



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

DETERMINACIÓN DE LA PROPENSIÓN AL AUMENTO DE CONSUMO CON
TARJETA DE CRÉDITO DE CLIENTES DE UNA INSTITUCIÓN BANCARIA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

BENJAMÍN ALBERTO CALDERÓN TESTA

PROFESORA GUÍA:
ALEJANDRA PUENTE CHANDÍA

MIEMBROS DE LA COMISIÓN:
IGNACIO CALISTO LEIVA
PABLO MARÍN VICUÑA

SANTIAGO DE CHILE
2016

RESUMEN DE LA MEMORIA PARA OPTAR AL
TÍTULO DE: Ingeniero Civil Industrial
POR: Benjamín Alberto Calderón Testa
FECHA: 20/08/2016
PROFESORA GUÍA: Alejandra Puente Chandía

DETERMINACIÓN DE LA PROPENSIÓN AL AUMENTO DE CONSUMO CON
TARJETA DE CRÉDITO DE CLIENTES DE UNA INSTITUCIÓN BANCARIA

La industria bancaria se caracteriza por ofrecer a sus clientes diferentes medios de pago dentro de los que destacan las tarjetas de crédito, con las cuales se transa anualmente un monto equivalente al 7% del PIB Chileno. Por otro lado, los clientes del banco con el que se desarrolla este trabajo utilizan menos la tarjeta que los clientes de los bancos líderes y además, la rentabilidad del negocio de tarjetas está directamente ligada al consumo.

En base a lo anterior, el objetivo general de este proyecto es determinar las acciones con las cuales se puede incentivar el aumento de consumo con la tarjeta de crédito, usando información transaccional de los clientes.

Considerando esto, se busca responder dos preguntas básicas del negocio ¿a quienes incentivar? y ¿con qué acciones hacerlo? Para responder esto se comienza clasificando a los clientes según variables transaccionales, con lo cual se encuentran cinco niveles de consumo que pueden tener los clientes durante un periodo. Dentro de estos destacan los clientes de Medio Bajo consumo, los cuales representan el 54% de la cartera del banco y se caracterizan por transar principalmente en rubros básicos como supermercados o grandes tiendas.

Una vez clasificados los clientes, se realizan modelos predictivos para identificar a aquellos que aumentarán su nivel de consumo en el futuro, además de los principales factores que inducen este aumento. En base a esto se define que los clientes a incentivar son todos aquellos que no incrementarán su consumo, pues los otros lo harán con alta probabilidad sin necesidad de ser incentivados.

Con respecto a las acciones, los resultados indican que se debe incentivar a los clientes a ampliar la variedad de rubros en que consumen, además de la cantidad de semanas en que tienen transacciones. Por otro lado, el banco debe incentivar los avances en efectivo, las compras internacionales y algunos rubros como supermercados o grandes tiendas.

Se destaca además el hecho de que, al aplicar la metodología propuesta en este trabajo, el banco podría ganar 51 millones de pesos en un año, dejando de enviar promociones a clientes que no lo necesitan y generando nuevos aumentos de consumo.

Finalmente como trabajo futuro se propone complementar el estudio con información de las promociones del banco, para así analizar además la efectividad de las campañas y determinar si aumentan el consumo de los clientes a largo plazo o sólo en el periodo en que se realizan.

Dedicada al Igo.

Agradecimientos

En este momento en que culmina una de las etapas más importantes de mi vida, me gustaría agradecer a todas las personas que hicieron de mi vida universitaria un periodo lleno de alegrías y nuevas experiencias.

Quiero agradecer en primer lugar a mi familia, a mis padres y hermanos que siempre estuvieron presentes y me apoyaron en todo lo necesario para lograr mis objetivos. Agradezco además la enorme preocupación durante el periodo más difícil de la memoria, en donde siempre me sentí apoyado.

Me gustaría agradecer también a mi polola Nicole, que nunca se cansó de darme ánimo y confiar en mí, incluso cuando yo mismo no era capaz de ver el rumbo. Creo que no hubiese llegado a este momento sin sus consejos, su insistencia por que tomara vacaciones y su apoyo incondicional.

A los profesores de mi comisión, Alejandra Puente, Ignacio Calisto y Pablo Marín, muchísimas gracias por tener siempre la mejor disposición para reunirse conmigo e interesarse por que este trabajo fuera mejor. Quiero agradecer y reconocer también al profesor Marcel Goic, quien sin tener ninguna responsabilidad con este trabajo, se tomó el tiempo de orientarme en varias ocasiones.

A la gente de Innovapay, Sergio, Rodrigo y Fernando, muchas gracias por confiar en mí, por guiarme y por apoyarme en los momentos más difíciles de este proceso.

Finalmente quiero agradecer a mis amigos y a todos los que de alguna u otra forma estuvieron presente en esta etapa.

¡A todos ustedes Muchas gracias!

Tabla de contenido

| | | |
|----------|--|----|
| 1. | Antecedentes generales | 1 |
| 1.1. | Industria Bancaria chilena | 1 |
| 1.2. | Medios de pago en la industria bancaria..... | 1 |
| 1.3. | Tarjetas de crédito | 3 |
| 2. | Descripción del proyecto y justificación | 5 |
| 3. | Objetivos y alcances de proyecto..... | 8 |
| 3.1. | Objetivo general..... | 8 |
| 3.2. | Objetivos específicos..... | 8 |
| 3.3. | Alcances del proyecto | 8 |
| 4. | Marco conceptual | 9 |
| 4.1. | Proceso KDD | 9 |
| 4.2. | Segmentación y clusterización | 10 |
| 4.3. | Variables RFM | 11 |
| 4.4. | Regresión logística..... | 12 |
| 4.5. | Árboles de decisión | 13 |
| 4.6. | Análisis ROC | 13 |
| 5. | Metodología..... | 15 |
| 6. | Desarrollo de la metodología | 17 |
| 6.1. | Clasificación de clientes por niveles de consumo | 17 |
| 6.1.1. | Definición de los periodos de estudio..... | 17 |
| 6.1.2. | Selección de datos | 17 |
| 6.1.3. | Transformación de variables | 18 |
| 6.1.3. | Pre procesamiento | 21 |
| 6.1.4. | Minería de datos | 22 |
| 6.1.5. | Análisis de resultados | 24 |
| 6.1.5.1. | Caracterización clientes inactivos..... | 24 |
| 6.1.5.2. | Caracterización segmento Bajo consumo | 25 |
| 6.1.5.3. | Caracterización segmento Medio Bajo consumo..... | 26 |
| 6.1.5.4. | Caracterización segmento Medio Alto consumo | 28 |
| 6.1.5.5. | Caracterización segmento Alto consumo..... | 30 |
| 6.2. | Modelos de propensión | 34 |
| 6.2.1. | Selección de datos y pre procesamiento..... | 34 |
| 6.2.2. | Transformación de variables | 34 |

| | | |
|----------|---|----|
| 6.2.3. | Minería de datos | 35 |
| 6.2.3.1. | Modelo de propensión clientes Inactivos | 36 |
| 6.2.3.2. | Modelo de propensión clientes Bajo consumo | 39 |
| 6.2.3.3. | Modelo de propensión clientes Medio Bajo consumo..... | 44 |
| 6.2.4. | Análisis de resultados | 50 |
| 6.2.4.1. | Recomendaciones clientes Inactivos | 52 |
| 6.2.4.2. | Recomendaciones clientes Bajo consumo | 53 |
| 6.2.4.3. | Recomendaciones clientes Medio Bajo consumo..... | 54 |
| 7. | Evaluación económica | 56 |
| 7.1. | Evaluación económica modelo Inactivos | 57 |
| 7.2. | Evaluación económica modelo Bajo consumo | 59 |
| 7.3. | Evaluación económica modelo Medio Bajo consumo | 62 |
| 8. | Conclusiones..... | 65 |
| 8.1. | Conclusiones del proyecto..... | 65 |
| 8.2. | Limitaciones y trabajos futuros..... | 66 |
| 9. | Bibliografía | 68 |
| 10. | Anexos | 70 |

Índice de Tablas

| | |
|---|----|
| Tabla 1: Crecimiento anual (CAGR) medios de pago 2000-2011..... | 2 |
| Tabla 2: Penetración de tarjetas 2012..... | 2 |
| Tabla 3: Periodos de estudio | 17 |
| Tabla 4: Coeficientes de correlación entre variables RFM | 20 |
| Tabla 5: Variables RFM por clúster obtenido en Periodo 6 | 23 |
| Tabla 6: Niveles de consumo periodo 6 | 24 |
| Tabla 7: Distribución tipos de pago clientes Bajo consumo | 26 |
| Tabla 8: Estructura de consumo por rubros clientes Bajo consumo | 26 |
| Tabla 9: Distribución tipos de pago clientes Medio Bajo consumo | 27 |
| Tabla 10: Estructura de consumo por rubros clientes Medio Bajo consumo | 28 |
| Tabla 11: Distribución tipos de pago clientes Medio Alto consumo..... | 29 |
| Tabla 12: Estructura de consumo por rubros clientes Medio Alto consumo | 30 |
| Tabla 13: Distribución tipo de pago clientes Alto consumo | 31 |
| Tabla 14: Estructura de consumo por rubros clientes Alto consumo | 32 |
| Tabla 15: Resumen información transaccional..... | 32 |
| Tabla 16: Indicadores ROC clientes Inactivos | 38 |
| Tabla 17: Indicadores principales análisis ROC clientes Bajo consumo | 42 |
| Tabla 18: Variables principales regresión logística clientes Bajo consumo | 43 |
| Tabla 19: Indicadores principales análisis ROC clientes Medio Bajo consumo | 47 |
| Tabla 20: Principales variables modelo Medio Bajo consumo | 49 |
| Tabla 21: Tasa de respuesta según tipo de promoción clientes Inactivos..... | 57 |
| Tabla 22: Evaluación económica clientes Inactivos escenario Realista..... | 58 |
| Tabla 23: Tasas de respuesta a promociones escenarios Pesimista y Optimista clientes Inactivos | 59 |
| Tabla 24: Cantidad de clientes Bajo consumo a los que se les envía cada combinación promocional | 59 |
| Tabla 25: Tasas de respuesta según tipo de promoción escenario realista clientes Bajo consumo | 60 |
| Tabla 26: Respuesta a promociones clientes Bajo consumo | 60 |
| Tabla 27: Evaluación Económica clientes Bajo consumo escenario Realista | 61 |
| Tabla 28: Tasas de respuesta a promociones escenarios Pesimista y Optimista clientes Bajo consumo | 61 |
| Tabla 29: Cantidad de clientes Medio Bajo consumo a los que se les envía cada combinación promocional | 62 |
| Tabla 30: Tasas de respuesta según tipo de promoción escenario Realista clientes Medio Bajo consumo | 62 |
| Tabla 31: Respuesta a promociones clientes Medio Bajo consumo | 63 |
| Tabla 32: Evaluación Económica clientes Medio Bajo consumo escenario Realista..... | 63 |
| Tabla 33: Tasas de respuesta a promociones escenarios Pesimista y Optimista clientes Medio Bajo consumo | 64 |

Índice de Ilustraciones

| | |
|---|----|
| Ilustración 1: Participación en colocaciones por banco 2015 | 1 |
| Ilustración 2: Porcentaje de montos transados con tarjeta de crédito octubre 2015 | 3 |
| Ilustración 3: Monto promedio transado por tarjetas activas entre enero 2015 y octubre 2015 | 5 |
| Ilustración 4: Monto promedio transado por tarjetas de débito entre enero 2015 y octubre 2015 | 6 |
| Ilustración 5: Rentabilidad de los clientes del Banco según decil de consumo | 6 |
| Ilustración 6: Proceso KDD..... | 9 |
| Ilustración 7: Diagrama de posibles valores de la variable Aumento | 16 |
| Ilustración 8: Histograma variable Recency para clientes con transacciones Periodo 6 .. | 19 |
| Ilustración 9: Histograma variable Frequency para clientes con transacciones Periodo 6 | 19 |
| Ilustración 10: Histograma variable Monetary Value para clientes con transacciones Periodo 6 | 20 |
| Ilustración 11: Gráfico de dispersión variables F y M periodo 6 | 21 |
| Ilustración 12: Estadísticos principales variables RFM sin outliers periodo 6 | 22 |
| Ilustración 13: Distancia total vs número de clusters..... | 23 |
| Ilustración 14: Información de cuenta clientes Inactivos | 24 |
| Ilustración 15: Información de cuenta clientes Bajo consumo | 25 |
| Ilustración 16: Información de cuentas clientes Medio Bajo consumo | 27 |
| Ilustración 17: Información de cuentas clientes Medio Alto consumo | 29 |
| Ilustración 18: Información de cuentas clientes Alto consumo | 31 |
| Ilustración 19: Distribución de consumo por nivel | 33 |
| Ilustración 20: Compras internacionales clientes Inactivos | 36 |
| Ilustración 21: Porcentaje de clientes según comportamiento clientes Inactivos | 37 |
| Ilustración 22: Curva ROC regresión logística clientes Inactivos | 37 |
| Ilustración 23: Semanas con transacciones clientes Bajo consumo | 40 |
| Ilustración 24: Revolving promedio mensual periodo 1 Clientes Bajo consumo | 40 |
| Ilustración 25: Porcentaje de clientes con transacciones en tipos importantes de transacción | 41 |
| Ilustración 26: Curvas ROC clientes Bajo consumo | 41 |
| Ilustración 27: Sensibilidad vs. Especificidad, encontrando el mejor punto de corte..... | 43 |
| Ilustración 28: Cantidad de semanas con transacciones clientes Medio Bajo consumo.. | 45 |
| Ilustración 29: Cantidad de rubros en los que consumen los clientes de Medio Bajo consumo | 45 |
| Ilustración 30: Porcentaje de clientes con transacciones en principales rubros | 46 |
| Ilustración 31: Curvas ROC modelos Medio Bajo consumo | 47 |
| Ilustración 32: Sensibilidad vs. Especificidad, encontrando el mejor punto de corte | 48 |

1. Antecedentes generales

1.1. Industria Bancaria chilena

Durante el año 2013 la industria bancaria aportó el 3,9% del PIB chileno, siendo el mayor aporte de esta industria desde 2009 [1]. Por otro lado, el crecimiento de la industria entre septiembre de 2014 y el mismo mes de 2015 fue de un 5,8% [2], alcanzando un retorno sobre el patrimonio de casi un 15%.

De lo anterior se puede observar la alta actividad de la industria bancaria chilena, la cual es la segunda más activa de América latina [2]. A pesar de lo anterior y de contar con más de 19 instituciones activas, es un mercado concentrado en el que 5 bancos cuentan con más del 73% de las colocaciones:

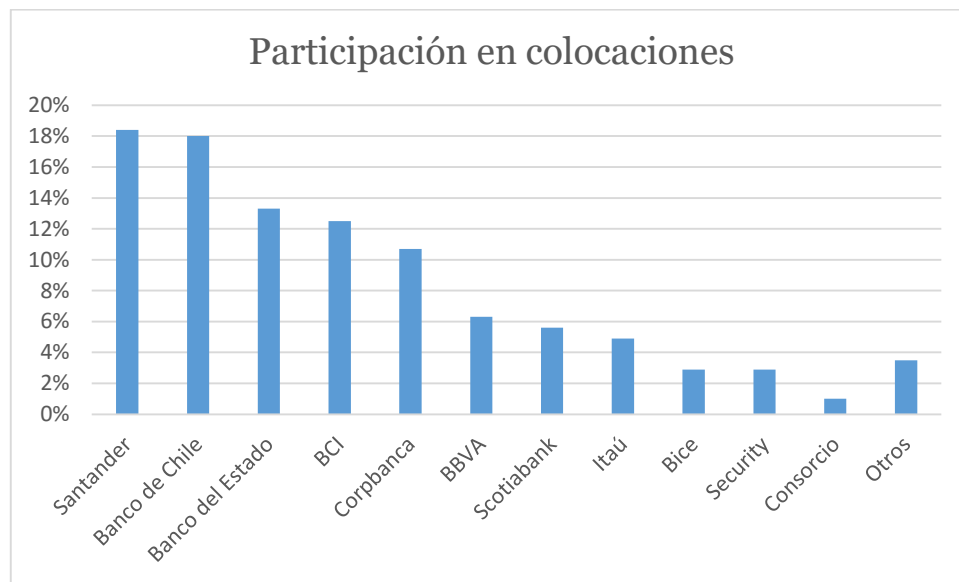


Ilustración 1: Participación en colocaciones por banco 2015
Fuente: Elaboración propia con datos de la SBIF

Por otro lado, dentro de la industria existen diferentes productos que se ofrecen a los clientes, como por ejemplo los créditos o las tarjetas de débito. Estos productos se incluyen dentro de las distintas áreas de negocio de los bancos y en este trabajo se trabaja exclusivamente con el área de medios de pago.

1.2. Medios de pago en la industria bancaria

Dentro de la industria bancaria chilena, una de las áreas más importantes es la de medios de pago, la que cuenta con productos que permiten a las personas pagar sin la necesidad de utilizar dinero en efectivo. Algunos de los principales productos dentro de los medios de pago son las transferencias electrónicas, los cheques, las tarjetas de débito y las tarjetas de crédito.

Por otro lado, como se puede apreciar en la siguiente tabla, el crecimiento anual compuesto (CAGR) de los diferentes medios de pago ha sido diferente en los últimos años [3]:

| Medio de pago | Variación del uso del instrumento |
|------------------------------------|--|
| <i>Transferencias electrónicas</i> | 33% |
| <i>Cheques</i> | -4% |
| <i>Tarjetas de débito</i> | 54% |
| <i>Tarjetas de crédito</i> | 10% |

Tabla 1: Crecimiento anual (CAGR) medios de pago 2000-2011
Fuente: Elaboración propia con datos del ministerio de hacienda

De lo anterior se puede observar que los cheques son un medio de pago cada vez menos utilizado, mientras que las otras opciones se han ido haciendo más importantes para los consumidores. Por otro lado, el aumento de las transferencias electrónicas demuestra que las personas están más dispuestas a utilizar métodos electrónicos para realizar pagos o depositar dinero.

Con respecto a las tarjetas de débito y crédito, estas tienen un funcionamiento completamente diferente entre ellas. Mientras que con la primera se pueden realizar pagos siempre que existan fondos en la cuenta, la tarjeta de crédito permite al consumidor realizar transacciones incluso cuando no tenga dinero para hacerlo y por lo tanto, otorgan mayor libertad al usuario.

Por otro lado, si bien en Chile ambas tarjetas tienen niveles de penetración similares, con respecto a Latinoamérica y el mundo, la penetración en Chile de la tarjeta de crédito es mayor [3], como se ve en la siguiente tabla:

| | Población mayor de 15 años con tarjeta de débito | Población mayor de 15 años con tarjeta de crédito |
|-----------------------|---|--|
| <i>Chile</i> | 25% | 22% |
| <i>América latina</i> | 28% | 18% |
| <i>Mundo</i> | 20% | 14% |

Tabla 2: Penetración de tarjetas 2012
Fuente: Elaboración propia con datos del ministerio de hacienda

Al mes de octubre de 2015 existían más de 20 millones de tarjetas de débito activas, las cuales tuvieron transacciones por 900 mil millones de pesos¹ durante ese mes [5]. Por otro lado, en el mismo mes existían más de 12 millones de tarjetas de crédito con operaciones totales superiores a 1 billón de pesos durante octubre [4].

Si bien ambos son productos en crecimiento y con un alto nivel de transaccionalidad, existe una gran diferencia en el modelo de negocio de ambas. Mientras la tarjeta de débito deja ingresos al banco por concepto de comisiones a los comercios en los que se usa, el modelo de la tarjeta de crédito es más complejo pues incluye diferentes fuentes de ingreso. Este trabajo se centra en la tarjeta de crédito, por lo que a continuación se presentan datos relevantes y se detalla su modelo de negocios.

¹ Monto no incluye retiros de dinero en cajeros

1.3. Tarjetas de crédito

En el mercado de las tarjetas de crédito existen actualmente 19 emisores, es decir, instituciones facultadas para ofrecer tarjetas, de las cuales 5 concentran el 80% de los montos transados durante el mes de octubre de 2015 [4]. Esta situación se repite a la hora de evaluar la cantidad de transacciones realizadas durante ese mes.

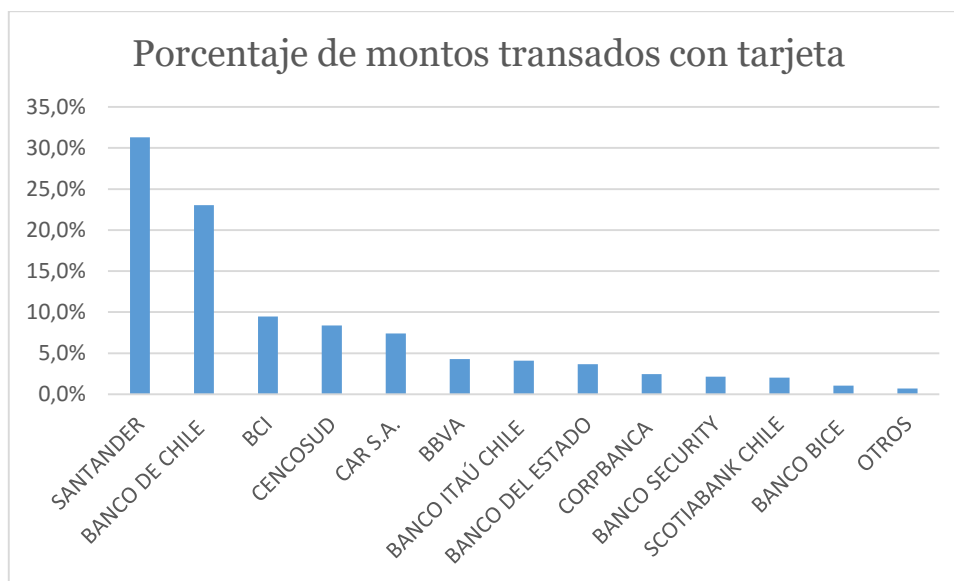


Ilustración 2: Porcentaje de montos transados con tarjeta de crédito octubre 2015
Fuente: Elaboración propia con datos de la SBIF

Como se puede ver en el gráfico, existen dos bancos que tienen sobre el 50% de los montos transados con tarjetas y por lo tanto es importante para el resto de los bancos generar acciones para incentivar el uso de la tarjeta por parte de sus clientes.

Si se analiza el monto de las transacciones realizadas con tarjeta de crédito durante el año 2015, se puede notar que este supera el 7% del PIB de Chile, sin embargo, sólo durante el mes de octubre de ese año más del 65% de las tarjetas emitidas no tuvo actividad. Si este comportamiento se analiza en un periodo de cuatro meses, el número se acerca al 50%.

Lo anterior demuestra que existe en la industria cerca de un millón de tarjetas pertenecientes a clientes activos del banco, pero que sin embargo no utilizan su tarjeta todos los meses y por lo tanto podrían aumentar su uso.

Como se mencionó en la parte 1.2, es el modelo de negocios lo que hace especial a la tarjeta de crédito, ya que genera ingresos para el banco a través de 5 fuentes diferentes:

- 1) En primer lugar, el cliente debe pagar un costo de mantención en cada mes en que utilice la tarjeta. Este cobro es independiente de la cantidad de veces que se utilice la tarjeta en el mes.
- 2) En segundo lugar, está el cobro que se hace por los avances de efectivo, donde cada vez que un cliente retira efectivo con su tarjeta de crédito, debe pagar un cobro fijo e intereses asociados a la transacción.

- 3) En tercer lugar, se encuentra el cobro que se realiza a los comercios cuando un cliente paga con tarjeta. Este cobro es variable pues representa un porcentaje de la venta.
- 4) Otro cobro importante es el de intereses, el cual se cobra cuando el cliente realiza transacciones en cuotas.
- 5) Por último, la tarjeta de crédito también genera ingresos para el banco por conceptos de revolving, que ocurre cuando un cliente no paga el total de su estado de cuenta al momento del vencimiento y por lo tanto se generan intereses sobre la deuda no paga. La tasa de interés aplicada en estos casos suele ser mayor a la tasa de pago en cuotas.

Cabe destacar que tanto los avances de efectivo como la deuda de revolving generan intereses desde el momento en que comienzan, a diferencia de la deuda común que no genera intereses si se paga al momento del vencimiento.

Como ya se ha visto, las tarjetas de crédito son un producto importante dentro de la industria bancaria y es quizás por eso que existen un gran número de trabajos relacionados a ella, donde los principales se dividen en tres categorías:

- Segmentación y clusterización: En estos trabajos se busca encontrar a grupos de clientes con características similares. Su utilidad radica en que para los bancos es más efectivo separar a los clientes para generar acciones diferentes en cada grupo. Además, se basan en que grupos con características similares deberían tener comportamientos similares y por lo tanto reaccionar a los mismos estímulos.
- Predicción de fuga y retención: En este tipo de trabajos se busca encontrar a los clientes con alta probabilidad de dejar de utilizar la tarjeta para siempre. Además buscan encontrar formas de retener a esos clientes para que no abandonen el banco o el uso de la tarjeta.
- Incentivo del uso de la tarjeta: La idea principal de estos trabajos es determinar en qué clientes se debe incentivar el uso de la tarjeta de crédito y de qué manera se puede hacer esto.

El presente trabajo incluye tanto modelos de segmentación y clusterización como modelos para incentivar el consumo mediante tarjetas de crédito por parte de los clientes de un banco privado y se lleva a cabo en conjunto con la consultora Innovapay, la cual trabaja con el banco y por lo tanto se tienen datos a nivel de cuenta, cliente y transacción realizada con la tarjeta.

Cabe destacar que si bien la consultora se fundó en 2015, sus socios tienen más de treinta años de experiencia en el rubro bancario y por lo tanto, son la fuente principal de conocimiento experto utilizada en este trabajo.

2. Descripción del proyecto y justificación

Como se mencionó en la sección anterior, existe un 15% de tarjetas en la industria bancaria que pertenecen a clientes que si bien están activos (no fugados), no tienen transacciones en un mes específico. Esta situación lleva a plantear la pregunta ¿Cómo se puede hacer para que esos clientes realicen transacciones?

Por otro lado, actualmente en el banco en que se desarrolla el proyecto se envían las mismas acciones y promociones a todos los clientes de la cartera, independiente de su nivel de actividad o la forma en la que usan la tarjeta comúnmente. Debido a esto ocurren dos problemas:

- En primer lugar, se están haciendo esfuerzos de marketing y promociones sin tener la seguridad de estar ofreciendo los incentivos correctos para aumentar el uso de la tarjeta por parte de los clientes. Además, las promociones no tienen ningún nivel de personalización y por lo tanto se le está ofreciendo el mismo estímulo a personas con intereses diferentes.
- En segundo lugar, al ofrecer las promociones a todos los clientes, se da también el caso de clientes que aumentarían de consumo sin la necesidad de la promoción, pero que al recibirla la utilizan generando gastos “innecesarios” para el banco.

Lo anterior indica que si se lograra identificar a los clientes que aumentarían su consumo, el banco podría dejar de incurrir en costos innecesarios y además determinar de mejor forma que tipo de promociones enviar.

Otro motivo que demuestra la necesidad del banco por incentivar el uso de su tarjeta por parte de los clientes es el hecho de que los montos transados con la tarjeta de crédito del banco son menos de un 15% de bancos como Santander o Banco de Chile. Se podría pensar que esta diferencia se debe a que los clientes de dichos bancos tienen un mayor ingreso, sin embargo, al analizar el monto promedio transado por tarjetas de crédito y de débito se obtiene lo siguiente:

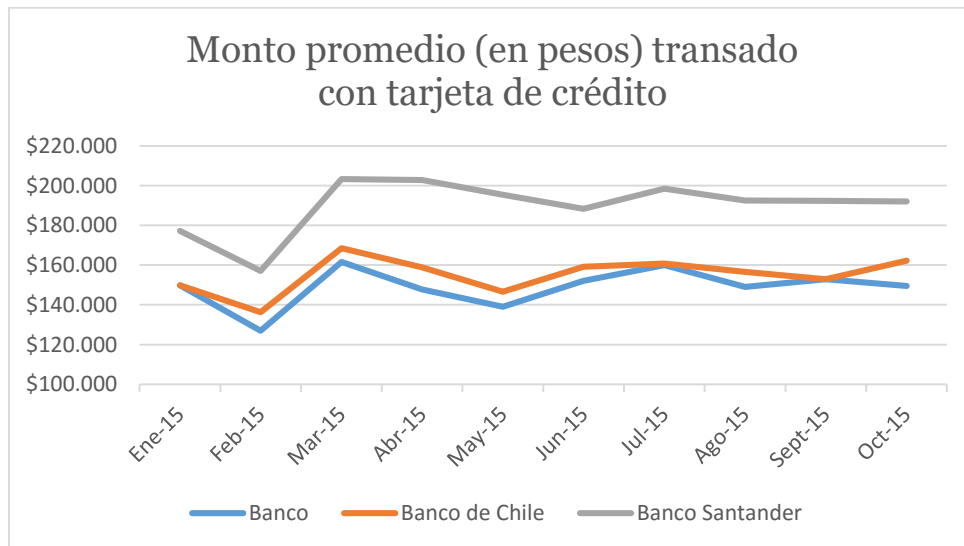


Ilustración 3: Monto promedio transado por tarjetas activas entre enero 2015 y octubre 2015
Fuente: Elaboración propia con datos de la SBIF

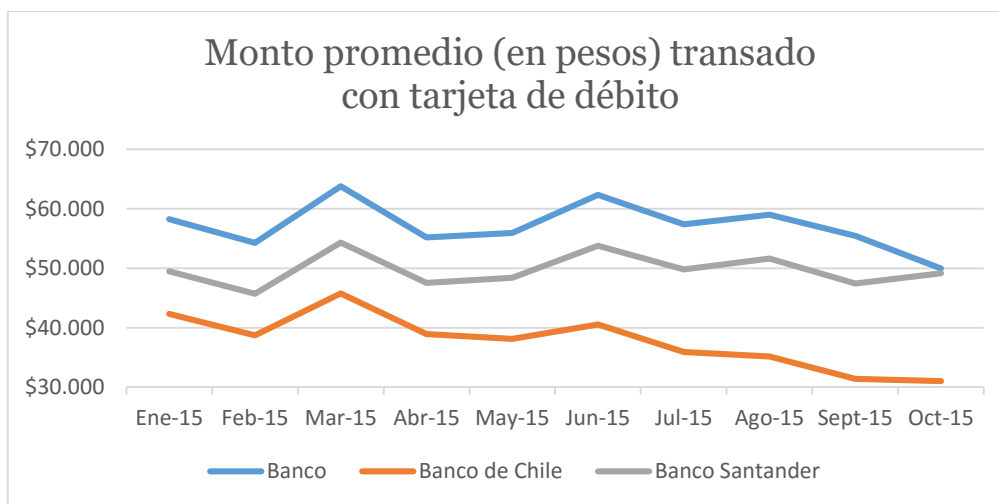


Ilustración 4: Monto promedio transado por tarjetas de débito entre enero 2015 y octubre 2015
Fuente: Elaboración propia con datos de la SBIF

De los gráficos se puede observar que el nivel de consumo que los clientes del banco con tarjeta de crédito es constantemente menor que el de los bancos líderes en el producto, sin embargo, al analizar este comportamiento en la tarjeta de débito, los clientes del banco tienen un mayor consumo promedio. De lo anterior se puede concluir que el bajo nivel de utilización de la tarjeta de crédito por parte de los clientes del banco no se debe a que no compren, sino que se debe a que no siempre – o no todos- utilizan este medio de pago.

De esta forma se puede notar que hay dos razones para creer que se puede incentivar el uso de la tarjeta de crédito por parte de los clientes del banco. En primer lugar, la experiencia de otros bancos demuestra que se pueden tener mejores resultados generando las promociones correctas y en segundo lugar, los clientes del banco tienen la posibilidad de aumentar el uso de su tarjeta pues la razón de su consumo no es una restricción de presupuesto, sino que una elección de medio de pago.

Además de lo anterior, es interesante preguntarse por qué el banco querría aumentar el consumo de sus clientes y no enfocar sus esfuerzos en otras acciones. Para responder esto se puede observar el siguiente gráfico:

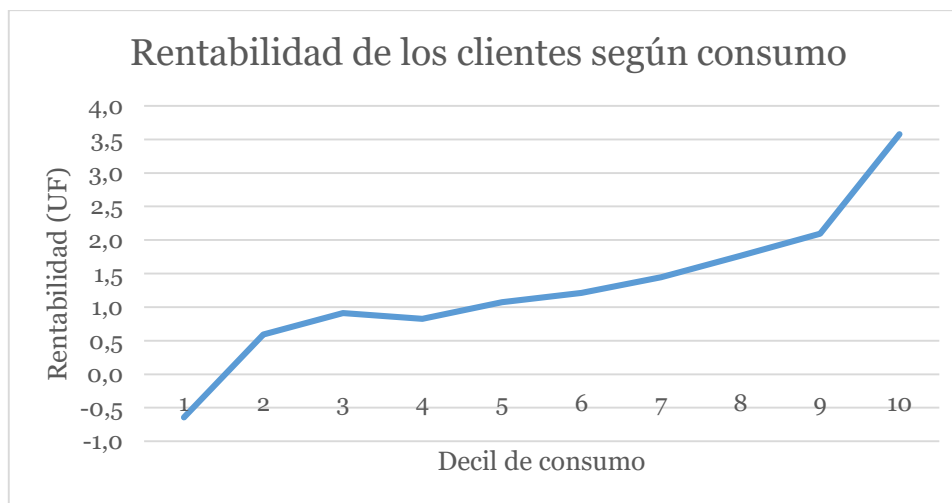


Ilustración 5: Rentabilidad de los clientes del Banco según decil de consumo
Fuente: Elaboración propia con datos del banco

Como se puede apreciar en el gráfico, la rentabilidad del negocio de tarjetas de crédito está directamente ligada al consumo de los clientes y por lo tanto al incentivar a los clientes a utilizar más su tarjeta, se los está incentivando a ser más rentables.

En base a lo anterior, en este trabajo se responden dos preguntas fundamentales para el banco, las cuales son ¿a qué clientes se debe incentivar? y ¿con qué acciones incentivar a cada uno ellos?

Para lograr responder estas preguntas, el proyecto a desarrollar consiste en identificar a los clientes que aumentarán su consumo y a la vez, determinar cuáles son las principales variables que influyen en ese aumento, para así tomar acciones que permitan incentivar el aumento de consumo en otros clientes.

Para esto y en base a la información brindada por la empresa se consideran dos hipótesis, las cuales son presentadas a continuación:

- En primer lugar, se plantea que existen diferencias en el uso de la tarjeta de crédito por los diferentes estados en los que se encuentre un cliente. Esto quiere decir que clientes con diferentes niveles de consumo no sólo compran montos distintos, sino que lo hacen en cosas distintas.
- En segundo lugar, se plantea que existen transacciones que inducen el aumento de consumo de un cliente. Esto es, que la compra en ciertos rubros –o medios de pago– específicos puede ser más determinante que la compra en otros para que un cliente aumente su consumo.

Considerando los puntos anteriores, se comienza identificando los diferentes niveles de consumo que puede tener un cliente para luego en cada uno de estos, determinar las acciones que inducen al aumento del uso de la tarjeta de crédito.

Por otro lado, con respecto al gran número de clientes no fugados sin transacciones, para los bancos es importante entender qué ofrecerle a estos clientes para que vuelvan a utilizar su tarjeta. Se recalca este punto ya que como se ve en el anexo A, la rentabilidad que genera un cliente aumenta mientras más tiempo está ligado a la empresa, por lo que es más conveniente mantener a un cliente que atraer a uno nuevo.

3. Objetivos y alcances de proyecto

Considerando lo expuesto anteriormente, se definen los objetivos y alcances que guían el desarrollo del proyecto.

3.1. Objetivo general

Determinar acciones para incentivar el aumento de consumo con la tarjeta de crédito por parte de los clientes del banco.

3.2. Objetivos específicos

1. Clasificar a los clientes del banco según su información transaccional, determinar sus niveles de consumo y caracterizarlos según sus estructuras de consumo.
2. Identificar las variables que inducen el aumento de consumo y cuantificar su impacto.
3. Determinar perfiles de clientes con alta propensión a aumentar su consumo.
4. Generar recomendaciones para el banco sobre la manera de incentivar el uso de la tarjeta para los diferentes grupos.

3.3. Alcances del proyecto

En primer lugar, el proyecto se llevará a cabo para la “banca de personas” del banco, por lo que no incluirán datos de banca privada ni banca empresas.

En segundo lugar, cabe destacar que en este trabajo no se desarrollará un modelo de propensión de deuda, es decir, se busca determinar con qué acciones aumentar el consumo independiente de que los clientes que consuman no sean los que más se endeudan en revolving.

Por último, se debe considerar que no se diseñarán campañas, sino que se harán recomendaciones generales sobre las acciones que se deben tomar para aumentar el uso de la tarjeta en base a las variables que sean más importantes según los resultados del modelo.

4. Marco conceptual

4.1. Proceso KDD

Es un modelo que entrega una forma de trabajar con datos, entregando un procedimiento establecido y por lo tanto un proceso de descubrimiento en base de datos (la sigla representa su nombre en inglés Knowledge Discovery in Databases). El proceso busca descubrir información útil en grandes cantidades de datos. Se dice que es un procedimiento no trivial pues se pueden usar herramientas avanzadas para encontrar los patrones, que de otra manera no serían visibles.

La metodología se divide en cinco etapas, las cuales son:

- 1) Selección de los datos: Consiste en determinar qué información se va a utilizar. Para esto se deben extraer los datos de las fuentes y elegir cuáles serán finalmente los datos útiles, eliminando campos no relacionados al estudio.
- 2) Pre procesamiento: Es la etapa en la que se limpian los datos, manejando datos faltantes (llenándolos o eliminándolos) y datos fuera de rango. La idea final de esta fase es obtener una base de datos limpia con la que se pueda comenzar a generar nuevas variables y modelos.
- 3) Transformación: Consiste en generar nuevas variables útiles a partir de las existentes. La idea es que en esta fase se crean o transforman todas las variables necesarias para utilizar en las fases siguientes.
- 4) Minería de datos: Es la fase en que se aplican modelos y herramientas con el fin de extraer los patrones “ocultos” de los datos.
- 5) Interpretación y evaluación: Consiste en analizar los patrones obtenidos, viendo cuales son realmente importantes para el problema. Además, en esta fase se evalúan los modelos de la fase anterior, en base a métricas determinadas. Esta es la etapa más cercana al usuario final del proyecto.

Gráficamente el proceso se ve de la siguiente manera:

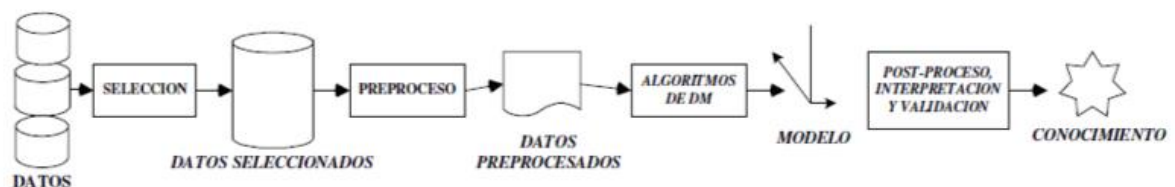


Ilustración 6: Proceso KDD
Fuente: RTD Ibermatica

En este trabajo el proceso KDD se utiliza en dos etapas, primero para clasificar a los clientes según su información transaccional y luego para determinar las acciones con las que incentivar a estos clientes.

4.2. Segmentación y clusterización

Cuando se tiene una base de datos de clientes, es común que los clientes sean diferentes y por lo tanto requieran un trato diferente por parte de la empresa. De esta manera las técnicas que permiten diferenciar entre clientes son útiles pues permiten encontrar grupos más homogéneos que el grupo original.

Para realizar esta agrupación de clientes existen diferentes formas, las cuales se agrupan en dos tipos de procedimientos, segmentación supervisada y clusterización o segmentación no supervisada.

La segmentación supervisada se utiliza cuando se tiene un subconjunto de los clientes en que se sabe a qué segmento pertenecen, sin embargo, no se sabe bien que los caracteriza y por lo tanto, no se puede extender fácilmente al resto de los clientes que no han sido clasificados. Para solucionar esto se utilizan modelos de clasificación que ayudan a determinar las variables más importantes de segmentación. Luego, se puede extender la segmentación a toda la base de clientes.

Por otro lado, también existen casos en que se sabe qué variables generan diferencias entre clientes y por lo tanto la segmentación consiste en separar a los clientes según la variable importante.

Con respecto a la segmentación no supervisada, esta se utiliza cuando no se tienen segmentos previos sobre los que se pueda utilizar un método de clasificación. Es útil cuando no se tiene información a priori sobre la forma que podrían tener los segmentos ni el número de estos.

Los métodos de clusterización se dividen en dos grupos principales, jerárquicos y no jerárquicos. La primera es ideal para casos en que existen relaciones jerárquicas entre los datos. Por otro lado están los métodos de agrupamiento no jerárquico, los cuales forman grupos buscando minimizar la distancia de las observaciones al centro del grupo correspondiente [8].

Además, existen diferentes métricas de homogeneidad para generar clusters, sin embargo, la más utilizada es la distancia euclidiana utilizada por el método K-medias [8].

K-medias

La idea de este algoritmo es que encuentra k centros y adjudica cada dato al centro más cercano.

Es importante destacar que el algoritmo no encuentra el número óptimo de segmentos, sino que recibe este número como parámetro. De esta forma, el algoritmo comienza con k centros aleatorios y asigna cada dato un centro, minimizando la suma de cuadrados dentro de los grupos.

Para entender el algoritmo, es importante definir cómo se mide la distancia entre una observación y el centro, la cual se define como:

- Suponiendo que se tienen n observaciones (x_1, x_2, \dots, x_n) , donde cada observación es un vector de d dimensiones.

- Luego el algoritmo selecciona un centro y asocia cada observación a un centro, eligiendo el k que minimiza $\|x_i - c_k\|$, donde c_k es un vector de d dimensiones que representa del clúster k .
- Una vez que se asigna una nueva observación a un grupo, el centro de este se mueve de la siguiente manera:

$$c'_k = \frac{c_k * n_k + x_i}{n_k + 1}$$

Donde n_k es el número de observaciones del grupo k .

- Este algoritmo se repite hasta que no haya movimiento de observaciones entre grupos.

El algoritmo K medias se utiliza en este trabajo para clasificar a los clientes en base a su información transaccional, en la primera fase del KDD.

4.3. Variables RFM

Son variables transaccionales que buscan determinar el consumo de un cliente. La variable *Recency* mide el tiempo transcurrido desde la última compra del cliente. La variable *Frequency*, mide el número promedio de veces que el cliente compra en un periodo determinado. Por último, la variable *Monetary value* mide el monto del gasto de un cliente.

No es difícil notar que las variables *Frequency* y *Monetary value* debiesen tener una muy alta correlación, pues es esperable que a medida que un cliente compra más veces, el monto total gastado aumente. Por esta razón es que se puede introducir un cambio en que la variable *Monetary value* represente el monto de la transacción promedio del cliente [9]. Esto se hace pues si existiera una correlación muy alta entre las variables, entonces se estaría considerando una dimensión que no aportaría valor.

Una de las ventajas de estas variables, es que evita medir el consumo de un cliente sólo en una dimensión, es decir, no sólo considera el monto que un cliente compró en un periodo, sino que incluye más información relevante para determinar el consumo.

Cabe destacar que estas son las tres variables más utilizadas en la industria bancaria, sin embargo, existen casos en que se agregan nuevas variables para segmentar como el caso de Nan-Chen Hsieh (2004) quién diferenció a los clientes de un banco según sus valores de R, F, M y comportamiento de pago con la tarjeta de crédito, encontrando que existen clientes de alto consumo que utilizan el revolving y clientes de alto consumo que no lo hacen.

En este trabajo no se utiliza el comportamiento de pago pues lo que se busca es encontrar niveles de consumo, sin importar si la persona se endeuda o no. Esto se relaciona directamente con la sección de alcances, en donde se resalta que no se realiza un modelo de propensión al endeudamiento.

Estas variables son especialmente conocidas por el método RFM de quintiles, el cual consiste en generar cinco divisiones sobre cada una de estas tres variables y asignar a cada cliente un valor de 1 a 5 para R, para F y para M. En este trabajo no se utilizará el método RFM pues este se basa en generar grupos de igual tamaño para cada variable y por lo tanto los grupos no son necesariamente los más homogéneos pues puede pasar que dos clientes tengan valores muy similares de alguna variable pero queden en quintiles diferentes.

Cabe destacar que en este trabajo se utilizan las variables RFM como entrada para el algoritmo K medias, de forma de encontrar grupos de clientes con comportamiento transaccional similar.

4.4. Regresión logística

Es un modelo de regresión que clasifica observaciones en variables categóricas. Es un método supervisado que requiere datos en que se sepa el resultado de la clasificación, para así entrenarse y determinar las variables importantes a la hora de clasificar.

Existe tanto el modelo de logit simple, como el multivariado, donde la principal diferencia es que en el primero, sólo existen dos opciones de elección (1 y 0), mientras que en el segundo, la observación tiene la opción de quedar clasificada en más de dos opciones.

Lo que busca el modelo, es determinar cómo afectan diferentes factores a la hora de clasificar una observación, es decir, cuanto más probable es que ocurra un suceso, dado cierto factor. Considerando lo anterior, el modelo de logit simple se puede escribir de la siguiente manera:

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right) = \beta' \cdot X_i$$

De esta manera, la probabilidad de ocurrencia de un suceso es:

$$p_i = \frac{\exp(\beta \cdot X_i)}{1 + \exp(\beta \cdot X_i)}$$

Donde X_i representa a las variables independientes asociadas a la observación i y β es el vector de parámetros a encontrar, lo cuales representan el efecto de cada variable sobre la variable dependiente. Luego, para dos variables de orden similar, se puede decir que aquella con $|\beta|$ mayor será más importante en la clasificación.

El modelo logit es muy utilizado como modelo de clasificación principalmente por su simpleza y fácil interpretación, pues a diferencia de otros modelos, permite conocer el impacto de las variables que se utilizan.

En este trabajo se utiliza la regresión logística para identificar a los clientes que aumentan su nivel de consumo, definiendo la variable dependiente como 1 cuando el cliente aumenta su consumo.

4.5. Árboles de decisión

Al igual que la regresión logística, los árboles de decisión son modelos supervisados de clasificación utilizados para predecir el resultado de una variable generalmente categórica. El método consiste en dividir la base de datos según el cumplimiento de condiciones por parte de las observaciones.

Algunas de las ventajas que tienen los modelos de árboles de decisión por sobre las regresiones logísticas son las siguientes:

- Cada división de los datos se aplica sobre aquellos que cumplen las condiciones anteriores y por lo tanto se incorpora de forma natural la interacción de las variables. Esto quiere decir que el al predecir un resultado, este es válido para los clientes que cumplen una condición A y una condición B simultáneamente. De esta manera se hace fácil encontrar grupos de clientes que cumplen ciertas condiciones y tienen un resultado similar.
- Al ser un sistema basado en reglas de decisión, este tipo de modelos es interpretable por cualquier persona sin la necesidad de tener conocimientos avanzados de minería de datos.

En este trabajo se prueban dos tipos de árboles de decisión como modelos de pensión al aumento de consumo, los árboles CART y los ID3.

CART

El árbol de clasificación y regresión (CART por su nombre en inglés) es un modelo de árbol de decisión que genera divisiones en los datos en base a al índice de Gini, es decir, cada división busca aumentar la diferencia en las proporciones de las clases. De esta forma, el algoritmo CART no generará una división en que ambos grupos sean igual de homogéneos que el original.

ID3

Al igual que el CART, el algoritmo ID3 genera estructuras en forma de árbol para clasificar clientes, sin embargo, utiliza como criterio de división el aumento de la ganancia de información o disminución de la entropía. De esta forma, el algoritmo genera grupos de observaciones similares en cuanto al comportamiento de la variable dependiente.

4.6. Análisis ROC

El análisis ROC se utiliza para la comparación de modelos predictivos y se basa en los diferentes indicadores que se obtienen de la matriz de confusión. Dentro de este análisis destaca la curva ROC, la cual se construye graficando el ratio de verdaderos positivos (sensibilidad) versus uno menos el ratio de verdaderos negativos (especificidad). De esta manera un modelo perfecto tiene valores de sensibilidad y especificidad iguales a 1 y por lo tanto el área bajo la curva ROC también lo es.

Cabe destacar que el análisis ROC es independiente de la proporción de respuestas positivas que haya en la variable dependiente y por lo tanto es un buen análisis para bases desbalanceadas. En este trabajo se utilizan seis indicadores para la comparación del mejor modelo, los cuales representan lo siguiente:

- Sensibilidad: Representa la capacidad del modelo de detectar a los clientes que aumentarán su consumo. Es decir, cuántos de los clientes que realmente aumentan, el modelo clasifica como tales.
- Especificidad: Representa la capacidad de detectar a aquellos que no aumentan su consumo.
- Predicción clase negativa: Indica el porcentaje de clientes que son clasificados como no aumento de consumo y que realmente no aumentan.
- Predicción clase positiva: Indica cuántos de los clientes que el modelo clasifica como aumentos de consumo, finalmente tienen este comportamiento.
- Predicción general: Representa el poder predictivo del modelo considerando a las dos clases de clientes.
- AUC: Es el área bajo la curva ROC. Mientras más cercano a 1 sea su valor, mejor es la calidad del modelo.

Se debe considerar que los seis indicadores pueden tomar valores entre 0 y 1, donde el mayor es el mejor.

5. Metodología

Como lo que se busca en este trabajo es determinar las variables que inducen el aumento de consumo de los clientes, es necesario generar periodos para así entender si un cliente aumenta su consumo entre dos periodos. Para esto es importante comenzar definiendo el largo de los periodos y las fechas de inicio y término de cada uno de ellos.

Por otro lado, como ya se mencionó, la metodología de este trabajo se basa en el proceso KDD, el cual se realiza en dos etapas:

Etapa 1

En esta etapa se busca agrupar a los clientes según sus niveles de consumo con la tarjeta. Para esto es necesario definir el consumo y determinar los niveles. Esta etapa sirve como input para la segunda etapa pues una vez que se tienen los niveles para cada periodo, se puede ver qué clientes aumentaron su consumo pasando a un nivel mayor entre dos periodos. En base a lo anterior, el primer proceso KDD se realiza de la siguiente manera:

Selección de datos: En esta fase se determinan las características que deben cumplir los clientes que integran el estudio. Además se definen las variables representantes del nivel de consumo, es decir, las variables Recency, Frequency y Monetary Value explicadas anteriormente.

Transformación y Pre procesamiento: En este proceso se mezclan estas fases pues como no se tienen las variables RFM en la base original, primero se deben generar para luego ser analizadas y tratar los outliers y valores faltantes.

Minería de datos: En esta fase se busca encontrar los niveles de consumo existentes en la base de datos para cada periodo, para lo cual se utiliza el algoritmo K-medias, el cual agrupa a los clientes según sus variables R, F y M. Se utilizan estas variables pues representan las tres dimensiones principales del consumo [9], [12]. Por otro lado, la combinación de RFM con K-medias fue utilizada por Dimitris Christodoulakis (2005) para demostrar la existencia de una estructura piramidal en los niveles de consumo de los clientes de un banco [20]. Lo que se hace en esta fase entonces es asignar un nivel de consumo a cada cliente en cada uno de los periodos del trabajo.

Análisis de resultados: Finalmente se caracteriza a cada clúster en base al uso que le dan a la tarjeta de crédito, encontrando perfiles de consumo.

Cabe destacar que la determinación de los niveles de consumo de los clientes se realiza para cada periodo del estudio, con lo que se eliminan los efectos de la estacionalidad, ya que si en un periodo los clientes consumen más en promedio, entonces todos los niveles tendrán consumos promedio más altos.

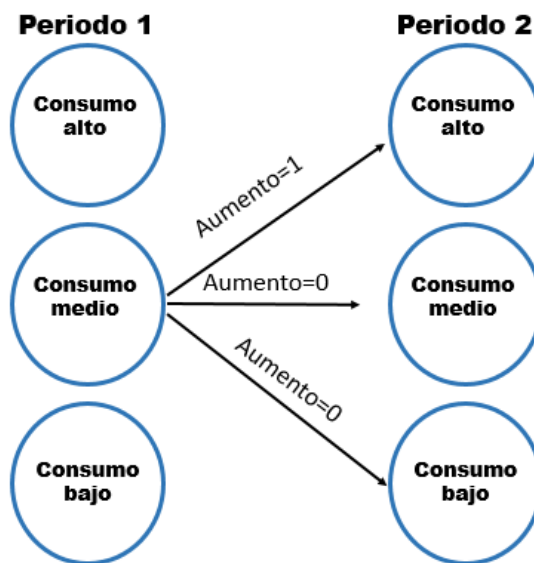
Etapa 2

En esta etapa se busca identificar a los clientes que aumentarán su consumo y determinar cuáles son las variables más importantes que inducen este aumento.

Selección de datos: Se selecciona a todos los clientes que estén en una transición completa, es decir, que estén activos durante dos periodos.

Pre procesamiento: No se realiza pre procesamiento pues este ya se hizo en las etapas anteriores.

Transformación: Se generan las variables independientes de los modelos de propensión, para lo cual se utilizan datos demográficos, características de la cuenta bancaria e información transaccional. Además, se define la variable dependiente de los modelos, la cual toma el valor 1 si el cliente pasó de un nivel de menor consumo a uno de mayor entre dos periodos. En la siguiente ilustración se puede observar la forma en que se define la variable dependiente de los modelos de propensión.



*Ilustración 7: Diagrama de posibles valores de la variable Aumento
Fuente: Elaboración propia*

Minería de datos: Esta es la fase principal de este trabajo pues es donde se determinan las variables que inducen el aumento de consumo. La fase comienza con un análisis exploratorio de la información transaccional de los clientes, con el fin de determinar que variables tienen alta correlación con el aumento de consumo. Luego, con esta información se generan modelos de propensión para determinar la probabilidad de que un cliente aumente su consumo. Cabe destacar que en esta fase se prueban dos tipos de modelos diferentes, Regresión logística y Árboles de decisión, para cada uno de los cuales se evalúa el poder predictivo en base a los indicadores del análisis ROC. Esto se hace en la etapa de validación, para lo cual se usa una base de datos diferente a la utilizada en la construcción de los modelos.

Análisis de resultados: Finalmente se analizan los valores de los coeficientes y las variables importantes que inducen al aumento de consumo según el mejor modelo de cada nivel de consumo y se generan recomendaciones para incentivar el uso de la tarjeta de crédito.

Cabe destacar que Sung Ho Ha (2007) propone que los segmentos de clientes van cambiando y que además, los clientes no siempre tienen el mismo comportamiento y por lo tanto se mueven entre los segmentos. Esta es la razón principal por la que en este trabajo no se realiza un modelo de propensión clásico, si no que se busca las formas de hacer que los clientes cambien a segmentos de más alto consumo [13].

6. Desarrollo de la metodología

6.1. Clasificación de clientes por niveles de consumo

6.1.1. Definición de los periodos de estudio

Para determinar el largo de los periodos a utilizar, es importante considerar que cada periodo debe tener el tiempo suficiente para que un cliente alcance a realizar una transacción y así no ser considerado como inactivo. Además, si se considera un tiempo muy corto, por ejemplo un mes, aumenta la probabilidad de que ese comportamiento sea excepcional y no represente el comportamiento real del cliente. Luego, considerando que el tiempo promedio en que un cliente se activa en el banco es de más de 3 meses, se define que el largo mínimo de cada periodo debe ser de 4 meses.

Por otro lado, periodos muy largos generan pocas transiciones (entre el periodo 1 y el 2 existe una transición) y por lo tanto se define el largo de cada periodo como 4 meses. De esta manera, se cuenta con los siguientes 6 periodos para el estudio:

| Periodo | Mes inicial | Mes final |
|----------------|--------------------|------------------|
| 1 | Dic-13 | Mar-14 |
| 2 | Abr-14 | Jul-14 |
| 3 | Ago-14 | Nov-14 |
| 4 | Dic-14 | Mar-15 |
| 5 | Abr-15 | Jul-15 |
| 6 | Ago-15 | Dic-16 |

Tabla 3: Periodos de estudio

Una vez que se tienen los periodos definidos, se busca encontrar los diferentes niveles de consumo a los que pueden pertenecer los clientes del banco en cada periodo. Para esto se define el consumo como la combinación de las variables R, F y M.

6.1.2. Selección de datos

Para realizar este trabajo se tienen datos transaccionales y de cuenta desde diciembre de 2013 hasta noviembre de 2015, es decir, dos años de información de los clientes del banco con lo que se pueden formar seis periodos de cuatro meses como los detallados en la parte anterior.

Además, los clientes que integran el estudio deben cumplir con las siguientes características al menos en un periodo completo:

- Deben ser clientes pertenecientes al banco y no a alguna de sus filiales.
- Para evitar que se trate de empresas mal ingresadas o problemas de data, sólo se consideran clientes de género masculino y femenino, y entre 18 y 100 años de edad.

- Sólo se consideran clientes con tipos de tarjeta común, es decir, se excluye a clientes con tarjetas internas del banco pues tienen condiciones especiales y podrían distorsionar el estudio.
- Este estudio no está diseñado para descubrir la forma de hacer que clientes fugados vuelvan al banco, sino que busca incentivar el uso de la tarjeta en clientes activos. Por esta razón sólo se consideran clientes con al menos una transacción durante alguno de los 24 meses del estudio.
- Además, para que un cliente esté activo en un periodo, debe haber tenido al menos una cuenta vigente durante los 4 meses de ese periodo.

Dado esto, de los más de 130.700 clientes del banco en noviembre de 2015, el estudio considera a 60.675 mil clientes, equivalente al 46,4% de la cartera. Cabe destacar que este es el número total de clientes, lo que no quiere decir que en cada periodo haya la misma cantidad.

Por otro lado, para definir las variables R, F y M se utilizan datos transaccionales de los clientes del banco, más específicamente, datos sobre transacciones nacionales. No se consideran transacciones que contemplen el pago de la tarjeta de crédito ya que esto no es un nuevo consumo. Con esto, se consideran todos los gastos que un cliente tenga en alguno de los siguientes ítems:

- Compras nacionales
- Avances de efectivo
- Avances en cuotas

Cabe destacar que no se consideran compras internacionales debido a que el banco incentiva que la deuda internacional sea pagada con el cupo nacional pues de esa manera genera una doble ganancia y por lo tanto cerca del 70% de las transacciones internacionales están contabilizadas en las transacciones nacionales. A pesar de esto, tanto compras como avances internacionales se utilizan como variables independientes en los modelos de propensión para analizar si son capaces de influir en el consumo nacional de los clientes.

Para la variable F se utiliza el número de transacciones del cliente en cada periodo, para la variable M se utiliza el monto y número total de las transacciones, además del valor de la UF en cada periodo. Se utiliza la UF para así evitar considerar que los clientes consumen más por temas de alza de precios. Finalmente para la variable R se considera la fecha de las transacciones realizadas durante el periodo.

6.1.3. Transformación de variables

En un proceso KDD común se deben pre procesar los datos antes de transformarlos, sin embargo, en la base de datos con la que se desarrolla este trabajo no existen las variables R, F y M por lo que deben ser creadas antes de poder procesarlas.

Luego, para crear la variable F se cuentan todas las transacciones validas del cliente en el periodo de cuatro meses y se dividen en 4. De esta forma se obtiene el número promedio de transacciones mensuales. Con respecto a la variable M, se divide el monto total

transado en el periodo en el número total de transacciones. Además este valor se divide en el valor de la UF promedio durante el mes. Finalmente la variable R se crea calculando los días entre la fecha de la última transacción del cliente y el último día del periodo. Luego, esta variable nunca puede ser superior a 120 o 121 días dependiendo del periodo.

A continuación se presenta un análisis descriptivo de las variables R, F y M utilizando como referencia el periodo 6 del trabajo, por ser el más cercano a la actualidad.

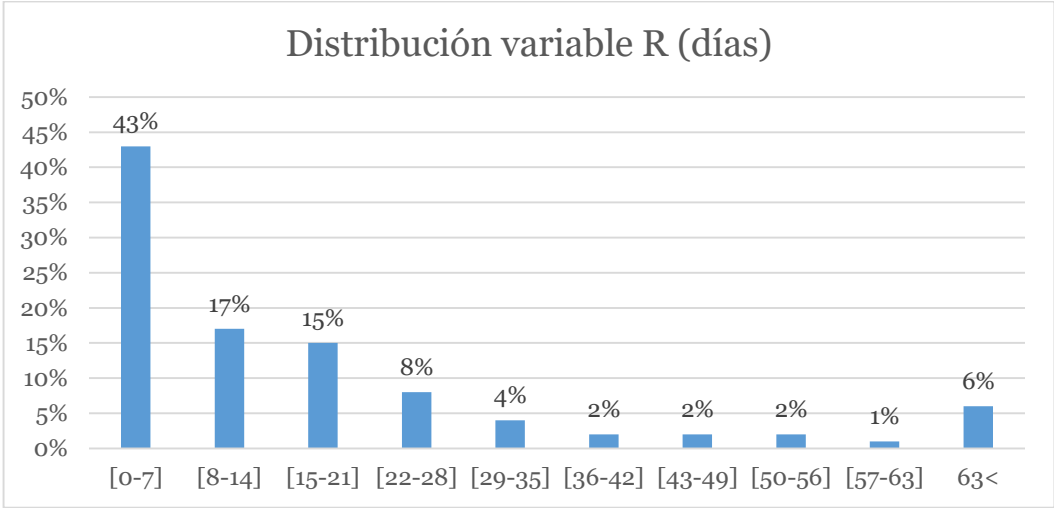


Ilustración 8: Histograma variable Recency para clientes con transacciones Periodo 6
Fuente: Elaboración propia

Del gráfico se observa que el 43% de los clientes consumió durante la última semana del periodo, mientras que cerca del 80% de ellos lo hicieron al menos una vez durante el último mes. Por otro lado el 7% consumió por última vez 60 días antes del fin del periodo. Cabe destacar que esto se debe a que en el histograma sólo se incluyen clientes con transacciones durante este espacio de tiempo.

A continuación se presenta el histograma de la variable Frequency, nuevamente para los clientes con transacciones durante el periodo:

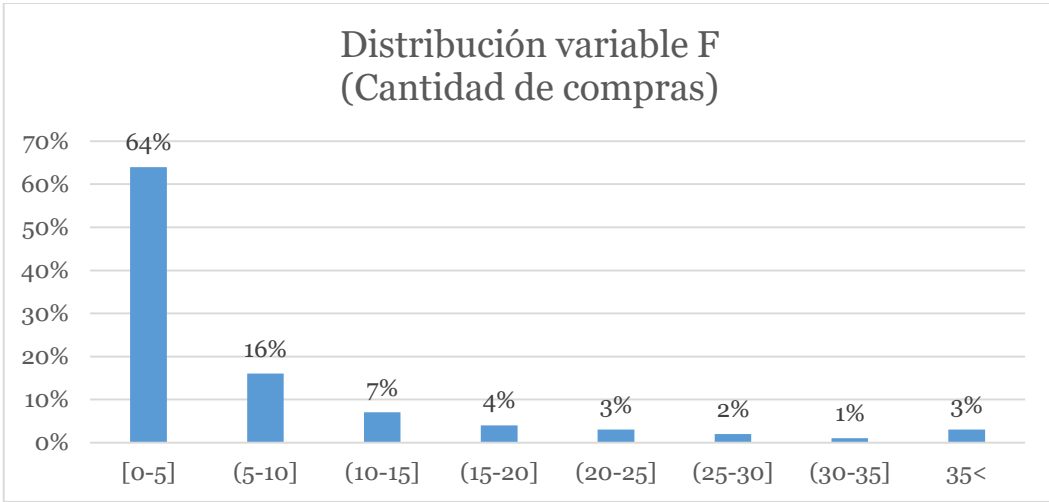


Ilustración 9: Histograma variable Frequency para clientes con transacciones Periodo 6
Fuente: Elaboración propia

Como se puede ver en el gráfico, el 64% de los clientes tienen menos de 5 transacciones durante un mes promedio, es decir, utilizan su tarjeta cada 6 días. Por otro lado, existe un 20% de clientes que realizan más de 10 compras mensuales, los que son considerados como clientes más activos. Cabe destacar que el histograma muestra los datos previos al tratamiento de outliers y por lo tanto es esperable tener un 3% de clientes con un nivel de uso de la tarjeta muy superior al promedio.

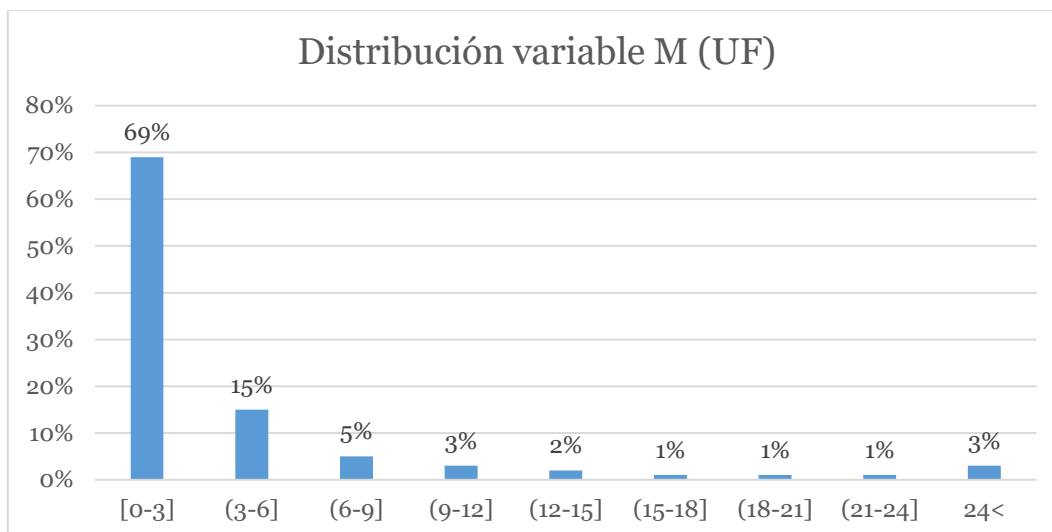


Ilustración 10: Histograma variable Monetary Value para clientes con transacciones Periodo 6
Fuente: Elaboración propia

Del histograma se puede notar que la gran mayoría de los clientes tienen un ticket promedio inferior a las 3 UF, sin embargo, existe un 16% de clientes que gastan al menos 6 UF por cada compra. Al igual que en el gráfico anterior, no se han tratado los outliers de esta variable y por lo tanto es esperable encontrar clientes con comportamientos diferentes al común de la cartera. Por último, es importante señalar que este histograma representa el valor del ticket promedio y no el gasto total, por lo que es posible que clientes con un M muy bajo tengan un alto consumo durante el periodo.

Como se mencionó anteriormente, una de las ventajas de definir el consumo en base a las variables RFM es que evita definirlo solamente como el monto transado o la cantidad de compras que realiza un cliente. Sin embargo, para poder afirmar esto se requiere que la correlación entre las variables no sea alta, pues de lo contrario existiría una variable redundante. Es importante señalar que la redundancia no produciría un problema en la metodología, pero sí aportaría menos información de la deseada.

A continuación se presentan las correlaciones entre las variables R, F y M tomando como referencia el periodo 6:

| | R | F | M |
|----------|----------|----------|----------|
| R | 1 | -0,33 | 0,20 |
| F | -0,33 | 1 | -0,13 |
| M | 0,20 | -0,13 | 1 |

Tabla 4: Coeficientes de correlación entre variables RFM
Fuente: Elaboración propia

De la tabla se puede apreciar que las correlaciones son bajas, sin embargo, llama la atención el coeficiente de las variables F y R el cual si bien no es cercano a 1, sí es más alto que el resto. Cabe destacar que efectivamente es esperable que un cliente que consume más veces lo haya hecho más recientemente. Es importante mencionar que si bien las relaciones de las variables no son necesariamente lineales, los coeficientes de correlación entregan una intuición sobre los comportamientos de dichas variables.

Con respecto a los signos de los coeficientes, es esperable que mientras más veces consuma un cliente, el valor de ticket promedio disminuya. Esto además explica el signo positivo del coeficiente de correlación entre las variables R y M.

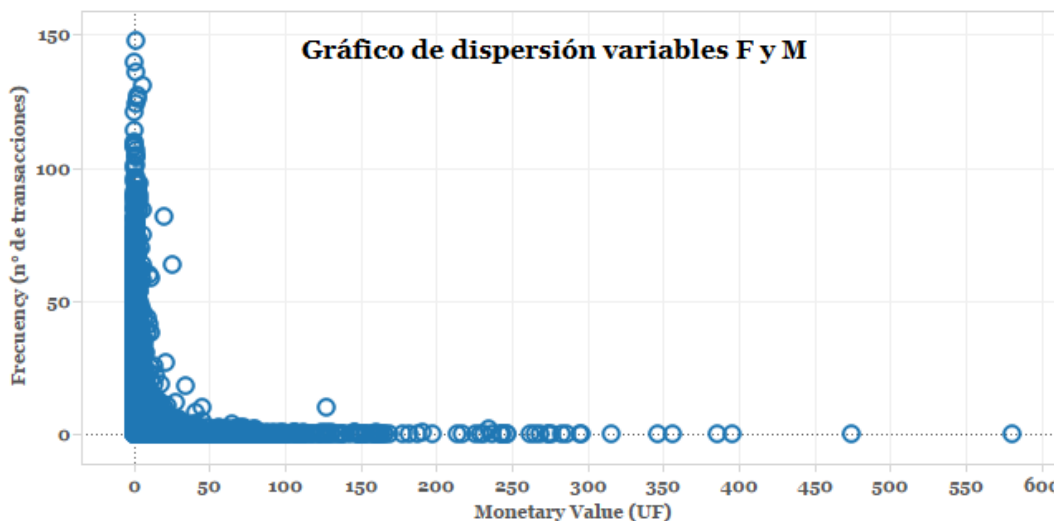


Ilustración 11: Gráfico de dispersión variables F y M periodo 6
Fuente: Elaboración propia

Del gráfico se puede apreciar que existen diferentes tipos de clientes. Por un lado están aquellos que consumen pocas veces y con montos bajos. Por otro lado están los que consumen pocas veces pero con montos altos y aquellos que lo hacen con montos bajos pero muchas veces. Esto es importante pues significa que existen grupos que utilizan de forma diferente su tarjeta durante el periodo.

6.1.3. Pre procesamiento

En esta fase se pone especial énfasis en los valores faltantes y outliers de las variables R, F y M. Cabe destacar que este proceso se realiza para cada periodo del estudio.

Dada la forma en que se construyen las variables, no existen valores faltantes, pues cada transacción cuenta con su respectiva fecha en la base de datos. Con respecto a los outliers, estos sólo se encontraron en las variables F y M pues si un cliente compró en un periodo entonces R no puede ser mayor a 120 o 121 dependiendo del periodo.

En el caso de los clientes con frecuencia o montos muy altos, estos pueden distorsionar los resultados debido a que no se trata de clientes comunes y por lo tanto se decide eliminarlos del estudio. Para esto se considera que aquellos clientes con valores mayores

a la media de la variable más 2 desviaciones estándar, están fuera de rango. Esto es para cada uno de los 6 periodos del estudio.

Con esto se eliminan cerca del 5% de los clientes de cada periodo, quedando en el último 48.154 personas. Finalmente se presentan los estadísticos más importantes de las variables R, F y M durante el periodo 6, una vez sacados los outliers, incluyendo tanto a clientes con y sin transacciones tanto a clientes con y sin transacciones durante el periodo para las variables F y M, mientras que la variable R incluye sólo a clientes con transacciones.

| | R (Días) | F (n°) | M (UF) |
|----------------------------|---------------------|-------------------|-------------------|
| Promedio | 16 | 3,9 | 2,7 |
| Desviación estándar | 21 | 5,4 | 4,3 |
| Mínimo | 0 | 0,0 | 0,0 |
| Máximo | 121 | 27,3 | 33,3 |

*Ilustración 12: Estadísticos principales variables RFM sin outliers periodo 6
Fuente: Elaboración propia*

Se puede notar que la desviación estándar es mayor al promedio de cada una de las tres variables, lo que indica que existe una gran dispersión, lo que justifica la necesidad de encontrar grupos más homogéneos,

6.1.4. Minería de datos

Antes de comenzar con la minería de datos se debe recordar que en este trabajo no se incluyen clientes que no hayan realizado transacciones durante 24 meses por considerarse fugados, sin embargo, sí existen clientes que durante un periodo completo no tengan transacciones y a estos se les llama Clientes Inactivos, los cuales son de gran importancia para el banco, el cual busca entender de qué forma activarlos. Por esta razón es que aquellos clientes que no tengan transacciones durante un periodo no ingresan a la clusterización durante ese periodo, sino que pasan directamente al segmento de clientes inactivos.

Lo anterior quiere decir que además de los niveles de consumo que se encuentren, existe un segmento más llamado Inactivos. Cabe destacar que un cliente no puede estar inactivo durante todos los periodos, pero sí puede estarlo por más de uno.

Dicho lo anterior, en esta fase se busca encontrar los diferentes clusters que representan los niveles de consumo a los que puede pertenecer un cliente en un periodo determinado. Para esto se utiliza el algoritmo K-medias tomando como “input” las variables RFM de los clientes.

Es importante destacar que al encontrar los clusters, el criterio para definir cuál tiene mayor nivel de consumo es el valor de la multiplicación de las variables F y M. Esto se hace ya que con el algoritmo K-medias se encuentran grupos de clientes similares, pero no necesariamente las tres variables irán en la misma dirección.

Como se comentó en el marco teórico, el algoritmo K-medias requiere que se le entregue el número de clusters a buscar y en base a eso asigna cada observación a un grupo. Para evaluar este algoritmo se utiliza la distancia total sumada de cada observación al centro del grupo que se le asignó. Luego, para determinar el número de clusters óptimos se utiliza la “regla del codo” [23].

La “regla del codo” consiste en graficar el valor de la distancia total para distintos números de clusters. De esta forma existirá un número de grupos en que seguir dividiendo ya no aporta un valor significativo y es ahí donde se observa una punta en el gráfico. Utilizando esta metodología se realiza la búsqueda de los clusters, la cual se presenta a continuación.

Para determinar el número de clusters se utiliza el algoritmo K-medias 9 veces, probando desde 2 a 10 clusters, con lo que se obtiene el siguiente gráfico:

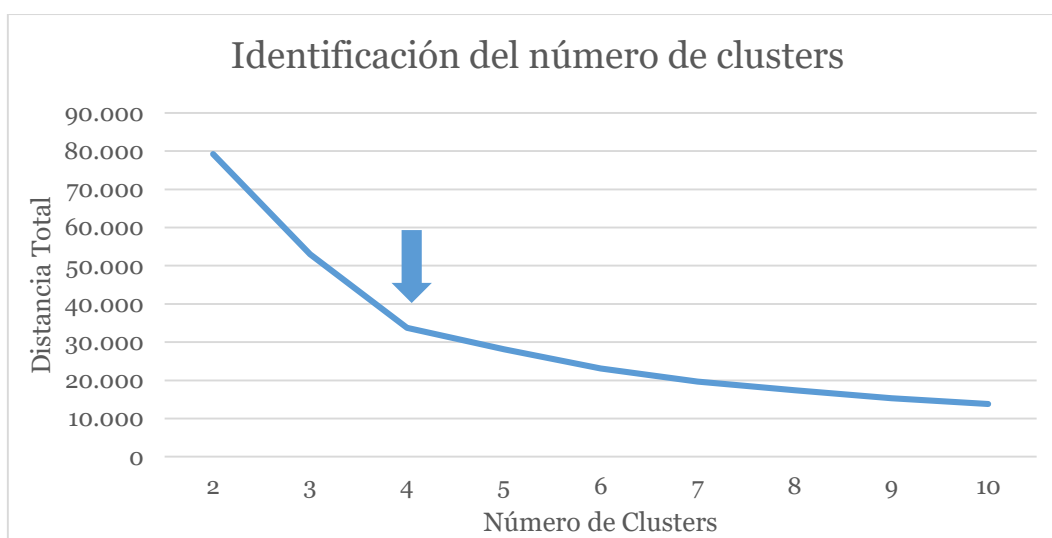


Ilustración 13: Distancia total vs número de clusters
Fuente: Elaboración propia

Del gráfico se puede observar que el número de segmentos es de 4, obteniéndose los siguientes promedios de las variables R, F y M:

| | R (días) | F (N°) | M (UF) |
|------------------|---------------------|-------------------|-------------------|
| Clúster 1 | 72 | 1 | 3,6 |
| Clúster 2 | 12 | 3,1 | 2,4 |
| Clúster 3 | 27 | 1,2 | 17,8 |
| Clúster 4 | 4 | 15,9 | 1,6 |

Tabla 5: Variables RFM por clúster obtenido en Periodo 6
Fuente: Elaboración propia

Si bien estos clusters sólo representan al periodo 6, durante el resto de los periodos se obtienen resultados similares tanto en los valores las variables RFM como en la distribución de clientes de cada nivel de consumo. En lo que sigue de este informe los clusters pasan a ser niveles de consumo donde el 1 es el nivel más bajo y el 4 es el más alto.

6.1.5. Análisis de resultados

De esta manera, los niveles de consumo que puede tener un cliente durante el periodo 6 se presentan en la siguiente tabla:

| Nivel de consumo | R | F | M | F x M | # clientes | % de clientes de la cartera |
|-------------------|----|------|------|-------------|------------|-----------------------------|
| Inactivo | - | 0 | 0 | 0 | 9.851 | 20% |
| Bajo | 72 | 1,0 | 3,6 | 3,6 | 3.689 | 8% |
| Medio Bajo | 12 | 3,1 | 2,4 | 7,4 | 25.976 | 54% |
| Medio Alto | 27 | 1,2 | 17,8 | 21,3 | 2.388 | 5% |
| Alto | 4 | 15,9 | 1,6 | 25,4 | 6.264 | 13% |

Tabla 6: Niveles de consumo periodo 6

De la tabla se puede apreciar que la mayoría de los clientes tienen un nivel Medio Bajo de consumo, utilizando la tarjeta cada 10 días en promedio y gastando 7,4 UF mensuales. Por otro lado se puede observar que a excepción de los clientes de Medio Alto consumo, a mayor número de transacciones mayor es el consumo, es decir, los mejores clientes utilizan la tarjeta más veces y no necesariamente lo hacen en montos más grandes. Además, del anexo B se puede ver que las desviaciones estándar de cada variable bajan.

Luego de obtener los grupos con diferentes niveles de consumo, es interesante entender qué los caracteriza en cuanto a su información transaccional y demográfica. Esta última es importante pues si bien los clientes van cambiando su nivel de consumo, podría darse el caso de que clientes con determinadas características fueran más propensos a pertenecer a algún grupo en específico. Cabe destacar que si bien se presenta la caracterización de los grupos para el último periodo de este trabajo, las distribuciones de consumo y características demográficas se mantienen en cada uno de ellos.

6.1.5.1. Caracterización clientes inactivos

Son 9.851 los clientes que cumplen las condiciones para ingresar al estudio que no tienen transacciones durante el periodo 6. De estos el 62% son hombres y el 38% mujeres. Además, la edad promedio de los clientes es 43 años.

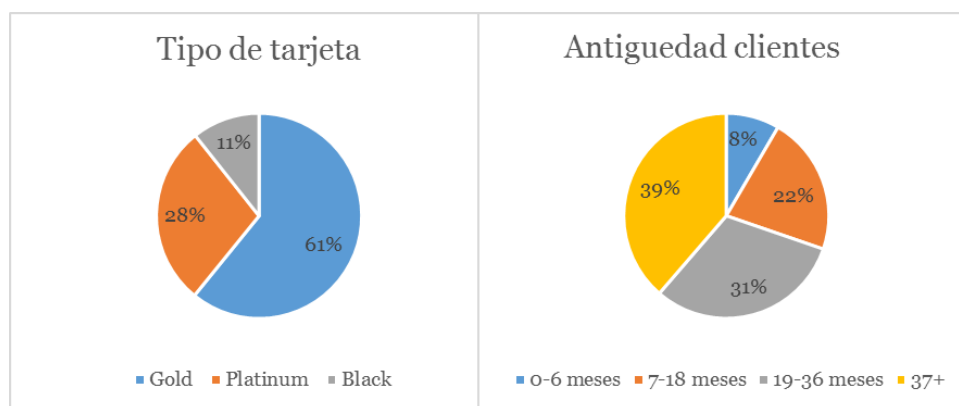


Ilustración 14: Información de cuenta clientes Inactivos

De los gráficos se puede notar que la mayoría de los clientes inactivos tienen tarjetas doradas y además llevan en el banco al menos un año y medio. Además, el cupo promedio de estos clientes es de 2,8 millones de pesos.

Con respecto a la cantidad de cuentas, 45% de los clientes inactivos tiene una, mientras que el 53% tiene dos. Cabe destacar además que en promedio los clientes inactivos tienen una deuda de colocación (revolving) de 106 mil pesos.

Por último cabe destacar que para los clientes inactivos no se presenta información transaccional pues no la tienen durante el periodo.

6.1.5.2. Caracterización segmento Bajo consumo

En el último periodo del estudio 3.689 clientes tienen un nivel de consumo bajo, de los cuales el 59% mientras que el 41% son mujeres y en la edad promedio del grupo completo es de 44 años.

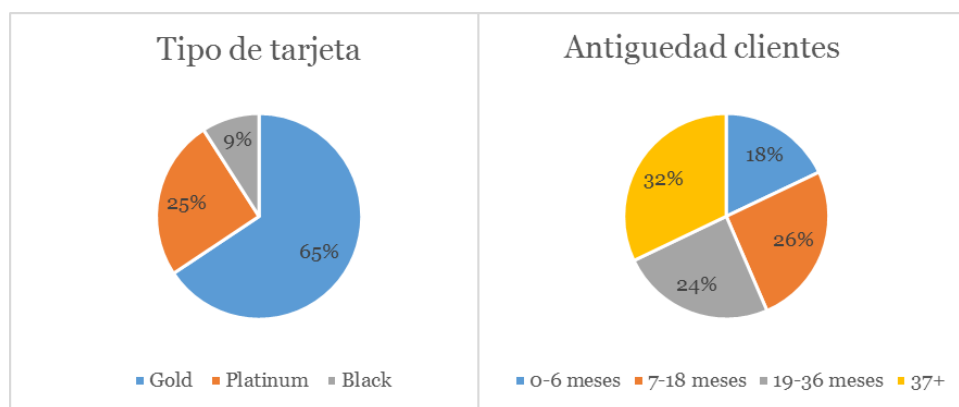


Ilustración 15: Información de cuenta clientes Bajo consumo

De los gráficos se puede notar que los clientes de Bajo consumo utilizan en general tarjetas doradas. Además, se puede notar que el porcentaje de clientes nuevos es mayor que en el grupo anterior, siendo más equilibrada la distribución de la antigüedad. Por otro lado, el cupo promedio de este grupo es más bajo que en los clientes inactivos llegando a 2,5 millones.

Con respecto a la cantidad de cuentas de los clientes de Bajo consumo, la distribución es similar al grupo anterior pues el 47% tiene sólo una cuenta, el 51% tiene dos y sólo el 2% de los clientes tiene tres o más cuentas. Además la deuda de colocación de estos clientes es de 204 mil pesos en promedio, es decir, 92% más que los clientes inactivos.

Con respecto a la información transaccional de los clientes de bajo consumo, esta se presenta a continuación, donde el porcentaje se calcula utilizando el promedio de los clientes:

| | Porcentaje de uso |
|------------------------------|--------------------------|
| Cuotas precio contado | 36% |
| Cuotas con interés | 19% |
| Compra sin cuotas | 45% |

Tabla 7: Distribución tipos de pago clientes Bajo consumo

Se puede observar que los clientes de bajo consumo prefieren tipos de pago que no generen interés.

Para entender el uso que le dan los clientes de Bajo consumo a su tarjeta de crédito, es interesante observar los rubros en los que compran. Para esto se presenta a continuación una tabla con el porcentaje del monto total consumido que destinan a cada uno de los tipos de rubros, en los que destacan las Líneas aéreas, los supermercados y las tiendas por departamento. Además, los montos transados en avances de efectivo son similares a los rubros más consumidos.

| Rubro | Porcentaje de uso |
|----------------------------------|--------------------------|
| Tiendas por departamentos | 12,1% |
| Líneas aéreas | 9,6% |
| Supermercados | 8,4% |
| Avances en efectivo | 8,3% |
| Servicios financieros | 5,4% |
| Restaurantes | 3,4% |
| Pago de impuestos | 1,1% |
| Bencina | 3,3% |
| Vestuario | 3,2% |
| Clínicas | 2,5% |
| Farmacias | 2,2% |
| Reparación y acces. Autos | 1,9% |
| Servicios básicos | 1,7% |
| Pago de impuestos | 1,1% |
| Educación superior | 0,9% |
| Compraventa autos | 0,3% |
| Colegios | 0,2% |
| Otros rubros | 34,6% |

Tabla 8: Estructura de consumo por rubros clientes Bajo consumo

Es importante notar de la tabla que estos clientes consumen principalmente en rubros habituales o básicos. Por otro lado, cabe destacar que se presenta esta agrupación de rubros pues es la utilizada por la consultora Innovapay.

6.1.5.3. Caracterización segmento Medio Bajo consumo

Son 25.976 los clientes de Medio Bajo consumo durante el periodo 6, es decir, la mayoría de los clientes del banco tienen este nivel de consumo, lo cual calza con el

conocimiento previo de la consultora y del banco. De estos clientes el 58% son hombre y el 42% mujeres. Por otro lado la edad promedio de este grupo es de 46 años.

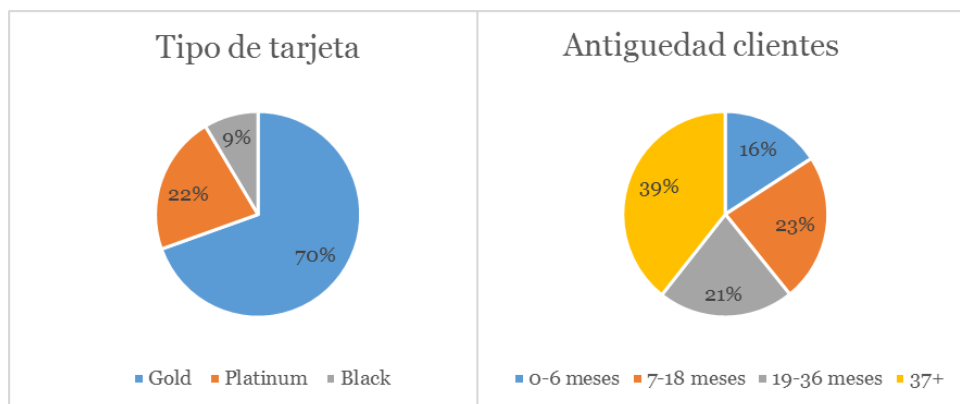


Ilustración 16: Información de cuentas clientes Medio Bajo consumo

De la tabla se puede ver que en comparación a los grupos anteriores, el porcentaje de clientes de Medio Bajo consumo que utilizan tarjetas doradas es mayor. Además, la distribución por antigüedad vuelve a indicar que en general son clientes que llevan al menos 18 meses en el banco. Con respecto al cupo permitido promedio es de 2,8 millones de pesos. La información de cuentas de la cartera completa se encuentra en el anexo D.

Con respecto a la deuda de colocación, esta es un 48% más alta que la de los clientes de bajo consumo, llegando a los 302 mil pesos promedio mensual. A continuación se presenta la información transaccional y la estructura de consumo en base a rubros de los clientes de Medio Bajo consumo:

| | Porcentaje de uso |
|------------------------------|-------------------|
| Cuotas precio contado | 35% |
| Cuotas con interés | 17% |
| Compra sin cuotas | 48% |

Tabla 9: Distribución tipos de pago clientes Medio Bajo consumo

Nuevamente se aprecia que los clientes de Medio Bajo consumo prefieren tipos de pago que no generen interés, siendo el principal la compra sin cuotas, conocida también como compra normal.

| Rubro | Porcentaje de uso |
|----------------------------------|--------------------------|
| Avances en efectivo | 12,0% |
| Servicios financieros | 9,9% |
| Tiendas por departamentos | 9,6% |
| Supermercados | 9,0% |
| Líneas aéreas | 5,5% |
| Bencina | 3,9% |
| Servicios básicos | 3,8% |
| Vestuario | 3,0% |
| Restaurantes | 2,7% |
| Clínicas | 2,7% |
| Farmacias | 2,4% |
| Reparación y acces. Autos | 1,6% |
| Pago de impuestos | 1,0% |
| Colegios | 0,8% |
| Educación superior | 0,8% |
| Viajes | 0,7% |
| Compraventa Autos | 0,4% |
| Otros | 30,3% |

Tabla 10: Estructura de consumo por rubros clientes Medio Bajo consumo

Se puede notar que con respecto al grupo de Bajo consumo, los clientes de Medio Bajo utilizan más su tarjeta para realizar avances de efectivo, compra en supermercados y servicios financieros. Por otro lado, si bien las compras en tiendas por departamento siguen siendo una gran parte de las compras totales, su importancia es un 3% menor que la de los clientes del grupo anterior. Los montos transados por cada nivel de consumo en cada rubro y sus desviaciones estándar se encuentran en el anexo F, en donde se incluye tanto a clientes con y sin transacciones en cada rubro, lo que hace crecer los valores de la dispersión.

6.1.5.4. Caracterización segmento Medio Alto consumo

Los clientes de Medio Alto consumo son el grupo más pequeño de los cinco, siendo sólo 2.388, de los cuales el 64% son hombres y el 36% mujeres, es decir, el porcentaje de hombres es mayor que en el resto de los niveles de consumo. La edad promedio de estos clientes es similar a la del resto de los grupos siendo 44 años.

Con respecto a la información de las cuentas, el porcentaje de tarjetas plateadas y negras es mayor que en los otros grupos, mientras que la distribución de antigüedad vuelve a indicar que la mayoría de los clientes de Medio Alto consumo llevan más de 18 meses en el banco, sin embargo, cerca del 20% son clientes nuevos. Además de lo anterior, el cupo promedio de los clientes es de 4 millones de pesos, es decir, un 42% más alto que el de los clientes de Medio Bajo consumo.

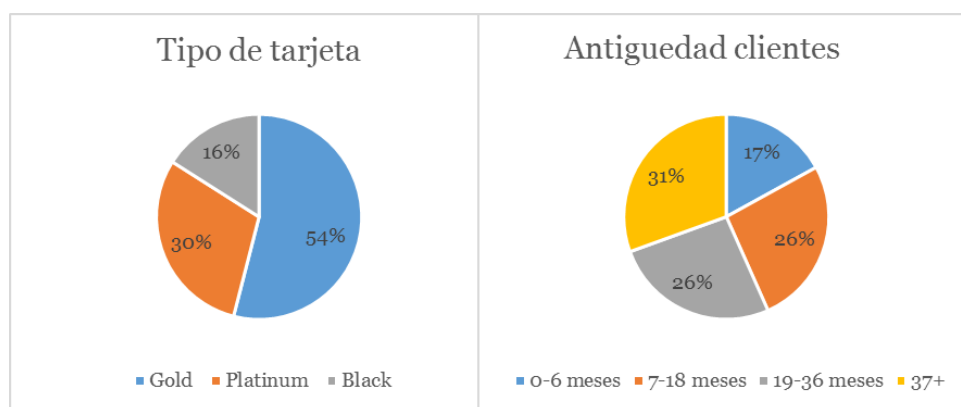


Ilustración 17: Información de cuentas clientes Medio Alto consumo

Cabe destacar el hecho de que los clientes de Medio Alto consumo utilizan en general mejores tarjetas que el resto de los grupos, tienen más cuentas y además tienen cupos permitidos más altos, por lo que se puede suponer que son clientes con mayores ingresos.

Con respecto a la cantidad de cuentas, este grupo presenta una diferencia importante con el resto, pues el porcentaje de clientes con más de una cuenta alcanza casi el 65%. Además, la deuda de colocación promedio mensual de este grupo es de 330 mil pesos, es decir, 10% más que los clientes de Medio Bajo consumo.

A continuación se presenta la información transaccional de los clientes de Medio Alto consumo:

| | Porcentaje de uso |
|------------------------------|-------------------|
| Cuotas precio contado | 30% |
| Cuotas con interés | 39% |
| Compra sin cuotas | 31% |

Tabla 11: Distribución tipos de pago clientes Medio Alto consumo

De la tabla se puede ver que los clientes de este grupo utilizan principalmente las cuotas como tipo de pago, teniendo como medio de pago predilecto la compra en cuotas con interés. Esto es importante pues los diferencia del resto de los grupos presentados, además dado el modelo de negocios de la tarjeta de crédito, estos clientes generan ingresos al banco tanto por su mayor nivel de consumo como por la forma en que utilizan su tarjeta.

Con respecto a los rubros en los que consumen los clientes, se puede apreciar que la distribución de consumo es completamente diferente a la del resto de los clientes, donde los rubros básicos como bencina, supermercado, vestuario o farmacias no representan más del 2% (por separado) del monto consumido. El último rubro habitual, Tiendas por departamentos, tiene una importancia mayor que el resto, sin embargo, es menor que la importancia que tiene en otros segmentos de consumo. Por otro lado, destaca que estos clientes utilizan la tarjeta para rubros en los que comúnmente se realizan grandes compras como Líneas aéreas, educación o clínicas. Cabe destacar también la gran participación que tienen en el consumo los avances en efectivo, llegando casi al 40%.

Lo anterior se puede ver en la siguiente tabla:

| Rubro | Porcentaje de uso |
|----------------------------------|--------------------------|
| Avances en efectivo | 38,0% |
| Líneas aéreas | 13,2% |
| Tiendas por departamentos | 6,5% |
| Viajes | 4,2% |
| Colegios | 3,6% |
| Clínicas | 3,2% |
| Educación superior | 2,2% |
| Reparación y acces. Autos | 1,8% |
| Supermercados | 1,8% |
| Compraventa Autos | 1,6% |
| Pago de impuestos | 1,3% |
| Servicios financieros | 0,9% |
| Vestuario | 0,8% |
| Servicios básicos | 0,6% |
| Restaurantes | 0,4% |
| Farmacias | 0,2% |
| Bencina | 0,1% |
| Otros | 19,6% |

Tabla 12: Estructura de consumo por rubros clientes Medio Alto consumo

Al analizar también las desviaciones estándar de los diferentes rubros del anexo F, se puede ver que estas son más altas que en otros segmentos, lo que se debe a que los clientes de Medio Alto consumo tienen pocas transacciones en pocos rubros y por lo tanto en cada rubro hay muchos clientes que no consumen y aumentan la dispersión.

Por otro lado, lo expuesto en la tabla tiene relación con que según los valores de R, F y M obtenidos del algoritmo K-medias, ya que este grupo de clientes se caracteriza por tener una baja frecuencia de compras y un alto ticket promedio.

6.1.5.5. Caracterización segmento Alto consumo

Los clientes de más alto consumo del banco durante el periodo 6 del estudio son 6.264, de los cuales el 61% son hombres y el 39% mujeres. La edad promedio de estos clientes es 48 años, superior al resto de los clientes.

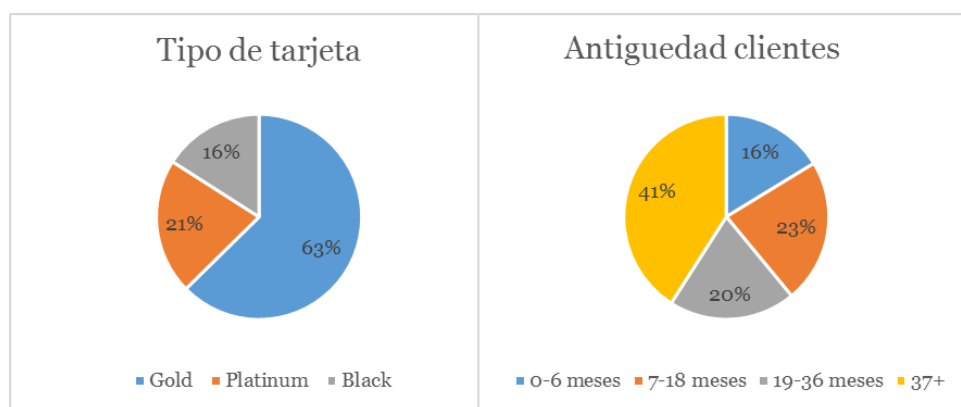


Ilustración 18: Información de cuentas clientes Alto consumo

De los gráficos se puede notar que en cuanto al tipo de tarjeta, son más los clientes con tarjetas negras en este grupo que en los de más bajo consumo. Con respecto a la antigüedad, se aprecia que es el grupo con mayor porcentaje de clientes con más de tres años en el banco. Además es importante destacar que el cupo promedio de los clientes de Alto consumo es 4,2 millones de pesos, es decir, es sólo un 5% más alto que los clientes de Medio Alto consumo.

Con respecto a la cantidad de cuentas, la distribución es similar al resto de los grupos, sin embargo, destaca el hecho de que un 5% de los clientes tiene más de tres cuentas activas con el banco. Por otro lado la deuda de colocación promedio mensual de estos clientes es 412 mil pesos, es decir, un 25% más alto que los clientes de Medio Alto consumo y el más alto de todos grupos de clientes presentados.

La información transaccional de los clientes de alto consumo se presenta a continuación:

| | Porcentaje de uso |
|------------------------------|-------------------|
| Cuotas precio contado | 34% |
| Cuotas con interés | 11% |
| Compra sin cuotas | 55% |

Tabla 13: Distribución tipo de pago clientes Alto consumo

Como se puede ver, los clientes de alto consumo utilizan su tarjeta principalmente para pagar sin cuotas y en caso de pagar con ellas, prefieren cuotas de precio contado, es decir, sin intereses.

| Rubro | Porcentaje de uso |
|----------------------------------|--------------------------|
| Supermercados | 18,6% |
| Tiendas por departamentos | 8,6% |
| Bencina | 7,5% |
| Avances en efectivo | 5,9% |
| Restaurantes | 5,1% |
| Servicios básicos | 4,9% |
| Farmacias | 4,3% |
| Clínicas | 3,2% |
| Líneas aéreas | 2,8% |
| Vestuario | 2,5% |
| Servicios financieros | 2,1% |
| Pago de impuestos | 1,9% |
| Colegios | 1,5% |
| Reparación y acces. Autos | 1,3% |
| Educación superior | 1,2% |
| Compraventa autos | 0,8% |
| Viajes | 0,8% |
| Otros | 27,2% |

Tabla 14: Estructura de consumo por rubros clientes Alto consumo

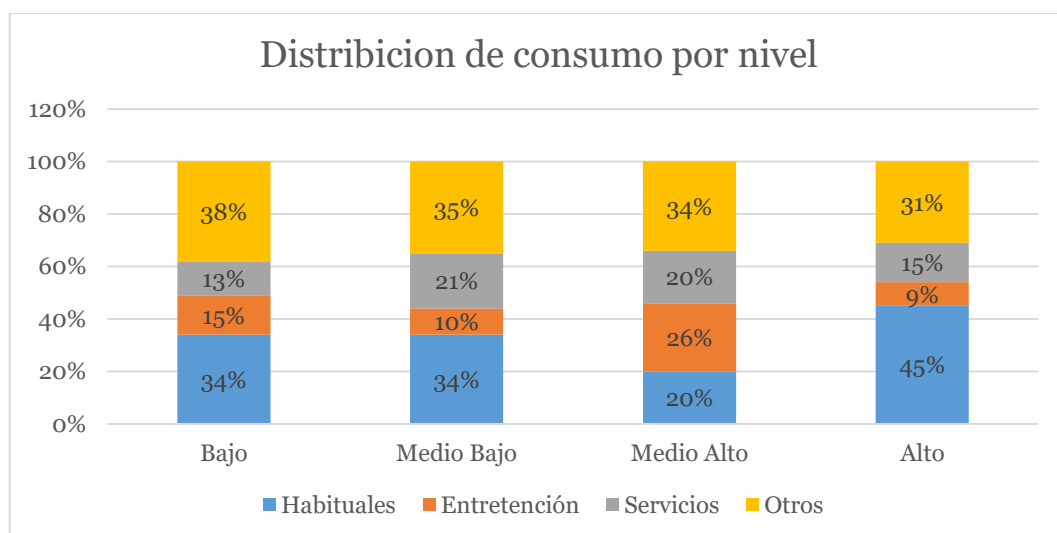
Como se puede observar en la tabla, los clientes de Alto consumo utilizan su tarjeta principalmente en los rubros habituales, destacando supermercados, tiendas por departamento y bencina. Por otro lado, el porcentaje de gasto en avances de efectivo es menor que en los otros grupos de clientes. Todo esto tiene relación con las desviaciones estándar presentadas en el anexo F las cuales son menores que en otros grupos y menores aun en los rubros habituales.

Lo anterior es esperable considerando que los clientes de Alto consumo destacan principalmente en el número de compras que realizan, por sobre el monto promedio de cada una de ellas (a diferencia de los clientes de Medio Alto consumo).

A continuación se presenta un resumen de la información transaccional de los diferentes segmentos y las distribuciones de consumo de los diferentes segmentos en base a la división propuesta por la SBIF (Anexo C):

| Nivel de consumo | # Transacciones | Monto total transado | # Avances | Monto Avances |
|-------------------------|------------------------|-----------------------------|------------------|----------------------|
| Bajo | 3,6 | \$224.824 | 0,5 | \$77.106 |
| Medio Bajo | 11,7 | \$583.309 | 0,8 | \$118.260 |
| Medio Alto | 3,9 | \$1.355.981 | 1,3 | \$776.785 |
| Alto | 62 | \$2.368.518 | 1,4 | \$213.736 |

Tabla 15: Resumen información transaccional



*Ilustración 19: Distribución de consumo por nivel
Fuente: Elaboración propia*

Al ver los resúmenes y las distribuciones de consumo se puede notar que los segmentos de Bajo, Medio Bajo y Alto consumo transan principalmente en rubros habituales, mientras que los clientes de Medio Alto consumo utilizan su tarjeta mayoritariamente en rubros de entretenimiento.

Lo anterior está relacionado a la cantidad de compras de cada nivel, donde se puede notar que los clientes de Medio Alto consumo realizan menos de la mitad de las transacciones que hacen los clientes de Medio Bajo, sin embargo, lo hacen por montos muy superiores.

Con respecto a los avances, es interesante notar que el grupo de Medio Alto consumo es el que más los realiza, sin embargo, el grupo que más gasta en este ítem es el de Medio Alto consumo.

En base a lo anterior se pueden obtener algunas conclusiones relevantes, las cuales se describen a continuación:

- En primer lugar, la distribución de consumo de los clientes de Medio Alto consumo es completamente diferente a las otras pues se basa en pocas transacciones con altos montos, mientras que los otros grupos siguen un patrón que indica que, a mayor frecuencia, mayor consumo.
- En base a lo anterior se considera que pasar del grupo de Bajo al de Medio Bajo o Alto implica una profundización o aumento del nivel de consumo, mientras que hacerlo del grupo Medio Alto al Alto implica un cambio en la estructura.
- Por otro lado, los grupos de clientes consumen en rubros diferentes y lo hacen con diferentes tipos de pago, por lo que los modelos de propensión que se presentan a continuación son diferentes para cada nivel de consumo

6.2. Modelos de propensión

Para la construcción de los modelos de propensión se utiliza nuevamente el proceso KDD, esta vez poniendo especial énfasis en los pasos de transformación de variables, minería de datos y análisis de resultados.

Cabe destacar que para la generación de los modelos se utilizan transiciones, las cuales cuentan con tres periodos móviles que se definen a continuación:

- Periodo 2: Es el periodo en el que se mide si el cliente aumentó su consumo. Como lo que se busca es predecir este aumento, no se utilizan variables independientes de este periodo ya que al momento de hacer la predicción real, estas no se conocerán.
- Periodo 1: Es el periodo previo al aumento de consumo, es decir, si el nivel de consumo del periodo 2 es mayor al del periodo 1, entonces el cliente aumentó su consumo.
- Periodo 0: Es el periodo previo al periodo 1, es decir, no se relaciona directamente con el aumento de consumo, sin embargo, se utiliza con el fin de generar variables independientes y de incluir la historia transaccional del cliente.

De esta manera, de los seis periodos del estudio se obtienen cuatro transiciones válidas, que se detallan en la sección de minería de datos más adelante.

6.2.1. Selección de datos y pre procesamiento

Para la creación de las variables independientes de los modelos de propensión se selecciona la información transaccional de los clientes, considerando tanto las compras en los diferentes rubros como los tipos de pago de cada transacción. Además se utiliza la información demográfica disponible y datos sobre las cuentas de los clientes. En esta última categoría se incluyen los tipos de tarjeta, cupos, cantidad de cuentas, etc.

Con respecto a la variable dependiente, para su creación se utilizan los niveles de consumo de cada cliente en cada periodo del estudio, los cuales se ordenan de menor a mayor asignándose un valor de 0 a 4.

En esta etapa no se observan valores faltantes, ya que la información se construye a partir de la base de datos de transacciones de los clientes y por lo tanto si un cliente no tiene transacciones, se considera que consume 0. Por otro lado, no se tratan outliers pues de esa manera se permite que los clientes tengan comportamientos marcados en ciertos rubros.

6.2.2. Transformación de variables

En esta etapa se generan las variables independientes con las que se predice el aumento de consumo de los clientes. Para esto se crean diferentes tipos de variables:

- Variables *monto* y *número*: Resumen la información transaccional de los clientes considerando el número y monto total consumido durante el periodo de cuatro meses en cada rubro, agrupación de rubros y tipos de pago. Además se crean otras variables

como la cantidad de semanas en que consume un cliente o la cantidad de rubros en que lo hace.

- Variables *dummies*: Toman el valor 1 si el cliente incurre en ciertos comportamientos. A modo de ejemplo están las variables *dummies* que indican si el cliente consume en rubros específicos, si tiene compras por internet o si tiene un tipo de tarjeta específico.
- Variables *cambio*: Buscan rescatar los cambios en la utilización de la tarjeta por parte de los clientes entre los periodos 0 y 1. Estas variables pueden ser de tipo binaria o numérico, por ejemplo pueden tomar el valor 1 si el cliente aumenta el cupo de su tarjeta o tomar directamente el valor del aumento del cupo.

Con respecto a la variable dependiente, esta toma valores de 1 o 0, de forma que si un cliente pasa a un nivel de mayor consumo en una transición válida, la variable Aumento toma el valor 1. Por otro lado, si el cliente mantiene su nivel de consumo o lo disminuye, la variable toma el valor 0.

Cabe destacar que si bien podría ser interesante definir la variable dependiente según el nivel de consumo al que llega el cliente, esto generaría dificultades pues son pocos los clientes que aumentan más de un nivel y aún menos los que lo hacen más de dos. A modo de ejemplo se considera el caso del nivel de consumo Bajo, desde el cual sólo el 1% de los clientes llega al nivel de consumo Alto. Este porcentaje es aún menor para los clientes inactivos y por lo tanto los modelos no contarían con suficiente información para determinar la importancia de las variables independientes. El porcentaje de aumentos hacia cada nivel se puede ver en el anexo Ñ.

Por otro lado, no se realiza un modelo para los niveles Medio Alto y Alto consumo pues la tasa de aumentos del primero es sólo un 1,8% y por otro lado, la rentabilidad de ambos niveles es muy similar, siendo mayor la de los clientes de Medio Alto consumo. Con respecto a los clientes de Alto consumo, si bien estos pueden ser incentivados, no existe un clúster al que puedan aumentar.

6.2.3. Minería de datos

Antes de correr los modelos para cada nivel de consumo, se divide la base de datos en dos partes, utilizando las cuatro transiciones válidas. Esto se hace pues con el primer grupo de datos se entrenan o calibran los modelos para obtener los parámetros, mientras que con el segundo grupo se valida que dichos parámetros sean capaces de predecir si los clientes aumentan su consumo.

La razón de la división por lo tanto es evitar que el modelo esté sobre ajustado a los datos con los que se construye. De esta manera las bases de datos se construyen de la siguiente forma:

Base de entrenamiento

Considera las 3 primeras transiciones válidas del estudio, es decir:

- Periodo 1 (dic 13-mar 14), periodo 2 (abr 14-jul 14) y periodo 3 (ago 14-nov 14)
- Periodo 2 (abr 14-jul 14), periodo 3 (ago 14-nov 14) y periodo 4 (dic 14-mar 15)

- Periodo 3 (ago 14-nov 14), periodo 4 (dic 14-mar 15) y periodo 5 (abr 15-jul 15)

Base de validación

Considera la última transición del estudio ya que se busca que sea lo más similar al periodo actual, es decir:

- Periodo 4 (dic 14-mar 15), periodo 5 (abr 15-jul 15) y periodo 6 (ago 15-nov 15)

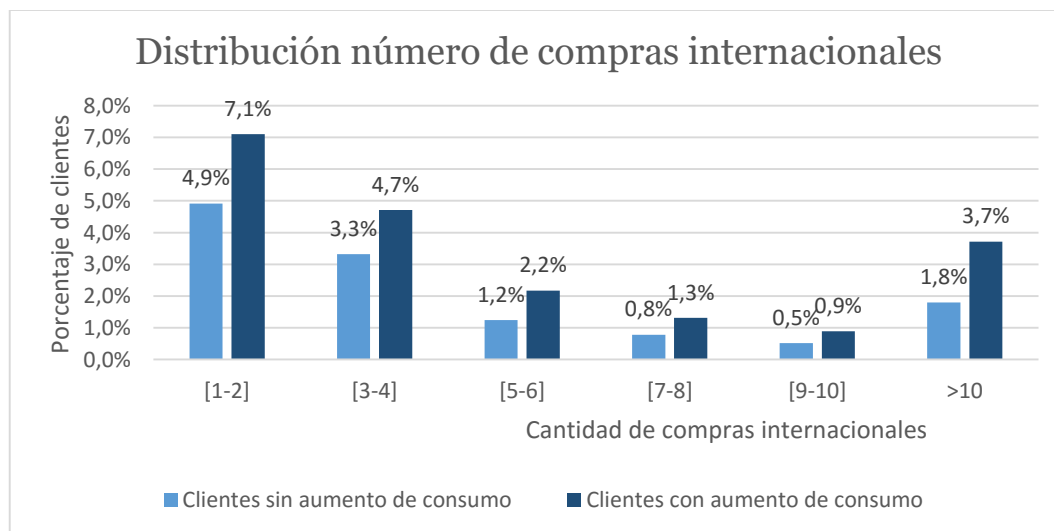
De esta forma se puede ver que la validación de los modelos se realizará con las predicciones del periodo seis y utilizando variables independientes de los periodos cuatro y cinco.

6.2.3.1. Modelo de propensión clientes Inactivos

Para la confección del modelo de propensión al aumento de consumo de los clientes Inactivos se cuenta con una base de datos de entrenamiento de 25.580 registros, de los cuales el 35% aumentan su consumo, por lo que no es necesario balancear la base de datos.

Se presenta a continuación un análisis descriptivo de las principales variables a estudiar, con el fin de entender que efectos pueden tener en la variable dependiente y la forma en que se relacionan con ella.

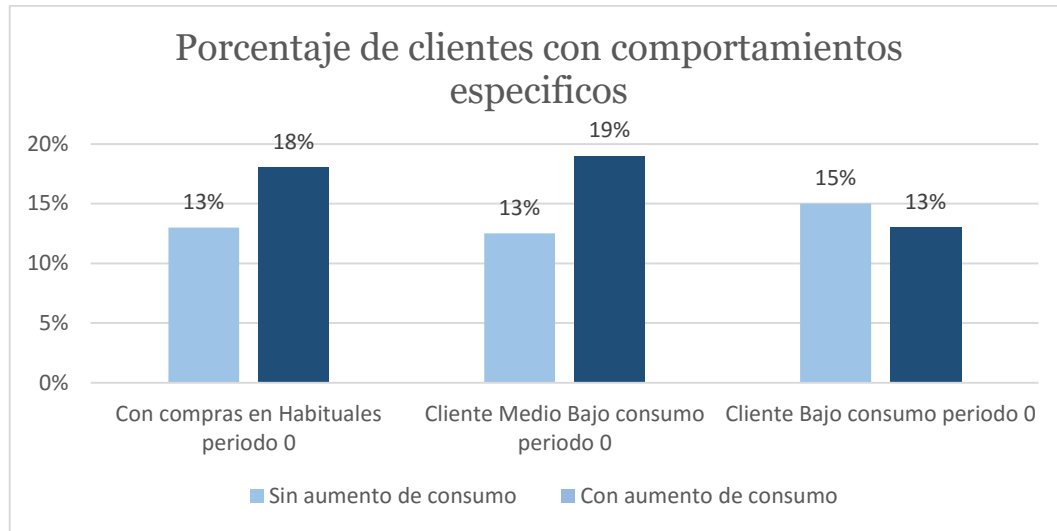
Compras internacionales durante periodo de inactividad



*Ilustración 20: Compras internacionales clientes Inactivos
Fuente: Elaboración propia*

Con respecto al gráfico es necesario aclarar que no se muestra a los clientes sin compras internacionales debido a que son la mayoría y distorsionan la escala del histograma, sin embargo, de igual manera se puede notar que los clientes que aumentan su consumo realizan más compras internacionales durante su periodo de inactividad. Se debe recordar además que el concepto de inactivo sólo incluye transacciones nacionales.

Comportamientos específicos



*Ilustración 21: Porcentaje de clientes según comportamiento clientes Inactivos
Fuente: Elaboración propia*

Como se puede observar en los gráficos, los clientes que se inactivan luego de tener un nivel de consumo Medio Bajo tienen a aumentar más que aquellos que tuvieron un nivel Bajo. Por otro lado, con respecto a las compras en rubros habituales, durante el periodo 0 son más los clientes con compras en estos rubros y que sí aumentan su consumo que los que compran y no aumentan. Como se puede ver en el anexo E, la diferencia de los clientes provenientes del nivel Bajo no es significativa, pero ayuda a ilustrar su comportamiento.

Tal como se mencionó en la metodología, en este trabajo se realizan tres tipos de modelos de propensión, los que luego son comparados en base al análisis ROC. Los resultados de los modelos se presentan a continuación:

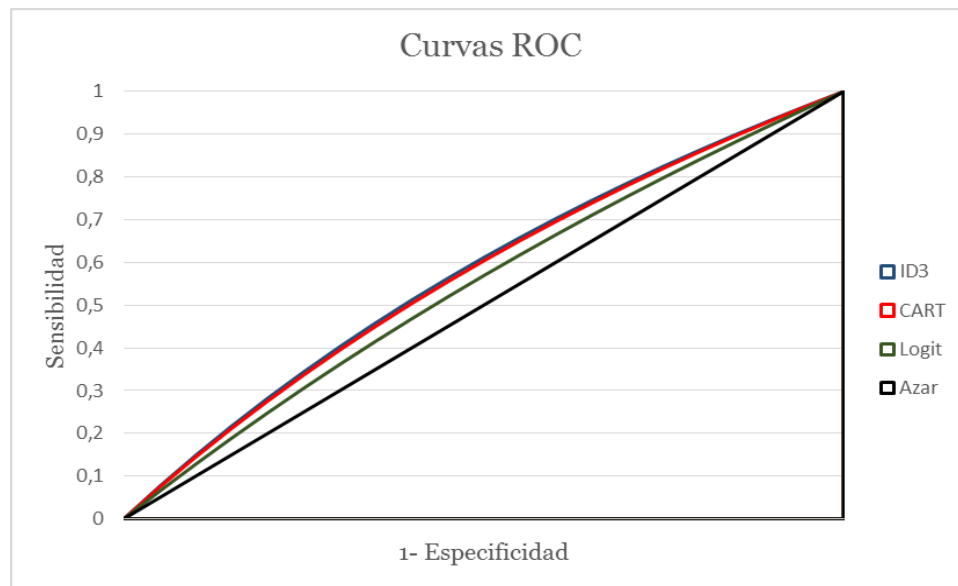


Ilustración 22: Curva ROC regresión logística clientes Inactivos

De las curvas ROC se puede apreciar que los modelos basados en árboles de decisión permiten predecir de mejor manera que la regresión logística a aquellos clientes que aumentan su consumo. Para determinar de mejor manera qué modelo utilizar, se presentan los indicadores principales del análisis ROC:

| | Regresión logística | CART | ID3 |
|-----------------------------------|--------------------------------|-------------|------------|
| Especificidad | 56% | 67% | 67% |
| Sensibilidad | 58% | 50% | 51% |
| Predicción No aumentos | 70% | 70% | 71% |
| Predicción aumentos | 46% | 46% | 48% |
| Predicción general | 57% | 61% | 61% |
| AUC | 59% | 61% | 62% |

Tabla 16: Indicadores ROC clientes Inactivos

De la tabla se puede notar que el modelo que permite predecir de mejor manera es el ID3, pues si bien tiene una menor sensibilidad que la regresión logística, en el resto de los indicadores es superior. La matriz de confusión del modelo de clientes Inactivos se encuentra en el anexo H.

Con respecto a las variables que influyen en el aumento de consumo de los clientes Inactivos, se puede decir que aquellas que aparecen más arriba en el árbol, tienen mayor importancia. En base a eso, las variables que se incluyen en los tres primeros niveles del árbol son:

- **Número de compras internacionales:** Es la variable más importante pues permite diferenciar entre clientes que están inactivos en todos los usos de la tarjeta de aquellos que la usan en viajes y que por lo tanto demuestran no estar fugados del banco. Luego, clientes con compras internacionales son más propensos a activarse al siguiente periodo.
- En segundo lugar, aparece el porcentaje de cupo utilizado al final del periodo de inactividad, donde aquellos clientes con cupos utilizados muy bajos son menos propensos a aumentar su consumo. Esto se puede deber a que son clientes que dejaron de consumir con la tarjeta y por lo tanto no la consideran como medio de pago válido.
- En tercer lugar, el modelo indica que aquellos clientes que durante el periodo o tuvieron un nivel de consumo Medio Bajo son más propensos a aumentar su consumo.
- Por otro lado, el árbol distingue entre aquellos clientes que realizaban compras en rubros habituales durante el periodo o de aquellos que no. Cabe destacar que no se puede decir que alguno de estos grupos sea más propenso, pues el efecto de esta variable depende la interacción con los niveles más bajos del árbol.
- Finalmente aparece como variable importante la cantidad de pagos (disminución de deuda) que realizan los clientes a la tarjeta durante su periodo sin consumo. Esto quiere decir que aquellos clientes que no compran pero que sí pagan su deuda utilizada, son más propensos a volver a comprar con la tarjeta.

Cabe destacar además que de acuerdo al modelo, aquellos clientes inactivos que cierran cuentas o cancelan sus contratos PAT, son los menos propensos a aumentar su consumo, pues son clientes que están terminando su relación con el banco.

Además de las variables principales, es interesante entender los perfiles más importantes de los clientes que aumentan su consumo, pues el banco podría incentivar a los clientes a tener comportamientos similares a los perfiles y por lo tanto aumentar la propensión al aumento de consumo.

En base a lo anterior, se determinan los tres principales perfiles de los clientes que aumentan su consumo:

1) En primer lugar están los clientes que durante el periodo 0 tuvieron un nivel de consumo Medio Bajo y realizaron al menos una transacción en rubros habituales. Estos clientes representan al 7,4% de los clientes inactivos.

2) En segundo lugar están aquellos clientes que no realizan compras internacionales, pero que en el periodo anterior tuvieron un nivel de consumo Medio Bajo y se mantienen pagando sus deudas con la tarjeta tanto durante el periodo 0 como durante la inactividad. El 8,1% de los clientes inactivos entran en este grupo.

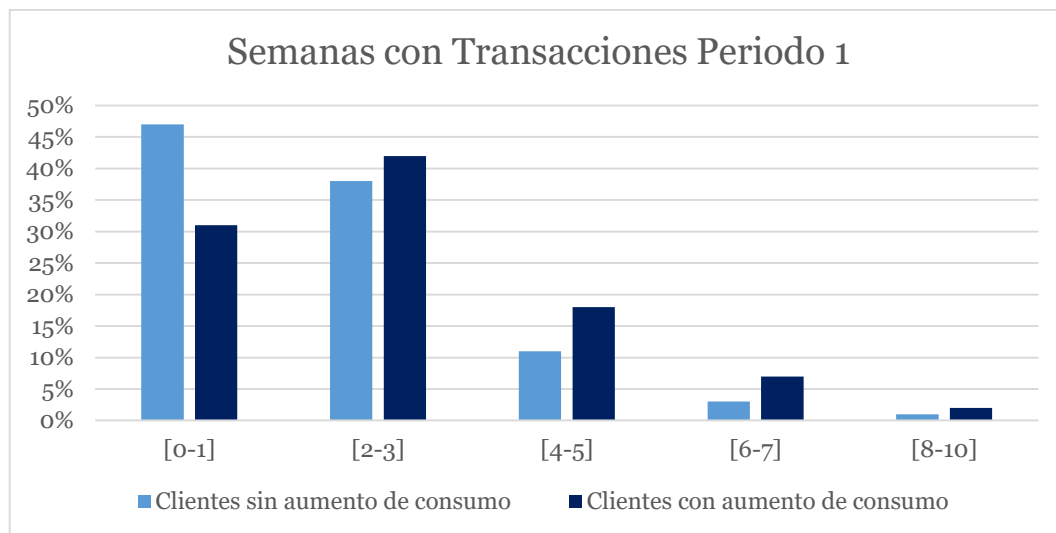
3) Están además aquellos clientes que no transan en rubros internacionales ni fueron del nivel de Medio Bajo consumo durante el periodo 0, pero que sin embargo, se mantienen pagando sus deudas con la tarjeta y tenían transacciones en más de dos semanas del periodo anterior. Cabe destacar que la cantidad de semanas habla de que tan recurrente era el cliente. El 12% de los clientes inactivos son de este grupo.

El árbol generado por el modelo ID3 se puede encontrar en el anexo I de este trabajo, mientras que las variables que utiliza se encuentran en el anexo U.

6.2.3.2. Modelo de propensión clientes Bajo consumo

La base de entrenamiento para la creación del modelo de clientes de Bajo consumo cuenta con 9.593 registros, de los cuales el 46% tienen aumento de consumo y por lo tanto es una base que no requiere ser balanceada. Con respecto a estos clientes, se presenta a continuación un análisis descriptivo de ciertas variables importantes para el modelo de propensión.

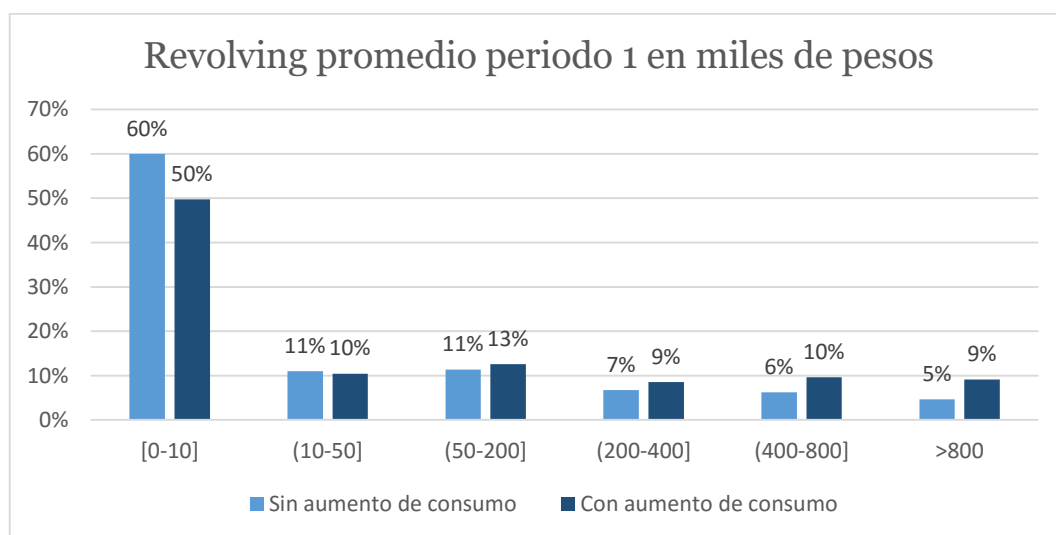
Cantidad de semanas con transacciones



*Ilustración 23: Semanas con transacciones clientes Bajo consumo
Fuente: Elaboración propia*

Del gráfico se puede notar que la gran mayoría de los clientes sin aumento de consumo realizan transacciones menos de cuatro semanas en el periodo, mientras que la mayoría lo hace sólo una semana. Por otro lado, los clientes con aumento de consumo utilizan su tarjeta más semanas durante el periodo de cuatro meses, existiendo un porcentaje significativo de ellos que la utiliza al menos seis semanas del periodo. Esto hace suponer que los clientes de Bajo consumo que aumentan el uso de la tarjeta, son clientes más acostumbrados a utilizarla como medio de pago.

Deuda de colocación

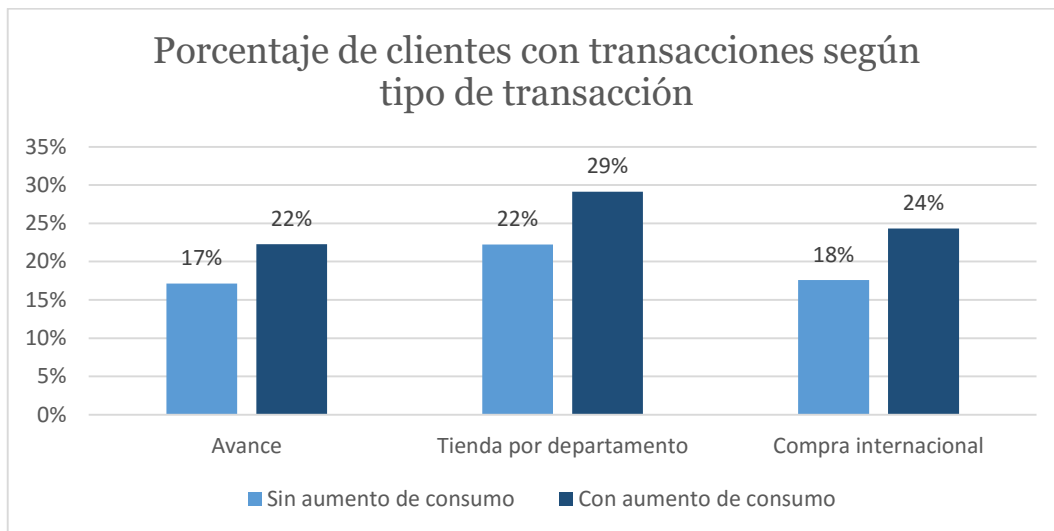


*Ilustración 24: Revolving promedio mensual periodo 1 Clientes Bajo consumo
Fuente: Elaboración propia*

Del histograma se puede apreciar que si bien las diferencias son pequeñas, los clientes que aumentan su consumo utilizan más la deuda de colocación. Esto puede deberse a que

estos clientes utilizan más formas de la tarjeta de crédito que aquellos que sólo la utilizan para pagar y no endeudarse.

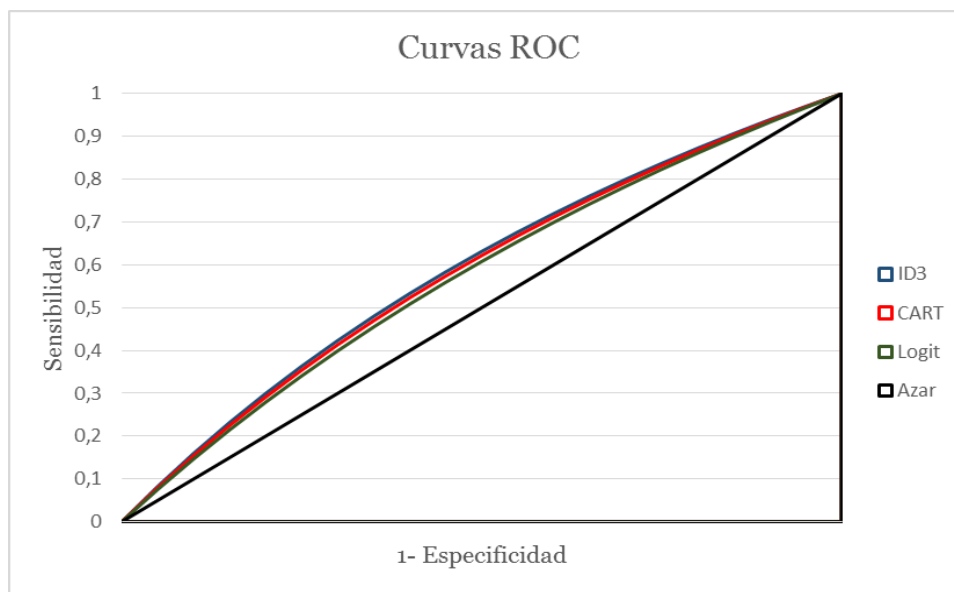
Tipos de transacción importantes



*Ilustración 25: Porcentaje de clientes con transacciones en tipos importantes de transacción
Fuente: Elaboración propia*

Como se puede observar en el gráfico, los clientes que aumentan su consumo utilizan más los avances y las compras internacionales. Así mismo son clientes que compran más en tiendas por departamento que aquellos que no aumentan su consumo. De esta manera, se espera que clientes con estas transacciones sean más propensos a aumentar su consumo ya que como se ve en el anexo E, todas las diferencias son significativas.

A continuación se presentan los resultados de los modelos de propensión para los clientes de Bajo consumo:



*Ilustración 26: Curvas ROC clientes Bajo consumo
Fuente: Elaboración propia*

Al igual que en el segmento de clientes inactivos, al ver las curvas ROC se puede notar que los modelos basados en árboles aportan mayor información que la regresión logística. Esto se puede confirmar al observar los principales indicadores:

| | Regresión logística | CART | ID3 |
|-------------------------------|----------------------------|-------------|------------|
| Especificidad | 60% | 65% | 64% |
| Sensibilidad | 63% | 61% | 62% |
| Predicción No aumentos | 63% | 63% | 64% |
| Predicción aumentos | 60% | 62% | 62% |
| Predicción general | 61% | 62% | 63% |
| AUC | 66% | 67% | 67% |

Tabla 17: Indicadores principales análisis ROC clientes Bajo consumo

Se puede notar del análisis ROC que el mejor modelo en cuanto al poder predictivo es el árbol ID3, sin embargo, este modelo utiliza principalmente variables no manejables para el negocio pues son de información del cliente durante el periodo o demográfica. Esto es importante pues cabe recordar que el objetivo principal de este trabajo es la determinación de acciones con las cuales incentivar el aumento de consumo, para lo cual se requieren variables manejables.

Con respecto a lo anterior, la regresión logística utiliza más información sobre rubros específicos y por lo tanto entrega información sobre qué acciones se debe incentivar en los clientes. En base a lo anterior se selecciona el modelo de regresión logística como el mejor para los clientes de Bajo consumo.

Por otro lado, los modelos logit asumen que la variable dependiente toma el valor 1 cuando la probabilidad del suceso es mayor a 0,5, sin embargo, esto no es necesariamente cierto. Para solucionar esto se encuentra el punto óptimo de corte en el que equilibran los dos errores que puede cometer el modelo: indicar que un cliente aumentará su consumo cuando no lo hará e indicar que un cliente no aumentará su consumo cuando en realidad sí lo hará.

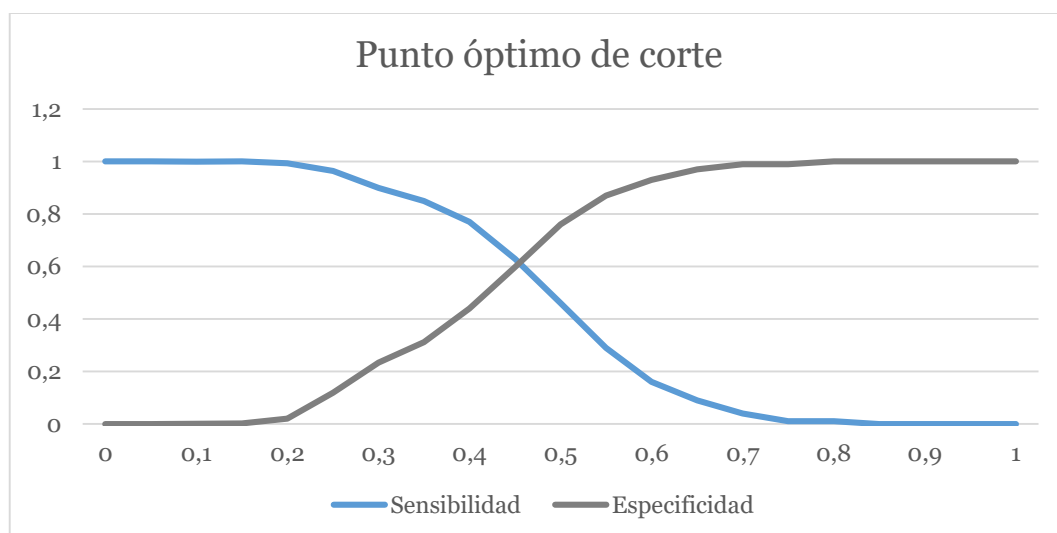


Ilustración 27: Sensibilidad vs. Especificidad, encontrando el mejor punto de corte
Fuente: Elaboración propia

Del gráfico se puede notar que el punto óptimo de corte es 0,45, es decir, clientes con probabilidades menores a este punto son considerados como clientes que no aumentarán su nivel de consumo. La matriz de confusión de la regresión logística con punto de corte 0,45 se puede encontrar en el anexo J.

A continuación se presentan las variables del modelo, ordenadas según su coeficiente estandarizado, el cual representa el impacto de aumentar en una desviación estándar cada variable y por esta razón permite comparar la importancia de variables medidas en diferentes dimensiones. Las variables presentadas incluyen sólo a aquellas cuyo coeficiente estandarizado es mayor a 0,08, sin embargo, la lista completa de variables se encuentra en el anexo K de este trabajo.

| Variable | Coefficiente | P-Valor | Coefficiente estandarizado | Medida |
|--|----------------------|---------|----------------------------|---------|
| Aumento periodo 0 | -0,64 | <0,01 | -0,27 | SI/NO |
| Intercepto | -1,11 | <0,01 | -0,16 | -- |
| Semanas con TRX periodo 1 | 0,09 | <0,01 | 0,16 | Número |
| Con avance | 0,30 | <0,01 | 0,12 | SI/NO |
| Semanas con TRX periodo 0 | 0,03 | <0,01 | 0,12 | Número |
| Revolving promedio mensual | 0,23 | <0,01 | 0,11 | B SI/NO |
| Cantidad de rubros periodo 0 | 0,05 | <0,01 | 0,1 | Número |
| Con compra internacional | 0,24 | <0,01 | 0,1 | SI/NO |
| Monto compra sin cuotas | $-5,8 \cdot 10^{-7}$ | <0,01 | -0,1 | Monto |
| Aumenta compra en cuotas | 0,20 | <0,01 | 0,1 | SI/NO |
| Aumenta N° de compras en rubros habituales | 0,02 | <0,01 | 0,09 | SI/NO |

Tabla 18: Variables principales regresión logística clientes Bajo consumo

Se puede observar que un cliente que en el periodo 0 estaba inactivo es menos propenso a aumentar su consumo, es decir, es difícil que aumente dos periodos seguidos. Además,

como se vio en el análisis descriptivo, mientras más sean las semanas con transacciones del cliente, más propenso es a aumentar su consumo.

Por otro lado, se puede apreciar que los clientes del grupo Bajo que utilizan más el pago sin cuotas son menos propensos a aumentar su nivel de consumo, mientras que aquellos clientes que aumentan su compra en cuotas tienen mayor probabilidad de aumentar su consumo al periodo siguiente.

Es importante destacar además el alto impacto que tienen los avances y las compras internacionales en este tipo de clientes, ya que ambos aumentan la propensión a aumentar el consumo. Finalmente, con respecto a las variables que no se presentan en esta tabla, se destaca el hecho de que clientes con tarjeta dorada son más propensos a subir su consumo que el resto.

Utilizando estas variables, se determinan los tres perfiles de clientes con alta probabilidad de aumentar su consumo:

1) Personas que no hayan aumentado en el periodo anterior (es decir, no hayan sido inactivos) y que hayan realizado al menos un avance con su tarjeta de crédito durante el periodo 1. El 15,6% de los clientes de Bajo consumo son de este perfil.

2) Clientes que realizaron al menos un avance, consumieron al menos en cuatro semanas del periodo de Bajo consumo y compraron al menos una vez en cuotas. Lo último es importante pues al ser clientes que vienen del segmento de Inactivos, estos clientes están aumentando su compra en cuotas. El 4% de los clientes de Bajo consumo se comportan de esta manera.

3) Finalmente están los clientes que realizan compras internacionales y compras en grandes tiendas (o tiendas por departamento). El 6,7% de los clientes del nivel Bajo realizaron estas acciones durante el periodo.

6.2.3.3. Modelo de propensión clientes Medio Bajo consumo

Para la generación del modelo de los clientes de Medio Bajo consumo se cuenta con una base de datos de entrenamiento de 63.427 registros, sin embargo, sólo 9,7% de ellos aumentan su consumo y por lo tanto es necesario balancear la base. Para esto se selecciona una muestra aleatoria considerando todos los registros con aumentos de consumo tal que representen el 30% de la base. Además se impone la restricción de que se debe mantener la proporción de clientes que bajan a Inactivos, a Bajo consumo y los que se mantienen en el nivel Medio Bajo. De esta forma la base balanceada queda con 20.610 registros únicos.

Es importante señalar que el grupo de clientes de Medio Bajo consumo es el más importante para el banco por dos razones:

- En primer lugar, estos clientes representan el 54% de la banca personas y por lo tanto al descubrir como incentivarlos se puede atacar a la mayoría de los clientes del banco.
- En segundo lugar, la diferencia promedio de consumo entre los clientes de Medio Bajo y Medio Alto es la más alta y por lo tanto para pasar a un nivel mayor deben aumentar más su consumo.

A continuación se presenta un análisis descriptivo de las principales variables dependientes utilizadas en los modelos de propensión:

Número de semanas con transacciones en periodo 0

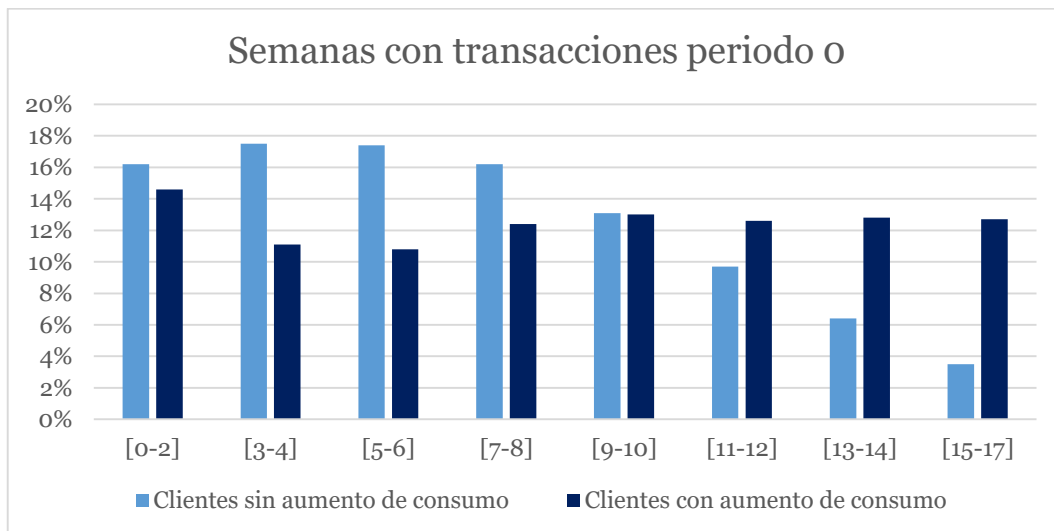


Ilustración 28: Cantidad de semanas con transacciones clientes Medio Bajo consumo
Fuente: Elaboración propia

Al comparar la distribución de semanas con transacciones de ambos tipos de clientes se puede notar que los clientes con aumento de consumo tienen transacciones durante un mayor número de semanas que los clientes sin aumento. Esto es importante pues permite suponer que al incluir esta variable en el modelo de propensión, esta tendría un $\beta > 0$ y que de esta forma, al incentivar a los clientes a comprar durante todas las semanas, se podría incentivar el aumento de consumo. A pesar de lo anterior, esta información debe ser corroborada en el modelo.

Cantidad de rubros en los que compra un cliente

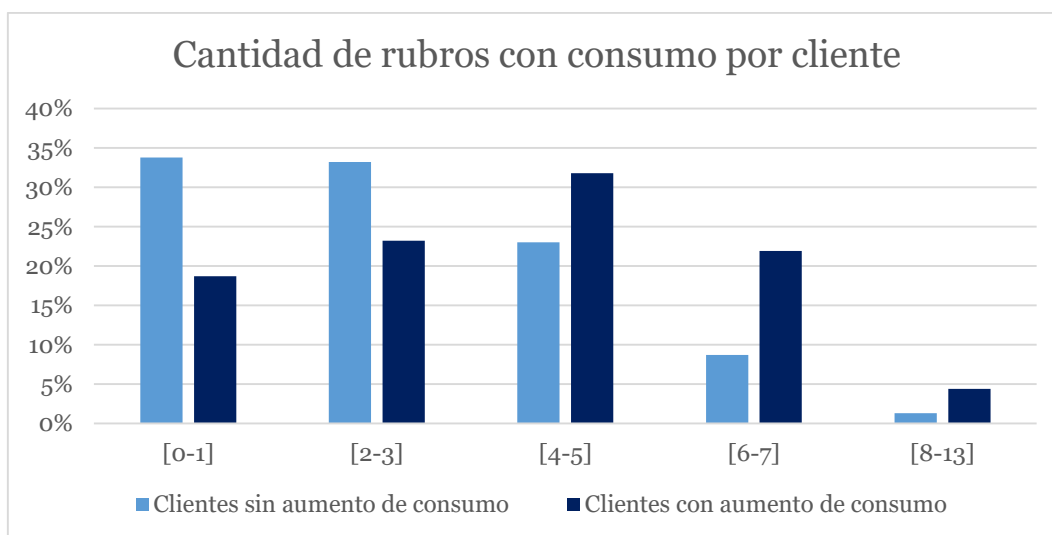
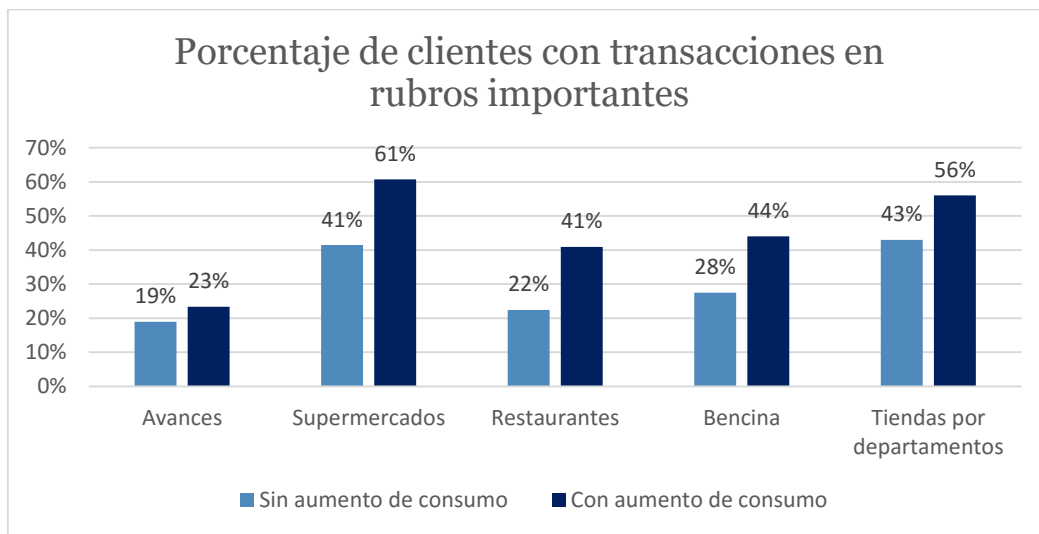


Ilustración 29: Cantidad de rubros en los que consumen los clientes de Medio Bajo consumo
Fuente: Elaboración propia

Al comparar a los clientes sin aumento de consumo a los con aumento, se puede notar que nuevamente los últimos tienen transacciones en un mayor número de rubros, donde la mayoría compra en al menos cuatro, mientras que los clientes sin aumento de consumo no pasan este número en su mayoría. De esta forma se espera que al aumentar el número de rubros en los que consume un cliente, aumente su propensión a incrementar su nivel de consumo.

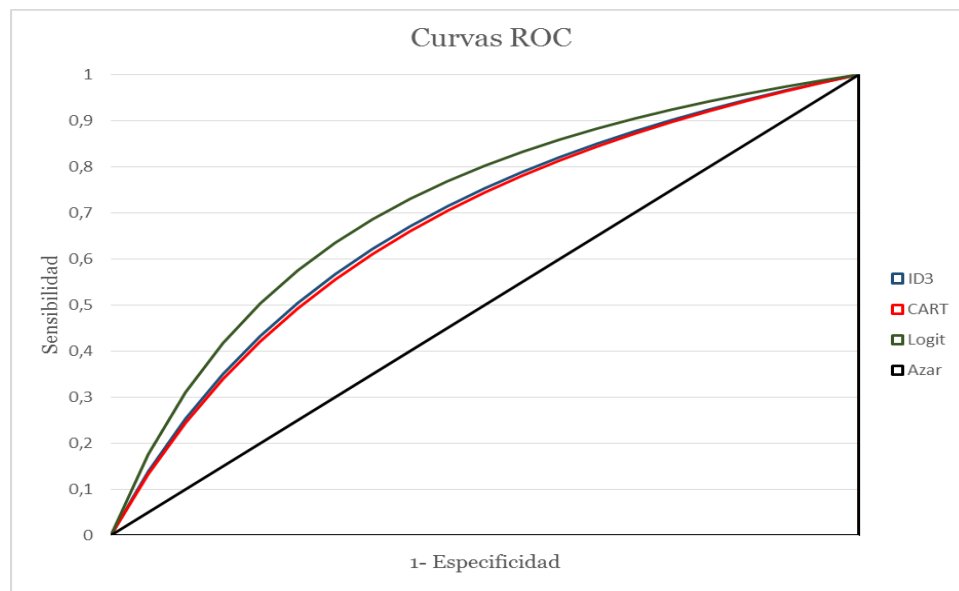
Rubros importantes



*Ilustración 30: Porcentaje de clientes con transacciones en principales rubros
Fuente: Elaboración propia*

Del gráfico nuevamente se puede notar que los clientes con aumento de consumo tienen más transacciones durante el periodo anterior (periodo 1) en los rubros importantes. Por otro lado, el rubro de Supermercados demuestra ser importante ya que es en el que se aprecia la mayor diferencia y además, tiene un alto nivel de transaccionalidad y por lo tanto hay suficientes datos para que el modelo pueda entrenarse. Porcentualmente hablando, la mayor diferencia se aprecia en el rubro de Restaurantes, por lo que se espera que sea un rubro inductor de aumento de consumo. En el anexo E se puede apreciar que las diferencias son estadísticamente significativas.

A continuación se presentan los resultados del modelo de propensión para los clientes de Medio Bajo consumo:



*Ilustración 31: Curvas ROC modelos Medio Bajo consumo
Fuente: Elaboración propia*

A diferencia de los segmentos anteriores, al comparar las curvas ROC de los modelos, se puede notar que la regresión logística obtiene mejores resultados que los modelos de árboles de decisión. Para confirmar esto se observan los principales indicadores del análisis ROC:

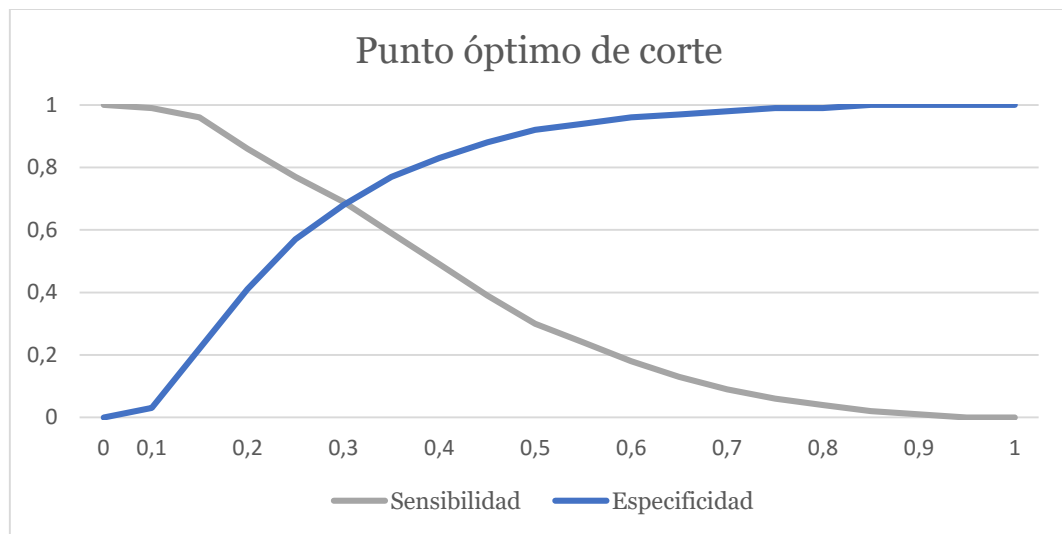
| | Regresión logística | CART | ID3 |
|-------------------------------|----------------------------|-------------|------------|
| Especificidad | 69% | 88% | 87% |
| Sensibilidad | 68% | 37% | 40% |
| Predicción No aumentos | 95% | 93% | 93% |
| Predicción aumentos | 19% | 24% | 24% |
| Predicción general | 69% | 83% | 83% |
| AUC | 74% | 70% | 72% |

Tabla 19: Indicadores principales análisis ROC clientes Medio Bajo consumo

De la tabla se puede observar que si bien la regresión logística no es el modelo con mejor poder predictivo general, es el que tiene los indicadores más semejantes y por lo tanto obtiene buenos valores tanto en sensibilidad como en especificidad.

Por otra parte cabe destacar que los tres modelos son capaces de predecir correctamente cuando los clientes no aumentarán su consumo, lo que es importante si se recuerda que cerca del 90% de los clientes del segmento Medio Bajo tienen este comportamiento.

De esta forma se determina que el mejor modelo para predecir los aumentos de consumo en los clientes de Medio Bajo consumo es la regresión logística. Luego, para encontrar el punto de corte del modelo, se presenta el siguiente gráfico:



*Ilustración 32: Sensibilidad vs. Especificidad, encontrando el mejor punto de corte
Fuente: Elaboración propia*

Del gráfico se puede observar que el punto de corte que balancea de mejor forma la sensibilidad y la especificidad del modelo es 0,305, es decir, el modelo indica que cualquier cliente con probabilidad de aumentar mayor a este número aumentará su consumo. La matriz de confusión del modelo se encuentra en el anexo L de este trabajo.

Finalmente se presentan las variables que se utilizan en el modelo, su importancia y el impacto que tienen con respecto al aumento de consumo de un cliente. Por otro lado cabe destacar que se presentan sólo las variables cuyo coeficiente estandarizado sea mayor a 0,1, sin embargo, en el anexo M puede encontrarse la tabla completa.

| Variable | Coefficiente | P-Valor | Coefficiente estandarizado | Medida |
|---------------------------|---------------------|----------------|-----------------------------------|---------------|
| Intercepto | -2,20 | <0,01 | -0,95 | -- |
| Monto transado periodo 0 | $3,1 \cdot 10^{-7}$ | <0,01 | 0,31 | Monto |
| Aumento periodo 0 | 0,66 | <0,01 | 0,23 | SÍ/NO |
| Cantidad de rubros | 0,10 | <0,01 | 0,21 | Número |
| Con avance | 0,45 | <0,01 | 0,18 | SÍ/NO |
| Nº compras "Otros rubros" | 0,04 | <0,01 | 0,18 | Número |
| Cupo utilizado final | -0,51 | <0,01 | -0,17 | % |
| Con TRX en Supermercado | 0,29 | <0,01 | 0,14 | SÍ/NO |
| Con tarjeta GOLD | -0,29 | <0,01 | -0,13 | SÍ/NO |
| Semanas con TRX periodo 0 | 0,03 | <0,01 | 0,12 | Número |
| Con compra por internet | 0,25 | <0,01 | 0,12 | SÍ/NO |

Tabla 20: Principales variables modelo Medio Bajo consumo

Como se puede ver en la tabla las dos variables más importantes tienen relación al periodo 0. En el caso del monto transado, este indica qué tan activo era el cliente previo a evaluar su aumento de consumo, mientras que un cliente que aumentó su consumo en la transición anterior es más propenso a hacerlo de nuevo.

Por otro lado, se confirma lo observado en el análisis descriptivo ya que mientras mayor sea el número de rubros en los que compra un cliente, mayor será la propensión a aumentar su consumo. Este resultado se repite con las semanas de compra en el periodo anterior. Además, el hecho de que el cliente realice avances o compras en supermercados lo hace más propenso a aumentar su consumo.

Además de lo anterior se puede notar que los clientes con tarjetas doradas son menos propensos a aumentar su consumo, lo cual es esperable si se considera que estos clientes en general tienen menores ingresos y por lo tanto es menos probable que tengan niveles tan altos de consumo. Finalmente en la tabla se puede notar que aquellos clientes con compras por internet son más propensos a aumentar su consumo, lo que calza con lo expuesto por Gensler et. Al (2012), quienes demuestran que clientes que compran por internet utilizan más su tarjeta [21].

Con respecto a las variables no mostradas destaca el hecho de que clientes con transacciones en Restaurantes o con aumentos de cupo son más propensos a aumentar su consumo. Además, con respecto a la bencina, lo más importante no es si el cliente consume o no, sino que cuánto gasta en el rubro, lo que se repite para las compras sin cuotas.

Si se comparan los impactos de las variables entre los clientes de Bajo y Medio Bajo consumo, se pueden encontrar algunas diferencias interesantes:

- Cuando los clientes de Bajo consumo compran más sin cuotas, disminuyen su propensión a aumentar el consumo, sin embargo, para los clientes de Medio Bajo consumo esto es al revés.

- Para los clientes de Bajo consumo es difícil aumentar de consumo si ya lo hicieron en el periodo anterior, sin embargo, los clientes de Medio Bajo consumo son más propensos si ya aumentaron anteriormente.
- Clientes de bajo consumo con tarjeta Gold son más propensos a aumentar su consumo, sin embargo, tener tarjeta Gold en Medio Bajo consumo disminuye la propensión al aumento de consumo.

Además de estas existen otras diferencias menos notorias como el impacto de tener compras por internet o compras internacionales.

En base a lo anterior, se generan tres perfiles de clientes con alta propensión a aumentar su consumo, de manera de entender qué hacen:

1) En primer lugar están los clientes realizan transacciones en restaurantes y tienen transacciones en al menos cuatro rubros durante el periodo 1. Este grupo representa al 18% de los clientes de Medio Bajo consumo.

2) En segundo lugar están los clientes con transacciones en al menos cuatro semanas del periodo 0, que realizan compras principalmente en supermercados, bencina y restaurantes. Estos son los clientes denominados por el banco como habituales. Este perfil representa al 22% de los clientes del segmento.

3) El tercer perfil con alta probabilidad de aumentar su consumo es el de los clientes que realizan transacciones en avances de efectivo y compras por internet y supermercados. Estos clientes son aquellos que utilizan la tarjeta de diferentes formas y no sólo en los rubros comunes y representan al 6% del nivel Medio Bajo.

6.2.4. Análisis de resultados

Si bien la idea de los modelos de propensión es identificar a los clientes que aumentarán su consumo, el hecho de entender cómo las diferentes variables afectan a este comportamiento permite generar recomendaciones al banco sobre cómo incentivar a los clientes a aumentar su consumo. En base a esto se presenta a continuación un análisis que incluye tanto la calidad de los modelos a nivel general, como recomendaciones para el banco según cada grupo de clientes.

Al ver los resultados del modelo de clientes Inactivos se puede notar que el poder predictivo no sobrepasa al 61% y por lo tanto no representa una gran mejora con respecto al azar. La razón principal de esto es que los modelos utilizan principalmente información transaccional, sin embargo, los clientes inactivos no tienen esta información durante un periodo completo. De esta manera, a diferencia de los otros grupos de clientes, el modelo de Inactivos se alimenta en gran parte de la información del periodo 0, la que no alcanza a predecir correctamente los aumentos de consumo.

Con respecto a los clientes de Bajo consumo, la situación anterior vuelve a afectar pues si bien estos clientes cuentan con más información transaccional, consumen poco y por lo tanto las variables independientes no logran predecir correctamente. A pesar de lo anterior, destaca el hecho de que el modelo de regresión logística utiliza más información

transaccional que el resto y por lo tanto permite identificar de mejor manera acciones con las que incentivar a los clientes.

Por último los resultados del modelo de clientes Medio Bajo consumo demuestran ser mejores que los anteriores pues permiten diferenciar con más certeza entre clientes con y sin aumento de consumo. A pesar de lo anterior, se puede ver que el error de tipo I o los falsos positivos representan cerca de un 28% de los clientes del grupo. Esto se debe en parte a que la base de datos original es desbalanceada y por lo tanto el modelo indica que más clientes aumentarán su consumo.

Dado lo anterior, cabe preguntarse si los resultados de los modelos pueden ser mejorados de alguna manera, para lo cual se plantea la recolección de nuevos datos no necesariamente transaccionales, con el fin de tener más información sobre los clientes Inactivos y de Bajo consumo. Algunas recomendaciones de datos a recolectar son:

- Variables demográficas: Kim y DeVaney (2001) demuestran que las personas casadas son más propensas a ser buenos clientes de la tarjeta de crédito y que además, mientras mayor sea el tamaño familiar, mayor la propensión a ser un buen cliente [22]. Debido a esto se considera que estos datos podrían aportar a mejorar los modelos de propensión, sin embargo, la recolección de estos datos podría ser costosa para el banco.
- Ingreso de los clientes: El ingreso es una de las variables más importantes que determinan la estructura de consumo de los clientes [6] y clientes con mayor ingreso consumen más con tarjeta de crédito [22].
- Otros medios de pago: Si bien los distintos medios de pago del banco son áreas de negocio diferentes, podría ser útil conocer la evolución de los clientes Inactivos con otros medios de pago como la tarjeta de débito o los cheques. De esta manera se podría saber además si el cliente está Inactivo en el banco o sólo con la tarjeta de crédito. Rocco (2010) demuestra la importancia de estas variables en el aumento de consumo [15].
- Deuda sistema bancario: La SBIF cuenta con información sobre la deuda total de los clientes en el sistema bancario. Esta variable podría aportar información sobre cuán fácil es para un cliente seguir consumiendo o qué tan acostumbrado está a comprar con crédito. Esta variable no se utiliza en este trabajo pues al trabajar con una consultora, no se tienen los RUT reales de los clientes y por lo tanto no se podría saber qué cliente tiene qué deuda. Al igual que la información de otros medios de pago, Rocco (2010) demuestra la importancia de estas variables.
- Incentivos externos: Una variable que podría influir en la decisión de los clientes de aumentar su consumo son las campañas promocionales a las que son expuestos por parte del banco. De esta forma además se podría responder la pregunta de qué grupo de clientes es más propenso a interactuar con las promociones.

Por último, como se menciona al comienzo de este trabajo, lo que se busca es incentivar a los clientes a aumentar su consumo, para lo cual se debe responder las interrogantes sobre a quién incentivar y con qué acciones hacerlo. A continuación se presentan las recomendaciones finales para el banco con respecto a ambos temas.

6.2.4.1. Recomendaciones clientes Inactivos

Si bien los resultados del modelo no permiten predecir con certeza si un cliente aumentará su consumo, la importancia de las variables sí entrega información relevante sobre qué características tienen los clientes más propensos a incrementar su nivel de consumo y por lo tanto, de qué maneras se podría incentivar a los clientes a para que se activen.

Como primera recomendación se propone generar promociones tanto a clientes que aumentarán como a aquellos que no lo harán, ya que tener clientes inactivos por ocho meses puede ser riesgoso pues aumenta la posibilidad de fuga. Debido a lo anterior se recomienda generar dos tipos de promociones:

- Dirigidas a clientes que no aumentarán su consumo: estas deben ser más importantes que el promedio pues buscan generar un cambio de comportamiento en los clientes inactivos.
- Dirigidas a clientes que aumentarán su consumo: buscan asegurar que el cliente efectivamente se active y por lo tanto se propone que sean más baratas que el promedio.

En cuanto a transacciones específicas, el banco debería incentivar las compras internacionales de los clientes inactivos y la única forma de hacer esto es disminuyendo temporalmente la comisión especial que tienen estas compras, ya que no se cuenta con las herramientas para incentivar rubros internacionales específicos.

Por otro lado, clientes con mayores niveles de cupo utilizado son más propensos a activarse y por lo tanto se recomienda generar promociones especiales dirigidas a ellos. De igual forma se destaca que si un cliente realiza pagos de su deuda durante el periodo de actividad, se le deben enviar promociones o acciones de marketing para reactivarlo.

Al analizar el árbol completo, se puede notar que clientes de comunas más adineradas, con cupos más altos o con mejores tipos de tarjeta, son menos propensos a aumentar su consumo si están en el nivel de consumo Inactivos. Esto es importante ya que estas tres variables están relacionadas al ingreso de los clientes y por lo tanto podrían indicar que aquellos de mayor ingreso son menos propensos a activarse. Cabe recordar que si bien en este trabajo no se conocen los ingresos de los clientes, el banco sí cuenta con esta información y por lo tanto podría generar promociones especiales a clientes de ingresos medios y bajos.

Además de lo anterior, existe otra acción no transaccional que induce el aumento de consumo: los cambios de tipo de tarjeta. Este cambio implica entregarle a un cliente una tarjeta con más beneficios sin la necesidad de pagar con la tarjeta y por lo tanto tiene costos directos para el banco. Por esta razón se recomienda cambiar la tarjeta sólo a los clientes que lo soliciten y no realizar campañas masivas sobre esto.

Con respecto a los clientes que cancelan sus contratos PAT o cierran alguna de sus cuentas, se recomienda tomar acciones instantáneas sobre esto, debido que esta podría ser la primera señal de que el cliente no pretende seguir usando su tarjeta. Estas acciones

pueden ir desde promociones personalizadas, regalos de puntos del club de lealtad o llamados del ejecutivo de cuenta.

Por último se destaca que el 57% de los clientes inactivos que aumentan su consumo pasan directamente al nivel Medio Bajo de consumo y por lo tanto se recomienda generar promociones en supermercados, avances de efectivo y grandes tiendas, pues en esos tres tipos de compra transan principalmente los clientes de Medio Bajo consumo.

Cabe destacar que la última propuesta no induciría un aumento de consumo desde el periodo 1, pero sí aportaría a que los clientes comenzaran a utilizar su tarjeta en el periodo de activación.

6.2.4.2. Recomendaciones clientes Bajo consumo

Para aumentar el consumo de este grupo de clientes el banco debe considerar que lo que hace menos propenso a un cliente es el hecho de haber aumentado su consumo en el periodo anterior y por lo tanto se deben generar acciones específicas con estos clientes. Esto es importante pues indica que aquellos clientes Inactivos que aumentan su nivel de consumo, prefieren hacerlo de forma lenta y no aumentar durante periodos consecutivos, es decir, es un proceso de mediano plazo y por lo tanto no se deben enviar promociones agresivas continuamente a este tipo de clientes.

Por otro lado se puede notar que las semanas en que consume un cliente son de gran importancia tanto en el periodo 1 como en el periodo 0 y por lo tanto se deben generar promociones en esta línea. A modo de ejemplo se podría incentivar a los clientes a consumir al menos una vez a la semana a cambio de puntos del club de lealtad.

Al igual que en los clientes inactivos, el hecho de que un cliente tenga un alto nivel de revolving lo hace más propenso a aumentar su consumo, lo que va ligado al hecho de utilizar de distintas formas la tarjeta de crédito. Por esta razón se debe incentivar en los clientes de Bajo consumo las compras que generen este tipo de deuda con mayor probabilidad. Algunas de estas pueden ser:

- Compras grandes: Rubros como educación o líneas aéreas
- Pagos en cuotas
- Avances de efectivo

Estos dos últimos son doblemente importantes pues aquellos clientes que aumentan sus compras en cuotas y que realizan avances son más propensos a aumentar su consumo, por lo que se recomienda incentivar estos comportamientos.

Con respecto al tipo de tarjeta, se debe tener en cuenta que aquellos clientes con tarjeta dorada son más propensos a aumentar. Además, estos representan a la gran mayoría de los clientes del banco y por lo tanto se recomienda generar promociones especiales a ellos y a tarjetas platinum, pues estas no aumentan ni disminuyen la propensión. Además, como se puede ver del modelo, los clientes jóvenes que tienen Bajo consumo son menos propensos y por lo tanto las promociones deben ser diferenciadas según esta característica.

En cuanto a rubros específicos, se recomienda incentivar las compras en tiendas por departamento, los avances, las compras internacionales y los rubros habituales en general. Con lo anterior se puede responder la pregunta de con qué acciones incentivar a los clientes, sin embargo, falta responder a quién se debe incentivar.

Para responder a la pregunta se debe considerar que los clientes de este grupo ya están consumiendo y por lo tanto no corren el riesgo de no activarse (como el grupo anterior). Por esta razón se recomienda generar campañas promocionales sólo a aquellos clientes que el modelo indique que no aumentarán su consumo, para de esa forma incentivar que lo hagan. Cabe destacar que en este punto se utiliza el supuesto de que todos los clientes reaccionan de igual manera a las promociones, sin importar su propensión al aumento de consumo.

Con respecto a los clientes que el modelo indica que sí aumentarán, enviarle promociones significaría perder dinero pues esos clientes aumentarán su consumo de igual manera, sin embargo, se recomienda enviar recordatorios de la tarjeta y acumulación de puntos para así generar presencia de la tarjeta.

6.2.4.3. Recomendaciones clientes Medio Bajo consumo

Con respecto a los clientes de Medio Bajo consumo, se propone enviar promociones solamente a aquellos clientes que no vayan a aumentar su consumo de acuerdo al modelo. Se propone esto por las siguientes razones:

- En primer lugar estos clientes representan cerca del 90% de los clientes de Medio Bajo consumo y por lo tanto cualquier mejora puede ser significativa.
- En segundo lugar, los clientes de Medio Bajo consumo utilizan su tarjeta 3 veces al mes en promedio y por lo tanto no son “malos” clientes a los que haya que convencer de utilizar la tarjeta.
- Si los clientes que aumentarán su consumo de manera natural toman promociones, el banco incurre en un gasto innecesario.

Para estos, en primer lugar se recomienda generar promociones especiales a los clientes que durante el periodo anterior hayan aumentado su nivel de consumo, debido a que estos son más propensos a seguir aumentando. Esto se corrobora al notar que clientes que durante el periodo o consumieron más son más propensos a incrementar su nivel. Lo anterior es importante pues indica que se debe ser agresivo en cuanto al marketing con aquellos clientes que vengán en aumento con respecto a sus niveles de transaccionalidad.

Por otro lado el banco debiera poner especial énfasis en incentivar que los clientes realicen transacciones en diferentes rubros. Para esto se podrían generar promociones en que los clientes ganen puntos del club de lealtad por cada rubro diferente en el que consumen. De esta forma los clientes podrían acostumbrarse a utilizar su tarjeta para diferentes tipos de compras.

Actualmente las promociones del banco están enfocadas en gran parte en entregar la posibilidad de pagar en más cuotas, sin embargo, para este grupo de clientes esto sería contraproducente pues es justamente la compra sin cuotas la que aumenta la propensión

a subir el nivel de consumo. Por otro lado, los clientes tienden a pagar más en cuotas en rubros de compras altas como líneas aéreas o educación, sin embargo, los rubros o tipos de transacciones que se deben incentivar en los clientes de Medio Bajo consumo son los siguientes:

- Bencina: Se debe incentivar que los clientes transen mayores montos y no solamente un número mayor de veces. Para esto se pueden proponer metas de consumo a cambio de puntos en el club de fidelización.
- Avances: Se deben incentivar independiente del monto y número de transacciones
- Supermercados: Se debe incentivar que los clientes realicen al menos una compra en este rubro.
- Restaurantes: Se debe incentivar que los clientes realicen al menos una compra en este rubro con la tarjeta.

Se debe tener cuidado además con los tipos de tarjeta de los clientes, ya que si bien se debe incentivar a los clientes que no aumentarán su consumo, los clientes con tarjetas doradas son menos propensos a aumentar su consumo, lo que se puede deber a que el nivel Medio Bajo es su tope de consumo posible y por lo tanto no aumentarán aunque se los incentive. A pesar de esto, los clientes con este tipo de tarjeta representan al 70% de los clientes de Medio Bajo consumo y por lo tanto no se pueden dejar de lado. De la misma forma los clientes con tarjeta negra son más propensos y por tanto se deberían incentivar también.

Al igual que con el tipo de tarjeta se debe considerar que aquellos clientes con un mayor porcentaje de cupo utilizado tienen menos probabilidad de aumentar su consumo, lo que se puede explicar con que no pueden seguir consumiendo. Si a esto se le suma el hecho de que el aumento de cupo aumenta la propensión a incrementar el consumo, se recomienda analizar la idea de ofrecer aumentos de cupo a clientes con altos cupos utilizados.

Por otro lado se propone incentivar a los clientes a comprar por internet, debido a que esto aumenta la propensión a aumentar el consumo. Esto se puede hacer a través de campañas que indiquen este medio de pago como seguro o entregando beneficios especiales a los clientes que lo hagan. Esto se propone especialmente para los clientes de Medio Bajo consumo pues es en este grupo las compras por internet tienen el mayor impacto.

A diferencia de los otros grupos de clientes, los clientes del grupo Medio Bajo que canjean sus puntos del club de lealtad son más propensos a aumentar su consumo y por lo tanto se propone enviar recordatorios a los clientes con posibilidad de canjear y que hayan bajado sus niveles transaccionales.

En cuanto a temas demográficos, se destaca que los hombres de entre 35 y 55 años son más propensos a aumentar su consumo y por lo tanto se propone generar promociones especiales para este tipo de clientes.

Finalmente se recomienda, al igual que en el grupo de bajo consumo, enviar recordatorios de la tarjeta y acumulación de puntos a los clientes que vayan a aumentar su consumo.

7. Evaluación económica

En la sección anterior se presentan las recomendaciones al banco en base a los resultados de los modelos, sin embargo, cabe preguntarse cuál es el beneficio que obtendría este al utilizar esta metodología. Para esto se considera que, en base a conversaciones con el banco, este no puede generar promociones diferenciadas para cada cliente y además, considera más factible generar promociones para consumir en rubros o avances, que otras campañas enfocadas en aumentar montos.

En base a lo anterior, se define que para los clientes de Bajo y Medio Bajo consumo se generan tres promociones sólo a aquellos clientes que no vayan a aumentar su consumo, mientras que a los Inactivos se les envía promociones diferenciadas según si aumentan su consumo o no. Se compara además los resultados de la metodología con el envío de tres promociones genéricas en la situación actual, las cuales son enviadas a todos los clientes del grupo.

Por otro lado, no se tiene claridad sobre las tasas de respuesta a las promociones del banco y por lo tanto se generan tres evaluaciones económicas en las que varía la respuesta de los clientes a las promociones:

- Escenario realista: Se utilizan tasas de respuesta a las promociones estimadas en conjunto con la consultora Innovapay, las cuales son diferentes para cada promoción. La tasa de respuesta para la situación actual es el promedio de las tasas utilizadas en el modelo.
- Escenario pesimista: Considera que los clientes del banco son menos propensos a tomar promociones y por lo tanto se utilizan tasas de respuesta más bajas. Las tasas utilizadas son la mitad que en el escenario realista, sin embargo, se considera además que la respuesta a la situación actual es ligeramente mejor que la del modelo y por lo tanto su tasa es 1% mejor que el promedio.
- Escenario optimista: Considera tasas de respuesta iguales a 1,5 veces las tasas del escenario realista. Además, la tasa de la situación actual es 1% menor que el promedio de las tasas del modelo.

A continuación se detallan los costos considerados en la evaluación, los cuales varían entre la situación actual y la aplicación de la metodología:

- Costo de envío: Actualmente todas las promociones y recordatorios dirigidos a la banca personas del banco se envían a través de correo electrónico. El costo de cada envío ronda los 100 pesos chilenos.
- Costo de diseño: La generación de las promociones requiere de la definición de los términos legales, diseño de la promoción desde el departamento de marketing y diseño de la pieza que se envía en el correo electrónico. Esto se realiza para cada promoción o recordatorio y tiene un costo promedio de 500 mil pesos.
- Costo directo de la promoción: Ya sea a través de la devolución de dinero o el canje de puntos, las promociones del banco tienen un costo promedio de 4 mil pesos por cliente que la utiliza.

- Descuento adquirente: Para hacer posibles las transacciones en los comercios, los bancos se afilian a empresas adquirentes (en Chile es Transbank), las cuales cobran un 30% del ingreso por comisión a comercios.

Con respecto a los ingresos que percibe el banco por las promociones, estos también son diferentes en el caso de la situación actual y el modelo. Además, se debe destacar que en esta evaluación no se consideran ingresos relacionados a la recordación de marca por parte de los clientes. En base a eso los ingresos son:

- Comisión a comercios: Es el 1,5% del valor de las transacciones que se realizan con la tarjeta de crédito en cualquier comercio. Se considera una transacción promedio de 50 mil pesos en base a información de la consultora.
- Aumentos de consumo: Los clientes de los diferentes niveles de consumo tienen rentabilidades diferentes y por lo tanto al inducir el aumento de consumo, aumenta la rentabilidad de los clientes.

Finalmente se debe tener en cuenta que la evaluación económica se realiza utilizando periodo 6 de los datos, el cual va entre agosto y noviembre de 2015.

7.1. Evaluación económica modelo Inactivos

Los resultados del modelo de propensión indican que 5.256 clientes no aumentarán su consumo y 3.455 sí lo harán. De esta forma, al primer grupo se le deben enviar promociones más caras que al segundo.

Por otro lado, al incentivar las compras internacionales de estos clientes se incentiva al aumento de consumo y por lo tanto una de las tres promociones que se envía es en este tipo de compra. Se considera además que para este grupo de clientes, las dos promociones restantes son iguales a las de la situación actual, es decir, promociones genéricas que se diseñan para otros niveles de consumo como supermercados o tiendas por departamento.

Luego en base a lo analizado con la consultora, las tasas a utilizar en el escenario realista son:

| Tipo de promoción | Tasa de respuesta |
|-----------------------------------|--------------------------|
| Genérica: Situación actual | 8% |
| Genérica: Modelo NO aumentan | 10% |
| Genérica: Modelo SÍ aumentan | 6% |
| Internacional: Modelo NO aumentan | 5% |
| Internacional: Modelo SÍ aumentan | 4% |

Tabla 21: Tasa de respuesta según tipo de promoción clientes Inactivos

La tabla indica que en la situación actual, el 8% de los clientes Inactivos toman una promoción, mientras que en la situación donde se aplica la metodología, el 6% de los clientes que aumentarán su consumo toman una promoción genérica (no de compra internacional). Cabe destacar que las tasas de los clientes que no aumentan son más altas pues las promociones son más caras y por lo tanto más atractivas.

Por otro lado, se debe notar el hecho de que las promociones en compras internacionales tienen tasas de respuesta más bajas, pues para el banco es más difícil incentivar este tipo de transacciones.

En base a lo anterior, tanto en la situación actual como en la aplicación de la metodología se envían tres promociones a cada cliente, es decir, 26.613 envíos. Cabe destacar que en la situación actual se diseñan tres promociones para todos los clientes, mientras que en la aplicación de la metodología se deben diseñar tres en cada grupo de clientes. Para incorporar esto en la evaluación, se considera el diseño de una promoción en cada grupo, es decir, el costo de diseño en la situación actual de cada grupo es de 500 mil pesos.

Con respecto a los costos de devolución de dinero, para la situación actual estos son de 4 mil pesos, mientras que en el modelo se utilizan dos valores: 6 mil pesos para clientes que no aumentan su consumo (promoción cara) y 2 mil para los que sí (promoción barata). Además, se consideran transacciones promedio de 40 mil y 60 mil para cada promoción.

Por otro lado, se debe estimar el número de promociones que son tomadas por los clientes, para lo cual se debe considerar que cada cliente puede utilizar de cero a tres promociones. Luego, considerando las tasas de respuesta, en la situación actual se toman 2091 promociones, mientras en la aplicación del modelo, 1052 clientes toman una promoción genérica cara, 415 toman una genérica barata, 263 una promoción de compra internacional cara y 138 toman una barata.

Para determinar los ingresos que se obtienen por el aumento de consumo se calcula el diferencial de ingreso, considerando sólo a los clientes que aumentan debido a la aplicación y no aquellos que lo hacen de manera natural, por lo que se define que el ingreso de la situación actual es 0. Por otro lado, de los 263 clientes que tomarían la promoción cara en compras internacionales, 256 aumentarían su consumo de acuerdo al modelo, sin embargo, el poder predictivo de aumentos de los clientes Inactivos es del 43%, por lo tanto 110 clientes aumentarían su consumo debido a la promoción.

Por último, como se puede ver en el anexo O, el aumento ponderado de la rentabilidad de cada cliente que aumenta su consumo es de \$33.878. De esta manera se obtienen los siguientes resultados para un escenario realista en un periodo de cuatro meses:

| | Situación actual | Aplicación metodología |
|-----------------------------|-------------------------|-------------------------------|
| Costo envío | \$2.613.300 | \$2.613.300 |
| Costo diseño | \$500.000 | \$500.000 |
| Costo devolución | \$8.362.560 | \$8.989.600 |
| Comisión a comercios | \$1.567.980 | \$916.860 |
| Descuento adquirente | \$470.394 | \$275.058 |
| Aumento de consumo | \$0 | \$3.734.300 |
| Total | -\$11.378.274 | -\$8.726.798 |
| Ganancia Metodología | | \$2.651.476 |

Tabla 22: Evaluación económica clientes Inactivos escenario Realista

De la tabla se puede notar que al aplicar la metodología diseñada para los clientes inactivos, el banco podría ganar más de 2,6 millones de pesos en un periodo de 4 meses, es decir 7,9 millones en un año.

Para realizar el análisis por escenarios se consideran variaciones en las tasas de respuesta de los clientes hacia las promociones del banco según lo explicado en el punto anterior. De esta forma las tasas a utilizar para el cada escenario son las siguientes:

| Tipo de promoción | Tasas de respuesta: Escenario pesimista | Tasas de respuesta: Escenario optimista |
|-----------------------------------|--|--|
| Genérica: Situación actual | 5% | 11% |
| Genérica: Modelo NO aumentan | 6% | 14% |
| Genérica: Modelo SÍ aumentan | 4% | 8% |
| Internacional: Modelo NO aumentan | 2,5% | 7,5% |
| Internacional: Modelo SÍ aumentan | 2% | 6% |

Tabla 23: Tasas de respuesta a promociones escenarios Pesimista y Optimista clientes Inactivos

De esta forma la ganancia por aplicar la metodología para un periodo de cuatro meses en el escenario pesimista sería de \$1.512.563, mientras que en el escenario optimista sería de \$3.790.388. La desagregación de ambos escenarios se encuentra en el anexo P de este trabajo.

7.2. Evaluación económica modelo Bajo consumo

Al utilizar el modelo de propensión en los clientes de Bajo consumo, este indica que 1.631 de los 3.353 clientes no aumentará su consumo y por lo tanto es a ellos a los que se debe incentivar. Por otro lado, se determina que los tres tipos de transacción que se deben incentivar son los avances de efectivo, las compras internacionales y las compras en tiendas por departamento, donde lo que se busca con cada promoción es que los clientes tengan al menos una transacción en el rubro.

De esta forma, se envía una promoción de cada tipo a los clientes que no aumentan su consumo y que a la vez no tienen transacciones en el rubro de la promoción. En la siguiente tabla se puede ver el número de clientes que reciben cada promoción o combinación de estas:

| Tipo de promoción | # Clientes que reciben |
|--|-------------------------------|
| Avance de efectivo | 28 |
| Compra internacional | 5 |
| Tiendas por departamento | 12 |
| Avance de efectivo y compra internacional | 173 |
| Avance de efectivo y tiendas por depto. | 150 |
| Compra internacional y tiendas por depto. | 171 |
| Avance, internacional y tiendas por depto. | 1.089 |

Tabla 24: Cantidad de clientes Bajo consumo a los que se les envía cada combinación promocional

De la tabla se puede observar que de los clientes que no aumentarían su consumo de acuerdo al modelo, 1.089 no realizaron transacciones en ninguno de los rubros a incentivar y por lo tanto reciben tres promociones. Por otro lado, 28 clientes realizaron compras en tiendas por departamentos y compras internacionales, pero no avances y por esa razón reciben sólo una promoción en avances de efectivo.

Se puede notar además que al utilizar el modelo se envían 4.300 promociones, mientras que en la situación actual el banco envía tres promociones a 3.353 clientes, es decir 10.059. Además, a los 1.722 clientes que no aumentarían se les envía un recordatorio vía correo electrónico.

Con respecto a la situación actual, la aplicación del modelo en el grupo de clientes de Bajo consumo implica la creación de una comunicación extra, pues además de las tres promociones, se debe diseñar el recordatorio a los clientes que aumentan su consumo. Además, el costo de diseño de la promoción de en compras internacionales se incluye en los costos de los clientes Inactivos (al ser una promoción en el mismo tipo de transacción, sólo se diseña una vez).

Las tasas de respuesta utilizadas en el escenario realista son estimadas en conjunto con la consultora, quedando de la siguiente manera:

| Tipo de promoción | Tasa de respuesta |
|--------------------------|--------------------------|
| Situación actual | 7% |
| Avance de efectivo | 5% |
| Compra internacional | 5% |
| Tiendas por departamento | 10% |

Tabla 25: Tasas de respuesta según tipo de promoción escenario realista clientes Bajo consumo

En base a las tasas de respuesta y a la cantidad de promociones que se envían, la cantidad de clientes que toman cada combinación de promociones se presenta en la siguiente tabla:

| Incentivos | # clientes que toman | Modelo predice aumento | Aumento Real |
|--|-----------------------------|-------------------------------|---------------------|
| Avance de efectivo | 63 | 27 | 16 |
| Compra internacional | 63 | 23 | 14 |
| Tiendas por departamento | 130 | 26 | 16 |
| Avance de efectivo y compra internacional | 3 | 2 | 1 |
| Avance de efectivo y tiendas por depto. | 6 | 3 | 2 |
| Compra internacional y tiendas por depto. | 6 | 3 | 2 |
| Avance, internacional y tiendas por depto. | 0 | 0 | 0 |

Tabla 26: Respuesta a promociones clientes Bajo consumo

De la tabla se puede notar que de todos los clientes a los que se les enviaría una promoción de avance en efectivo, 63 la tomarían de manera exclusiva, es decir, sin tomar

otra promoción. Además, el modelo indica que 27 de estos 63 clientes aumentarían su consumo luego de tomar el avance, sin embargo, la predicción de los aumentos es de 60% y por lo tanto serían 16 los clientes que aumentarían su consumo.

De esta manera, al aplicar la metodología serían 50 los clientes que aumentarían su consumo debido a las promociones. Además, en base al anexo O, la rentabilidad producida por un cliente de bajo consumo que aumenta su nivel de consumo es de \$16.783.

En base a lo anterior, se presenta a continuación la tabla comparativa de la situación actual y la aplicación de la metodología:

| | Situación actual | Aplicación metodología |
|---------------------------------|-------------------------|-------------------------------|
| Costo envío promoción | \$1.005.900 | \$430.000 |
| Costo envío recordatorio | \$0 | \$172.200 |
| Costo diseño | \$500.000 | \$1.500.000 |
| Costo devolución | \$2.011.800 | \$572.200 |
| Comisión a comercios | \$377.213 | \$107.288 |
| Descuento adquirente | \$113.164 | \$32.186 |
| Aumento de consumo | \$0 | \$427.191 |
| Total | -\$4.302.302 | -\$2.172.107 |
| | | |
| Ganancia Metodología | | \$2.051.253 |

Tabla 27: Evaluación Económica clientes Bajo consumo escenario Realista

De la tabla se puede observar que el banco dejaría de gastar más de 2 millones de pesos en cuatro meses por aplicar el modelo de propensión para clientes de Bajo consumo y 6,1 millones en un año.

Con respecto a los escenarios optimista y pesimista, estos consideran que los clientes del banco son más o menos propensos a tomar promociones, lo que se ve reflejado en las tasas de respuesta que se presentan a continuación:

| Tipo de promoción | Tasas de respuesta: Escenario pesimista | Tasas de respuesta: Escenario optimista |
|--------------------------|--|--|
| Situación actual | 5% | 11% |
| Avances de efectivo | 3% | 8% |
| Compra internacional | 3% | 8% |
| Tiendas por departamento | 5% | 15% |

Tabla 28: Tasas de respuesta a promociones escenarios Pesimista y Optimista clientes Bajo consumo

En base a la tabla, la ganancia que obtendría el banco en un escenario pesimista es de \$1.081.544 en cuatro meses, mientras que en un escenario optimista sería de \$3.012.004. La desagregación de ambos escenarios se encuentra en el anexo Q.

Finalmente cabe destacar que si bien los resultados predictivos del modelo de Bajo consumo son mejores que los del segmento de Inactivos, la ganancia que genera es menor pues el grupo de Inactivos tiene 2,5 veces los clientes del nivel Bajo.

7.3. Evaluación económica modelo Medio Bajo consumo

Al analizar las variables obtenidas del modelo, se puede notar que en cuanto a rubros -o tipos de compra- las más importantes son los avances de efectivo, supermercados, bencina y restaurantes. No obstante, cómo se mencionó anteriormente, para el banco es más fácil generar promociones que busquen que el cliente realice una compra y no buscar que aumenten montos, por lo que se eligen los avances, supermercados y farmacias.

Por otro lado, el modelo indica que 14.279 clientes de Medio Bajo consumo no aumentarán su consumo en el último periodo y por lo tanto es a ellos a quienes se les enviarían las promociones. Con respecto a los 7.463 clientes que sí aumentarían, estos recibirían un recordatorio vía correo electrónico.

En base a lo anterior cada cliente podría recibir entre cero y tres promociones, pues sólo se envía si el cliente no ha realizado ya alguna transacción en el rubro durante el periodo. De esta forma el número de clientes que recibirían cada combinación de promociones sería:

| Tipo de promoción | # clientes que reciben |
|-----------------------------------|-------------------------------|
| Supermercados | 95 |
| Restaurant | 497 |
| Avances de efectivo | 610 |
| Avances de efectivo y Restaurant | 2.679 |
| Avance de efectivo y Supermercado | 675 |
| Restaurant y Supermercados | 1.520 |
| Avance, Restaurant y Supermercado | 8.121 |

Tabla 29: Cantidad de clientes Medio Bajo consumo a los que se les envía cada combinación promocional

De la tabla se puede ver que existen 8.121 clientes que no aumentarán su consumo y que a la vez no realizaron transacciones en ninguno de los tres tipos de compra propuestos y que por lo tanto, reciben tres promociones. Además el número de promociones enviadas sería de 35.313, mientras que en la situación actual serían tres por cada cliente de Medio Bajo consumo, es decir, 65.226.

Con respecto al diseño de promociones se puede notar que se repite el envío de campañas de avances de efectivo y del recordatorio, por lo que sólo se deben diseñar una para supermercados y una para restaurantes. Por otro lado, para determinar los otros costos e ingresos se utilizan las siguientes tasas de respuesta:

| Tipo de promoción | Tasa de respuesta |
|--------------------------|--------------------------|
| Situación actual | 8% |
| Avance de efectivo | 6% |
| Supermercados | 12% |
| Restaurantes | 6% |

Tabla 30: Tasas de respuesta según tipo de promoción escenario Realista clientes Medio Bajo consumo

De esta manera el número de clientes que tomarían la promoción y que aumentarían su consumo debido a eso se muestra a continuación:

| Incentivos | # clientes que toman | Modelo predice aumento | Aumento Real |
|-----------------------------------|-----------------------------|-------------------------------|---------------------|
| Supermercados | 1.120 | 189 | 36 |
| Restaurant | 664 | 112 | 21 |
| Avances de efectivo | 626 | 113 | 21 |
| Avances de efectivo y Restaurant | 35 | 11 | 2 |
| Avance de efectivo y Supermercado | 60 | 53 | 10 |
| Restaurant y Supermercados | 66 | 33 | 6 |
| Avance, Restaurant y Supermercado | 4 | 3 | 1 |

Tabla 31: Respuesta a promociones clientes Medio Bajo consumo

De esta manera la cantidad de promociones que se harían efectivas con la aplicación del modelo serían 2.743, mientras que en la situación actual serían 5.218. Por otro lado, los clientes que realmente aumentarían su consumo serían 97, donde la rentabilidad ponderada de cada uno según el anexo O sería de \$22.412.

En base a lo anterior, la ganancia que genera para el banco la aplicación del modelo en un escenario realista se puede observar en la siguiente tabla:

| | Situación actual | Aplicación metodología |
|---------------------------------|-------------------------|-------------------------------|
| Costo envío promoción | \$6.522.600 | \$3.531.300 |
| Costo envío recordatorio | \$0 | \$746.300 |
| Costo diseño | \$500.000 | \$1.000.000 |
| Costo devolución | \$20.872.320 | \$10.973.760 |
| Comisión a comercios | \$3.913.560 | \$2.057.580 |
| Descuento adquirente | \$1.174.068 | \$617.274 |
| Aumento de consumo | \$0 | \$2.179.762 |
| Total | -\$25.155.428 | -\$10.631.292 |
| | | |
| Ganancia Metodología | | \$12.524.136 |

Tabla 32: Evaluación Económica clientes Medio Bajo consumo escenario Realista

De la tabla se puede notar que al aplicar la metodología en el grupo de clientes de Medio Bajo consumo, el banco puede ahorrar 12,5 millones de pesos en cuatro meses y más de 37 millones en un año. Además, se puede notar que la principal diferencia entre la situación actual y la aplicación del modelo es la cantidad de personas a las que se les envía cada promoción y junto con eso, los montos devueltos o de canje de puntos.

Lo anterior se debe en parte a que no se consideran fuentes de ingreso como la cantidad de clientes nuevos que llegan al banco por tener más promociones, ni el posible aumento de consumo de los clientes luego del periodo de cuatro meses.

Con respecto a las tasas de respuesta de los escenarios pesimista y optimista, estas se presentan en la siguiente tabla:

| Tipo de promoción | Tasas de respuesta: Escenario pesimista | Tasas de respuesta: Escenario optimista |
|--------------------------|--|--|
| Situación actual | 5% | 11% |
| Avances de efectivo | 3% | 9% |
| Supermercados | 6% | 18% |
| Restaurantes | 3% | 9% |

Tabla 33: Tasas de respuesta a promociones escenarios Pesimista y Optimista clientes Medio Bajo consumo

Luego, utilizando estas tasas el valor económico de aplicar la metodología en los clientes de Medio Bajo consumo es de 9,3 millones de pesos en un escenario pesimista y 15,4 millones en un escenario optimista, medido en un periodo de cuatro meses. La desagregación de los escenarios se encuentra en el anexo R.

En base a las tres evaluaciones presentadas se puede notar que en un escenario realista, el banco podría ganar \$51.680.596 en un año al utilizar la metodología propuesta en este trabajo. Por otro lado, la mayor fuente de ahorro se produce al dejar de ofrecer promociones a clientes que aumentarán su consumo de manera natural, pues el costo de las promociones del banco no se recupera con la comisión a los comercios.

En cuanto a la rentabilidad de la aplicación en los diferentes grupos, se puede notar que en el nivel de Medio Bajo consumo es donde se genera una mayor ganancia pues es el grupo más grande y además, es el segmento en que se obtuvieron los mejores resultados a nivel predictivo.

Finalmente, se puede notar que sin importar el escenario, la aplicación de la metodología resultaría rentable para el banco e independiente de la cantidad de aumentos de consumo que genere, representaría una disminución en los costos promocionales. La ganancia total que percibiría el banco en cada escenario se presenta en el anexo S.

8. Conclusiones

8.1. Conclusiones del proyecto

El principal objetivo de este trabajo consiste en identificar y cuantificar los comportamientos capaces de inducir aumentos de consumo en el uso de la tarjeta de crédito en los clientes de un banco. La importancia de esto radica en que la rentabilidad del negocio de las tarjetas está directamente relacionada al consumo de los clientes.

Lo anterior se basa en el conocimiento del banco sobre las diferencias en el uso de la tarjeta de crédito por parte de los diferentes niveles de consumo del banco. Además se considera la hipótesis de que las transacciones en rubros específicos son capaces de inducir el aumento de consumo. De esta forma, se busca responder a las preguntas de ¿a quién incentivar? y ¿con qué acciones hacerlo?

El primer paso entonces es determinar los niveles de consumo de los clientes del banco, de lo cual se ve que estos se pueden clasificar correctamente en cinco grupos en base a sus principales variables transaccionales. De esta forma se concluye que las variables RFM son capaces de distinguir entre clientes con un alto número de transacciones y aquellos que utilizan la tarjeta sólo para compras grandes.

En relación a esto destaca el hecho de que a excepción de un grupo, el nivel de consumo de los clientes aumenta al aumentar el número de transacciones y no el monto promedio de estas. Esto se suma al hecho de que los clientes de Alto consumo, utilizan su tarjeta 16 veces en un mes promedio, es decir, más de una vez cada dos días.

Por otro lado se concluye que si bien no es el grupo más rentable, los clientes de Medio Bajo consumo son los que requieren mayor atención por parte del banco pues representan el 54% de la cartera y además, al aumentar su consumo son los que aumentan más el monto consumido. Con respecto a los clientes del grupo Medio Alto se puede decir que al cambiar al grupo de Alto consumo están cambiando su estructura de consumo y no profundizándola como el resto de los grupos.

Con respecto a los grupos encontrados en este estudio, destaca el hecho de que tanto las distribuciones de consumo como los valores de las variables transaccionales se mantienen durante los seis periodos, lo que indica que estos son los estados naturales de los clientes y por lo tanto valida los resultados encontrados.

En cuanto a los resultados obtenidos de los modelos de propensión, estos permiten cumplir con el objetivo de este trabajo, pues logran identificar transacciones y comportamientos capaces de inducir el aumento de consumo de los clientes.

Con respecto a lo anterior, destaca el hecho de que tanto la cantidad de rubros como las semanas en las que consume un cliente, debiesen ser incentivadas por el banco pues aportan a la probabilidad de aumentar el consumo

Por otro lado los modelos permiten concluir que los avances de efectivo, las compras por internet, compras en supermercados o grandes tiendas son tipos de transacciones que

inducen a los clientes a aumentar su consumo sólo con realizarlas y por lo tanto no se requiere incentivar a los clientes a transar grandes montos.

Además de lo anterior, el modelo permite identificar a los clientes que aumentarán su consumo y por lo tanto ayuda a determinar a qué clientes se debe incentivar, pues no se debe incentivar a los clientes que aumentarán su consumo de igual manera, pues se incurre en pérdidas debido al costo de las promociones.

Finalmente se concluye que en caso de utilizar la metodología propuesta en este trabajo, el banco podría ganar 51 millones de pesos en un año, generando ingresos por los nuevos aumentos de consumo y por otro lado, dejando de gastar recursos en clientes que no lo requieren.

8.2. Limitaciones y trabajos futuros

La principal limitación de este trabajo consistió en la falta de datos completos sobre las promociones históricas del banco, por lo que no se pudo incluir el efecto de estas en el aumento de consumo de los clientes. A causa de esto, en este trabajo se debe considerar el supuesto de que toda la propensión a aumentar el consumo se debe a los factores inductores, sin embargo, no se consideran los efectos de las campañas. En caso de contar con esta información, se podría ampliar el modelo actual para entender la efectividad promocional de las campañas, es decir, entender si estas tienen un efecto sólo en el periodo en que se envían o también en el periodo posterior.

Además de lo anterior, sería interesante saber si los clientes con mayor probabilidad de aumentar su consumo deben ser incentivados de la misma forma que aquellos con baja propensión. Para esto se propone el diseño e implementación de experimentos con diversos tipos de incentivos a cada grupo, para analizar si existen diferencias en las respuestas de cada tipo de cliente.

Por otro lado, en este estudio no se contó con datos sobre el ingreso de los clientes, sin embargo, se sabe que la segmentación con esta variable es la principal en la industria bancaria [6]. En base a lo anterior, de conseguir esta información con el banco se podría generar una segmentación previa a la clasificación por niveles de consumo, para así tratar de diferente manera a los clientes según su nivel de ingreso.

En este trabajo se determinó la importancia de los avances de efectivo, sin embargo, sería interesante saber si todos los clientes del banco son igualmente propensos a realizar este tipo de transacciones o si en realidad existe un perfil específico de personas que lo hacen. Lo anterior también aplica a las compras internacionales y compras por internet.

Como se menciona en los alcances, en este proyecto sólo se considera a los clientes de la banca personas, sin embargo, sería interesante entender si los clientes de la banca privada -que tienen mayores ingresos- se comportan de una manera similar y si los resultados obtenidos son extrapolables a ese tipo de clientes.

De la misma forma, en los alcances se menciona que no se determina la propensión al endeudamiento, sin embargo, este es una fuente de ingresos de la tarjeta y por lo tanto

para el banco sería útil entender cómo es la relación entre el aumento de consumo y el endeudamiento.

Por último, en este trabajo se busca predecir si los clientes pasarán a un nivel de mayor consumo, sin embargo, para el banco podría ser útil entender los movimientos que tienen los clientes entre los diferentes clusters a través del tiempo. Esto es, determinar si los clientes de Bajo consumo suben directamente a niveles altos o pasan primero un periodo teniendo consumo Medio Bajo.

Con lo anterior se podrían determinar caminos de bajada para entender las razones por las que los clientes disminuyen su consumo. En esta línea, Sung Ho Ha (2007) determina los caminos que toman en promedio los clientes, sin embargo, podría ampliarse a nivel individual.

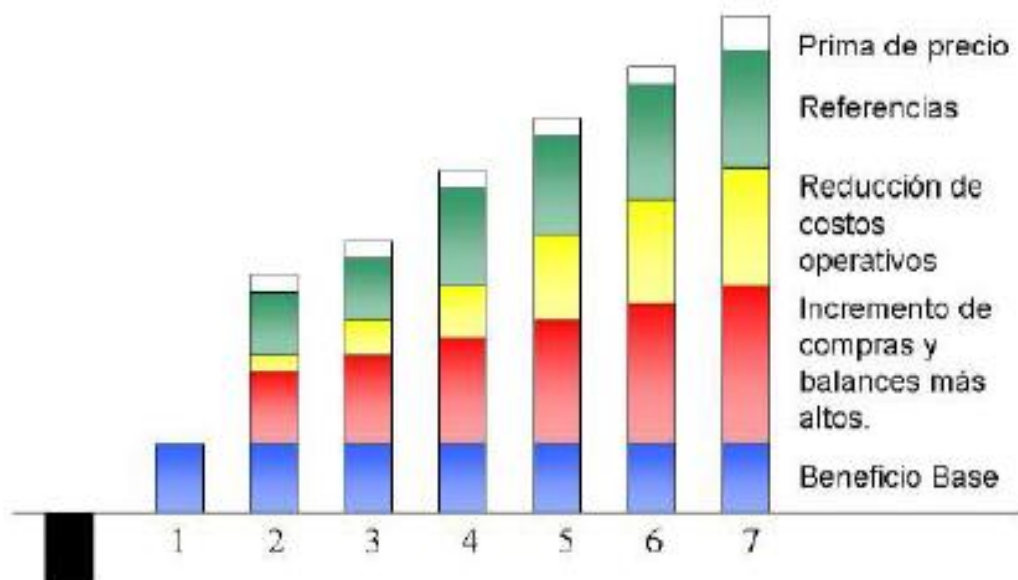
9. Bibliografía

- [1] ABIF INFORMA. 2014. El PIB del sector bancario alcanzó a US\$ 10.900 millones en 2013. [En línea]. < <http://www.abif.cl/wp-content/uploads/2015/02/ABIF-Infirma-N27.pdf>> [consulta: 06 abril 2016]
- [2] Sbif. 2016. Panorama de la Industria Bancaria al tercer trimestre de 2015. [En línea]. <www.sbif.cl>. [Consulta: 06 abril 2016]
- [3] Ministerio de hacienda. 2013. Inclusión Financiera y Medio de Pago Electrónicos. [En línea]. <<http://www.economia.gob.cl>> [Consulta: 06 abril de 2016]
- [4] Sbif. 2016. Informe de Tarjetas de Crédito. Emisores Bancarios. [En línea]. <www.sbif.cl>. [Consulta: 06 abril 2016]
- [5] Sbif. 2016. Informe de Tarjetas de débito. Emisores Bancarios. [En línea]. <www.sbif.cl>. [Consulta: 06 abril 2016]
- [6] Retención y Fidelización de Clientes. 2010. Banco de Chile. 5to congreso nacional de crédito y cobranzas.
- [7] RDT Ibermatica. 2011. Minería de datos sobre ontologías. [En línea]. <www.rtdibermatica.com>. [Consulta: 06 Abril de 2016]
- [8] XU, R. WUNSCH, D. 2005. Survey of Clustering Algorithms. IEEE transactions on neural networks 16(3): 645-678.
- [9] WEI, J. LIN, S. WU, H. 2010. A review of the application of RFM model. African Journal of Business Management 4: 4199-4206.
- [10] Ministerio de desarrollo social. 2013. Encuesta Casen. [En línea]. Santiago, Chile. <http://www.ministeriodesarrollosocial.gob.cl/basededatoscasen.php> > [consulta: 20 abril 2016]
- [11] Criteria Research. 2016. Grupos socioeconómicos 2015. [En línea]. <http://www.criteria.cl/wp-content/files/GRUPO_SOCIO_ECO_2015.pdf>. [Consulta: 20 abril 2016]
- [12] BIRANT, D. 2011. Data Mining Using RFM Analysis. Knowledge-Oriented Applications in Data Mining, 91-108.
- [13] HO HA, S. 2007. Applying knowledge engineering techniques to customer analysis in the service industry. Advanced Engineering Informatics 21: 293-301.
- [14] HSIEH, N. 2004. An integrated data mining and behavioral scoring model for analyzing bank customers. Expert Systems with Applications 27: 623-633.
- [15] ROCO, C. 2010. Modelamiento Predictivo para el aumento de consumo de tarjeta de crédito sobre el análisis de comportamiento transaccional de clientes de una institución financiera. Memoria de Ingeniería Civil Industrial. Santiago, Chile. Universidad de Chile, Facultad de ciencias físicas y matemáticas. 141p.

- [16] REICHHELD, F. SASSER, W. 1990. Zero Defections: Quality Comes to Services. *Harvard Business Review* 68: 105-111.
- [17] WEBER, R. MIRANDA J. REY, P. 2005. Predicción de fugas de clientes para una institución financiera mediante support vector machines. *Ingeniería de sistemas*, vol XIX.
- [18] NGAI, E. XIU, L. CHAU, D. 2009. Application of data mining techniques in customer relationship management: A literature review and classification. *Experts systems with applications* 36: 2592-2602.
- [19] BERRY, M, LINOFF, G. 2004. *Data mining techniques for Marketing, Sales and Customer Relationship Management*. 2^a ed. Indiana. Wiley Publishing. 672p.
- [20] AGGELIS, V. CHRISTODOULAKIS D. 2005. Customer clustering using RFM analysis. ICCOMP'05 Proceedings of the 9th WSEAS International Conference on Computers.
- [21] GENSLER, S. LEEFLANG, P. SKIERA, B. 2012. Impact of online channel use on customer revenues and costs to serve: Considering product portfolios and self-selection. *Intern journal of research in marketing*. 29: 192-201.
- [22] KIM, H. DEVANEY, S. 2001. The determinants of outstanding balances among credit card revolvers. *Financial Counseling and Planning*. 67-79.
- [23] HALKIDI, M. BATISTAKIS, Y. VAZIRGLANNIS, M. 2001. On Clustering Validation Techniques. *Journal of Intelligent Information Systems*. 17: 107-145.

10. Anexos

Anexo A: Beneficio de un cliente Bancario en el tiempo



Fuente: Zero Defections: Quality Comes to Services. HBR

Anexo B: Estadísticos descriptivos RFM por niveles de consumo

Variable R (días)

| Variable R | Bajo | Medio Bajo | Medio Alto | Alto |
|-----------------|------|------------|------------|------|
| Promedio | 72 | 12 | 26 | 4 |
| SD | 22 | 10 | 25 | 6 |
| Mínimo | 39 | 0 | 0 | 0 |
| Máximo | 121 | 45 | 119 | 70 |

Variable F (cantidad de compras)

| Variable F | Bajo | Medio Bajo | Medio Alto | Alto |
|-----------------|------|------------|------------|------|
| Promedio | 1,0 | 3,1 | 1,3 | 15,8 |
| SD | 1,4 | 2,4 | 1,7 | 4,9 |
| Mínimo | 0,3 | 0,3 | 0,3 | 9,3 |
| Máximo | 15,8 | 10,3 | 26,8 | 27,3 |

Variable M (UF)

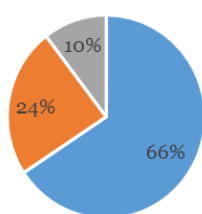
| Variable M | Bajo | Medio Bajo | Medio Alto | Alto |
|-----------------|------|------------|------------|------|
| Promedio | 3,6 | 2,4 | 17,9 | 1,6 |
| SD | 3,4 | 2,1 | 6,1 | 1,3 |
| Mínimo | 0,0 | 0,0 | 9,4 | 0,01 |
| Máximo | 19,6 | 11,0 | 33,3 | 14,1 |

Anexo C: Agrupación de rubros SBIF

| Habituales o básicos | Entretención | Servicios |
|------------------------------------|-----------------------|-------------------|
| Supermercados | Restaurantes | Recaudación |
| Farmacias | Líneas aéreas | |
| Combustibles | Hoteles | Pago de impuestos |
| Tiendas por departamentos | Operadores Rent a car | Educación |
| Vestuario y Calzado | | Clínicas |
| Accesorios y servicios automotoras | | |
| Muebles y artículos electrónicos | | |

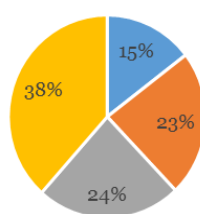
Anexo D: Información de cuentas cartera completa

Tipo de tarjeta



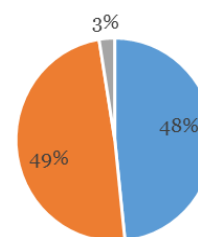
■ Gold ■ Platinum ■ Black

Antigüedad



■ 0-6 meses ■ 7-18 meses ■ 19-36 meses ■ 37+

Cantidad de cuentas



■ 1 ■ 2 ■ 3+

Anexo E: Test de igualdad de medias clientes con y sin aumento de consumo

| Nivel de consumo | Variables | Estadístico | Valor crítico (95%) | Conclusión |
|-------------------------|--------------------------|--------------------|----------------------------|--|
| Inactivos | Compras habituales Per o | -4,7 | 1,96 | Rechaza hipótesis de igualdad de medias |
| Inactivos | Cliente Medio Bajo Per o | -5,6 | 1,96 | Rechaza hipótesis de igualdad de medias |
| Inactivos | Cliente Bajo Per o | 1,9 | 1,96 | No se rechaza hipótesis de igualdad |
| Bajo | Avances de efectivo | -2,9 | 1,96 | Rechaza hipótesis de igualdad de medias |
| Bajo | Tiendas por depto. | -4,1 | 1,96 | Rechaza hipótesis de igualdad de medias |
| Bajo | Compra internacional | -3,5 | 1,96 | Rechaza hipótesis de igualdad de medias |
| Medio Bajo | Avances de efectivo | -5,9 | 1,96 | Rechaza hipótesis de igualdad de medias |
| Medio Bajo | Supermercados | -29,5 | 1,96 | Rechaza hipótesis de igualdad de medias |
| Medio Bajo | Restaurantes | -28,1 | 1,96 | Rechaza hipótesis de igualdad de medias |
| Medio Bajo | Bencina | -23,6 | 1,96 | Rechaza hipótesis de igualdad de medias |
| Medio Bajo | Tiendas por depto. | -19,2 | 1,96 | Rechaza hipótesis de igualdad de medias |

Anexo F: Montos y número de compras promedio por rubro

Clientes Bajo consumo

| Rubro | Monto por cliente (UF) | SD (UF) | Nro. de transacciones por cliente | SD (#) | Clientes con transacciones |
|----------------------------------|-------------------------------|----------------|--|---------------|-----------------------------------|
| Avances | 2,96 | 10,06 | 0,49 | 1,37 | 781 |
| Bencina | 0,23 | 1,00 | 0,26 | 0,97 | 439 |
| Serv. financieros | 0,13 | 1,01 | 0,14 | 0,58 | 267 |
| Clínicas | 0,30 | 1,95 | 0,09 | 0,41 | 215 |
| Colegios | 0,05 | 1,59 | 0,01 | 0,05 | 10 |
| Compraventa Autos | 0,07 | 1,16 | 0,01 | 0,08 | 20 |
| Educación superior | 0,22 | 3,29 | 0,02 | 0,19 | 50 |
| Farmacias | 0,13 | 0,55 | 0,16 | 0,60 | 370 |
| Impuestos | 0,17 | 2,25 | 0,03 | 0,20 | 73 |
| Líneas aéreas | 1,09 | 4,40 | 0,18 | 0,58 | 444 |
| Tiendas por depto. | 1,22 | 3,76 | 0,42 | 1,02 | 882 |
| Reparación y acces. Autos | 0,25 | 2,02 | 0,05 | 0,29 | 137 |
| Restaurantes | 0,26 | 1,39 | 0,23 | 1,02 | 437 |
| Servicios básicos | 0,18 | 1,39 | 0,10 | 0,69 | 153 |
| Supermercados | 0,79 | 2,81 | 0,51 | 1,49 | 804 |
| Vestuario | 0,28 | 1,45 | 0,13 | 0,45 | 368 |
| Viajes | 0,12 | 1,40 | 0,04 | 0,25 | 117 |
| Otros | 3,13 | 6,72 | 1,28 | 2,17 | 1.959 |

Clientes Medio Bajo consumo

| Rubro | Monto por cliente (UF) | SD (UF) | Nro. de transacciones por cliente | SD (#) | Clientes con transacciones |
|----------------------------------|-------------------------------|----------------|--|---------------|-----------------------------------|
| Avances | 4,54 | 14,98 | 0,79 | 2,18 | 5.904 |
| Bencina | 0,85 | 3,05 | 0,90 | 2,32 | 7.002 |
| Serv. financieros | 0,79 | 2,94 | 1,19 | 2,31 | 7.201 |
| Clínicas | 0,85 | 4,28 | 0,25 | 0,80 | 3.652 |
| Colegios | 0,59 | 6,13 | 0,04 | 0,34 | 486 |
| Compraventa Autos | 0,22 | 3,32 | 0,02 | 0,16 | 376 |
| Educación superior | 0,41 | 4,22 | 0,04 | 0,30 | 615 |
| Farmacias | 0,50 | 1,65 | 0,58 | 1,38 | 6.698 |
| Impuestos | 0,44 | 4,64 | 0,07 | 0,52 | 1.074 |
| Líneas aéreas | 1,91 | 7,86 | 0,28 | 0,98 | 3.699 |
| Tiendas por depto. | 2,67 | 6,44 | 1,13 | 1,97 | 11.190 |
| Reparación y acces. Autos | 0,49 | 2,89 | 0,11 | 0,52 | 2.006 |
| Restaurantes | 0,68 | 2,20 | 0,58 | 1,46 | 6.470 |
| Servicios básicos | 0,79 | 3,18 | 0,60 | 2,09 | 3.879 |
| Supermercados | 2,36 | 5,73 | 1,57 | 2,86 | 11.308 |
| Vestuario | 0,79 | 2,60 | 0,43 | 1,02 | 6.235 |
| Viajes | 0,44 | 4,71 | 0,06 | 0,35 | 1.031 |
| Otros | 7,61 | 12,43 | 3,94 | 4,16 | 19.937 |

Clientes Medio Alto consumo

| Rubro | Monto por cliente (UF) | SD (UF) | Nro. de transacciones por cliente | SD (#) | Clientes con transacciones |
|----------------------------------|-------------------------------|----------------|--|---------------|-----------------------------------|
| Avances | 29,81 | 62,38 | 1,14 | 2,17 | 1.066 |
| Bencina | 0,25 | 3,91 | 0,13 | 0,85 | 134 |
| Serv. financieros | 1,31 | 18,50 | 0,33 | 1,41 | 213 |
| Clínicas | 2,34 | 14,79 | 0,11 | 0,47 | 188 |
| Colegios | 4,77 | 26,36 | 0,12 | 0,64 | 122 |
| Compraventa Autos | 2,77 | 34,81 | 0,04 | 0,24 | 69 |
| Educación superior | 1,77 | 16,25 | 0,07 | 0,45 | 87 |
| Farmacias | 0,20 | 2,81 | 0,07 | 0,38 | 108 |
| Impuestos | 1,69 | 15,72 | 0,08 | 0,53 | 88 |
| Líneas aéreas | 9,10 | 31,65 | 0,39 | 1,39 | 515 |
| Tiendas por depto. | 3,69 | 16,27 | 0,35 | 0,96 | 473 |
| Reparación y acces. Autos | 1,42 | 14,90 | 0,07 | 0,33 | 131 |
| Restaurantes | 0,35 | 2,73 | 0,13 | 0,80 | 177 |
| Servicios básicos | 1,46 | 22,08 | 0,15 | 1,06 | 100 |
| Supermercados | 2,17 | 19,37 | 0,26 | 1,03 | 276 |
| Vestuario | 0,54 | 4,59 | 0,10 | 0,45 | 167 |
| Viajes | 4,23 | 22,45 | 0,19 | 0,80 | 260 |
| Otros | 13,96 | 39,72 | 1,31 | 2,64 | 1.095 |

Clientes Alto Consumo

| Rubro | Monto por cliente (UF) | SD (UF) | Nro. de transacciones por cliente | SD (#) | Clientes con transacciones |
|----------------------------------|-------------------------------|----------------|--|---------------|-----------------------------------|
| Avances | 8,20 | 36,65 | 1,40 | 4,53 | 1.455 |
| Bencina | 5,96 | 9,80 | 6,35 | 7,19 | 4.842 |
| Serv. financieros | 2,15 | 9,10 | 1,79 | 3,34 | 2.271 |
| Clínicas | 3,38 | 12,94 | 1,41 | 2,31 | 3.023 |
| Colegios | 2,66 | 16,14 | 0,15 | 0,76 | 387 |
| Compraventa Autos | 1,65 | 16,69 | 0,08 | 0,41 | 344 |
| Educación superior | 1,52 | 9,51 | 0,16 | 0,68 | 474 |
| Farmacias | 3,27 | 5,28 | 4,47 | 4,91 | 5.162 |
| Impuestos | 2,62 | 16,44 | 0,42 | 2,07 | 1.068 |
| Líneas aéreas | 14,19 | 3,68 | 0,47 | 1,53 | 1.280 |
| Tiendas por depto. | 8,17 | 14,60 | 4,54 | 4,73 | 5.352 |
| Reparación y acces. Autos | 1,49 | 8,13 | 0,38 | 1,15 | 1.318 |
| Restaurantes | 4,45 | 6,95 | 4,77 | 6,15 | 4.873 |
| Servicios básicos | 4,66 | 15,49 | 3,42 | 6,36 | 2.938 |
| Supermercados | 15,21 | 17,98 | 13,25 | 11,03 | 5.985 |
| Vestuario | 2,15 | 4,37 | 1,53 | 2,10 | 3.642 |
| Viajes | 1,23 | 9,68 | 0,11 | 0,49 | 445 |
| Otros | 26,63 | 34,65 | 18,69 | 10,41 | 6.244 |

Anexo G: Montos y número de compras promedio por tipo de transacción

Clientes Bajo consumo

| Tipo de pago | Monto por cliente (UF) | SD (UF) | Nro. de transacciones por cliente (#) | SD (#) |
|------------------------------|-------------------------------|----------------|--|---------------|
| Cuotas precio contado | 3,47 | 7,07 | 1,27 | 2,61 |
| Cuotas con interés | 2,14 | 6,89 | 0,35 | 0,91 |
| Compra sin cuotas | 2,76 | 7,48 | 2,02 | 3,91 |

Clientes Medio Bajo consumo

| Tipo de pago | Monto por cliente (UF) | SD (UF) | Nro. de transacciones por cliente (#) | SD (#) |
|------------------------------|------------------------|---------|---------------------------------------|--------|
| Cuotas precio contado | 8,73 | 14,20 | 3,73 | 5,19 |
| Cuotas con interés | 3,85 | 10,08 | 0,73 | 1,66 |
| Compra sin cuotas | 8,93 | 16,36 | 7,53 | 7,58 |

Clientes Medio Alto consumo

| Tipo de pago | Monto por cliente (UF) | SD (UF) | Nro. de transacciones por cliente (#) | SD (#) |
|------------------------------|------------------------|---------|---------------------------------------|--------|
| Cuotas precio contado | 15,44 | 36,45 | 1,21 | 2,47 |
| Cuotas con interés | 12,86 | 30,48 | 0,55 | 1,07 |
| Compra sin cuotas | 17,14 | 52,85 | 2,26 | 4,73 |

Clientes Alto consumo

| Tipo de pago | Monto por cliente (UF) | SD (UF) | Nro. de transacciones por cliente (#) | SD (#) |
|------------------------------|------------------------|---------|---------------------------------------|--------|
| Cuotas precio contado | 28,95 | 35,35 | 14,95 | 16,29 |
| Cuotas con interés | 11,92 | 31,49 | 1,89 | 4,03 |
| Compra sin cuotas | 50,03 | 58,68 | 45,13 | 24,99 |

Anexo H: Matriz de confusión árbol ID3 clientes inactivos

| | | Predicho | |
|------|----------|----------|----------|
| | | 0 | 1 |
| Real | 0 | 3.726 | 1.841 |
| | 1 | 1.530 | 1.614 |

Anexo I: Árbol ID3 clientes inactivos

Cientes con compra internacional: Primera mitad



Cientes sin compra internacional: Segunda mitad



Anexo J: Matriz de confusión Modelo Bajo consumo

| | | Predicho | |
|------|----------|----------|----------|
| | | 0 | 1 |
| Real | 0 | 1.024 | 697 |
| | 1 | 607 | 1.025 |

Anexo K: Resultados regresión logística Bajo consumo

| Variable | Coefficiente | P-Valor | Coefficiente estandarizado | Medida |
|--|-----------------------|---------|----------------------------|---------|
| Aumento periodo 0 | -0,64 | <0,01 | -0,27 | Binaria |
| Intercepto | -1,11 | <0,01 | -0,16 | -- |
| Semanas con TRX periodo 1 | 0,09 | <0,01 | 0,16 | Entero |
| Con avance | 0,30 | <0,01 | 0,12 | Binaria |
| Semanas con TRX periodo 0 | 0,03 | <0,01 | 0,12 | Entero |
| Revolving promedio mensual | 0,23 | <0,01 | 0,11 | Binaria |
| Cantidad de rubros periodo 0 | 0,05 | <0,01 | 0,1 | Entero |
| Con compra internacional | 0,24 | <0,01 | 0,1 | Binaria |
| Monto compra sin cuotas | $-5,8 \times 10^{-7}$ | <0,01 | -0,1 | CLP |
| Aumenta compra en cuotas | 0,20 | <0,01 | 0,1 | Binaria |
| Aumenta N° de compras en rubros habituales | 0,02 | <0,01 | 0,09 | Binaria |
| Edad | $5,8 \times 10^{-3}$ | <0,01 | 0,07 | Entero |
| Con tarjeta GOLD | 0,15 | <0,01 | 0,07 | Binaria |
| Con compra en "Otros" | 0,12 | 0,01 | 0,06 | Binaria |
| Con compra tiendas por departamento | 0,13 | 0,01 | 0,06 | Binaria |
| Con compras por internet | 0,11 | 0,04 | 0,05 | Binaria |

Anexo L: Matriz de confusión regresión logística Medio Bajo consumo

| | | Predicho | |
|------|----------|----------|----------|
| | | 0 | 1 |
| Real | 0 | 13.627 | 6.079 |
| | 1 | 652 | 1.384 |

Anexo M: Resultados regresión logística Medio Bajo consumo

| Variable | Coefficiente | P-Valor | Coefficiente estandarizado | Medida |
|--|----------------------|---------|----------------------------|---------|
| Intercepto | -2,20 | <0,01 | -0,95 | -- |
| Monto transado periodo 0 | $3,1 \times 10^{-7}$ | <0,01 | 0,31 | CLP |
| Aumento periodo 0 | 0,66 | <0,01 | 0,23 | Binaria |
| Cantidad de rubros | 0,10 | <0,01 | 0,21 | Entero |
| Con avance | 0,45 | <0,01 | 0,18 | Binaria |
| N° compras "Otros rubros" | 0,04 | <0,01 | 0,18 | Entero |
| Cupo utilizado final | -0,51 | <0,01 | -0,17 | % |
| Con TRX en Supermercado | 0,29 | <0,01 | 0,14 | Binaria |
| Con tarjeta GOLD | -0,29 | <0,01 | -0,13 | Binaria |
| Semanas con TRX periodo 0 | 0,03 | <0,01 | 0,12 | Entero |
| Con compra por internet | 0,25 | <0,01 | 0,12 | Binaria |
| Monto transado en Bencina | $9,7 \times 10^{-7}$ | <0,01 | 0,10 | CLP |
| Monto compra sin cuotas | $2,1 \times 10^{-7}$ | <0,01 | 0,10 | CLP |
| Aumento de cupo | 0,45 | <0,01 | 0,10 | Binaria |
| Con TRX en Restaurantes | 0,20 | <0,01 | 0,09 | Binaria |
| Es hombre | 0,13 | <0,01 | 0,06 | Binaria |
| Tiene entre 35 y 55 años | 0,09 | <0,01 | 0,04 | Binaria |
| Con tarjeta BLACK | 0,17 | 0,01 | 0,04 | Binaria |
| Cantidad de canjes de puntos periodo 0 | 0,07 | 0,08 | 0,03 | Binaria |

Anexo N: Rentabilidad por nivel de consumo

| Nivel de consumo | Rentabilidad 4 meses (UF) |
|-------------------|---------------------------|
| Inactivos | 0,2 |
| Bajo | 0,9 |
| Medio Bajo | 1,4 |
| Medio Alto | 2,4 |
| Alto | 2,2 |

Anexo Ñ: Porcentaje de aumento hacia cada nivel

| Nivel de origen | Nivel de destino | # aumentos | % del segmento |
|------------------------|-------------------------|-------------------|-----------------------|
| Inactivos | Bajo | 712 | 7% |
| Inactivos | Medio Bajo | 1.772 | 18% |
| Inactivos | Medio Alto | 602 | 6% |
| Inactivos | Alto | 58 | 1% |
| Bajo | Medio Bajo | 1.395 | 38% |
| Bajo | Medio Alto | 188 | 5% |
| Bajo | Alto | 49 | 1% |
| Medio Bajo | Medio Alto | 563 | 2% |
| Medio Bajo | Alto | 1.473 | 6% |
| Medio Alto | Alto | 44 | 2% |

Anexo O: Rentabilidad ponderada del aumento de consumo por nivel

| Origen | Rentabilidad (UF) |
|-------------------|--------------------------|
| Inactivos | 1,26 |
| Bajo | 0,64 |
| Medio Bajo | 0,86 |

Anexo P: Desagregación escenarios metodología clientes Inactivos

Escenario pesimista

| | Actual | Metodología |
|-----------------------------|---------------|--------------------|
| Costo envío | \$2.613.300 | \$2.613.300 |
| Costo diseño | \$500.000 | \$500.000 |
| Costo devolución | \$5.226.600 | \$5.263.720 |
| Comisión a comercios | \$979.988 | \$526.464 |
| Descuento adquirente | \$293.996 | \$157.939 |
| Aumento de consumo | \$0 | \$1.867.150 |
| Total | -\$7.653.909 | -\$6.141.345 |
| Ganancia total | | \$1.512.563 |

Escenario optimista

| | Actual | Metodología |
|-----------------------------|---------------|--------------------|
| Costo envío | \$2.613.300 | \$2.613.300 |
| Costo diseño | \$500.000 | \$500.000 |
| Costo devolución | \$11.498.520 | \$12.715.480 |
| Comisión a comercios | \$2.155.973 | \$1.307.256 |
| Descuento adquirente | \$646.792 | \$392.177 |
| Aumento de consumo | \$0 | \$5.601.450 |
| Total | -\$13.102.639 | -\$9.312.251 |
| | | |
| Ganancia total | | \$3.790.388 |

Anexo Q: Desagregación escenarios metodología clientes Bajo consumo

Escenario Pesimista

| | Actual | Metodología |
|---------------------------------|---------------|--------------------|
| Costo envío | \$1.005.900 | \$430.000 |
| Costo envío recordatorio | \$0 | \$172.200 |
| Costo diseño piezas | \$500.000 | \$1.500.000 |
| Costo devolución | \$2.011.800 | \$572.200 |
| Comisión a comercios | \$377.213 | \$107.288 |
| Descuento adquirente | \$113.164 | \$32.186 |
| Aumento de consumo | \$0 | \$427.191 |
| Costo Total | -\$3.253.651 | -\$2.172.107 |
| | | |
| Ganancia Modelo | | \$1.081.544 |

Escenario Optimista

| | Actual | Metodología |
|---------------------------------|---------------|--------------------|
| Costo envío | \$1.005.900 | \$430.000 |
| Costo envío recordatorio | \$0 | \$172.200 |
| Costo diseño piezas | \$500.000 | \$1.500.000 |
| Costo devolución | \$4.425.960 | \$1.716.600 |
| Comisión a comercios | \$829.868 | \$321.863 |
| Descuento adquirente | \$248.960 | \$96.559 |
| Aumento de consumo | \$0 | \$1.254.548 |
| Costo Total | -\$5.350.953 | -\$2.338.949 |
| | | |
| Ganancia Modelo | | \$3.012.004 |

Anexo R: Desagregación escenarios metodología clientes Medio Bajo consumo

Escenario Pesimista

| | Actual | Metodología |
|---------------------------------|---------------|--------------------|
| Costo envío | \$6.522.600 | \$3.531.300 |
| Costo envío recordatorio | \$0 | \$746.300 |
| Costo diseño piezas | \$500.000 | \$1.000.000 |
| Costo devolución | \$13.045.200 | \$5.486.880 |
| Comisión a comercios | \$2.445.975 | \$1.028.790 |
| Descuento adquirente | \$733.793 | \$308.637 |
| Aumento de consumo | \$0 | \$1.054.493 |
| Costo Total | -\$17.905.618 | -\$6.989.834 |
| | | |
| Ganancia Modelo | | \$9.365.783 |

Escenario Optimista

| | Actual | Metodología |
|---------------------------------|---------------|---------------------|
| Costo envío | \$6.522.600 | \$3.531.300 |
| Costo envío recordatorio | \$0 | \$0 |
| Costo diseño piezas | \$500.000 | \$1.000.000 |
| Costo devolución | \$28.699.440 | \$17.998.680 |
| Comisión a comercios | \$5.381.145 | \$3.374.753 |
| Descuento adquirente | \$1.614.344 | \$1.012.426 |
| Aumento de consumo | \$0 | \$3.673.220 |
| Costo Total | -\$31.955.239 | -\$14.494.433 |
| | | |
| Ganancia Modelo | | \$15.460.806 |

Anexo S: Resumen evaluación económica por escenarios

Escenario Pesimista

| Pesimista | Valor 4 meses | Valor anual |
|--|----------------------|----------------------|
| Valor Modelo Inactivos | \$ 1.512.563 | \$ 4.537.690 |
| Valor Modelo Bajo Consumo | \$ 1.081.544 | \$ 3.244.631 |
| Valor Modelo Medio Bajo Consumo | \$ 9.365.783 | \$ 28.097.350 |
| Total | \$ 11.959.891 | \$ 35.879.672 |

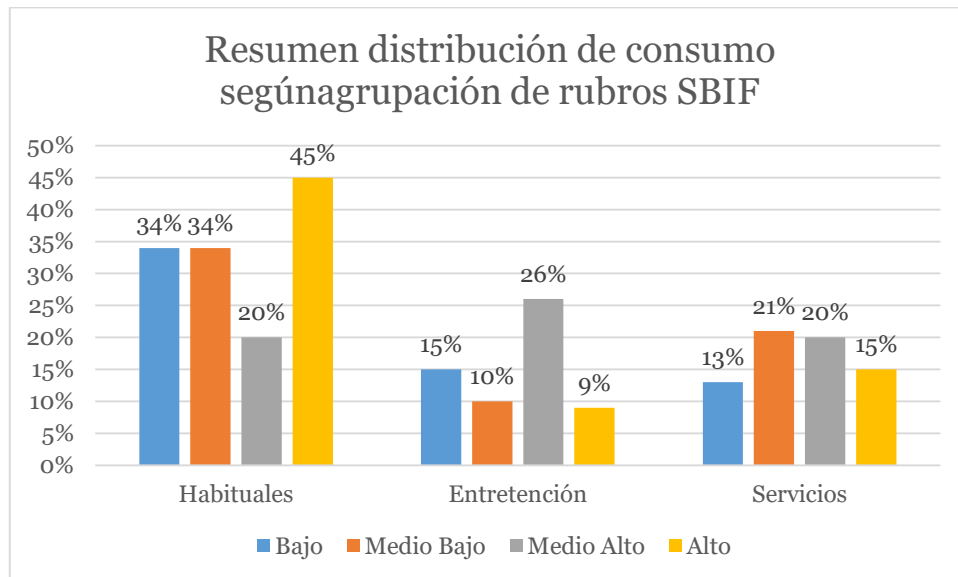
Escenario Realista

| Realista | Valor 4 meses | Valor anual |
|--|----------------------|----------------------|
| Valor Modelo Inactivos | \$ 2.651.476 | \$ 7.954.427 |
| Valor Modelo Bajo Consumo | \$ 2.051.253 | \$ 6.153.759 |
| Valor Modelo Medio Bajo Consumo | \$ 12.524.136 | \$ 37.572.409 |
| Total | \$ 17.226.865 | \$ 51.680.596 |

Escenario Optimista

| Optimista | Valor 4 meses | Valor anual |
|--|----------------------|----------------------|
| Valor Modelo Inactivos | \$ 3.790.388 | \$ 11.371.165 |
| Valor Modelo Bajo Consumo | \$ 3.012.004 | \$ 9.036.012 |
| Valor Modelo Medio Bajo Consumo | \$ 15.460.806 | \$ 46.382.417 |
| Total | \$ 22.263.198 | \$ 66.789.594 |

Anexo T: Resumen distribución de consumo por clúster



Anexo U: Variables árbol ID3 clientes inactivos

| Variable | Medida |
|----------------------------------|------------------|
| # compras internac. per 1 | Cantidad |
| Cupo utilizado | % |
| # compras en habituales per 0 | Cantidad |
| # compras por internet per 0 | Cantidad |
| Revolving final | Monto CLP |
| Antigüedad | Tramos por meses |
| # compras internac. per 0 | Cantidad |
| Con trx supermerc. per 0 | Sí/No |
| Monto transado vestuario per 0 | Monto CLP |
| Cliente Medio Bajo consumo per 0 | Sí/No |
| Monto transado restaurant per 0 | Monto CLP |
| Monto transado en cuotas per 0 | Monto CLP |
| Con Tarjeta Gold | Sí/No |
| Monto Mora grave | Monto CLP |
| Pagos a la tarjeta per 1 | Cantidad |
| Con avance per 0 | Sí/No |
| Monto tiendas por depto. Per 0 | Monto CLP |
| Pagos PAT per 0 | Cantidad |
| Cierra cuenta | Sí/No |
| Monto transado en serv. Básicos | Monto CLP |
| Mejora tipo de tarjeta | Sí/No |
| Semanas con trx per 0 | Cantidad |
| Con trax nacional per 0 | Sí/No |
| Meses con pago a la trj. per 0 | Cantidad |
| Nro. Colegios 0 | Cantidad |
| Puntos canjeados per 0 | Cantidad |
| Frecuencia per 0 | Cantidad |