



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

PERFILAMIENTO DE CLIENTES INFLUENCIABLES EN CAMPAÑAS DE
PRODUCTOS FINANCIEROS EN UNA EMPRESA DE RETAIL FINANCIERO

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL

VALENTINA IGNACIA AGUIRRE SAN MARTÍN

PROFESOR GUÍA:
ALEJANDRA PUENTE CHANDÍA

MIEMBROS DE LA COMISIÓN:
CAROLINA SEGOVIA RIQUELME
JOSE ANTONIO NALDA

SANTIAGO DE CHILE
2017

Perfilamiento de clientes influenciables en campañas de productos financieros en una empresa de retail financiero

La eficiencia y efectividad de las campañas en la industria del retail financiero son de suma importancia debido a la cantidad de actores en el rubro. Primero, es fundamental en la realización de campañas la existencia de un grupo de control que permita evitar que las apariencias lleven a conclusiones erróneas. Segundo, ¿todos necesitan recibir la campaña para que tengan el comportamiento deseado?, es probable que la respuesta sea no.

Debido a lo anterior, se propone mejorar la eficiencia de las campañas de súper avances a través del perfilamiento de clientes que efectivamente se pueden persuadir con diferentes estímulos de marketing. Esto último se pretende lograr mediante la implementación de modelos uplift, que tienen como resultado la segmentación de los clientes en 4 grupos: los influenciables, los seguros, las causas perdidas y los *sleeping dogs*, quienes reaccionan de manera contraria al comportamiento deseado. Esta segmentación se conoce como la segmentación fundamental de las campañas y es de suma importancia al momento de hacer una, ya que su implementación mejora las tasas de respuesta incrementales, reduce los costos, reduce los efectos negativos del marketing y mejora la comunicación con los clientes. Es importante mencionar que para la confección de los modelos uplift es necesario la realización de experimentos que permitan obtener la data necesaria con un grupo de control adecuado.

Primero, la confección de los experimentos tiene la finalidad de testear dos cosas: 1) si el sugerir un posible uso del crédito mejora los resultados de una campaña y 2) responder qué tipo de estímulo es mejor para las diferentes categorías de clientes.

De lo anterior se concluye que el sugerir un uso del crédito no afecta significativamente las tasas de respuesta y además, que independiente de que los clientes *elite* (clientes de alto valor para la empresa) respondan significativamente mejor cuando se trata de un regalo de puntos, los clientes en general no siempre mejoran sus tasas de respuesta cuando se les ofrece algún gancho.

Una vez realizados los modelos (uno para las campañas genéricas, otro para puntos y otro para giftcard), se concluye que no es necesario enviar estímulos a todos los clientes, es decir, efectivamente existe un óptimo de clientes a tocar, los que me permitirán obtener los beneficios mencionados anteriormente.

Finalmente, se recomienda a la empresa utilizar siempre un grupo de control adecuado y además, utilizar el modelo confeccionado evitando así el hostigar a los clientes. Es importante considerar que pasado un tiempo prudente, los modelos necesitarán ser calibrados nuevamente debido al cambio de las condiciones de las campañas.

Tabla de Contenido

Antecedentes generales.....	1
1. Características de la empresa y sector industrial.....	1
2. Glosario.....	3
2.1 Productos financieros (PF).....	3
2.2 Características de clientes.....	4
2.3 Campañas.....	4
3. Campañas de productos financieros vía email.....	5
4. Gráfica y componentes de campañas.....	6
Justificación del tema.....	9
1. Antecedentes de las campañas.....	9
2. Área de la empresa donde se desarrollará la memoria.....	11
3. Oportunidad.....	13
Objetivos.....	14
1. Objetivo general.....	14
2. Objetivos específicos.....	14
Marco conceptual.....	15
1. Evaluación de campañas.....	15
1.1 Tasa de apertura.....	15
1.2 Tasa de respuesta.....	15
2. Diseño de experimentos.....	15
2.1 Grupo tratamiento y grupo de control.....	15
3. Contraste de hipótesis.....	16
3.1 Test z de dos proporciones.....	16
4. Tamaño muestral.....	16
4.1 Comparación de dos proporciones.....	17

5. Modelos de predicción.....	17
5.1 Modelamiento uplift	18
Metodología.....	19
1. Definición del proyecto y revisión bibliográfica	20
2. Análisis de la situación actual, selección de datos e información.....	20
2.1 Análisis descriptivo	20
2.2 Modelo predictivo convencional	20
2.3 Diseño e implementación de experimentos.....	21
2.4 Evaluación de campañas y análisis de experimentos	21
3. Confección modelo uplift	21
4. Elaboración de recomendaciones para la empresa	22
Alcances.....	22
Resultados esperados.....	23
Desarrollo metodológico.....	23
1. Análisis descriptivo	23
2. Modelo predictivo convencional	27
3. Diseño de experimentos.....	28
3.1 Experimento usos del crédito	28
3.2 Experimento "ganchos"	30
3.3 Cálculo de los tamaños muestrales.....	32
3.4 Creación de bases.....	32
3.5 Creación de los grupos de control.....	33
4. Evaluación y análisis de experimentos.....	34
4.1 Experimento ganchos.....	34
4.2 Experimento usos del crédito	36
5. Modelo Uplift	37

5.1 Selección de datos	37
5.2 Preprocesamiento y transformación de variables.....	39
5.3 Modelamiento	40
5.4 Interpretación y evaluación de modelos	47
5.5 Análisis de riesgo clientes influenciables.....	60
6. Recomendaciones empresa y trabajo futuro	62
6.1 Recomendaciones.....	62
6.2 Trabajo futuro	63
Conclusiones.....	64
Bibliografía	65
Anexos	67

Índice de ilustraciones

Ilustración 1: Organigrama Empresa. Fuente: Memoria anual Empresa 2015.....	3
Ilustración 2: Envíos de campañas de productos financieros. Fuente: Elaboración propia. (*) Campañas especiales hace referencia a las que se hacen en una fecha particular, por ejemplo: navidad, cyberday, generalmente en estas campañas se ofrece algún gancho.....	5
Ilustración 3: Imágenes utilizadas en las diversas campañas de productos financieros. .	7
Ilustración 4: Ejemplo campaña de bienvenida.	9
Ilustración 5:Proceso de campañas de súper avances. Elaboración propia.	10
Ilustración 6: Segmentación fundamental de las campañas propuesta por el modelamiento uplift. Fuente: Blog de data mining y webminig de Webmining Consultores.	11
Ilustración 7: Espectro de propensión de clientes a optar por un producto financiero. Elaboración propia.	11
Ilustración 8: Organigrama área de CBI. Elaboración propia.	12
Ilustración 9: Proceso de creación de un modelo uplift, enfoque confección de dos modelos. Fuente: Rzepakowski y Szymon Jaroszewicz (2012), uplift modeling in direct marketing.....	18
Ilustración 10: Definición de divergencias a utilizar. P y Q son distribuciones distintas.	19
Ilustración 11: Distintos enfoques para el modelamiento uplift. Elaboración propia.	19
Ilustración 12: A) Ventas totales de SAV en UF. B) Venta promedio SAV por persona en UF. Elaboración propia.....	24
Ilustración 13: Porcentaje de Clientes que sacó un SAV en el periodo y la cantidad promedio sacada por cada uno. Elaboración propia.	24
Ilustración 14: Tasa de transacciones por rango etario. Elaboración propia con datos transaccionales del 2016.....	25
Ilustración 15: Tasa de transacciones por género. Elaboración propia con datos transaccionales del 2016.....	25
Ilustración 16: Tasa de transacciones por GSE. Elaboración propia con datos transaccionales del 2016.....	25

Ilustración 17: Tasa de transacciones por producto (tipo de tarjeta). Elaboración propia con datos transaccionales del 2016.	26
Ilustración 18: Gráficas utilizadas en el experimento de usos del crédito.	29
Ilustración 19: Gráficas utilizadas en el experimento de ganchos.	31
Ilustración 20: Importancia relativa de las variables del modelo genérico. Elaboración propia en R.	41
Ilustración 21: Curva uplift basada en Random Forest con el método de interacciones ordenado por deciles. Elaboración propia.	42
Ilustración 22: Importancia relativa de las variables del modelo de giftcard. Elaboración propia en R.	45
Ilustración 23: Curva uplift basada en Random Forest con el método de interacciones ordenado por deciles. Elaboración propia.	45
Ilustración 24: Importancia relativa de las variables del modelo de puntos. Elaboración propia en R.	46
Ilustración 25: Curva uplift basada en Random Forest con el método de interacciones ordenado por deciles. Elaboración propia.	47
Ilustración 26: Gráfico de validación del modelo genérico con campaña realizada en mayo del 2017. Elaboración propia.	48
Ilustración 27: Gráfico de validación del modelo genérico con campaña realizada en julio del 2017. Elaboración propia.	48
Ilustración 28: Gráfico de validación modelo puntos realizado con datos generados por experimento realizado en mayo del año. Elaboración propia.	49
Ilustración 29: Gráfico de validación modelo giftcard realizado con datos generados por experimento realizado en mayo del 2017. Elaboración propia.	50
Ilustración 30: Gráfico validación modelo puntos con campaña masiva realizada en julio del 2017. Elaboración propia.	51
Ilustración 31: Comparación campaña julio del 2017: sin aplicar modelo uplift y aplicando el modelo. Elaboración propia.	59
Ilustración 32: Evolución puntaje score cartera de clientes con disponible. Elaboración propia.	60
Ilustración 33: Variación del interés por mora y revolving promedio. Elaboración propia.	61

Índice de Tablas

Tabla 1: Porcentaje de venta a través de la tarjeta de la empresa. Fuente: Memoria anual Empresa 2016.	2
Tabla 2: Costos por canal y ejemplo de campaña con 70.000 clientes a contactar. Elaboración propia.	22
Tabla 3: Tabla resumen enero 2017 para campañas de Súper avance. Se muestra el tamaño del grupo target (GM), las tasas de lectura (TL) y las tasas de respuesta (TR). Elaboración propia.	26
Tabla 4: Descripción y tipo de variables utilizadas en el modelo. Elaboración propia.	27
Tabla 5: Condiciones experimentales experimento usos del crédito. Elaboración propia.	30
Tabla 6: Condiciones experimentales experimento ganchos. Elaboración propia.	31
Tabla 7: Valores para el regalo de puntos. CP: costo por punto. Elaboración propia. ...	32
Tabla 8: Indicadores de campañas para correo electrónico, tasa de respuesta y tasa de lectura. Elaboración propia en base a información obtenida de Exact Target y de las bases transaccionales de la empresa.	34
Tabla 9: Tasas de respuesta, tasas de respuesta incrementales y p-valor para cada tratamiento y grupo de control. Elaboración propia.	35
Tabla 10: Tasas de respuesta por categoría de clientes y porcentaje de estas. Elaboración propia.	35
Tabla 11: Indicadores de campañas para correo electrónico, tasa de respuesta y tasa de lectura. Elaboración propia en base a información obtenida de Exact Target y de las bases transaccionales de la empresa.	36
Tabla 12: Tasas de respuesta, tasas de respuesta incrementales y p-valor para cada tratamiento y grupo de control. Elaboración propia.	37
Tabla 13: Descripción variables modelo genérico. Elaboración propia.	41
Tabla 14: Q de Qini para los distintos criterios de divergencia para Random Forest. Elaboración propia en R.	42
Tabla 15: Tabla resumen modelo uplift basado en Random Forest ordenado en deciles. Elaboración propia en R.	42
Tabla 16: Descripción variables modelo giftcard. Elaboración propia.	45

Tabla 17: Tabla perfilamiento con variables de comportamiento y fidelidad. Elaboración propia.	52
Tabla 18: Tabla perfilamiento con variables de situación económica. Elaboración propia.	53
Tabla 19: Tabla perfilamiento con variables de conocimiento del producto. Elaboración propia.	54
Tabla 20: Tabla perfilamiento con variables de comportamiento y fidelidad. Elaboración propia.	55
Tabla 21: Tabla perfilamiento con variables de situación económica. Elaboración propia.	56
Tabla 22: Tabla perfilamiento con variables de conocimiento del producto. Elaboración propia.	57
Tabla 23: Tabla perfilamiento con variables diversas. Elaboración propia.	57
Tabla 24: Tabla resumen perfilamiento. Elaboración propia.	58
Tabla 25: Variación en puntos porcentuales del puntaje score para los distintos segmentos de clientes. Se toma la variación comparando mayo del 2016 con julio del 2017. Elaboración propia.	61
Tabla 26: Variación en puntos porcentuales de la proporción de clientes que pagó intereses por mora para los distintos segmentos de clientes. Se toma la variación comparando mayo del 2016 con julio del 2017. Elaboración propia.	62

Antecedentes generales

1. Características de la empresa y sector industrial

La empresa donde se realiza el trabajo de título es el mayor operador de tarjetas de crédito de casa comercial de Sudamérica, con presencia en los mercados de Chile, Argentina, Perú y Colombia. Además, la empresa es la encargada de la tarjeta de crédito asociada a todas las empresas asociadas a su holding.

La compañía pertenece a la industria de servicios financieros, que abarca un amplio rango de organismos que gestionan los fondos, incluidas las cooperativas de crédito, bancos, compañías de tarjetas de crédito, compañías de seguros, compañías de financiamiento al consumo, *brokers* de bolsa, fondos de inversión y muchas otras.

Los servicios financieros tienen un papel importante para la mayoría de las Tiendas por Departamento chilenas, ya que son una herramienta fundamental para fidelizar al cliente. Esto es lo que llamamos retail financiero, que se refiere a la unión entre el sector de venta minorista o retail, con los servicios financieros comunes de las instituciones bancarias, como los créditos de consumo y tarjetas de crédito.

Existen varios organismos o cuerpos legales que regulan el quehacer de las casas comerciales y el retail financiero. Los organismos reguladores del sector son:

- Superintendencia de Valores y Seguros (SVS),
- Superintendencia de Bancos e Instituciones Financieras (SBIF)
- Dirección del Trabajo
- Servicio Nacional del Consumidor (SERNAC)

Además, cabe mencionar que el retail cuenta con un marco regulatorio que norma su operación y la relación con los clientes. En el anexo 1 podemos ver las principales leyes y normativas que conforman el marco regulatorio en el que se desenvuelven las empresas del retail financiero.

A diciembre del 2016, la empresa contaba con más de 2,5 millones de tarjetas activas y más del 45% de las ventas en las tiendas por departamento fueron realizadas con su propia tarjeta (Ver tabla 1). Además, la compañía está en constante crecimiento llegando a un 13,3% de participación en las ventas de transbank, aumentando sus ingresos y utilidades en un 16,2% y un 7,3% respectivamente¹.

¹ Memoria Anual empresa 2016.

Tabla 1: Porcentaje de venta a través de la tarjeta de la empresa. Fuente: Memoria anual Empresa 2016.

CIFRAS EN MM\$	2016	2015	2014	2013	2012
Tiendas por departamento	45,3%	45,7%	46,6%	46,9%	47,6%
Mejoramiento del hogar	27,5%	26,1%	26,5%	27,1%	27,7%
Supermercados	20,0%	20,1%	19,4%	18,5%	19,7%

Cabe destacar que la empresa cuenta con una asociación con la tarjeta Visa y Mastercard, obteniendo de esta forma dos nuevos productos: la tarjeta Visa y Mastercard de la empresa, dando la oportunidad a sus clientes de utilizar esta tarjeta fuera de la casa comercial, generando aún más lealtad hacia su marca. En el 2015, Cencosud contaba con 1,3 millones de tarjetas de crédito y el 36% de las ventas en sus propias tiendas eran realizadas por este medio de pago. Ripley contaba con 1,1 millones de tarjetas y un 50% de las compras en sus propias tiendas eran utilizando este medio de pago. Mientras Hites, La Polar y ABCDIN tenían 519.000; 453.000 y 708.679 tarjetas respectivamente.

De acuerdo a los principios y valores que inspiran a esta empresa de retail financiero está la misión que se declara como "hacer posible las aspiraciones de las personas, mejorar su calidad de vida y superar sus expectativas a través de una oferta integrada de servicios, potenciada por los beneficios del mundo del holding".

La Visión queda expresada como "ser la compañía preferida por las personas, generando relaciones de largo plazo, a partir de: ser líderes por la transparencia, simplicidad y conveniencia; atraer, desarrollar y motivar un equipo de excelencia, comprometido, colaborativo y apasionado por los clientes y finalmente ser valorados por el aporte a las comunidades donde trabajan".

Con respecto a los valores, estos son:

- Transparencia "para generar relaciones de confianza a largo plazo, sobre la base de la honestidad, el diálogo y el compromiso".
- Conveniencia "orientada al beneficio mutuo, a partir de la cercanía y la proactividad".
- Simplicidad "en la aproximación y promesas hacia nuestros clientes internos y externos, haciendo eficientes y ágiles nuestros procesos, con miras a la competitividad y, por lo tanto, la sostenibilidad de nuestras actividades".

Finalmente el objetivo social de la empresa es la "emisión de tarjetas de crédito y la realización de todas las actividades y operaciones complementarias del giro principal". La administración de la empresa está a cargo de 1.928 empleados y 101 sucursales y está compuesta por la siguiente estructura organizacional:

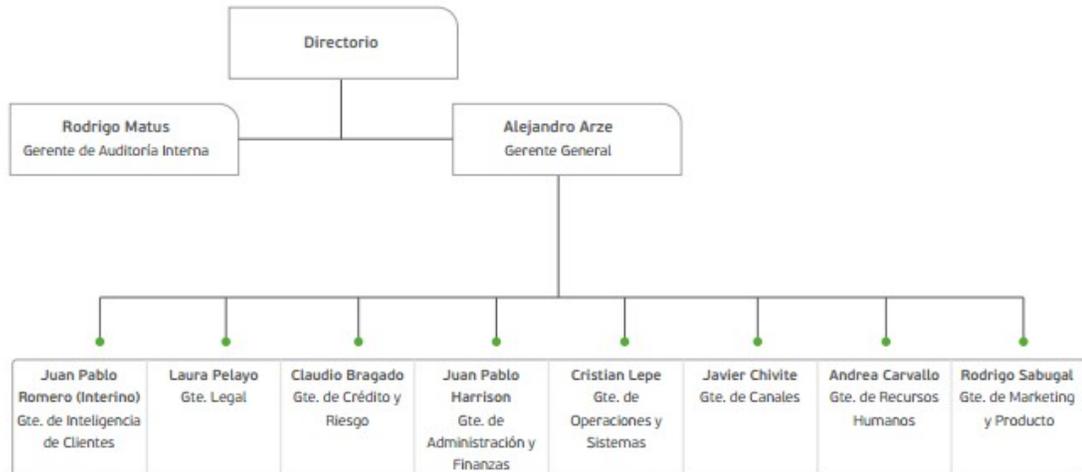


Ilustración 1: Organigrama Empresa. Fuente: Memoria anual Empresa 2015.

La propuesta de valor de la empresa es el otorgamiento de crédito a sus clientes, principalmente a los pertenecientes a los grupos socioeconómicos medios de la población. Esto permite un mejoramiento de su calidad de vida a través de una mayor acceso a nuevos productos y servicios. Con respecto a los productos y servicios ofrecidos, estos son:

- Tarjetas de crédito
- Avances en efectivo
- Créditos de consumo o súper avances
- Cargo automático a tarjeta de crédito (PAT)
- Recarga telefónica
- Programa de Puntos de la Empresa
- Seguros
- Tarjeta BIP Postpago

2. Glosario

A continuación se definen los productos financieros ofrecidos por la empresa y otros conceptos considerados importantes para un entendimiento a cabalidad del proyecto.

2.1 Productos financieros (PF)

2.1.1 Avance en efectivo (Av)

Es dinero en efectivo que se puede sacar de la tarjeta de crédito de la empresa hasta en 24 cuotas. El monto máximo del avance corresponde al cupo de la tarjeta con la que se pide el producto.

2.1.2 Súper avance o crédito de consumo (SAv)

Este producto corresponde a un cupo adicional al cupo de compra con la tarjeta de crédito. Este monto va desde \$100.000 a \$6.000.000 y se puede pedir desde 12 a 48 cuotas.

2.1.3 Tarjetas

Las tarjetas de crédito que administra la empresa corresponden al medio de pago que permite al cliente comprar en todas las tiendas pertenecientes al holding, así como también permite la realización de compras en una amplia red de comercios adheridos a lo largo del país.

- T1: corresponde a la tarjeta que permite comprar 'sólo' en las tiendas pertenecientes al holding.
- T2: tarjeta Visa 'básica' de la empresa. Es importante mencionar que la T8 es una Visa (T2) pero con una nueva tecnología incorporada: contactless.
- T3: tarjeta Mastercard.
- T6: tarjeta Visa destinada a clientes *premium* de la empresa.
- T7: tarjeta Visa destinada a clientes de la empresa.

2.2 Características de clientes

2.2.1 Cliente accionable

Que un cliente sea accionable hace referencia a que dicho cliente sea recurrente, es decir, que haya tenido compras con la tarjeta en los últimos tres meses.

2.2.2 Cliente del blacklist

Los clientes que pertenecen a esta lista negra (o *Blacklist*), no pueden ser tocados, es decir, no se les puede enviar publicidad mediante las campañas de marketing directo. Existen distintos motivos para estar en el *Blacklist*, siendo uno de estos el haber acudido al SERNAC para evitar recibir publicidad.

2.3 Campañas

2.3.1 Gancho

Un gancho se define como una promoción realizada con la finalidad de lograr que un cliente opte por un producto financiero, es decir, acciones que buscan enganchar al cliente para que logre el comportamiento deseado.

Los ganchos que actualmente utiliza la empresa son de tres tipos: descuento en tasas, regalo de puntos y regalo de una giftcard.

3. Campañas de productos financieros vía email

La empresa realiza mensualmente entre 7 y 8 campañas masivas de productos financieros vía email, de las cuales 2 corresponden a campañas de Avances en Efectivo y 5 o 6 a campañas de Súper Avances.

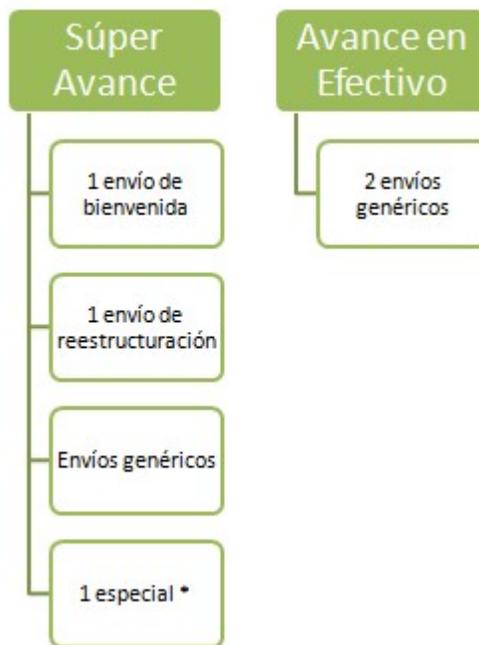


Ilustración 2: Envíos de campañas de productos financieros. Fuente: Elaboración propia. () Campañas especiales hace referencia a las que se hacen en una fecha particular, por ejemplo: navidad, cyberday, generalmente en estas campañas se ofrece algún gancho.*

Reestructuración: son campañas para ofrecer al cliente que tenga más de un Súper Avance vigente, la posibilidad de consolidarlos en una sola cuota. Las tasas de una reestructuración dependen del monto y la cantidad de cuotas. La idea de este servicio es ofrecer al cliente la posibilidad de ordenarse al juntar sus cuotas de Súper Avances, quedando solo con una cuota mensual y teniendo mayor claridad de sus créditos. Además se ofrece la posibilidad de ajustar las cuotas a las necesidades del cliente, es decir, se puede escoger un nuevo plazo de pago entre 12 y 48 meses.

Bienvenida: son campañas destinadas a las personas que nunca han sacado un Súper Avance. Estas campañas se caracterizan por tener un información de qué es un Súper Avance.

Genéricas: son campañas destinadas a todos los clientes que tienen disponible para un crédito de consumo. Cabe destacar que la única diferencia entre las campañas de bienvenida y las genéricas es la información adicional que se agrega con respecto a qué es un súper avance.

Especiales: estas campañas se realiza en los momentos que las personas aumentan considerablemente su gasto, por ejemplo, en el *cyberday*, 18 de septiembre, navidad, etc. El objetivo es que la gente escoja a esta empresa como las que los provea de financiamiento en los momentos que aumentan su gasto.

4. Gráfica y componentes de campañas

A continuación se describen las componentes de las campañas realizadas vía mail con la finalidad de contextualizar el proyecto.

Nombre: siempre aparece el nombre de la persona a la que va dirigido el mail.

Asunto: en los Avances el asunto siempre es %%Nombre%%, pide tu Avance en Efectivo. En el caso de los SAV de bienvenida y genéricos el asunto es %%Nombre%%, tienes disponible un Súper Avance de %%Disponible%%. Finalmente cuando la campaña es de reestructuración, el asunto es %%Nombre%%, ¡consolida tus cuotas de Súper Avance y ordénate!

Mensaje con el disponible de cada persona: para el caso de los SAV va siempre un mensaje estándar (excepto para cuando la campaña es de reestructuración): Tienes aprobado un Súper Avance de %%Disponible%%. Para el caso de los Avances en Efectivo, este mensaje no va.

Mensaje principal según tipo de campaña:

- Avance: ¡Disfruta %%mes%% con tu Avance en Efectivo!
- SAV bienvenida y genérico: ¡Pide tu **Súper Avance** y relájate en %%mes%%!, ¡Despreocúpate en %%mes%% con tu **Súper Avance**!, ¡Sigue disfrutando en %%mes%% con tu **Súper Avance**!
- SAV reestructuración: Si tienes más de un **Súper Avance** vigente, consolídalos en una sola cuota mensual hecha para ti.
- Los mensajes para las campañas especiales dependen de la fecha, por ejemplo si es para navidad el mensaje puede ser: Disfruta esta navidad con tu súper avance.

Imagen: las imágenes son principalmente de mujeres dado los modelos que tiene la empresa, existe solo un modelo que es hombre. A continuación se muestran ejemplos de las imágenes que se utilizan en las campañas de productos financieros:

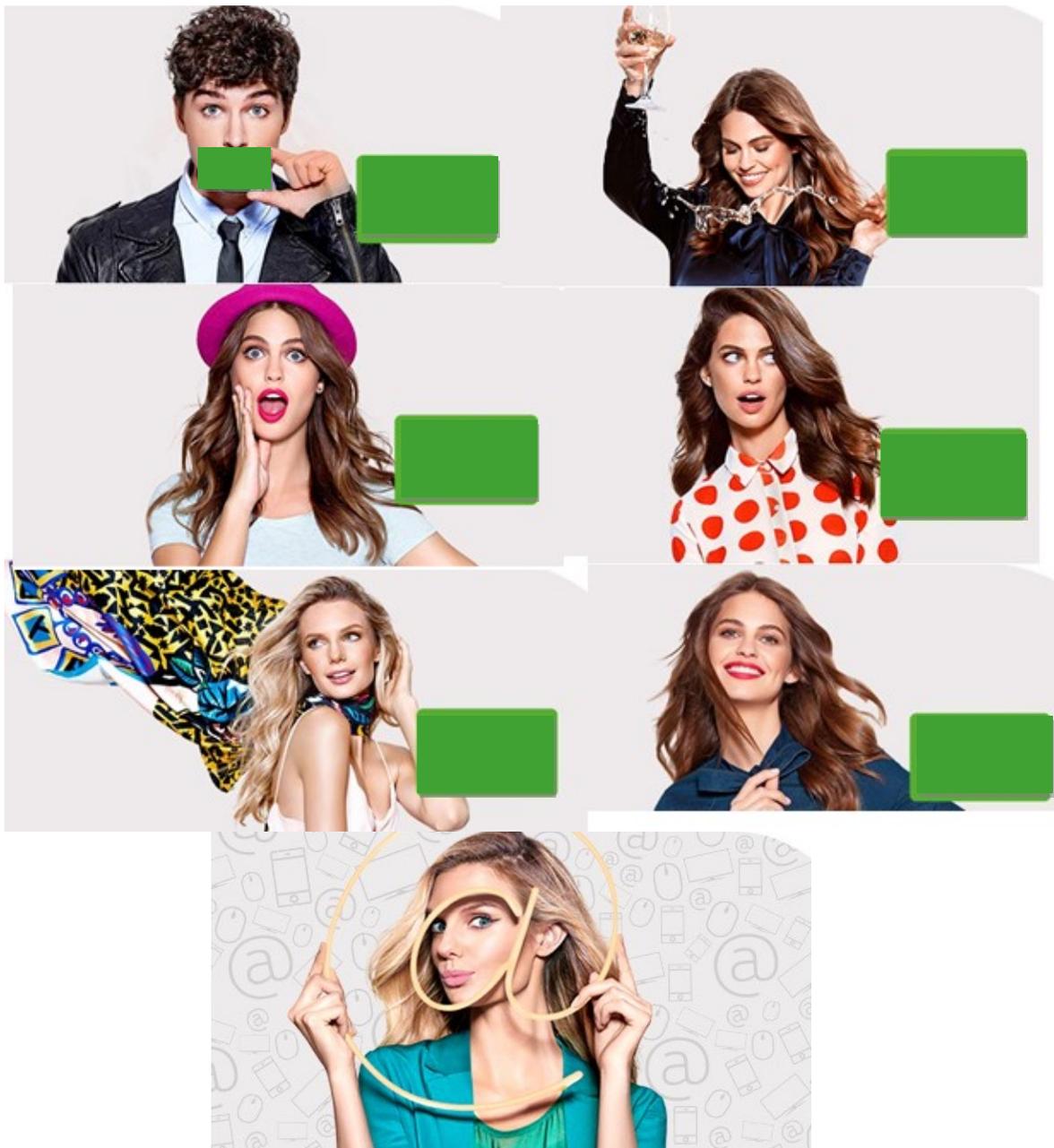


Ilustración 3: Imágenes utilizadas en las diversas campañas de productos financieros.

Información del SAV de un millón: esta información aparece en las campañas genéricas de SAV y las de bienvenida. Además este mensaje aparece independiente de si la persona tiene el disponible suficiente para obtenerlo, ya que en la misma campaña aparece información para poder ampliar ese monto. Cabe destacar que siempre sale la información para un SAV de 1 millón, ya que este es el producto estrella que ofrece la empresa en SAV, es decir, que tiene tasas promocionales.

Para el caso de los avances en efectivo, sale la información del avance de \$100.000, ya que al igual que el caso anterior, este es el producto estrella.

Gancho: Para el caso que sea una campaña especial, en vez de poner la información del SAV de 1 millón, se pone la información del gancho que se va a ofrecer (descuento en tasas, regalo de una giftcard o regalo de puntos). La información que se entrega es a partir de que monto se hace válido el beneficio y a partir de cuantas cuotas.

¿Qué es un SA?: esta información aparece sólo cuando la campaña es de bienvenida, de hecho, una campaña de bienvenida con una genérica difieren sólo en este punto, es decir, la genérica no lo lleva. Para el caso de los Avances, siempre aparece la información de qué es un Avance en Efectivo.

Información de crédito de diferentes montos: esta información se añade, ya que no toda la gente puede acceder al SAV de 1 millón o porque existen personas que requieren de un monto mayor de financiamiento. Para el caso de los Avances, ganchos o reestructuración no se muestra la tabla con información para otros montos.

¿Dónde pedirlo?: siempre va este apartado, ya que es información fundamental para el cliente si es que decide optar por un producto financiero ya sea un Súper Avance o un Avance en Efectivo.

A continuación se muestra a modo de ejemplo una campaña de bienvenida de Súper Avance.

The image shows a digital campaign for a Super Advance (SA) with several key elements highlighted by red boxes and blue arrows pointing to descriptive labels:

- Personalización con el nombre:** A green header bar containing the text "Hola, %%NOMBRE%%".
- Mensaje estándar con el disponible de cada persona:** A white box with a red border containing the text "Tienes aprobado un Súper Avance de \$%%disponible_CC%%⁽¹⁾".
- Mensaje:** A white box with a red border containing the text "¡Sigue disfrutando en noviembre con tu Súper Avance".
- Imagen:** A photograph of a smiling woman with long brown hair, wearing a blue top, positioned next to a green square graphic.
- Información del SA de un millón:** A white box with a red border containing the text: "Aprovecha esta oportunidad", "\$1.000.000", "36 cuotas de \$38.990⁽²⁾", "CAE 24,13%", and "Costo total crédito: \$1.411.640⁽²⁾".
- ¿Qué es un SA?:** A white box with a red border containing the text: "¿Qué es un Súper Avance?", "• Es un crédito con cupo adicional a tu tarjeta de crédito", and "• Puedes pedirlo desde 12 a 48 cuotas".
- Footnote:** A small text box at the bottom with a red border containing the text: "(1) Si necesitas aumentar el monto de tu Súper Avance, acércate a nuestras sucursales".



Ilustración 4: Ejemplo campaña de bienvenida.

Justificación del tema

1. Antecedentes de las campañas

El proceso de creación de campañas de Súper Avances comienza con la definición de los objetivos de la campaña y su segmentación. La segmentación la realiza el área de riesgo y se hace de acuerdo a la propensión de un cliente a optar por un producto financiero, clasificando a los clientes en 13 grupos, siendo el grupo 1 el más propenso a sacar un producto financiero y 13 el menos propenso. Además se realiza una segmentación de acuerdo al disponible que tiene cada cliente y se hace una marca para determinar si un cliente es nuevo o antiguo, refiriéndose como nuevo a un cliente que nunca ha sacado un producto financiero. Existen otras segmentaciones de las bases, como por ejemplo, clientes con buena tasa de lectura, clientes que se estimularán solo por marketing directo (inserto² o carta directa), entre otras. En total, el área de riesgo realiza 16 bases mensuales incluyendo una que deja como grupo de control mensual. Es importante mencionar que cada base es conocida como una pieza y la pieza 0 es la que corresponde al grupo de control.

Una vez definidos los objetivos y segmentada la BD, el área de riesgo define una regla la cual se utiliza para determinar el cupo que le darán a cada persona. Luego, el área de sistemas debe generar la BD teniendo en cuenta la regla definida por riesgo, la cual se envía a la sub-área de gestión de clientes que finalmente debe cargar la BD en Exact Target. Una vez arriba la BD se realiza el envío de correos masivo y, transcurrido el plazo de la campaña, el área de gestión de clientes se encarga de analizar los datos obtenidos. Por su parte, el área de riesgo también realiza la evaluación de las campañas.

En la ilustración 5 se puede ver un diagrama simplificado del proceso de las campañas de Súper Avances.

² Inserto es la publicidad que va junto a los estados de cuenta enviados al hogar.



Ilustración 5: Proceso de campañas de súper avances. Elaboración propia.

Es importante destacar que la evaluación de las campañas realizada por riesgo, es simplemente el cálculo de las tasas de respuesta para cada una de las piezas que genera mensualmente, las cuales son 16 bases de datos (incluyendo al grupo de control). Es de gran relevancia enfatizar que la base del grupo de control es representativa de toda la base de clientes con disponible, no así de las piezas generadas por el área de riesgo, ya que cada una de las piezas cumple ciertas restricciones: clientes nuevos, con un mínimo de disponible, clientes con buena lectura, entre otras, por lo tanto, comparar la tasa de respuesta de la pieza 0 con cualquier otra pieza no es suficiente para decir que las campañas de PF están teniendo efectos positivos.

De acuerdo a la evaluación que realiza el área de gestión de clientes, esta corresponde al cálculo de las tasas de respuesta por campaña, en particular se calcula una tasa de respuesta para los clientes que leyeron el mail, una para los que no leyeron el mail y una tasa de respuesta en general. Una campaña puede tener clientes de más de una de las bases de datos generadas por el área de riesgo. Dado lo anterior, no existen tasas de respuesta incrementales por campaña, por lo que no se puede hacer un análisis sobre el rendimiento de la inversión realizada (ROI) y más aún, es probable que se obtengan conclusiones erradas acerca de la efectividad de campañas en productos financieros.

Para explicar la oportunidad existente se muestra la ilustración 6, que corresponde a la segmentación fundamental de las campañas propuesta por el modelamiento uplift, la cual no trata a los clientes de manera homogénea, si no que se percata que los clientes son heterogéneos en cuanto a las respuestas que tienen a los diferentes estímulos de marketing.

Según esta segmentación, el único segmento que proporciona una verdadera respuesta incremental es el segmento de los influenciables y destinar recursos a los otros segmentos es inútil, ya que no se verán estimulados e incluso en el caso de los *sleeping dogs*, se podría ver un efecto contrario y provocar la fuga de potenciales

clientes o generar disconformidad con la marca. Los influenciables se definen como aquellos clientes que no iban a hacer 'nada' pero dado el estímulo de marketing, cambiarán su comportamiento.

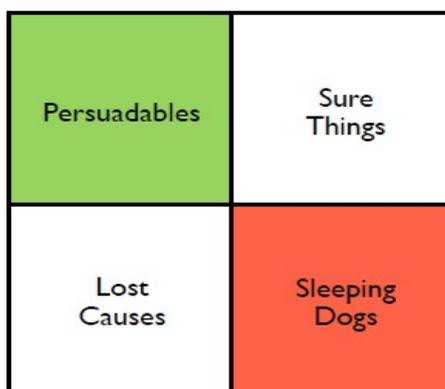


Ilustración 6: Segmentación fundamental de las campañas propuesta por el modelamiento uplift. Fuente: Blog de data mining y webminig de Webmining Consultores.

En la ilustración 7 se observa un espectro de propensión de clientes a optar por un producto financiero. Actualmente, se le envían mails a todas las personas del espectro que posean disponible, es decir, que puedan sacar un crédito de consumo. Por ejemplo, si se envía una campaña a los que tienen propensión entre 7 y 13, se están destinando recursos a personas que no necesitan un estímulo para que adopten el comportamiento deseado, ya que la decisión está tomada, es decir, según la segmentación fundamental de las campañas antes mencionadas, pertenecería al segmento de los *sure things*.



Ilustración 7: Espectro de propensión de clientes a optar por un producto financiero. Elaboración propia.

2. Área de la empresa donde se desarrollará la memoria

El área donde se desarrollará la memoria es el área de inteligencia de clientes (CBI) específicamente en el sub-área de data mining. A continuación se muestra el organigrama de la gerencia correspondiente:

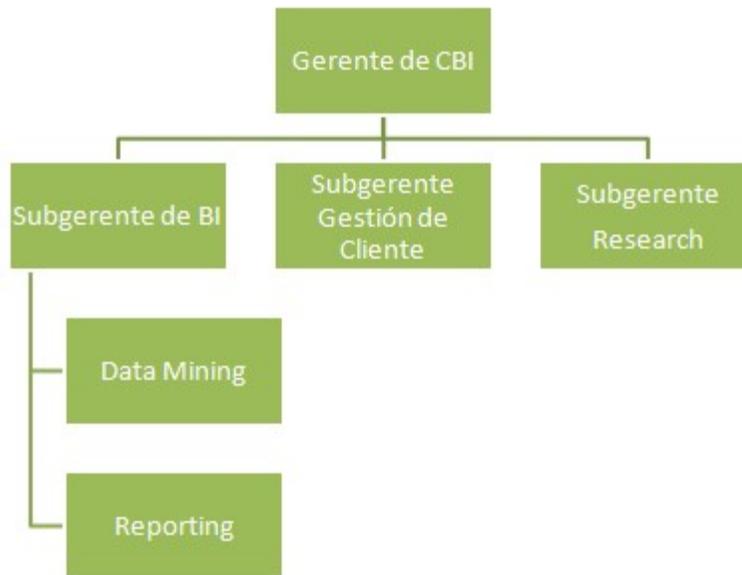


Ilustración 8: Organigrama área de CBI. Elaboración propia.

El área de data mining es bastante pequeña y cuenta con 4 ingenieros civiles industriales (ingenieros de data mining) y 2 informáticos encargados de las bases de datos, automatización de procesos, creación de aplicaciones y creación de reportes. Estas 6 personas se encargan de analizar las bases de datos generando información a partir de los datos transaccionales para luego generar conocimiento valioso para un mejor desempeño de la empresa. Cabe destacar que uno de los principios de área es tener al cliente en el centro de todo análisis y proyecto a realizar.

Las bases de datos de la empresa se alimentan cuando los clientes utilizan la tarjeta para realizar cualquier tipo de transacción. Para el caso de las compras On Us (compras en los comercios del holding), se conocen todos los detalles de la transacción, ya que se puede acceder hasta los datos del producto comprado. Cuando la compra se realiza OnThem (compras fuera del holding), se conoce el nombre del comercio donde se utilizó la tarjeta, el monto transado y la fecha en la que se realizó la transacción. El valor se crea al momento en que se extrae de los datos la información útil y valiosa para el negocio, como, por ejemplo, detectar oportunidades únicas de crecimiento, mejorar la personalización de las campañas, crear nuevos productos, entre otras.

Las otras dos sub-áreas del área de CBI son:

Research: En esta área se encargan de todos los estudios que requiere la gerencia. Para llevar a cabo esto se realizan encuestas, focus-groups, investigaciones de mercado, etc. Actualmente el área está formada por tres personas. Los clientes de esta área son todas las áreas que tengan requerimientos de algún estudio.

Gestión de Clientes: Son los encargados de realizar las campañas publicitarias en conjunto con la el área comercial. Básicamente comercial envía a gestión de clientes un requerimiento de campaña y luego estos se encargan de generar las bases de datos con los clientes que cumplen requisitos y posteriormente, luego de realizada la

campaña, se encargan de recoger la información y realizar un análisis de efectividad de la publicidad. Son 7 las personas que conforman esta sub-área. El cliente de esta área es el área comercial, a los cuales se les entregan las BD y además se le realizan las evaluaciones de las campañas. Otro servicio que presta esta área, es un análisis de clientes, en donde se debe realizar una recomendación de perfiles de clientes para ciertas campaña.

Los "clientes" del área donde se desarrollará la memoria son el área comercial, de canales, de marketing, entre otras, donde el servicio que se ofrece finalmente es la información extraída de las bases de datos que es considerada valiosa para el crecimiento del negocio.

El solicitante del trabajo de título es el gerente de CBI, el cuál ve una oportunidad de mejora en las campañas que se están realizando, específicamente, las campañas de productos financieros. Se cree que se destinan recursos de manera ineficiente, ya que al tocar al 100% de los clientes que pueden optar por un producto financiero y según lo propuesto por el modelamiento uplift, se están destinando recursos a personas que no cambiaran su comportamiento dado las intervenciones de marketing.

Cabe destacar que las áreas que se verán afectadas con este trabajo de memoria serán: CBI, marketing y comercial. CBI, ya que se encuentra el área de gestión de clientes, encargados de las bases de datos y los softwares de administración de las campañas, marketing dado que deben hacer las gráficas de los experimentos necesarios para la realización del proyecto y finalmente comercial que son los encargados de las campañas.

3. Oportunidad

Anteriormente se mencionó que se está llegando al 100% de las personas que tienen disponible un súper avance, y es en este punto donde existe una oportunidad de mejora en la realización y evaluación de campañas. Lo anterior se pretende lograr a través de una **asignación que contemple sólo a clientes influenciables y una evaluación que mida el impacto genuino de las campañas**. Esto permitirá que se destinen los recursos de manera eficiente dejando fuera a los que no cambiaran su comportamiento y a los que reciben de manera negativa un estímulo, los conocidos *sleeping dogs*.

Viéndolo en un ejemplo, cuando la empresa ofrece ganchos, ya sea un descuento en tasas, regalos, etc., los costos de los estímulos se incrementan considerablemente: si la campaña fuese para 40.000 clientes que tienen tasas de respuesta de un 3%, y el regalo es una giftcard de \$100.000, entonces el gasto se eleva a \$120.000.000. Actualmente no podemos saber si esos 120MM destinados a la campaña tuvieron un efecto sobre las ventas de productos financieros, ya que no se posee un grupo de control para medir el incremento y además, es claro que al enviar el estímulo a todos los clientes, se generan ineficiencias en las campañas.

Teóricamente, utilizar un modelo incremental para una campaña genera una asignación más eficiente de estos recursos, ya que se puede simultáneamente mejorar la tasa incremental de respuesta hasta en 3 veces y disminuir los costos de la campaña hasta

en un 40%, además, se reducen los efectos negativos de la campaña dejando de tocar a clientes que no le gusta recibir publicidad [18].

Para lograr determinar cuáles son los clientes que efectivamente se verán influenciados por las campañas de marketing vía email, es necesario tener cierta información de campañas realizadas. La información requerida es tanto del grupo de tratamiento como del grupo del control (el cual es inexistente en nuestro caso), ya que para la correcta construcción del modelo que predecirá a los influenciables es necesario trabajar con ambas muestras.

Como se mencionó anteriormente, para las campañas de productos financieros, el área de riesgo realiza 16 bases de datos que en total tienen ~1,3MM de clientes. Una de ellas se define como grupo de control, que es una muestra aleatoria del total y es de ~100M clientes. Las otras 15 bases se construyen de acuerdo a ciertos criterios, por ejemplo, clientes con disponible mayor a 1,5MM y con buena tasa de lectura. Dado lo anterior, las 15 bases creadas no son comparables al grupo de control y por lo tanto, la información existente no es apta para la construcción del modelo.

Debido a lo anterior se propone la realización de experimentos que permitan generar los datos necesarios para alimentar el modelo, lo cual tiene como principal ventaja el conocimiento que generará para la empresa, sobre todo, para el área comercial, ya que les permitirá mejorar las campañas en términos de los factores modificables de cada una de ellas.

Objetivos

1. Objetivo general

Perfilar a los clientes influenciables de las campañas de productos financieros a través de un experimento de campo y un modelo predictivo.

2. Objetivos específicos

- ❖ Determinar qué variables influyen en la probabilidad de que un cliente opte o no por un producto financiero.
- ❖ Diseñar experimentos con el objetivo de levantar datos para la confección del modelo.
- ❖ Construir un modelo uplift que permita identificar clientes que efectivamente se pueden influenciar.
- ❖ Perfilar a los clientes de acuerdo al segmento al que pertenecen según la segmentación fundamental de las campañas.
- ❖ Elaborar recomendaciones a la empresa a partir del conocimiento generado.

Marco conceptual

1. Evaluación de campañas

1.1 Tasa de apertura

Corresponde al porcentaje de correos abiertos en una campaña. Se calcula como:

$$TA = \frac{\textit{Total de correos abiertos}}{\textit{Total de correos efectivamente enviados}} \quad (1)$$

1.2 Tasa de respuesta

Se define como el porcentaje de destinatarios que han abierto el correo enviado y además completaron la acción deseada, es decir, sacaron un producto financiero.

Para este caso, la tasa de conversión se calcula por las personas pertenecientes al grupo target que realizan la acción deseada en el periodo en que la campaña está vigente, por lo tanto, este es un indicador que posee bastante ruido, ya que una persona puede haber tenido la acción deseada por motivos externos, por ejemplo, un comercial en la televisión.

La forma de calcular el indicador es:

$$TR = \frac{\textit{Total de conversiones}}{\textit{Total de correos efectivamente enviados}} \quad (2)$$

2. Diseño de experimentos

Los diseños de experimentos tienen el objetivo de averiguar si ciertos factores (que son las variables independientes en el modelo) influyen en alguna variable de interés (variable dependiente) y, si existe influencia de algún factor, cuantificar dicha influencia. La idea principal es la manipulación deliberada de las variables independientes para medir el efecto que tiene sobre la dependiente, detectando la causalidad de tal efecto [6].

2.1 Grupo tratamiento y grupo de control

Para diseñar un experimento y medir su impacto se deben confeccionar dos grupos: el grupo tratamiento y el grupo de control, los cuales se generan de manera aleatoria ya que deben ser 'iguales' con respecto a las diferentes proporciones de variables potencialmente relevantes, por ejemplo, deben tener igual proporción de mujeres, de clientes con edad superior a 60 años, entre otras.

El objetivo de la confección de un grupo de control, aquel que no recibe intervención alguna, es permitir comparar al grupo que recibe tratamiento, siendo la diferencia de ambos resultados lo que se le atribuye al efecto de la intervención.

3. Contraste de hipótesis

También denominado test de hipótesis o prueba de significación, se define como un procedimiento para contrastar la hipótesis nula con la hipótesis alternativa con el objetivo de determinar con cierta significancia estadística si se rechaza o no la hipótesis nula. Existen variados test dependiendo de que se pretende testear y de los parámetros que se conocen de la muestra experimental [17].

3.1 Test z de dos proporciones

Este test se utiliza para contrastar las proporciones de dos poblaciones en dos muestras independientes y determinar si estas son diferentes. Para este caso, se contrastará el grupo de tratamiento y grupo de control.

El grupo tratamiento y el grupo de control tienen proporciones P_{GT} y P_{GC} , por lo tanto el contraste anterior es equivalente a formular:

$$H_0 := P_{GT} = P_{GC}$$

$$H_1 := P_{GT} \neq P_{GC} \text{ (o bien } < \text{ ó } > \text{)}$$

Por lo tanto el estadístico de contraste para el test será:

$$z = \frac{P'_{GT} - P'_{GC}}{\sqrt{P'_P * (1 - P'_P) * \left(\frac{1}{n_{GT}} + \frac{1}{n_{GC}}\right)}} \quad (3)$$

donde una buena aproximación de P'_P puede ser la estimación de la porción completa de éxitos de las poblaciones combinadas.

$$P'_P = \frac{n_{GT} * P'_{GT} + n_{GC} * P'_{GC}}{n_{GT} + n_{GC}} \quad (4)$$

Donde:

- n_{GT} y n_{GC} son los tamaños del grupo de tratamiento y de control respectivamente

4. Tamaño muestral

Para el contraste de hipótesis, dónde el objetivo es comparar si las medias/proporciones de las muestras son diferentes, se requiere: [15]

- Magnitud de la diferencia a detectar, es decir, la diferencia entre las medias/proporciones de ambas muestras
- Tener una idea aproximada de los parámetros de la variable que se estudian
- Seguridad del estudio (riesgo de cometer un error tipo I)
- Poder estadístico (riesgo de cometer un error tipo II)
- Definir si la hipótesis va a ser unilateral o bilateral. Una hipótesis bilateral es más conservadora y disminuye el riesgo de cometer un error tipo I.

4.1 Comparación de dos proporciones

La siguiente fórmula sirve para calcular los sujetos necesarios para cada una de las muestras experimentales:

$$n = \frac{[Z_{\alpha} * \sqrt{2p(1-p)} + Z_{\beta} * \sqrt{p_{GT}(1-p_{GT}) + p_{GC}(1-p_{GC})}]^2}{p_{GT} - p_{GC}} \quad (5)$$

Donde

$$p = \frac{p_{GT} + p_{GC}}{2} \quad (6)$$

Cabe destacar que para todo estudio es preciso estimar las posibles pérdidas (por ejemplo, que el correo no esté actualizado o que el correo este mal escrito), por lo que se debe incrementar el tamaño muestral respecto a dichas pérdidas.

El tamaño muestral ajustado a pérdidas se calcula como:

$$n_{ajustado} = n * \frac{1}{1-R} \quad (7)$$

Donde R es la proporción esperada de pérdidas.

5. Modelos de predicción

Los modelos de predicción son representaciones simplificadas de la realidad que ayudan a explicar y predecir patrones que se observan cotidianamente. Una de las grandes utilidades de estos modelos en marketing es que ayudan a caracterizar a los clientes y a estimar sus respuestas cuando son intervenidos con acciones de marketing. Cabe destacar que el modelo que se utiliza en una situación dada depende mucho de los objetivos del que construirá el modelo, los datos con que se cuente y el fenómeno que se desea explicar [3].

5.1 Modelamiento uplift

Están enfocados en la realización de un modelo incremental, es decir, están hechos para medir el impacto directo y genuino de una intervención de marketing. Lo más importante en la construcción de estos modelos es la definición de un 'grupo de control', con la finalidad de hacer un modelo predictivo que se centre en la respuesta gradual, permitiendo medir el efecto incremental que presenta un segmento de clientes como consecuencia de una intervención de marketing [4].

El output de estos modelos es la segmentación de la población en cuatro grupos:

- Los influenciables: aquellos que su probabilidad de compra aumenta si es que son intervenidos en alguna campaña de marketing.
- Los seguros: los que no necesitan un estímulo para que adopten el comportamiento deseado, es decir ya tomaron la decisión.
- Los resistentes: los que no cambiarán su decisión de por ejemplo no sacar un crédito de consumo independiente de la cantidad y contenido de los estímulos.
- Los no molestar: Personas a las cuales las campañas de marketing, tienen un efecto negativo en ellas.

Esta segmentación se realiza con la finalidad de realizar campañas de marketing solo a los clientes a los cuales estas campañas generan un impacto positivo, es decir, a los influenciables.

Cualquier acción de marketing realizada sobre los otros tres grupos representa mayores costos e incluso un impacto negativo de la campaña[13].

Existen dos enfoques para el modelamiento uplift, el primero consiste en construir dos modelos por separado: uno para el grupo de tratamiento y otro para el grupo de control. Luego se restan ambos modelos, lo que se traduce en una estimación de la diferencia de comportamiento causada por la acción (ver ilustración 9). Este enfoque de modelamiento tiene una gran desventaja, ya que se está construyendo un modelo incremental a partir de otros no incrementales [19].

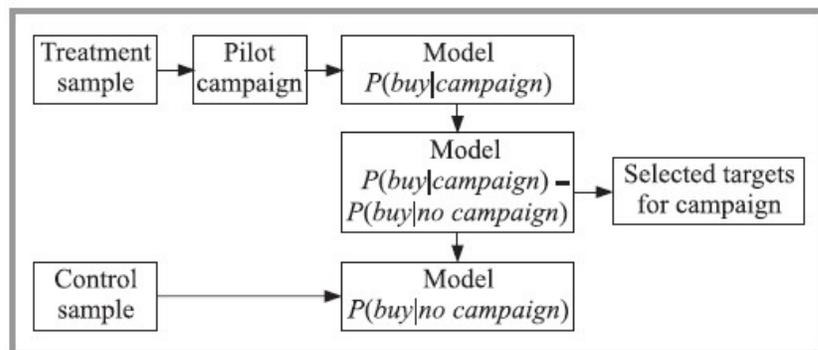


Ilustración 9: Proceso de creación de un modelo uplift, enfoque confección de dos modelos. Fuente: Rzepakowski y Szymon Jaroszewicz (2012), uplift modeling in direct marketing.

El segundo enfoque de modelamiento busca modelar directamente la diferencia entre las probabilidades del grupo target y el grupo de control. Los árboles de decisión han sido adaptados para el modelamiento uplift usando para esto diferentes criterios de separación basados en distribuciones de divergencia. Estas se definen como una medida de cuánto dos distribuciones de probabilidad difieren. El objetivo del modelamiento uplift es maximizar la diferencia entre la distribución del grupo de control y el grupo target [20]. Para este trabajo se utilizarán tres medidas de divergencia: la divergencia de Kullback-Leibler, la distancia euclidiana y la divergencia chi-cuadrado. A continuación, se definen las divergencias:

$$KL(P : Q) = \sum_i p_i \log \frac{p_i}{q_i},$$

$$E(P : Q) = \sum_i (p_i - q_i)^2,$$

$$\chi^2(P : Q) = \sum_i \frac{(p_i - q_i)^2}{q_i}.$$

Ilustración 10: Definición de divergencias a utilizar. P y Q son distribuciones distintas.

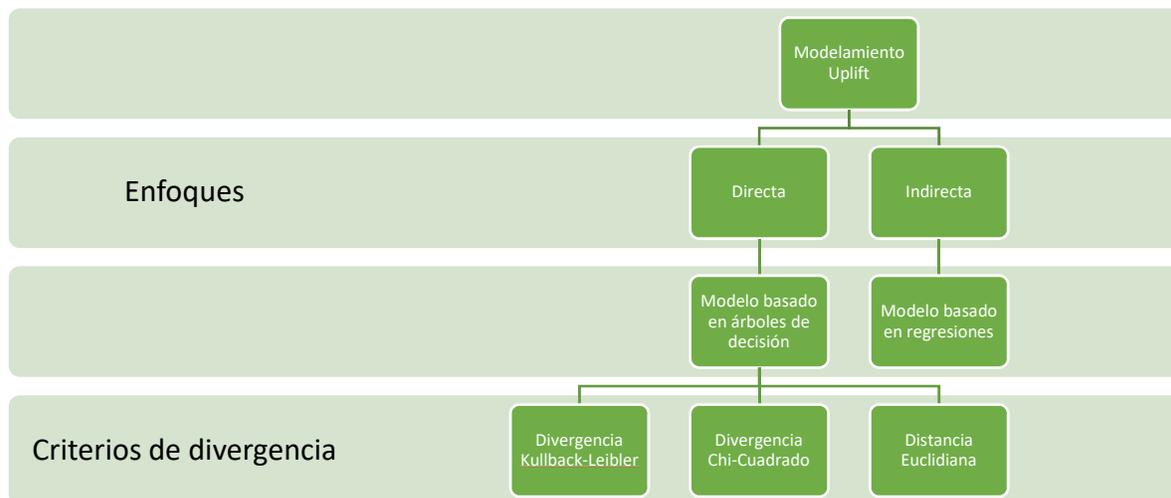


Ilustración 11: Distintos enfoques para el modelamiento uplift. Elaboración propia.

Metodología

La metodología que se utiliza en el presente trabajo de título es la metodología KKD: se parte con la definición del proyecto y la selección de datos, para luego continuar con el preprocesamiento de la data. Una vez procesados los datos se confeccionan los modelos correspondientes, finalizando con los análisis de resultados y evaluaciones. Es importante mencionar que para la selección de datos se llevan a cabo experimentos y para la etapa de la confección de modelos (etapa donde se aplica *data mining* en sí), se utiliza nuevamente la metodología KKD.

Se describe a continuación cada uno de los pasos mencionados anteriormente:

1. Definición del proyecto y revisión bibliográfica

La metodología del trabajo de memoria comienza con la definición del problema/oportunidad para luego seguir con la revisión bibliográfica, la cual permitirá conocer qué estudios y modelos se han utilizado para lograr una mayor eficiencia en las campañas de marketing. En esta etapa se debe lograr:

- Contextualizar la industria y la empresa
- Plantear el problema
- Plantear proyecto
 - Objetivos
 - Alcances
 - Marco conceptual
 - Posibles soluciones
 - Metodología

2. Análisis de la situación actual, selección de datos e información

2.1 Análisis descriptivo

En esta parte de la metodología se busca observar la evolución de los productos financieros viendo las ventas, transacciones, clientes, entre otras características.

- Selección, limpieza y transformación de datos para análisis
- Análisis descriptivo

2.2 Modelo predictivo convencional

Una vez realizado el análisis descriptivo, se realizará un modelo predictivo convencional que permita identificar qué variables son las que influyen en si una persona saca o no un producto financiero.

La finalidad del análisis descriptivo y el modelo logit binario permitirá apoyar la selección de variable y la posible transformación de otras para la realización del modelo uplift.

Para esta parte del desarrollo metodológico se utilizarán variables socio demográficas, de comportamiento y transaccionales, teniendo como resultado:

- Variables que incluyen significativamente en si un cliente saca o no un producto financiero

2.3 Diseño e implementación de experimentos

Para el diseño de los experimentos a realizar, se debe pensar principalmente en los objetivos y qué es lo que se busca con su realización. Una vez hecho esto se definió en conjunto con la empresa (específicamente el área comercial y el área de marketing), los factores con sus respectivos niveles que son factibles de ser estudiados. Luego, se definieron las distintas hipótesis a testear. Una vez visto el tema de factibilidad, se procedió con la realización de las gráficas, la determinación de los grupos muestrales y finalmente la implementación.

- Determinar los factores a estudiar y sus distintos niveles
- Planteamiento de las hipótesis a testear
- Determinar los distintos tratamientos a realizar
- Construcción de las distintas gráficas
- Determinar los diferentes grupos muestrales
- Evaluar en términos de costos la realización del experimento
- Envío de campañas

2.4 Evaluación de campañas y análisis de experimentos

En este punto de la metodología se va a realizar una estimación del efecto de las campañas a través de la comparación del grupo target y el grupo control para cada experimento, luego se procederá a realizar el contraste de hipótesis a través de test de proporciones, el cual permitirá analizar si se rechaza/aceptan las diferentes hipótesis que fueron formuladas en la etapa de diseño e implementación de experimentos.

Se debe lograr:

- Realizar evaluación económica de las campañas
- Cálculo de indicadores de efectividad de las campañas: tasas de apertura y tasas de conversión.
- Testeo de hipótesis y análisis por segmentos de interés

3. Confección modelo uplift

De acuerdo a las lecturas realizadas, el modelo que mejor se ajusta a los requerimientos del proyecto es el modelo uplift, el cual permitirá detectar aquellos clientes que efectivamente son influenciados por las acciones de marketing y además, realizar una evaluación genuina de las campañas.

Para la realización de este modelo se define como variable dependiente la variación de probabilidad de optar por un producto financiero cuando el cliente es tratado y cuando no. Además, se definen como variables independientes:

- Variables significativas según modelo de propensión realizado para determinar que variables influyen en si un cliente saca o no un producto financiero
- Variables relacionadas con los tratamientos de las campañas. Ejemplo: regalo de puntos o regalo de una giftcard.
- Variables creadas para el testeo de hipótesis

Cabe destacar que para la confección del modelo se utilizará el proceso KDD (*knowledge discovery in databases*), el cual se divide en 5 etapas:

1. Selección de datos
2. Procesamiento
3. Transformación
4. Modelamiento
5. Interpretación y evaluación

4. Elaboración de recomendaciones para la empresa

Finalmente se elaboraran recomendaciones de acuerdo al conocimiento generado con los resultados de los experimentos y los resultados de la implementación del modelo uplift.

- Comparación del antes y el después
- Evaluar aplicabilidad del modelo
- Confección de conclusiones y recomendaciones

Alcances

El proyecto presentado en esta memoria abarca el estudio de los diferentes factores que influyen en la efectividad de las campañas de productos financieros, todo esto a través del diseño experimental. Se trabajará con las campañas realizadas vía mail dado la factibilidad de implementación. La tabla 2 muestra los costos de envíos unitarios por canal y el costo de un experimento donde se deban contactar a 70.000 clientes.

Tabla 2: Costos por canal y ejemplo de campaña con 70.000 clientes a contactar. Elaboración propia.

Canal	Costo	Costo campaña (70.000 clientes)	Restricciones canal
Email	\$ 0,25	\$17.500	
SMS	\$ 28	\$1.960.000	
Call Center	\$40	\$2.800.000	Límite de contactabilidad mensual
Inserto	\$ 95	\$ 6.650.000	
Carta Directa	\$ 100	\$ 7.000.000	

Como se observa en la tabla, la realización de un experimento vía mail es 112 veces más barato que vía SMS, 160 veces más que call center, 380 que inserto y finalmente 400 veces más barato que carta directa.

Además, no existe información de las tasas incrementales de respuesta para SMS, inserto y carta directa, implicando que es imposible el cálculo del tamaño muestral para llevar a cabo un experimento. Sumado a lo anterior, para call center existe una restricción mucho mayor, que es que existe un límite de contactos a realizar en el mes. Esto implica que es inviable un experimento por este canal, ya que dado las tasas de respuesta (menores a un 1%), el tamaño muestral sería extremadamente grande.

Cabe destacar que se trabajará con los créditos de consumo, es decir, súper avances, esto debido a que los resultados que se puedan obtener de la realización de la memoria, serían de mayor valor para la empresa. Lo anterior debido a que implican mayores ingresos para la empresa y mejorar la eficiencia en estas campañas, sería de mayor impacto. A diciembre del 2016, las ventas para los avances en efectivo eran de 1.136.375 UF correspondientes a 164.006 transacciones y para súper avances las ventas se elevaban a 1.615.361 UF correspondientes a 47.880 transacciones.

Finalmente, el proyecto contempla la evaluación del modelo en campañas pasadas, pero no contempla la implementación del modelo en futuras campañas, ya que este proceso requiere un tiempo mayor al de los plazos establecidos para la entrega de la memoria.

Resultados esperados

Con la realización de los experimentos y el modelo se busca:

- ❖ Levantar la data necesaria para la confección de los modelos y además determinar los factores que influyen significativamente en las tasas de respuesta de una campaña.
- ❖ Proponer un modelo uplift para determinar qué clientes son influenciados por diferentes tratamientos.
- ❖ Perfilar a los clientes de acuerdo a la segmentación fundamental de las campañas propuesta por el modelamiento uplift.
- ❖ Confeccionar propuesta para la empresa, incluyendo principales resultados, recomendaciones, formas de implementación y próximos pasos.

Desarrollo metodológico

1. Análisis descriptivo

En una primera instancia se comienza con un análisis descriptivo el cual busca responder cómo ha evolucionado el uso de los créditos de consumo y realizar un análisis del uso de SAV para los diferentes segmentos.

A continuación se muestran dos gráficos que muestra la evolución de las ventas de súper avances.

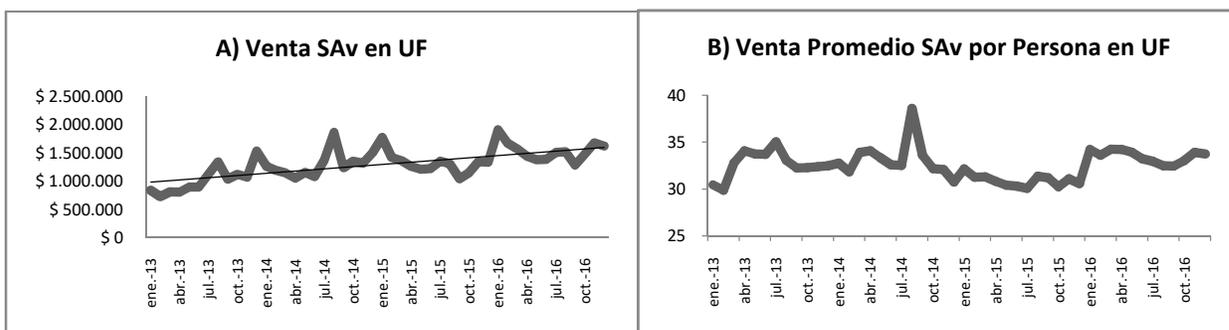


Ilustración 12: A) Ventas totales de SAv en UF. B) Venta promedio SAv por persona en UF. Elaboración propia.

Las ventas de SAv han aumentado en un 94% y se observan peaks en los meses de julio/agosto y diciembre/enero que corresponden a las vacaciones de invierno y verano respectivamente. Con respecto a las ventas promedios por persona, se observa un gran peak en julio del 2014, que se puede atribuir al mundial de Brasil realizado en ese año.

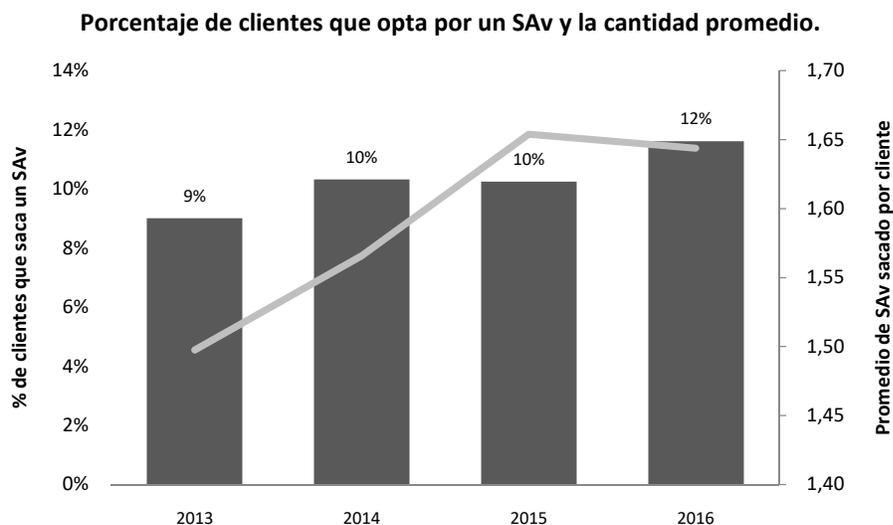


Ilustración 13: Porcentaje de Clientes que sacó un SAv en el periodo y la cantidad promedio sacada por cada uno. Elaboración propia.

En el gráfico anterior, se observa que los clientes que optan por SAv han aumentado en un 3% mientras que la cantidad promedio sacada por cada uno de estos clientes ha aumentado en un 0,15.

Para la construcción de los siguientes gráficos se calculó el ratio $\frac{\text{Total de transacciones del segmento}}{\text{Total de clientes del segmento}}$, el cual indica la cantidad promedio de SAv sacado por cada cliente perteneciente al segmento. La información para calcular el ratio corresponden a datos transaccionales de diciembre del 2016.

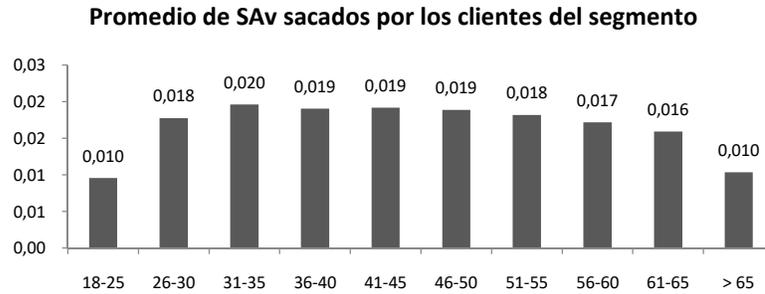


Ilustración 14: Tasa de transacciones por rango etario. Elaboración propia con datos transaccionales del 2016.

En el gráfico se observa que los grupos con menos transacciones promedio por cliente son de extremos y que los con mayor están en el rango de 26 a 55 años. Lo anterior se puede deber a que los clientes entre 26 y 55 años tienen un mayor nivel de gasto que los segmentos de los extremos (menores a 25 años y mayores a 55 años).

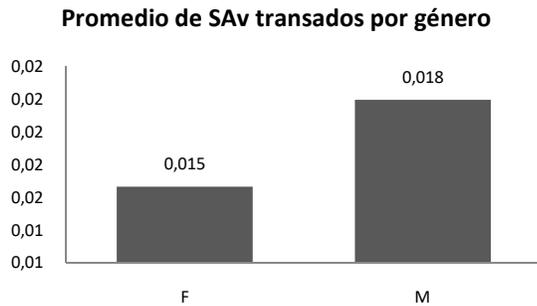


Ilustración 15: Tasa de transacciones por género. Elaboración propia con datos transaccionales del 2016.

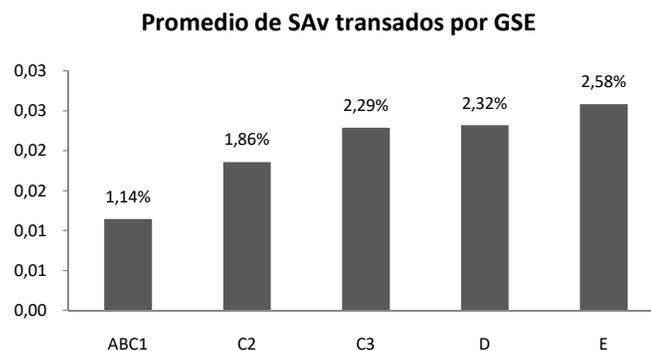


Ilustración 16: Tasa de transacciones por GSE. Elaboración propia con datos transaccionales del 2016.

De la ilustración 15 se observa que son los hombres los que más sacan créditos de consumo siendo que representan un 48% de los clientes. De la ilustración 16 se puede decir que entre más bajo el grupo socio económico, más participativos son al momento

de sacar un crédito de consumo. Cabe destacar que el grupo E representa el 1,57% del total de clientes.

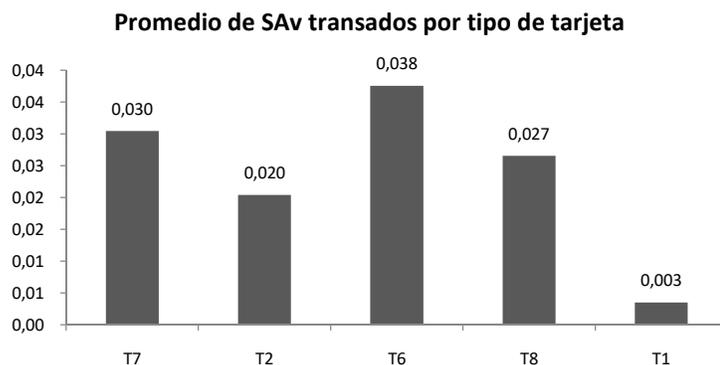


Ilustración 17: Tasa de transacciones por producto (tipo de tarjeta). Elaboración propia con datos transaccionales del 2016.

Del gráfico anterior se observa que los que tienen tarjeta T7 y T6 son los que más sacan productos financieros aunque ellos representan el 1,67% y el 3,39% de los clientes de la empresa respectivamente. Lo anterior es de gran relevancia, ya que los clientes que poseen estos productos son clientes *Premium* (T6) y *Elite* (T7), es decir, son clientes fieles y participativos de la empresa.

Con respecto a las campañas en sí, la tabla 3 a continuación muestra el tamaño promedio del grupo mail y las tasas de lectura y respuesta promedios para el mes de enero del 2017:

Tabla 3: Tabla resumen enero 2017 para campañas de Súper avance. Se muestra el tamaño del grupo target (GM), las tasas de lectura (TL) y las tasas de respuesta (TR). Elaboración propia.

Campaña	GM	TL	TR
Genérica	70.894	26,5%	2,3%
Bienvenida	2.171	36,2%	7,5%
Reestructuración	550	40,5%	9,3%
Gancho	245.653	28,1%	0,4%

Como se puede observar en la tabla N°3, las tasas de respuesta para las campañas de reestructuración, que son las campañas destinadas a los clientes con dos o más SAV's para que puedan consolidar las distintas cuotas, son las que tienen mayores tasas de respuesta.

Opuesto a lo que se podría pensar, las tasas de respuesta de las campañas con ganchos son las más pequeñas. Lo anterior se puede deber a que los montos en estas campañas son montos iguales o superiores a \$1.500.000 y/o que se están enviando los ganchos a clientes *no influenciables*. El monto en créditos de consumo más transado en la empresa es el SAV de 1 millón, por lo tanto, los diferentes ganchos se ofrecen para montos superiores a ese.

Cabe mencionar que las campañas tienen refuerzo (un segundo y hasta tercer envío), los que se consideran como campañas completamente diferentes, por ejemplo, se envía una campaña genérica el 1 de enero del 2017 que tiene validez por todo el mes de enero. El día 10 del mes se envía un refuerzo el cual se considera los clientes del envío realizado el 1 de enero más algunos que para este envío cumplan los requisitos. La manera de evaluar estas campañas es considerando los clientes que responden entre la fecha de envío y el fin de mes, por lo tanto, un cliente que estuvo en ambos envíos y sacó un PF el 17 de enero se estará considerando para la evaluación de ambas campañas.

2. Modelo predictivo convencional

Con la finalidad de complementar el análisis descriptivo se realiza un modelo que busca determinar qué variables influyen en la propensión de un cliente a sacar un producto financiero. El modelo realizado es un logit binario y se confecciona utilizando el software R. Se utilizan variables sociodemográficas, de comportamiento y transaccionales. A continuación, se muestra la tabla 4 con las variables utilizadas en el modelo.

Tabla 4: Descripción y tipo de variables utilizadas en el modelo. Elaboración propia.

Variable	Descripción	Tipo
Producto	tarjeta que posee el cliente	Catagórica
Mes	Mes en el que se saca el super avance	Catagórica
GSE	Grupo socioeconómico del cliente según dirección declarada	Catagórica
# Sav Pagando	Determina el número del SA que está pagando	Entera
Contactable	1 si el cliente es contactable vía email, 0 si no	Catagórica
Rango edad	Rango etario al que pertenece el cliente. <35; 35-55; 55>	Catagórica
Categoría	Si el cliente es elite, normal o premium	Catagórica
Deuda Sav	Deuda que se tiene por créditos de consumo	Catagórica
Monto deuda	Monto que se adeuda con la tarjeta	Entera
Monto cupo Sav	Cupo disponible para Sav	Entera
Monto cupo	Cupo de la tarjeta	Entera
Monto mora	Monto que se adeuda fuera de los plazos establecidos para pagar	Entera
Puntaje score	Indicador generado por área de riesgo. Entre más alto mejor cliente	Entera
# adicionales	Cantidad de adicionales que posee un cliente	Entera
Situación	Puede ser normal, tener mora leve o grave	Catagórica
Antigüedad	Que tan antiguo es el cliente. Se calcula con la fecha de apertura de la cuenta	Catagórica
Pagando Sav	1 si está pagando un súper avance, 0 si no	Catagórica
Ratio Max	Puede ser inicio (que está empezando a pagar un Sav) al medio o final (que está pagando las últimas cuotas de un SAV)	Catagórica

Para la confección se utilizó una muestra de 100.000 registros con un 20% de éxitos, es decir, con 20.000 casos en donde efectivamente se transó un súper avance. Los coeficientes estimados y su significancia se pueden ver en el anexo 2, donde se destaca que las tarjetas T8, T6 y T7 aumentan significativamente la probabilidad de que un cliente opte por un PF, lo que refuerza lo encontrado en el análisis descriptivo. Sumado a lo anterior se refuerza el hecho de que existe estacionalidad en las transacciones de productos financieros, siendo los meses correspondientes a las vacaciones de invierno y verano los que influyen más positivamente en la probabilidad estudiada. Cabe destacar la influencia de la variable pagando SAV, la que afecta negativamente en la probabilidad, sin embargo, la variable Ratio_Max en el nivel "final" afecta positivamente, es decir, que un cliente este pagando un producto financiero

afecta de manera negativa, pero si este cliente está pagando las últimas cuotas de un súper avance, es más probable que saque otro en comparación a si se encuentra pagando las primeras cuotas o las de al medio. Finalmente, los clientes de los grupos socioeconómicos más bajos (E, D, C3) son los que más transan créditos de consumo.

En anexos se puede ver la matriz de correlaciones del modelo.

3. Diseño de experimentos

La elección de los factores con sus distintos niveles es un proceso iterativo en el que se escogen las variables que se quieren analizar y luego se debe evaluar la factibilidad de la implementación del estudio. En una primera instancia se propone estudiar la mayoría de los factores mencionados en el punto de *antecedentes de las campañas*, pero dado la cantidad de tratamientos que esto implica se realiza una segunda propuesta que busca probar dos cosas de interés para la empresa. La primera es realizar una prueba con distintas combinaciones de imágenes con mensajes que representen los posibles usos de un crédito de consumo y la segunda es probar los distintos "ganchos" que tiene la empresa para que la gente opte por un crédito de consumo.

La realización de este proyecto implica la implementación de dos experimentos, los que pretenden probar las siguientes hipótesis:

H_1 := El sugerir un posible uso de un crédito de consumo afecta positivamente las tasas de respuesta

H_2 := Un regalo de puntos tiene mejores tasas de respuesta cuando el cliente es elite

H_3 := Regalar una giftcard tiene mejores tasas de respuesta cuando el cliente es normal o premium

3.1 Experimento usos del crédito

Este experimento busca estudiar si el sugerir un posible uso para un crédito de consumo tiene algún efecto en las tasas de respuesta de una campaña. La idea es probar distintas imágenes y mensajes que representen los distintos usos de un crédito de consumo.

Para la determinación de los distintos tratamientos a testear se analizaron los principales usos que se les da a un súper avance. Esta información se obtuvo de un estudio de súper avance realizado por el área de *research* de la gerencia de inteligencia de clientes en junio del 2016. El estudio consiste en una entrevista telefónica realizada a 331 personas y uno de los objetivos de este, era captar para que se utilizó el crédito de consumo. Las principales razones de uso se enumeran a continuación (ver anexo 4):

1. Pago de deudas
2. Compra de un automóvil
3. Financiar la remodelación de una casa
4. Financiar un viaje
5. Financiar el capital de trabajo de un negocio

La hipótesis que se busca testear con este experimento es:

H_1 := El sugerir un posible uso de un crédito de consumo afecta positivamente las tasas de respuesta

3.1.1 Diseño experimento usos del crédito

Para el testeo de la hipótesis se utilizó el formato de las campañas genéricas, en las que como ya se mencionó, son las campañas en las que se le ofrece al cliente el producto estrella de la empresa a tasas preferenciales: el súper avance de 1 millón. Para testear los diferentes usos que se querían sugerir se crearon 4 campañas donde el tratamiento estaba dado por la combinación de dos factores: la imagen y el mensaje. De acuerdo a lo anterior y dado los intereses de la empresa, se probaron los siguientes usos:

- Realización de un proyecto
- Financiamiento de un viaje
- Financiamiento de un automóvil
- Tratamiento genérico

El tratamiento genérico que se utilizó son las campañas que actualmente utiliza la empresa y además, esta fue la condición experimental que se utilizó como grupo de control.

A continuación se muestran las gráficas que se utilizaron para la realización del experimento:

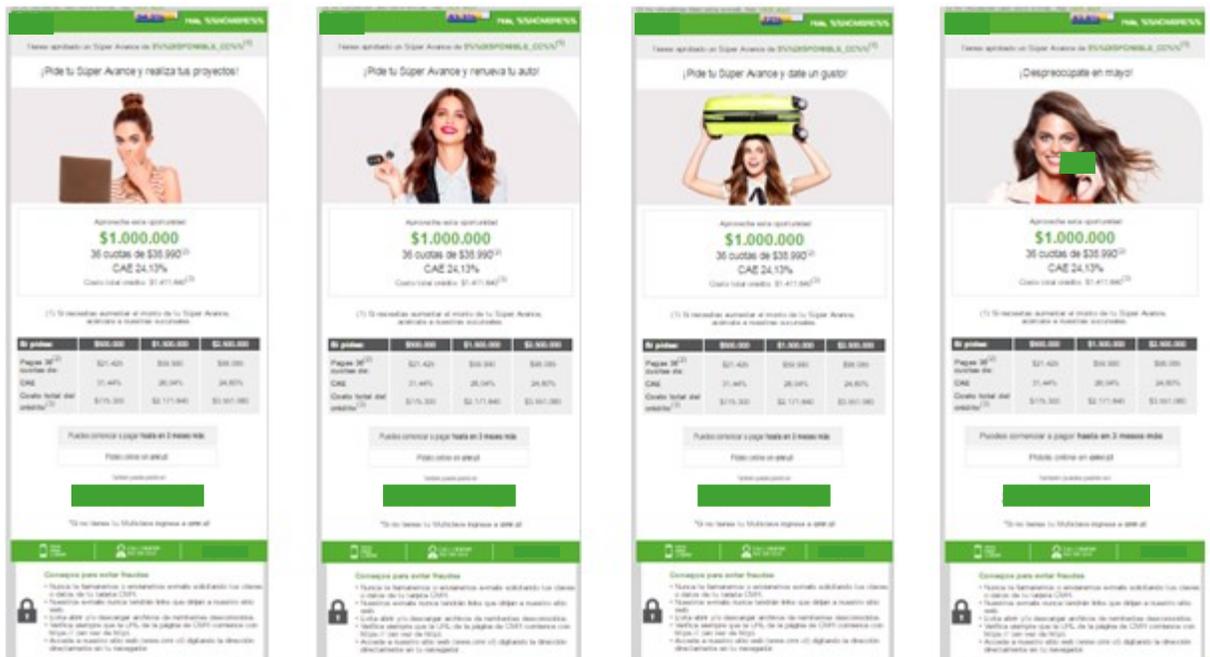


Ilustración 18: Gráficas utilizadas en el experimento de usos del crédito.

Es importante mencionar que una restricción al momento de elaborar las diferentes condiciones experimentales fue la imagen. Lo anterior debido a que las modificaciones que se pueden hacer son sólo a los objetos que tiene la modelo, no al fondo de la imagen, es decir, la imagen no podría haber sido de la modelo en el Caribe. Debido a lo anterior y a que en algunos casos podía no entenderse el uso que se estaba sugiriendo, se varió además el factor del mensaje, ya que de alguna forma se debía reforzar la idea de que se estaba sugiriendo un uso particular del crédito de consumo. Se observa en la ilustración 18 que por ejemplo para la primera imagen donde la modelo aparece con un computador en la mano, era estrictamente necesario un mensaje que reforzara lo que se quería comunicar: para este caso, pide tu súper avance y podrás realizar tus proyectos.

Las condiciones experimentales se resumen en el siguiente cuadro:

Tabla 5: Condiciones experimentales experimento usos del crédito. Elaboración propia.

	Realización de un proyecto	Financiamiento de un auto	Financiamiento de un viaje	Tratamiento genérico
Imagen	Modelo con computador	Modelo con llaves	Modelo con maleta	Modelo con tarjeta de la empresa
Mensaje	¡Pide tu súper avance y realiza tus proyectos!	¡Pide tu súper avance y renueva tu auto!	¡Pide tu súper avance date un gusto!	¡Despreocúpate en mayo!

Este experimento corrió desde el 04 al 31 de mayo.

3.2 Experimento "ganchos"

Como se mencionó anteriormente, se define como gancho a las distintas promociones que tiene la empresa para lograr que un cliente saque un producto financiero. Los ganchos que se buscan probar con el experimento son:

- Regalo de una giftcard
- Regalo de puntos del programa de fidelidad de la empresa

Es importante mencionar que el gancho de descuento en tasas no se puede aplicar en esta oportunidad, ya que la empresa no cuenta con un motor para lograr tasas diferenciadas para los distintos clientes, es decir, si se hace el descuento en tasas para este experimento, todos los clientes obtendrán el descuento y por ende los clientes que reciban la promoción de puntos o giftcard también tendrán el beneficio de tasas preferenciales, teniendo finalmente un doble beneficio.

Las hipótesis que se buscan testear con este experimento son:

H_2 := Un regalo de puntos tiene mejores tasas de respuesta cuando el cliente es elite

H_3 := Regalar una giftcard tiene mejores tasas de respuesta cuando el cliente es normal o *premium*

De esta forma se busca saber cuál de los dos ganchos más utilizados por la empresa en cuanto a productos financieros tiene un mayor efecto sobre las tasas de respuesta de las campañas para las distintas categorías de clientes (normal, *premium*, *elite*).

3.2.1 Diseño experimento "ganchos"

Para testear las dos hipótesis antes mencionadas se crean dos tratamientos, uno por cada gancho, en donde cada uno será el control del otro.

A continuación, se muestran las gráficas utilizadas para el experimento:

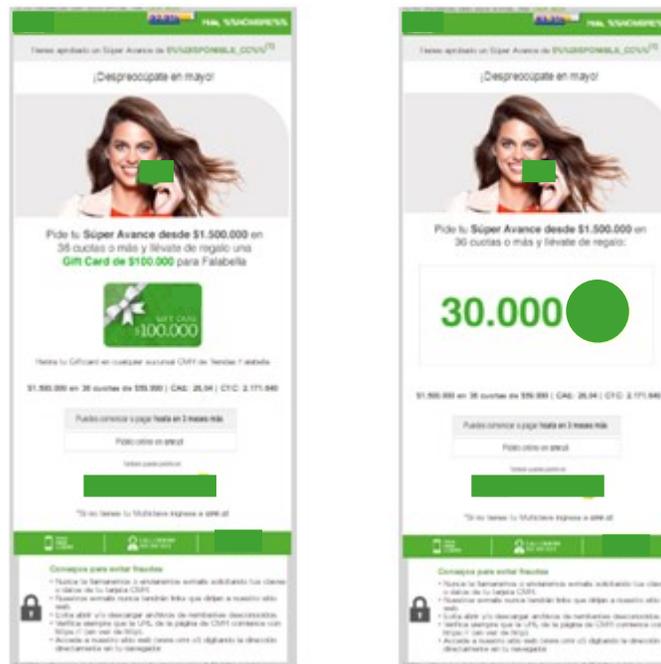


Ilustración 19: Gráficas utilizadas en el experimento de ganchos.

Para la realización de estas dos gráficas se tuvieron ciertas consideraciones, ya que la finalidad del experimento es obtener el efecto puro de los diferentes tratamientos. Dado lo anterior, la única variación en las gráficas es el regalo ofrecido, buscando además que el valor monetario de ambos tratamientos a probar sea equivalente. A continuación, se muestran las condiciones experimentales del experimento:

Tabla 6: Condiciones experimentales experimento ganchos. Elaboración propia.

	Regalo de puntos	Regalo de giftcard
Gancho	Regalo de 30.000 puntos	Regalo de una giftcard de \$100.000

Existe una restricción con respecto a igualar los valores monetarios de ambos regalos: el valor de un punto del programa de fidelidad no es el mismo para los diferentes clientes, ya que depende de la categoría del cliente y de la cantidad de puntos que este tiene acumulado. La tabla 7 muestra el valor mínimo y máximo por un regalo de 30.000 puntos.

Tabla 7: Valores para el regalo de puntos. CP: costo por punto. Elaboración propia.

	Min	Max	Prom
CP	2,91	6,3	4,2
Total	\$ 87.300	\$ 189.000	\$ 126.000

El experimento estuvo vigente desde el 04 al 12 de mayo.

Es importante mencionar que el legal de la campaña decía que se entregarían hasta 30 premios o hasta que se acabe la fecha de la promoción.

3.3 Cálculo de los tamaños muestrales

Lo principal para el cálculo del tamaño muestral es determinar las respuestas esperadas tanto del grupo control como el grupo experimental. Estos valores son de suma importancia dado que entre más pequeña sea la diferencia, más grande es el número de individuos requeridos para el experimento.

Para estimar los valores recién mencionados se analizó la información histórica de campañas, en particular una genérica para el cálculo del tamaño muestral del experimento *usos del crédito*, y dos campañas con descuento en tasas para estimar el tamaño muestral del experimento *ganchos*. Las campañas que se utilizaron fueron campañas realizadas en enero del 2017.

Una vez determinado lo anterior, se define la confianza y el poder de la muestra los cuales son del 95% y 80% respectivamente. Finalmente, con un ajuste del 10% correspondiente a los registros que se caen al momento de subir la base a Exact Target, más los emails que rebotan al momento del envío masivo de emails, se obtienen los siguientes resultados:

- Para el experimento de ganchos, el mínimo de individuos requeridos por muestra es 13.387. Para tener un margen se utilizarán 17.000 individuos por grupo muestral, es decir, un total de 34.000 clientes.
- Para el experimento de imagen, se requiere como mínimo 31.318 individuos, dando un total de 125.272 clientes a utilizar.

3.4 Creación de bases

Actualmente, el área de riesgo es la encargada de realizar las distintas bases para las campañas de créditos de consumo. Se generan 15 bases mensuales las cuales se segmentan de acuerdo a distintas variables. La base que se utiliza para la realización de los experimentos es la base genérica con disponible correspondiente a la pieza 11.

Esta base tiene aproximadamente 380.000 clientes y corresponden a clientes con buena lectura del estado de cuenta y/o comercial, es decir, tienen al menos una apertura correspondiente al 70% de los mails enviados. Es importante considerar que esta marca de buena lectura se realizó hace aproximadamente 2 años, por lo tanto, puede ser una característica que el cliente ya no posea. Estos clientes con buena lectura corresponden a aproximadamente el 80% de la pieza 11, el otro 20% se completa con clientes que sean contactables vía email.

El primer paso para la realización de las bases fue aplicar los filtros que utiliza el área de gestión de clientes, es decir, para el envío de email el cliente debe:

- Ser accionable: recurrente
- No estar en el *blacklist*
- No pertenecer al grupo de control anual, el cual determina la empresa que son clientes que no se deben contactar de ninguna forma
- Ser cliente contactable, es decir, que tenga correo electrónico
- No tener mora
- No ser empleado de la empresa
- No ser parte del directorio ni familiar de estos

Estos últimos dos filtros se realizan por el hecho de ser experimento. Las campañas oficiales no reciben estos filtros.

Una vez realizados los filtros de gestión de clientes quedan 255.583 clientes. Luego se procede con la generación de dos bases de 17.000 individuos de manera aleatoria. Una vez realizado esto se verifica la igualdad de ambos grupos muestrales con respecto a distintas variables:

- Categoría del cliente (normal, *premium o elite*).
- Contrato de súper avance en el mismo periodo del año anterior.
- Clientes nuevos: clientes que nunca han sacado un crédito de consumo.
- Si se está o no pagando actualmente un súper avance o crédito de consumo.
- Propensión a sacar un SAV.
- Disponible que posee el cliente para un súper avance.
- Clientes con buena lectura.

Para la generación de las bases para el experimento usos del crédito se procede a tomar los clientes que quedaron y dividirlos en cuatro, esto ya que como se mencionó anteriormente, se debe tocar al 100% de los clientes. Finalmente se forman tres bases de 55.396 clientes y una de 55.395, quedando igualadas en las variables antes mencionadas.

3.5 Creación de los grupos de control

La creación de un grupo de control para cada experimento es de suma importancia ya que este es requerido para la elaboración de los modelos. Para esto se ocupará el grupo de control anual de la empresa y la pieza 0 (base generada mensualmente por el

área de riesgo), que corresponde al grupo de control mensual para productos financieros.

Al igual que para la creación de las bases para los experimentos, se aplican todos los filtros que realiza gestión de clientes (menos el grupo de control anual) y luego se estratifica la base de acuerdo a las variables de importancia mencionadas anteriormente para igualarla a las bases de envíos en cada uno de los dos experimentos.

4. Evaluación y análisis de experimentos

Una vez transcurrido el tiempo de experimentación se continúa con la evaluación de las campañas, es decir, se obtienen los indicadores de importancia para la empresa: tasas de respuesta 'TR' y de lectura 'TL' (esta última indica la apertura del correo), tanto generales como para los distintos segmentos de interés dentro de las bases. La comparación estadística se realiza utilizando el software R.

Cabe destacar que para los test de contraste de proporciones se utiliza un nivel de significancia del 95%.

4.1 Experimento ganchos

Los primeros indicadores a analizar cuando una campaña se realiza vía correo electrónico son las tasas de lectura o apertura del correo y las tasas de respuesta. En la tabla 8 se pueden apreciar estos indicadores para ambas campañas y además para el grupo de control, el cual no fue estimulado de ninguna manera.

Tabla 8: Indicadores de campañas para correo electrónico, tasa de respuesta y tasa de lectura. Elaboración propia en base a información obtenida de Exact Target y de las bases transaccionales de la empresa.

	Recibidos	Aperturas	TL	Respuestas	Resp/Aper	TR	TR/Aperturas
Puntos	16.079	6.347	39,47%	23	20	0,14%	0,32%
GiftCard	15.960	5.901	36,97%	11	9	0,07%	0,15%
Grupo Control	4.083	-	0,00%	4	-	0,10%	0,00%

A pesar de las bajas tasas de respuesta podemos decir que las tasas de respuesta para ambos tratamientos son significativamente diferentes (p -valor=0,0416), sin embargo, las tasas de apertura (TL) presentan diferencias significativas con un p -valor de $4,134e-6$. Debido a esto, para realizar el testeo de las hipótesis asociadas a este experimento, se hará la evaluación con respecto a email aperturados, básicamente porque no se sabe si puntos es mejor que giftcard por el hecho de que más clientes abrieron el correo o porque efectivamente regalar puntos sea mejor que regalar una giftcard. Es importante notar que el grupo de control no presenta el indicador de tasa de lectura ni la cantidad de respuestas de clientes que aperturaron el correo, ya que por ser grupo de control, no recibieron estímulo alguno.

Evaluando ahora con respecto a los emails aperturados se tiene que:

- No se puede decir que las tasas de respuesta de ambos tratamientos son estadísticamente diferentes (test de dos colas con un p-valor=0,06432).
- Regalar puntos tienen mejores tasas de respuesta que regalar una giftcard. La existencia de diferencia significativa de las proporciones implica un cambio de comportamiento en la variable de transar un producto financiero para cuando el tratamiento implica el regalo de puntos del programa de fidelidad de la empresa versus si el regalo implica una giftcard. El p-valor de este test de proporciones es de 0,03216.

Tabla 9: Tasas de respuesta, tasas de respuesta incrementales y p-valor para cada tratamiento y grupo de control. Elaboración propia.

Campaña	TR	TR Incremental	P-valor
Puntos	0,14%	0,04%	0,244
Giftcard	0,07%	-0,03%	0,270
Grupo Control	0,10%	-	

Ahora si comparamos ambos tratamientos con el grupo de control, el cual no recibió estímulo alguno, se concluye que en ningún caso existen diferencias significativas con el grupo de control. Ver tabla 9.

Para poder testear las hipótesis es necesario obtener las tasas de respuesta para las diferentes categorías de clientes. El cuadro a continuación muestra los distintos valores para las tasas de respuesta.

Tabla 10: Tasas de respuesta por categoría de clientes y porcentaje de estas. Elaboración propia.

	Categoría					
	Puntos			Giftcard		
	Respuestas	Envíos	TR	Respuestas	Envíos	TR
Normal	7	7810	0,090%	1	7673	0,013%
Premium	6	4983	0,120%	7	4983	0,140%
Elite	10	3286	0,304%	3	3304	0,091%
Total	23	16079	0,143%	11	15960	0,069%

Para testear la hipótesis de que un regalo de puntos tiene mejores tasas de respuesta cuando el cliente es *elite*, se realiza el test de proporciones para lo que se obtiene un p-valor de 0,0254. Lo anterior implica que se rechaza la hipótesis nula de igualdad de proporciones, es decir, existe evidencias significativas de que la tasa de respuesta para el tratamiento de puntos es mejor que para el tratamiento de regalo de una giftcard para cuando el cliente es *elite*.

La segunda hipótesis a testear fue que regalar una giftcard tiene mejores tasas de respuesta cuando el cliente es *premium* o normal, la que se rechaza inmediatamente dado que las tasas de respuesta son mayores para el caso de un regalo de puntos. Al

hacer el test para ver lo contrario, es decir que los clientes normales o *premium* responden mejor cuando el regalo consiste en puntos del programa de fidelidad de la empresa, se tiene que se rechaza la hipótesis nula con un p-valor de 0,286.

4.1.1 Otros análisis

Analizando al detalle las respuestas en los distintos segmentos (23 para puntos y 11 para giftcard) se tiene que existe evidencia significativa para decir que:

- Para el caso de la campaña de puntos, un 26,1% de las transacciones fueron realizadas por clientes *premium* mientras que para la campaña de giftcard fueron un 63,6%. Al realizar un test de proporciones sobre estos resultados se obtiene que la proporción de clientes *premium* que responden para el tratamiento de regalo de una giftcard es significativamente mayor que para cuando el regalo es de puntos.

4.2 Experimento usos del crédito

A continuación se muestra la tabla 11 con los resultados de la campaña:

Tabla 11: Indicadores de campañas para correo electrónico, tasa de respuesta y tasa de lectura. Elaboración propia en base a información obtenida de Exact Target y de las bases transaccionales de la empresa.

	Recibidos	Aperturas	TL	Respuestas	Resp/Aper	TR	TR/Aper
Renueva tu auto	44.243	14.575	32,94%	224	113	0,51%	0,78%
Viaje	44.279	14.935	33,73%	227	107	0,51%	0,72%
Realiza tus proyectos	44.427	14.773	33,25%	235	117	0,53%	0,79%
Genérica	44.537	15.038	33,77%	266	136	0,60%	0,90%
Grupo Control	46.336	-	-	211	0	0,46%	-

Al igual que para el caso anterior, existen diferencias significativas entre algunas de las tasas de lectura (al comparar el tratamiento genérico con el tratamiento de renueva tu auto) por lo tanto, para el testeado de la hipótesis se hará el análisis con respecto a emails aperturados.

Para el testeado de la hipótesis se realizaron test de una cola comparando cada tratamiento que sugiriera un posible uso del crédito (financiamiento de un auto, financiamiento de un viaje y realización de un proyecto) con el tratamiento genérico. Al hacer la comparación, se concluye que el sugerir un posible uso del crédito de consumo no mejora las tasas de respuesta de las campañas. Es importante destacar que esta afirmación no es estricta debido a que los cambios realizados en la imagen fueron pequeños, por ejemplo, el poner unas llaves en la mano de la modelo. Ver ilustración 18.

Finalmente al comparar cada tratamiento con el grupo de control, el cual no recibió estímulo alguno, se tiene:

Tabla 12: Tasas de respuesta, tasas de respuesta incrementales y p-valor para cada tratamiento y grupo de control. Elaboración propia.

Campaña	TR	TR Incremental	P-valor
Genérica	0,60%	0,14%	0,002
Proyecto	0,53%	0,07%	0,067
Auto	0,51%	0,05%	0,152
Viaje	0,51%	0,05%	0,123
Grupo Control	0,46%	-	

El único tratamiento que genera mejoras en las tasas de respuesta al comparación con el grupo de control es el tratamiento genérico con un p-valor de 0,002.

5. Modelo Uplift

Se confeccionarán tres modelos uplift, uno para determinar los clientes influenciables de las campañas genéricas, donde se ofrece el producto estrella definido por la empresa como el súper avance de \$1.000.000 a tasas preferenciales, otro para detectar los clientes influenciables de las campañas de puntos y, finalmente, uno para las campañas de giftcard.

De acuerdo al marco conceptual, se realizarán los modelos según el enfoque directo, es decir, basado en árboles de decisión. Lo anterior debido a que estos tienen un mejor rendimiento que los basados en regresiones (enfoque indirecto). Es importante mencionar que se realizarán utilizando el software R, específicamente utilizando *random forest* debido a su eficiencia para trabajar grandes bases de datos. R ofrece un paquete para construir y testear modelos uplift, el cual permite confeccionar un modelo utilizando distintos criterios de divergencia. Por lo tanto, se buscará encontrar el modelo que presente un mejor rendimiento basado en su criterio de separación, los cuales son: método de interacciones, *kullback-Leibler*, distancia euclidiana y divergencia *chi-cuadrado*.

5.1 Selección de datos

Esta etapa de la metodología KDD es de suma importancia en el sentido que se determina en cierto modo si las etapas sucesivas van a ser capaces de extraer conocimiento válido y útil para la empresa. Para este caso los datos a utilizar vendrán directamente del *data warehouse* de la compañía.

Dado la diversidad de variables a utilizar, se agruparon en cuatro grupos:

- Variables sociodemográficas
- Variables de comportamiento
- Variables de fidelidad
- Variables de conocimiento del producto

- Variables de entorno

A continuación se describe cada uno de los grupos de variables:

5.1.1 Variables sociodemográficas

Corresponden a características personales de los clientes. En este grupo de variables encontramos por ejemplo, la edad, el género, el GSE, entre otras. En anexos se muestran las variables utilizadas para este grupo.

5.1.2 Variables de comportamiento

Estas variables tienen la finalidad de detectar tanto el comportamiento transaccional del cliente como su comportamiento respecto al club de fidelidad de la empresa. Además, con este conjunto de variables se pretende identificar cambios en los patrones de comportamiento, por ejemplo, el hecho de que un cliente aumente su gasto promedio con la tarjeta o que el cliente decida aumentar su cupo.

Con respecto al comportamiento transaccional se busca capturar el nivel de gasto del cliente, el número de transacciones, entre otra información y con respecto a lo que se busca capturar con respecto al club de fidelidad de la empresa es por ejemplo si el cliente tiene un buen manejo y conocimiento de este. Una variable que se incluye en este último punto es el máximo nivel de canje del cliente, ya que se cree que el comportamiento de un cliente que siempre canjea el mínimo nivel (5.000 puntos) tiene un comportamiento completamente distinto a uno que ha canjeado por ejemplo 60.000 puntos.

En anexos se muestran las variables de este grupo con su respectiva descripción.

5.1.3 Variables de fidelidad

Estas variables se pueden dividir en dos: variables de fidelidad a la empresa (tarjeta) o variables de fidelidad al holding. Dentro de estas variables tenemos: la participación On Us del cliente, indicador que captura el porcentaje de gasto del cliente en comercios del holding, el SOW indicador que es estimado no solo para la empresa si no que para los distintos rubros. Por ejemplo, SOW supermercados se calcula como el gasto con la tarjeta en supermercados partido en el potencial de gasto del cliente en el rubro de supermercado. Finalmente se tiene una variable que indica si el cliente está enganchado. Son cuatro variables binarias: una para indicar si está enganchado a la empresa, otra para indicar si está enganchado por el club de fidelidad de la empresa, es decir los puntos, otra para ver si lo está de las promociones y finalmente ver si está enganchado por las opciones de financiamiento que le ofrece la empresa. Ver anexo 7.

5.1.4 Variables de conocimiento del producto

Son variables que indican si el cliente ya está en el mundo de los créditos de consumo y sabe cómo funcionan. Además estas variables pretenden capturar si existe cierta adicción del cliente a estar endeudado. Variables de ese grupo son: si el cliente ha sacado alguna vez un crédito de consumo en la empresa, si el cliente está o no actualmente pagando un SAV, entre otras. Ver en anexos otras variables que se incluyen en este grupo.

5.1.5 Variables de entorno

Estas variables caracterizan hitos o acontecimientos especiales del cliente durante su permanencia en la compañía. Dentro de estas variables esta la antigüedad del cliente, si es un cliente nuevo con disponible, es decir, si hace poco tiempo es un cliente que tiene la oportunidad de sacar un súper avance, entre otras variables. Ver anexos.

5.2 Preprocesamiento y transformación de variables

Algunas veces la manera en que aparecen los datos no es la mejor forma de representar la información disponible en una base de datos. Debido a lo anterior, se deben preprocesar los datos con la finalidad de corregir errores e inconsistencias y luego se pueden realizar transformaciones que permitan un enriquecimiento de la información entregada.

5.2.1 Etapa de preprocesamiento

En esta etapa se consideró el tratamiento de valores faltantes y el tratamiento de outliers (valores fuera de rango). Para esto se trató cada variable de una manera en particular:

- El GSE poseía 7 niveles: ABC1, C2, C3, D, E, OTRO y <null>. Como el GSE es de acuerdo a la dirección informada por el cliente al momento de aperturar la tarjeta, existen direcciones que no se pueden clasificar a un grupo socioeconómico debido a que a la cuadra a la que pertenecen es muy heterogénea. Por lo tanto, todas las direcciones que son <null> se dejaron en el grupo de OTRO.
- Para algunas variables se eliminaron los registros nulos, ya que estos casos correspondían a clientes que habían cerrado la tarjeta y por ende no contaban con la opción de sacar un PF. Ejemplos de estas variables son: recency y enganche del cliente
- Algunas de las variables que contenían registros nulos se rellenaron con ceros. Ejemplos de estas variables son: máximo nivel de canje, puntos acumulados en el último año calendario, días con compra en los últimos 3 meses y la participación *On Us* en los últimos 6 meses.

5.2.2 Etapa de transformación

En una primera etapa se recolecto toda la data necesaria para la realización del modelo sin considerar que algunas de estas variables seleccionadas no se encontraban en condición de ser utilizadas en su estado original. Debido a esto, se realizó un conjunto de transformaciones de la data original, las cuales buscan enriquecer la información que se puede obtener del modelo y por supuesto hacerlas compatibles con las técnicas de modelamiento.

Para la transformación de las variables se consideró un análisis descriptivo previo en dónde por ejemplo, para capturar el efecto de la edad se realizaron 3 intervalos debido a que esta se asemeja a una función cuadrática.

- Transformación de variables categóricas a binarias. Ejemplo: si una persona es un cliente nuevo o antiguo en súper avance.
- Transformación de variables numéricas a categóricas. Ejemplo: Máximo nivel de canje, se pasó a 3 niveles: bajo (clientes que habían canjeado a los más 5.000 puntos), medio (clientes que habían canjeado entre 5.000 y 24.000), alto (clientes que habían canjeado más de 24.000 puntos).
- Transformación de la fecha de nacimiento en años y fecha apertura cuenta en antigüedad del cliente.
- Estandarización de variables numéricas dentro de un rango específico [0,1]. Ejemplo: Participación *On Us* durante los últimos 3/6 meses.
- Generación de nuevas variables a partir de las variables iniciales. Ejemplo: variación porcentual de la saturación de la tarjeta con respecto al mes anterior y la saturación del cupo de la tarjeta en sí.

5.3 Modelamiento

La construcción de los modelos se lleva a cabo en dos etapas, una de entrenamiento y otra de evaluación. Para cada etapa se considera un subconjunto de clientes que en este caso es un 80% para el grupo de entrenamiento y 20% para el de testeo.

Para la primera etapa, la de entrenamiento, se utiliza el modelado basado en árboles de decisión, en particular *Random Forest*. Se utilizarán 4 criterios de divergencia y se escogerá el que presente una mayor ganancia de información representado por el coeficiente de Qini, el cual mide la ganancia de información con respecto a una elección realizada al azar: área entre ambas curvas.

Una vez escogido el modelo, se continúa con la confección de la curva uplift. Para esto se ordenan los clientes de acuerdo al *lift* generado, recordando que este se calcula como $lift = P(y = 1|tratamiento = 1) - P(y = 1|tratamiento = 0)$. Una vez ordenados se prosigue con la agrupación de los clientes en deciles, siendo el primer decil el con una mayor *lift* promedio y el último decil el con el menor. Ya con los deciles formados se calcula el uplift, que es la tasa de respuesta incremental del decil: tasa de respuesta del grupo target menos la tasa del grupo de control. Finalmente, este último valor está representando el incremento en porcentaje de respuestas debido a la existencia de tratamiento.

5.3.1 Modelo campaña genérica

Para la creación de este modelo se utilizará la data generada con el experimento, donde se considerará el tratamiento genérico y el tratamiento de financiamiento de un proyecto, ya que estos son los que tienen mayores tasas incrementales con respecto al grupo de control (de un 0,14% y un 0,07% respectivamente).

Se cuenta con 135.300 registros, de los cuales 44.537 son del tratamiento genérico, 44.427 del tratamiento de financiamiento de un proyecto y 46.336 son registros del grupo de control. Además, se tiene un total de 712 respuestas por lo que al calibrar la base con un 5% de éxitos se obtiene un total de 14.212 registros.

Para la realización del modelo se partió utilizando todas las variables que se tenían a disposición para luego ir eliminándolas una a una de acuerdo a la importancia que se le atribuye a la variable por la mejora en el criterio de separación. El comando utilizado en R Studio es *varImportance()*.

Finalmente, nos quedamos con ocho variables las que se muestran en el gráfico a continuación según su importancia relativa:

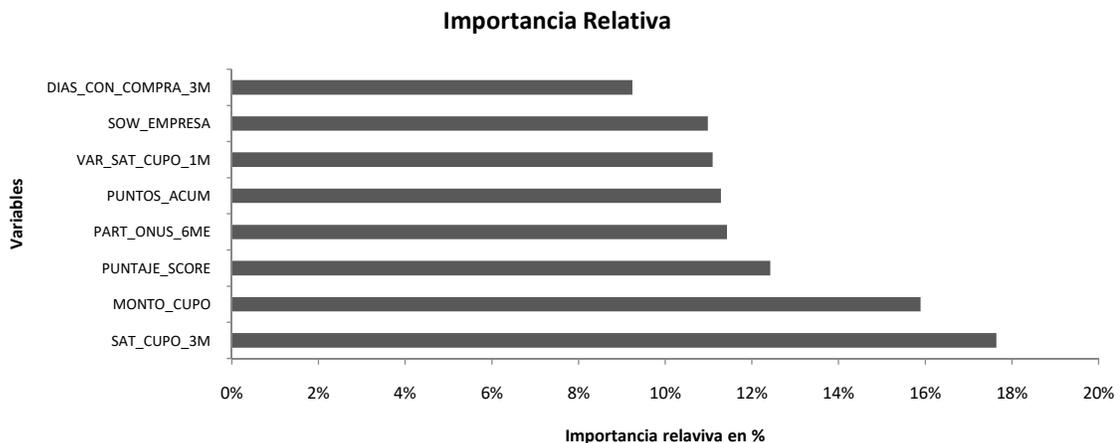


Ilustración 20: Importancia relativa de las variables del modelo genérico. Elaboración propia en R.

Donde cada variable significa:

Tabla 13: Descripción variables modelo genérico. Elaboración propia.

Variable	Descripción
DIAS_CON_COMPRA_3M	Días con compra en los últimos 3 meses
SOW_EMPRESA	Share of Wallet de la empresa
VAR_SAT_CUPO_1M	Variación porcentual de la saturación del cupo con respecto al mes anterior
PUNTOS_ACUM	Puntos acumulados en el último año calendario
PART_ONUS_6ME	Participación On Us del cliente durante 6 meses
PUNTAJE_SCORE	Indicador generado por el área de riesgo
MONTO_CUPO	Cupo de la tarjeta
SAT_CUPO_3M	Saturación de la tarjeta tres meses antes

El modelo que obtuvo el mejor rendimiento basado en su criterio de divergencia fue el método de interacciones, la tabla 14 muestra el coeficiente de Qini para cada criterio de divergencia de *Random Forest*. Ver en anexos la curva de ganancia de información.

Tabla 14: Q de Qini para los distintos criterios de divergencia para *Random Forest*. Elaboración propia en R.

	Divergencia Kullback-Leibler	Distancia Euclidiana	Divergencia Chi-Cuadrado	Método de Interacciones
Q de Qini	1,53%	1,46%	1,42%	1,66%

A continuación se muestra la tabla 15 que resume los resultados del modelo uplift para campañas genéricas. Se puede observar que en cada decil existen 1.100 clientes (sumando el grupo target y grupo de control). Con estos resultados se puede ver lo que realmente pasó dentro de cada decil, es decir, el número de clientes de cada grupo y además el número de sus respectivas respuestas. Finalmente, esto permite el cálculo del uplift, calculado como la TR del grupo target menos la TR del grupo de control. En la ilustración 21 se muestra la curva uplift que se genera.

Tabla 15: Tabla resumen modelo uplift basado en *Random Forest* ordenado en deciles. Elaboración propia en R.

Decil	N° clientes tratados (GT)	N° clientes no tratados (GC)	N° respuestas GT	N° respuestas GC	TR GT	TR GC	Uplift
1	558	542	176	3	31,54%	0,55%	30,99%
2	696	404	100	1	14,37%	0,25%	14,12%
3	698	402	46	1	6,59%	0,25%	6,34%
4	694	406	25	0	3,60%	0,00%	3,60%
5	760	340	19	3	2,50%	0,88%	1,62%
6	760	340	17	1	2,24%	0,29%	1,94%
7	783	317	10	3	1,28%	0,95%	0,33%
8	779	321	5	4	0,64%	1,25%	-0,60%
9	794	306	8	23	1,01%	7,52%	-6,51%
10	778	322	3	113	0,39%	35,09%	-34,71%

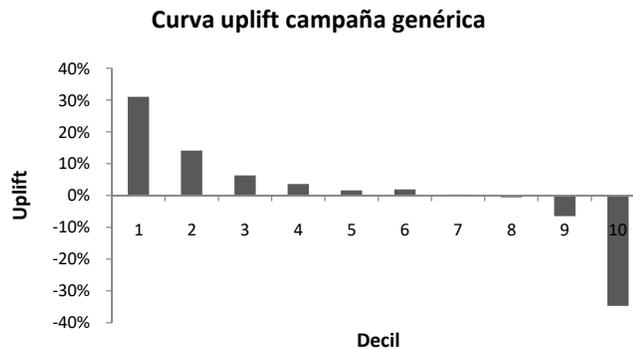


Ilustración 21: Curva uplift basada en *Random Forest* con el método de interacciones ordenado por deciles. Elaboración propia.

Del gráfico se observa que existen claras diferencias entre los distintos deciles. Resaltan los dos primeros deciles con uplift superiores al 10% y por el otro lado, se observa el decil 10, el cual tiene un evidente decaimiento del uplift (cercano al -30%). Existen además deciles intermedios, del decil 3 al 9, donde se observa un incremento o decaimiento muy bajo.

La manera de interpretar la curva es considerar a los dos primeros deciles como clientes influenciados a las intervenciones de marketing, en particular, a las campañas genéricas realizadas. Esto debido a que la campaña genera un considerable incremento en su probabilidad de sacar un producto financiero.

El decil 10 se puede considerar como el decil de los clientes *sleeping dogs*, es decir son clientes que tienen una elevada probabilidad a sacar un PF cuando no son intervenidos. Lo anterior puede deberse a que el cliente se molesta por los envíos o simplemente la campaña hizo que el cliente cambiara su comportamiento de manera negativa para la empresa: por ejemplo, que cliente decidiera informarse acerca de los créditos y haya preferido sacar el crédito de consumo en la competencia.

Finalmente tenemos del decil 3 al 6 con un leve incremento del *lift*, estos clientes pueden ser considerados como los clientes seguros (*sure things*), clientes que sacarían el súper avance sin la necesidad de la intervención de marketing. Por último está del decil 7 al 9, clientes que consideraremos como causas perdidas (*lost causes*), ya que son clientes que en ningún caso van a sacar un crédito de consumo.

5.3.2 Modelos ganchos

Como se puede observar en el apartado de evaluación y análisis de campañas, las respuestas obtenidas en el experimento de ganchos son muy bajas para la realización del modelo, por lo que este se realizará con data histórica de la empresa. Los resultados de los experimentos se utilizarán para realizar una validación de los modelos.

Recordando que las campañas de PF no se hacen con un grupo de control adecuado, al momento de decidir que los modelos se harán con data histórica es indispensable además pensar en la construcción de un grupo de control.

Es importante mencionar que las campañas de ganchos se hacen con filtros por disponible, es decir, la manera de armar la base de envío es simplemente tomando a toda la gente que posee disponible súper avance y aplicarle el filtro correspondiente. Al realizar esto, los clientes del grupo de control mensual de riesgo (pieza 0) son tocados vía email, es decir, no se pueden usar esos clientes para armar el grupo de control necesario para la confección de los modelos.

Debido a que la confección de un grupo de control es indispensable para la realización del modelo, la solución más adecuada al problema es:

- Acceder a los clientes no contactables vía email ya que estos son los únicos que no fueron tocados mediante este canal.

- Obtener clientes de la base de las campañas que rebotaron o que se cayeron al momento de subir la base a *exact target*. La idea de esto es aumentar el número de individuos en el grupo de control considerando lo pequeñas que son las tasas de respuesta (entre un 0,2% y un 0,6%).

5.3.2.1 Modelo campaña de giftcard

Para la construcción de este modelo se utilizará una campaña de giftcard realizada en febrero del año 2017, la cual constó de dos envíos realizados el mismo día. Esta campaña corresponde a un regalo de una giftcard de \$120.000 por sacar un SAV de \$2.500.000 o más en 36 cuotas o más.

Se cuenta con 74.437 registros para el grupo target de los cuales se tienen 175 respuestas. Para el grupo de control se tienen 66.625 registros: recordemos que una parte del grupo de control fue creada con los correos que se cayeron al momento de subir la base de la campaña a Exact Target incluyendo los rebotes (17.503 registros) y una segunda parte de este grupo está compuesta por clientes no contactables (49.122 registros). Para el grupo de control se obtienen 81 respuestas, 47 son de los clientes no contactables y 34 de la base de envío que se cayeron, por lo tanto de clientes contactables. La tasa de respuesta incremental para esta campaña fue de un 0,11% la cual esta sobreestimada debido a que el grupo de control está conformado por aproximadamente un 75% de clientes no contactables los cuales se comportan diferente, es decir, sacan menos productos financieros (ver resultados de modelo predictivo convencional donde la contactabilidad de un cliente vía email influye positivamente en la probabilidad de que un cliente opte por un producto financiero).

Considerando que se tienen tan solo 81 respuestas para el grupo de control y 175 para el grupo target, se realiza el balanceo triplicando los casos positivos para mejorar la performance del modelo.

La base se balancea con un 5% de éxitos, teniéndose así un total de 15.360 registros (siendo 768 casos de éxito) para modelar.

Este modelo basado en *Random Forest* obtuvo su mejor rendimiento con el método de interacciones con un coeficiente de Qini igual a 3,01%. En anexos se puede ver el coeficiente de Qini para cada uno de los criterios de divergencia y además se muestra la curva de ganancia de información para el caso del método de interacciones.

Para la selección de variables se ocupa el comando *varImportance()* de R, quedando finalmente 10 variables:

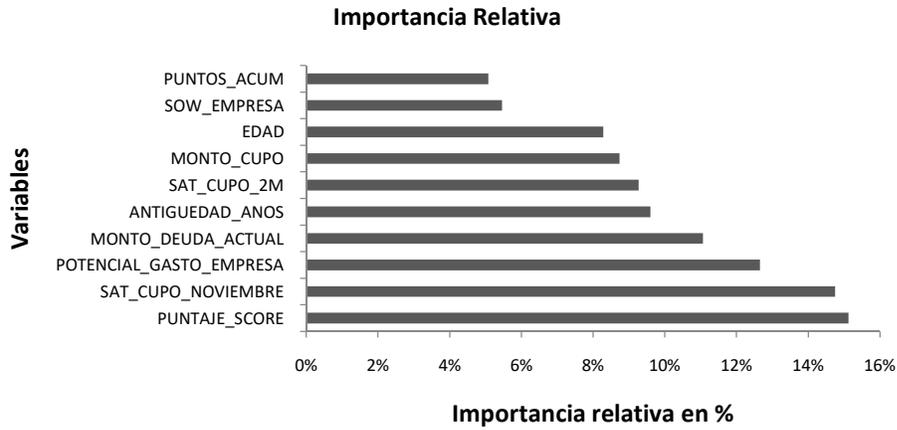


Ilustración 22: Importancia relativa de las variables del modelo de giftcard. Elaboración propia en R.

A continuación se describe cada una de las variables del modelo:

Tabla 16: Descripción variables modelo giftcard. Elaboración propia.

Variable	Descripción
PUNTOS_ACUM	Puntos acumulados en el último año calendario
SOW_EMPRESA	Share of Wallet de la empresa
EDAD	Edad del cliente
MONTO_CUPO	Cupo de la tarjeta
SAT_CUPO_2M	Variación porcentual de la saturación del cupo con respecto a dos meses antes
ANTIGUEDAD_ANOS	Antigüedad del cliente en años
MONTO_DEUDA_ACTUAL	Monto total adeudado del cliente
POTENCIAL_GASTO_EMPRESA	Gasto potencial del cliente en CMR
SAT_CUPO_NOVIEMBRE	Saturación de la tarjeta dos meses antes
PUNTAJE_SCORE	Indicador generado por el área de riesgo

La curva uplift para la campaña de giftcard queda como sigue:

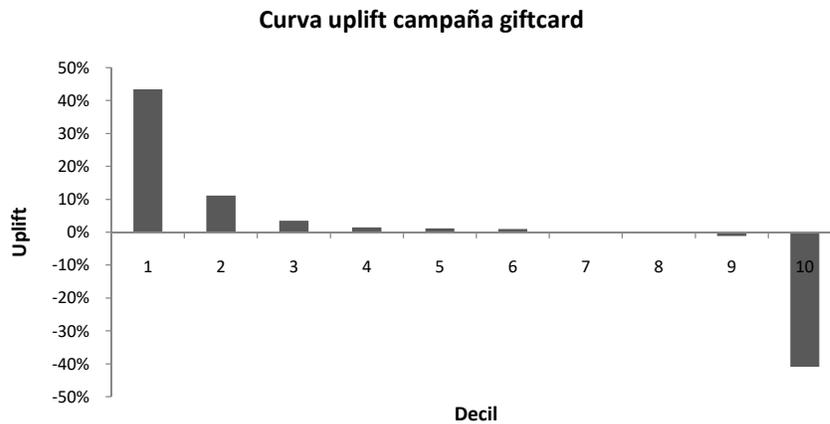


Ilustración 23: Curva uplift basada en Random Forest con el método de interacciones ordenado por deciles. Elaboración propia.

En el anexo 13 se puede ver la tabla resumen del modelo uplift basado en *Random Forest* para el método de interacciones ordenado en deciles.

Al igual que para el caso de las campañas genéricas se considerarán los dos primeros deciles como clientes influenciables y el último como los *sleeping dogs*. Del decil 3 al 9 serán los *sure things* y los *lost causes*.

5.3.2.2 Modelo campaña de puntos

Para este modelo se ocupa una campaña masiva realizada en enero el 2017, la campaña es por todo el mes de enero pero el envío masivo se realizó en 04 del mes, por lo tanto, como se está viendo el efecto de la publicidad vía email, se tomarán como respuestas afirmativas todos los clientes que sacaron un crédito de consumo del 4 de enero en adelante. La campaña corresponde a un regalo de 36.000 puntos por sacar un SAV de \$2.500.000 o más en 36 cuotas o más.

En este caso se tienen 267.464 registros para el grupo target y 42.456 para el grupo de control (formado al igual que para el caso de la campaña de giftcard con clientes no contactables y con clientes que se cayeron de la base de envío al momento de subirla más los rebotes). Las respuestas del GT son 1.334 mientras que las del GC son solo 115, siendo con estos resultados la tasa de respuesta incremental de un 0,23% (sobrestimada).

Debido a que se tienen pocas respuestas positivas para el caso del grupo de control, se realiza el balanceo duplicando los casos afirmativos de este, quedando finalmente 230 respuestas afirmativas en el grupo de control.

Finalmente, al igual que para los dos modelos anteriores, la base se balancea con un 5% de éxitos quedando un total 31.280 registros para modelar.

Al igual que los dos modelos anteriores, el método de interacciones es el que tiene un mejor rendimiento con un coeficiente de Qini de 1,57%. En anexos están los coeficientes de Qini para cada método de separación y además está la curva de ganancia de información.

Las variables del modelo según su importancia relativa son:

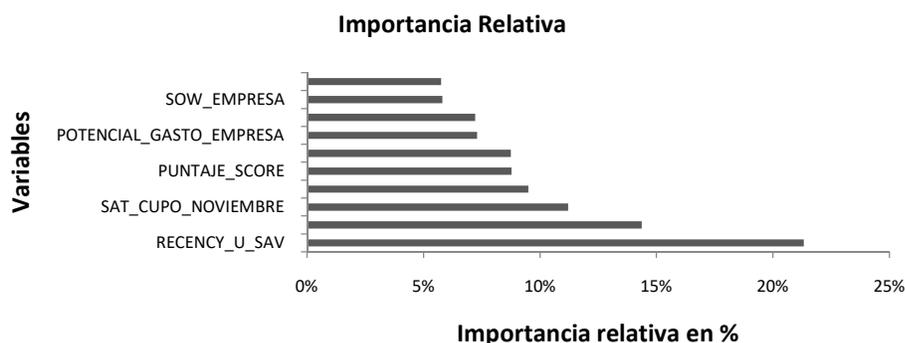


Ilustración 24: Importancia relativa de las variables del modelo de puntos. Elaboración propia en R.

Dónde *recency_u_sav* es una variable que captura cuánto ha pasado en meses de la última vez que saco un SAV. Las demás variables están descritas anteriormente.

A continuación, la curva uplift para el modelo de puntos:

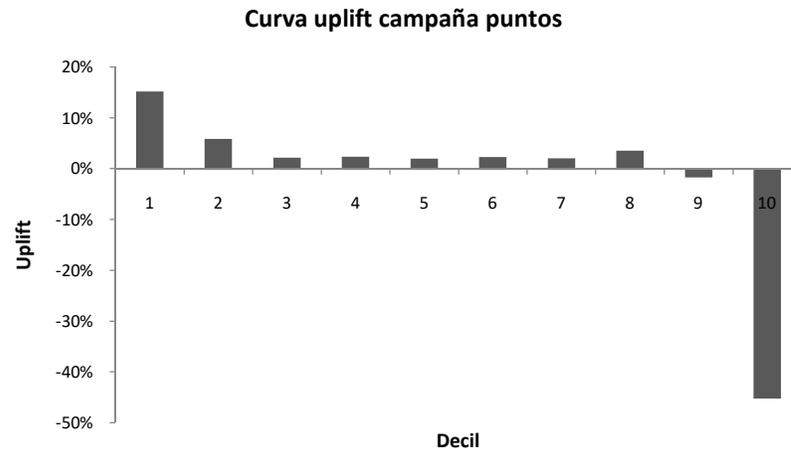


Ilustración 25: Curva uplift basada en *RandomForest* con el método de interacciones ordenado por deciles. Elaboración propia.

Al igual que en los dos casos anteriores, los dos primeros deciles serán considerados como los clientes influenciables mientras que el último, el decil 10, será considerado como el decil de los clientes *sleeping dogs*.

5.4 Interpretación y evaluación de modelos

5.4.1 Validación modelos

Para la validación de los modelos se utilizaron campañas diferentes y se analizaron a posteriori con el objetivo de analizar los resultados que se habrían obtenido si es que se hubiera utilizado el modelo.

Para el caso del modelo de las campañas genéricas se utilizaron dos campañas, una realizada en el mismo mes en que se realizaron los experimentos (mayo del 2017) y otra realizada en julio del 2017.

Para el caso de la campaña realizada en mayo, se tuvo una tasa de respuesta incremental del 0,05% con respecto a su grupo de control, grupo formado a posteriori debido a que las campañas no lo poseen. Es importante mencionar que esta validación puede estar sobre ajustada debido a que los clientes a utilizar como grupo de control fueron los mismos que se utilizaron para la confección del modelo (recordar que se forman utilizando la pieza 0).

A continuación se muestran los gráficos con las tasas de respuesta incremental que se hubiesen obtenido si es que en la campaña se hubiese tocado a un porcentaje diferente de clientes, es decir, se hubiese ocupado el modelo uplift.

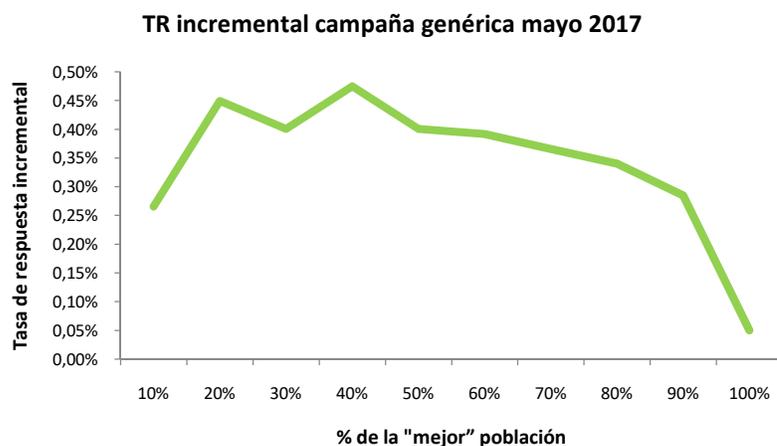


Ilustración 26: Gráfico de validación del modelo genérico con campaña realizada en mayo del 2017. Elaboración propia.

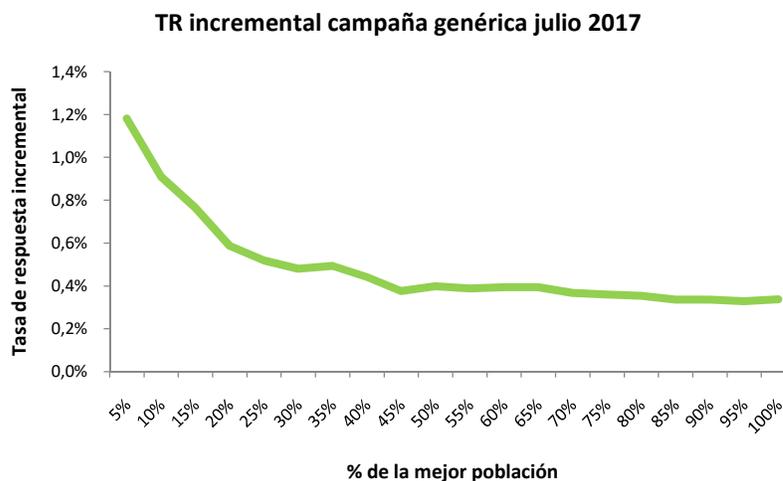


Ilustración 27: Gráfico de validación del modelo genérico con campaña realizada en julio del 2017. Elaboración propia.

En la ilustración 26 se observa claramente que tocando a un porcentaje menor de la población se obtienen mejores resultados que tocando al 100%, por ejemplo:

- Si hubiésemos tocado al 20% mejor de los clientes, se habría tenido una tasa de respuesta incremental 8,5 veces mejor que la que realmente se obtuvo (0,05%).
- Si se hubiese tocado al 40% mejor, la tasa de respuesta incremental habría sido 9,5 veces la tasa de respuesta incremental real de la campaña.

Es importante mencionar que para la validación realizada con la campaña de mayo del 2017 (ilustración 26), los clientes del grupo de control que se utilizaron fueron los mismos que se ocuparon para la confección del modelo, por lo tanto es probable que

esta curva tiende a parecerse en demasía a la curva de Qini generada por el modelo (ver anexo 10).

En la ilustración 27 podemos ver que el panorama cambia, es decir, los primeros deciles concentran la mayor cantidad de clientes influenciados encontrándose el peak antes del 20% mejor de la población. Para este caso tenemos que:

- Si se hubiese tocado al 5% mejor de los clientes, se habría obtenido una tasa de respuesta incremental del 1,18% en la campaña.
- Si se hubiese tocado al 10% mejor de los clientes, se habría obtenido una tasa de respuesta incremental del 0,91% en la campaña.

Es importante tener en cuenta que el único costo de la realización de una campaña genérica es el costo del envío del email, es decir, \$0,25 por envío. Lo que dicen los resultados de ambas validaciones es que en ningún caso el óptimo de envío de una campaña sería tocar al 100% de los clientes. Para el primer caso se observa un óptimo en los primeros 4 deciles y en el segundo caso, el óptimo se observa en el primer decil. Dicho esto, al momento de realizar el corte, este debería ser a lo más en el 40% mejor de los clientes.

Para la validación de los modelos de puntos y giftcard se utilizan los experimentos realizados, ya que estos no pudieron ser ocupados para el modelamiento debido a las bajas tasas de respuesta que se obtuvieron. Además se utilizará una campaña masiva de puntos realizada en julio del 2017 para la validación del modelo de puntos, la cual consiste en un regalo de 24.000 puntos del club de fidelidad de la empresa por sacar un súper avance desde \$2.500.000 en 48 cuotas.

Primero se mostrarán las validaciones realizadas con los experimentos llevados a cabo en el trabajo de título, luego se muestra la validación que se hizo con la campaña de puntos de julio del 2017. Para el caso de giftcard no se pudo llevar a cabo una segunda validación ya que no existían más campañas para su realización.

A continuación se muestra el gráfico con las tasas de respuestas incrementales para la validación realizada con el experimento de ganchos (puntos):

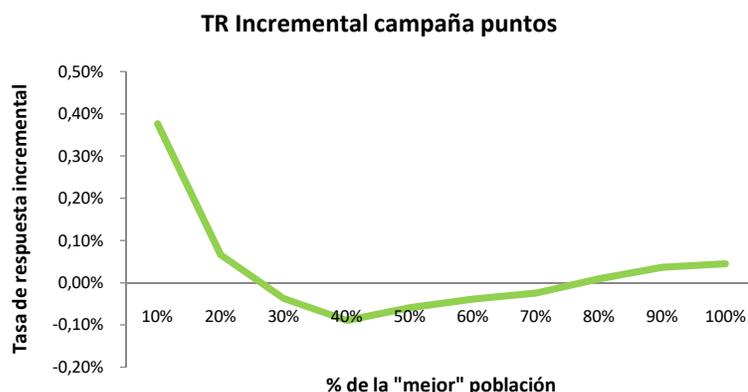


Ilustración 28: Gráfico de validación modelo puntos realizado con datos generados por experimento realizado en mayo del año. Elaboración propia.

Recordando que la tasa de respuesta de esta campaña fue de 0,14% y la tasa de respuesta incremental fue de un 0,04%, tenemos que en la ilustración 28 se observa que si en la misma campaña se hubiese tocado solo al 10% mejor de la población, la tasa de respuesta incremental hubiese sido 2,6 veces la tasa de respuesta del experimento y 8,5 veces su tasa de respuesta incremental.

El gráfico a continuación muestra las tasas de respuesta incrementales para el caso del experimento con el tratamiento de regalo de una giftcard:

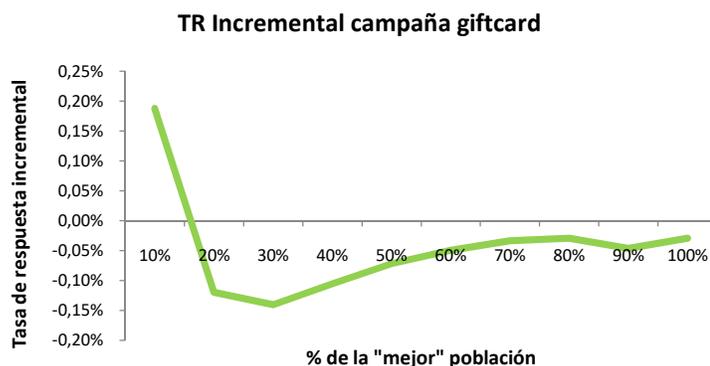


Ilustración 29: Gráfico de validación modelo giftcard realizado con datos generados por experimento realizado en mayo del 2017. Elaboración propia.

Al igual que en el caso anterior se observa que si se hubiese enviado la campaña solo al 10% mejor de los clientes, la tasa de respuesta incremental de la campaña hubiese sido 2,8 veces la tasa de respuesta y 7,6 veces la tasa de respuesta incremental. Recordar que para este caso la tasa de respuesta fue de un 0,07% mientras que la incremental fue de un -0,03%.

Es importante mencionar que el experimento de ganchos tuvo respuestas bastante pequeñas (23 para puntos, 11 para giftcard y 4 para el grupo de control), por eso es que se observa cierta inestabilidad en las curvas. Sin embargo es importante notar, que aunque las respuestas fueron pocas y por ende era probable que se tuviera mucho error en la validación, el modelo funciona, concentrando en el primer decil la mayor cantidad de clientes influenciados.

A continuación se muestra el gráfico de las tasas de respuesta incremental que se hubiesen obtenido si se hubiera utilizado el modelo uplift realizado para puntos en la campaña masiva realizada en julio del 2017:

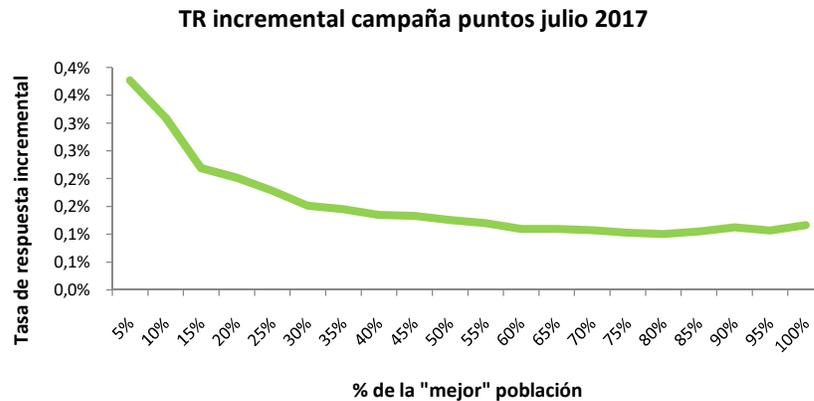


Ilustración 30: Gráfico validación modelo puntos con campaña masiva realizada en julio del 2017. Elaboración propia.

En la ilustración 30 se observa que las tasas de respuesta incrementales para esta campaña (julio 2017), tienen el mismo comportamiento que para la validación realizada con los experimentos, es decir, efectivamente tenemos un peak en el primer decil.

Finalmente es importante mencionar que cuando se dice el X% mejor de la población hace referencia a los clientes con mejor *lift*. Por ejemplo, para el 50% mejor de la población, tomamos al 50% con mejor *lift* para el grupo target y el grupo de control. De esta forma maximizamos las respuestas, ya que el grupo target responderá debido a la publicidad y el grupo de control tendrá muy pocas respuestas en el sentido que estos clientes sólo responden cuando existe un estímulo de marketing (definición de clientes influenciables).

5.4.2 Perfilamiento de clientes influenciables

Para el perfilamiento de clientes influenciables se realiza una comparación con los clientes *sleeping dogs* de su respectiva campaña. Además se agrupan las variables de acuerdo a:

- Fidelidad y comportamiento
- Conocimiento del producto
- Situación financiera
- Sociodemográficas
- Oportunidad

Con la finalidad de comparar si las proporciones/medias para los influenciables y *sleeping dogs* son estadísticamente significativas, se realizarán test de hipótesis de dos colas con un nivel de significancia el 95%.

Clientes influenciables campañas genéricas

Con respecto a las variables de fidelidad y comportamiento tenemos:

Tabla 17: Tabla perfilamiento con variables de comportamiento y fidelidad. Elaboración propia.

Variables de comportamiento y fidelidad	Campaña Genérica			
	Influenciables	Sleeping Dogs	Diferencia	P-valor
Proporción de clientes enganchados a la empresa	79,7%	71,1%	8,6%	2,72e-08
Proporción de clientes enganchados por puntos	62,1%	48,9%	13,2%	1,48e-13
SOW Empresa	35,8%	30,9%	4,9%	7,97e-9
Proporción de clientes con un nivel de canje mayor a 24.000 puntos	7,36%	7,88%	-0,5%	0,584
Proporción de clientes con un nivel de canje menor o igual a 5.000 puntos	51,68%	61,82%	10,7%	1,48e-8
Promedio de puntos acumulados	13257	14158	-901	0,24
Días con compra últimos 3 meses	13	11	2	8,21e-05
Participación On Us	32,2%	44,2%	-12,0%	1,03e-13
Proporción de clientes elite	5,16%	6,94%	-1,8%	0,0348
Proporción de clientes premium	28,79%	22,13%	6,7%	2,42e-5
Proporción de clientes normal	66,05%	70,93%	-4,9%	0,0037
Saturación del cupo 2 meses	31%	26%	4,9%	6,93e-9
Variación porcentual de la saturación del cupo 2 meses	62%	43%	19%	1,37e-17

Analizando la tabla 17 tenemos que los clientes influenciables son clientes más fieles a la empresa: tienen mayor proporción de clientes enganchados y además poseen un *share of wallet* de la empresa 4,9% mayor que los clientes *sleeping dogs*. Con respecto a las variables de puntos tenemos que la acumulación no presenta diferencias significativas (p-valor de 0,24), pero son los clientes *sleeping dogs* los que tienen una mayor proporción de clientes que realiza canjes por montos bajos (menores o igual a 5.000 puntos). Esto último tiene sentido ya que los clientes influenciables son significativamente más conocedores del programa de fidelidad de la empresa.

La variable de participación On Us nos indica que a pesar de que exista una mayor fidelidad a la empresa (a la tarjeta), esta fidelidad no es tan fuerte hacia los comercios del holding: como se observa, los clientes *sleeping dogs* tienen en promedio un 12% más de participación en las empresas del holding.

Es importante notar que la acumulación de puntos es levemente más baja (no significativa) para los clientes influenciables lo que puede deberse a su menor fidelidad a las empresas del holding. Notar que por cada \$150 pesos gastados *On Us* se acumula un punto mientras que *On Them* el gasto debe ser de \$300 para la acumulación de ese punto.

Finalmente decir que los clientes *premium* son significativamente más influenciables para el caso de las campañas genéricas siendo en proporción 6,7% más que los *sleeping dogs*: p-valor de 2,42e-5.

Con respecto al comportamiento se puede decir que los clientes influenciables tienen una frecuencia de compra y una saturación del cupo hace dos meses significativamente mayor. Además es importante notar que la variación de esta saturación del cupo es en promedio un 19% mayor para los clientes influenciables, es decir son clientes que se están activando en el uso de su tarjeta principal.

A continuación se muestra una tabla resumen para las variables consideradas que describen el estado económico de un cliente:

Tabla 18: Tabla perfilamiento con variables de situación económica. Elaboración propia.

Variables de estado económico	Campaña Genérica			
	Influenciables	Sleeping Dogs	Diferencia	P-valor
Proporción de clientes ABC1	8,40%	6,22%	2,2%	0,02
Proporción de clientes C2	18,42%	16,78%	1,6%	0,23
Proporción de clientes C3	11,41%	11,86%	-0,4%	0,7
Proporción de clientes D	31,00%	30,95%	0,05%	0,97
Proporción de clientes E	1,80%	1,88%	-0,1%	0,86
Monto cupo	\$ 2.126.029 (163%)	\$ 1.307.695 (100%)	\$ 818.334 (63%)	2,2e-16
Promedio potencial de gasto con la tarjeta	\$ 691.460 (121%)	\$ 572.309 (100%)	\$ 119.151 (21%)	8,95e-14

Se observa en la tabla 18 que los clientes influenciables tienen en promedio un nivel socioeconómico mayor: la proporción de clientes ABC1 es significativamente mayor en los clientes influenciables y además, se refuerza este hecho viendo las diferencias de la variable monto cupo, donde los clientes influenciables tienen un cupo de aproximadamente \$800.000 mayor que los clientes considerados *sleeping dogs*. Con respecto al potencial de gasto, este es casi \$120.000 mayor para los influenciables.

Con respecto a las variables de conocimiento del producto se tiene:

Tabla 19: Tabla perfilamiento con variables de conocimiento del producto. Elaboración propia.

Variables de conocimiento del producto	Campaña Genérica			
	Influenciables	Sleeping Dogs	Diferencia	P-valor
Proporción de clientes pagando Sav	39%	22%	17%	< 2,2e-15
Proporción de clientes pagando cuotas intermedias	15,53%	8,82%	6,7%	2,02e-8
Recency u sav	97	114	-17	1e-13
Deuda súper avance	\$ 550.494 (204%)	\$ 269.718 (100%)	\$ 280.776 (104%)	< 2,2e-15
Cliente nuevo	41%	60%	-19%	< 2,2e-15

Los clientes influenciados son más conocedores del producto y en cierto modo, adictos a este. Se observa en la tabla 19 que cerca del 40% de los clientes influenciados ya estaba pagando un súper avance versus un 22% existente en los *sleeping dogs*. Además, con respecto a la variable que me indica si un cliente es nuevo o antiguo, es decir, si el cliente ha sacado alguna vez el producto, tenemos que la proporción de clientes nuevos es casi un 20% mayor para los clientes *sleeping dogs*.

Otro aspecto relevante de los clientes influenciados es que tienen una deuda por créditos de consumo significativamente mayor y de acuerdo al promedio de la variable recency último súper avance, podemos decir que los influenciados tienen en promedio 17 meses menos.

Finalmente es importante mencionar con respecto a la variable cliente nuevo con disponible (clientes que hace menos de 6 meses pueden por primera vez sacar un SAV) que esta es en promedio 4,5% menor para los clientes influenciados, es decir, los clientes nuevos con disponible son más propensos a responder naturalmente, sin publicidad de por medio, ya que esta finalmente hace que cambien su comportamiento de manera negativa para la empresa.

Todas las variables pertenecientes al grupo de las variables de conocimiento del producto son estadísticamente significativas (test de dos colas).

Cientes influenciados campañas de puntos y giftcard

La tabla 20 muestra las variables de comportamiento y fidelidad para los clientes influenciados y *sleeping dogs* para el caso de puntos y giftcard.

Tabla 20: Tabla perfilamiento con variables de comportamiento y fidelidad. Elaboración propia.

Variables de comportamiento y fidelidad	Campaña Giftcard				Campaña Puntos			
	Influenciables	Sleeping Dogs	Diferencia	P-valor	Influenciables	Sleeping Dogs	Diferencia	P-valor
Proporción de clientes enganchados a la empresa	70%	58%	12%	< 2,2e-16	83%	62%	21%	< 2,2e-16
Proporción de clientes enganchados por puntos	47%	30%	17%	< 2,2e-16	71%	42%	29%	< 2,2e-16
SOW Empresa	33%	19%	14%	< 2,2e-16	44%	23%	20%	< 2,2e-16
Proporción de clientes con un nivel de canje mayor a 24,000 puntos	8,09%	5,33%	3%	0,0003	16,58%	8,06%	9%	< 2,2e-16
Proporción de clientes con un nivel de canje menor o igual a 5,000 puntos	50,66%	72,28%	-12%	3,37e-16	36,51%	62,42%	-26%	< 2,2e-16
Promedio de puntos acumulados	14.664	12.216	2.448	0,14	21.799	16.243	5.555	5,9e-5
Días con compra últimos 3 meses	12	8	4	< 2,2e-16	18	10	8	< 2,2e-16
Participación On Us	35,25%	30,55%	4,70%	5,69e-06	37,55%	31,60%	5,95%	7,74e-15
Proporción de clientes elite	5,06%	4,38%	0,69%	0,28	13,14%	9,17%	3,97%	4,23e-07
Proporción de clientes premium	28,70%	14,39%	14,31%	< 2,2e-16	43,93%	19,23%	24,70%	< 2,2e-16
Saturación del cupo 2 meses	17,70%	-4,06%	21,76%	< 2,2e-16	23,87%	5,81%	18,06%	< 2,2e-16
Monto deuda total	\$ 782.991 (179%)	\$ 437.058 (100%)	\$ 345.932 (79%)	< 2,2e-16	\$ 1.584.360 (198%)	\$ 799.367 (100%)	\$ 784.993 (98%)	< 2,2e-16
Variación porcentual de la saturación del cupo 2 meses	11%	-25%	36%	< 2,2e-16	-4%	-172%	168%	< 2,2e-16

Los influenciables por el gancho de puntos son clientes muy fieles a la empresa: son clientes que están enganchados a esta, en particular están más enganchados por puntos con una proporción casi un 30% mayor para los clientes influenciables. Para el caso de las campañas donde el gancho es el regalo de una giftcard, pasa lo mismo, aunque la diferencia no es tan marcada como para el caso de puntos, con un 17% más de clientes enganchados por puntos en el segmento de los influenciables. Ambas diferencias son significativas con un p-valor menor a 2,2e-16.

Con respecto a la variable *SOW_EMPRESA* que representa el *share of wallet* de la empresa, podemos ver que este es aproximadamente un 30% mejor para los clientes influenciables en el caso de puntos y un 14% más para el caso de giftcard. Es importante mencionar que el *share of wallet* es uno de los mayores indicadores de lealtad que una empresa puede tener.

Otros aspectos que refuerzan el hecho de que los clientes influenciables son más fieles para ambos casos son su acumulación de puntos, los niveles de canje y la categoría: la acumulación es en promedio 5.500 puntos mayor para el caso de puntos con una diferencia estadísticamente significativa. Por el contrario, la diferencia para el caso de giftcard no es significativa aunque se tiene una acumulación promedio de 2.500 puntos mayor (significativo al 93% aplicando un test de una cola). Con respecto a los canjes altos (mayor a 24.000 puntos), estos son el doble para los clientes influenciables en el caso de puntos y un 3% mayor para el caso de giftcard. De la mano con esto, está la categoría del cliente, donde la proporción de clientes *elite* y *premium* es casi un 30% mayor y un 15% mayor para el caso de puntos y giftcard respectivamente. Es importante mencionar que la proporción de clientes elite para el caso de giftcard no es estadísticamente significativa (p-valor de 0,28).

Con respecto a las variables de comportamiento se puede decir que los clientes influenciados tienen una mayor frecuencia de compra, siendo esta diferencia más marcada para el caso de puntos. El monto de la deuda es mayor en los influenciados, lo que tiene relación con el uso de la tarjeta y, por último, podemos decir que los influenciados son clientes que se han estado activando en el uso de su tarjeta, lo que se puede ver en la variable de variación porcentual de la saturación del cupo de la tabla 20.

Finalmente, es importante observar la variable participación *On Us*, donde los clientes influenciados tienen en promedio una mayor participación, es decir, son clientes que son estadísticamente más fieles a los comercios del holding.

A continuación se muestra la tabla 21 con variables de situación económica:

Tabla 21: Tabla perfilamiento con variables de situación económica. Elaboración propia.

Variables de estado económico	Campaña Giftcard				Campaña Puntos			
	Influenciados	Sleeping Dogs	Diferencia	P-valor	Influenciados	Sleeping Dogs	Diferencia	
Proporción de clientes ABC1	8,77%	15,69%	-6,93%	9,21e-13	10,09%	9,68%	0,41%	0,16
Proporción de clientes C2	20,25%	24,70%	-4,45%	0,0004	23,86%	22,85%	1,01%	0,33
Proporción de clientes C3	12,44%	10,16%	2,27%	0,017	12,16%	13,00%	-0,84%	0,31
Proporción de clientes D	28,94%	23,34%	5,60%	2,39e-05	26,32%	25,73%	0,59%	0,59
Proporción de clientes E	1,32%	0,75%	0,56%	0,069	1,06%	1,29%	-0,22%	0,40
Monto cupo	\$ 1.997.653 (75%)	\$ 2.658.435 (100%)	-\$ 660.781 (-25%)	<2,2e-16	\$ 2.752.289 (116%)	\$ 2.366.045 (100%)	\$ 386.244 (16%)	<2,2e-16
Monto disponible súper avance	\$ 3.300.466 (85%)	\$ 3.884.316 (100%)	-\$ 583.850 (-15%)	<2,2e-16	\$ 3.875.743 (103%)	\$ 3.745.731 (100%)	\$ 130.012 (3%)	6,26e-07

Para este tipo de variables, los clientes influenciados para ambos casos difieren. Los influenciados para el caso de giftcard son clientes que en promedio tienen un nivel socioeconómico menor que sus respectivos clientes *sleeping dogs*. Lo anterior se ve al observar las proporciones de clientes para cada grupo socioeconómico. Para el caso de puntos, el GSE no presenta diferencias significativas.

Observando la variable de monto cupo, se tiene que los clientes que son influenciados para giftcard tienen un cupo de \$660.000 menor que los *sleeping dogs*, en cambio para el caso de puntos, los influenciados tienen un cupo promedio de aproximadamente \$380.000 mayor que los *sleeping dogs*. Lo mismo se observa con la variable de monto disponible súper avance. Finalmente podemos decir que para las campañas de puntos, los clientes influenciados tienen un nivel socioeconómico levemente mayor que los *sleeping dogs* respectivos, lo que se puede suponer viendo las variables del monto cupo y disponible SAV.

La siguiente tabla muestra las variables de conocimiento del producto. Es importante mencionar que la campaña de giftcard tenía casi un 100% de clientes nuevos por lo tanto no se muestra información con respecto a estas variables.

Tabla 22: Tabla perfilamiento con variables de conocimiento del producto. Elaboración propia.

Variables de conocimiento del producto	Campaña Giftcard				Campaña Puntos			
	Influenciables	Sleeping Dogs	Diferencia	P-valor	Influenciables	Sleeping Dogs	Diferencia	P-valor
Proporción de clientes pagando Sav					24%	16%	8,09%	6,42e-16
Proporción de clientes pagando cuotas intermedias					10,66%	4,60%	6,06%	< 2,2e-16
Recency u sav					87	92	-4	0,001
Deuda súper avance					\$ 232.073 (102%)	\$ 228.352 (100%)	3.721 (2%)	0,6725
Cliente nuevo	98,53%	98,89%	-0,37%	0,28	48,68%	54,40%	-5,72%	3,63e-06

Se observa en la tabla que los clientes influenciados para el caso de puntos podrían ser más conocedores del producto y en cierto modo ser más adictos al crédito, esto se puede inferir a partir de la variable que indica la proporción de clientes que está pagando un súper avance, donde en los influenciados esta proporción es el doble. Además se observa esto en las variables de cliente nuevo y recency último SAV, las cuales presentan diferencias significativas.

A continuación la tabla 23 incluye edad, antigüedad del cliente y si es nuevo con disponible:

Tabla 23: Tabla perfilamiento con variables diversas. Elaboración propia.

Otras variable	Campaña Giftcard				Campaña Puntos			
	Influenciables	Sleeping Dogs	Diferencia	P-valor	Influenciables	Sleeping Dogs	Diferencia	P-valor
Nuevo con disponible	2%	4%	-2%	0,0006	1%	3%	-2%	0,0004
Edad	52	43	9	< 2,2e-16	50	41	9	< 2,2e-16
Antigüedad años	13	12	1	7,552e-06	17	12	5	< 2,2e-16

Por último es importante mencionar las variables edad y cliente nuevo con disponible y antigüedad años. Con respecto a esto podemos decir que:

- Los clientes influenciados son en promedio casi 10 años mayor que los *sleeping dogs* (p-valor menor a 2,2e-16).
- Los clientes nuevos con disponible, es decir, los clientes que hace 6 meses o menos tienen por primera vez la oportunidad de sacar un súper avance, tienden a responder naturalmente a estos productos sin la necesidad de un estímulo de marketing, ya que debido a estos, cambian su comportamiento de manera negativa para la empresa.
- Para el caso de puntos, los influenciados son en promedio clientes más antiguos, es por esto que puede ser que para este caso, las variables de fidelidad y comportamiento sean más intensas que en los influenciados giftcard.

La tabla de abajo muestra una comparación con respecto a los diferentes grupos de variables para los casos de puntos y giftcard.

Tabla 24: Tabla resumen perfilamiento. Elaboración propia.

	Clientes influenciables	
	Puntos	Giftcard
Variables de fidelidad	++	+
Variables de comportamiento	++	+
Variables de conocimiento del producto	+	Sin Información
variables de situación económica	o	-

Finalmente podemos decir que los clientes influenciables son más fieles y tienen un mejor comportamiento desde el punto de vista de la empresa que los *sleeping dogs*. Esta diferencia, como mencionamos anteriormente es mucho más marcada para el caso de los influenciables por puntos. Otro aspecto relevante aquí es que los clientes influenciables por giftcard tienen una situación económica peor que los *sleeping dogs*, cuestión que puede explicar el hecho de que prefieran el regalo de una giftcard (es dinero que tendrán a su disposición por sacar un PF). En cambio para el caso de puntos, estos clientes son concedores del club de fidelidad de la empresa y en cierto modo se puede decir que son adictos a este.

5.4.3 Ejemplo práctico

En este apartado se analiza la campaña masiva de julio del 2017, recordando que en esta campaña se regalaban 24.000 puntos del programa de fidelidad de la empresa a los clientes que sacaran un SAV de \$2.500.000 en 48 cuotas. La finalidad de este análisis es tener una mirada más comercial y analizar el impacto monetario (ventas) del proyecto de título.

Recordando las validaciones realizadas, tenemos que esta campaña tuvo una TR incremental del 0,12%, pero tocando sólo al 10% mejor de los clientes, la TR incremental llega a ser 0,38%.

	Campaña Masiva Julio 2017	Aplicando el Modelo
Grupo Target	N = 185.066 TR = 0,23% Respuestas = 428	N = 18.507 TR = 0,58% Respuestas = 107
Grupo Control	N = 55.920 TR = 0,11% Respuestas = 64	N = 5.592 TR = 0,27% Respuestas = 15
TR Incremental Campaña	0,12%	0,38%
No Tocados	N = 0 TR = 0,15% Respuestas = 0	N = 216.887 TR = 0,15% Respuestas = 325
total Respuestas	492	447
Costo Total Campaña	\$ 30,862,267	\$ 7,708,627

Ilustración 31: Comparación campaña julio del 2017: sin aplicar modelo uplift y aplicando el modelo. Elaboración propia.

Para la construcción de la ilustración 31 se considera un universo de 240.986 y un costo por punto de \$3. Analizando por parte tenemos que en la campaña masiva se tocaron a 185.066 clientes los cuales tuvieron una TR de 0,23% lo que se traducen en 428 respuestas. Si se hubiesen aplicado los resultados del modelo uplift, es decir, se hubiese tocado al 10% mejor de los clientes, se hubiese tocado a 18.507 clientes, los cuales hubieran respondido a una tasa de 0,58%, es decir, 107 respuestas. Pasando al comportamiento del grupo de control, tenemos que para el caso de la campaña realizada en julio del 2017, el grupo de control (realizado a posteriori) está compuesto por 55.920 clientes con una tasa de respuesta del 0,11% lo que equivale a 64 respuestas y para el caso hipotético en que se hubiese aplicado el modelo, este grupo estaría compuesto por 5.593 clientes con una TR de 0,27% equivalente a 15 respuestas. Sacando el cálculo tenemos que las TR incremental para la campaña real y para el caso dónde se aplica el modelo es de 0,12% y 0,38% respectivamente. La siguiente sección "No Tocados" es el análisis realizado a los clientes que se dejaron de tocar debido a la aplicación del modelo, por lo tanto, para el caso de la campaña masiva, este número es 0. Viendo el caso hipotético de la aplicación del modelo, en este grupo estarían 216.887 clientes (90% de clientes no tocados) con una TR natural del 0,15% (TR natural a este producto financiero) lo que equivale a 325 respuestas. Sumando todas las respuestas de los 240.986 clientes analizados para ambos casos se tienen: 492 respuestas para la campaña real con un costo total de la campaña (costo de envío más el costo del regalo de 24.000 puntos por cliente que saca el SAV) de \$30.862.267, mientras que si se hubiera implementado el modelo se hubiesen tenido 447 respuestas pero el costo de campaña hubiese sido de \$7.708.627, es decir, la diferencia de dinero invertida en ambos casos es de más de 23 millones y la diferencia en respuestas es solo de 45, por lo tanto, para obtener cada una de esas respuestas se debería pagar un monto que supera los 500 mil pesos, lo cual evidentemente no es conveniente.

5.5 Análisis de riesgo clientes influenciables

Se busca analizar cómo evoluciona el riesgo de los clientes influenciables desde el momento que sacan un PF. La idea es ver si es rentable el destinar los recursos de la empresa solo en este segmento de clientes.

Dicho esto, se analizará cómo evolucionan 3 variables que indican si el cliente varía su riesgo. Estas variables son: si el cliente ha generado intereses por mora, si el cliente ha generado intereses por *revolving* (que es realizar el pago mínimo) y finalmente el puntaje score, recordando que esta última variable es creada por el área de riesgo (mientras más alto, menos riesgoso el cliente).

Para tener una idea de cómo se comportan estas variables en general, primero se analiza la evolución de la cartera de clientes con disponible comparando los clientes que sacaron un PF (Y=1) versus una que no (Y=0). El gráfico a continuación muestra un análisis realizado a clientes con disponible en mayo del 2016, por lo tanto, se considera si el cliente sacó el PF o no solo en el mes de mayo.

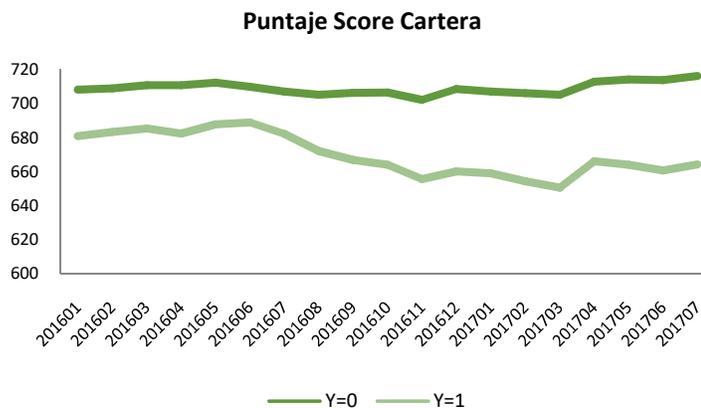


Ilustración 32: Evolución puntaje score cartera de clientes con disponible. Elaboración propia.

En el gráfico anterior se observa claramente que un cliente empeora su riesgo si es que opta por un crédito de consumo. Lo anterior se observa ya que para el caso de la curva de los clientes que sacaron el PF (Y=1), existe un quiebre en el mes de junio del 2016, un mes después de haber sacado el súper avance. La variación de su puntaje score es de aproximadamente 3,5 puntos porcentuales.

A continuación se muestra cómo varía el interés por mora y por revolving una vez sacado un crédito de consumo:



Ilustración 33: Variación del interés por mora y revolving promedio. Elaboración propia.

Es claro que una vez que un cliente decide sacar un producto financiero, su riesgo aumenta, lo cual además se puede ver en como aumentan los intereses promedio que han generado. Ver ilustración 33.

Una vez claro el comportamiento general de la cartera de clientes con respecto a su riesgo, se prosigue con un análisis que pretende comparar la evolución las mismas variables para los distintos segmentos encontrados con el modelo: clientes influenciables, *sleeping dogs* y causas perdidas, el cual también incluye a los *sure things*.

Tabla 25: Variación en puntos porcentuales del puntaje score para los distintos segmentos de clientes. Se toma la variación comparando mayo del 2016 con julio del 2017. Elaboración propia.

	Influenciables	Causas Perdidas	Sleeping Dogs
No contrata PF (Y=0)	-3	-6	-7,5
Contrata PF (Y=1)	-11	-9,5	-10,8
Diferencia	8	3,5	3,3

La tabla 25 muestra la variación entre mayo del 2016 y julio del 2017 de la variable puntaje score en puntos porcentuales. De la tabla es importante notar que:

- Los clientes influenciables tienen una mayor variación del puntaje score en comparación a los otros dos segmentos (aproximadamente 4.5 puntos porcentuales más).
- Es importante destacar que en promedio los *sleeping dogs* son peores clientes que los influenciables, esto se observa viendo el comportamiento para los clientes que no optaron por un PF (Y=0).

Es importante mencionar que la diferencia de la variación en puntos porcentuales del puntaje score para la cartera de clientes es de 4.5, mientras que los influenciables tienen una diferencia de 8.

La tabla 26 muestra la variación en puntos porcentuales de la proporción de clientes que pagó intereses por mora para los distintos segmentos de clientes:

Tabla 26: Variación en puntos porcentuales de la proporción de clientes que pagó intereses por mora para los distintos segmentos de clientes. Se toma la variación comparando mayo del 2016 con julio del 2017. Elaboración propia.

	Influenciables	Causas Perdidas	Sleeping Dogs
No contrata PF (Y=0)	1	2	1
Contrata PF (Y=1)	9	10	14
Diferencia	8	8	13

Se observa que la mayor variación en puntos porcentuales de la proporción de clientes que pagó intereses por mora está en los clientes *sleeping dogs* (13). Esta diferencia se mantiene constante para los segmentos de los influenciables y las causas perdidas.

Finalmente se concluye que a pesar de que los clientes influenciables tienen una diferencia en puntos porcentuales mayor cuando se trata del puntaje score, esta no se debe a que están generando mayores intereses por mora. Esto último es de gran relevancia, ya que el cliente influenciable tiene mejores hábitos de pago que los otros dos segmentos y la diferencia que existe con respecto a la variable puntaje score puede deberse a que el monto de la deuda total de un cliente es aproximadamente el doble que la de uno que no lo es.

6. Recomendaciones empresa y trabajo futuro

A partir de la realización de los experimentos y la etapa de modelamiento del trabajo de título se pueden obtener diversas recomendaciones y pasos a seguir para la empresa.

6.1 Recomendaciones

Lo primero que se recomienda es tener un grupo de control adecuado para cada una de las campañas que se realizan en productos financieros. Es de suma importancia este grupo debido a que se evita llegar a conclusiones erradas de las campañas. Por ejemplo, una de las conclusiones que se tenían en el área comercial de la empresa era que el tocar al 100% de los clientes era siempre lo mejor, ya que aumentaban las ventas. Otra conclusión era que siempre que se ofrece un gancho las transacciones aumentan considerablemente. Como se ve tanto en los resultados de los experimentos como en el modelo uplift, tenemos que tocar el 100% de los clientes no es el óptimo y que no siempre el hecho de ofrecer un gancho mejora las tasas de respuesta (ver resultados experimento gancho)

De acuerdo al modelo se puede decir que:

- Cuando se decide confeccionar un modelo uplift es importante tener en cuenta los tamaños muestrales (sobre todo si el modelo es basado en árboles de decisión), ya que el número de individuos restringe la creación de nuevos nodos al momento de aplicar un nuevo test para la separación. La población más

pequeña activa la restricción, la cual generalmente es el grupo de control. Tener en cuenta que los tamaños de la población afectan la profundidad de un árbol y el número de nodos.

- Al momento de escoger el corte óptimo de clientes a tocar, no siempre hay que considerar el punto que maximiza el número de respuestas incrementales, ya que hay que tener en cuenta que el objetivo es aumentar las ventas dado la inversión realizada. Por ejemplo, para el caso de las campañas genéricas se puede tener holgura en la elección del corte (la inversión solo es el costo de envío de un correo electrónico), pero para el caso de algún gancho, se debe ser estricto dado que la inversión es considerablemente mayor.
- Es importante que al realizar un modelo se realice una validación del mismo. Lo anterior debido a que por ejemplo el modelo puede decir que el óptimo es el 60% pero en la realidad el corte podría ser en el 40%. Es importante saber que esto ocurre debido a que la clasificación de datos distintos a los utilizados para el entrenamiento del modelo tenderá a ser más débil que la clasificación de los datos de entrenamiento.

Finalmente se recomienda tocar al 40% mejor de los clientes para el caso de que la campaña a realizar sea genérica, y para el caso de que en la campaña se ofrezca algún gancho, tocar a lo más al 20% mejor de los clientes.

6.2 Trabajo futuro

Transcurrido un tiempo prudente, es estrictamente necesario calibrar nuevamente el modelo para las campañas genéricas, ya que es muy probable que pasado un tiempo ocupando los resultados del modelo uplift (tocar sólo a un porcentaje de los clientes con disponible) las condiciones cambien. Actualmente, como se mencionó anteriormente las campañas genéricas les llegan a toda persona que posea disponible, es decir, todos los meses les llega a los clientes esta "promoción". Es claro que el comportamiento de los clientes una vez que sepan que la promoción no es válida todo el año, podría cambiar. Básicamente pasaría a ser una campaña más personalizada, no algo que todo el mundo puede tener.

Una vez que se comiencen a realizar las campañas con un grupo de control adecuado, se deben calibrar nuevamente los modelos realizados para ganchos, esto debido a que la construcción de los grupos de control para estos modelos se realizó con clientes no contactables lo que para este caso sobrestima los resultados.

Una vez que se tenga un motor diferenciador de tasas, es recomendable realizar un nuevo experimento (probar regalo de puntos, regalo de giftcard y el descuento en tasas) con la finalidad de confeccionar un solo modelo uplift para ganchos. Lo anterior permitiría a la empresa determinar cuál es el mejor gancho para un determinado cliente.

Conclusiones

Una vez finalizado el proyecto de título se puede concluir lo siguiente:

- El hecho de ofrecer algún incentivo a los clientes no necesariamente aumenta las ventas o las transacciones de algún producto (resultados experimentos). Es por esto que la realización de los modelos uplift son de suma importancia para poder destinar el presupuesto de marketing de manera eficiente, es decir, destinarlo a los clientes que efectivamente se verán influenciados por diferentes estímulos.
- De acuerdo a las validaciones realizadas para el modelo de las campañas con tratamiento genérico, el corte óptimo se debería encontrar dentro de los 4 primeros deciles. De esta forma, si tomásemos el mejor 40% de los clientes para la realización de campañas, se mejoraría considerablemente la eficiencia en el sentido que se liberarían cerca de 500.000 mil clientes mensuales a los cuales se les puede realizar algún otro tipo de comunicación y además, porque de esta forma evitamos el perder las ventas de los *sleeping dogs*.
- Con respecto a las validaciones de los modelos de giftcard y puntos, los óptimos se encuentran siempre en el primer decil, es decir, tocando al mejor 10% de los clientes, estaríamos maximizando las tasas de respuesta incremental de una campaña con gancho. Para este caso, es de suma importancia el respetar el corte debido a la inversión que se debe realizar para llevar a cabo estos estímulos.
- Es predecible que los óptimos para todos los casos se encuentren en los primeros deciles, ya que un crédito de consumo es un producto que se obtiene más por necesidad que por la existencia de estímulos.
- Los clientes influenciados a los estímulos de marketing para los diferentes modelos realizados, son particularmente clientes fieles a la empresa y activos en el uso de sus productos (tarjetas).
- Aplicar los resultados obtenidos en el presente trabajo tiene mejoras considerables en el desempeño de las campañas de productos financieros, ya que se mejora la personalización en la comunicación y además, se logra ahorrar aproximadamente 20 millones de pesos cuando la campaña se trata del ofrecimiento del algún gancho (puntos o giftcard).
- Finalmente es importante mencionar que es eficiente destinar recursos a clientes influenciados, ya que estos mantienen mejores hábitos de pago que los otros segmentos. Lo anterior puede deberse a la alta fidelidad de los influenciados tanto a la empresa en sí como a las compañías asociadas al holding.

Bibliografía

- [1] SBIF. Informe de tarjetas de crédito no bancarias (2015).
- [2] Memoria Anual Empresa.
- [3] Roberts, J. H., &Lilien, G. L. (1993). Explanatory and predictive models of consumer behavior. Handbooks in operations research and management science, 5, 27-82.
- [4] Radcliffe, N. J., &Surry, P. D. (2011). Real-worldupliftmodellingwithsignificance-baseduplifttrees. White Paper TR-2011-1, Stochastic Solutions.
- [5] Rubin, D. B. (1997). Estimating Causal Effects from Large Data Sets Using Propensity Scores. Annals of Internal Medicine. Vol. 127(8S), 757-763.
- [6] Kuhfeld, W. F., Tobias, R. D., &Garratt, M. (1994). Efficient experimental design with marketing research applications. Journal of Marketing Research, 545-557.
- [7] Rzepakowski, P., &Jaroszewicz, S. (2010). Decision trees for uplift modeling. In Data Mining (ICDM), 2010 IEEE 10th International Conference on (pp. 441-450). IEEE.
- [8] Fuenzalida, B.B. (2012). Pronóstico de efectividad de promociones sobre clientes, dadas sus características y respuestas pasadas. Memoria de Título, Universidad de Chile.
- [9] Calderón, B. A. (2016). Determinación de la propensión al aumento de consumo con tarjeta de crédito de clientes de una institución bancaria.
- [10] Zoccola, O. P. (2015). Estudio experimental de envío de correos automatizados basados en el comportamiento reciente de clientes en el sitio Web de un programa de lealtad de una línea aérea.
- [11] Wilson, D. (2013). Modelos de propensión integrados para la optimización de campañas de marketing.
- [12] Araya Abadie, M. A. (2011). Estimación de la Efectividad Promocional en un Supermercado Mayorista.
- [13] Radcliffe, N. J. (2007). Generating Incremental Sales: Maximizing the Incremental Impact of Cross-Selling, Up-Selling and Deep-Selling Through Uplift Modelling. Stochastic Solutions Limited.

- [14] Montoya R. (2014). Apuntes IN4601: Marketing I[Diapositivas]. FCFM. Departamento de Ingeniería Civil Industrial. Universidad de Chile.
- [15] Fernández, P. (1996). Determinación del tamaño muestral. *Cad Aten Primaria*, 3, 138-141.
- [16] Damon Jr, R. A., & Harvey, W. R. (1987). *Experimental design, ANOVA, and regression* (No. 311.2 D163). Harper&Row.
- [17] Villafranca, R. R., & Ramajo, L. R. Z. (2013). *Métodos estadísticos para Ingenieros*. Editorial Universitat Politècnica de València.
- [18] Ryan Zhao, (2012). *Improve Marketing Campaign ROI using Uplift Modeling*. Analytics Resourcing Centre.
- [19] Rzepakowski, P., & Jaroszewicz, S. (2012). Uplift modeling in direct marketing. *Journal of Telecommunications and Information Technology*, 43-50.
- [20] Rzepakowski, P., & Jaroszewicz, S. (2012). Decision trees for uplift modeling with single and multiple treatments. *Knowledge and Information Systems*, 32(2), 303-327.
- [21] Peralta, F. (2016). *Desarrollo y evaluación de programa de incentivos para aumentar fidelización en clientes en cuponera online*.

Anexos

Promulgación Ley Libre Competencia (Oficio 12.763)

Proyecto de ley que permite la emisión de medios de pago con provisión de fondos por entidades no bancarias

Sernac Financiero (Ley N°20.555)

Ley de Protección de la Vida Privada (Ley N°19.812)

Circular N°40 para Emisores y Operadores de Tarjetas de Crédito

Ley Tasa Máxima Convencional (Ley N° 18.010)

Ley de Cobranzas extrajudiciales (Ley N° 19.659)

Ley de Sociedades Anónimas (Ley N° 18.046)

Ley del Consumidor (Ley N° 19.496)

Ley de Principio de la Finalidad del Dato (Ley Dicom -Ley N° 20.575)

Ley de Protección de la Vida Privada (Ley N° 19.628)

Ley de quiebras personales (Ley N°20720)

Ley Harboe (Ley N°20575)

Acuerdo de Autoregulación Mejores Prácticas en Comercialización de Seguros.

Capítulo III.J.1 del Compendio de Normas Financieras (Banco Central).

Ley N° 20.552 Que Moderniza y Fomenta la Competencia del Sistema Financiero.

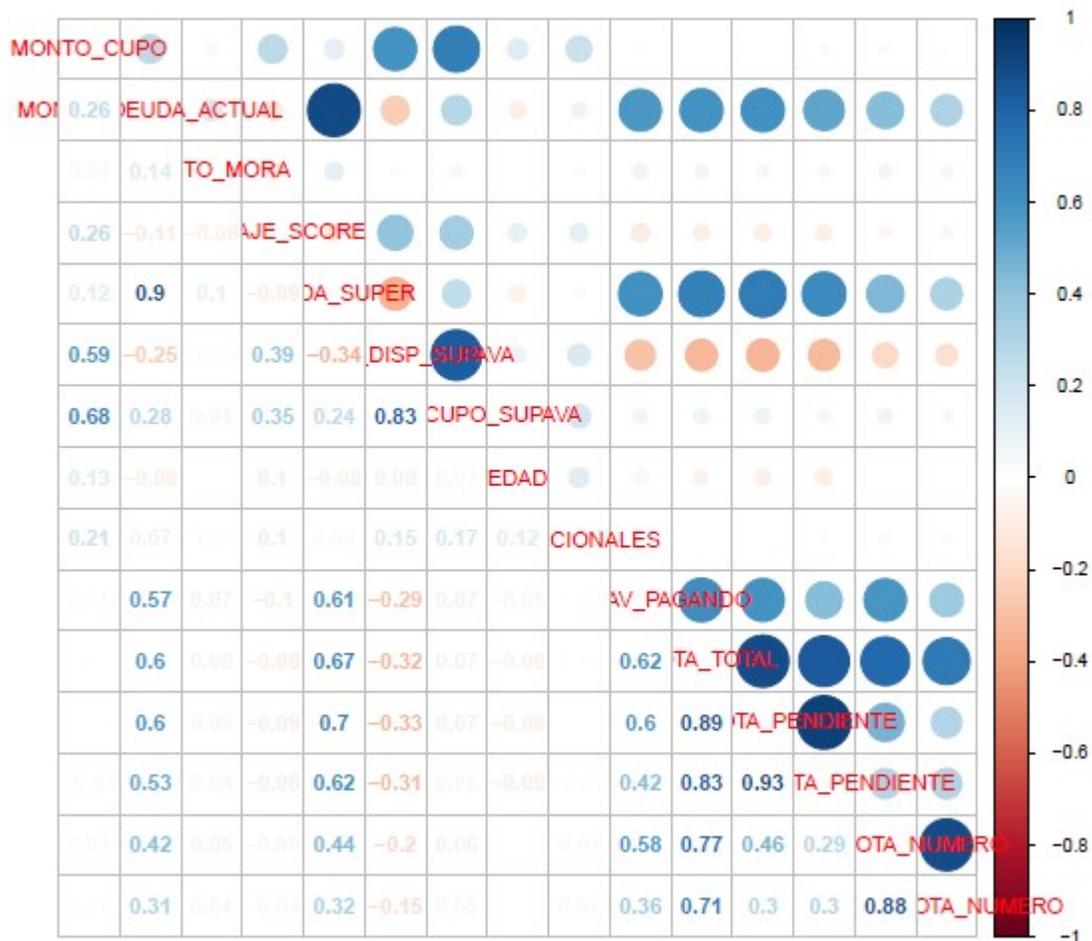
Ley 20.454 Que Incentiva el Precontrato.

Ley de Crédito Universal (Ley N°20.448)

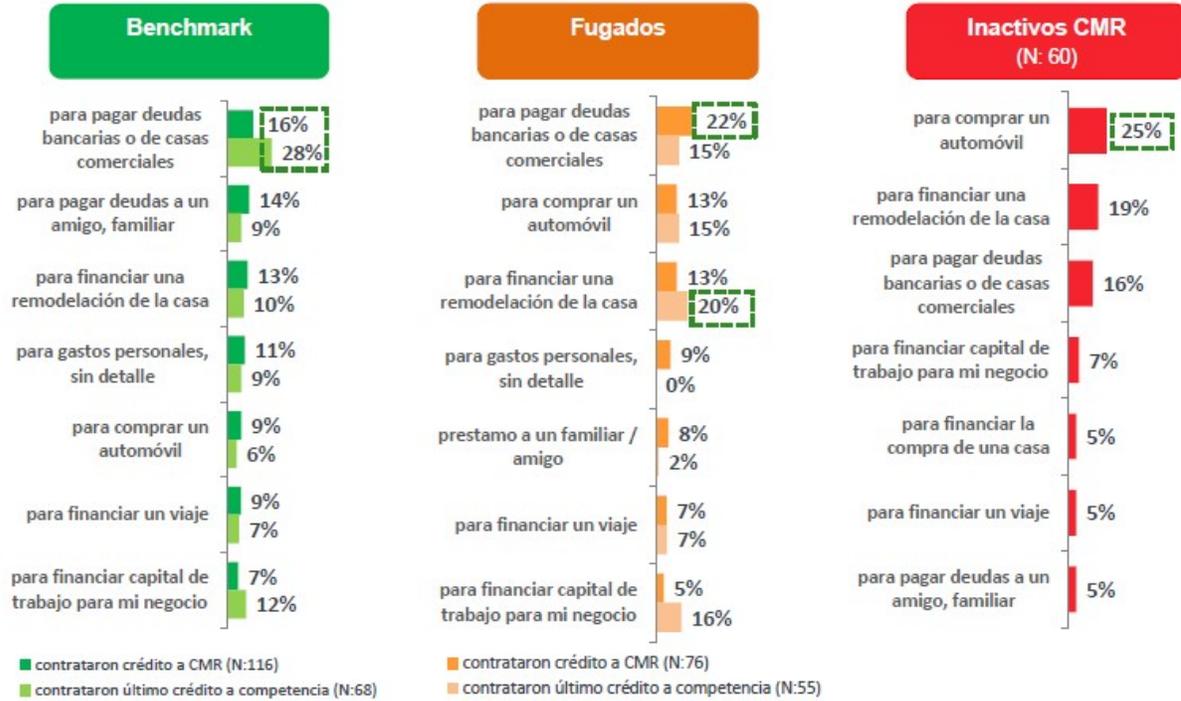
Anexo 1: tabla con principales leyes y normativas que conforman el marco regulatorio en el que se desenvuelven las empresas del retail financiero.

Coefficients:	Estimate	Std. Error	z value	Pr(> z)	Coefficients:	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.67e+00	1.59e-01	10.48	< 2e-16 ***	MONTO_CUPO_SUPAVA	-4.16e-07	1.91e-08	-21.76	< 2e-16 ***
ID_AGNO_MES201606	2.56e-01	7.10e-02	3.61	0.00031 ***	ID_PREMIUM1	4.39e-02	4.44e-02	0.99	0.32336
ID_AGNO_MES201607	4.46e-01	6.98e-02	6.39	1.6e-10 ***	ID_PREMIUM4	-2.12e-01	1.47e-01	-1.45	0.14759
ID_AGNO_MES201608	4.05e-01	6.97e-02	5.82	5.9e-09 ***	CONTACTABLE1	1.28e-01	3.34e-02	3.83	0.00013 ***
ID_AGNO_MES201609	2.73e-01	7.13e-02	3.83	0.00013 ***	RANGO_EDADBAJA	6.65e-02	4.30e-02	1.55	0.12166
ID_AGNO_MES201610	4.00e-01	6.99e-02	5.72	1.0e-08 ***	RANGO_EDADMEDIA	-9.57e-03	3.39e-02	-0.28	0.77746
ID_AGNO_MES201611	4.95e-01	6.95e-02	7.11	1.1e-12 ***	ID_SITUACTA1	-7.21e-02	1.28e-01	-0.56	0.57401
ID_AGNO_MES201612	5.17e-01	6.99e-02	7.41	1.3e-13 ***	ID_SITUACTA2	-1.38e+00	1.43e-01	-9.63	< 2e-16 ***
ID_AGNO_MES201701	4.17e-01	7.03e-02	5.94	2.9e-09 ***	GSEC2	2.32e-01	8.39e-02	2.76	0.00576 **
ID_AGNO_MES201702	3.05e-01	7.15e-02	4.27	2.0e-05 ***	GSEC3	2.37e-01	8.58e-02	2.77	0.00569 **
ID_AGNO_MES201703	1.89e-01	7.25e-02	2.61	0.00916 **	GSED	2.99e-01	7.97e-02	3.76	0.00017 ***
ID_AGNO_MES201704	1.50e-01	7.33e-02	2.05	0.04032 *	GSEE	3.84e-01	1.28e-01	2.99	0.00281 **
MONTO_CUPO	-6.72e-07	2.71e-08	-24.77	< 2e-16 ***	GSEOTRO	1.97e-01	8.03e-02	2.45	0.01443 *
MONTO_DEUDA_ACTUAL	6.87e-07	2.76e-08	24.86	< 2e-16 ***	NADICIONALES	-5.10e-03	3.13e-02	-0.16	0.87047
MONTO_MORA	-2.12e-06	4.76e-07	-4.45	8.5e-06 ***	NUM_SAV_PAGANDO	1.80e-01	1.67e-02	10.81	< 2e-16 ***
PUNTAJE_SCORE	-8.79e-04	9.78e-05	-8.99	< 2e-16 ***	PAGANDO_SAV1	-3.54e+00	6.74e-02	-52.45	< 2e-16 ***
ID_DMC_ANTIGUEDAD3	-1.08e-01	9.54e-02	-1.13	0.25886	PRODUCTO_ST2	3.92e-01	5.23e-02	7.48	7.2e-14 ***
ID_DMC_ANTIGUEDAD4	-1.82e-01	9.09e-02	-2.01	0.04458 *	PRODUCTO_ST6	6.69e-01	8.46e-02	7.91	2.5e-15 ***
ID_DMC_ANTIGUEDAD5	-3.40e-01	9.09e-02	-3.74	0.00018 ***	PRODUCTO_ST7	5.65e-01	1.81e-01	3.13	0.00174 **
ID_DMC_ANTIGUEDAD6	-2.47e-01	9.10e-02	-2.71	0.00663 **	PRODUCTO_ST8	7.63e-01	5.93e-02	12.88	< 2e-16 ***
ID_DMC_ANTIGUEDAD7	-3.11e-01	8.79e-02	-3.54	0.00039 ***	RATIOMAXFINAL	3.46e-01	4.20e-02	8.24	< 2e-16 ***
ID_DMC_ANTIGUEDAD8	-2.40e-01	9.52e-02	-2.52	0.01166 *	RATIOMAXINICIO	-3.06e-01	4.30e-02	-7.12	1.1e-12 ***
ID_DMC_ANTIGUEDAD9	-3.09e-01	1.01e-01	-3.07	0.00217 **	RATIOMAXNO PAGA	-5.54e-01	4.82e-02	-11.49	< 2e-16 ***
ID_DMC_ANTIGUEDAD10	-1.66e-01	1.06e-01	-1.57	0.11644	---				
DEUDA_SUPER	8.92e-07	3.68e-08	24.20	< 2e-16 ***	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Anexo 2: Resumen modelo logit utilizando software R. Elaboración propia.



Anexo 3: Matriz de correlaciones modelo predictivo convencional. Elaboración propia utilizando R.



Q15 / Q27.- De las razones que me acaba de mencionar, ¿cuál diría usted que fue la principal?

Anexo 4: Resultados encuesta telefónica realizada por research sobre las principales razones de uso de un crédito de consumo.

Nombre	Descripción
Edad	Edad del cliente al momento de recibir la campaña.
Género	Género del cliente.
GSE	Grupo socioeconómico del cliente según dirección declarada.
Id cliente	Identificador del cliente.

Anexo 5: Descripción variables sociodemográficas del cliente. Elaboración propia.

Nombre	Descripción
Categoría	Corresponde a la categoría del cliente. Puede ser normal, premium o elite.
Nivel de canje	Máximo nivel de canje del cliente.
Puntos acumulados	Considera el último año calendario. Para este caso fue el año 2016.
Días con compra	Cantidad de días con compra en los últimos 3 meses.
Monto deuda actual	Deuda que posee el cliente al momento de la campaña.
Saturación cupo	Variación porcentual promedio del cupo utilizado en los últimos 3/2/1 meses.
Puntaje score	Puntaje generado por el área de riesgo, entre más alto menos riesgoso es el cliente.
Gasto	Variación porcentual promedio del gasto en el rubro de mejoramiento del hogar, automotriz, viajes, vivienda, salud y educación.
Cambio de cupo	Binaria: 1 si el cliente aumentó su cupo en los últimos 3/6 meses.
Aumento de cupo	Si hubo aumento de cupo, indica la cantidad.
Productos	Binaria que indica si el cliente posee algún otro producto: PAT, Bip, Avance en efectivo, etc.
Con producto	Binaria que indica si se contrato algún producto en los últimos 3/6 meses.
Cambio de producto pr	Binaria: 1 si el cliente cambió de producto en los últimos 3/6 meses.
Adicionales	Binaria: 1 si el cliente posee adicionales al momento del envío de la campaña.
Reclamo	Binaria: 1 si el cliente ha realizado algún reclamo en los últimos 3/6 meses.

Anexo 6: Descripción variables de comportamiento. Elaboración propia.

Nombre	Descripción
Participación On Us	Indicador que muestra el % del gasto en los últimos 3/6 meses con la tarjeta en las empresas pertenecientes al holding.
Enganchado	Variable categórica que indica por qué esta enganchado un cliente: Club de puntos, financiamiento o promociones. Puede no estarlo.
SOW	Corresponde al share of wallet: existe un indicador para la empresa y uno por rubro, algunos de los rubros son: tiendas por departamento, automotriz, combustible, mejoramiento del hogar, farmacias, supermercados entre otros.

Anexo 7: Descripción variables de fidelidad. Elaboración propia.

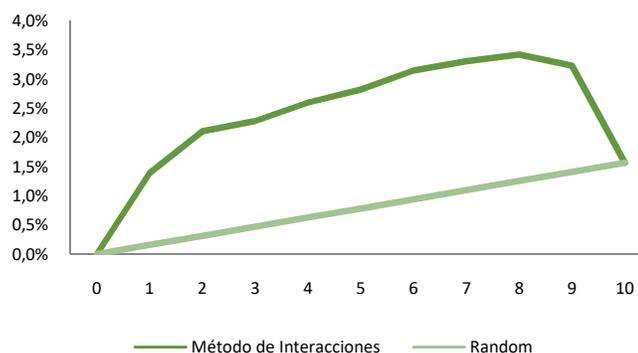
Nombre	Descripción
Pagando Sav	Binaria: 1 si el cliente está pagando un Sav en el momento del envío de la campaña.
Deuda Sav	Deuda del cliente debido a un súper avance.
Cuota pagando	Indicador entre 0 y 1 que indica si esta pagando cuotas iniciales, intermedias o finales. Puede no estar pagando cuotas.
Monto deuda	Deuda del cliente debido a un súper avance.
Cliente nuevo	Binaria: 1 si el cliente nunca ha sacado un PF.
Recency último súper avance	Cantidad de meses desde cuando saco el último súper avance.

Anexo 8: Descripción variables de conocimiento del producto. Elaboración propia.

Nombre	Descripción
Contrato	Número del contrato del cliente. Se puede tener más de uno, ya que aparte de tener el producto principal, se puede tener la Mastercard.
Antigüedad	Antigüedad en años del cliente.
Producto Principal	Tarjeta principal que posee el cliente, ya que este también puede tener la mastercard.
Cliente nuevo disponible	Binaria: 1 si el cliente tiene disponible Sav hace menos de 3/6 meses.
Monto cupo	Monto de cupo que tiene el cliente en la tarjeta.

Anexo 9: Descripción variables de entorno. Elaboración propia.

Curva de ganancia de información

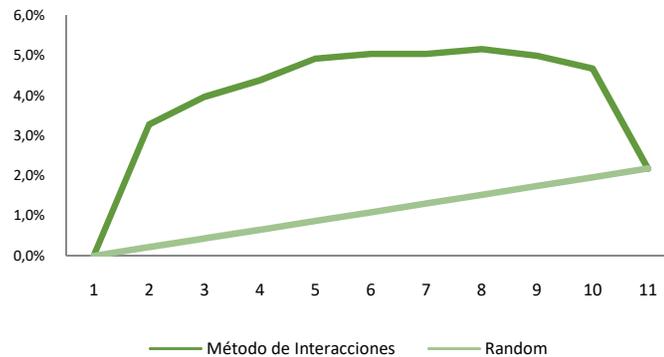


Anexo 10: Curva e ganancia de información modelo campañas genéricas. Elaboración propia utilizando R.

	Divergencia Kullback-Leibler	Distancia Euclidiana	Divergencia Chi-Cuadrado	Método de Interacciones
Q de Qini	2,47%	2,81%	2,34%	3,01%

Anexo 11: Q de Qini para cada uno de los criterios de divergencia utilizados para la confección del modelo de campañas genéricas. Elaboración propia utilizando R.

Curva de ganancia de información



Anexo 12: Curva de ganancia de información modelo campaña giftcard. Elaboración propia utilizando R.

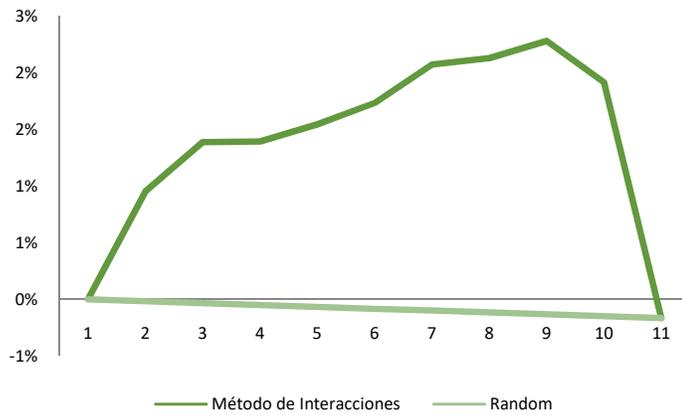
Decil	N° clientes tratados (GT)	N° clientes no tratados (GC)	N° respuestas GT	N° respuestas GC	TR GT	TR GC	Uplift
1	709	520	308	-	43,4%	0,0%	43,4%
2	560	669	63	1	11,3%	0,1%	11,1%
3	539	690	19	-	3,5%	0,0%	3,5%
4	562	666	8	-	1,4%	0,0%	1,4%
5	624	605	7	-	1,1%	0,0%	1,1%
6	610	619	6	-	1,0%	0,0%	1,0%
7	653	575	4	3	0,6%	0,5%	0,1%
8	724	505	4	2	0,6%	0,4%	0,2%
9	796	433	7	9	0,9%	2,1%	-1,2%
10	797	432	2	178	0,3%	41,2%	-41,0%

Anexo 13: Tabla resumen modelo upliftgiftcardenRandomForest ordenado en deciles. Elaboración propia en R.

	Divergencia Kullback-Leibler	Distancia Euclidiana	Divergencia Chi-Cuadrado	Método de Interacciones
Q de Qini	1,40%	1,21%	1,34%	1,57%

Anexo 14: Q de Qini para cada uno de los criterios de divergencia utilizados para la confección del modelo de campañas con regalo de giftcard. Elaboración propia utilizando R.

Curva de ganancia de información



Anexo 15: Curva de ganancia de información modelo campaña puntos. Elaboración propia utilizando R.

Decil	N° clientes tratados (GT)	N° clientes no tratados (GC)	N° respuestas GT	N° respuestas GC	TR GT	TR GC	Uplift
1	2079	373	327	2	16%	1%	15%
2	2016	435	118	0	6%	0%	6%
3	2044	407	44	0	2%	0%	2%
4	2068	384	49	0	2%	0%	2%
5	2103	348	42	0	2%	0%	2%
6	2114	337	48	0	2%	0%	2%
7	2172	280	68	3	3%	1%	2%
8	2172	279	84	1	4%	0%	4%
9	2180	271	123	20	6%	7%	-2%
10	2156	296	139	153	0,064471	52%	-45%

Anexo 16: Tabla resumen modelo uplift puntos en RandomForest ordenado en deciles. Elaboración propia en R.