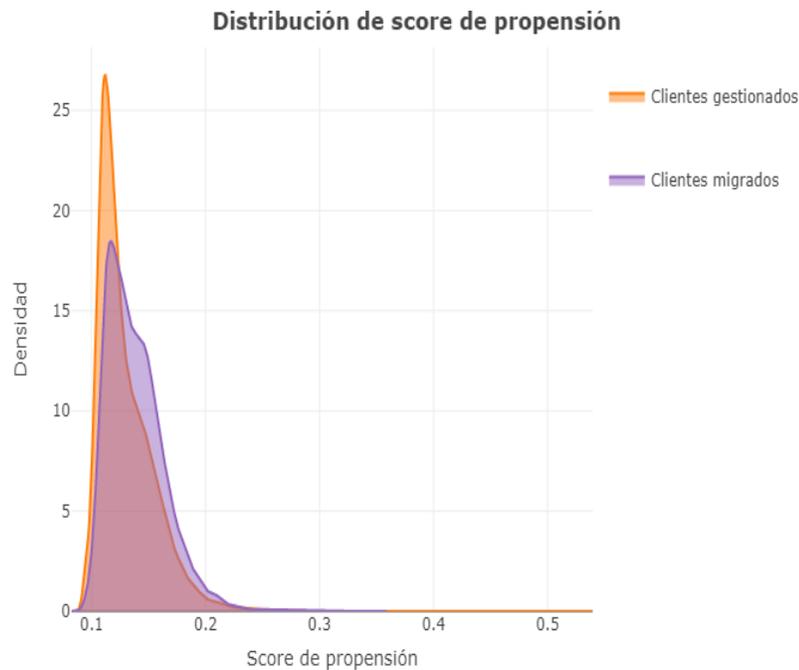
A partir de la tabla recién mencionada, se observa que, para los primeros tres deciles con mayor probabilidad de contratar un plan, la tasa de migración es mayor para los clientes con monto de recarga superior a \$4.000, sin embargo, a partir del cuarto decil, la tasa de migración de los clientes de bajo valor es superior.

<b>Deciles de propensión</b>	<b>Tasa de migración según rango de recarga</b>	
	<b>&lt; \$4.000</b>	<b>[\$4.001, \$15.000]</b>
1	12,1%	23,2%
2	12,3%	18,6%
3	11,7%	14,5%
4	11,7%	11%
5	10,8%	9,6%
6	8,3%	7,9%
7	7,3%	6,9%
8	6,4%	5,7%
9	5,6%	4,4%
10	4,9%	2,7%

*Tabla 20: Tasa de migración según deciles de propensión y rango de recarga*

## 8.4. Análisis según score de propensión

Al analizar la distribución de score de los clientes migrados hacia suscripción en comparación a la base gestionada por la campaña, se aprecia que los clientes migrados poseen una distribución desplazada hacia la derecha, lo anterior es consistente con lo esperado a partir de los modelos, ya que estos les asignan una mayor probabilidad de migrar hacia el mercado de pospago.



*Figura 63: Distribución de densidad de clientes gestionados y migrados*

*Fuente: Elaboración propia en R Studio*

## 8.5. Análisis de cumplimiento del plan sugerido

Tal como se explicó en la sección 6.1.4, tras los análisis realizados que muestran que los clientes no migran hacia el plan sugerido, por lo tanto, se indagó si con las nuevas instrucciones de gestión dadas al call center la tasa de incumplimiento de la campaña mejoró.

Para ello, se comparó el porcentaje de incumplimiento del último mes previo a los cambios de gestión, correspondiente a septiembre, con la campaña del mes de diciembre. Esta métrica se calcula como el porcentaje de clientes que migraron hacia un plan con cargo fijo menor a los dos planes sugeridos inicialmente.

En el mes de septiembre, la tasa de incumplimiento, considerando el total de los clientes migrados vía outbound es del 23%, sin embargo, esto incluye a clientes con rango de recarga superior a \$15.000, los cuales no fueron gestionados en el mes de diciembre. Es importante mencionar que este grupo de clientes en su mayoría migra hacia un plan no sugerido de menor valor, por lo que considerar este grupo incrementa significativamente la tasa de incumplimiento.

Dejando fuera del análisis a los clientes cuyo rango de recarga es superior a \$15.000, la tasa de incumplimiento para el mes de septiembre es de un 13,5%.

Por otro lado, para la campaña del mes de diciembre, la tasa experimenta una baja significativa, teniendo un incumplimiento de tan solo un 7,5%. Sin embargo, parte de esta baja se debe a que en el mes de diciembre se gestionó un menor porcentaje de clientes con rango de recarga alto.

## 8.6. Análisis de utilidad de la campaña

La campaña de migraciones busca aumentar el valor que generan los clientes para la compañía por medio de su traspaso al mercado de pospago, lo anterior ocurre si el cliente contrata un plan cuyo valor es mayor al monto de recarga promedio que realizaba durante su estadía en el mercado de prepago. Para evaluar lo anterior, se calculó la utilidad de un cliente migrado como la diferencia entre el valor del plan que elige y el promedio de recarga de los últimos tres meses en el mercado prepago previo a la migración.

La Tabla 21 compara la utilidad de la campaña realizada en diciembre con la del mes de septiembre. En esta se muestra la utilidad generada por cada segmento de recarga, junto a la cantidad de clientes migrados.

En base a la Tabla 21, se aprecia que en el mes de septiembre la utilidad total de la campaña corresponde a \$35.050.000, lo cual se logra con 19.998 migraciones. Además, se observa que a partir del rango de recarga entre \$10.001 y \$12.000, las utilidades para todos los segmentos comienzan a ser negativas, y en su conjunto acumulan una pérdida de \$14.640.000.

En comparación, la utilidad total de la campaña para el mes de diciembre es de \$47.650.000, monto obtenido a partir de 12.985 migraciones. Si bien se observan

utilidades negativas para rangos de recarga sobre \$10.000, estas son menores a las pérdidas obtenidas en el mes de septiembre, lo cual se debe a que existen menos clientes gestionados en esos tramos, y, por ende, menos clientes migrados hacia planes no sugeridos.

A partir de los resultados explicados previamente, para la campaña de septiembre se obtiene una utilidad promedio de \$1.753 por cada cliente migrado, mientras que para la campaña de diciembre la utilidad promedio por cliente es de \$3.670.

	<b>Septiembre</b>		<b>Diciembre</b>	
<b>Rango de recarga</b>	<b>Utilidad generada</b>	<b>Clientes migrados</b>	<b>Utilidad generada</b>	<b>Clientes migrados</b>
< \$4.000	\$17,26 M	3.210	\$33,8 M	5.858
[\$4.001, \$6.000]	\$24,26 M	6.133	\$12,68 M	3.626
[\$6.001, \$8.000]	\$7,12 M	3.389	\$2,4 M	1.662
[\$8.001, \$10.000]	\$1,04 M	2.274	\$0,06 M	1.085
[\$10.001, \$12.000]	\$-1,59 M	1.287	\$-0,89 M	577
[\$12.001, \$15.000]	\$-3,77 M	1.085	\$ -0,41 M	177
[\$15.001, \$20.000]	\$-5,26 M	774	-	-
Más de \$20.000	\$-4,02 M	1.846	-	-
<b>Total utilidad</b>	<b>\$35,05 M</b>	<b>19.998</b>	<b>\$47,65 M</b>	<b>12.985</b>

Tabla 21: Comparación de utilidad para campañas de septiembre y diciembre 2019

## 9. Conclusiones

El objetivo principal del trabajo de tesis consistió en mejorar la campaña de migración mediante la incorporación de técnicas analíticas que permitan definir el grupo de clientes prepago a los cuales ofrecer la opción de contratar un plan. Campaña que, dado el contexto de la industria, resulta de gran relevancia para la compañía, pues de hacerlo de buena manera, permite un incremento sostenible de la base de clientes pospago.

Para cumplir con lo anterior, primero se realizó un análisis exhaustivo de la gestión actual de la campaña. A partir de este, se descubrió que la mayoría de los clientes migran hacia planes cuyo cargo fijo es bajo puesto que no existe un control que fiscalice el cumplimiento de la oferta de los planes sugeridos por la campaña. Esta situación genera que todos los clientes con un rango de recarga alto, al migrar hacia suscripción disminuyan considerablemente el valor que generan para la compañía.

El primer análisis mencionado llevó a realizar una serie de cambios en la gestión de la campaña, lo cual permitió reducir la tasa de incumplimiento de los planes sugeridos de un 13,5% a un 7,5%. Sumado a lo anterior, se logró también un aumento en el ingreso promedio percibido por cada cliente migrado dada la mejora en la gestión, y la incorporación de más clientes con rango de recarga bajo en la campaña. Esto último, duplicó el ingreso promedio de cada cliente, pasando de una utilidad promedio de \$1.753 a \$3.670.

La inclusión del modelo de migración espontánea permitió encontrar un nuevo segmento de clientes propensos a migrar que anteriormente no era considerado, este grupo, además de tener altas tasas de migración, incrementa necesariamente el valor que genera cada cliente para la compañía, debido a que los clientes migran hacia un plan cuyo valor es mayor a su monto de recarga promedio previo.

Por otra parte, la utilización de los modelos para definir la base de clientes a gestionar en la campaña entregó mejores tasas de migración en comparación a la gestión actual de esta, es decir, los modelos desarrollados son capaces de identificar de mejor forma a los clientes con una mayor probabilidad de migrar hacia el mercado de pospago, aumentando así la tasa de efectividad de la campaña en relación con la situación previa a la incorporación de técnicas analíticas. Lo anterior se refleja en que la gestión utilizando los modelos analíticos desarrollados deja fuera a clientes que normalmente se cargan a la campaña (Campaña y no modelo) y poseen una baja efectividad (3,9%), y los reemplaza por un grupo de clientes no considerados previamente (Modelo y no campaña) con una mayor tasa de efectividad (11,4%).

Por lo tanto, tras el trabajo realizado, es posible concluir que la incorporación de técnicas analíticas a la campaña outbound genera valor para la compañía, puesto

que, mediante la aplicación de los modelos, se puede incrementar la cantidad de migraciones gestionando un mismo número de clientes mensualmente.

## 10. Recomendaciones y trabajo futuro

### 10.1. Recomendaciones

Dado los buenos resultados obtenidos a partir del despliegue realizado, se recomienda utilizar los modelos para definir el grupo total de clientes a cargar en la campaña mensualmente. De esta forma, la base enviada al call center corresponderá a los clientes a los cuales el modelo les asigne una mayor probabilidad de migrar hacia suscripción y permanecer en este. De acuerdo con el experimento realizado, esta base posee una mejor tasa de migración en comparación a la gestión actual.

Por otro lado, también es importante evaluar y monitorear el desempeño de la campaña mes a mes. Para ello, se sugieren calcular distintas métricas que fueron utilizadas para evaluar la campaña, las cuales se describen a continuación:

- Tasa de incumplimiento: Porcentaje de clientes que migra hacia un plan no sugerido de menor cargo fijo.
- Tasa de migración por tramo: Porcentaje de clientes migrados, sobre el total de clientes gestionados en el mismo rango de recarga.
- Utilidad según segmento: Se obtiene a partir de la suma de las utilidades generadas por cada cliente migrado perteneciente a un mismo rango de recarga. Esta utilidad se calcula como la diferencia entre el valor del plan al cual migra el cliente, y el promedio de recarga en los tres últimos meses.
- Utilidad total de la campaña: Suma de las utilidades de los distintos rangos de recarga.

En base a lo anterior, se recomienda definir un umbral aceptable respecto a la tasa de incumplimiento con el fin de alertar a los encargados del call center en caso de que esta sea superior al umbral establecido. Si la tasa de incumplimiento comienza a incrementar a valores inaceptables según lo establecido por el negocio, se sugiere dividir la gestión del call center en dos, teniendo por un lado una cantidad de ejecutivos que mantengan la gestión actual pero solo llamen a los clientes con rango de recarga bajo de tal manera que el incumplimiento sea mínimo. Y, por otro lado, contar con una cantidad de ejecutivos que solo se encarguen de contactar a clientes

con rango de recarga alto, pagándoles la migración solamente si el cliente contrata uno de los planes sugeridos por la campaña.

Teniendo en consideración que el modelo se construyó con información de las migraciones hasta el mes de mayo, dado que se analiza la permanencia en los cuatro meses siguientes (en el caso de mayo, hasta septiembre del 2019) y con el objetivo de capturar la información más actualizada de los clientes, se sugiere reentrenar el modelo mensualmente, agregando el mes más reciente disponible, y eliminando a su vez el mes más antiguo. Por ejemplo, siguiendo la metodología descrita, para utilizar el modelo en el mes de febrero del 2020, lo óptimo sería entrenar el modelo con las migraciones entre mayo y julio del 2019, y posteriormente validar los resultados de este con las migraciones asociadas al mes de agosto, lo cual incluye analizar la permanencia de estos clientes hasta el mes de diciembre.

De esta forma, el modelo incorpora y aprende de la información más reciente, lo cual podría incluir nuevas tendencias en el comportamiento de los consumidores o cambios en la gestión del call center, como, por ejemplo, ofrecer solo el plan sugerido.

Dado que con el trabajo realizado se perfila a la campaña una base de clientes con mayor propensión a aceptar la oferta, es posible mejorar la eficiencia del call center, logrando con una misma cantidad de clientes gestionados una mayor cantidad de migraciones. Esta mejora en la eficiencia se mide según la cantidad de ventas por hora por cada ejecutivo del call center. Por lo tanto, si la compañía desea mantener la cantidad de migraciones mensuales, es posible gestionar esta campaña con una menor cantidad de ejecutivos operando, lo cual deja espacio para contar con un mayor potencial de ejecutivos en otras campañas que la compañía quisiera potenciar, como lo son el mercado hogar, u otras campañas relacionadas al mercado de pospago.

Redistribuyendo la dotación de los ejecutivos del call center es posible obtener mayor valor de otras campañas valiosas para la compañía, como lo son mantener su posición líder en el mercado de pospago y aumentar su participación de mercado en internet hogar, con énfasis en el internet de fibra óptica.

## 10.2. Trabajo futuro

El modelo considera una migración exitosa como un cliente que migra y permanece al menos en los cuatro meses siguientes en el mercado de suscripción. Por ende, la propensión del modelo se relaciona con un cliente con una alta probabilidad de

migrar y permanecer en el mercado de pospago en el corto plazo. Para comprobar lo anterior, es necesario evaluar y comparar las tasas de fuga en los próximos meses de los clientes migrados y perfilados por el modelo en el mes de diciembre.

Por lo tanto, como trabajo futuro se debe analizar la tasa de fuga de los clientes en función de la propensión que le asigna el modelo a cada uno. A partir de lo anterior, es posible evaluar si la tasa de fuga es menor para los clientes que prioriza el modelo, y en qué porcentaje se disminuye la fuga natural de los clientes migrados por esta vía.

Otro aspecto importante por analizar consiste en estudiar el comportamiento de los clientes prepago cuyo rango promedio de recarga es superior a \$20.000, usuarios que debido a las nuevas reglas de negocio fueron excluidos de la campaña de migración. Resulta interesante estudiar los clientes con este comportamiento que a partir del mes de octubre no han sido gestionados, con el fin de ver cómo ha sido su comportamiento durante los últimos meses, analizando su fuga de la compañía, tasa de migración espontánea hacia suscripción, y cambios en su nivel de recarga.

Lo anterior, dependiendo de los resultados del análisis, podría llevar a tomar acciones comerciales sobre este grupo de clientes, como, por ejemplo, gestionar una campaña aparte para ellos, u ofrecerles migrar hacia suscripción mediante otra vía.

## 11. Bibliografía

- [1] Subtel | Reporte telecomunicaciones, “Sector Telecomunicaciones Primer Trimestre 2019,” p. 41, 2019.
- [2] Entel, “Memoria Corporativa,” 2018
- [3] T. Subsecretar and T. Marzo, “Sector Telecomunicaciones Cierre 2018 Ministerio de Transportes y Telecomunicaciones Subsecretaría de Telecomunicaciones,” 2019.
- [4] I-cuadrado, “Encuestas portabilidad entel,” 2019.
- [5] “CRISP-DM, still the top methodology for analytics, data mining, or data science projects.” [Online]. Available: <https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html>. [Accessed: 18-Oct-2019].
- [6] A. Azevedo and M. F. Santos, “KDD, semma and CRISP-DM: A parallel overview,” *MCCSIS’08 - IADIS Multi Conf. Comput. Sci. Inf. Syst. Proc. Informatics 2008 Data Min. 2008*, no. June, pp. 182–185, 2008.
- [7] “CRISP-DM: La metodología para poner orden en los proyectos - Sngular.” [Online]. Available: <https://www.sngular.com/es/data-science-crisp-dm-metodologia/>. [Accessed: 18-Oct-2019].
- [8] K. E. Train, “Discrete choice methods with simulation,” *Discret. Choice Methods with Simul.*, vol. 9780521816, pp. 1–334, 2003.
- [9] B. Concepts, D. Trees, and M. Evaluation, “<Classification Decision Trees Etc.Pdf>.”
- [10] “Understanding Random Forest - Towards Data Science.” [Online]. Available: <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>. [Accessed: 18-Oct-2019].
- [11] A. Natekin and A. Knoll, “Gradient boosting machines, a tutorial,” *Front. Neurorobot.*, vol. 7, no. DEC, 2013.
- [12] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*,

vol. 13-17-Augu, pp. 785–794, 2016.

- [13] “Performance and Prediction – H2O 3.26.0.8 documentation.” [Online]. Available: <http://docs.h2o.ai/h2o/latest-stable/h2o-docs/performance-and-prediction.html>. [Accessed: 18-Oct-2019].
- [14] M. Bekkar, H. K. Djemaa, and T. A. Alitouche, “Evaluation Measures for Models Assessment over Imbalanced Data Sets,” *J. Inf. Eng. Appl.*, vol. 3, no. 10, pp. 27–38, 2013.

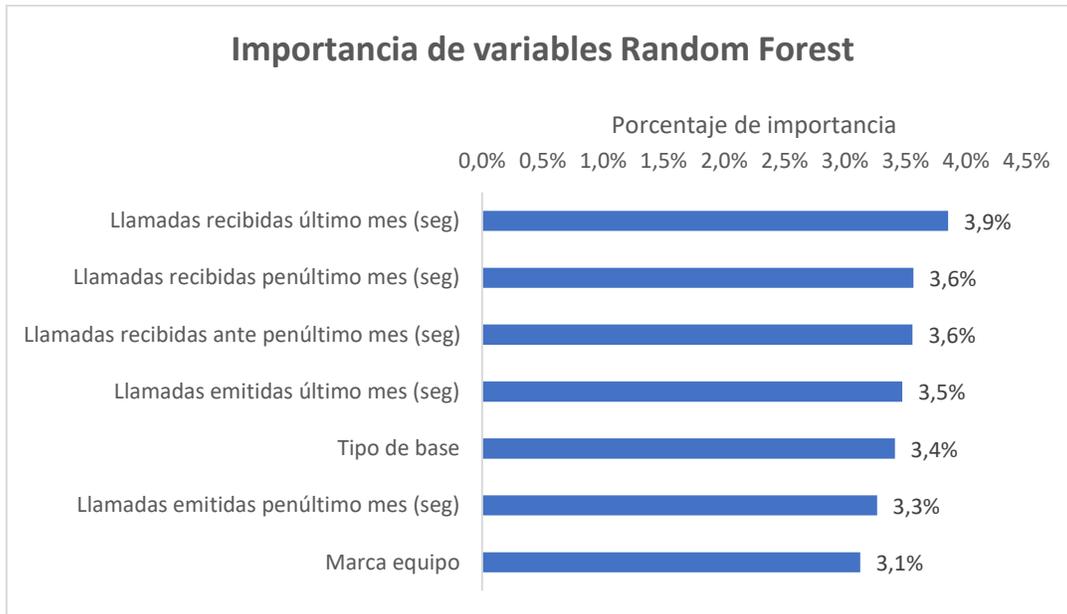
## 12. Anexos

### 12.1 Anexo 1: Detalle de aplicaciones utilizados en el modelo

<b>Aplicaciones utilizadas en el modelo</b>							
1	Facebook	18	Bancobci	35	Chilevisioncl	52	Snapchat
2	Netflix	19	Safertaxi	36	Mercadolibre	53	Amazoncloud
3	Google	20	Youtubekid	37	Easycl	54	Bancoestado
4	Applemaps	21	Onedrive	38	Espn	55	Cabify
5	Directv	22	Tvncl	39	Waze	56	Playstation
6	Easytaxi	23	Yapoel	40	Spotify	57	Dropbox
7	Twitch	24	Homecenter	41	Gmail	58	Megacl
8	Linkedin	25	Estadiocdf	42	Bancochile	59	Luncl
9	Cuevana2tv	26	Speedtest	43	Uber	60	Pariscl
10	Falabellacl	27	Whatsapp	44	Outlook	61	Hbogo
11	Ripleycl	28	Applemusic	45	Yahoomail	62	Abastible
12	Foxplay	29	Itunes	46	Pelispediatv		
13	Wikipedia	30	Bancosantander	47	Emol		
14	Twitter	31	Bancofalabella	48	Groupon		
15	Youtube	32	Icloud	49	Lapolarcl		
16	Googleplay	33	Pokemongo	50	Transantiago		
17	Googlemaps	34	Soundcloud	51	Instagram		

## 12.2 Anexo 2: Importancia de variables modelo de migración vía campaña

### **Random Forest**

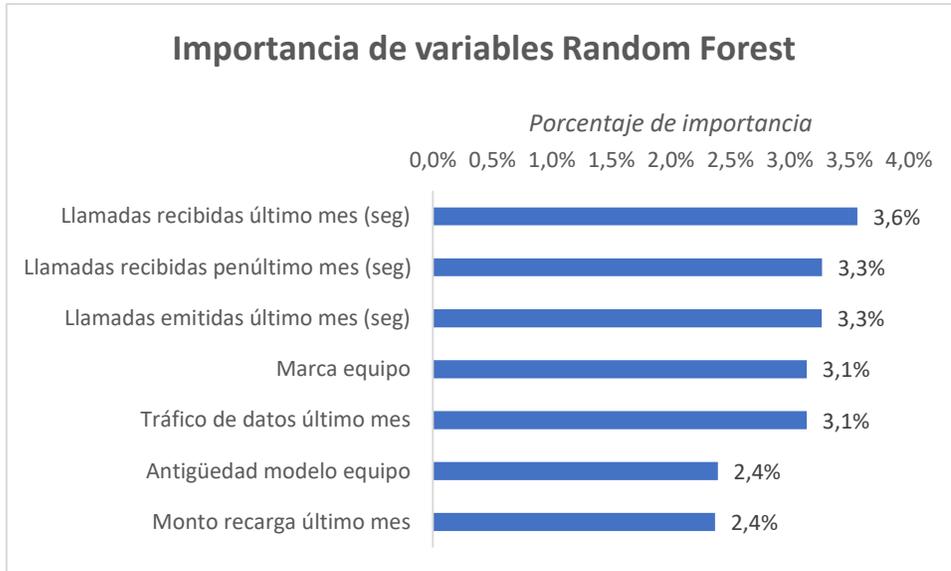


### **GBM**

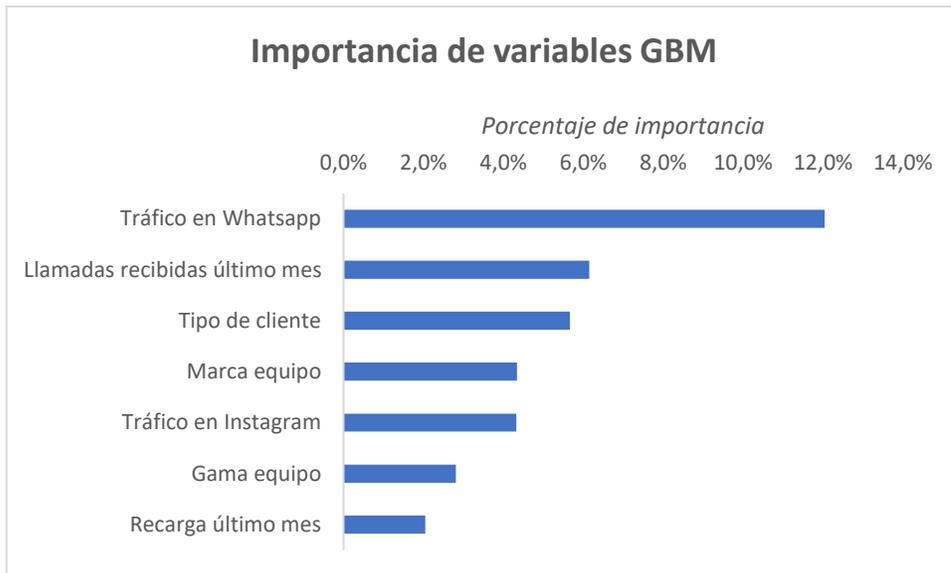


## 12.3 Anexo 3: Importancia de variables modelo de migración espontánea

### **Random Forest**



### **GBM**



12.4 Anexo 4: Test de hipótesis para comparar diferencias de proporciones

<b>Test de Hipótesis</b>		
[\$4.001, \$6.000]	<b>Intersección campaña y modelo</b>	<b>Campaña y no modelo</b>
N ° migrados	3.149	477
N ° clientes	31.036	12.683
Tasa de migración	10,1%	3,8%
p-valor	< 2.2e-16	

<b>Test de Hipótesis</b>		
[\$6.001, \$8.000]	<b>Intersección campaña y modelo</b>	<b>Campaña y no modelo</b>
N ° migrados	1.445	217
N ° clientes	15.902	6.065
Tasa de migración	9,1%	3,6%
p-valor	< 2.2e-16	

<b>Test de Hipótesis</b>		
[\$8.001, \$10.000]	<b>Intersección campaña y modelo</b>	<b>Campaña y no modelo</b>
N ° migrados	962	123
N ° clientes	11.469	3.346
Tasa de migración	8,4%	3,7%
p-valor	< 2.2e-16	

<b>Test de Hipótesis</b>		
[\$10.001, \$12.000]	<b>Intersección campaña y modelo</b>	<b>Campaña y no modelo</b>
N ° migrados	542	35
N ° clientes	6.092	1.359
Tasa de migración	8,9%	2,6%
p-valor	2.489e-15	

<b>Test de Hipótesis</b>		
[\$12.001, \$15.000]	<b>Intersección campaña y modelo</b>	<b>Campaña y no modelo</b>
N ° migrados	161	16
N ° clientes	1.745	420
Tasa de migración	9,2%	3,8%
p-valor	0.0002013	

<b>Test de Hipótesis</b>		
<4000	<b>Intersección campaña y modelo</b>	<b>Campaña y no modelo</b>
N ° migrados	1.744	255
N ° clientes	15.551	5.002
Tasa de migración	11,2%	5,1%
p-valor	< 2.2e-16	

<b>Test de Hipótesis</b>		
<4000	<b>Intersección campaña y modelo</b>	<b>Modelo y no campaña</b>
N ° migrados	1.744	3.857
N ° clientes	15.551	33.737
Tasa de migración	11,2%	11,4%
p-valor	0,4884	

<b>Test de Hipótesis</b>		
<4000	<b>Campaña y no modelo</b>	<b>Modelo y no campaña</b>
N ° migrados	255	3.857
N ° clientes	5.002	33.737
Tasa de migración	5,1%	11,4%
p-valor	< 2.2e-16	