



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

DESARROLLO DE MODELO DE PROBABILIDAD DE INCUMPLIMIENTO PARA LA  
CARTERA DE CONSUMO DEL BANCO BCI

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

ALEXANDER OSVALDO SANTIBÁÑEZ HUIAQUIMILLA

PROFESOR GUÍA:  
JOSÉ MIGUEL CRUZ GONZÁLEZ

MIEMBROS DE LA COMISIÓN:  
CARLOS PULGAR ARATA  
FERNANDO ORDOÑEZ PIZARRO

SANTIAGO DE CHILE  
2018

**RESUMEN DE LA MEMORIA PARA OPTAR AL  
TÍTULO DE:** Ingeniero Civil Industrial  
**POR:** Alexander Santibáñez Huaiquimilla  
**FECHA:** 11/06/2018  
**PROFESOR GUÍA:** José Miguel Cruz González

**DESARROLLO DE UN MODELO DE PROBABILIDAD DE  
INCUMPLIMIENTO PARA LA CARTERA DE CONSUMO DEL BANCO BCI**

El objetivo general de esta memoria fue proponer un modelo de probabilidad de incumplimiento para la cartera de consumo del banco, el cual posea una mayor capacidad predictiva que el modelo actual, a través de la incorporación de variables no tradicionales.

En primer lugar, se realizó la reconstrucción del modelo actual que el banco posee para contar con la situación base y poder comparar con los resultados del nuevo modelo que se propone. Al finalizar esta etapa, no se llegó a un modelo idéntico al del banco. Sin embargo, la réplica es bastante similar en el desempeño, medido en KS, con un valor de 65.83%, con una diferencia de -0,03% al resultado del banco.

La segunda etapa fue, construir las variables no tradicionales que se incluirían a las propuestas de modelos. Para ello, lo que se realizó fue la interacción de productos, para capturar el efecto no lineal. Para ello se multiplicó variables como: deuda consumo con deuda hipotecaria, monto en acciones con monto en fondos mutuos, edad con deuda, entre otros. De este proceso se crearon 182 nuevas variables.

Luego, se realiza 3 propuestas de modelos: incluir nuevas variables a segmentación actual del banco, generar segmentos nuevos de la actual segmentación del banco y una segmentación completamente nueva.

Posteriormente, se comparó las 3 propuestas de modelo con la réplica del modelo que se construyó. Esto se realizó mediante el estadístico KS, ROC y estimación de las provisiones. Donde una de las propuestas obtuvo mejor desempeño que la réplica, con un KS de 66,38%.

Para finalizar, se concluye que la segmentación actual del banco es la óptima y que la incorporación de variables no tradicionales tuvo un efecto positivo en el desempeño del modelo que se propone.

## Tabla de contenido

<b>Capítulo I. Introducción</b>	1
1.1 Antecedentes	1
1.2 Objetivos	3
1.2.1 Objetivo general	3
1.2.2 Objetivos específicos	3
1.3 Alcances y delimitaciones	3
<b>Capítulo II. Marco conceptual</b>	4
2.1 Proceso KDD	5
2.2 Árboles de decisión	6
2.3 Regresión logística	7
2.4 Medición de desempeño	9
2.4.1 Estadístico Kolmogorov-Smirnov	9
2.4.2 Curva ROC	10
<b>Capítulo III. Metodología</b>	11
3.1 Comprensión del negocio	11
3.2 Comprensión de los datos	11
3.3 Elaboración de base de datos	12
3.4 Segmentación	12
3.5 Modelamiento	14
3.5.1 Análisis univariado	14
3.5.2 Análisis de correlación	14
3.5.3 Análisis multivariado	15
3.5.4 Análisis de contribución a logverosimilitud	15
3.6 Medición de desempeño y elección de modelo	16
<b>Capítulo IV. Preparación de datos</b>	17
4.1 Periodo de información	17
4.2 Construcción de variable dependiente	17
4.3 Información disponible	18
4.4 Tratamiento de datos	19
<b>Capítulo V. Modelo actual del banco BCI</b>	20
5.1 Segmentación de réplica del modelo actual del banco	20
5.2 Modelamiento de réplica modelo actual del banco	23

5.3	Desempeño replica modelo actual del banco	23
5.4	Comparación resultados del modelo replicado	24
<b>Capítulo VI. Diseño de nuevos modelos</b>		32
6.1	Propuesta de modelo n°1	33
6.1.1	Segmentación propuesta de modelo n°1	33
6.1.2	Modelamiento de propuesta de modelo n°1	33
6.1.3	Medición de desempeño de propuesta de modelo n°1	37
6.2	Propuesta de modelo n°2	38
6.2.1	Segmentación de propuesta de modelo n°2	38
6.2.2	Modelamiento de propuesta de modelo n°2	45
6.2.3	Medición de desempeño de propuesta de modelo n°2	49
6.3	Propuesta de modelo n°3	50
6.3.1	Segmentación de propuesta de modelo n°3	50
6.3.2	Modelamiento de propuesta de modelo n°3	61
6.3.3	Medición de desempeño de propuesta de modelo n°3	65
<b>Capítulo VII. Comparación y elección de modelo</b>		66
<b>Capítulo VIII. Conclusiones y recomendaciones</b>		72
<b>Bibliografía</b>		74
<b>Anexos</b>		75

## Índice de tablas

Tabla 1 - Matriz de confusión .....	10
Tabla 2 - Ejemplo de variable tramificada.....	13
Tabla 3 - Número de variables a incluir en el modelo de cada segmento del modelo replicado. ....	23
Tabla 4 - Comparación número de variables por segmento de modelo del banco y modelo replicado.....	24
Tabla 5 - Comparación de variables de modelo actual replicado con variables del modelo actual del banco. Segmento 1. ....	25
Tabla 6 - Comparación de variables de modelo actual replicado con variables del modelo actual del banco. Segmento 2. ....	26
Tabla 7 - Comparación de variables de modelo actual replicado con variables del modelo actual del banco. Segmento 3. ....	26
Tabla 8 - Comparación de variables de modelo actual replicado con variables del modelo actual del banco. Segmento 4. ....	27
Tabla 9 - Comparación de variables de modelo actual replicado con variables del modelo actual del banco. Segmento 5. ....	28
Tabla 10 - Comparación de variables de modelo actual replicado con variables del modelo actual del banco. Segmento 6.....	28
Tabla 11 - Comparación de variables de modelo actual replicado con variables del modelo actual del banco. Segmento 1.....	29
Tabla 12 - Comparación de desempeño de modelos mediante KS.....	29
Tabla 13 - tabla comparación de desempeño de modelo replicado con variables del banco y su información entregada para desempeño de su modelo .....	30
Tabla 14 - Número de variables a incluir en modelo propuesta N°1. ....	33
Tabla 15 - Variables y coeficientes incluidos en segmento 1 de propuesta de modelo N°1.....	34
Tabla 16 - Variables y coeficientes incluidos en segmento 2 de propuesta de modelo N°1.....	34
Tabla 17 - Variables y coeficientes incluidos en segmento 3 de propuesta de modelo N°1.....	35
Tabla 18 - Variables y coeficientes incluidos en segmento 4 de propuesta de modelo N°1.....	35
Tabla 19 - Variables y coeficientes incluidos en segmento 5 de propuesta de modelo N°1.....	36
Tabla 20 - Variables y coeficientes incluidos en segmento 6 de propuesta de modelo N°1.....	37
Tabla 21 - Variables y coeficientes incluidos en segmento 7 de propuesta de modelo N°1.....	37
Tabla 22 - Ranking IV con primeras las 5 variables para segmentación del segmento 1. Propuesta N°2.....	39
Tabla 23 - Ranking IV con primeras las 5 variables para segmentación del segmento 3. Propuesta N°2.....	41

Tabla 24 - Ranking IV con primeras las 5 variables para segmentación del segmento 5. Propuesta N°2.....	43
Tabla 25 - Número de variables a incluir en modelo, propuesta N°2 .....	46
Tabla 26 - Variables y coeficientes incluidos en segmento 1.1 de propuesta de modelo N°2 .....	46
Tabla 27 - Variables y coeficientes incluidos en segmento 1.2 de propuesta de modelo N°2 .....	47
Tabla 28 - Variables y coeficientes incluidos en segmento 3.1 de propuesta de modelo N°2 .....	47
Tabla 29 - Variables y coeficientes incluidos en segmento 3.2 de propuesta de modelo N°2 .....	48
Tabla 30 - Variables y coeficientes incluidos en segmento 5.1 de propuesta de modelo N°2 .....	49
Tabla 31 - Variables y coeficientes incluidos en segmento 5.2 de propuesta de modelo N°2 .....	49
Tabla 32 - Ranking IV con primeras las 5 variables para segmentar a segmento "No Renegociados". Propuesta N°3.....	51
Tabla 33 - Ranking IV con primeras las 5 variables para segmentar a segmento "No Renegociados - Renta<=41UF". Propuesta N°3. ....	53
Tabla 34 - Ranking IV con primeras las 5 variables para segmentar a segmento "No Renegociados - 41UF<Renta". Propuesta N°3.....	53
Tabla 35- Ranking IV con primeras las 5 variables para segmentar a segmento "No Renegociados - Renta<=41UF - Con Avances últimos 12 meses". Propuesta N°3.....	55
Tabla 36 - Ranking IV con primeras las 5 variables para segmentar a segmento "No Renegociados - Renta<=41UF - Sin Avances últimos 12 meses". Propuesta N°3.....	56
Tabla 37 - Ranking IV con primeras las 5 variables para segmentar a segmento "No Renegociados - 41UF<Renta - #TDC<=2". Propuesta N°3.....	56
Tabla 38 - Ranking IV con primeras las 5 variables para segmentar a segmento "No Renegociados - 41UF<Renta - 2<#TDC". Propuesta N°3.....	57
Tabla 39 - Número de variables a incluir en modelo propuesta N°3.....	61
Tabla 40 - Variables y coeficientes incluidos en segmento 1 de propuesta de modelo N°3.....	61
Tabla 41 - Variables y coeficientes incluidos en segmento 2 de propuesta de modelo N°3.....	62
Tabla 42 - Variables y coeficientes incluidos en segmento 3 de propuesta de modelo N°3.....	62
Tabla 43 - Variables y coeficientes incluidos en segmento 4 de propuesta de modelo N°3.....	63
Tabla 44 - Variables y coeficientes incluidos en segmento 5 de propuesta de modelo N°3.....	63

Tabla 45 - Variables y coeficientes incluidos en segmento 6 de propuesta de modelo N°3.....	64
Tabla 46 - Variables y coeficientes incluidos en segmento 7 de propuesta de modelo N°3.....	65
Tabla 47- Estimación de provisión bruta de cada modelo. ....	69
Tabla 48 – Mejor modelo dependiendo del criterio escogido. ....	71

## Índice de ilustraciones

Ilustración 1 – Proceso KDD.....	5
Ilustración 2 - Ejemplo de árbol de decisión .....	6
Ilustración 3 - Ejemplo de estadístico KS .....	9
Ilustración 4 - Ejemplo de curva ROC.....	10
Ilustración 5 - Construcción de variable de desempeño. ....	12
Ilustración 6 - Segmento luego de cuarta segmentación. Fuente: banco BCI. .....	21
Ilustración 7 – Segmentación propuesta n°2.....	45
Ilustración 8 – Resultado de primera segmentación generada. Propuesta N°3. .....	51
Ilustración 9 – Resultado de segunda segmentación generada. Propuesta N°3.....	54
Ilustración 10 – Resultado de tercera segmentación generada. Propuesta N°3 .....	57
Ilustración 11 – Resultado de segmentación final generada. Propuesta N°3	60



## Índice de gráficos

Gráfico 1 – Evolución de la distribución de las colocaciones del banco BCI. ..	1
Gráfico 2 – Ejemplo de análisis de contribución a la logverosimilitud .....	16
Gráfico 3 – Distribución de clientes por mes.....	17
Gráfico 4 – Distribución de clientes que caen en incumplimiento y los que no. .....	18
Gráfico 5 - Cantidad de clientes por segmento y porcentaje que representan de la cartera de consumo. Elaboración propia a partir de BD. ....	22
Gráfico 6 – Bad rate asociado a cada segmento. Modelo actual del banco. .	22
Gráfico 7 – Desempeño medido en Ks y ROC del modelo actual replicado. .	23
Gráfico 8 – Comparación de KS para los diferentes modelos. ....	30
Gráfico 9 – Medición de desempeño en KS y ROC de propuesta de modelo N°1.....	38
Gráfico 10 - Análisis de tamaño de segmentos. ....	39
Gráfico 11- Distribución de variable “porcentaje de uso de línea de crédito” en el segmento 1. ....	40
Gráfico 12 – Tamaño de segmentos generados por la segmentación a partir de la variable “porcentaje de uso de línea de crédito” .....	40
Gráfico 13 – Comparación de Bad Rate de segmento 1 versus nuevos segmentos generados en la propuesta N°2.....	41
Gráfico 14 – Distribución de clientes con y sin línea de crédito en segmento 3.....	42
Gráfico 15 – Tamaño segmentos generados por la segmentación a partir de la variable “Marca solicitó avances en los últimos 12 meses” .....	42
Gráfico 16 – comparación de Bad Rate segmento 3 versus nuevos segmentos generados en la propuesta N°2.....	43
Gráfico 17 – Distribución de clientes con y sin línea de crédito en segmento 5.....	44
Gráfico 18 – tamaño de segmentos generados por la segmentación a partir de la variable “Marca solicitó avances en los últimos 6 meses”.....	44
Gráfico 19 – Comparación de Bad Rate segmento 5 versus nuevos segmentos generados en la propuesta N°2.....	45
Gráfico 20 - Medición de desempeño en KS y ROC de propuesta de modelo N°2. ....	50
Gráfico 21 – Distribución de renta en UF de los clientes del segmento “No Renegociados”.....	51
Gráfico 22 – Comparación de Bad Rate de primera segmentación. Propuesta N°3. ....	52
Gráfico 23 – Evolución temporal de Bad Rate de primera segmentación. Propuesta N°3. ....	52
Gráfico 24 – Comparación Bad Rate de segunda segmentación. Propuesta N°3.....	54
Gráfico 25 – Evolución temporal de Bad Rate de segunda segmentación. Propuesta N°3. ....	55

Gráfico 26 – Comparación de Bad Rate de tercera segmentación. Propuesta N°3 .....	58
Gráfico 27 – Evolución temporal de Bad Rate de tercera segmentación. Propuesta N°3 .....	58
Gráfico 28 – Evolución temporal de Bad Rate con segmentación final. Propuesta N°3 .....	59
Gráfico 29 – Comparación Bad Rate de segmentación final. Propuesta N°3.	60
Gráfico 30 - Medición de desempeño en KS y ROC de propuesta de modelo N°3 .....	65
Gráfico 31 – Comparación de modelo actual replicado con propuesta N°1, mediante KS y ROC.....	66
Gráfico 32 Comparación de modelo actual replicado con propuesta N°2, mediante KS y ROC.....	67
Gráfico 33 – Evolución temporal de Bad Rate se propuesta n°2. ....	68
Gráfico 34 - Comparación de modelo actual replicado con propuesta N°3, mediante KS y ROC.....	68
Gráfico 35 – Comparación Bad Rate Real vs propuesta N°1. Cartera total. .	70
Gráfico 36 - Comparación Bad Rate Real vs propuesta N°3. Cartera total. .	71

# Capítulo I. Introducción

## 1.1 Antecedentes

El Banco de Créditos e Inversiones (BCI) es una institución que se desempeña en el sistema financiero de Chile, en particular, en la industria bancaria, cuyo giro básico es, captar dinero del público con el objetivo de darlo en préstamo (colocaciones).

Para septiembre del 2017, BCI se ubicó en el tercer lugar de la industria bancaria chilena (medida en colocaciones) con un 15.37% de participación de mercado, de un total de 23 bancos que se encontraron activos en dicho periodo.

El negocio de créditos del banco BCI se divide en 3 grandes carteras: comercial, vivienda y consumo, de las cuales, a continuación, se muestra cómo ha ido evolucionando su participación en las colocaciones del banco.

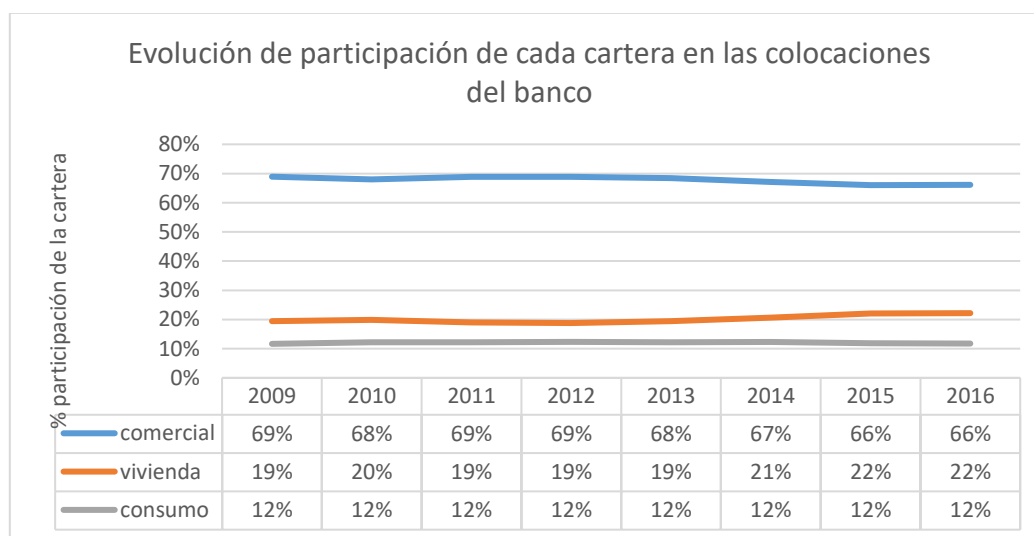


Gráfico 1 – Evolución de la distribución de las colocaciones del banco BCI.  
Fuente: Memoria 2016 del banco BCI

Como se aprecia en el Gráfico 1, la contribución de cada cartera a las colocaciones se ha mantenido estable entre el 2009 y el 2016, siendo la de mayor participación la cartera comercial, con cerca del 66%, seguida por la de vivienda, con alrededor del 22% y finalmente la de consumo, con un 12%.

Pero con cada colocación que realiza el banco viene, asociado un riesgo de que el cliente al que le fue otorgado el préstamo deje de cumplir con sus obligaciones al plazo esperado y no siga pagando. Ésto se conoce como el riesgo de crédito.

Dada la existencia del riesgo de crédito es que cada banco busca resguardarse frente a posibles pérdidas, además, existen normativas que tienen como finalidad que los bancos se resguarden de forma adecuada, y así cada banco pueda responder con sus obligaciones con quienes le prestan dinero. Esto es conocido como costo esperado en un año.

La entidad encargada de fiscalizar la industria es la Superintendencia de Bancos e Instituciones Financieras (SBIF), y su accionar se enmarca en la Ley General de Bancos. El propósito de dicha institución es proteger a los depositantes u otros acreedores y el interés público, cuidando el buen funcionamiento del sistema financiero.

La normativa vigente señala que las instituciones deben los modelos o métodos de evaluación más apropiado para cada tipo de cartera, estos pueden ser: análisis individual de los deudores o análisis grupal.

Dado que la cartera que se desea evaluar en esta memoria es la de consumo, resulta más adecuado un modelo de análisis grupal, pues los montos individuales son bajos y pertenecientes a personas naturales.

Para desarrollar dichos modelos existen dos métodos, los cuales son, estimar pérdidas esperadas a través de la experiencia recogida, mediante ejecución de garantías y acciones de cobranza, el otro método que pueden utilizar los bancos es, segmentar a sus deudores en grupos homogéneos y a cada segmento asociarle una probabilidad de incumplimiento y porcentaje de recuperación, basado en un análisis histórico.

Para poder desarrollar un modelo interno, algunos de los requisitos que se deben cumplir son: metodologías internas estiman provisiones por tipo de deudos en los segmentos que sean pertinente a cada cartera. Las segmentaciones son deseables si es que permiten una mayor discriminación del riesgo. Los scoring deben poseer un alto grado de predictibilidad de la calidad crediticia de los deudores, además para su construcción se debe poseer un horizonte de mínimo 5 años. Las variables evaluadas en la metodología se basan en atributos del deudos y características de sus créditos. Elaboración de un diccionario de variables que contiene la base de datos.

Entre los beneficios que trae para el banco poseer un modelo capaz de identificar cuál es el riesgo que representa cada cliente en una cartera tan masiva, se encuentra poseer una cantidad adecuada de provisiones, además mejorar la gestión del riesgo de la cartera mediante seguimiento a los segmentos con mayor probabilidad de caer en incumplimiento de sus obligaciones, ofrecer tasas diferenciadas, entre otros.

## **1.2 Objetivos**

### **1.2.1 Objetivo general**

El objetivo es proponer un modelo de probabilidad de incumplimiento para la cartera de consumo del banco BCI con una mayor capacidad predictiva que el modelo actual del banco.

### **1.2.2 Objetivos específicos**

Los objetivos específicos de la memoria se mencionan a continuación:

- i. Comprender el modelo que actualmente posee el banco, las variables que lo componen y su capacidad predictiva.
- ii. Generar nuevas variables que potencialmente pueden ser incorporadas en el modelo.
- iii. Seleccionar modelo para la cartera de consumo.

## **1.3 Alcances y delimitaciones**

Dentro del alcance que posee esta memoria es que, el modelo que se propondrá y los resultados obtenidos podrán ser aplicable solo a la cartera de consumo del Banco BCI, ya que todos los datos utilizados para su elaboración fueron obtenidos en esta institución.

Para la elaboración del modelo que se propondrá, no se considerará información del mercado, como indicadores macroeconómicos, que pudieran ayudar a elaborar un mejor modelo. El modelo será construido solo con las variables que posee el banco y la elaboración de nuevas variables en base a éstas.

Queda fuera del alcance del trabajo de título que el modelo propuesto sea implementado por la empresa. El motivo de lo mencionado anteriormente se debe a qué para poder implementar un modelo interno, primero debe ser revisada y aprobada por el Directorio, luego debe ser revisado por la SBIF, dicho proceso toma un tiempo considerable.

## Capítulo II. Marco conceptual

Antes de describir el marco conceptual, es necesario definir algunos conceptos de la industria bancaria, tales como:

- Mora: cantidad de días de atraso que lleva un cliente de la cuota impaga más antigua.
- Provisiones: fondos necesarios para cubrir pérdidas de valor del activo como para hacer frente a potenciales obligaciones que aún no se ha materializado.
- Incumplimiento: se define por convención que un deudor cae en incumplimiento de su deuda cuando tiene una mora mayor a 90 días en un horizonte de un año.
- Pérdida esperada (PE): pérdida promedio que se espera tener en un horizonte determinando (típicamente un año). Este es el valor que debe provisionar el banco y queda determinado por: probabilidad de incumplimiento (PD), porcentaje de pérdida dado el incumplimiento (LGD) y exposición al incumplimiento (EAD), mediante la siguiente expresión:

$$PE = PD \cdot LGD \cdot EAD$$

- Probabilidad de incumplimiento: es la probabilidad con la que un determinado deudor caerá en incumplimiento en 12 meses.
- Porcentaje de pérdida dado el incumplimiento: porcentaje de la exposición que no se recupera cuando un deudor cae en incumplimiento.
- Exposición dado el incumplimiento: deuda que posee el cliente al momento de caer en incumplimiento.

Para construir un modelo de probabilidad de incumplimiento, que cumpla con la normativa, es necesario trabajar una base de datos histórica y a partir de dichos datos lograr predecir el comportamiento de quienes caen en incumplimiento. Lo anterior puede realizarse mediante el proceso KDD, pues es una forma de descubrir conocimiento a partir de los datos. Además, se busca agrupar a los deudores en segmentos homogéneos, esto se realizará mediante árboles de decisión, luego para modelar el comportamiento de los clientes que caen en incumplimiento y los que no lo hacen, es decir, una variable *dummy*, una técnica apropiada es el modelo de regresión logística. Finalmente, es necesario poseer métricas de desempeño del modelo construido.

## 2.1 Proceso KDD

El proceso KDD, de la sigla *Knowledge Discovery in databases*, que traducida al español sería, descubrimiento de conocimientos en bases de datos, es el proceso busca descubrir conocimientos e información útil dentro de los datos que se poseen.

El proceso se divide en cinco etapas: selección, preprocesamiento, transformación, minería de datos e interpretación, y evaluación. En la Ilustración 1 se muestran las etapas mencionadas.

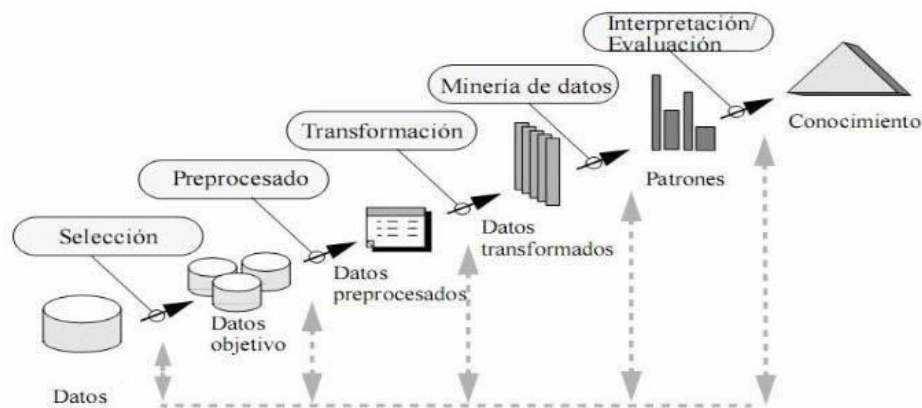


Ilustración 1 – Proceso KDD

A continuación, se describen las cinco etapas del proceso KDD.

- Selección: la primera etapa consiste en determinar las bases de datos que se utilizarán.
- Preprocesamiento: en esta etapa se realiza una limpieza de los datos. Es decir, se solucionan los problemas con los datos faltantes, inconsistencias u otros.
- Transformación: en esta etapa se realiza la construcción de variables a partir de los datos que se poseen.
- Minería de datos: esta es la etapa de modelamiento, donde se utiliza la técnica más acorde al problema que se enfrenta. Los métodos entregarán los patrones encontrados.
- Interpretación y evaluación: se interpretan los resultados de la etapa previa y se evalúan los resultados que se obtuvo.

## 2.2 Árboles de decisión

Los árboles de decisión son un método de clasificación, el cual descompone un conjunto de datos en dos o más subconjuntos de datos, donde exista diferencia estadísticamente significativa de la variable dependiente entre los subconjuntos. Para ello se debe seleccionar una variable como objetivo o dependiente y variables independientes con las que se generarán las "hojas" del árbol. El árbol define tramos para las variables continuas y asocia características para variables nominales. El objetivo del árbol de decisión es generar segmentos que tengan diferencias estadísticamente significativas entre sí en cuanto a la distribución de la variable dependiente.

Un ejemplo de la visualización de un árbol de decisión de aprecia en la Ilustración 2.

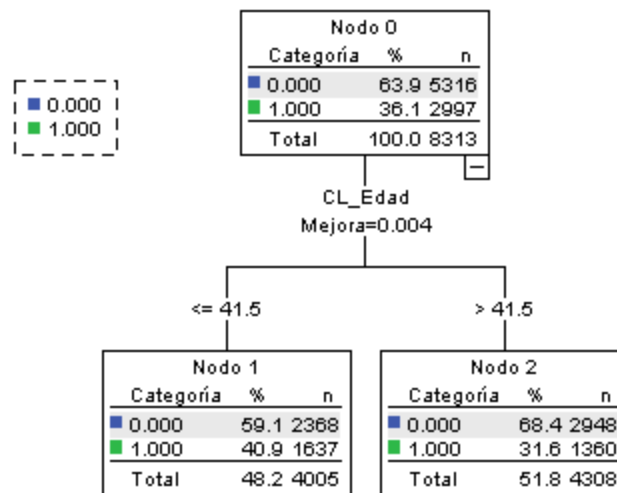


Ilustración 2 - Ejemplo de árbol de decisión

La interpretación del ejemplo anterior es: se tiene una variable dependiente binaria, que toma valores 0 y 1. El número de casos de la variable dependientes es de 8.313, donde 5.316 son valores 0 y 2.997 son valores 1. La variable independiente que se utilizó es "CL\_Edad" que es la edad del cliente, el árbol a partir de ella generó 2 nodos, segmentos o tramos. El punto de corte del tramo es el valor 41,5, es decir, si posee una Edad ≤ 41,5 pertenece al nodo 1, de lo contrario pertenece al nodo 2. La distribución de la variable dependiente dentro del nodo 1 es, 59,1% de valores 0 y 40,9% de valores 1. Mientras que el nodo 2 tiene una distribución de 68,4% de valores 0 y 31,6% de valores 1.



## 2.3 Regresión logística

Los modelos de regresión son herramientas estadísticas utilizadas para describir la relación entre una variable objetivo y un conjunto de variables explicativas. En particular, la regresión logística se utiliza cuando la variable que se desea modelar es dicotómica (Hosmer & Lemeshow, 2000) y toma los valores 1 y 0, como se aprecia a continuación:

$$y_i = \begin{cases} 1 & \text{cuando el fenómeno ocurre} \\ 0 & \text{cuando el fenómeno no ocurre} \end{cases}$$

Donde  $y_i$  es la variable a explicar e "i" representa cada observación

Por lo tanto, la regresión logística es un modelo estadístico de clasificación binaria que entrega la probabilidad de pertenencia a uno de los dos grupos definidos, utilizando para ello un conjunto de variables  $x_i \in \mathbb{R}$ .

La probabilidad de pertenencia a cada uno de los grupos se obtiene mediante la expresión:

$$p(x_i) = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i x_i)}}$$

Donde  $\beta_i$  corresponde al parámetro que se busca estimar mediante la regresión logística

Para la estimación de los parámetros, es decir, la calibración del modelo, se utiliza el método de máxima verosimilitud (Hosmer & Lemeshow, 2000). En este se busca maximizar la probabilidad estimada de obtener los resultados categorizados según  $y_i$ . La función de verosimilitud es la siguiente:

$$\mathcal{L}(\beta) = \prod_i f(x_i, \beta)$$

Donde  $f(x_i, \beta)$  corresponde a la función de densidad de probabilidad de  $x_i$  que en este caso correspondería al modelo de regresión logística. En este caso, la función de verosimilitud puede ser expresada como:

$$\mathcal{L}(\beta) = \prod_i p(x_i)^{y_i} [1 - p(x_i)]^{1-y_i}$$

Usualmente se trabaja con el logaritmo de la función de verosimilitud, pues es más simple de abordar matemáticamente. La expresión es la siguiente:

$$\mathcal{LL}(\beta) = \text{Ln}[\mathcal{L}(\beta)] = \sum_i \{y_i \text{Ln}[p(x_i)] + (1 - y_i) \text{Ln}[1 - p(x_i)]\}$$

Los estimadores de máxima verosimilitud se calculan al aplicar condiciones de primer orden a la función de verosimilitud. Las ecuaciones obtenidas, conocidas como ecuaciones de verosimilitud son:

$$\begin{aligned}\sum [y_i - p(x_i)] &= 0 \\ \sum x_i [y_i - p(x_i)] &= 0\end{aligned}$$

En la práctica, el uso que se le da a este modelo consiste en seleccionar un punto de corte para el valor de la probabilidad, tal que para valores mayores a ese punto de corte se determine que el valor esperado para la variable en estudio sea 1, es decir, el cliente caerá en incumplimiento. En caso que la probabilidad sea menor al corte se asigna valor 0, que significa que el cliente no caerá en incumplimiento. En términos matemáticos:

$$\begin{aligned}p(x_i) \geq \text{punto de corte} &\rightarrow y_i = 1 \\ p(x_i) < \text{punto de corte} &\rightarrow y_i = 0\end{aligned}$$

## 2.4 Medición de desempeño

Como se mencionó anteriormente, luego de construido el modelo es necesario medir su desempeño, para ello Lyn Thomas (2009) presenta diferentes métricas para evaluar el rendimiento y capacidad predictiva de un modelo, las cuales se revisan a continuación:

### 2.4.1 Estadístico Kolmogorov-Smirnov

El test Kolmogorov-Smirnov (KS), es un test de hipótesis que se utiliza para contrastar la distribución de probabilidad de dos muestras. Este test puede ser modificado para que pueda medir el desempeño de un modelo. Lo anterior se realiza definiendo el siguiente estadístico:

$$KS \triangleq \max_p (F(p|\text{malos clientes}) - F(p|\text{buenos clientes}))$$

Donde  $p$  es la probabilidad.

$F(p|\text{buenos clientes}) = Pr(\text{probabilidad} < p | \text{buenos clientes})$  es la función de distribución acumulada condicional de los buenos clientes, es decir, no caen en incumplimiento).

$F(s|B) = Pr(\text{puntuación} < s | B)$  es la función de distribución acumulada condicional de los malos clientes, es decir, caen en incumplimientos.

Mientras mayor es el  $KS$  mejor discriminación, con un valor máximo de 1. Empíricamente, un modelo con estadístico KS superior al 0.4 es considerado como bueno.

En la Ilustración 2 se puede observar un ejemplo del estadístico KS.

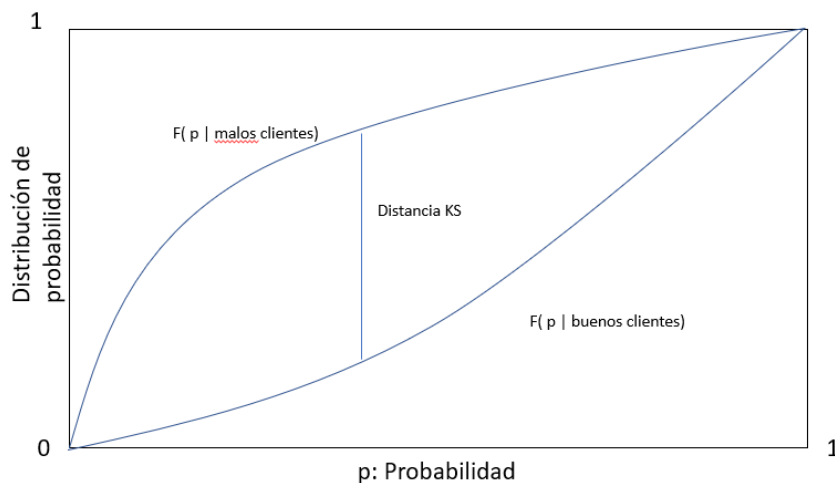


Ilustración 3 - Ejemplo de estadístico KS

## 2.4.2 Curva ROC

Es la curva de ROC (*Receiver Operating Characteristic*) construida a partir de la distribución de errores de clasificación de un modelo. Los errores son clasificados en dos tipos:

- Error tipo I: clientes buenos son clasificados como malos clientes.
- Error tipo II: clientes malos son clasificados como buenos clientes.

	Estimación: cliente bueno	Estimación: cliente malo
Cliente bueno	Cliente bien clasificado	Cliente mal clasificado
Cliente malo	Cliente mal clasificado	Cliente bien clasificado

Tabla 1 - Matriz de confusión

La curva se construye al graficar  $F(p|\text{buenos clientes})$  en el eje "x" y  $F(p|\text{malos clientes})$  en el eje "y", como en la Ilustración 3. Donde  $F(p|\text{buenos clientes}) = Pr(\text{probabilidad} < p | \text{buenos clientes})$  es la función de distribución acumulada condicional de los buenos clientes; y  $F(p|\text{malos clientes}) = Pr(\text{probabilidad} < p | \text{malos clientes})$  es la función de distribución acumulada condicional de los malos clientes.

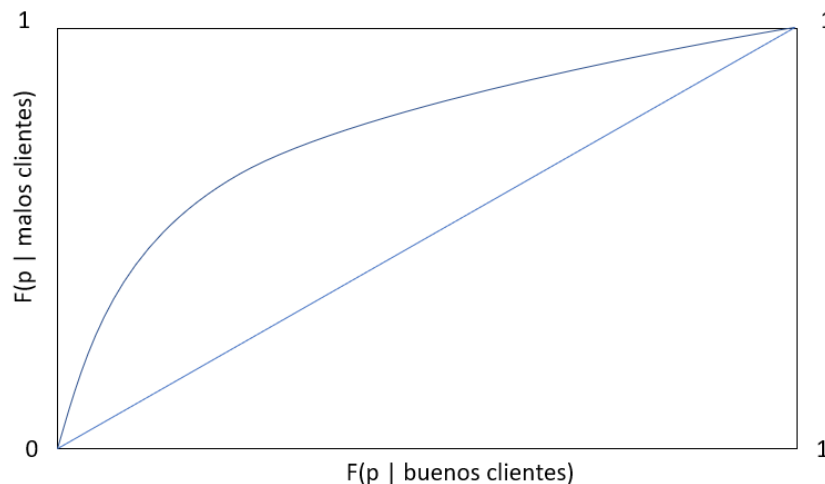


Ilustración 4 - Ejemplo de curva ROC.

Finalmente, el estadístico ROC es el área bajo la curva ROC. Mientras mayor sea dicho valor, mejor es la discriminación del modelo. Discriminación perfecta se alcanza cuando el estadístico es 1. Empíricamente, un modelo con estadístico superior al 0.75 es considerado como bueno.

## **Capítulo III. Metodología**

Para el desarrollo del modelo de probabilidad de incumplimiento se utiliza la metodología propuesta en *Credit Risk Scorecards* (Siddiqi, 2006).

La metodología que se propone se basa en siete etapas que se describen a continuación.

### **3.1 Comprensión del negocio**

En esta etapa, se busca conocer las razones que fundamentan la realización de esta memoria y cuáles son los resultados esperados. En este sentido, se busca alinear a las partes involucradas y definir en conjunto los objetivos que tiene.

Por otro lado, se busca comprender la situación actual de la institución respecto al modelo de probabilidad de incumplimiento que posee, para lo cual se identifican:

- Variables que construyen los segmentos que posee actualmente la cartera en cuestión.
- Variables con las que modelan el comportamiento de sus clientes de cada uno de los segmentos actuales del banco.
- Capacidad predictiva del modelo.

### **3.2 Comprensión de los datos**

Esta etapa consiste en estudiar la información que posee el banco de sus clientes. Esta etapa es fundamental para las etapas posteriores, pues si los datos no son confiables los resultados obtenidos tampoco lo serán.

Es importante conocer y entender el significado de cada una de las variables de información que posee el banco. Además, es necesario realizar un análisis de la calidad de las bases de datos proporcionadas por la institución. Es muy probable encontrar errores e inconsistencias en los datos. Por lo anterior, se debe realizar una inspección de la información proporcionada.

- Valores perdidos.
- Valores fuera de rango.
- Error de tipeo.
- Incoherencia entre nombre de variable y valor asignado a dicho campo.

### 3.3 Elaboración de base de datos

Para comenzar, se debe solucionar los problemas encontrados en la etapa anterior respecto a la calidad de los datos.

En segundo lugar, se debe construir la variable dependiente o de desempeño. Esta es binaria y que toma valor "1" si el cliente en cuestión cae en incumplimiento<sup>1</sup> en la ventana de tiempo de 12 meses. Toma valor "0" si el individuo no cae en incumplimiento. Lo anterior se representa en la Ilustración 5.

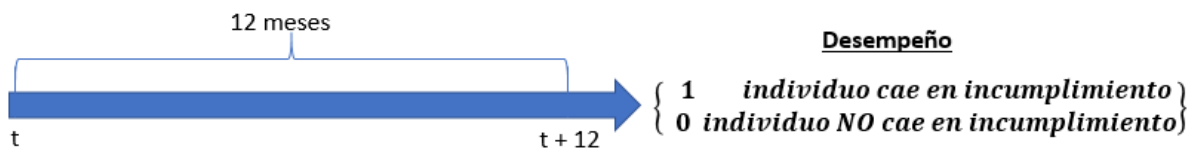


Ilustración 5 - Construcción de variable de desempeño.

Posterior a ello, se crean las nuevas variables en base a la información que el banco posee de los clientes y que pueden ser interesantes para explicar su comportamiento de pago en el tiempo.

### 3.4 Segmentación

El banco posee una gran cantidad de clientes, los cuales poseen características y comportamientos diferentes unos de otros. Dado lo anterior, en esta etapa se busca agrupar clientes en segmentos con características similares. Los segmentos que se originen deben ser homogéneos entre sí y heterogéneos entre segmentos, en cuanto al nivel de riesgo (*Bad Rate* o tasa de incumplimiento). Esto se realiza en base a características observables y que hagan sentido al negocio.

Para identificar las variables que mejor segmentan, se realiza un análisis de discriminancia (o análisis univariado) a cada una de las variables seleccionadas mediante árboles de decisión.

A continuación, se detallan los pasos del análisis univariado.

---

<sup>1</sup> Mora mayor o igual a 90 días

- i. Seleccionar una variable.
- ii. Realizar el árbol de decisión, colocando como variable a dependiente la variable de desempeño y como variable independiente la variable escogida en i.

Lo anterior arroja como resultado una tramificación de la variable independiente. A cada tramo se le asocia la cantidad de clientes que no caen en incumplimiento (buenos) y los que si caen en incumplimiento (malos).

- iii. Con los resultados del árbol de decisión, se calculan los siguientes estadísticos:

a.  $BadRate_i = \frac{Cantidad\ de\ clientes\ malos\ del\ tramo\ i}{Total\ de\ clientes\ en\ rango\ i}$ , es el BadRate de cada tramo que se generó a partir del árbol.

b.  $\%buenos_i = 100 * \frac{cantidad\ de\ clientes\ buenos\ del\ rango\ i}{total\ de\ clientes\ buenos}$ , es el porcentaje de clientes buenos de cada tramo respecto del total de clientes buenos.

c.  $\%malos_i = 100 * \frac{cantidad\ de\ clientes\ malos\ del\ rango\ i}{total\ de\ clientes\ malos}$ , es el porcentaje de clientes malos de cada tramo respecto del total de clientes malos.

d.  $WOE_i = Ln(\frac{\%buenos_i}{\%malos_i})$ , es el valor WOE (Weight of Evidence) asociado a cada tramo de la variable.

e.  $Information\ Value\ (IV) = 100 * \sum_i [(\%buenos_i - \%malos_i) * WOE_i]$ , es el valor de Information Value asociado a la variable y representa que tan buena es la variable para discriminar el riesgo de los diferentes tramos generados.

Un ejemplo de lo descrito se aprecia en la tabla 2.

Tramo	Buenos	Malos	BadRate	WoE	IV
(-inf,31]	10551	1005	8,70%	-0,49	3,69
(31,49]	41546	2809	6,33%	-0,14	1,09
(49,+inf)	36653	1364	3,59%	0,44	6,72
TOTAL	88750	5178	5,51%		11,51

Tabla 2 – Ejemplo de variable tramificada

- iv. Iterar para cada una de las variables seleccionadas como posible variable para segmentar a los clientes.
- v. Realizar ranking de variables a partir del *Information Value*, ordenando de mayor a menor las variables según su IV. Donde la

variable con mayor IV es la que tiene un mayor poder de discriminancia.

- vi. Del ranking de *Information Value* se eliminan las variables que tengan valores menores a 2, pues se considera que no poseen poder de discriminancia entre buenos y malos clientes.

### **3.5 Modelamiento**

En esta etapa se busca identificar aquellas variables que mejor predicen el comportamiento de los clientes de cada segmento. Para ello se realizan cuatro subetapas: análisis univariado, análisis de correlación, análisis multivariado y análisis de contribución a la logverosimilitud. A continuación, se detalla cada subetapa.

#### **3.5.1 Análisis univariado**

En esta primera subetapa se busca identificar el poder discriminante de cada variable, que potencialmente podría ingresar al modelo. Para esto se realizan los pasos mencionados en capítulo 3.4.

Luego de tener confeccionado el ranking de *Information Value*, se eliminan todas aquellas variables cuyo IV sea menor o igual a 2, pues se considera que no poseen un poder de discriminancia relevante.

Además, desde esta etapa en adelante, se trabajará con las variables tramificadas. A cada tramo que se generó, para cada variable, se le asigna el valor WOE correspondiente a dicho tramos.

#### **3.5.2 Análisis de correlación**

Esta subetapa busca eliminar variables que sean redundantes, es decir, variables que estén explicando el mismo efecto. Para esto, se calcula la matriz de correlaciones de todas las variables tramificadas.

La forma de eliminación de variables con un alto nivel de correlación, en este caso se utilizó un valor de 0,5, se realiza de la siguiente forma:

- i. Seleccionar la variable con mayor IV, pues es la que mejor discrimina.
- ii. Eliminar todas las variables que posean una correlación, con la variable seleccionada en i., mayor a 0,5 pues al estar altamente correlacionadas se está explicando el mismo efecto, por lo tanto, se deja la que tiene un mayor poder explicativo.



- iii. Seleccionar la siguiente variable con mayor IV que no fue eliminada e iterar, de esta forma quedaran solo variables con poca correlación entre si.

### 3.5.3 Análisis multivariado

En esta subetapa se busca determinar el aporte de cada variable a la realización de un evento, en este caso, que tanto aporta a modelar el comportamiento de incumplimiento de los clientes. Dado que la variable a explicar es binaria, se utiliza la regresión logística.

La técnica escogida para realizar la regresión es *Forward Wald*. Mediante este método, se ingresa variable por variable al modelo, haciendo que interactúen con la variable dependiente. Una variable ingresa al modelo si su estadístico de puntuación es menor que un valor de entrada determinado, en este caso 0,05. Cuando ninguna variable cumple la restricción se detiene el proceso.

En cada paso de la iteración, se obtiene un modelo con las variables ingresadas hasta ese momento. Además, se puede obtener la logverosimilitud de cada etapa, que será fundamental para la siguiente subetapa, en la cual se reduce la cantidad de variables a incluir en el modelo.

### 3.5.4 Análisis de contribución a logverosimilitud

Como resultado de la subetapa anterior, se puede obtener un modelo con un número de variables que sería difícil de interpretar sus resultados. Por esta razón, en esta subetapa se busca determinar un número de variables razonable, de entre 10 a 20 variables por modelo.

Para determinar el número de variables a incluir, se mide el porcentaje de contribución a la logverosimilitud en cada etapa del proceso *Forward Wald*. El valor, para cada iteración, se obtiene mediante la siguiente expresión:

$$\% \text{ Contribución de iteración } i = 1 - \frac{\text{Logverosimilitud iteración } i}{\text{Logverosimilitud iteración } (i - 1)} \cdot 100$$

La primera iteración,  $i=1$ , es el modelo con solo una variable.

Como la contribución tiene una tendencia decreciente, llega un momento en que el aporte de ingresar una nueva variable al modelo no es relevante. El criterio que se utiliza para definir que el aporte de ingreso de una variable no significativo es: porcentaje de contribución a la logverosimilitud menor a 0,5%.

Finalmente, el número de variables que se incluyen al modelo es la cantidad indicada en la última iteración donde el porcentaje de contribución es significativo.

A continuación, en el gráfico 2 se aprecia un ejemplo de lo descrito anteriormente. En este caso, las variables a incluir en el modelo es 14, pues al ingresar la variable número 15, su contribución no es significativa.

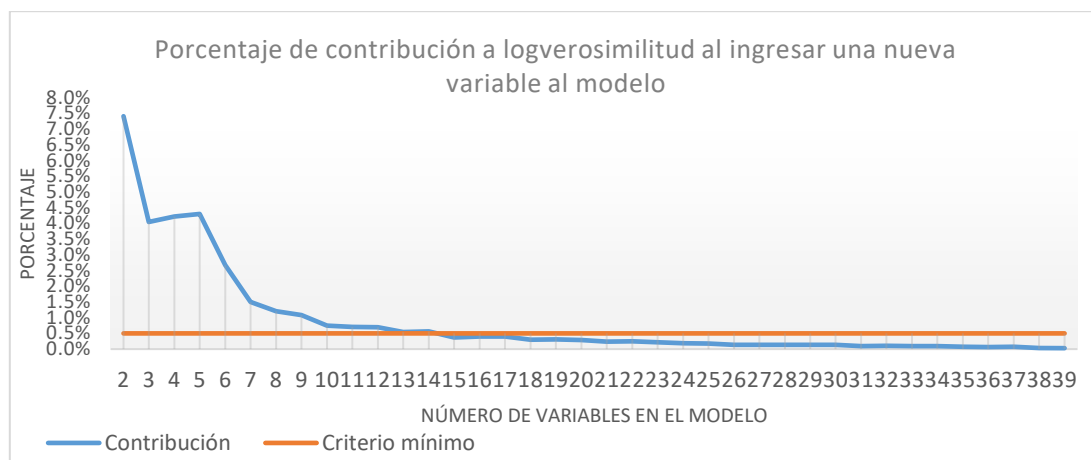


Gráfico 2 – Ejemplo de análisis de contribución a la logverosimilitud

### 3.6 Medición de desempeño y elección de modelo

En esta etapa final, se elaboran los estadísticos KS y ROC, para cada modelo que se desarrolla en la memoria. De esta forma se mide el desempeño que posee cada modelo. Luego, se comparan entre sí. Donde, a mayor valor del estadístico KS y ROC, mejor es el modelo.

Además de realizar una comparación entre estadísticos, se realiza una estimación del impacto en las provisiones que posee cada modelo propuesto respecto a la línea base que se construye en el capítulo V.

Otra forma en que se evalúan los modelos que se elaboran en el capítulo VI, es mediante *backtesting*, donde se compararán las tasas de incumplimiento predichas por los modelos versus la tasa de incumplimiento real.

Finalmente, el criterio de elección de modelo es una combinación entre el desempeño del modelo y el impacto que tiene a las provisiones. Pues se podría llegar a un modelo con altos estadísticos, pero que los clientes en los que equivoca son los más caros para el banco. Lo anterior quiere decir que, estima una alta probabilidad de incumplimiento a aquellos que no lo harán, y además, el monto de su deuda es elevada.

## Capítulo IV. Preparación de datos

### 4.1 Periodo de información

El conjunto de datos proporcionados por el banco posee información de los clientes desde el 31-01-2008 al 31-01-2016. Dentro de este periodo, se encuentran todos los clientes de la cartera de consumo del banco, pero cada uno de estos aparece una sola vez. Es decir, no se encontrará un cliente en el periodo 31-01-2010 y en el periodo 31-08-2013.

### 4.2 Construcción de variable dependiente

Para construcción de la variable de desempeño, es necesario que cada cliente se encuentre solo una vez. De no ser así, se estaría midiendo y modelando el comportamiento del mismo cliente 2 o más veces. Lo anterior podría producir multicolinealidad y no se obtendrían buenos resultados en el modelamiento.

Como ya se posee a cada cliente una única vez en toda la ventana de información, lo que se realizó fue medir el desempeño de dicho cliente. El desempeño es medido en los próximos 12 desde el periodo que se encuentra el cliente. Esto se realiza de acuerdo con lo expuesto en el capítulo "3.3 Elaboración de base de datos".

La base de datos cuenta con 469.798 clientes. Estos clientes se ubicaron en la ventana de tiempo, con ello se construyó el gráfico 3 que muestra la cantidad de clientes por mes.

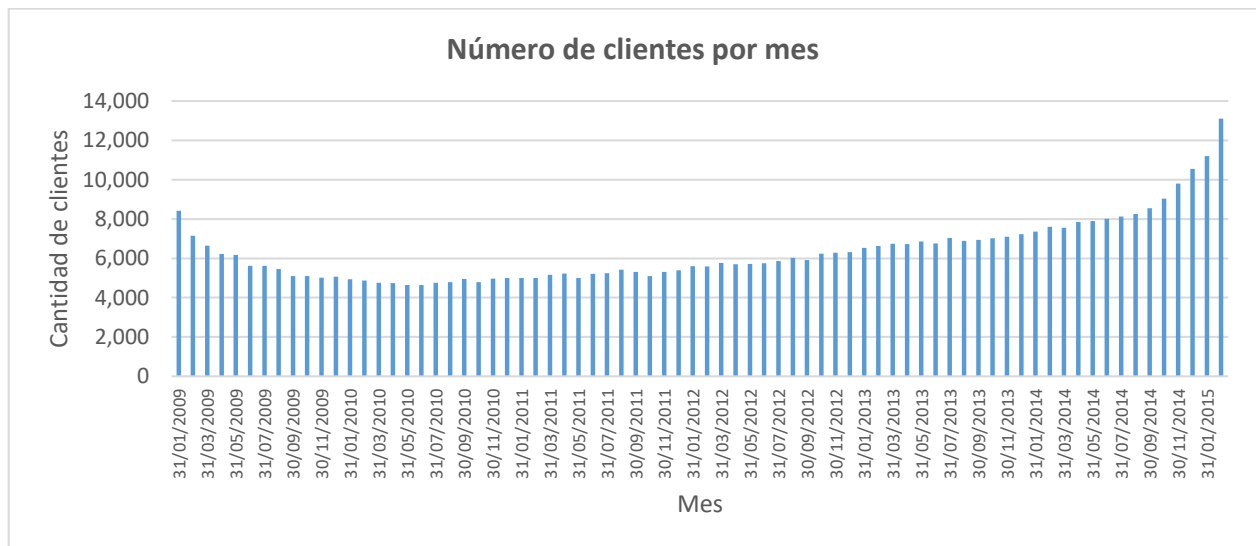


Gráfico 3 - Distribución de clientes por mes.

Los resultados de la construcción de la variable dependiente, donde "0" representa a un buen cliente, es decir, no cae en incumplimiento. Mientras que "1" representa a un mal cliente, es decir, cae en incumplimiento, lo anterior se aprecia en el gráfico 4.

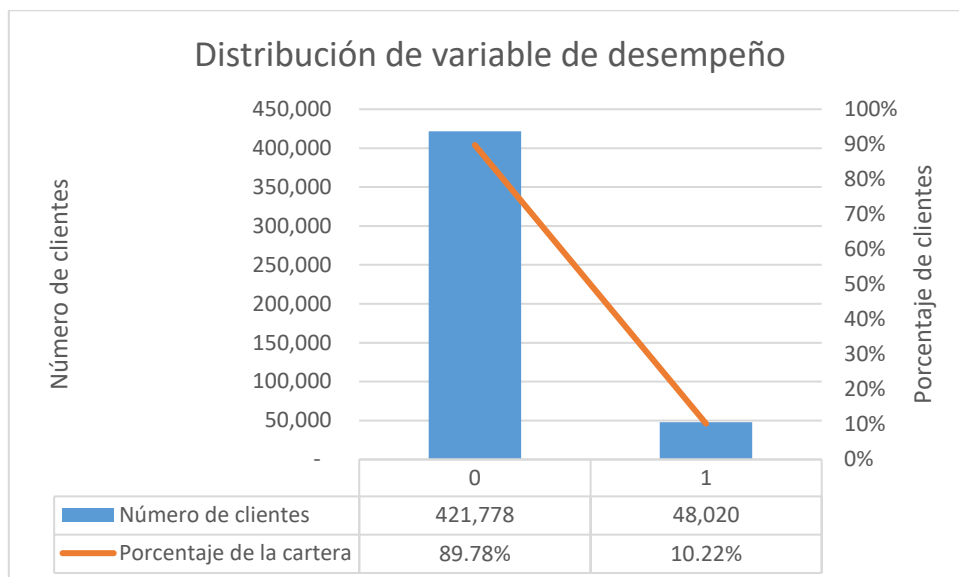


Gráfico 4 - Distribución de clientes que caen en incumplimiento y los que no.

### 4.3 Información disponible

Además de contar con la información de desempeño de los clientes, es necesario poseer información para tratar de explicar y modelar el comportamiento de incumplimiento. Dado lo anterior, el banco facilitó las siguientes bases de datos: (Mas detalle de la información que posee cada base se encuentra en Anexo 1)

- Cientes: donde se encuentran variables demográficas de los clientes, antigüedad de los clientes en los productos, montos de deuda, entre otras.
- Consumo: esta base posee información respecto a los créditos de consumo en cuota de los clientes.
- Cuenta corriente: aquí se encuentra información respecto a los ingresos y egresos de las cuentas corriente de los clientes.
- Tarjeta de crédito: en esta base se encuentra la información relacionada a las características y uso de la tarjeta de crédito que poseen los clientes.
- Línea de crédito: en esta base se encuentra la información relacionada a las características y uso de la línea de crédito que poseen los clientes.

- Facturación tarjeta de crédito: aquí se encuentra los movimientos facturados de la tarjeta de crédito de los clientes, tales como, los avances en efectivo que solicito u otros.
- Inversiones: en esta base se ubica la información relacionada la actividad del cliente en inversiones, como depósitos a plazo, acciones u otros.

#### **4.4 Tratamiento de datos**

Parte fundamental del modelamiento es poseer información confiable y de calidad, por lo que en primera instancia se estudia que los datos estén en igual formato, que las variables numéricas efectivamente sean consideradas de esa forma.

Luego de corregir errores de formatos (cambiar comas por puntos decimales en algunas variables), se analiza que los registro en las variables estén dentro de los rangos correspondientes, esto se realiza contrastando los datos con el diccionario que posee el banco.

Posteriormente se ejecuta el cruce de bases de datos, donde a cada cliente se le asigna su valor correspondiente para cada una de las variables. En el cruce de bases de datos ocurre que algunos clientes no poseen información asociada a alguna variable, en estos casos se le asigna el valor "-999" lo que representa dicha situación.

## **Capítulo V. Modelo actual del banco BCI**

El objetivo de este capítulo es realizar la reconstrucción del modelo que posee actualmente el Banco. El fin de lo anterior es, poder comparar los resultados que se obtuvo con el modelo actual del banco replicado por el estudiante respecto a los resultados que el banco facilitó al medir el desempeño de su modelo actual.

Por otra parte, los resultados que se obtienen en este capítulo serán la línea base con la que se contrasta los resultados de los modelos que se proponen en el capítulo VI.

### **5.1 Segmentación de réplica del modelo actual del banco**

En esta primera etapa, se buscó generar los segmentos que actualmente posee el banco. Para ello, se utilizó las variables que fueron entregadas por la contraparte. El motivo de utilizar directamente las variables, y no realizar el análisis univariado para llegar a éstas, fue que ellos complementaron su análisis estadístico con el criterio de expertos, de esta forma seleccionaron las variables.

Las variables que determinan la segmentación actual del banco son las siguientes:

- Renegociado: indica si el cliente renegocio o no su deuda de consumo.
- Crédito de consumo en cuotas: indica si el cliente posee un crédito de consumo en cuotas vigente.
- Cliente antiguo: indica si la persona es cliente de la cartera de consumo del banco BCI hace más de 36 meses.
- Sueldo abonado: indica si al cliente le es abonado el sueldo en su cuenta corriente del banco BCI.

El orden utilizado para generar la segmentación de la cartera de consumo fue:

1. Se diferenció entre clientes que renegociaron su deuda de consumo con los que no la han renegociado.
2. Los clientes que no han renegociado su deuda se separaron en clientes que poseen crédito de consumo en cuotas vigente de los que no lo poseen.
3. Para ambos segmentos anteriores, con y sin crédito de consumo en cuotas vigentes, se dividieron en clientes con antigüedad en la cartera de consumo menor o igual a 36 meses y mayor a 36 meses.
4. Finalmente, a los clientes con antigüedad menor o igual a 36 meses, se separaron en clientes a los que su sueldo es abonado en el banco y a los que no.

Lo anterior, dio origen a un total de 7 segmentos para la cartera de consumo del banco BCI. La segmentación final se aprecia en la Ilustración 6.

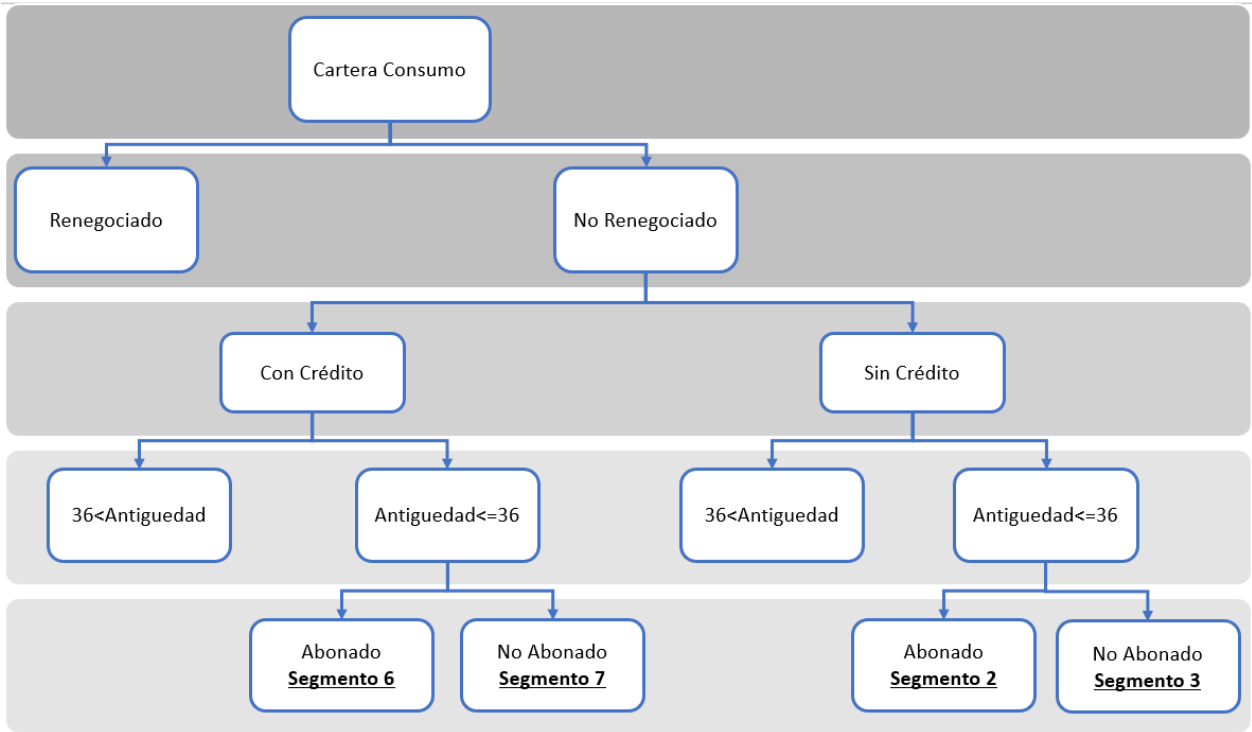


Ilustración 6 - Segmento luego de cuarta segmentación. Fuente: banco BCI.

A continuación, en el Gráfico 5 se aprecia el tamaño que representa cada segmento del total de clientes de la cartera de consumo y porcentaje de exposición asociada a cada segmento. Se identifica que, en cuanto a la cantidad de clientes, el segmento más grande es el Segmento 1, con un 28.5% del total de la cartera. Sin embargo, el segmento que mayor exposición posee es el Segmento 5, con un 48.3% del total.

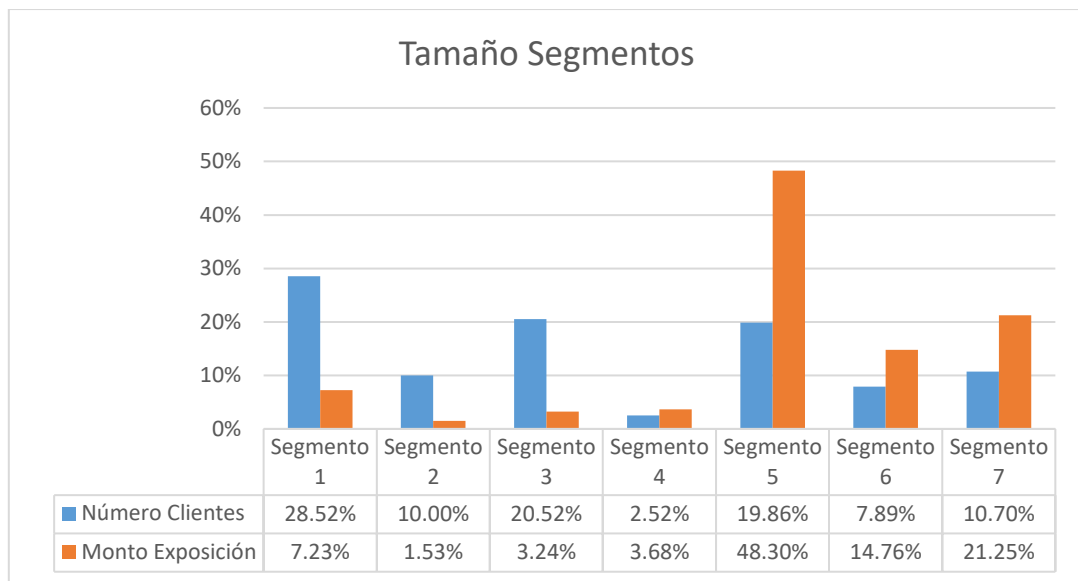


Gráfico 5 - Cantidad de clientes por segmento y porcentaje que representan de la cartera de consumo. Elaboración propia a partir de BD.

Los Bad Rates que poseen cada uno de los segmentos finales se aprecian en el gráfico 6.

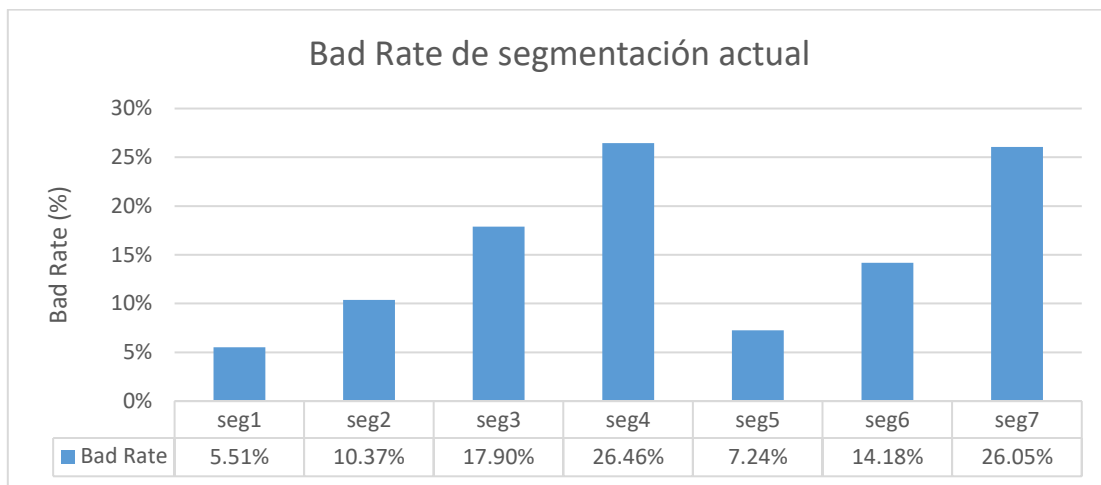


Gráfico 6 - Bad rate asociado a cada segmento. Modelo actual del banco.

Se observa que los segmentos generados poseen distintos grados de riesgo, tomando como medida de riesgo el *Bad Rate*. Tener un mayor *Bad Rate* significa que dentro del segmento un mayor porcentaje de clientes, de dicho segmento, cayó en incumplimiento. Dado lo anterior los segmentos más riesgosos para el banco son el 4 y 7. Mientras que los segmentos 1 y 5 son los menos riesgosos, pues son los que poseen los menores *Bad Rate*.



## 5.2 Modelamiento de réplica modelo actual del banco

Determinados los segmentos, la siguiente etapa busca identificar las variables que mejor predicen el comportamiento de incumplimiento de los clientes de cada uno de los siete segmentos. Para encontrar dichas variables, se utiliza la metodología descrita en "Capítulo 3.5".

A continuación, en la tabla 3 se observa la cantidad de variables que posee cada modelo, de cada segmento. (El detalle de los pasos para llegar al número de variables a incluir en cada modelo se encuentra en Anexo 2).

Segmento	Número de variables a incluir a modelo de cada segmento
1	14
2	14
3	16
4	15
5	14
6	16
7	15

Tabla 3 – Número de variables a incluir en el modelo de cada segmento del modelo replicado.

## 5.3 Desempeño replica modelo actual del banco

En esta etapa se midió el desempeño del modelo de cada segmento y del modelo en forma agregada. Las métricas que se utilizaron son el estadístico KS y la curva de ROC. Los resultados obtenidos para cada métrica se encuentran en el gráfico 7.

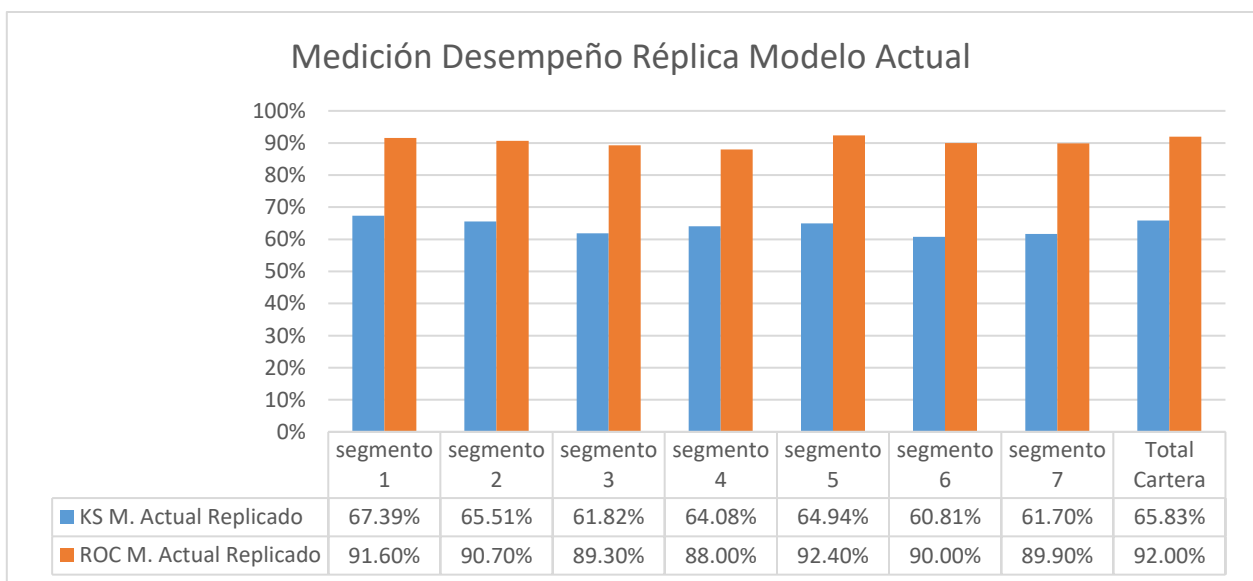


Gráfico 7 – Desempeño medido en Ks y ROC del modelo actual replicado.

Es posible apreciar en el gráfico anterior que, el modelo posee una gran capacidad predictiva, pues la literatura señala que con KS mayor a 40% es un buen modelo, mientras que para la curva de ROC considera un buen modelo al tener un valor mayor al 75%. Los resultados que se observan para el KS al evaluar toda a cartera, del modelo actual replicado es de 65.83% y para la curva de ROC es de 92%, valores considerablemente alto.

#### 5.4 Comparación resultados del modelo replicado

Luego de finalizar la réplica del modelo actual del banco, se procedió a comparar los resultados obtenidos con la información proporcionada por el banco. Dicha información es: variables que componen el modelo de cada segmento, su respectivo coeficiente y el valor del estadístico KS de cada segmento y de la cartera completa.

En la tabla 4, se observa una comparación de la cantidad de variables, del modelo de cada segmento, de la réplica y del modelo del banco. Se observa que solo en el segmento 5 la cantidad de variables es la misma.

Segmento	Cantidad de variables modelo replicado	Cantidad de variables modelo banco
<b>1</b>	14	13
<b>2</b>	14	16
<b>3</b>	16	14
<b>4</b>	15	14
<b>5</b>	14	14
<b>6</b>	16	14
<b>7</b>	15	16

Tabla 4 - Comparación número de variables por segmento de modelo del banco y modelo replicado

Sin embargo, en ninguno de los segmentos las variables que se incluyen son todas iguales. La diferencia en la cantidad de variables puede deberse a dos motivos: el criterio utilizado en el análisis de contribución a la logverosimilitud fue diferente y/o el banco utilizó un 80% de los datos para el modelamiento, mientras que en el modelo replicado se utilizó un 70% de los datos.

Respecto a las variables que componen el modelo actual del banco respecto a las que componen el modelo actual replicado, se señalan a continuación.

## 1. Segmento 1

En la tabla 5 se observan las variables incluidas en el segmento 1 del modelo actual replicado y variables del banco. De las 13 variables del modelo replicado, 8 de éstas se encuentran en el modelo del banco, son las variables que, asociadas a la actividad de los clientes en sus cuentas corrientes, avances solicitados y la peor situación del cliente en su deuda directa. Mientras que las diferencias se encuentran en que el modelo replicado no incluye variables del comportamiento en la línea de crédito y sobregiros no pactados.

Modelo Actual Replicado			Modelo Actual Banco	
N°	Variable	Beta	Variable	Beta
1	Antigüedad_CCT	-0.57	Antigüedad_CCT	-0.71
2	Evolucion_Monto_Disponible_CCT_T0_T12	-0.36	Antigüedad_Ultima_Actividad_Cliente	-0.38
3	Evolucion_Monto_Disponible_LDC_mes_T0_T12	-0.57	Deuda_Directa_Productos_Consumo_sist	-0.40
4	Evolucion_Monto_Egreso_Cheque_T0_T3	-0.53	Evolucion_Monto_Disponible_CCT_T0_T12	-0.41
5	Evolucion_Monto_Promedio_Uso_LDC_mes_T6_T12	-0.40	Evolucion_Monto_Pago_CCT_Lsb_T3_T6	-0.30
6	Evolucion_Monto_Protestos_T0_T12	-0.41	Evolucion_Monto_Transferido_Lem_CCT_T0_T12	-0.45
7	Evolucion_Monto_Transferido_Lem_CCT_T0_T12	-0.54	Máximo_Días_atraso_U6M	-0.61
8	Monto_Total_Abonado_en_CCT_U3M	-0.73	Monto_Total_Abona_en_CCT_mes	-0.35
9	Numero_Avances_Solicitados_U12M	-0.34	Numero_Avances_Solicitados_U12M	-0.22
10	Peor_Situacion_Deuda_Directa_U12	-0.44	Peor_Situacion_Deuda_Directa_U12M_sist	-0.31
11	Peor_Situacion_Deuda_Directa_U12M_sist	-0.48	Ratio_Monto_Egreso_Girado_Atm_T0_T12	-0.38
12	Ratio_Monto_Egreso_Girado_Atm_T0_T12	-0.41	Saldo_Minimo_CCT_Activas_Mes	-0.24
13	Saldo_Maximo_SGNP	-0.25	Total_de_Saldos_Disponibles	-0.35
14	Total_Saldos_Contractuales	-0.99	Constante	-2.84
15	Constante	-2.87		

Tabla 5 – Comparación de variables de modelo actual replicado con variables del modelo actual del banco. Segmento 1.

## 2. Segmento 2

En la tabla 6 se observan las variables incluidas en el segmento 2 del modelo actual replicado y variables del banco. Al comparar las variables incluidas en los modelos, 6 de éstas se encuentran en ambos. Las diferencias entre modelos se encuentran en que, el modelo actual del banco incluye más variables del comportamiento de los clientes en sus tarjetas de crédito y líneas de créditos.

Modelo Actual Replicado			Modelo Actual Banco	
N°	Variable	Beta	Variable	Beta
1	Cupo_Utilizado_TDC	-0.19	Antigüedad_Ultimo_Uso_CCT	-1.03
2	Deuda_Directa_Productos_Consumo_sist	-0.42	Cupo_Utilizado_TDC	-0.22
3	Evolucion_Monto_Abonado_Lem_CCT_T0_T6	-0.30	Deuda_Directa_Productos_Consumo_sist	-0.38
4	Evolucion_Monto_Disponible_LDC_mes_T0_T12	-0.34	Evolucion_Cantidad_Dias_Usada_LDC_T0_T12	-0.32
5	Evolucion_Monto_Girado_Atm_T3_T12	-0.46	Evolucion_Monto_Disponible_CCT_T0_T12	-0.44
6	Evolucion_Monto_Promedio_Uso_LDC_T3_T12	-0.49	Evolucion_Monto_Disponible_LDC_mes_T0_T12	-0.41
7	Máximo_Días_Atraso_U3M	-0.37	Evolucion_Monto_Egreso_CCT_T0_T3	-1.06
8	Monto_Total_Abona_en_CCT_mes	-0.82	Evolucion_Monto_Girado_Caja_T0_T12	-0.78
9	Nivel_Endudamiento_vigente	-0.39	Máximo_Días_Atraso_U3M	-0.51

10	Numero_Avances_Solicitados_U12M	-0.41	Monto_Utilizado_LDC_SBIF	-0.18
11	Ratio_Monto_Abonado_Transferencia_CCT_T6_T12	-1.05	Nivel_Educacional	-0.56
12	Ratio_Monto_Pago_CCT_Lsb_T6_T12	-1.32	Numero_Avance_Solicitados_U6M	-0.32
13	Saldo_Minimo_CCT_Activas_Mes	-0.27	Pago_Minimo_TDC	-0.38
14	Total_de_Saldos_Disponibles	-0.39	Peor_Situacion_Deuda_Sistema_U6M_sist	-0.26
15	Constante	-2.17	Ratio_Monto_Girado_Atm_T6_T12	-2.41
16			Total_de_Saldos_Disponibles	-0.32
17			Constante	-2.18

Tabla 6 - Comparación de variables de modelo actual replicado con variables del modelo actual del banco. Segmento 2.

### 3. Segmento 3

En la tabla 7 se observan las variables incluidas en el segmento 3 del modelo actual replicado y variables del banco. Al comparar las variables que componen ambos modelos, se aprecia que 6 de éstas están incluidas en ambos. Las diferencias se encuentran en que, el modelo replicado incluye más variables de actividad en las cuentas corrientes y no se incluyen variables de información del cliente (nivel educacional y tipo de vivienda) que si son incluidas en el modelo actual del banco.

Modelo Actual Replicado			Modelo Actual Banco	
N°	Variable	Beta	Variable	Beta
1	Antiguedad_Cliente	-1.63	Antiguedad_Cliente	-1.60
2	Cupo_Utilizado_TDC	-0.52	Cupo_Utilizado_TDC	-0.40
3	Evolucion_Monto_Abonado_T6_T12	-1.22	Deuda_Directa_Productos_Consumo_sist	-0.32
4	Evolucion_Monto_Disponible_LDC_mes_T0_T12	-0.42	Evolucion_Monto_Abonado_T6_T12	-1.60
5	Máximo_Días_atraso_U6M	-0.48	Evolucion_Monto_Disponible_CCT_T0_T12	-0.80
6	Monto_Total_Abonado_en_CCT_U3M	-0.46	Máximo_Días_atraso_U6M	-0.72
7	Monto_Total_Mora_Blanda	-0.27	Monto_Maximo_Adeudado_U3M_UF	-0.29
8	Monto_Total_utilizado_LDC	-0.37	Monto_Utilizado_LDC_SBIF	-0.46
9	Numero_Avances_Solicitados_U12M	-0.33	Nivel_Educacional	-0.51
10	Peor_Situacion_Deuda_Directa_U12M_sist	-0.45	Numero_Avances_Solicitados_U12M	-0.25
11	Ratio_Cupo_Disponible_LDC_T3_T6	1.54	Pago_Minimo_TDC	-0.92
12	Ratio_Monto_Abonado_Transferencia_CCT_T0_T12	-0.88	Peor_Situacion_Deuda_Directa_U12M_sist	-0.32
13	Ratio_Monto_Promedio_Saldos_Positivos_CCT_T6_T12	-0.82	Tipo_Vivienda	-0.66
14	Ratio_Monto_Transferido_Lem_CCT_T0_T6	-0.31	Total_de_Saldos_Disponibles	-0.39
15	Saldo_Minimo_CCT_Activas_Mes	-0.24	Constante	-1.51
16	Total_Cupo_LDC	-0.62		
17	Constante	-1.50		

Tabla 7 - Comparación de variables de modelo actual replicado con variables del modelo actual del banco. Segmento 3.

### 4. Segmento 4

En la tabla 8 se observan las variables incluidas en el segmento 4 del modelo actual replicado y variables del banco. Al comparar las variables incluidos en

estos modelos, 9 de éstas se encuentran en ambos. Mientras que el resto de las variables apuntan a capturar efectos similares (Antigüedad, Ratio/Evolución de montos en mora de crédito de consumo, protestos).

Modelo Actual Replicado			Modelo Actual Banco	
N°	Variable	Beta	Variable	Beta
1	Antigüedad_Ultima_TDC_Abierta	-0.60	Antigüedad_CCT	-0.35
2	Cantidad_TDC_vigentes	-0.49	Cantidad_TDC_vigentes	-0.58
3	Edad	-0.67	Deuda_Directa_Productos_Consumo_sist	-0.73
4	Evolucion_Monto_Disponible_CCT_T0_T12	-0.27	Edad	-0.71
5	Evolucion_Monto_IFRS_LDC_T0_T3	-0.40	Evolucion_Monto_Promedio_Saldos_Positivos_T0_T12	-0.37
6	Evolucion_Monto_Mora_Credito_Consumo_T0_T12	-0.63	Maxima_Mora_TDC_U6M	-0.84
7	Maxima_Mora_TDC_U6M	-0.49	Monto_Promedio_Adeudado_U6M_U3M	-0.58
8	Monto_Promedio_Adeudado_U6M_U3M	-0.66	Monto_Total_Abona_en_CCT_mes	-0.43
9	Monto_Total_Abona_en_CCT_mes	-0.56	Numero_días_SGNP_U3M	-0.53
10	Nivel_Endeudamiento_vigente	-0.61	Numero_Protesto_Aclarado_U12M_sist	-0.58
11	Numero_días_SGNP_U3M	-0.36	Peor_Situacion_Deuda_Directa_mes	-0.33
12	Peor_Situacion_Deuda_Directa_mes	-0.37	Peor_Situacion_Deuda_Directa_Sistema_mes	-0.32
13	Peor_Situacion_Deuda_Directa_Sistema_mes	-0.36	Peor_Situacion_Deuda_Directa_U12M_sist	-0.24
14	Peor_Situacion_Deuda_Directa_U12M_sist	-0.26	Ratio_Monto_Mora_Credito_Consumo_T0_T12	-0.67
15	Ratio_Monto_Protesto_T3_T12	-0.92	Constante	-0.95
16	Constante	-0.97		

Tabla 8 - Comparación de variables de modelo actual replicado con variables del modelo actual del banco. Segmento 4.

## 5. Segmento 5

En la tabla 9 se observan las variables incluidas en el segmento 5 del modelo actual replicado y variables del banco. Al comparar las variables incluidas en cada modelo, 10 de éstas se encuentran en ambos. Otra variable que no es idéntica pero sí es similar es el "ratio monto promedio saldos positivos" con "evolución monto promedio saldos positivos". Las diferencias entre las variables que no son iguales en ambos modelos son, tienen relación con el cupo utilizado de la tarjeta de crédito, deuda consumo en la SBIF que no es BCI y saldo mínimo en la cuenta corriente, dichas variables se encuentran en el modelo actual replicado, pero no en el modelo del banco, que en su lugar tiene a las variables del sueldo abonado, deuda directa de consumo en el sistema y el total del cupo de la tarjeta de crédito.

Modelo Actual Replicado			Modelo Actual Banco	
N°	Variable	Beta	Variable	Beta
1	Antigüedad_Ultima_CCT_Abierta	-0.6	Monto_sueldo_abonado	-0.45
2	Cupo_Utilizado_TDC	-0.54	Antigüedad_Ultima_CCT_Abierta	-0.38
3	Deuda_Producto_consumo_Sbif_No_BCI_sist	-0.32	Cupo_Utilizado_TDC	-0.51
4	Evolucion_Monto_Abonado_Lem_CCT_T0_T6	-0.55	Deuda_Directa_Productos_Consumo_sist	-0.41
5	Evolucion_Monto_Credito_consumo_T0_T3	-0.43	Evolucion_Monto_Abonado_Lem_CCT_T3_T6	-0.67
6	Evolucion_Monto_Disponible_LDC_mes_T0_T12	-0.41	Evolucion_Monto_Credito_consumo_T0_T3	-0.36
7	Máximo_Días_Atraso_U3M	-0.41	Evolucion_Monto_Promedio_Saldos_Positivos_T0_T12	-0.56

8	Monto_Promedio_Adeudado_U6M_U3M	-0.66	Máximo_Días_Atraso_U3M	-0.46
9	Monto_Total_Abonado_en_CCT_mes	-0.71	Monto_Promedio_Adeudado_U6M_U3M	-0.67
10	Numero_Avances_Solicitados_U12M	-0.34	Monto_Total_Abonado_en_CCT_mes	-0.33
11	Peor_Situacion_Deuda_Sistema_U6M_sist	-0.3	Numero_Avances_Solicitados_U12M	-0.43
12	Ratio_Monto_Promedio_Saldos_Positivos_CCT_T0_T12	-0.51	Peor_Situacion_Deuda_Sistema_U6M_sist	-0.36
13	Saldo_Minimo_CCT_Activas_Mes	-0.42	Total_Cupo_LDC	-0.17
14	Total_de_Saldos_Disponibles	-0.19	Total_de_Saldos_Disponibles	-0.27
15	Constante	-2.53	Constante	-2.51

Tabla 9 - Comparación de variables de modelo actual replicado con variables del modelo actual del banco. Segmento 5.

## 6. Segmento 6

En la tabla 10 se observan las variables incluidas en el segmento 6 del modelo actual replicado y variables del banco. Al comparar las variables incluidas en cada modelo, 8 de éstas se encuentran en ambos. Las principales diferencias se aprecian en que, el modelo actual incluye el cupo de la línea de crédito y el monto abonado en las cuentas corrientes que no están en el modelo replicado.

Modelo Actual Replicado			Modelo Actual Banco	
N°	Variable	Beta	Variable	Beta
1	Antigüedad_Cliente	-1.06	Antigüedad_Cliente	-1.31
2	Antigüedad_LDC	-0.20	Cupo_Utilizado_TDC	-0.30
3	Cupo_Utilizado_TDC	-0.48	Deuda_Directa_Productos_Consumo_sist	-0.44
4	Deuda_Producto_consumo_Sbif_No_BCI_sist	-0.73	Evolucion_Monto_Credito_consumo_T0_T3	-0.97
5	Evolucion_Monto_Consumo_T0_T6	-1.26	Evolucion_Monto_IFRS_LDC_T0_T3	-0.40
6	Evolucion_Monto_IFRS_LDC_T0_T3	-0.42	Maximo_Dias_Atraso_U3M	-0.61
7	Maximo_Dias_Atraso_U3M	-0.53	Monto_Total_utilizado_TDC	-0.50
8	Monto_Total_Abonado_en_CCT_U3M	-0.75	Monto_Utilizado_LDC_SBIF	-0.16
9	Nivel_Educacional	-0.54	Nivel_Educacional	-0.74
10	Numero_Avance_Solicitados_U6M	-0.44	Peor_Situacion_Deuda_Sistema_U6M_sist	-0.53
11	Peor_Situacion_Deuda_Sistema_U6M_sist	-0.53	Ratio_Monto_Disponible_LDC_T0_T6	-0.29
12	Ratio_Monto_Abonado_CCT_T0_T3	-0.62	Ratio_Monto_Promedio_Saldos_Positivos_CCT_T0_T6	-0.57
13	Ratio_Monto_Disponible_LDC_T0_T6	-0.26	Tipo_Vivienda	-0.74
14	Ratio_Monto_Promedio_Saldos_Positivos_CCT_T0_T6	-0.42	Total_de_Saldos_Disponibles	-0.28
15	Tipo_Vivienda	-0.88	Constante	-1.78
16	Total_Cupo_LDC	-0.79		
17	Constante	-1.79		

Tabla 10 - Comparación de variables de modelo actual replicado con variables del modelo actual del banco. Segmento 6.

## 7. Segmento 7

En la tabla 11 se observan las variables incluidas en el segmento 7 del modelo actual replicado y variables del banco. Este segmento es el que menor cantidad de variables están incluidas en ambos modelos, solo 5 son iguales, mientras que otras capturan un efecto similar ("ratio monto IFRS" con "evolución monto IFRS" y "antigüedad cliente" con "antigüedad cliente

en cuenta corriente” y “antigüedad línea de crédito”, “evolución monto crédito consumo” con “ratio monto consumo”, “evolución monto utilizado en línea de crédito” con “ratio monto disponible línea de crédito”). Es decir, si bien no tienen un gran número de variables idénticas, las variables distintas tratan de capturar comportamientos similares.

Modelo Actual Replicado			Modelo Actual Banco	
N°	Variable	Beta	Variable	Beta
1	Antigüedad_Cliente_en_CCT	-0.95	Antigüedad_Cliente	-1.18
2	Antigüedad_LDC	-0.29	Cupo_Utilizado_TDC	-0.31
3	Cupo_Utilizado_TDC	-0.39	Deuda_Directa_Hipotecaria_sist	-0.59
4	Deuda_Producto_consumo_Sbif_No_BCI_sist	-0.75	Deuda_Directa_Productos_Consumo_sist	-0.61
5	Evolucion_Monto_IFRS_LDC_T0_T6	-0.34	Evolucion_Cupo_LDC_T0_T3	-0.81
6	Evolucion_Monto_Promedio_Saldos_Positivos_T0_T12	-0.37	Evolucion_Monto_Credito_consumo_T0_T3	-0.88
7	Maximo_Atraso_Deuda_Directa_Mes	-0.47	Evolucion_Monto_Utilizado_LDC_T0_T12	-0.40
8	Monto_en_Mora_Blanda_Consumo	-0.37	Maximo_Dias_Atraso_U3M	-0.69
9	Monto_Promedio_Adeudado_U6M_U3M	-0.54	Monto_en_Mora_Blanda_Consumo	-0.43
10	Numero_Avances_Solicitados_U12M	-0.46	Monto_Utilizado_LDC_SBIF	-0.36
11	Peor_Situacion_Deuda_Sistema_U6M_sist	-0.52	Nivel_Educacional	-0.69
12	Ratio_Monto_Abonado_Transferencia_CCT_T0_T12	-0.92	Numero_Avances_Solicitados_U12M	-0.40
13	Ratio_Monto_Consumo_T0_T3	-0.43	Peor_Situacion_Deuda_Sistema_U6M_sist	-0.49
14	Ratio_Monto_Disponible_LDC_T3_T6	-0.28	Ratio_Monto_IFRS_LDC_T0_T12	-1.05
15	Total_de_Saldos_Disponibles	-0.44	Tipo_Vivienda	-0.66
16	Constante	-0.96	Total_de_Saldos_Disponibles	-0.26
17			Constante	-1.00

Tabla 11 - Comparación de variables de modelo actual replicado con variables del modelo actual del banco. Segmento 1.

En la tabla 12, se observa una comparación del desempeño del modelo actual replicado con el modelo actual del banco. El desempeño fue medido en KS. Dado que ni la cantidad de variables y no todas las variables son iguales en los modelos, los resultados de la medición de desempeño serían diferentes. Pese a lo anterior, es posible apreciar que ambos modelos al evaluar toda cartera difieren en un 0,03%, siendo el replicado el inferior.

Desempeño modelo actual (KS)			
Segmento	Replica modelo	Información del banco	Diferencia <sup>2</sup> (%)
1	67.39%	67.00%	0.58%
2	65.51%	67.82%	-3.41%
3	61.82%	58.81%	5.12%
4	64.08%	65.43%	-2.06%
5	64.94%	65.28%	-0.52%
6	60.81%	61.88%	-1.73%
7	61.70%	61.40%	0.49%
<b>Total</b>	<b>65.83%</b>	<b>65.85%</b>	<b>-0.03%</b>

Tabla 12 - Comparación de desempeño de modelos mediante KS.

<sup>2</sup> El cálculo de la diferencia porcentual es:  $100 * \left( \frac{KS \text{ modelo replicado}}{KS \text{ modelo actual}} - 1 \right)$

No se pudo realizar una comparación del estadístico ROC, pues no se contó con la información de dicha métrica por parte del banco.

Luego, se realizó el ejercicio de incluir las mismas variables del modelo actual del banco y se midió el desempeño de éste. Al igual que el caso anterior, se comparó los resultados obtenidos con los informados con el banco (ver tabla 13). Se aprecia que al evaluar toda la cartera existe una diferencia de 0,27% en los KS, siendo mayor la información proporcionada por el banco.

Desempeño modelo actual (KS)			
Segmento	Replica modelo con variables del banco	Información del banco	Diferencia (%)
1	67.24%	67.00%	0.36%
2	65.14%	67.82%	-3.95%
3	58.32%	58.81%	-0.83%
4	63.35%	65.43%	-3.18%
5	63.59%	65.28%	-2.59%
6	62.83%	61.88%	1.54%
7	60.75%	61.40%	-1.06%
<b>Total</b>	<b>65.67%</b>	<b>65.85%</b>	<b>-0.27%</b>

Tabla 13 – tabla comparación de desempeño de modelo replicado con variables del banco y su información entregada para desempeño de su modelo

En el gráfico 8, se observa una comparación del KS de cada modelo (modelo actual replicado, replica con las variables del banco e información del desempeño entregada por el banco de su modelo actual).

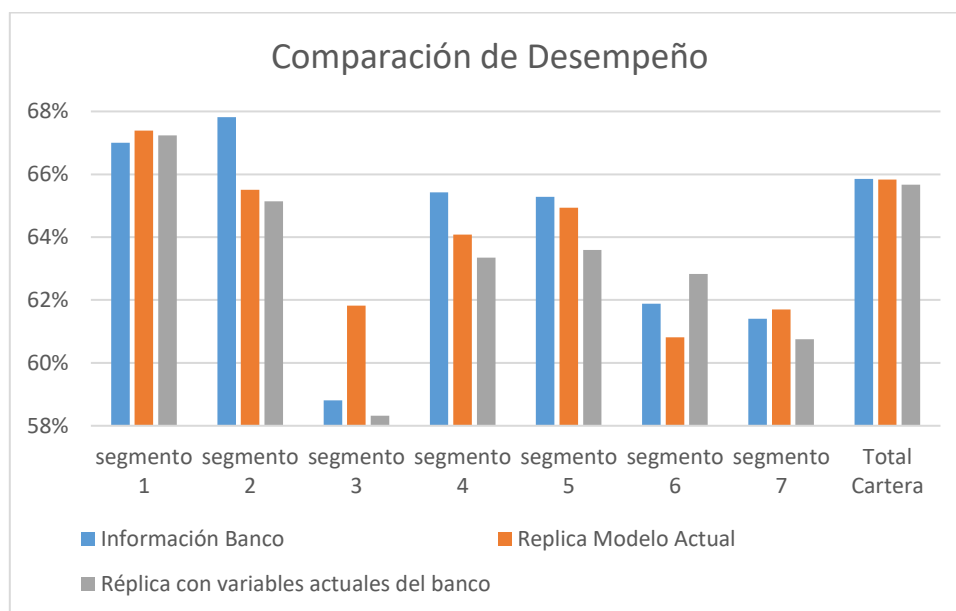


Gráfico 8 – Comparación de KS para los diferentes modelos.



De esta forma se da como finalizada la etapa de réplica del modelo actual del banco. Donde se concluye que, si bien hubo diferencias en la cantidad de variables, variables incorporadas y desempeño (medido en KS) para cada segmento, a nivel de modelo total, evaluando toda la cartera, la diferencia de desempeño entre modelos es mínima. Además, las diferencias encontradas en las variables que se incluyen en los modelos tratan de capturar efectos similares y estas discrepancias pueden ser ratio en vez de evolución o viceversa, antigüedad en un producto diferente (en línea de crédito en vez de tarjeta de crédito o cuenta corriente, etc.) o meses que incluye la variable (últimos 3 meses o últimos 6 meses, por ejemplo). Por lo tanto, esta fue la línea base de comparación para los modelos que se proponen en el capítulo VI y que se comparan en el capítulo VII.

## Capítulo VI. Diseño de nuevos modelos

El objetivo de este capítulo es el desarrollo, y posterior medición de desempeño, de diferentes modelos para la cartera de consumo del banco. El motivo es la incorporación de variables no tradicionales y observar su efecto.

Para la elaboración de los distintos modelos, lo primero que se realizó fue la elección de las variables que se utilizaron para realizar las diferentes segmentaciones. Estas fueron escogidas mediante el criterio: los segmentos que originen sean relativamente estables en el tiempo, es decir, que un cliente no cambie constantemente de segmento a otro en el corto plazo. Dada la anterior definición, quedan fuera variables tales como monto total utilizado en tarjeta de crédito, línea de crédito, monto abonado en cuenta corriente, número de acciones que posee o variables similares, pues estas difieren mes a mes y podrían generar alta movilidad de los clientes entre segmentos.

Las variables escogidas se dividen en tres categorías: *dummys*, nominales y continuas. En la primera categoría se encuentran las variables que señalan si el cliente posee o no una característica determinada, por ejemplo, si posee o no cuenta corriente, tiene o no un crédito de consumo vigente, posee o no acciones, entre otros. En la segunda categoría se encuentran variables del cliente, tal como se edad, renta en UF, sexo, estado civil, nivel educacional, entre otras. Finalmente, en la tercera categoría están las variables, como la edad, renta en UF, antigüedad del cliente en el banco, número de avances nacionales en los últimos 12 meses, etc. La cantidad total de variables utilizadas fue de 112 y el listado completo de estas se encuentra en Anexo 4.

Luego se realizó la construcción de las nuevas variables que serían incorporadas en la etapa "Modelamiento". Para ello, se creó variables a partir de la interacción de productos, tanto en BCI y en el sistema bancario (SBIF). Las variables creadas fueron: categóricas, suma de montos y multiplicación de montos. De esta forma se logra incorporar el efecto cruzado que se produce entre dichos productos y variables. El listado de variables creadas se muestra en Anexo 5.

Realizado lo anterior, se llevó a cabo el desarrollo de los distintos modelos y posterior medición de desempeño. Para la elaboración de cada modelo se utilizó el 70% de la base de datos, mientras que el 30% restante se ocupó para medir su desempeño.

A continuación, se detalla la construcción de los nuevos modelos propuestos. Cada modelo contempla las siguientes etapas: segmentación, modelamiento, y medición de desempeño.

## 6.1 Propuesta de modelo n°1

### 6.1.1 Segmentación propuesta de modelo n°1

El objetivo de la primera propuesta es medir el impacto que poseen las nuevas variables en la segmentación que actualmente posee el banco. Dado lo anterior, se tomó como base la segmentación actual del banco, ver Ilustración 6.

### 6.1.2 Modelamiento de propuesta de modelo n°1

Como se mencionó anteriormente, esta etapa busca identificar las variables que mejor predicen el comportamiento de los clientes de cada uno de los siete segmentos. En esta ocasión se constó con 1.444 variables (1.262 de las bases de datos otorgadas por el banco y 182 nuevas variables construidas). A continuación, en la Tabla 14 se observa el número de variables tramificadas para cada uno de los segmentos. (El detalle de los pasos para llegar al número de variables a incluir en cada modelo se encuentra en Anexo 5).

Segmento	Número de variables a incluir en modelos de propuesta N°1
1	13
2	16
3	18
4	15
5	13
6	15
7	17

Tabla 14 – Número de variables a incluir en modelo propuesta N°1.

A continuación, se detallan las variables que se incluyen en el modelo de cada segmento.

#### 1. Segmento 1

En la tabla 14, se encuentran las variables del modelo para el segmento 1 de la propuesta N°1. Se observa que una de las variables incorporadas es "Multiplicación de deuda consumo BCI por la deuda consumo No BCI"

N°	Variable	Beta
1	Antigüedad_CCT	-0.49
2	Antigüedad_Ultima_Actividad_Cliente	-0.51
3	Evolucion_Monto_Disponible_CCT_T0_T12	-0.30
4	Evolucion_Monto_Disponible_LDC_mes_T0_T12	-0.55
5	Evolucion_Monto_Promedio_Uso_LDC_mes_T6_T12	-0.25

6	Evolucion_Monto_Transferido_Lem_CCT_T0_T12	-0.53
7	Maximo_Dias_atraso_U6M	-0.52
8	Monto_Total_Abonado_en_CCT_U3M	-0.65
9	Multiplicacion_Deuda_Consumo_BCI_NoBCI	-0.37
10	Peor_Situacion_Deuda_Directa_U12M_sist	-0.38
11	Saldo_Minimo_CCT_Activas_Mes	-0.24
12	Total_de_Saldos_Disponibles	-0.27
13	Total_Saldos_Contractuales	-0.88
14	Constante	-2.86

Tabla 15 - Variables y coeficientes incluidos en segmento 1 de propuesta de modelo N°1.

## 2. Segmento 2

En la tabla 16, se encuentran las variables del modelo del segmento 2 de la propuesta N°1. Se observa que no se incluye ninguna variable no tradicional en este modelo, y que las variables que lo componen son las mismas del modelo actual replicado.

N°	Variable	Beta
1	Cupo_Utilizado_TDC	-0.20
2	Deuda_Directa_Productos_Consumo_sist	-0.45
3	Evolucion_Monto_Abonado_T6_T12	-0.93
4	Evolucion_Monto_Disponible_LDC_mes_T0_T12	-0.36
5	Evolucion_Monto_Girado_Atm_T3_T12	0.27
6	Evolucion_Monto_Promedio_Uso_LDC_T3_T12	-0.37
7	Marca_Productos_Posee_LDC_TDC_LEM_Sbif	-0.27
8	Maximo_Dias_Atraso_U3M	-0.37
9	Monto_Total_Abona_en_CCT_mes	-0.80
10	Nivel_Endeudamiento_vigente	-0.42
11	Numero_Avances_Solicitados_U12M	-0.42
12	Ratio_Monto_Abonado_Transferencia_CCT_T6_T12	-0.96
13	Ratio_Monto_Girado_Atm_T3_T12	-0.97
14	Ratio_Monto_Pago_CCT_Lsb_T6_T12	-1.38
15	Saldo_Minimo_CCT_Activas_Mes	-0.29
16	Total_de_Saldos_Disponibles	-0.35
17	Constante	-2.17

Tabla 16 - Variables y coeficientes incluidos en segmento 2 de propuesta de modelo N°1.

## 3. Segmento 3

En la tabla 17, se encuentran las variables del modelo del segmento 3 de la propuesta N°1. Se observa que en este modelo se incluye la variable "Marca productos posee: SGNP, TDC, LEM en el banco BCI", es decir, en este segmento existe diferencia en el comportamiento de incumplimiento según los productos que posean los clientes (sobregiros no pactados, tarjeta de crédito, línea de emergencia u alguna combinación entre ellos).

N°	Variable	Beta
1	Antiguedad_Cliente	-1.65
2	Cupo_Utilizado_TDC	-0.52
3	Deuda_Producto_consumo_Sbif_No_BCI_sist	-0.49
4	Evolucion_Monto_Abonado_CCT_T3_T12	1.15
5	Evolucion_Monto_Abonado_T6_T12	-1.19
6	Evolucion_Monto_Disponible_LDC_mes_T0_T12	-0.40
7	Marca_Productos_Posee_SGNP_TDC_LEM_BCI	-0.22
8	Maximo_Dias_atraso_U6M	-0.52
9	Monto_Total_Abonado_en_CCT_U3M	-0.46
10	Monto_Total_utilizado_LDC	-0.37
11	Numero_Avances_Solicitados_U12M	-0.32
12	Peor_Situacion_Deuda_Directa_U12M_sist	-0.41
13	Ratio_Cupo_Disponible_LDC_T3_T6	1.37
14	Ratio_Monto_Abonado_Transferencia_CCT_T0_T12	-0.80
15	Ratio_Monto_Pago_CCT_Lsb_T0_T12	-0.64
16	Ratio_Monto_Promedio_Saldos_Positivos_CCT_T6_T12	-0.79
17	Saldo_Minimo_CCT_Activas_Mes	-0.29
18	Total_Cupo_LDC	-0.70
19	Constante	-1.48

Tabla 17 - Variables y coeficientes incluidos en segmento 3 de propuesta de modelo N°1.

#### 4. Segmento 4

En la tabla 18, se encuentran las variables del modelo del segmento 4 de la propuesta N°1. Se observa que no se incluye ninguna variable no tradicional en este modelo, y que las variables que lo componen son las mismas del modelo actual replicado.

N°	Variable	Beta
1	Antiguedad_Ultima_TDC_Abierta	-0.60
2	Cantidad_TDC_vigentes	-0.49
3	Edad	-0.67
4	Evolucion_Monto_Disponible_CCT_T0_T12	-0.27
5	Evolucion_Monto_IFRS_LDC_T0_T3	-0.40
6	Evolucion_Monto_Mora_Credito_Consumo_T0_T12	-0.63
7	Maxima_Mora_TDC_U6M	-0.49
8	Monto_Promedio_Adeudado_U6M_U3M	-0.66
9	Monto_Total_Abona_en_CCT_mes	-0.56
10	Nivel_Endudamiento_vigente	-0.61
11	Numero_dias_SGNP_U3M	-0.36
12	Peor_Situacion_Deuda_Directa_mes	-0.37
13	Peor_Situacion_Deuda_Directa_Sistema_mes	-0.36
14	Peor_Situacion_Deuda_Directa_U12M_sist	-0.26
15	Ratio_Monto_Protesto_T3_T12	-0.92
16	Constante	-0.97

Tabla 18 - Variables y coeficientes incluidos en segmento 4 de propuesta de modelo N°1.

## 5. Segmento 5

En la tabla 19, se encuentran las variables del modelo del segmento 5 de la propuesta N°1. Se observa que en este modelo se incluye la variable "Marca productos posee: LDC y/o LEM en SBIF", es decir, en este segmento existe diferencia en el comportamiento de incumplimiento según los productos que posean los clientes (línea de crédito, línea de emergencia u alguna combinación entre ellos). Además, otra variable que destaca es "suma de montos deuda LDC, TDC y LEM SBIF" que es la suma de deuda del cliente en estos tres productos que han sido informados a la SBIF.

N°	Variable	Beta
1	Cupo_Utilizado_TDC	-0.40
2	Deuda_Producto_consumo_Sbif_No_BCI_sist	-0.52
3	Evolucion_Monto_Credito_consumo_T0_T3	-0.81
4	Evolucion_Monto_Disponible_LDC_mes_T0_T12	-0.31
5	Marca_Productos_Posee_LDC_LEM_sbif	-0.32
6	Maximo_Dias_Atraso_U3M	-0.53
7	Monto_Total_Abona_en_CCT_mes	-0.62
8	Peor_Situacion_Deuda_Sistema_U6M_sist	-0.42
9	Ratio_Monto_Promedio_Saldos_Positivos_CCT_T0_T12	-0.38
10	Saldo_Minimo_CCT_Activas_Mes	-0.22
11	Suma_Montos_Deuda_LDC_TDC_LEM_sbif	-0.90
12	Total_Cupo_LDC	-0.49
13	Total_de_Saldos_Disponibles	-0.28
14	Constante	-2.50

Tabla 19 - Variables y coeficientes incluidos en segmento 5 de propuesta de modelo N°1.

## 6. Segmento 6

En la tabla 20, se encuentran las variables del modelo del segmento 6 de la propuesta N°1. Se observa que en este modelo se incluye la variable "Marca productos posee: TDC, LDC y/o LEM en SBIF", es decir, en este segmento existe diferencia en el comportamiento de incumplimiento según los productos que posean los clientes (tarjeta de crédito, línea de crédito, línea de emergencia u alguna combinación entre ellos).

N°	Variable	Beta
1	Antiguedad_Cliente	-1.25
2	CL_Maximo_Dias_Atrasos_CON	-0.42
3	Cupo_Utilizado_TDC	-0.45
4	Deuda_Producto_consumo_Sbif_No_BCI_sist	-0.67
5	Evolucion_Monto_Consumo_T0_T6	-1.27
6	Evolucion_Monto_Disponible_LDC_mes_T0_T12	-0.43
7	Marca_Productos_Posee_LDC_TDC_LEM_Sbif	-0.33
8	Maxima_Mora_TDC_U6M	-0.37
9	Monto_Total_Abonado_en_CCT_U3M	-0.86

10	Numero_Avance_Solicitados_U6M	-0.40
11	Peor_Situacion_Deuda_Sistema_U6M_sist	-0.55
12	Ratio_Dias_MorA_SGN_LDC_T0_T6	-0.27
13	Ratio_Monto_Abonado_CCT_T0_T3	-0.68
14	Ratio_Monto_Promedio_Saldos_Positivos_CCT_T0_T6	-0.45
15	Total_Cupo_LDC	-0.98
16	Constante	-1.77

Tabla 20 - Variables y coeficientes incluidos en segmento 6 de propuesta de modelo N°1.

## 7. Segmento 7

En la tabla 21, se encuentran las variables del modelo del segmento 7 de la propuesta N°1. Se observa que en este modelo se incluye la variable "Tipo de producto que posee", es decir, en este segmento existe diferencia en el comportamiento de incumplimiento según los productos que posean los clientes en el banco BCI (tarjeta de crédito, sobregiros no pactados, tarjeta de crédito, línea de crédito, línea de emergencia u alguna combinación entre ellos).

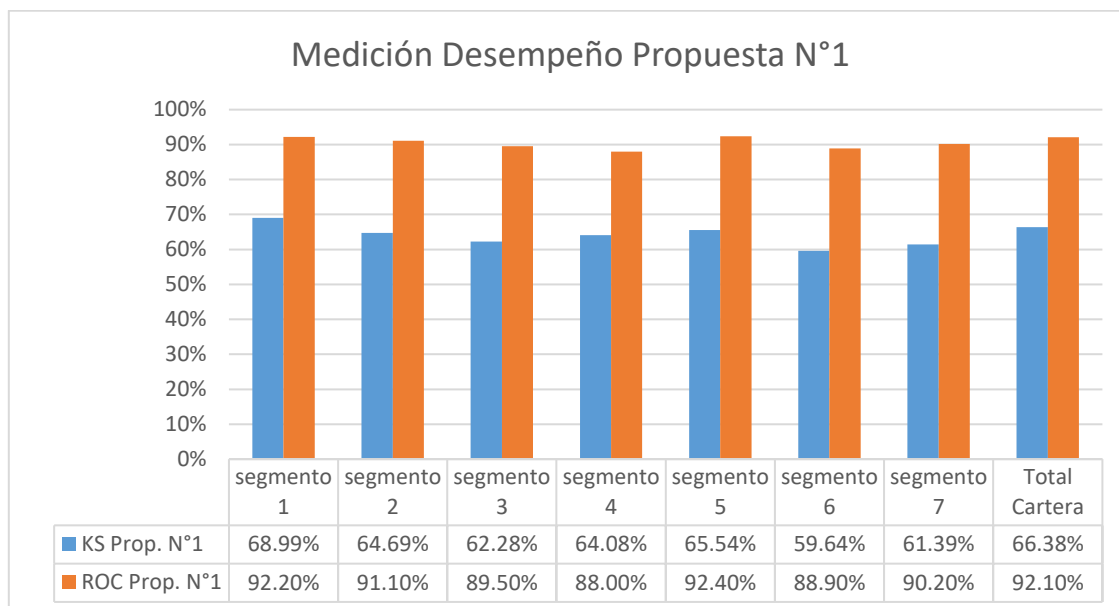
N°	Variable	Beta
1	Antiguedad_Cliente_en_CCT	-0.93
2	Antiguedad_LDC	-0.28
3	Cupo_Utilizado_TDC	-0.40
4	Deuda_Producto_consumo_Sbif_No_BCI_sist	-0.82
5	Evolucion_Monto_IFRS_LDC_T0_T6	-0.38
6	Evolucion_Monto_Promedio_Saldos_Positivos_T0_T12	-0.35
7	Maximo_Atraso_Deuda_Directa_Mes	-0.44
8	Monto_en_Mora_Blanda_Consumo	-0.37
9	Monto_Promedio_Adeudado_U6M_U3M	-0.57
10	Numero_Avances_Solicitados_U12M	-0.44
11	Peor_Situacion_Deuda_Sistema_U6M_sist	-0.53
12	Ratio_Monto_Abonado_Transferencia_CCT_T0_T12	-0.85
13	Ratio_Monto_Consumo_T0_T3	-0.44
14	Ratio_Monto_Disponible_LDC_T3_T6	-0.25
15	Tipo_Producto_posee	-0.61
16	Tipo_Vivienda	-0.80
17	Total_de_Saldos_Disponibles	-0.37
18	Constante	-0.95

Tabla 21 - Variables y coeficientes incluidos en segmento 7 de propuesta de modelo N°1.

### 6.1.3 Medición de desempeño de propuesta de modelo n°1

En esta etapa, se midió el desempeño del modelo de cada segmento y del modelo en la cartera completa. Las métricas que se utilizaron son el estadístico KS y la curva de ROC.

Los resultados obtenidos para la métrica KS y ROC se encuentran en el gráfico 9.



*Gráfico 9 – Medición de desempeño en KS y ROC de propuesta de modelo N°1.*

Es posible apreciar en el gráfico anterior que el modelo posee una gran capacidad predictiva. Al evaluar el desempeño en la cartera de consumo completa, mediante el estadístico KS y ROC, se obtuvo como resultados valores de 66,38% y 92,1%, respectivamente.

## **6.2 Propuesta de modelo n°2**

### **6.2.1 Segmentación de propuesta de modelo n°2**

La segunda propuesta tiene como objetivo segmentar algunos de los segmentos actuales que posee el banco. Esto se debe a que existen segmentos que poseen un gran número de clientes. En gráfico 10, se aprecia que entre el segmento 1, segmento 3 y segmento 5, acumulan un 69% de los clientes, con un 29%, 20% y 20% respectivamente.



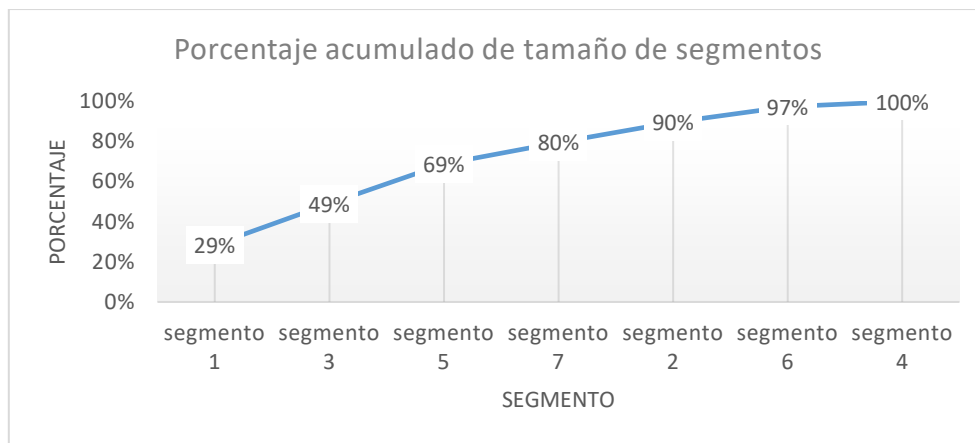


Gráfico 10 - Análisis de tamaño de segmentos.

Dado lo anterior, para los segmentos 1, 3 y 5, se realizó un análisis univariado, con las variables seleccionadas para segmentar. Luego, se elaboró un ranking mediante el *Information Value* de cada variable.

A continuación, se muestra el ranking para cada segmento y la elección de variable para sub-segmentar.

- Segmento 1

En la tabla 22 se encuentra las 5 primeras variables del ranking de *Information Value*.

Ranking	Variable	IV
1	Porcentaje_Uso_LDC	71.92
2	Antigüedad_LDC	38.04
3	CL_Mto_Linea_Disb_SBIF_Ult_Per	32.33
4	Marca_Cliente_Invirtio_mes	31.73
5	Marca_Cliente_Invirtio_U12M	31.73

Tabla 22 - Ranking IV con primeras las 5 variables para segmentación del segmento 1. Propuesta N°2.

Para este segmento se escogió la variable que encabeza el ranking. Al analizar la distribución de dicha variable, que se observa en el gráfico 11, la mayor cantidad de clientes se concentra en las colas de esta, es decir, los clientes o usan casi nada línea o la usan mucho.

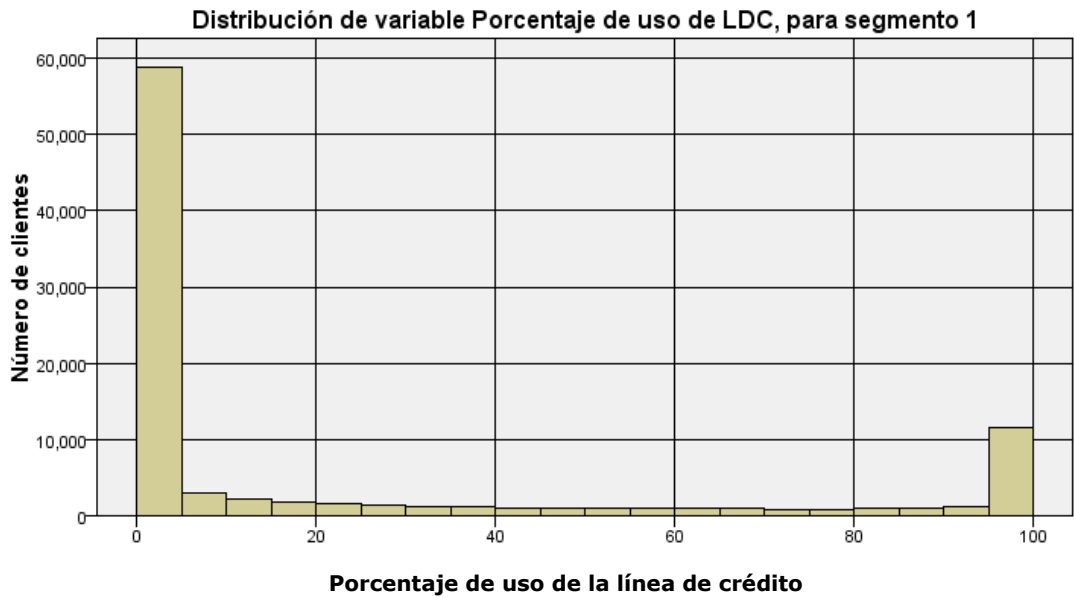


Gráfico 11- Distribución de variable "porcentaje de uso de línea de crédito" en el segmento 1.

Los segmentos que se generó a partir de dicha variable son: clientes con un porcentaje de uso de la línea de crédito menor o igual al 92% y clientes con un porcentaje de uso de la línea de crédito mayor al 92%.

En el gráfico 12, se aprecia el tamaño, en cantidad de clientes, de los segmentos generados.

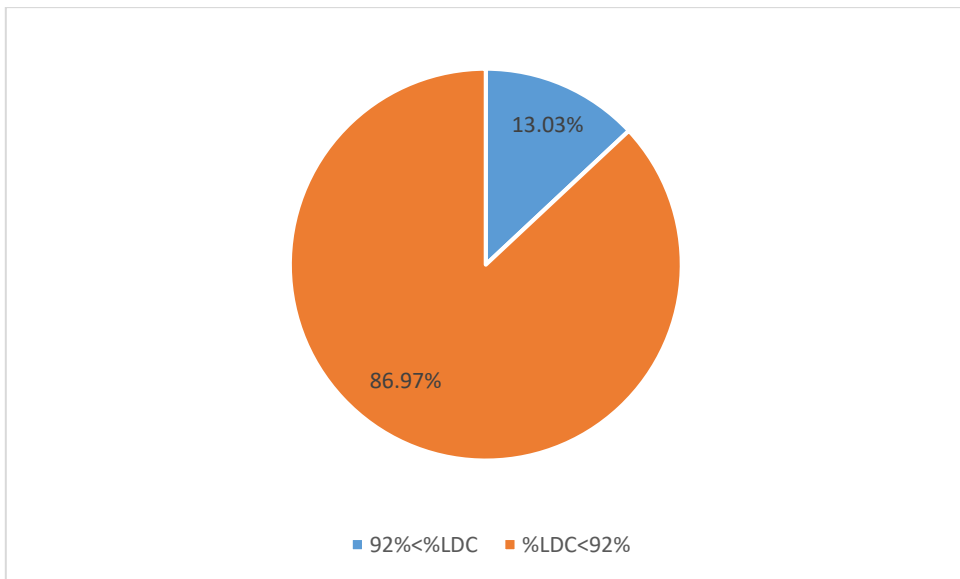


Gráfico 12 - Tamaño de segmentos generados por la segmentación a partir de la variable "porcentaje de uso de línea de crédito"

En el gráfico 13 se observa el Bad Rate tanto del segmento 1 y los segmento generados.

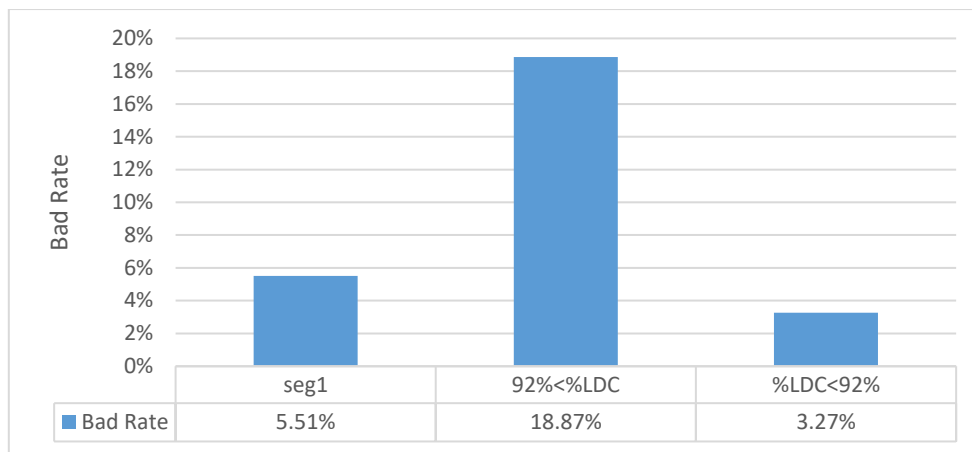


Gráfico 13 – Comparación de Bad Rate de segmento 1 versus nuevos segmentos generados en la propuesta N°2.

Se aprecia que los segmentos generados poseen niveles de *Bad Rate* estadísticamente diferentes. Identificando que los clientes que poseen altos niveles de uso de su línea de crédito (superior al 92% de cupo utilizado) tienen mayor tasa de incumplimiento, casi 6 veces mayor que la tasa incumplimiento con niveles de uso inferior al 92% de su cupo.

- Segmento 3

En la tabla 23 se encuentra las 5 primeras variables del listado del ranking de *Information Value*.

Ranking	Variable	IV
1	Antigüedad_LDC	55.96
2	Marca_Solicito_Avances_U12M	46.46
3	Numero_Avances_Solicitados_U12M	46.46
4	Marca_Solicito_Avances_U6M	44.33
5	Numero_Avance_Solicitados_U6M	44.33

Tabla 23 - Ranking IV con primeras las 5 variables para segmentación del segmento 3. Propuesta N°2.

Al analizar la variable con mayor *Information Value*, se observa que 30.144 clientes, que representan 44,75% del segmento, no posee Línea de Crédito (ver gráfico 14). Por lo anterior, se decidió no utilizar la variable “antigüedad en la línea de crédito” para segmentar, pues si bien estadísticamente existe discriminancia, no se puede atribuir que alguien sin el producto tenga similar comportamiento que un cliente que es nuevo en el producto (línea de crédito).

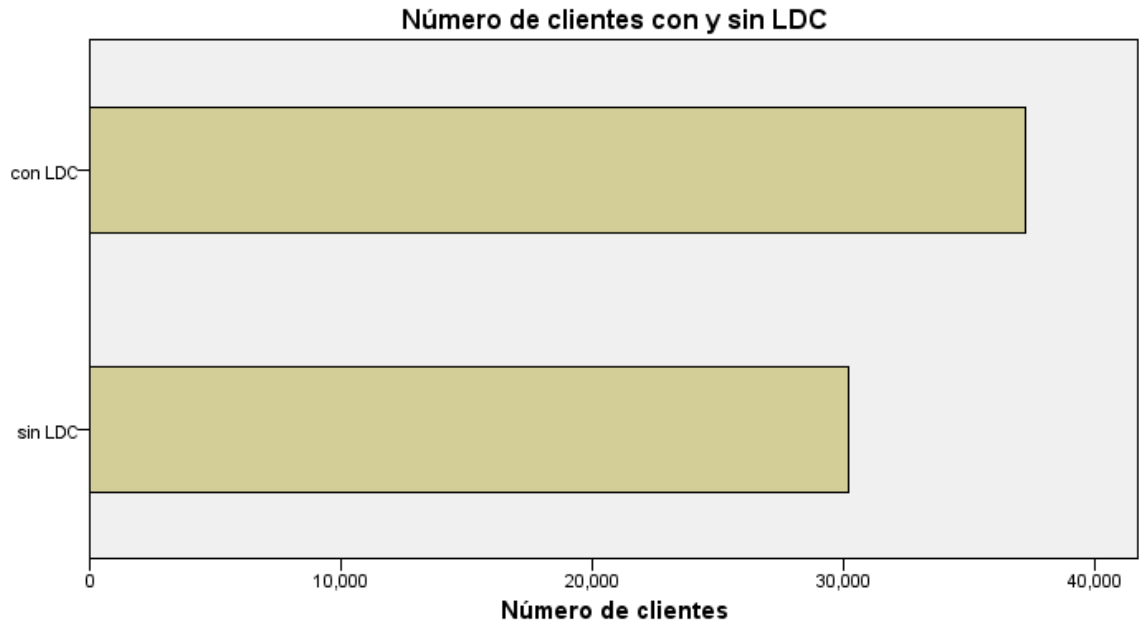


Gráfico 14 – Distribución de clientes con y sin línea de crédito en segmento 3.

Dado lo anterior, se pasó a utilizar la segunda variable del ranking para generar la segmentación, la cual es una variable *dummy* que señala si el cliente solicitó o no un avance en los últimos 12 meses.

Los segmentos generados son: cliente solicitaron avances en los últimos 12 meses y cliente no solicitaron avances en los últimos 12 meses.

En el gráfico 15, se aprecia el tamaño, en cantidad de clientes, de los segmentos generados.

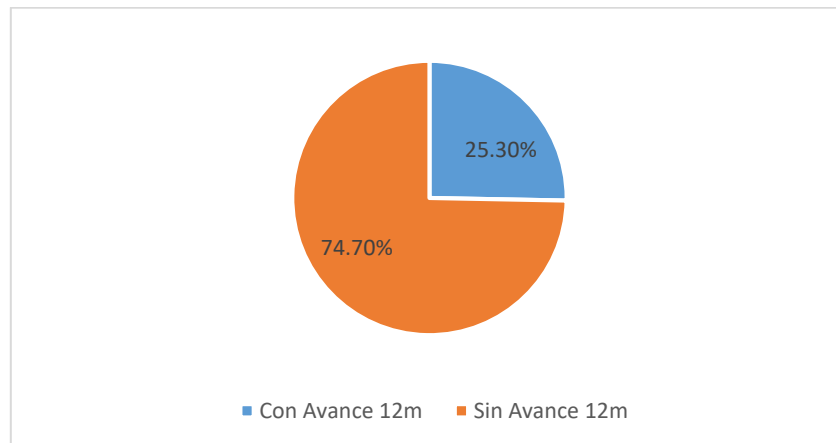


Gráfico 15 – Tamaño segmentos generados por la segmentación a partir de la variable "Marca solicitó avances en los últimos 12 meses"

En el gráfico 16 se observa el Bad Rate, tanto del segmento 3 como de los nuevos segmentos generados.

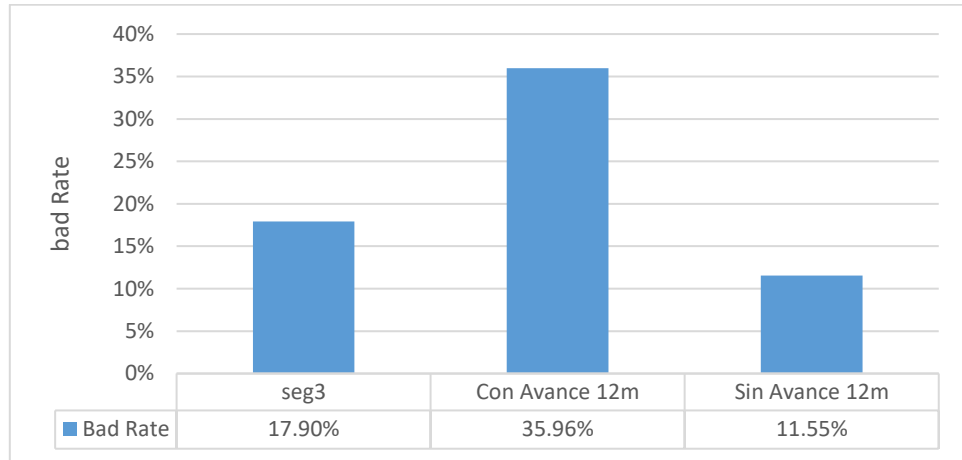


Gráfico 16 – comparación de Bad Rate segmento 3 versus nuevos segmentos generados en la propuesta N°2.

Los segmentos originados presentan niveles de *Bad Rate*, sumamente distintos. Se aprecia que los clientes que solicitaron avances nacionales en los últimos 12 meses, poseen casi el triple de tasa de incumplimiento mayor que quienes no lo hicieron.

- Segmento 5

En la tabla 19 se encuentra las 5 primeras variables del listado del ranking de *Information Value*.

Ranking	Variable	IV
1	Antigüedad_LDC	38.97
2	Marca_Solicito_Avances_U6M	26.38
3	Numero_Avance_Solicitados_U6M	26.38
4	Marca_Solicito_Avances_U12M	24.66
5	Marca_Solicito_Avances_U3M	24.55

Tabla 24 - Ranking IV con primeras las 5 variables para segmentación del segmento 5. Propuesta N°2.

Al igual que el caso anterior, al analizar la variable con mayor *Information Value*, se observa que 42.870 clientes, que representan 65,8% del segmento, no posee Línea de Crédito (ver gráfico 17). Por lo anterior, se decidió no utilizar la variable antigüedad en la Línea de crédito como para segmentar, pues si bien estadísticamente existe discriminancia, no se puede atribuir que alguien sin el producto tenga similar comportamiento que un cliente que es nuevo en el producto (línea de crédito).

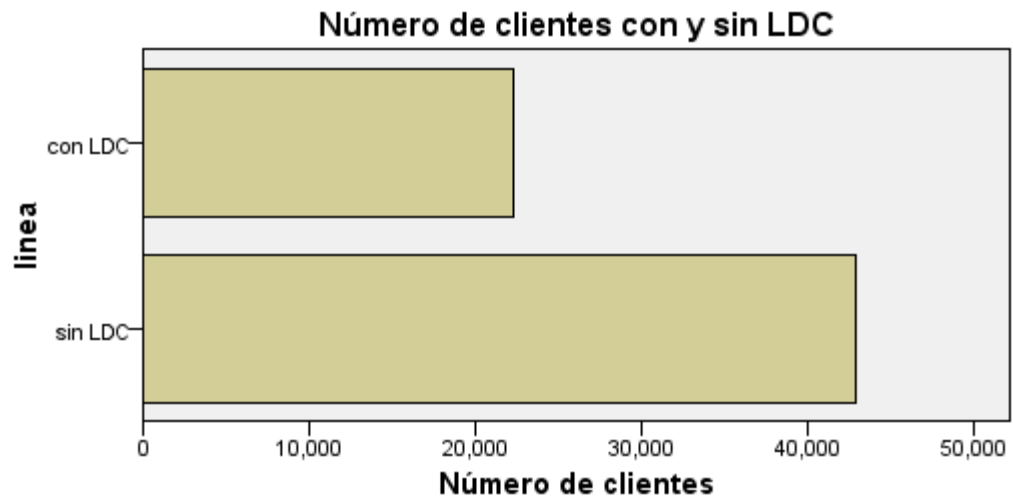


Gráfico 17 - Distribución de clientes con y sin línea de crédito en segmento 5.

Dado lo anterior, se pasó a utilizar la segunda variable del ranking para generar la nueva segmentación, la cual es una variable *dummy* que señala si el cliente solicitó o no un avance en los últimos 6 meses.

Los nuevos segmentos generados son: cliente solicitaron avances en los últimos 6 meses y cliente no solicitaron avances en los últimos 6 meses.

En el gráfico 18 se aprecia el tamaño, en cantidad de clientes, de los segmentos generados.

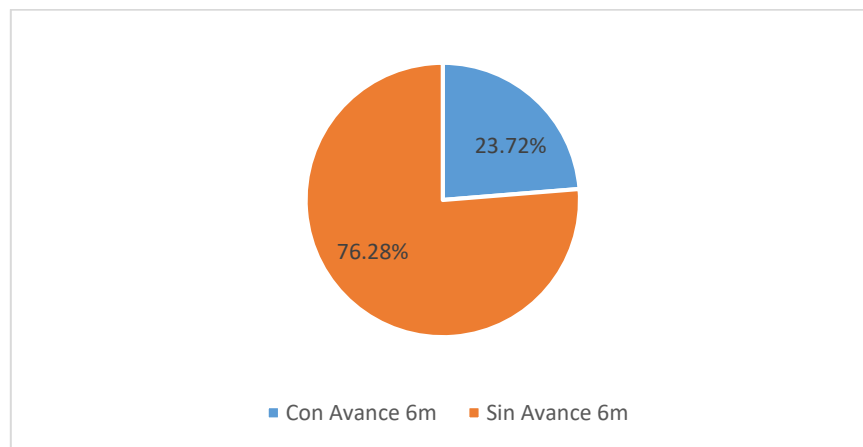


Gráfico 18 - tamaño de segmentos generados por la segmentación a partir de la variable "¿Marca solicitó avances en los últimos 6 meses?"

En el gráfico 19, se observa el *Bad Rate*, tanto del segmento 5 como de los segmentos generados.

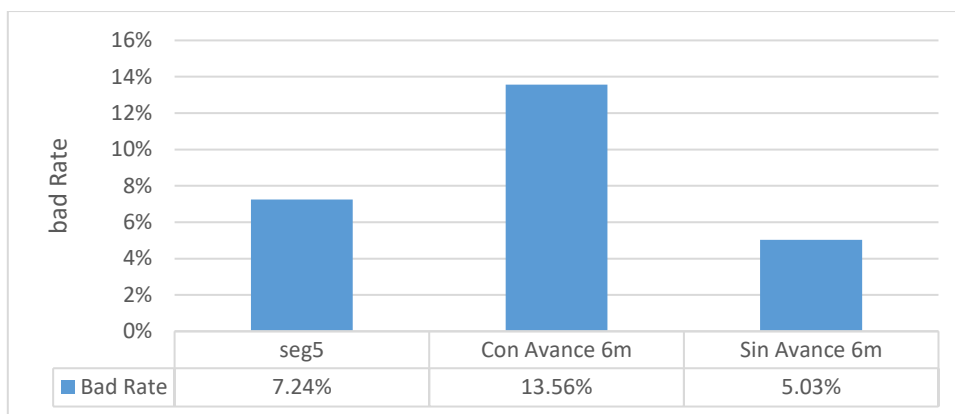


Gráfico 19 – Comparación de Bad Rate segmento 5 versus nuevos segmentos generados en la propuesta N°2.

Se aprecia que los clientes que solicitaron avances en los últimos 6 meses poseen casi un triple de tasa de incumplimiento que aquellos que no solicitan.

Finalmente, luego de la generación de nueva segmentación a partir de los segmentos 1, 3 y 5, del modelo actual del banco, se originó una nueva segmentación, que se aprecia en la Ilustración 7.

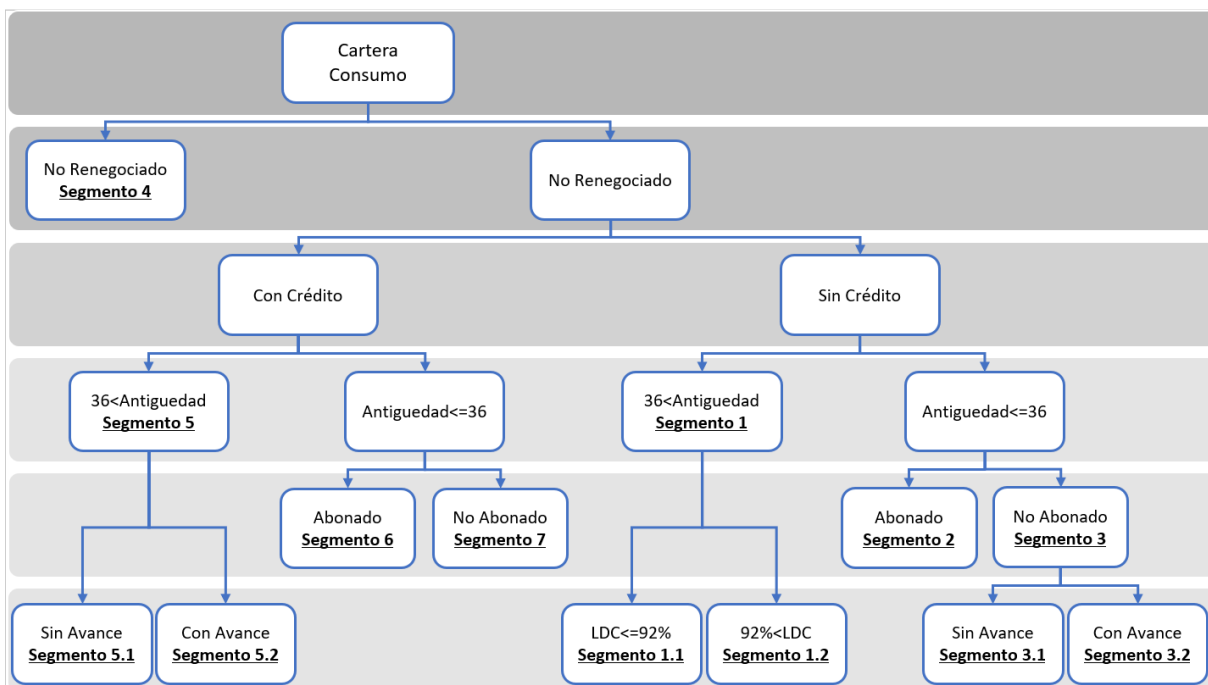


Ilustración 7 – Segmentación propuesta n°2

### 6.2.2 Modelamiento de propuesta de modelo n°2

De igual forma que para la propuesta anterior, para cada nuevo segmento originado se realizó el análisis univariado para las 1.446 variables. En la Tabla 25 se observan la cantidad de variables a incluir en cada modelo. (Detalle de cómo se llegó a estos números se encuentra en Anexo 6)

Segmento	Número de variables a incluir en modelos de propuesta N°2
<b>1.1</b>	13
<b>1.2</b>	13
<b>3.1</b>	14
<b>3.2</b>	18
<b>5.1</b>	14
<b>5.2</b>	14

Tabla 25 – Número de variables a incluir en modelo, propuesta N°2

A continuación, se señalan las variables que fueron incluidas en el modelo de cada segmento generado en la propuesta N°2, junto a su coeficiente respectivo.

### 1. Segmento 1.1

En la tabla 26, se encuentran las variables del modelo del segmento 1.1 de la propuesta N°2. Se observa que en este modelo se incluye la variable “Multiplicación Deuda Consumo por Deuda Comercial en sistema”.

N°	Variable	Beta
1	Antigüedad_Ultima_Actividad_Cliente	-0.52
2	Antigüedad_Ultima_CCT_Abierta	-0.63
3	Cupo_Utilizado_TDC	-0.74
4	Evolucion_Monto_Disponible_CCT_T0_T12	-0.38
5	Maximo_Dias_Atraso_U3M	-0.63
6	Monto_Total_Abona_en_CCT_mes	-0.60
7	Monto_Total_Depositos_CCT_mes	-0.46
8	Multiplicacion_Deuda_Consumo_Comercial_directa_sistema	-1.08
9	Numero_Avances_Solicitados_U12M	-0.99
10	Peor_Situacion_Deuda_Directa_U12M_sist	-0.60
11	Ratio_Monto_Pago_CCT_Lsb_T3_T6	-0.74
12	Ratio_Monto_Promedio_Saldos_Positivos_CCT_T3_T12	-0.34
13	Ratio_Monto_Protestos_Historicos_T3_T12	-1.04
14	Constante	-1.43

Tabla 26 - Variables y coeficientes incluidos en segmento 1.1 de propuesta de modelo N°2

### 2. Segmento 1.2

En la tabla 27, se encuentran las variables del modelo del segmento 1.2 de la propuesta N°2. Se observa que en este modelo se incluyen las variables “Suma cupo línea de crédito y línea de emergencia” y “suma de montos invertidos en fondos mutuos y depósitos a plazo en el mes”.



N°	Variable	Beta
1	Antiguedad_Ultima_CCT_Abierta	-0.67
2	Cupo_Utilizado_TDC	-0.37
3	Deuda_Producto_consumo_Sbif_No_BCI_sist	-0.33
4	Evolucion_Monto_Disponible_LDC_mes_T0_T12	-0.57
5	Evolucion_Monto_Promedio_Saldos_Positivos_T0_T12	-0.49
6	Monto_Total_Abonado_en_CCT_U3M	-0.54
7	Peor_Situacion_Deuda_Directa_U12M_sist	-0.47
8	Peor_Situacion_Deuda_Directa_U3	-0.50
9	Ratio_Dias_Moroso_SGN_LDC_T6_T12	-0.26
10	Ratio_Monto_IFRS_LDC_T0_T3	-0.53
11	Suma_Cupo_LDC_LEM	-0.65
12	Suma_Montos_Invertido_FondosMutuos_DAP_mes	-0.42
13	Total_Saldos_Contractuales	-1.89
14	Constante	-3.39

Tabla 27 - Variables y coeficientes incluidos en segmento 1.2 de propuesta de modelo N°2

### 3. Segmento 3.1

En la tabla 28, se encuentran las variables del modelo del segmento 3.1 de la propuesta N°2. Se observa que en este modelo se incluye la variable "marca productos que posee: línea de crédito, tarjeta de crédito, línea de emergencia u alguna combinación entre dichos productos"

N°	Variable	Beta
1	Cupo_Utilizado_TDC	-0.46
2	Evolucion_Monto_Disponible_CCT_T0_T12	-0.32
3	Evolucion_Monto_Disponible_LDC_mes_T0_T12	-0.50
4	Evolucion_Monto_Pago_CCT_Lsb_T0_T3	-0.67
5	Marca_Productos_Posee_LDC_TDC_LEM_Sbif	-0.73
6	Maxima_Mora_TDC_U12M	-0.31
7	Maximo_Atraso_Deuda_Directa_Mes	-0.51
8	Monto_Total_Abonado_en_CCT_U3M	-0.41
9	Monto_Total_Depositos_CCT_mes	-0.71
10	Peor_Situacion_Deuda_Directa_U12M_sist	-0.53
11	Ratio_Monto_Abonado_Lem_CCT_T3_T6	-0.50
12	Ratio_Monto_Abono_Transfencias_CCT_T3_T6	-0.74
13	Total_de_Saldos_Disponibles	-0.45
14	Total_Saldos_Contractuales	-0.60
15	Constante	-1.98

Tabla 28 - Variables y coeficientes incluidos en segmento 3.1 de propuesta de modelo N°2

### 4. Segmento 3.2

En la tabla 29, se encuentran las variables del modelo del segmento 3.2 de la propuesta N°2. Se observa que en este modelo se incluyen las variables

“Suma Deuda Consumo por Deuda Comercial en sistema” y “Tipo de productos que posee” es una variable nominal que señala cuáles son los productos que posee el cliente.

N°	Variable	Beta
1	Antiguedad_TDC_mas_Antigua	-1.11
2	Antiguedad_Ultimo_Uso_CCT	-1.09
3	Cupo_Utilizado_TDC	-0.51
4	Edad	-1.12
5	Evolucion_Monto_IFRS_LDC_T0_T12	-0.31
6	Evolucion_Monto_Protesto_T0_T3	-0.62
7	Evolucion_Monto_Transferido_CCT_T0_T6	-0.58
8	Maximo_Atraso_Deuda_Directa_Mes	-0.61
9	Monto_en_Mora_Blanda_TDC	-0.33
10	Monto_Total_Abonado_en_CCT_U3M	-0.65
11	Monto_Total_utilizado_LDC	-0.53
12	Numero_Avance_Solicitados_U3M	-0.59
13	Peor_Situacion_Deuda_Sistema_U6M_sist	-0.55
14	Ratio_Monto_Abonado_T0_T3	-0.49
15	Ratio_Monto_Disponible_CCT_T3_T6	-0.43
16	Ratio_Monto_Pago_LDC_T0_T3	-0.80
17	Suma_Deuda_Consumo_Comercial_directa_sistema	-0.70
18	Tipo_Producto_posee	-0.93
19	Constante	-0.55

Tabla 29 - Variables y coeficientes incluidos en segmento 3.2 de propuesta de modelo N°2

## 5. Segmento 5.1

En la tabla 26, se encuentran las variables del modelo del segmento 1.1 de la propuesta N°2. Se observa que en este modelo no se incluye ninguna variable creada.

N°	Variable	Beta
1	Antiguedad_Cliente_en_CCT	-0.62
2	Deuda_Directa_Productos_Consumo_sist	-0.86
3	Evolucion_Dias_Moroso_LDC_T3_T12	-0.81
4	Evolucion_Monto_Credito_consumo_T0_T3	-0.71
5	Evolucion_Monto_Disponible_LDC_mes_T0_T12	-0.35
6	Meses_Expiracion_LDC	-0.35
7	Monto_Promedio_Adeudado_U6M_U3M	-0.67
8	Monto_Total_Abona_en_CCT_mes	-0.64
9	Numero_eventos_SGNP_U3M	-0.38
10	Peor_Situacion_Deuda_Directa_mes	-0.50
11	Peor_Situacion_Deuda_Directa_U12M_sist	-0.46
12	Ratio_Monto_Abonado_Lem_CCT_T3_T6	-0.36

13	Ratio_Monto_Ingreso_CCT_T0_T12	-0.40
14	Total_de_Saldos_Disponibles	-0.34
15	Constante	-2.87

Tabla 30 - Variables y coeficientes incluidos en segmento 5.1 de propuesta de modelo N°2

## 6. Segmento 5.2

En la tabla 26, se encuentran las variables del modelo del segmento 1.1 de la propuesta N°2. Se observa que en este modelo se incluyen las variables "Suma deudas sobregiros no pactados, tarjeta de crédito y línea de emergencia que el cliente posee en el BCI" y "Tipo de productos que posee" es una variable nominal que señala cuáles son los productos que posee el cliente.

N°	Variable	Beta
1	Antiguedad_Ultima_CCT_Abierta	-0.66
2	Cupo_Utilizado_TDC	-0.48
3	Deuda_Comercial_Sbif_NoBCI	-0.53
4	Evolucion_Monto_Consumo_T0_T12	-1.08
5	Evolucion_Monto_Disponible_CCT_T3_T12	-0.45
6	Evolucion_Monto_Mora_Consumo_T0_T6	-0.30
7	Maximo_Dias_Atraso_U3M	-0.50
8	Maximo_Monto_Atraso_Deuda_Directa_mes	-0.33
9	Monto_Total_Abona_en_CCT_mes	-0.67
10	Peor_Situacion_Deuda_Sistema_U6M_sist	-0.57
11	Ratio_Monto_Abonado_T3_T12	-0.90
12	Ratio_Monto_Moroso_SGN_T6_T12	-0.34
13	Suma_Deuda_SGNP_TDC_LEM_BCI	-0.88
14	Tipo_Producto_posee	-0.67
15	Constante	-1.79

Tabla 31 - Variables y coeficientes incluidos en segmento 5.2 de propuesta de modelo N°2

### 6.2.3 Medición de desempeño de propuesta de modelo n°2

En esta etapa, se midió el desempeño de los modelos de los segmentos originados. Las métricas que se utilizaron son el estadístico KS y la curva de ROC. Los resultados obtenidos para las métricas KS y ROC se encuentran en el gráfico 20.

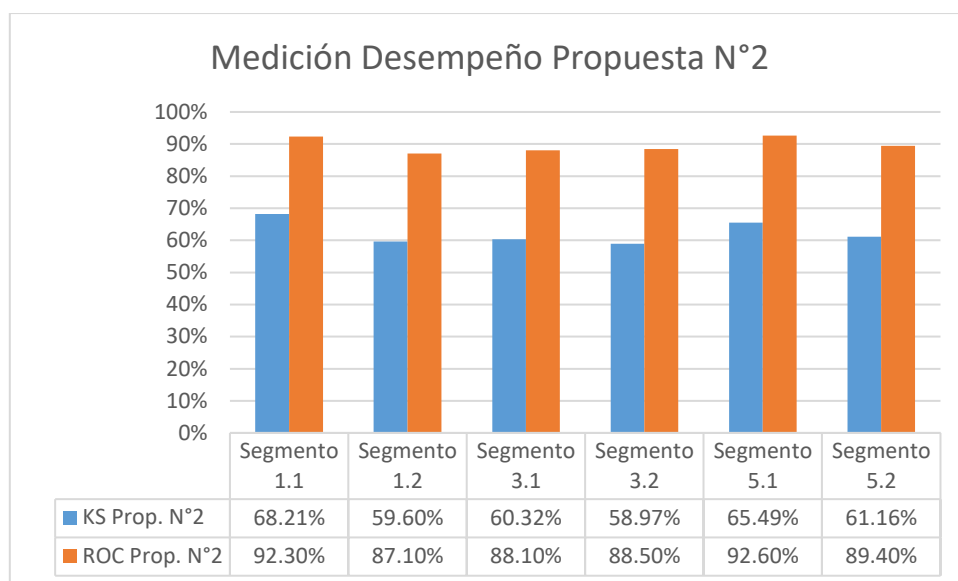


Gráfico 20 - Medición de desempeño en KS y ROC de propuesta de modelo N°2.

Es posible apreciar en el gráfico anterior que los modelos poseen una gran capacidad predictiva para uno de los dos segmentos originados (alto valor KS y ROC en segmento 1.1 y valor inferior en segmento 1.2, lo mismo para los segmentos 3.1-3.2 y 5.1-5.2, pero en menor medida que el primer caso). Pese a lo anterior, todos los modelos presentan desempeños buenos.

## 6.3 Propuesta de modelo n°3

### 6.3.1 Segmentación de propuesta de modelo n°3

El objetivo de la tercera propuesta es generar una segmentación completamente nueva. Para esto, se tomó la base de datos y se realizó análisis de discriminancia con las variables anteriormente seleccionadas para segmentar.

El primer paso fue segmentar al segmento “No Renegociado”, pues el segmento “Renegociado” se incluye en esta segmentación y por la cantidad de clientes que posee no se segmentará.

Al realizar en análisis univariado, incluyendo como restricción que los segmentos generados por el árbol de decisión posean un porcentaje de clientes mayor a un 33% (del segmento No Renegociado), se elaboró el ranking de *Information Value*. El listado del ranking con las primeras 5 variables se observa en la tabla 32.

Ranking	Variable	IV
1	Sueldo_liquido_en_UF	10.25
2	Antiguedad_TDC	10.2
3	Edad	9.35
4	Nivel_Educacional	8.44
5	Apalancamiento_cuotas_Crédito_consumo	3.8

Tabla 32 – Ranking IV con primeras las 5 variables para segmentar a segmento “No Renegociados”. Propuesta N°3.

En el primer lugar del ranking se encuentra la renta líquida en UF, su distribución se aprecia en el gráfico 21.

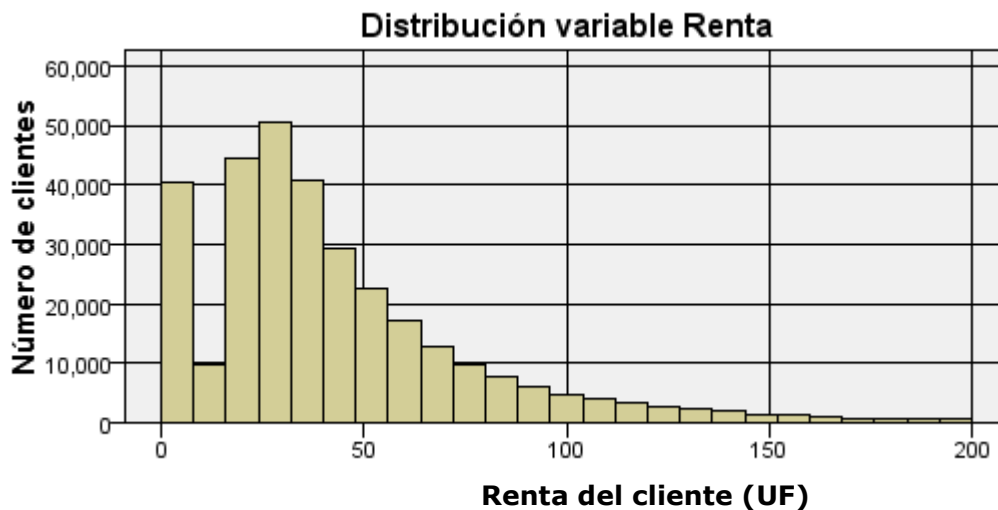


Gráfico 21 – Distribución de renta en UF de los clientes del segmento “No Renegociados”.

Al segmentar por la renta, se generó 2 segmentos: clientes con renta menor o igual a 41UF y clientes con renta mayor a 41UF, como se muestra en la Ilustración 8.

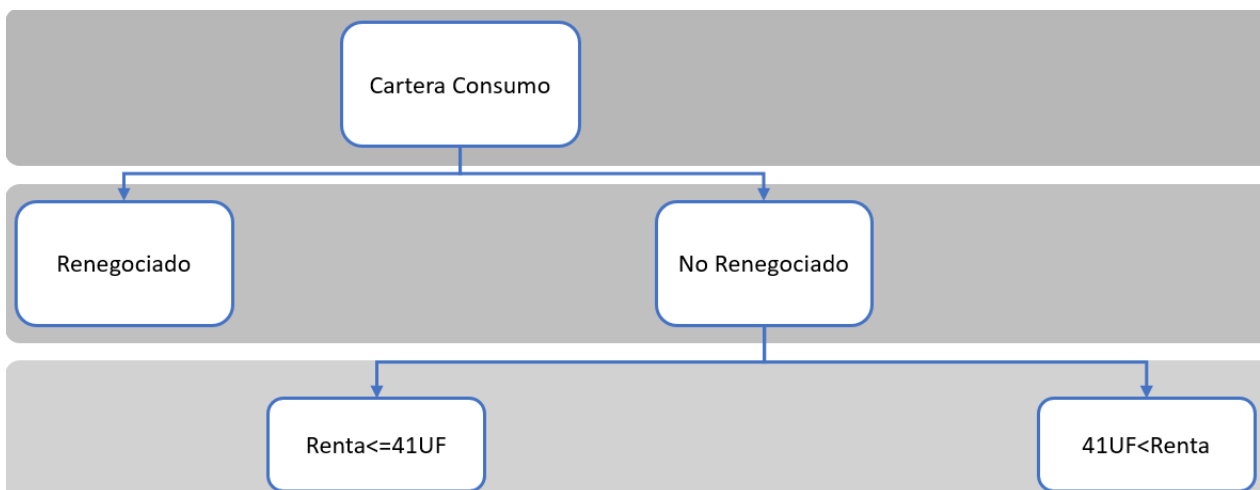


Ilustración 8 – Resultado de primera segmentación generada. Propuesta N°3.

En los gráficos 22 y 23, se encuentra el *Bad Rate* y evolución del *Bad Rate* de los 3 segmentos generados hasta el momento, cuales son:

- Renegociados
- No Renegociado con Renta  $\leq$  41UF
- No Renegociado con 41UF < Renta

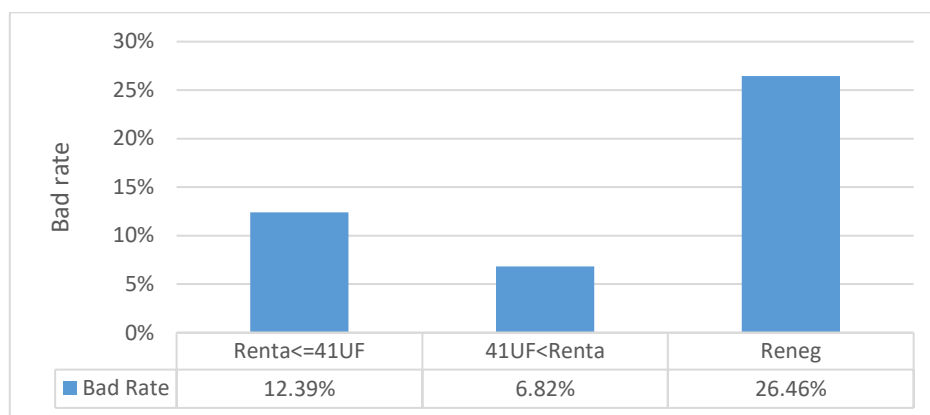


Gráfico 22 – Comparación de Bad Rate de primera segmentación. Propuesta N°3.

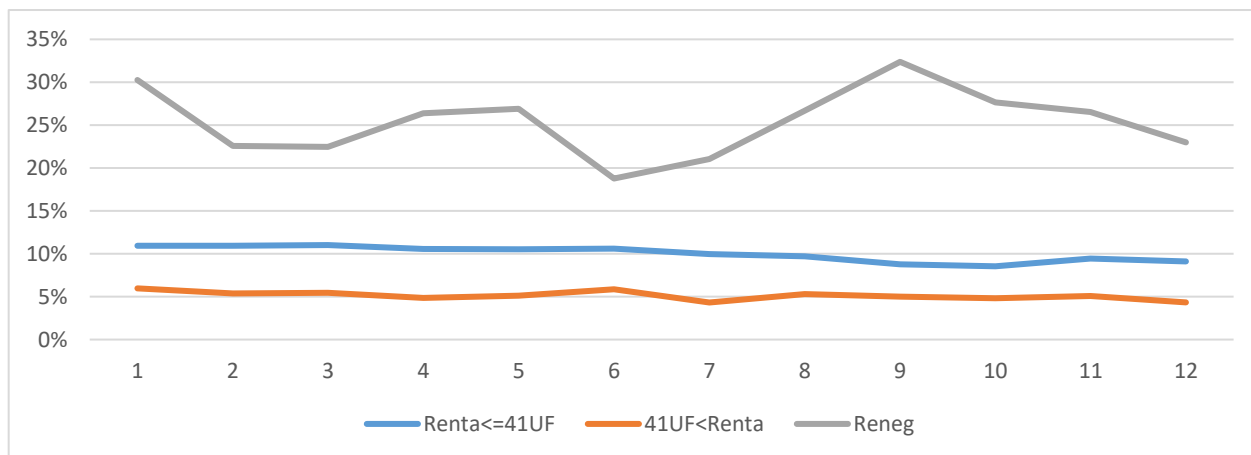


Gráfico 23 – Evolución temporal de Bad Rate de primera segmentación. Propuesta N°3.

Se aprecia que los segmentos originados, presentan diferencias estadísticamente significativas en sus niveles de *Bad Rate*, son estables en el tiempo y no existe traslape. Por lo tanto, es razonable esta segmentación

Luego, se procedió a realizar una nueva segmentación de los 2 segmentos originados anteriormente. Con el mismo procedimiento, el ranking de las 5 variables con mayor *Information Value* de cada segmento se encuentran en la Tabla 33 y Tabla 34.

- Ranking para segmento No renegociado con Renta  $\leq 41$  UF

Ranking	Variable	IV
1	Antigüedad_LDC	40.41
2	Marca_Solicito_Avances_U12M	33.44
3	Numero_Avances_Solicitados_U12M	33.44
4	Antigüedad_CCT	26.97
5	Monto_Linea_Disponible_Sbif	25.67

Tabla 33 – Ranking IV con primeras las 5 variables para segmentar a segmento "No Renegociados – Renta $\leq 41$ UF". Propuesta N°3.

Al analizar la cantidad de clientes que poseen línea de crédito, nuevamente tiene un alto porcentaje de los clientes del segmento no posee el producto. Luego, se tomó la segunda variable del ranking, "Marca solicito Avances en los últimos 12 meses". Con esta variable se originó dos nuevos segmentos:

- No renegociado, con Renta  $\leq 41$  UF y con Avances 12 meses.
- No renegociado, con Renta  $\leq 41$  UF y sin Avances 12 meses.

- Ranking para segmento No Renegociado con  $41UF < Renta$

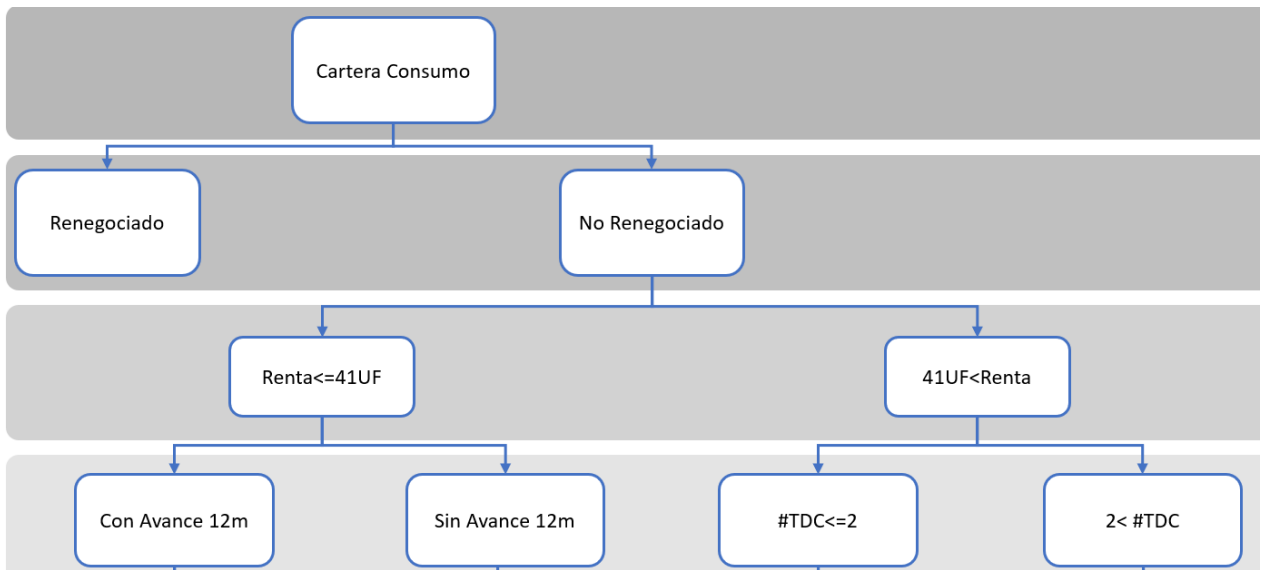
Ranking	Variable	IV
1	Antigüedad_LDC	41.19
2	Cantidad_TDC_vigentes	33.05
3	Marca_Prestamos_Otros_Bancos	18.93
4	Cantidad_LEM_vigentes	15.38
5	Nivel_Endeudamiento_vigente	11.27

Tabla 34 - Ranking IV con primeras las 5 variables para segmentar a segmento "No Renegociados –  $41UF < Renta$ ". Propuesta N°3.

Como en los casos anteriores, la cantidad de clientes que poseen línea de crédito nuevamente tiene un alto porcentaje de los clientes del segmento no posee el producto. Luego, se tomó la segunda variable del ranking, "Cantidad de tarjetas de crédito vigentes". Con esta variable se originó dos nuevos segmentos:

- No renegociado, con  $41UF < Renta$  y  $\#TDC \leq 2$ .
- No renegociado, con  $41UF < Renta$  y  $2 < \#TDC$ .

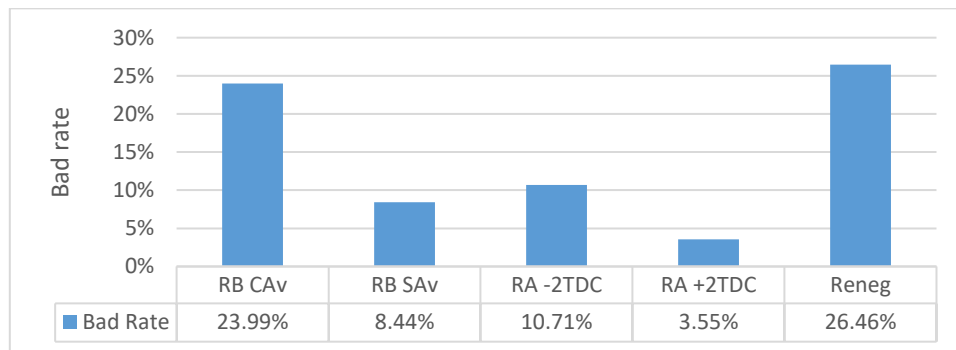
El resultado de la segmentación anterior se aprecia en la Ilustración 9.



*Ilustración 9 – Resultado de segunda segmentación generada. Propuesta N°3.*

En los gráficos 24 y 25, se encuentra el *Bad Rate* y evolución del *Bad Rate* de los 5 segmentos generados hasta el momento, cuales son:

- Renegociados (reneg).
- No renegociado con Renta  $\leq 41$  UF con Avances 12 meses (RB CAv).
- No renegociado con Renta  $\leq 41$  UF sin Avances 12 meses (RB SAv).
- No renegociado con  $41UF < Renta$  y  $\#TDC \leq 2$  (RA -2TDC).
- No renegociado con  $41UF < Renta$  y  $2 < \#TDC$  (RA +2TDC).



*Gráfico 24 – Comparación Bad Rate de segunda segmentación. Propuesta N°3.*



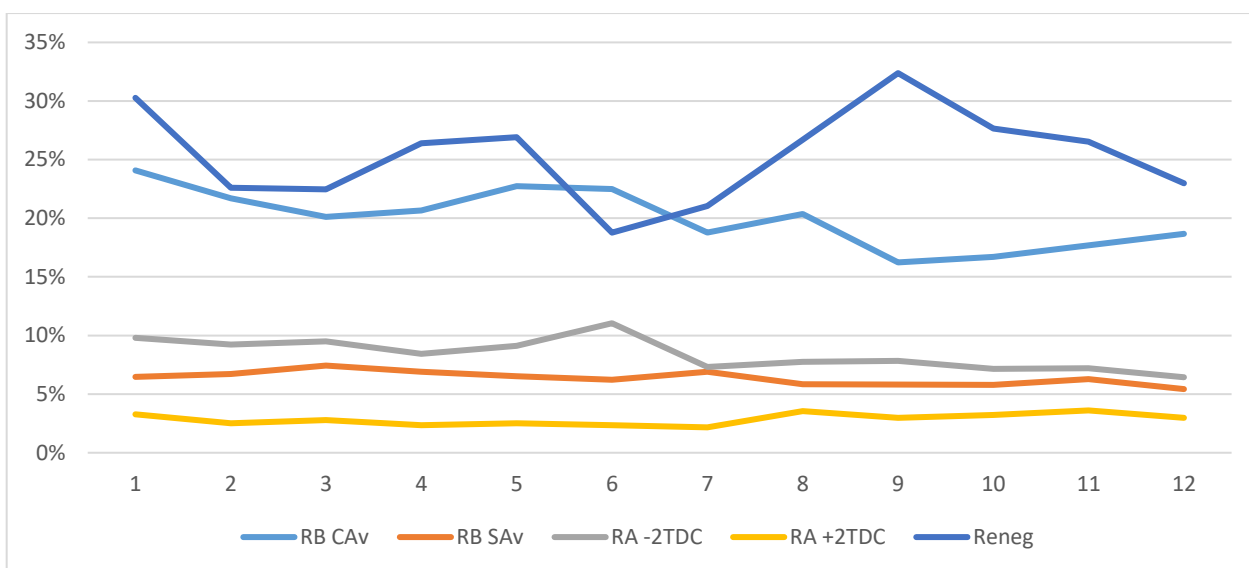


Gráfico 25 - Evolución temporal de Bad Rate de segunda segmentación. Propuesta N°3.

En esta ocasión, los niveles de *Bad Rate* son diferentes entre los segmentos generados, sin embargo, ya comienza a haber entrecruzamientos entre la evolución temporal de los *Bad Rate*. Como el cruce es solo entre el segmento renegociados y RB CAV, y ocurre en 2 periodos, se decide seguir realizando una nueva segmentación.

Luego, se procedió a realizar una nueva segmentación de los 4 segmentos originados anteriormente, con el mismo procedimiento. El ranking de las 5 variables con mayor *Information Value* de cada segmento se encuentran en la Tabla 35, 36, 37 y 38

- Ranking para segmento No renegociado, con Renta  $\leq$  41 UF y con Avances últimos 12 meses.

Ranking	Variable	IV
1	Antigüedad_CCT	35.27
2	Antigüedad_Ultima_CCT_Abierta	35.16
3	Meses_antigüedad_TDC_mas_antigua	34.45
4	Antigüedad_Cliente	34.43
5	Antigüedad_Cliente_en_CCT	34.31

Tabla 35- Ranking IV con primeras las 5 variables para segmentar a segmento "No Renegociados - Renta  $\leq$  41UF - Con Avances últimos 12 meses". Propuesta N°3.

Para este caso, como las primeras 5 variables del ranking hacen alusión a la antigüedad. Además, tomando en consideración que no todos los clientes poseen cuenta corriente y tarjeta de crédito, se decidió segmentar por la antigüedad cliente. Esta variable representa la antigüedad del cliente en cualquiera de los productos del banco. De esta manera se generó 2 segmentos. Clientes que poseen antigüedad

menor o igual a 27 meses y los que tienen antigüedad mayor a 27 meses.

- Ranking para segmento No renegociado, con Renta  $\leq$  41 UF y sin Avances últimos 12 meses

Ranking	Variable	IV
1	Antigüedad_LDC	36.94
2	Cantidad_TDC_vigentes	23.03
3	Antigüedad_Ultimo_Uso_LDC	22.08
4	Antigüedad_CCT	19.01
5	Antigüedad_Cliente	18.11

Tabla 36 - Ranking IV con primeras las 5 variables para segmentar a segmento "No Renegociados - Renta  $\leq$  41UF - Sin Avances últimos 12 meses". Propuesta N°3.

Como en los casos anteriores, la cantidad de clientes que poseen línea de crédito nuevamente tiene un alto porcentaje de los clientes del segmento no posee el producto. Luego, se tomó la segunda variable del ranking, "cantidad de tarjetas de crédito vigentes". Con dicha variable, se originó dos nuevos segmentos: clientes con 2 o más Tarjetas de crédito vigentes y los que no poseen o solo tienen 1 tarjeta de crédito vigente.

- Ranking para segmento No renegociado, con  $41UF < Renta$  y  $\#TDC \leq 2$

Ranking	Variable	IV
1	Marca_Solicitado_Avances_U6M	41.87
2	Numero_Avance_Solicitados_U6M	41.87
3	Marca_Solicitado_Avances_U12M	39.78
4	Numero_Avances_Solicitados_U12M	39.78
5	Antigüedad_LDC	37.05

Tabla 37 - Ranking IV con primeras las 5 variables para segmentar a segmento "No Renegociados -  $41UF < Renta$  -  $\#TDC \leq 2$ ". Propuesta N°3.

Para este segmento, se utilizó como variable para segmentar "marca solicitó avances últimos 6 meses", que ocupa el primer lugar del ranking. Con dicha variable se originó los segmentos de clientes que si solicitaron algún avance en los últimos 6 meses y los que no lo hicieron.

- Ranking para segmento No renegociado, con  $41UF < Renta$  y  $2 < \#TDC$

Ranking	Variable	IV
1	Marca_Solicitado_Avances_U12M	53.04
2	Antigüedad_LDC	33.09
3	Marca_Cliente_Invirtio_mes	27.89
4	Marca_Cliente_Invirtio_U12M	27.89
5	Marca_Cliente_Invirtio_U3M	27.89

Tabla 38 - Ranking IV con primeras las 5 variables para segmentar a segmento "No Renegociados -  $41UF < Renta - 2 < \#TDC$ ". Propuesta N°3.

Para este segmento, se utilizó como variable para segmentar "marca solicitó avances últimos 12 meses", que ocupa el primer lugar del ranking. Con ella se originó los segmentos de clientes que sí solicitaron algún avance en los últimos 12 meses y los que no lo hicieron.

El resultado de la segmentación anterior origina un árbol como el que se aprecia en la Ilustración 10.

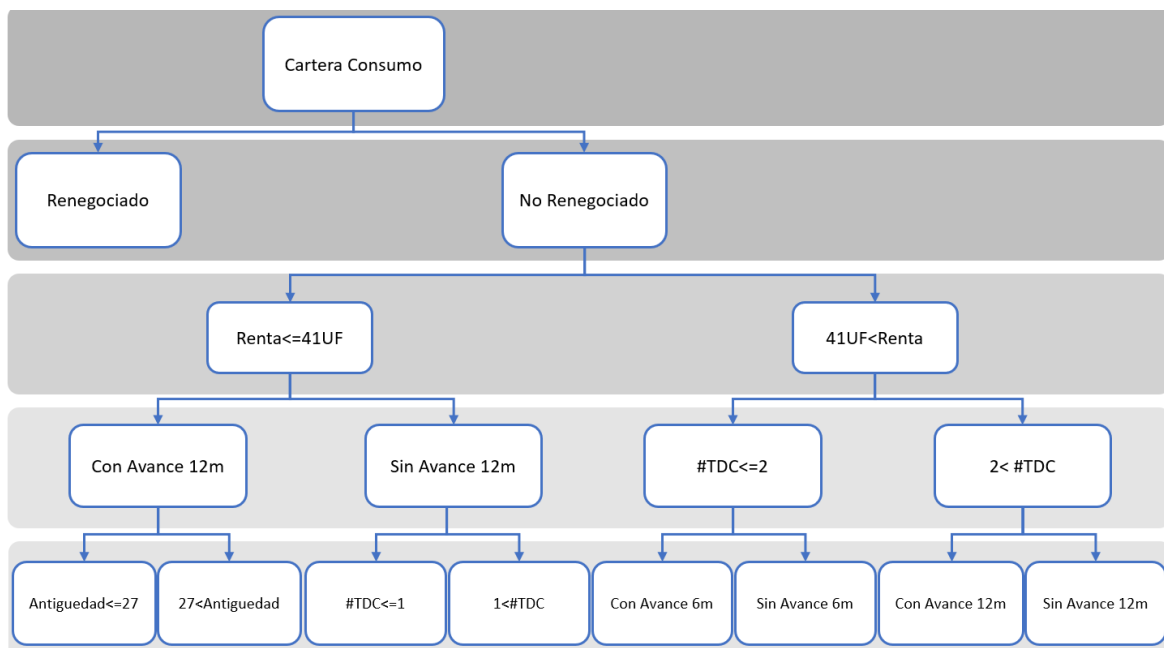


Ilustración 10 - Resultado de tercera segmentación generada. Propuesta N°3

En los gráficos 26 y 27, se encuentra el *Bad Rate* y evolución del *Bad Rate* de los 9 segmentos generados hasta el momento, cuales son:

- Renegociados (reneg).
- No renegociado con  $Renta \leq 41 UF$  con Avances 12 meses y antigüedad menor o igual a 27 meses (RB CA Nuevo).
- No renegociado con  $Renta \leq 41 UF$  con Avances 12 meses y antigüedad mayor a 27 (RB CA Ant).

- No renegociado con Renta  $\leq 41$  UF sin Avances 12 meses que no posee TDC o solo posee una (RB SA -1TDC).
- No renegociado con Renta  $\leq 41$  UF sin Avances 12 meses que posee 2 o más TDC (RB SA +2TDC).
- No renegociado con  $41UF < Renta$  y  $\#TDC \leq 2$  con Avance 6 meses (RA -2TDC CA).
- No renegociado con  $41UF < Renta$  y  $\#TDC \leq 2$  sin Avance 6 meses (RA -2TDC SA).
- No renegociado con  $41UF < Renta$  y  $2 < \#TDC$  con avance 12 meses (RA +2TDC CA).
- No renegociado con  $41UF < Renta$  y  $2 < \#TDC$  sin avance 12 meses (RA +2TDC SA).

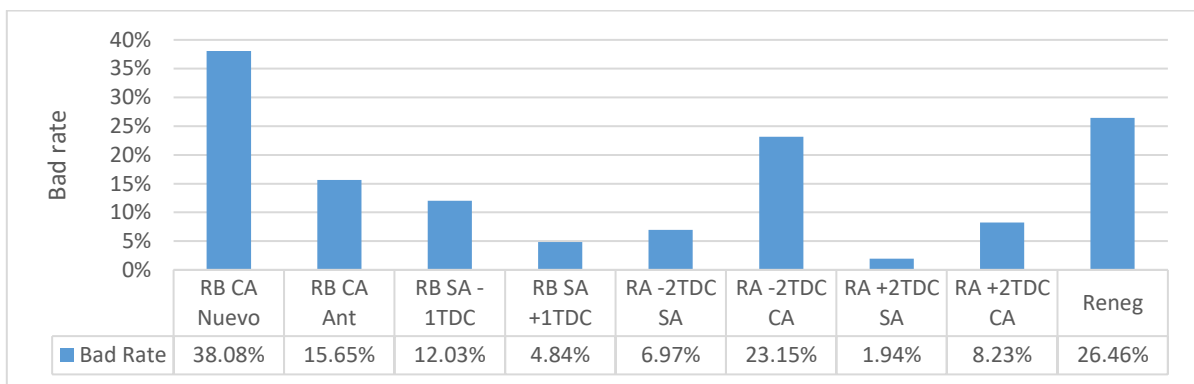


Gráfico 26 – Comparación de Bad Rate de tercera segmentación. Propuesta N°3

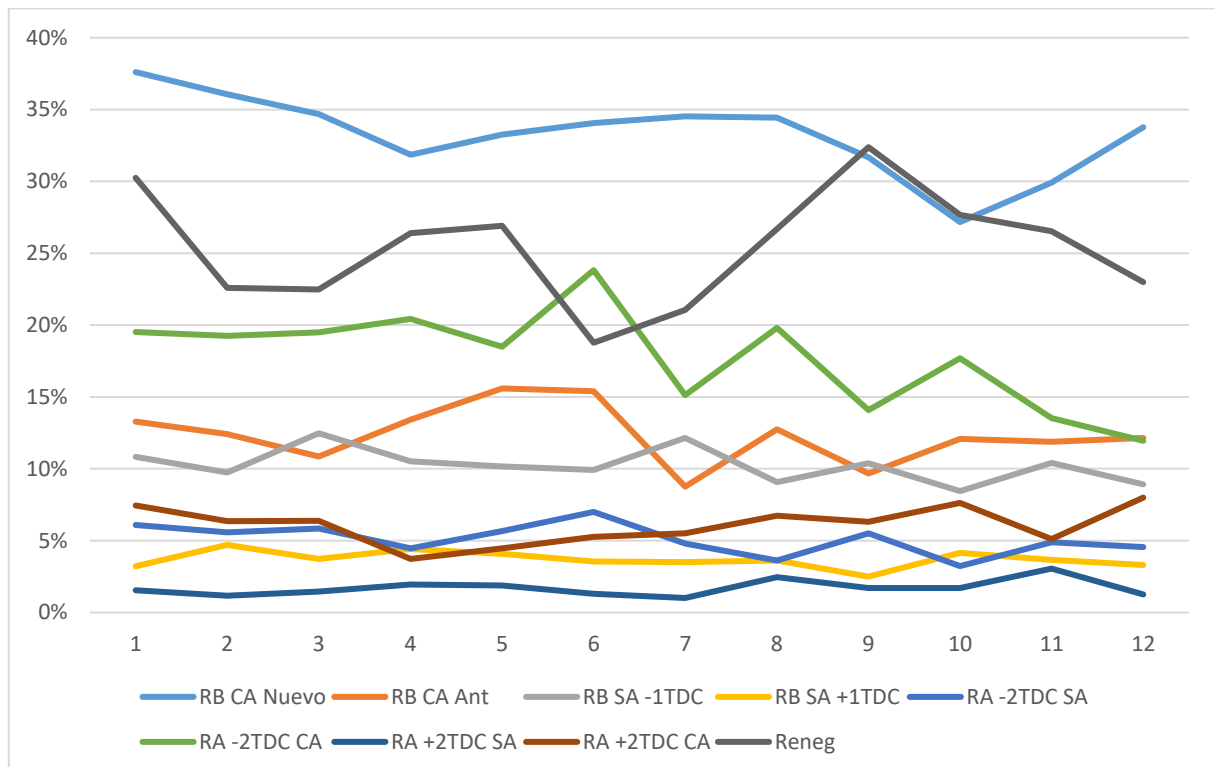


Gráfico 27 – Evolución temporal de Bad Rate de tercera segmentación. Propuesta N°3

Se aprecia que los *Bad Rate* de los segmentos poseen diferencias entre sí. Sin embargo, su evolución temporal posee muchos cruces entre los distintos segmentos. Dado lo anterior, se considera que los segmentos generados no poseen *Bad Rate* estables en el tiempo. Por esta razón se decidió reducir la cantidad de segmentos. El método para esto fue gráfico. El resultado que se obtuvo se aprecia en el gráfico 28. En éste se aprecia una estabilidad mayor de los *Bad Rate* en el tiempo.

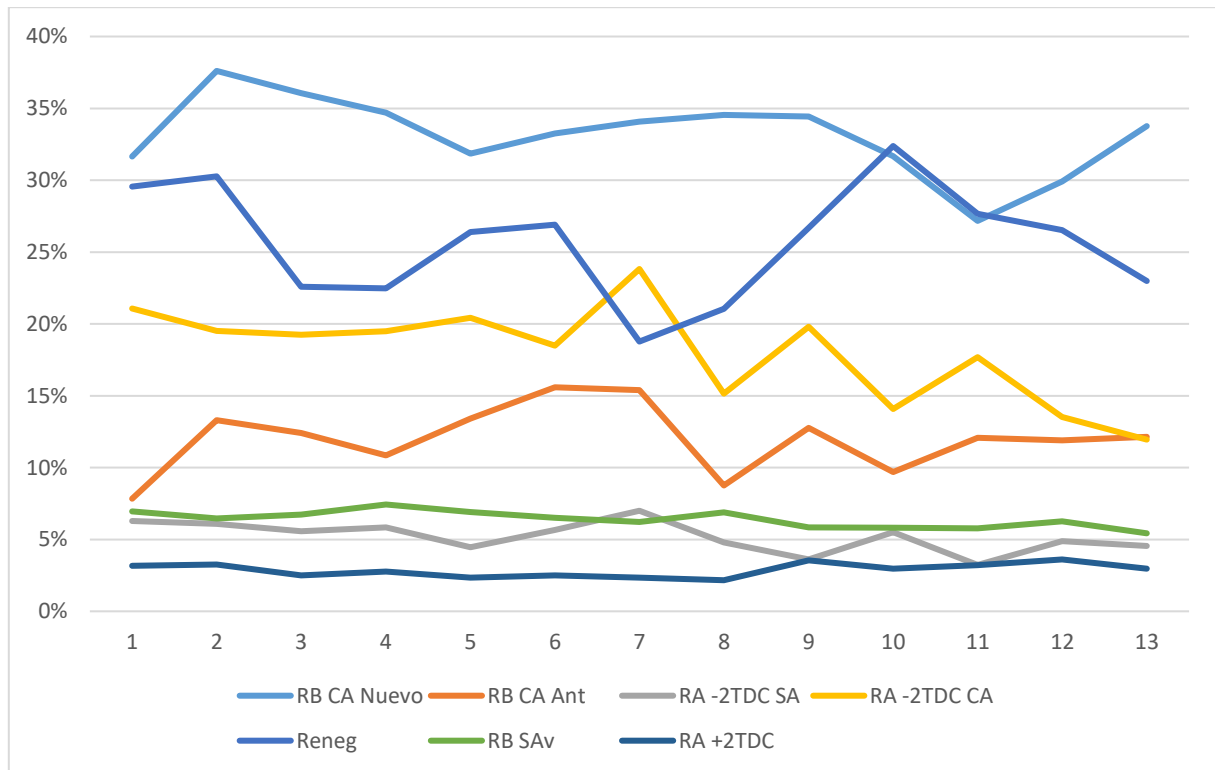


Gráfico 28 - Evolución temporal de Bad Rate con segmentación final. Propuesta N°3

Dado lo anterior, la segmentación final que se propone es la que aprecia en la Ilustración 11, la cual contempla los siguientes 7 segmentos:

- Renegociados (reneg).
- No renegociado, con Renta  $\leq 41$  UF, con Avances últimos 12 meses y antigüedad menor o igual a 27 meses (RB CA Nuevo).
- No renegociado, con Renta  $\leq 41$  UF, con Avances últimos 12 meses y antigüedad mayor a 27 (RB CA Ant).
- No renegociado, con Renta  $\leq 41$  UF y sin Avances últimos 12 meses (RB SA).
- No renegociado, con  $41UF < Renta$ ,  $\#TDC \leq 2$  y con Avance últimos 6 meses (RA -2TDC CA).
- No renegociado, con  $41UF < Renta$ ,  $\#TDC \leq 2$  y sin Avance 6 meses (RA -2TDC SA).
- No renegociado, con  $41UF < Renta$  y  $2 < \#TDC$  (RA +2TDC).

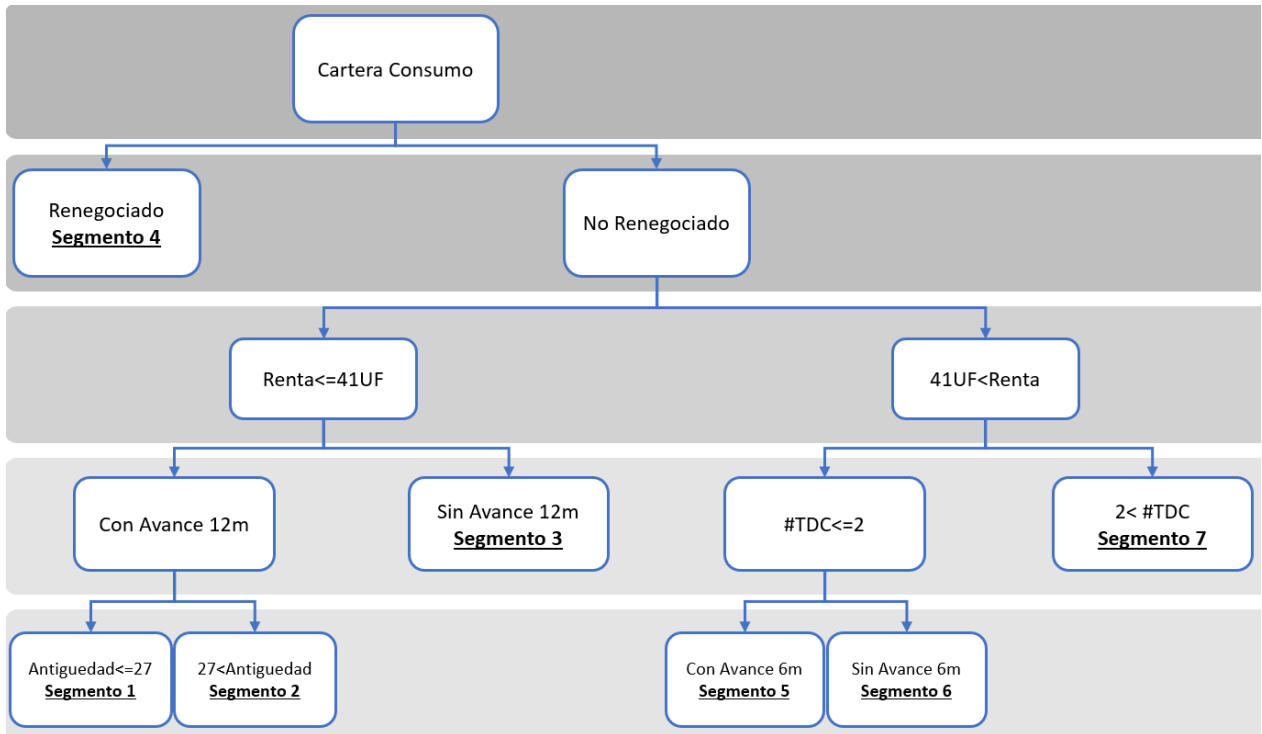


Ilustración 11 – Resultado de segmentación final generada. Propuesta N°3

Los *Bad Rate* de cada segmento se observan en el gráfico 29. Se observa que los niveles de *Bad Rate* presentan diferencias estadísticas entre los segmentos.

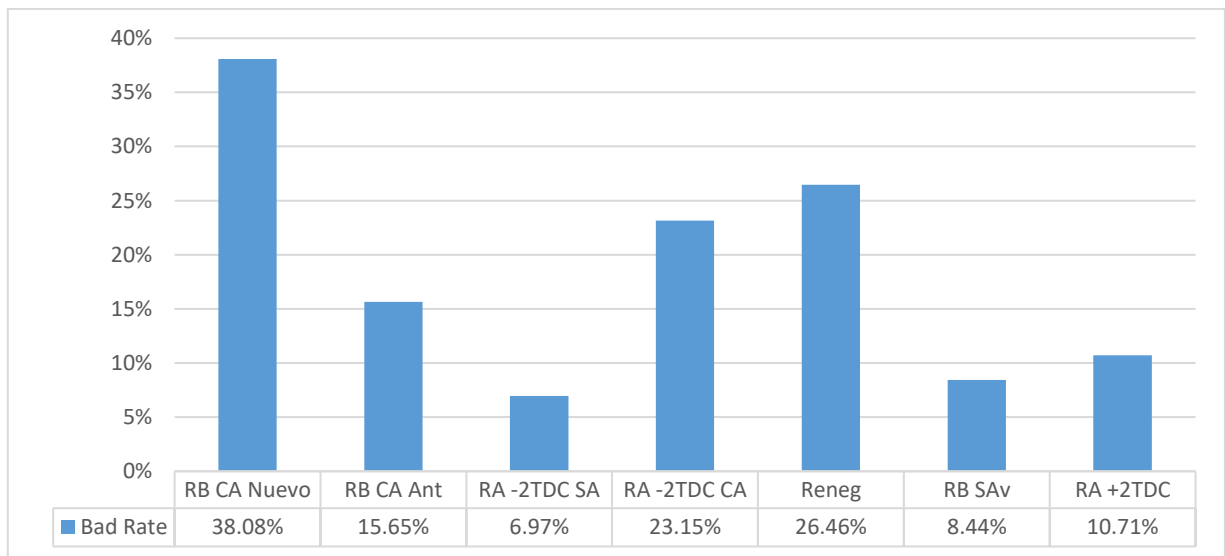


Gráfico 29 – Comparación *Bad Rate* de segmentación final. Propuesta N°3.

### 6.3.2 Modelamiento de propuesta de modelo n°3

De igual forma que para la propuesta anterior, para cada nuevo segmento originado se realizó el análisis univariado para las 1.446 variables. En la Tabla 39 se observan la cantidad de variables a incluir en cada modelo (detalle de cómo se llegó a estos números se encuentra en Anexo 7).

Segmento	Número de variables a incluir en modelos de propuesta N°3
<b>1</b>	15
<b>2</b>	12
<b>3</b>	13
<b>4</b>	15
<b>5</b>	12
<b>6</b>	13
<b>7</b>	15

Tabla 39 - Número de variables a incluir en modelo propuesta N°3

A continuación, se señalan las variables que fueron incluidas en el modelo de cada segmento generado en la propuesta N°3.

#### 1. Segmento 1

En la tabla 40, se encuentran las variables del modelo del segmento 1 de la propuesta N°3. Se observa que en este modelo se incluyen las variables "Suma deudas de línea de crédito, tarjeta de crédito y línea de emergencia que el cliente posee en el BCI" y "Tipo de productos que posee" es una variable nominal que señala cuáles son los productos que posee el cliente.

N°	Variable	Beta
1	Antigüedad_Cliente	-1.21
2	Antigüedad_LDC	-0.27
3	Cupo_Utilizado_TDC	-0.29
4	Deuda_Directa_Productos_Consumo_sist	-0.57
5	Evolucion_Monto_Credito_consumo_T0_T3	-0.95
6	Evolucion_Monto_IFRS_LDC_T0_T3	-0.39
7	Evolucion_Monto_Promedio_Saldos_Positivos_T0_T12	-0.40
8	Evolucion_Monto_Protesto_Historicos_T3_T12	-0.53
9	Maximo_Atraso_Deuda_Directa_Mes	-0.59
10	Monto_Total_Abona_en_CCT_mes	-0.79
11	Peor_Situacion_Deuda_Sistema_U6M_sist	-0.57
12	Ratio_Monto_Abonado_CCT_T3_T6	-0.89
13	Suma_Deuda_LDC_TDC_LEM_BCI	-0.90
14	Tipo_Producto_posee	-0.82
15	Total_de_Saldos_Disponibles	-0.37
16	Constante	-0.44

Tabla 40 - Variables y coeficientes incluidos en segmento 1 de propuesta de modelo N°3

## 2. Segmento 2

En la tabla 41, se encuentran las variables del modelo del segmento 2 de la propuesta N°3. Se observa que en este modelo se incluye la variable “suma deuda consumo con deuda comercial en el sistema”.

N°	Variable	Beta
1	Antigüedad_TDC	-1.08
2	Cupo_Utilizado_TDC	-0.54
3	Evolucion_Monto_Abonado_CCT_T0_T3	-0.77
4	Evolucion_Monto_Credito_consumo_T0_T3	-0.99
5	Evolucion_Monto_Disponible_CCT_T0_T12	-0.29
6	Evolucion_Monto_IFRS_LDC_T0_T12	-0.42
7	Maximo_Dias_Atraso_U3M	-0.66
8	Monto_Total_Abona_en_CCT_mes	-0.70
9	Numero_Protestos_Aclarados_U12M	-0.50
10	Peor_Situacion_Deuda_Sistema_U6M_sist	-0.50
11	Suma_Deuda_Consumo_Comercial_directa_sistema	-0.74
12	Total_Cupo_LDC	-0.52
13	Constante	-1.66

Tabla 41 - Variables y coeficientes incluidos en segmento 2 de propuesta de modelo N°3

## 3. Segmento 3

En la tabla 42, se encuentran las variables del modelo del segmento 3 de la propuesta N°3. Se observa que no se incluye ninguna variable no tradicional en este modelo.

N°	Variable	Beta
1	Antigüedad_CCT	-0.72
2	Deuda_Producto_consumo_Sbif_No_BCI_sist	-0.46
3	Evolucion_Monto_Consumo_T0_T6	-1.19
4	Evolucion_Monto_Disponible_CCT_T0_T12	-0.33
5	Evolucion_Monto_Disponible_LDC_mes_T0_T12	-0.56
6	Evolucion_Monto_IFRS_LDC_T6_T12	-0.58
7	Evolucion_Monto_Transferido_Lem_CCT_T0_T12	-0.50
8	Maximo_Dias_Atraso_U3M	-0.60
9	Monto_Promedio_Adeudado_U6M_U3M	-0.54
10	Monto_Total_Abona_en_CCT_mes	-0.74
11	Peor_Situacion_Deuda_Directa_U12M_sist	-0.37
12	Porcentaje_Uso_LDC	-0.14
13	Total_de_Saldos_Disponibles	-0.29
14	Constante	-2.39

Tabla 42 - Variables y coeficientes incluidos en segmento 3 de propuesta de modelo N°3



#### 4. Segmento 4

En la tabla 43, se encuentran las variables del modelo del segmento 4 de la propuesta N°3. Se observa que no se incluye ninguna variable no tradicional en este modelo, y que las variables que lo componen son las mismas del modelo actual replicado.

N°	Variable	Beta
1	Antigüedad_Ultima_TDC_Abierta	-0.60
2	Cantidad_TDC_vigentes	-0.49
3	Edad	-0.67
4	Evolucion_Monto_Disponible_CCT_T0_T12	-0.27
5	Evolucion_Monto_IFRS_LDC_T0_T3	-0.40
6	Evolucion_Monto_Mora_Credito_Consumo_T0_T12	-0.63
7	Maxima_Mora_TDC_U6M	-0.49
8	Monto_Promedio_Adeudado_U6M_U3M	-0.66
9	Monto_Total_Abona_en_CCT_mes	-0.56
10	Nivel_Endeudamiento_vigente	-0.61
11	Numero_dias_SGNP_U3M	-0.36
12	Peor_Situacion_Deuda_Directa_mes	-0.37
13	Peor_Situacion_Deuda_Directa_Sistema_mes	-0.36
14	Peor_Situacion_Deuda_Directa_U12M_sist	-0.26
15	Ratio_Monto_Protesto_T3_T12	-0.92
16	Constante	-0.97

Tabla 43 - Variables y coeficientes incluidos en segmento 4 de propuesta de modelo N°3

#### 5. Segmento 5

En la tabla 44, se encuentran las variables del modelo del segmento 5 de la propuesta N°3. Se observa que no se incluye ninguna variable no tradicional en este modelo.

N°	Variable	Beta
1	Antigüedad_CCT	-0.81
2	CCT_RatioIngMtoAbonoTransLemPRT6_T12	-0.37
3	Deuda_Directa_Productos_Consumo_sist	-0.37
4	Evolucion_Monto_Credito_consumo_T0_T3	-0.86
5	Evolucion_Monto_Disponible_CCT_T0_T12	-0.33
6	Evolucion_Monto_Disponible_LDC_mes_T0_T12	-0.51
7	Maximo_Dias_atraso_U6M	-0.53
8	Monto_Promedio_Adeudado_U6M_U3M	-0.50
9	Monto_Total_Abonado_en_CCT_U3M	-0.71
10	Peor_Situacion_Deuda_Directa_U12M_sist	-0.43
11	Saldo_Minimo_CCT_Activas_Mes	-0.26
12	Total_de_Saldos_Disponibles	-0.28
13	Constante	-2.55

Tabla 44 - Variables y coeficientes incluidos en segmento 5 de propuesta de modelo N°3

## 6. Segmento 6

En la tabla 45, se encuentran las variables del modelo del segmento 6 de la propuesta N°3. Se observa que no se incluye ninguna variable no tradicional en este modelo.

N°	Variable	Beta
1	Antigüedad_CCT	-0.83
2	Cupo_Utilizado_TDC	-0.51
3	Deuda_Producto_consumo_Sbif_No_BCI_sist	-0.61
4	Evolucion_Monto_Abonado_Lem_CCT_T3_T6	-0.52
5	Evolucion_Monto_Abonado_T6_T12	-0.75
6	Evolucion_Monto_Disponible_LDC_mes_T0_T12	-0.44
7	Maximo_Atraso_Deuda_Directa_Mes	-0.50
8	Maximo_Dias_Atraso_Consumo_U6M	-0.34
9	Monto_Total_Abona_en_CCT_mes	-0.59
10	Peor_Situacion_Deuda_Sistema_U6M_sist	-0.66
11	Ratio_Monto_Disponible_CCT_T6_T12	-0.38
12	Total_Cupo_LDC	-0.59
13	Total_de_Saldos_IFRS	-0.88
14	Constante	-1.18

Tabla 45 - Variables y coeficientes incluidos en segmento 6 de propuesta de modelo N°3

## 7. Segmento 7

En la tabla 46, se encuentran las variables del modelo del segmento 7 de la propuesta N°3. Se observa que en este modelo se incluye la variable "Suma deudas de línea de crédito, tarjeta de crédito y línea de emergencia que el cliente posee en el BCI".

N°	Variable	Beta
1	Antigüedad_CCT	-0.59
2	Cantidad_LDC_vigentes	-1.79
3	Deuda_Directa_Productos_Consumo_sist	-0.33
4	Evolucion_Monto_Abonado_T6_T12	-0.99
5	Evolucion_Monto_Credito_consumo_T0_T3	-0.50
6	Evolucion_Monto_Promedio_Saldos_Positivos_T0_T12	-0.53
7	Evolucion_Monto_Transferido_Lem_CCT_T0_T12	-0.24
8	Maximo_Dias_atraso_U6M	-0.50
9	Monto_Total_Abona_en_CCT_mes	-0.83
10	Numero_Acciones	-0.41
11	Numero_Avances_Solicitados_U12M	-0.31
12	Peor_Situacion_Deuda_Directa_U12M_sist	-0.45
13	Porcentaje_Uso_LDC	-0.38
14	Suma_Deuda_LDC_TDC_LEM_BCI	-0.65

15	Total_de_Saldos_Disponibles	-0.36
16	Constante	-3.32

Tabla 46 - Variables y coeficientes incluidos en segmento 7 de propuesta de modelo N°3

### 6.3.3 Medición de desempeño de propuesta de modelo n°3

En esta etapa, se midió el desempeño del modelo de cada segmento y de la cartera de consumo total. Las métricas que se utilizaron son el estadístico KS y la curva de ROC. Los resultados obtenidos para cada métrica se encuentran en el gráfico 30.

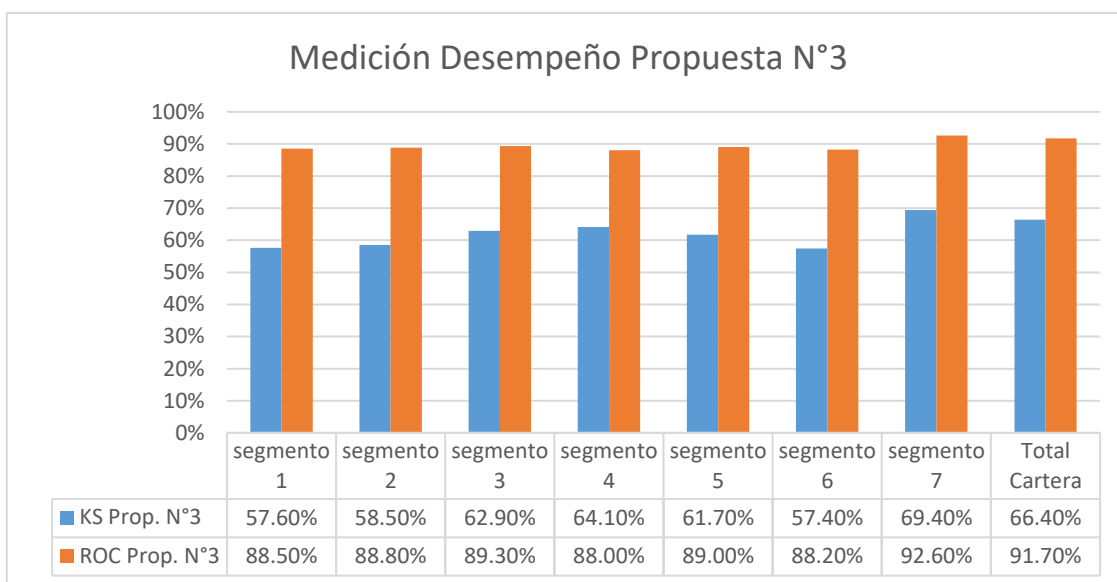


Gráfico 30 - Medición de desempeño en KS y ROC de propuesta de modelo N°3

Se aprecia que el desempeño KS de los segmentos 1, 2 y 6 son relativamente bajos, en comparación a los resultados que se ha obtenido en los modelos anteriormente propuestos. Sin embargo, el desempeño en la cartera de consumo, tanto en KS como en ROC, se encuentra en niveles altos y similares a los de propuestas anteriores.

## Capítulo VII. Comparación y elección de modelo

Como se mencionó, para la elección del modelo que se propondrá, se utilizará una combinación de criterios. Estos son, métricas de desempeño KS y ROC, complementado con un análisis de impacto en las provisiones que tendría cada modelo y un *backtest*, donde se compararan las tasas de incumplimiento reales con las predichas por los modelos.

En primer lugar, se comparó el desempeño de la propuesta de modelo n°1 respecto al modelo actual del banco replicado, medidos en KS y ROC.

Como la primera propuesta de modelo contempla la misma segmentación que posee el banco actualmente, se comparó el desempeño en los 7 segmentos y de la cartera total. Lo anterior se aprecia en el gráfico 31.

En cuanto a la métrica KS, se aprecia que los segmentos 1, 3 y 5 de la propuesta de modelo n°1 tiene un mejor desempeño, pues sus estadísticos KS son mayores. Sin embargo, en los segmentos 2, 6 y 7 tiene un mayor poder de discriminación el modelo actual replicado. Mientras que el segmento 4, posee el mismo desempeño. Al momento de evaluar toda la cartera, posee un mayor desempeño la propuesta n°1.

Respecto a la métrica ROC, la propuesta N°1 tiene un desempeño mayor en los segmentos 1, 2, 3 y 7. La réplica solo es superior en el segmento 6. Mientras que en el segmento 4 y 5 se obtuvo el mismo desempeño. Al comparar la cartera total, la propuesta 1 es levemente superior a la modelo actual replicado.

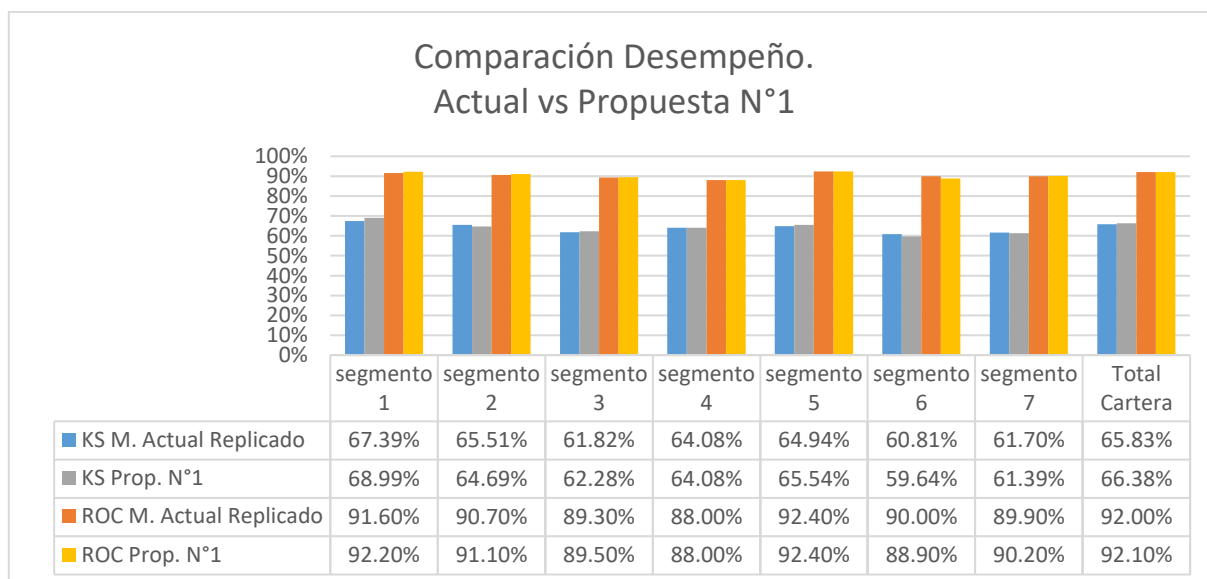


Gráfico 31 – Comparación de modelo actual replicado con propuesta N°1, mediante KS y ROC.

Para realizar una comparación de la propuesta de modelo n°2, se llevó a cabo de la siguiente forma, se tomó el resultado de la métrica KS y ROC de los segmentos en cuestión del modelo actual replicado. Estos valores se compararon con el resultado de medición de desempeño de forma agregada de los segmentos generados (medir KS para segmento 1.1 y 1.2 de forma conjunta, lo mismo para los demás segmentos originados).

En el gráfico 32 se comparan las métricas KS y ROC de los segmentos en cuestión. En cuanto al KS, se observa que la propuesta N°2 posee un mejor desempeño. Mientras que en la curva ROC, la propuesta N°2 es mejor en los segmentos 1 y 3, en el segmento 5 es superior el modelo actual replicado.

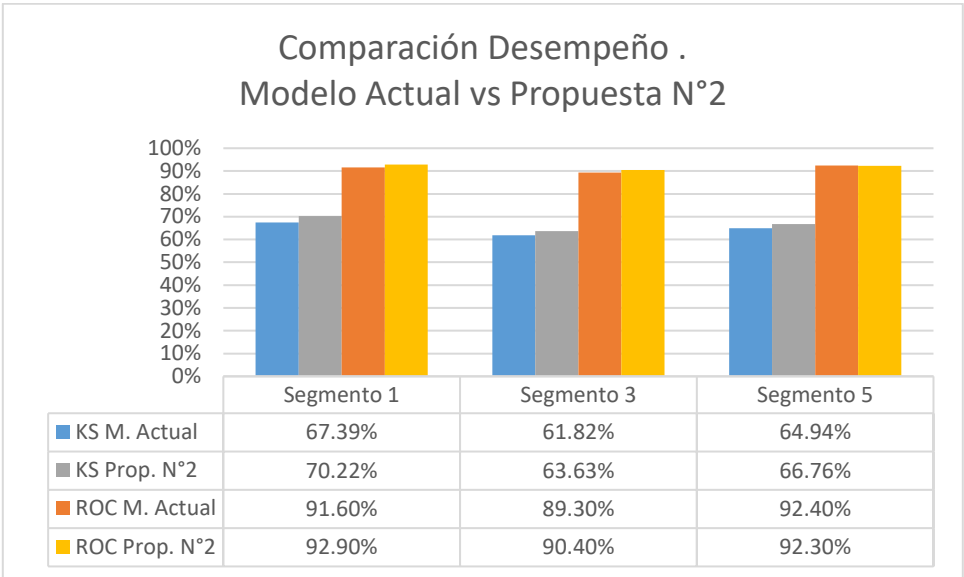


Gráfico 32 Comparación de modelo actual replicado con propuesta N°2, mediante KS y ROC.

Dado el resultado de comparación de propuesta 2, se podría decir que tiene sentido segmentar los segmentos 1, 3 y 5 del modelo actual del banco. Esto dado que las métricas obtenidas para la propuesta 2, para KS y ROC, son superiores que los resultados del modelo replicado (salvo ROC en segmento 5). Sin embargo, al realizar un análisis del comportamiento del *Bad Rate*, reemplazando los segmentos 1, 3 y 5, por 1.1, 1.2, 3.1, 3.2, 5.1 y 5.2 de la propuesta 2, se aprecia que existe un alto entrecruzamiento de *Bad Rate* en el tiempo (ver gráfico 33). Por el motivo anterior, se descarta la originación de los segmentos y con ello se descarta la propuesta N°2 como elección final de modelo.

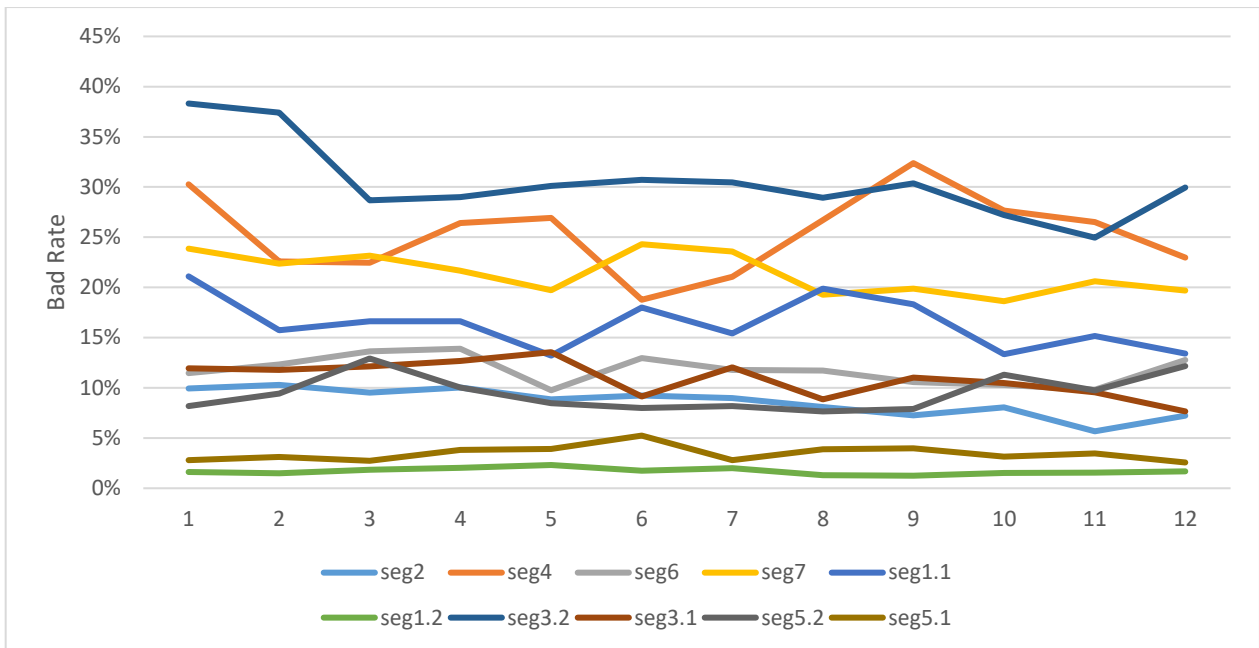


Gráfico 33 - Evolución temporal de Bad Rate se propuesta n°2.

Finalmente, se comparó la propuesta de modelo n°3 con la réplica. En esta ocasión solo se comparan los estadísticos KS y ROC del modelo agregado, pues la segmentación es diferente entre modelos. En el gráfico 34, se observa que en KS tiene mejor desempeño la propuesta N°3, pero en ROC es superior el modelo actual replicado.

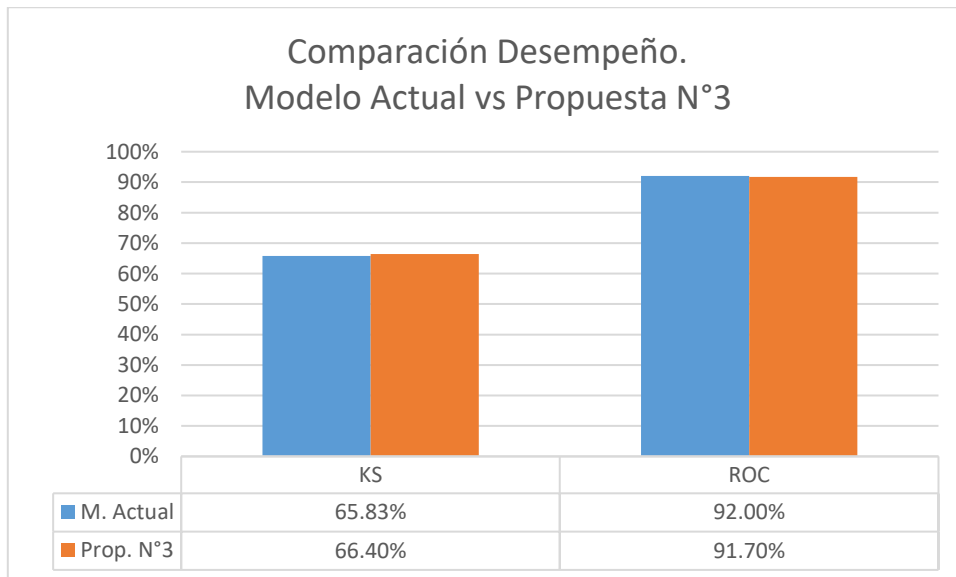


Gráfico 34 - Comparación de modelo actual replicado con propuesta N°3, mediante KS y ROC.

Como se mencionó, la comparación también se realizó en el impacto en las provisiones. La variable que se utilizó es deuda del cliente. Cabe resaltar que esta variable es la deuda del cliente en consumo dividida por su renta líquida. Por este motivo el resultado que se obtiene es solo una estimación, ya que existen clientes que no poseen renta y no se pudo realizar el ejercicio de multiplicar deuda por renta, y así obtener la deuda del cliente. Además, no se cuenta con el valor de la pérdida dado el incumplimiento, que es esencial para el cálculo de las provisiones, sin embargo, para los clientes no cambia dicho valor, sino que lo que varía es la probabilidad de incumplimiento que entrega cada propuesta de modelo.

El cálculo que se realizó es el siguiente:

$$\text{Estimación provisión modelo} = \sum_i \text{deuda}_i \cdot PI_i$$

Donde  $i$  es cada cliente de la cartera;  $\text{deuda}_i$  es la deuda del clientes  $i$ ;  
 $PI_i$  es la probabilidad de incumplimiento estimada por modelo

	<b>Estimación provisión bruta</b>
<b>Modelo actual replicado</b>	261.562.052.199
<b>Propuesta N°1</b>	259.021,569.187
<b>Propuesta N°3</b>	257.468.360.883

*Tabla 47- Estimación de provisión bruta de cada modelo.*

Se aprecia en la tabla 47 que la estimación es mayor para la réplica. Le sigue la propuesta 1, que es un 0.97% menor que el modelo actual replicado. Finalmente, el modelo que estima una menor provisión es la propuesta N°3, cuyo valor es 1,57% menor que la estimación del modelo actual replicado.

Otra medida utilizada para comparar los modelos fue el backtest, donde se comparan las tasas de incumplimiento reales con las que los modelos predicen.

- Backtest propuesta N°1

En el gráfico 35, se observa que la propuesta N°1 subestima las tasas de incumplimiento para todos los periodos. Los gráficos de comparación de *Bad Rate* reales versus los predicho por la propuesta N°1 por cada segmento se encuentran en Anexo 8.

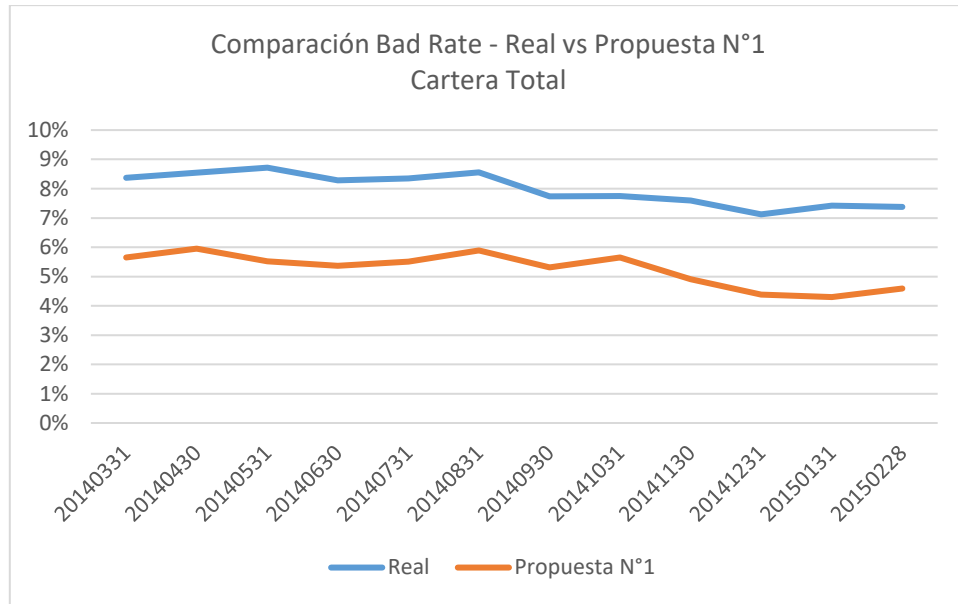


Gráfico 35 – Comparación *Bad Rate* Real vs propuesta N°1. Cartera total.

- Backtest propuesta N°3

En el gráfico 36, al igual que el modelo anterior, se observa que la propuesta N°3 subestima las tasas de incumplimiento para todos los periodos. Los gráficos de comparación de *Bad Rate* reales versus los predicho por la propuesta N°3 por cada segmento se encuentran en Anexo 9.



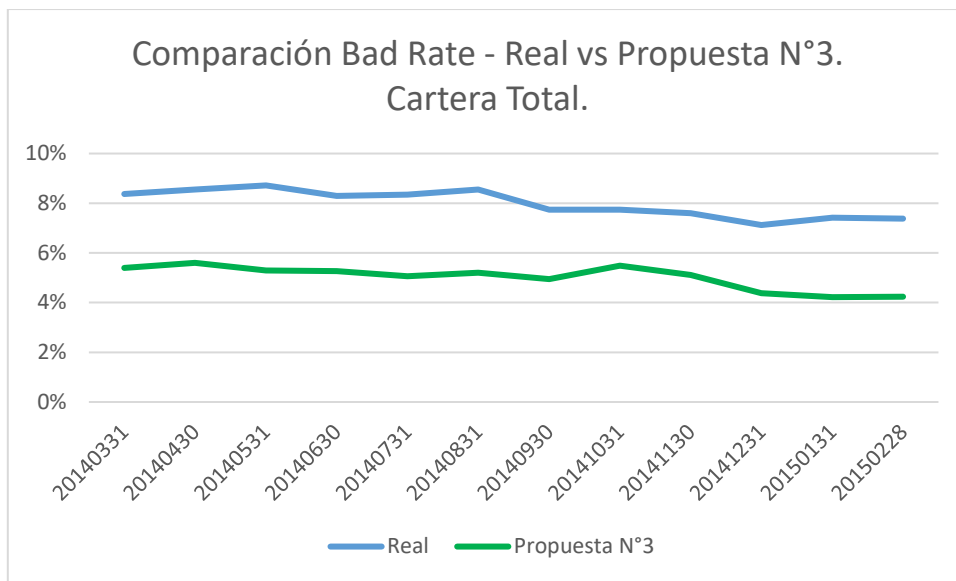


Gráfico 36 - Comparación Bad Rate Real vs propuesta N°3. Cartera total.

Para finalizar, se realizó la tabla resumen, donde se muestra que modelo es mejor en los primero dos criterios.

<b>Criterio</b>	<b>Modelo</b>
<b>Estadístico KS</b>	Propuesta 1 o 3
<b>Estadístico ROC</b>	Propuesta 1
<b>Menor estimación de provisión</b>	Propuesta 3

Tabla 48 - Mejor modelo dependiendo del criterio escogido.

Como el objetivo de la memoria es proponer un modelo con mayor capacidad predictiva, el modelo escogido es la propuesta n°1, pues es la que mayor capacidad predictiva posee, tanto en KS como en ROC. Además, la diferencia en estimación de provisión es similar a la de la propuesta 3.

## **Capítulo VIII. Conclusiones y recomendaciones**

El objetivo de este trabajo fue proponer un modelo para la cartera de consumo del banco BCI, que posea una mayor capacidad predictiva que el modelo actual que posee el banco, incorporando variables no tradicionales. Este objetivo se da por cumplido, pues se propone un modelo para la cartera de consumo del banco BCI, el cual posee un desempeño 66,38%, medido en KS, mayor a lo informado por el banco y la del modelo actual replicado. Además, posee un área bajo la curva ROC del 92,1%, esto es un resultado sumamente bueno. Pues en la literatura, modelos con ROC superior a 75% son considerados buenos modelos, este resultado no pudo ser contrastado con información del banco.

También se da por cumplido los objetivos específicos planteados. Pues se comprendió el modelo que actualmente posee el banco. Esto fue realizado al momento de generar la réplica del modelo actual del banco. Si bien no se obtuvo los mismos resultados, se llegó a métricas similares de desempeño, con un porcentaje de error del 0,03% en la métrica KS.

Respecto al segundo objetivo específico, se logró elaborar 182 variables nuevas, las cuales fueron incorporadas para la elaboración de los modelos propuestos. Estas variables creadas tienen un carácter innovador, pues son originadas a partir de la interacción de los productos que puede poseer un cliente, tanto en el mismo banco BCI o en alguna otra institución financiera.

Luego de elaborar 3 propuestas de modelos, se seleccionó la propuesta n°1. Con esta elección, se concluye que la segmentación actual del banco es óptima, considerando las variables que se utilizó en esta memoria. Esto no quiere decir, que en un futuro la segmentación pueda ser cambiada por otra mejor, si es que se incorporan nuevas variables de información de los clientes.

Respecto a las variables que componen el modelo que se seleccionó, en este se encuentran incluidas variables construidas en este trabajo, tales como: multiplicación de deuda de consumo y deuda comercial en el sistema; productos que posee el cliente; suma de cupo que posee en algunos productos; suma de deudas en los productos que posee en el banco o en el sistema (como deuda de línea de crédito, línea de emergencia, sobregiros y/o tarjeta de créditos).

Uno de los aprendizajes que dejó esta memoria, es que, si se estudian adecuadamente los datos, siempre se encontrará información útil. En este caso se encontraron variables que aportaron información valiosa a la etapa de modelamiento. Es así como se obtuvo un modelo con mejor desempeño.

Como se mencionó, algunos de los modelos que se proponen contienen variables que son la suma o multiplicación de deudas en el BCI o SBIF, cupos en los productos que tiene en el BCI o SBIF, esto quiere decir, que poseer más de un producto y las deudas y cupos a dichos productos, sirven para modelar diferentes comportamientos de incumplimiento de los clientes de la cartera de consumo.

No obstante, este trabajo de título no estuvo exento de dificultades. La principal problemática que se enfrentó fue, los tiempos de procesamiento computacional de los datos. El motivo de lo anterior fue que las bases de datos con las que se trabaja en este tipo de modelos son muy grandes.

Además, en un principio se buscó trabajar en el programa computacional R, pues se poseía conocimientos previos de este. Sin embargo, por el tamaño de las bases de datos, el programa tardaba mucho tiempo en solo abrir estas bases. Por este motivo, se realizó un cambio en el programa. El segundo, y definitivo, programa utilizado fue SPSS modeler. En un principio las complicaciones vinieron por el desconocimiento de dicha herramienta, pero luego de adquirir experiencia en el programa todo fluyó de forma más expedita y se pudo desarrollar los modelos que se pretendía.

La recomendación que se hace es, continuar probando nuevas variables no tradicionales, esto podría mejorar aún más los resultados del modelo de probabilidad de incumplimiento. Además de probar nuevas técnicas de modelamiento, y no solo modelos logit. Esto no pudo ser realizado en el trabajo de título, dado lo acotado del tiempo para la realización.

## Bibliografía

1. Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.
2. Siddiqi, N. (2012). *Credit risk scorecards: developing and implementing intelligent credit scoring* (Vol. 3). John Wiley & Sons.
3. Thomás, L. C. (2009). *Consumer credit models: pricing, profit and portfolios: pricing, profit and portfolios*. OUP Oxford.
4. BCI, Memoria anual – 2016. [en línea]. <https://www.bci.cl/investor-relations/memoria-anual/files/memoria-anual-2016> [consulta: 04 de mayo de 2017]
5. SBIF, Superintendencia de Bancos e Instituciones Financieras. [en línea]. [www.sbif.cl](http://www.sbif.cl) [consulta: 04 de mayo de 2017]
6. Fayyad, U (1996). *Data mining and knowledge discovery- making sense out of data. IEEE Expert-Intelligent Systems and Their Applications*.

## **Anexos**

### **Anexo 1: Información de Bases de datos proporcionadas por el banco**

- Cientes  
En esta base de datos se cuenta con 126 variables. Algunas de ellas son:
  - Demográficas: edad, sexo, nivel educacional, tipo de vivienda, estado civil, entre otros.
  - Antigüedad del cliente: en cartera de consumo, cuenta corriente, tarjeta de crédito, crédito de consumo y último protesto.
  - Cantidad vigente de: cuentas corrientes, créditos de consumo en cuotas, tarjetas de créditos, líneas de crédito y líneas de emergencia.
  - Monto de deuda, tanto BCI como SBIF: consumo, comercial e hipotecaria.
  - Días de mora en cartera: comercial, consumo e hipotecario.
  
- Consumo  
Esta base de datos tiene 42 variables. Estas poseen información asociada a créditos de consumo en cuotas, tales como:
  - Antigüedad desde que se genera el crédito más antiguo.
  - Apalancamiento.
  - Ratios de montos de mora.
  - Si posee o no multicréditos.
  - Porcentaje de cuotas pagas.
  
- Cuenta corriente  
Esta base de datos cuenta con 381 variables. Entre las que se encuentran:
  - Meses desde: último uso, cierre y apertura
  - Egresos: información de ratios temporales de máximos, promedios y total. Esto para compras con tarjeta de débito, giros en ATM, giros en caja, pagos automáticos a productos BCI o servicios, pago líneas de emergencia, pago de línea sobregiro y transferencias.
  - -Ingresos: ratios temporales de máximos, promedios y total. Esto para abonos, transferencias, depósitos, traspasos de línea de emergencia y línea de sobregiro.
  - Ratios temporales de máximos, promedios y totales de montos disponibles a fin de mes, montos de egreso, montos de ingresos.
  
- Tarjeta de crédito  
En esta base de datos se encuentran 10 variables. Estas son:
  - Meses de apertura de la TDC más antigua.
  - Meses desde el último aumento de cupo.
  - Meses desde última disminución de cupo.

- Máxima mora en TDC (mes, 3 meses, 6 meses y 12 meses).
- Si paga más del mínimo, pero menos que total.
- Si paga el mínimo.
- Saldo total.
  
- Línea de crédito  
Aquí se encuentran 291 variables, tales como:
  - Antigüedad línea, meses hasta que vence, meses desde último uso.
  - Ratios temporales:
    - Días que uso línea.
    - Pagos de línea de crédito.
  - Montos disponibles.
  - Deuda efectiva.
  - Montos morosos.
  - Cupo de línea.
  - Montos utilizados.
  
- Facturación tarjeta de crédito  
Esta base de datos posee 303 variables, tales como:
  - % cupo nacional utilizado.
  - Ratios temporales:
    - Montos avances nacionales en cuotas.
    - Compras internacionales.
    - Compras nacionales.
    - Montos pagados (nacionales e internacionales).
  - Cantidad de avances nacionales.
  - Cantidad de avances internacionales.
  
- Inversiones  
Aquí se encuentran 108 variables. Ejemplos de estas son:
  - Ratios temporales:
    - Ahorro
    - Depósito a plazo
    - Fondos mutuos
    - Acciones
  - Montos del mes, 3 meses, 6 meses y 12 meses:
    - Ahorro
    - Depósito a plazo
    - Fondos mutuos
    - Acciones
  - Cantidad: acciones, operaciones DAP, operaciones de ahorro, operaciones de fondos mutuos

## Anexo 2: Detalle de pasos para modelamiento de modelo actual replicado.

En primer lugar, para cada segmento se realizó el análisis univariado 1.262 variables, provenientes de las bases de datos otorgadas por el banco. A continuación, se observa el número de variables tramificadas para cada uno de los segmentos.

Segmento	Número de variables tramificadas
1	973
2	788
3	872
4	1.150
5	931
6	918
7	940

Con las variables ya tramificadas, se procedió con cálculo del valor WOE para cada tramo de cada variable. Luego, para cada variable se calculó su *Information Value*. Posteriormente, para cada segmento se elaboró el ranking de *Information Value*.

Del ranking construido, se descartó las variables que posean un IV menor a 2, pues significa que su poder de discriminancia es baja. En la siguiente tabla se aprecia el número de variables para cada segmento luego de la eliminación por IV.

Segmento	Número de variables posterior a eliminación
1	973
2	759
3	840
4	1.086
5	921
6	851
7	846

El siguiente paso que se realizó fue el análisis de correlaciones, con el fin de eliminar las variables que sean redundantes para el modelo.

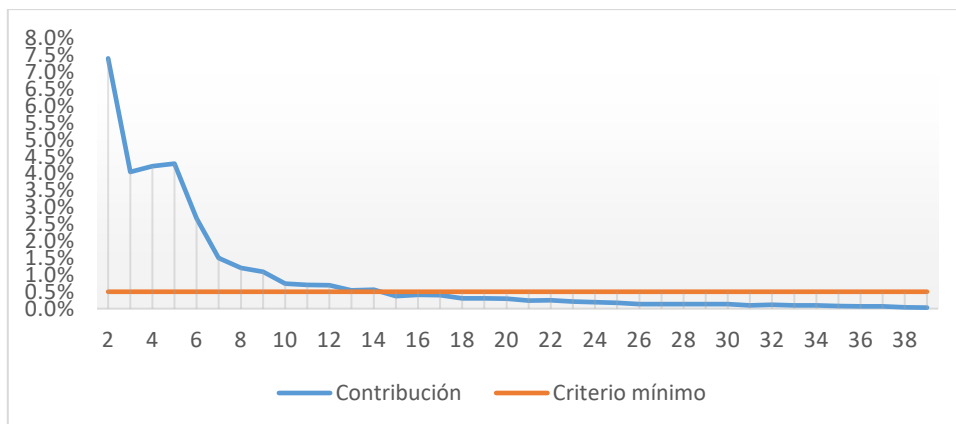
Segmento	Número de variables posterior a análisis de correlación
1	51
2	68
3	72
4	64
5	65
6	88
7	92

Luego, se realizó la regresión logística para cada segmento, con la técnica y parámetros descritos en capítulo 3.5.3. El número de variables que se obtuvo como resultado para cada segmento se aprecia en la siguiente tabla.

Segmento	Número de variables a incluir según regresión logística
1	39
2	55
3	58
4	30
5	43
6	47
7	62

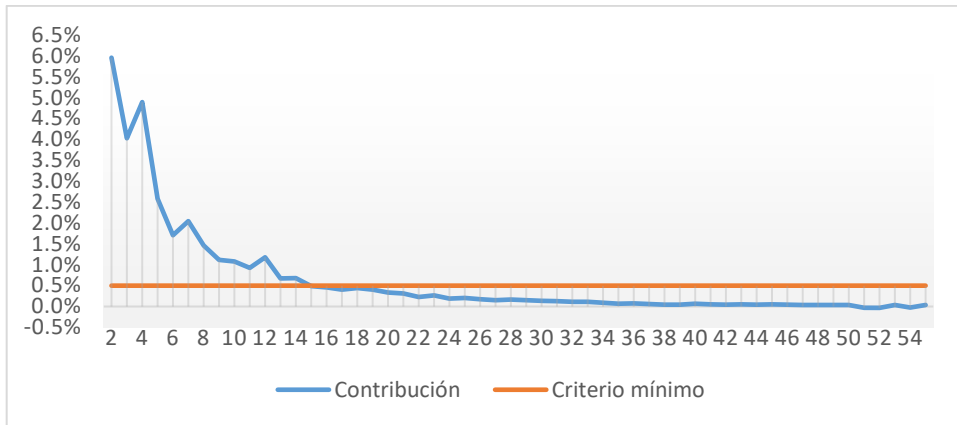
Se observa en la tabla anterior, es una gran cantidad de variables para cada segmento, por lo cual sería un modelo difícil de comprender e interpretar. Para reducir las variables se realizó un análisis de la contribución a la logverosimilitud, como se explica en el capítulo 3.5.4. El resultado del análisis determinó la cantidad de variables a incorporar para el modelo de cada segmento, los gráficos de contribución a la logverosimilitud se observan a continuación.

- Segmento 1

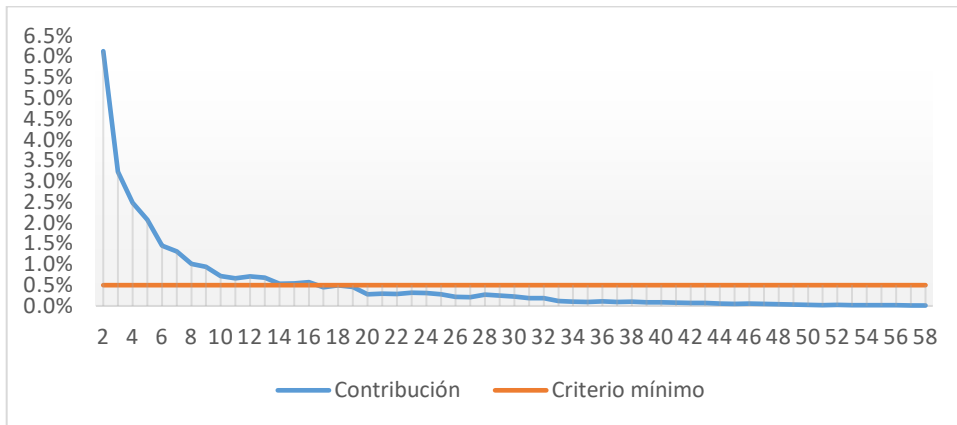




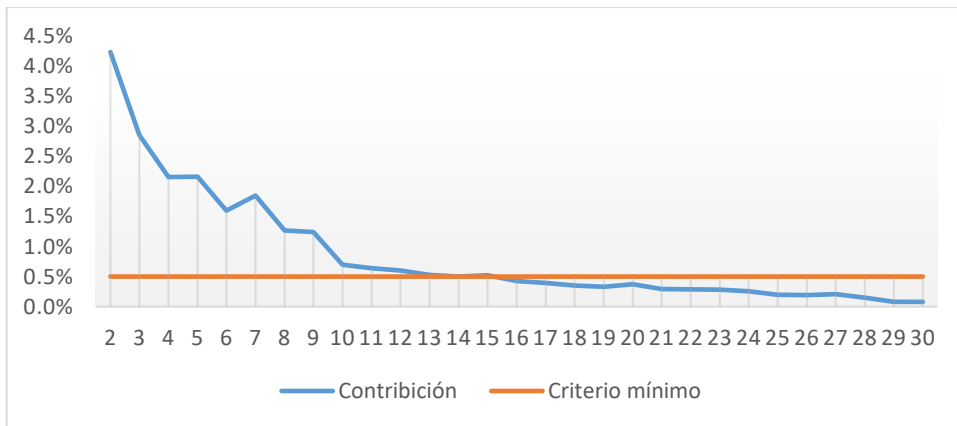
- Segmento 2



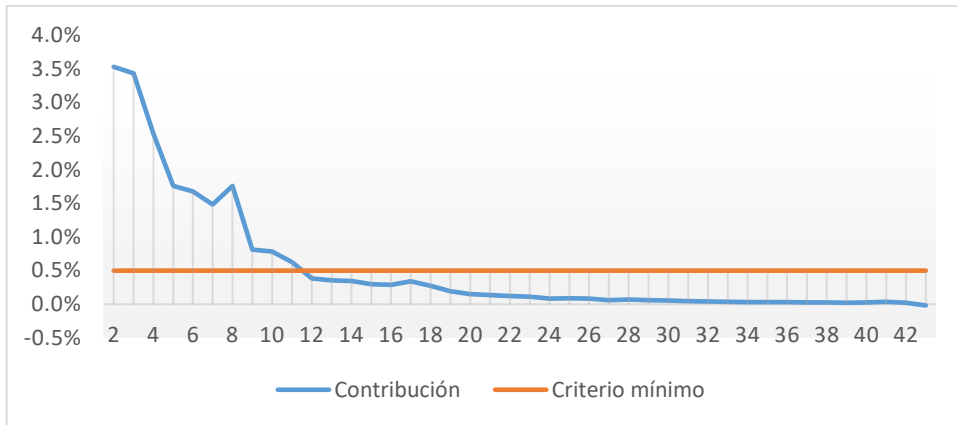
- Segmento 3



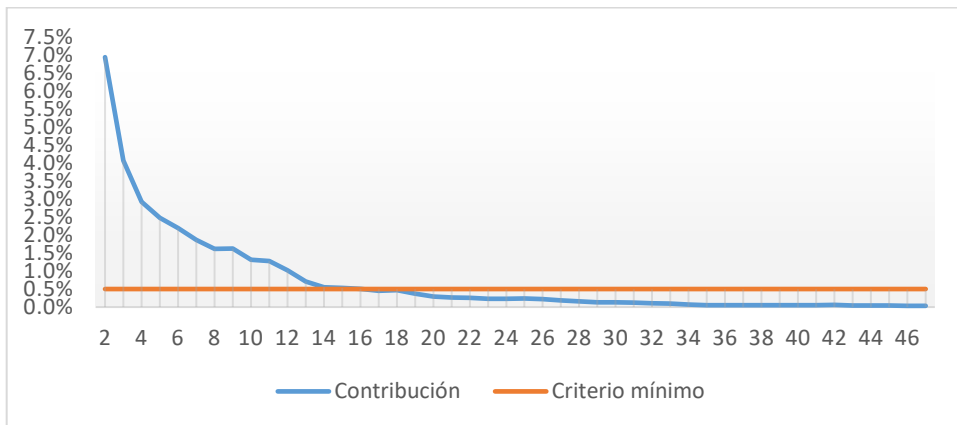
- Segmento 4



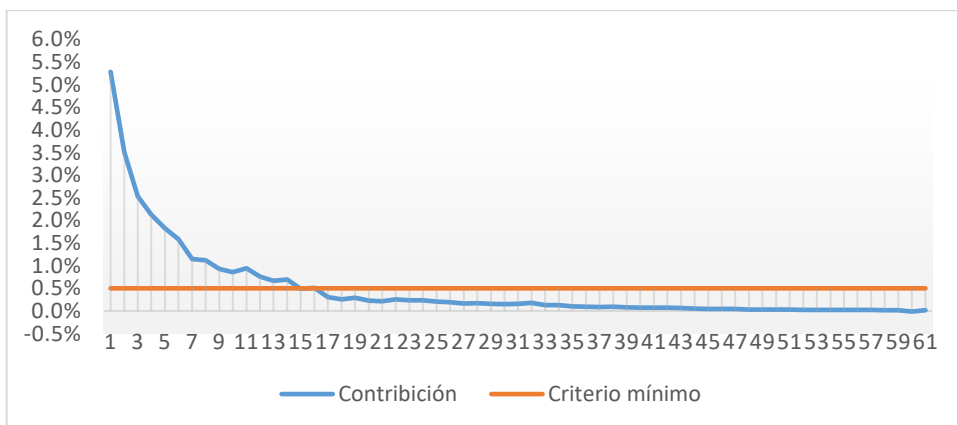
- Segmento 5



- Segmento 6



- Segmento 7



## Anexo 3: Variables seleccionadas para generar los segmentos

Variable	Descripción
CCT_AntiguedadCierre	Diferencia en meses entre la fecha de cierre de la cuenta corriente y la fecha actual
CCT_AntiguedadALTUSO	Diferencia en meses entre la fecha actual y la fecha del último uso de la cuenta corriente
CCT_AntiguedadCtaCte	Antigüedad de la cuenta corriente en meses
FACT_Numero_Avance_NAC	Número total de numero de avance nacional en el mes actual
FACT_Numero_Avance_INT	Número de avance internacional en el mes actual
FACT_Numero_Avance_NAC_SINCUO	Número de avance nacional sin cuotas en el mes
FACT_Numero_Avance_NAC_3CUO	Número de avances nacionales en 3 cuotas en el mes
FACT_Numero_Avance_NAC_6CUO	Número de avances nacionales en 6 cuotas en el mes
FACT_Numero_Avance_NAC_12CUO	Número de avances nacionales en 12 cuotas en el mes
FACT_Numero_Avance_NAC_24CUO	Número de avances nacionales en 24 cuotas en el mes
FACT_Numero_Avance_NAC_MÁS24CUO	Número de avances nacionales en más de 24 cuotas en el mes
FACT_Numero_Avance_NAC_U3M	Número total de avances nacionales de los últimos 3 meses
FACT_Numero_Avance_INT_U3M	Número de avances internacionales de los últimos 3 meses
FACT_Numero_Avance_NAC_SINCUO_U3M	Número de avances nacionales sin cuota de los últimos 3 meses
FACT_Numero_Avance_NAC_3CUO_U3M	Número de avances nacionales en 3 cuotas de los últimos 3 meses
FACT_Numero_Avance_NAC_6CUO_U3M	Número de avances nacionales en 6 cuotas de los últimos 3 meses
FACT_Numero_Avance_NAC_12CUO_U3M	Número de avances nacionales en 12 cuotas de los últimos 3 meses
FACT_Numero_Avance_NAC_24CUO_U3M	Número de avances nacionales en 24 cuotas de los últimos 3 meses
FACT_Numero_Avance_NAC_MÁS24CUO_U3M	Número de avances nacionales en más de 24 cuotas de los últimos 3 meses
FACT_Numero_Avance_NAC_U6M	Número total de avances nacionales de los últimos 6 meses
FACT_Numero_Avance_INT_U6M	Número total de avances internacionales de los últimos 6 meses
FACT_Numero_Avance_NAC_SINCUO_U6M	Número de avances nacionales sin cuota de los últimos 6 meses
FACT_Numero_Avance_NAC_3CUO_U6M	Número de avances nacionales en 3 cuotas de los últimos 6 meses
FACT_Numero_Avance_NAC_6CUO_U6M	Número de avances nacionales en 6 cuotas de los últimos 6 meses
FACT_Numero_Avance_NAC_12CUO_U6M	Número de avances nacionales en 12 cuotas de los últimos 6 meses
FACT_Numero_Avance_NAC_24CUO_U6M	Número de avances nacionales en 24 cuotas de los últimos 6 meses
FACT_Numero_Avance_NAC_MÁS24CUO_U6M	Número de avances nacionales en más de 24 cuotas de los últimos 6 meses
FACT_Numero_Avance_NAC_U12M	Número de avances nacionales realizados en los últimos 12 Meses
FACT_Numero_Avance_INT_U12M	Número total de avances internacionales de los últimos 12 meses
FACT_Numero_Avance_NAC_SINCUO_U12M	Número de avances nacionales sin cuota de los últimos 12 meses
FACT_Numero_Avance_NAC_3CUO_U12M	Número de avances nacionales en 3 cuotas de los últimos 12 meses
FACT_Numero_Avance_NAC_6CUO_U12M	Número de avances nacionales en 6 cuotas de los últimos 12 meses
FACT_Numero_Avance_NAC_12CUO_U12M	Número de avances nacionales en 12 cuotas de los últimos 12 meses
FACT_Numero_Avance_NAC_24CUO_U12M	numero de avances nacionales en 24 cuotas de los últimos 12 meses
FACT_Numero_Avance_NAC_MÁS24CUO_U12M	numero de avances nacionales en más de 24 cuotas de los últimos 12 meses
LDC_AntiguedadULTUSO	Antigüedad en meses desde la última vez que se usa la línea de crédito
LDC_AntiguedadLínea	Antigüedad de la línea de crédito
CL_Ant_Curse_Última_Ren_CON	Antigüedad de última Renegociación CON vigente del Cliente en meses.
CL_Ant_Ult_Act_Cliente_en_dia	Antigüedad de la última actividad de consumo del cliente.
CL_Ant_Ult_CCT_Abier_meses	Antigüedad última cuenta corriente BCI abierta en meses.
CL_Ant_Ult_Protesto_Vig_sist	Antigüedad del último protesto vigente en el sistema en días
CL_Ant_Ult_TDC_Abier_meses	Antigüedad última tarjeta de crédito BCI abierta en meses.
CL_Antiguedad_Cliente_en_CCT	Antigüedad primera CCT BCI en meses.
CL_Antiguedad_Cliente_meses	Antigüedad cliente en Bci en meses. Se refiere a la antigüedad del primer producto de consumo tomado por el cliente (CCT-TDC-CON).
CL_Cantidad_CCT_vigentes	Cantidad CCT BCI vigentes a fin de mes
CL_Cantidad_CON_vigentes	Cantidad de créditos de consumo BCI vigentes a fin de mes.
CL_Cantidad_LDC_vigentes	Cantidad de líneas de crédito BCI vigentes a fin de mes.
CL_Cantidad_LEM_vigentes	Cantidad de líneas de emergencias BCI vigentes a fin de mes.
CL_Cantidad_TDC_vigentes	Cantidad de tarjetas de crédito BCI vigentes a fin de mes
CL_Edad	Edad del cliente en años
CL_Estado_Civil	Estado civil del cliente
CL_Meses_ant_TDC_más_antigua	Meses de antigüedad de la primera tarjeta de crédito abierta.
CL_Mto_Línea_Dispon_SBIF_Ult_Per	Monto de línea disponible SBIF a fecha de proceso menos 2 meses dividida por la renta líquida en UF.
CL_Nivel_Educacional	Nivel educacional del cliente.
CL_Nivel_Endeud_vigente_CP	Nivel de endeudamiento vigente corto plazo en BCI al cierre de mes. Básicamente es la deuda de consumo, comercial e hipotecaria del cliente (deuda total del cliente).
CL_NProt_Flta_Fdos_S_Acla_sist	Número de Protestos por falta de fondos sin aclarar en el Sistema en el mes. Solo Cheques
CL_NProt_FltaFdosAcla_mes_sist	Número de Protestos por falta de fondos aclarados en el Sistema en el mes.
CL_NTot_Prot_Flta_Fon_Acla_mes	Cantidad de protestos por falta de fondos ACLARADOS durante el mes. Solo Cheques Bci.
CL_NTot_Prot_por_Falta_Fon_Vig	Acumula número de protestos vigentes por falta de fondos a través de todas sus cuentas corrientes en el mes BCI. Solo Cheques Bci.
CL_Num_Prot_Acla_U12M_sist	Número de Protestos por falta de fondos aclarados en el Sistema U12M.
CL_Num_Protestos_Acla_U3M_sist	Número de Protestos por falta de fondos aclarados en el Sistema U3M. Solo Cheques
CL_Num_Protestos_Acla_U6M_sist	Número de Protestos por falta de fondos aclarados en el Sistema U6M. Solo Cheques
CL_Num_Tot_Protestos_Acla_U3M	Cantidad de protestos por falta de fondos ACLARADOS durante los U3M. Solo Cheques Bci.
CL_Num_Tot_Protestos_Acla_U6M	Cantidad de protestos por falta de fondos ACLARADOS durante los U6M. Solo Cheques Bci.
CL_Num_Tot_Protestos_Acla_U12M	Cantidad de protestos por falta de fondos ACLARADOS durante los U12M. Solo Cheques Bci.
CL_Numero_eventos_SGNP_mes	Es la suma de los sobregiros no pactados en todas sus cuentas corrientes del cliente en el mes.
CL_Numero_eventos_SGNP_U3M	Es la suma de los sobregiros no pactados en todas sus cuentas corrientes del cliente en el mes.
CL_Porcentaje_Uso_LDC	Es el porcentaje de uso de las líneas de crédito del cliente.
CL_Prestamos_Otros_Bancos_sist	Marca que indica si el cliente tiene préstamos en otros bancos. Considera desde 2 meses de desfase según la fecha de proceso.
CL_Renegociado_COM	Cliente con al menos u operación renegociada en cualquiera de sus productos de crédito Comercial durante el mes.
CL_Renta_líquida_en_UF	Corresponde a la renta líquida del cliente en UF al cierre de mes.
CL_Sdo_Minimo_CCT_Activas_Mes	Saldo mínimo de todas las cuentas corrientes activas del Cliente durante el mes dividida por la renta líquida en UF.
CL_Sexo	Sexo del cliente
CL_Situacion_Laboral	Situación laboral del cliente.
CL_Tipo_Vivienda	Tipo de vivienda del cliente.
TDC_Ant_ULT_AUM_CUPO	Diferencia entre la fecha actual y la fecha del último aumento de cupo en la tarjeta de crédito

TDC_Ant_ULT_DSIM_CUPO	Diferencia entre la fecha actual y la fecha de la última disminución de cupo en la tarjeta de crédito
CON_AntCurCon	Antigüedad desde que se genera el crédito de consumo más antiguo
CON_Flag_Multi	Indica si el cliente tiene un multicrédito
CON_PorcentajePagado	Cuotas pagadas dividido por cuotas originales
CON_PorcentajeXPagar	Cuotas por vencer dividido por cuotas originales
CON_ApalancamientoxCuota	Suma de las cuotas de las operaciones de consumo dividido por la renta total (Renta fija + Renta Variable)
Acciones1	Indica si el cliente posee o no acciones en el último mes
Acciones3	Indica si el cliente posee o no acciones en los últimos 3 meses
Acciones6	Indica si el cliente posee o no acciones en los últimos 6 meses
Acciones12	Indica si el cliente posee o no acciones en los últimos 12 meses
Ahorro1	Indica si el cliente posee o no ahorro en el último mes
Ahorro3	Indica si el cliente posee o no ahorro en los últimos 3 meses
Ahorro6	Indica si el cliente posee o no ahorro en los últimos 6 meses
Ahorro12	Indica si el cliente posee o no ahorro en los últimos 12 meses
Fondos1	Indica si el cliente posee o no fondos mutuos en el último mes
Fondos3	Indica si el cliente posee o no fondos mutuos en los últimos 3 meses
Fondos6	Indica si el cliente posee o no fondos mutuos en los últimos 6 meses
Fondos12	Indica si el cliente posee o no fondos mutuos en los últimos 12 meses
DAP1	Indica si el cliente posee o no DAP en el último mes
DAP3	Indica si el cliente posee o no DAP en los últimos 3 meses
DAP6	Indica si el cliente posee o no DAP en los últimos 6 meses
DAP12	Indica si el cliente posee o no DAP en los últimos 12 meses
INV1	Indica si el cliente posee o no alguno de los siguientes productos: Acciones, Ahorro, Fondos Mutuos y/o DAP en el último mes
INV3	Indica si el cliente posee o no alguno de los siguientes productos: Acciones, Ahorro, Fondos Mutuos y/o DAP en los últimos 3 meses
INV6	Indica si el cliente posee o no alguno de los siguientes productos: Acciones, Ahorro, Fondos Mutuos y/o DAP en los últimos 6 meses
INV12	Indica si el cliente posee o no alguno de los siguientes productos: Acciones, Ahorro, Fondos Mutuos y/o DAP en los últimos 12 meses
AvNac	Indica si el cliente realizó avances nacionales el último mes
AvNac3	Indica si el cliente realizó avances nacionales en los últimos 3 meses
AvNac6	Indica si el cliente realizó avances nacionales en los últimos 6 meses
AvNac12	Indica si el cliente realizó avances nacionales en los últimos 12 meses
CtaCte?	Indica si cliente posee cuenta corriente
TDC?	Indica si cliente posee TDC
LDC?	Indica si cliente posee LDC
Con credito	Indica si cliente tiene un crédito de consumo vigente

## Anexo 4: Variables creadas

Variable	descripción
<b>cantidadproductos</b>	total de productos que posee el cliente
<b>cat_tipoproducto_posee</b>	variable categórica que representa las combinaciones posibles entre productos del banco
<b>sexo_estadocivil</b>	variable categórica que genera segmentos hombre
<b>EdadxRenta</b>	Multiplicación de edad por la renta
<b>EdadxDeudaCliente</b>	Multiplicación de edad por deuda
<b>cat_acc_fondos_mes</b>	Categórica que señala si el cliente posee: solo acciones; solo fondos mutuos; acciones y fondos mutuos a la vez; ninguno de los dos. Esto en el último mes
<b>cat_acc_fondos_3m</b>	Categórica que señala si el cliente posee: solo acciones; solo fondos mutuos; acciones y fondos mutuos a la vez; ninguno de los dos. Esto en los últimos 3 meses
<b>cat_acc_fondos_6m</b>	Categórica que señala si el cliente posee: solo acciones; solo fondos mutuos; acciones y fondos mutuos a la vez; ninguno de los dos. Esto en los últimos 6 meses
<b>cat_acc_fondos_12m</b>	Categórica que señala si el cliente posee: solo acciones; solo fondos mutuos; acciones y fondos mutuos a la vez; ninguno de los dos. Esto en los últimos 12 meses
<b>cat_acc_DAP_12m</b>	Categórica que señala si el cliente posee: solo acciones; solo DAP; acciones y DAP a la vez; ninguno de los dos. Esto en el último mes
<b>cat_acc_DAP_6m</b>	Categórica que señala si el cliente posee: solo acciones; solo DAP; acciones y DAP a la vez; ninguno de los dos. Esto en los últimos 3 meses
<b>cat_acc_DAP_3m</b>	Categórica que señala si el cliente posee: solo acciones; solo DAP; acciones y DAP a la vez; ninguno de los dos. Esto en los últimos 6 meses
<b>cat_acc_DAP_mes</b>	Categórica que señala si el cliente posee: solo acciones; solo DAP; acciones y DAP a la vez; ninguno de los dos. Esto en los últimos 12 meses
<b>cat_acc_aho_mes</b>	Categórica que señala si el cliente posee: solo acciones; solo Ahorro; acciones y Ahorro a la vez; ninguno de los dos. Esto en el último mes
<b>cat_acc_aho_3m</b>	Categórica que señala si el cliente posee: solo acciones; solo Ahorro; acciones y Ahorro a la vez; ninguno de los dos. Esto en los últimos 3 meses

<b>cat_acc_aho_6m</b>	Catagórica que señala si el cliente posee: solo acciones; solo Ahorro; acciones y Ahorro a la vez; ninguno de los dos. Esto en los últimos 6 meses
<b>cat_acc_aho_12m</b>	Catagórica que señala si el cliente posee: solo acciones; solo Ahorro; acciones y Ahorro a la vez; ninguno de los dos. Esto en los últimos 12 meses
<b>cat_fondos_DAP_12m</b>	Catagórica que señala si el cliente posee: solo fondos mutuos; solo DAP; fondos mutuos y DAP a la vez; ninguno de los dos. Esto en el último mes
<b>cat_fondos_DAP_6m</b>	Catagórica que señala si el cliente posee: solo fondos mutuos; solo DAP; fondos mutuos y DAP a la vez; ninguno de los dos. Esto en los últimos 3 meses
<b>cat_fondos_DAP_3m</b>	Catagórica que señala si el cliente posee: solo fondos mutuos; solo DAP; fondos mutuos y DAP a la vez; ninguno de los dos. Esto en los últimos 6 meses
<b>cat_fondos_DAP_mes</b>	Catagórica que señala si el cliente posee: solo fondos mutuos; solo DAP; fondos mutuos y DAP a la vez; ninguno de los dos. Esto en los últimos 12 meses
<b>cat_fondos_aho_mes</b>	Catagórica que señala si el cliente posee: solo fondos mutuos; solo Ahorro; fondos mutuos y Ahorro a la vez; ninguno de los dos. Esto en el último mes
<b>cat_fondos_aho_3m</b>	Catagórica que señala si el cliente posee: solo fondos mutuos; solo Ahorro; fondos mutuos y Ahorro a la vez; ninguno de los dos. Esto en los últimos 3 meses
<b>cat_fondos_aho_6m</b>	Catagórica que señala si el cliente posee: solo fondos mutuos; solo Ahorro; fondos mutuos y Ahorro a la vez; ninguno de los dos. Esto en los últimos 6 meses
<b>cat_fondos_aho_12m</b>	Catagórica que señala si el cliente posee: solo fondos mutuos; solo Ahorro; fondos mutuos y Ahorro a la vez; ninguno de los dos. Esto en los últimos 12 meses
<b>cat_aho_DAP_12m</b>	Catagórica que señala si el cliente posee: solo DAP; solo Ahorro; DAP y Ahorro a la vez; ninguno de los dos. Esto en el último mes
<b>cat_aho_DAP_6m</b>	Catagórica que señala si el cliente posee: solo DAP; solo Ahorro; DAP y Ahorro a la vez; ninguno de los dos. Esto en los últimos 3 meses
<b>cat_aho_DAP_3m</b>	Catagórica que señala si el cliente posee: solo DAP; solo Ahorro; DAP y Ahorro a la vez; ninguno de los dos. Esto en los últimos 6 meses
<b>cat_aho_DAP_mes</b>	Catagórica que señala si el cliente posee: solo DAP; solo Ahorro; DAP y Ahorro a la vez; ninguno de los dos. Esto en los últimos 12 meses
<b>sum_acc_fondos_mes</b>	suma de montos que cliente posee en acciones y fondos mutuos en el mes
<b>sum_acc_fondos_3m</b>	suma de montos que cliente posee en acciones y fondos mutuos en los últimos 3 meses
<b>sum_acc_fondos_6m</b>	suma de montos que cliente posee en acciones y fondos mutuos en los últimos 6 meses
<b>sum_acc_fondos_12m</b>	suma de montos que cliente posee en acciones y fondos mutuos en los últimos 12 meses
<b>sum_acc_DAP_12m</b>	suma de montos que cliente posee en acciones y DAP en el mes
<b>sum_acc_DAP_6m</b>	suma de montos que cliente posee en acciones y DAP en los últimos 3 meses
<b>sum_acc_DAP_3m</b>	suma de montos que cliente posee en acciones y DAP en los últimos 6 meses
<b>sum_acc_DAP_mes</b>	suma de montos que cliente posee en acciones y DAP en los últimos 12 meses
<b>sum_acc_aho_mes</b>	suma de montos que cliente posee en acciones y Ahorro en el mes
<b>sum_acc_aho_3m</b>	suma de montos que cliente posee en acciones y Ahorro en los últimos 3 meses
<b>sum_acc_aho_6m</b>	suma de montos que cliente posee en acciones y Ahorro en los últimos 6 meses
<b>sum_acc_aho_12m</b>	suma de montos que cliente posee en acciones y Ahorro en los últimos 12 meses
<b>sum_fondos_DAP_12m</b>	suma de montos que cliente posee en fondos mutuos y DAP en el mes
<b>sum_fondos_DAP_6m</b>	suma de montos que cliente posee en fondos mutuos y DAP en los últimos 3 meses
<b>sum_fondos_DAP_3m</b>	suma de montos que cliente posee en fondos mutuos y DAP en los últimos 6 meses
<b>sum_fondos_DAP_mes</b>	suma de montos que cliente posee en fondos mutuos y DAP en los

	últimos 12 meses
<b>sum_fondos_aho_mes</b>	suma de montos que cliente posee en fondos mutuos y Ahorro en el mes
<b>sum_fondos_aho_3m</b>	suma de montos que cliente posee en fondos mutuos y Ahorro en los últimos 3 meses
<b>sum_fondos_aho_6m</b>	suma de montos que cliente posee en fondos mutuos y Ahorro en los últimos 6 meses
<b>sum_fondos_aho_12m</b>	suma de montos que cliente posee en fondos mutuos y Ahorro en los últimos 12 meses
<b>sum_aho_DAP_12m</b>	suma de montos que cliente posee en DAP y Ahorro en el mes
<b>sum_aho_DAP_6m</b>	suma de montos que cliente posee en DAP y Ahorro en los últimos 3 meses
<b>sum_aho_DAP_3m</b>	suma de montos que cliente posee en DAP y Ahorro en los últimos 6 meses
<b>sum_aho_DAP_mes</b>	suma de montos que cliente posee en DAP y Ahorro en los últimos 12 meses
<b>mult_acc_fondos_mes</b>	multiplicación de montos que cliente posee en acciones y fondos mutuos en el mes
<b>mult_acc_fondos_3m</b>	multiplicación de montos que cliente posee en acciones y fondos mutuos en los últimos 3 meses
<b>mult_acc_fondos_6m</b>	multiplicación de montos que cliente posee en acciones y fondos mutuos en los últimos 6 meses
<b>mult_acc_fondos_12m</b>	multiplicación de montos que cliente posee en acciones y fondos mutuos en los últimos 12 meses
<b>mult_acc_DAP_12m</b>	multiplicación de montos que cliente posee en acciones y DAP en el mes
<b>mult_acc_DAP_6m</b>	multiplicación de montos que cliente posee en acciones y DAP en los últimos 3 meses
<b>mult_acc_DAP_3m</b>	multiplicación de montos que cliente posee en acciones y DAP en los últimos 6 meses
<b>mult_acc_DAP_mes</b>	multiplicación de montos que cliente posee en acciones y DAP en los últimos 12 meses
<b>mult_acc_aho_mes</b>	multiplicación de montos que cliente posee en acciones y Ahorro en el mes
<b>mult_acc_aho_3m</b>	multiplicación de montos que cliente posee en acciones y Ahorro en los últimos 3 meses
<b>mult_acc_aho_6m</b>	multiplicación de montos que cliente posee en acciones y Ahorro en los últimos 6 meses
<b>mult_acc_aho_12m</b>	multiplicación de montos que cliente posee en acciones y Ahorro en los últimos 12 meses
<b>mult_fondos_DAP_12m</b>	multiplicación de montos que cliente posee en fondos mutuos y DAP en el mes
<b>mult_fondos_DAP_6m</b>	multiplicación de montos que cliente posee en fondos mutuos y DAP en los últimos 3 meses
<b>mult_fondos_DAP_3m</b>	multiplicación de montos que cliente posee en fondos mutuos y DAP en los últimos 6 meses
<b>mult_fondos_DAP_mes</b>	multiplicación de montos que cliente posee en fondos mutuos y DAP en los últimos 12 meses
<b>mult_fondos_aho_mes</b>	multiplicación de montos que cliente posee en fondos mutuos y Ahorro en el mes
<b>mult_fondos_aho_3m</b>	multiplicación de montos que cliente posee en fondos mutuos y Ahorro en los últimos 3 meses
<b>mult_fondos_aho_6m</b>	multiplicación de montos que cliente posee en fondos mutuos y Ahorro en los últimos 6 meses
<b>mult_fondos_aho_12m</b>	multiplicación de montos que cliente posee en fondos mutuos y Ahorro en los últimos 12 meses
<b>mult_aho_DAP_12m</b>	multiplicación de montos que cliente posee en DAP y Ahorro en el mes
<b>mult_aho_DAP_6m</b>	multiplicación de montos que cliente posee en DAP y Ahorro en los últimos 3 meses
<b>mult_aho_DAP_3m</b>	multiplicación de montos que cliente posee en DAP y Ahorro en los últimos 6 meses
<b>mult_aho_DAP_mes</b>	multiplicación de montos que cliente posee en DAP y Ahorro en los últimos 12 meses
<b>cat_con_hip_directa_sist</b>	categoría que señala si el cliente posee deuda directa en el sistema: solo en consumo; solo hipotecario; hipotecaria y consumo; ninguna
<b>cat_con_com_directa_sist</b>	categoría que señala si el cliente posee deuda directa en el sistema: solo

	en consumo; solo comercial; comercial y consumo; ninguna
<b>cat_com_hip_directa_sist</b>	categórica que señala si el cliente posee deuda directa en el sistema: solo en hipotecaria; solo comercial; comercial e hipotecaria; ninguna
<b>cat_con_hip_com_directa_sist</b>	categórica que señala si el cliente posee deuda directa en el sistema: solo en hipotecaria; solo comercial; solo consumo; comercial e hipotecaria; comercial y consumo; hipotecaria y consumo; comercial, consumo e hipotecaria; ninguna
<b>sum_con_hip_directa_sist</b>	suma de monto deuda directa en sistema consumo e hipotecaria
<b>sum_con_com_directa_sist</b>	suma de monto deuda directa en sistema consumo e comercial
<b>sum_com_hip_directa_sist</b>	suma de monto deuda directa en sistema hipotecaria e comercial
<b>sum_con_hip_com_directa_sist</b>	suma de monto deuda directa en sistema consumo, hipotecaria y comercial
<b>mult_con_hip_directa_sist</b>	multiplicación de monto deuda directa en sistema consumo e hipotecaria
<b>mult_con_com_directa_sist</b>	multiplicación de monto deuda directa en sistema consumo e comercial
<b>mult_com_hip_directa_sist</b>	multiplicación de monto deuda directa en sistema hipotecaria e comercial
<b>mult_con_hip_com_directa_sist</b>	multiplicación de monto deuda directa en sistema consumo, hipotecaria y comercial
<b>cat_con_BCI_NoBCI</b>	categórica que indica si cliente posee deuda en consumo: solo BCI; solo NoBCI; BCI y NoBCI; ninguna
<b>sum_con_BCI_NoBCI</b>	suma monto deuda consumo BCI y no BCI
<b>mult_con_BCI_NoBCI</b>	multiplicación monto deuda consumo BCI y no BCI
<b>cat_com_BCI_NoBCI</b>	categórica que indica si cliente posee deuda en comercial: solo BCI; solo NoBCI; BCI y NoBCI; ninguna
<b>sum_com_BCI_NoBCI</b>	suma monto deuda comercial BCI y no BCI
<b>mult_com_BCI_NoBCI</b>	multiplicación monto deuda comercial BCI y no BCI
<b>cat_hip_BCI_NoBCI</b>	categórica que indica si cliente posee deuda en hipotecaria: solo BCI; solo NoBCI; BCI y NoBCI; ninguna
<b>sum_hip_BCI_NoBCI</b>	suma monto deuda hipotecaria BCI y no BCI
<b>mult_hip_BCI_NoBCI</b>	multiplicación monto deuda hipotecaria BCI y no BCI
<b>cat_LDC_TDC_Bci</b>	categórica que indica cliente posee deuda en BCI: solo LDC; solo TDC; LDC y TDC; ninguna
<b>cat_LDC_LEM_BCI</b>	categórica que indica cliente posee deuda en BCI: solo LDC; solo LEM; LDC y LEM; ninguna
<b>cat_TDC_LEM_BCI</b>	categórica que indica cliente posee deuda en BCI: solo TDC; solo LEM; TDC y LEM; ninguna
<b>cat_LDC_TDC_LEM_BCI</b>	categórica que indica si cliente posee deuda BCI: solo LDC; solo TDC; solo LEM; LDC y TDC; LDC y LEM; TDC y LEM; LDC, LEM y TDC; ninguna
<b>sum_LDC_TDC_Bci</b>	suma monto deuda bci LDC y TDC
<b>sum_LDC_LEM_BCI</b>	suma monto deuda bci LDC y LEM
<b>sum_TDC_LEM_BCI</b>	suma monto deuda bci TDC y LEM
<b>sum_LDC_TDC_LEM_BCI</b>	suma monto deuda bci LDC, LEM y TDC
<b>mult_LDC_TDC_Bci</b>	multiplicación monto deuda bci LDC y TDC
<b>mult_LDC_LEM_BCI</b>	multiplicación monto deuda bci LDC y LEM
<b>mult_TDC_LEM_BCI</b>	multiplicación monto deuda bci TDC y LEM
<b>mult_LDC_TDC_LEM_BCI</b>	multiplicación monto deuda bci LDC, LEM y TDC
<b>cat_LDC_TDC_sbif</b>	categórica que indica cliente posee deuda en sbif: solo LDC; solo TDC; LDC y TDC; ninguna
<b>cat_LDC_LEM_sbif</b>	categórica que indica cliente posee deuda en sbif: solo LDC; solo LEM; LDC y LEM; ninguna
<b>cat_TDC_LEM_sbif</b>	categórica que indica cliente posee deuda en sbif: solo TDC; solo LEM; TDC y LEM; ninguna
<b>cat_LDC_TDC_LEM_sbif</b>	categórica que indica si cliente posee deuda sbif: solo LDC; solo TDC; solo LEM; LDC y TDC; LDC y LEM; TDC y LEM; LDC, LEM y TDC; ninguna
<b>sum_LDC_TDC_sbif</b>	suma monto deuda sbif LDC y TDC
<b>sum_LDC_LEM_sbif</b>	suma monto deuda sbif LDC y LEM

<b>sum_TDC_LEM_sbif</b>	suma monto deuda sbif TDC y LEM
<b>sum_LDC_TDC_LEM_sbif</b>	suma monto deuda sbif LDC, LEM y TDC
<b>mult_LDC_TDC_sbif</b>	multiplicacion monto deuda sbif LDC y TDC
<b>mult_LDC_LEM_sbif</b>	multiplicacion monto deuda sbif LDC y LEM
<b>mult_TDC_LEM_sbif</b>	multiplicacion monto deuda sbif TDC y LEM
<b>mult_LDC_TDC_LEM_sbif</b>	multiplicacion monto deuda sbif LDC, LEM y TDC
<b>cat_LDC_SGNP_Bci</b>	categórica que indica cliente posee deuda en bci: solo LDC; solo SGPN; LDC y SGPN; ninguna
<b>cat_TDC_SGNP_BCI</b>	categórica que indica cliente posee deuda en bci: solo SGPN; solo TDC; SGPN y TDC; ninguna
<b>cat_SGNP_LEM_BCI</b>	categórica que indica cliente posee deuda en bci: solo SGPN; solo LEM; SGPN y LEM; ninguna
<b>cat_LDC_TDC_SGNP_BCI</b>	categórica que indica si cliente posee deuda bci: solo LDC; solo TDC; solo SGNP; LDC y TDC; LDC y SGNP; TDC y SGNP; LDC, TDC y SGNP; ninguna
<b>cat_LDC_SGNP_LEM_BCI</b>	categórica que indica si cliente posee deuda bci: solo LDC; solo LEM; solo SGNP; LDC y LEM; LDC y SGNP; LEM y SGNP; LDC, LEM y SGNP; ninguna
<b>cat_SGNP_TDC_LEM_BCI</b>	categórica que indica si cliente posee deuda bci: solo TDC; solo LEM; solo SGNP; TDC y LEM; TDC y SGNP; LEM y SGNP; TDC, LEM y SGNP; ninguna
<b>sum_SGNP_TDC_Bci</b>	suma monto deuda BCI SGPN y TDC
<b>sum_LDC_SGNP_BCI</b>	suma monto deuda BCI SGPN y LDC
<b>sum_SGNP_LEM_BCI</b>	suma monto deuda BCI SGPN y LEM
<b>sum_LDC_TDC_SGNP_BCI</b>	suma monto deuda BCI LDC, TDC y SGNP
<b>sum_LDC_SGNP_LEM_BCI</b>	suma monto deuda BCI LDC, LEM y SGNP
<b>sum_SGNP_TDC_LEM_BCI</b>	suma monto deuda BCI LEM, TDC y SGNP
<b>mult_LDC_SGNP_Bci</b>	multiplicacion monto deuda BCI SGPN y TDC
<b>mult_SGNP_LEM_BCI</b>	multiplicacion monto deuda BCI SGPN y LDC
<b>mult_TDC_SGNP_BCI</b>	multiplicacion monto deuda BCI SGPN y LEM
<b>mult_LDC_TDC_SGNP_BCI</b>	multiplicacion monto deuda BCI LDC, TDC y SGNP
<b>mult_LDC_SGNP_LEM_BCI</b>	multiplicacion monto deuda BCI LDC, LEM y SGNP
<b>mult_SGNP_TDC_LEM_BCI</b>	multiplicacion monto deuda BCI LEM, TDC y SGNP
<b>cat_LDC_SGNP_sbif</b>	categórica que indica cliente posee deuda en sbif: solo LDC; solo SGPN; LDC y SGPN; ninguna
<b>cat_SGNP_LEM_sbif</b>	categórica que indica cliente posee deuda en sbif: solo SGPN; solo TDC; SGPN y TDC; ninguna
<b>cat_TDC_SGNP_sbif</b>	categórica que indica cliente posee deuda en sbif: solo SGPN; solo LEM; SGPN y LEM; ninguna
<b>cat_LDC_SGNP_LEM_sbif</b>	categórica que indica si cliente posee deuda sbif: solo LDC; solo TDC; solo SGNP; LDC y TDC; LDC y SGNP; TDC y SGNP; LDC, TDC y SGNP; ninguna
<b>cat_TDC_SGNP_LEM_sbif</b>	categórica que indica si cliente posee deuda sbif: solo LDC; solo LEM; solo SGNP; LDC y LEM; LDC y SGNP; LEM y SGNP; LDC, LEM y SGNP; ninguna
<b>cat_LDC_SGNP_TDC_sbif</b>	categórica que indica si cliente posee deuda sbif: solo TDC; solo LEM; solo SGNP; TDC y LEM; TDC y SGNP; LEM y SGNP; TDC, LEM y SGNP; ninguna
<b>sum_LDC_SGNP_sbif</b>	suma monto deuda sbif SGPN y TDC
<b>sum_SGNP_LEM_sbif</b>	suma monto deuda sbif SGPN y LDC
<b>sum_TDC_SGNP_sbif</b>	suma monto deuda sbif SGPN y LEM
<b>sum_LDC_SGNP_LEM_sbif</b>	suma monto deuda sbif LDC, TDC y SGNP
<b>sum_SGNP_TDC_LEM_sbif</b>	suma monto deuda sbif LDC, LEM y SGNP
<b>sum_LDC_TDC_SGNP_sbif</b>	suma monto deuda sbif LEM, TDC y SGNP
<b>mult_LDC_SGNP_sbif</b>	multiplicacion monto deuda sbif SGPN y TDC



<b>mult_SGNP_LEM_sbif</b>	multiplicacion monto deuda sbif SGNP y LDC
<b>mult_TDC_SGNP_sbif</b>	multiplicacion monto deuda sbif SGNP y LEM
<b>mult_LDC_SGNP_LEM_sbif</b>	multiplicacion monto deuda sbif LDC, TDC y SGNP
<b>mult_SGNP_TDC_LEM_sbif</b>	multiplicacion monto deuda sbif LDC, LEM y SGNP
<b>mult_LDC_TDC_SGNP_sbif</b>	multiplicacion monto deuda sbif LEM, TDC y SGNP
<b>sum_LDC_TDC_cupo</b>	suma cupo LDC y TDC
<b>sum_LDC_LEM_cupo</b>	suma cupo LDC y LEM
<b>sum_TDC_LEM_cupo</b>	suma cupo TDC y LEM
<b>sum_LDC_TDC_LEM_cupo</b>	suma cupo TDC, LDC y LEM
<b>mult_LDC_TDC_cupo</b>	multiplicacion cupo LDC y TDC
<b>mult_LDC_LEM_cupo</b>	multiplicacion cupo LDC y LEM
<b>mult_TDC_LEM_cupo</b>	multiplicacion cupo TDC y LEM
<b>mult_LDC_TDC_LEM_cupo</b>	multiplicacion cupo TDC, LDC y LEM
<b>sum_morablanda_CCT_CON</b>	suma mora blanda CCT y credito de consumo
<b>sum_morablanda_CCT_TDC</b>	suma mora blanda CCT y TDC
<b>sum_morablanda_CON_TDC</b>	suma mora blanco TDC y credito de consumo
<b>sum_morablanda_CON_CCT_TDC</b>	suma mora blanda CCT, credito de consumo y TDC
<b>mult_morablanda_CCT_CON</b>	multiplicacion mora blanda CCT y credito de consumo
<b>mult_morablanda_CCT_TDC</b>	multiplicacion mora blanda CCT y TDC
<b>mult_morablanda_CON_TDC</b>	multiplicacion mora blanco TDC y credito de consumo
<b>mult_morablanda_CON_CCT_TDC</b>	multiplicacion mora blanda CCT, credito de consumo y TDC
<b>sum_moradura_CCT_CON</b>	suma mora dura CCT y credito de consumo
<b>sum_moradura_CCT_TDC</b>	suma mora dura CCT y TDC
<b>sum_moradura_CON_TDC</b>	suma mora duraTDC y credito de consumo
<b>sum_moradura_CON_CCT_TDC</b>	suma mora dura CCT, credito de consumo y TDC
<b>mult_moradura_CCT_CON</b>	multiplicacion mora dura CCT y credito de consumo
<b>mult_moradura_CCT_TDC</b>	multiplicacion mora dura CCT y TDC
<b>mult_moradura_CON_TDC</b>	multiplicacion mora duraTDC y credito de consumo
<b>mult_moradura_CON_CCT_TDC</b>	multiplicacion mora dura CCT, credito de consumo y TDC

## **Anexo 5: Detalle de pasos para modelamiento de modelo de propuesta N°1.**

En primer lugar, para cada segmento se realizó el análisis univariado para las 1.444 variables (1.262 de las bases de datos otorgadas por el banco y 182 nuevas variables). A continuación, se observa el número de variables tramificadas para cada uno de los segmentos.

<b>Segmento</b>	<b>Número de variables tramificadas</b>
<b>1</b>	1.089
<b>2</b>	898
<b>3</b>	989
<b>4</b>	1.266
<b>5</b>	1.045
<b>6</b>	1.017
<b>7</b>	1.013

Con las variables ya tramificadas, se procedió con cálculo del valor WOE para cada tramo de cada variable. Luego, para cada variable se calculó su *Information Value*. Posteriormente, para cada segmento se elaboró el ranking de *Information Value*. Del ranking construido, se descartó las variables que posean un IV menor a 2, pues significa que su poder de discriminancia es bajo. En la siguiente tabla, se aprecia el número de variables para cada segmento luego de la eliminación.

Segmento	Número de variables posterior a eliminación
1	1.089
2	869
3	955
4	1.183
5	1.032
6	949
7	916

El siguiente paso que se realizó fue el análisis de correlaciones, con el fin de eliminar las variables que sean redundantes para el modelo. El resultado de este análisis redujo la cantidad de variables, en cada segmento, al número que se aprecia en la siguiente tabla:

Segmento	Número de variables posterior a análisis de correlación
1	57
2	75
3	73
4	73
5	66
6	89
7	95

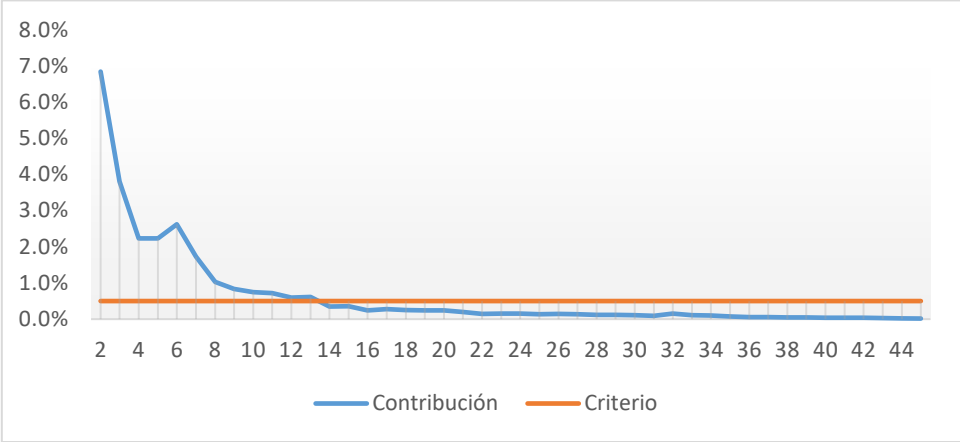
Luego, se realizó la regresión logística para cada segmento. El número de variables que se obtuvo como resultado para cada segmento se aprecia a continuación.

Segmento	Número de variables a incluir
1	45
2	51
3	58
4	34
5	42
6	52
7	64

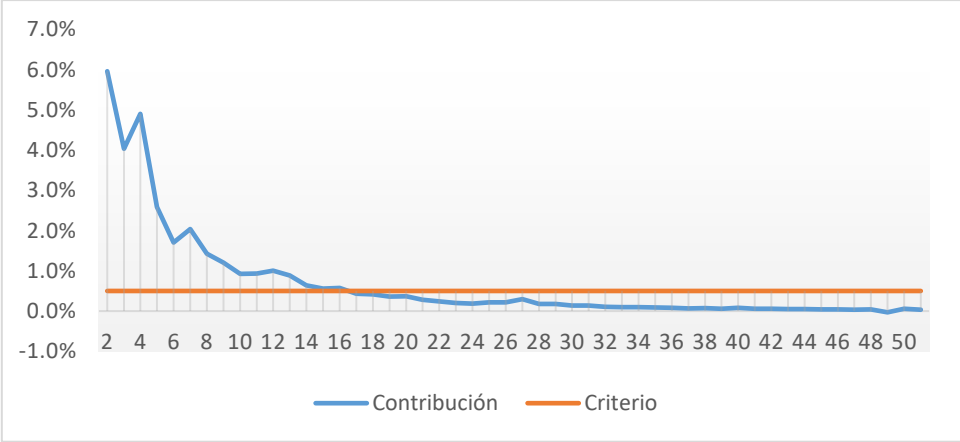
Se observa en la tabla anterior, que es una gran cantidad de variables para cada segmento, por lo cual sería un modelo difícil de comprender e interpretar. Para reducir las variables se realizó un análisis del aporte a la

logverosimilitud. El resultado del análisis determinó la cantidad de variables a incorporar para el modelo de cada segmento se muestran en los siguientes gráficos de contribución a la logverosimilitud

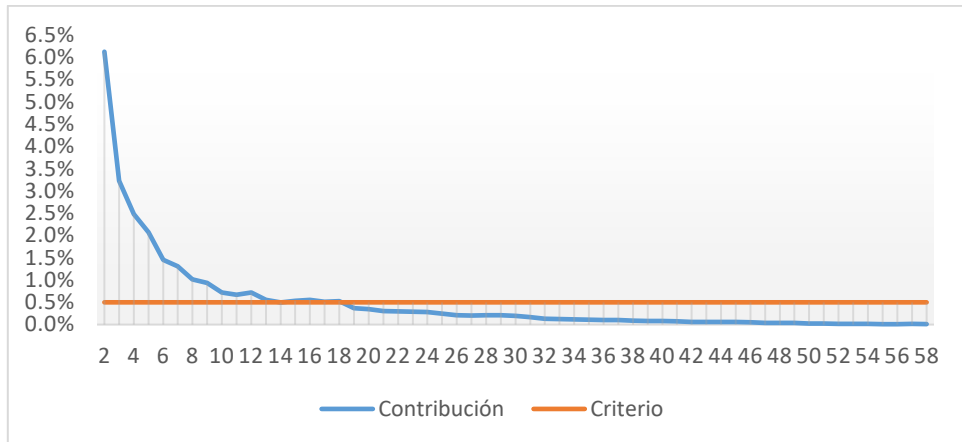
- Contribución a logverosimilitud, segmento 1



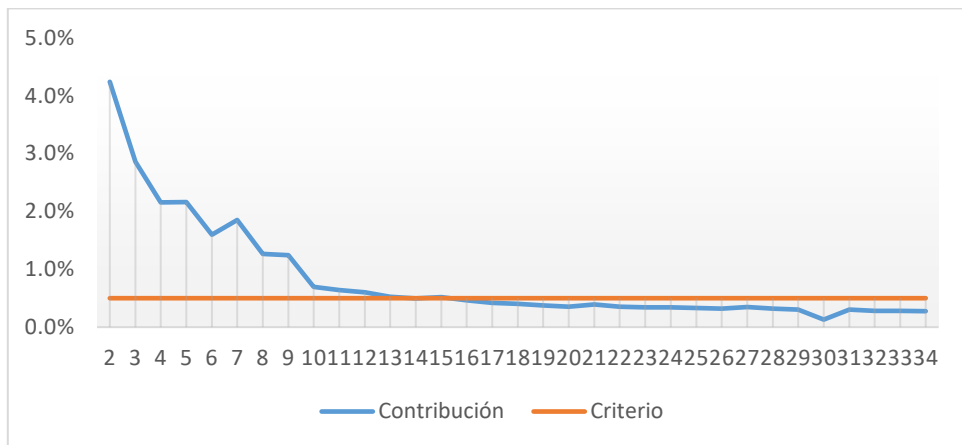
- Contribución a logverosimilitud, segmento 2



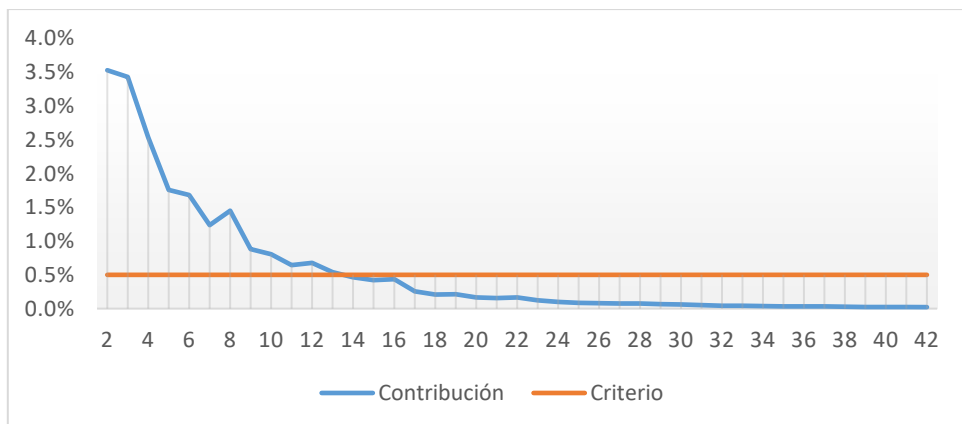
- Contribución a logverosimilitud, segmento 3



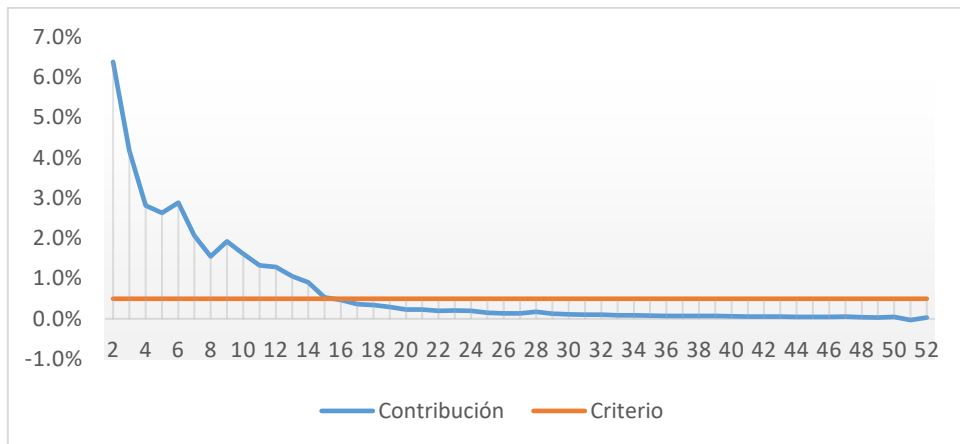
- Contribución a logverosimilitud, segmento 4



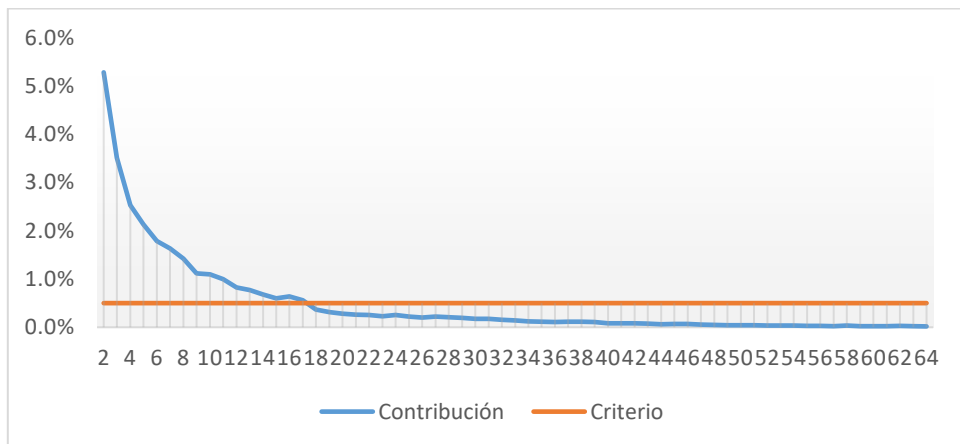
- Contribución a logverosimilitud, segmento 5



- Contribución a logverosimilitud, segmento 6



- Contribución a logverosimilitud, segmento 7



## Anexo 6: Detalle de pasos para modelamiento de modelo de propuesta N°2

De igual forma que para la propuesta anterior, en primer lugar, para cada segmento originado se realizó el análisis univariado para las 1.446 variables. A continuación, se observa el número de variables tramificadas para cada uno de los segmentos.

Segmento	Número de variables tramificadas
<b>1.1</b>	903
<b>1.2</b>	1.136
<b>3.1</b>	828
<b>3.2</b>	969
<b>5.1</b>	973
<b>5.2</b>	1.039

Luego, se construyó, para cada segmento, el ranking de *Information Value*. Del ranking construido, se descartó las variables que posean un IV menor a 2. En la siguiente tabla se aprecia el número de variables para cada segmento luego de la eliminación por IV.

Segmento	Número de variables posterior a eliminación
<b>1.1</b>	903
<b>1.2</b>	816
<b>3.1</b>	788
<b>3.2</b>	878
<b>5.1</b>	972
<b>5.2</b>	892

El siguiente paso que se realizó fue el análisis de correlaciones. El resultado de este análisis redujo la cantidad de variables en cada segmento, como se aprecia en la siguiente tabla.

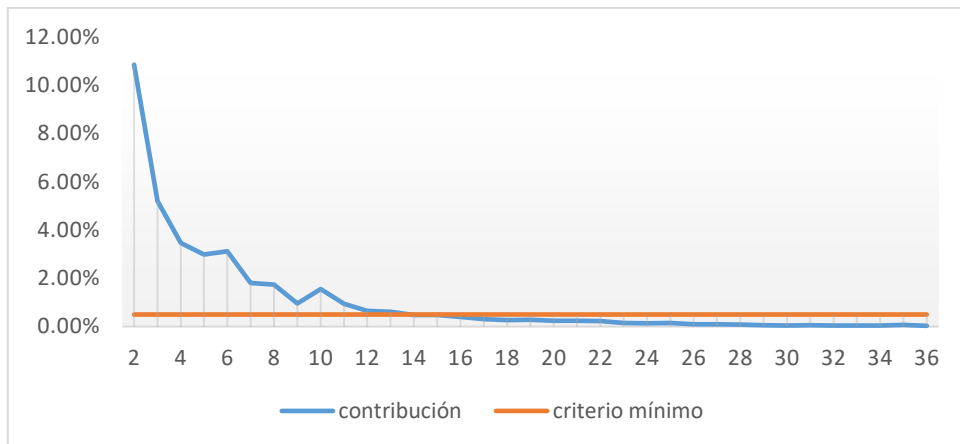
Segmento	Número de variables posterior a análisis de correlación
<b>1.1</b>	108
<b>1.2</b>	48
<b>3.1</b>	86
<b>3.2</b>	131
<b>5.1</b>	94
<b>5.2</b>	119

Luego, se realizó la regresión logística para cada segmento. El número de variables que se obtuvo como resultado para cada segmento se aprecia en siguiente tabla.

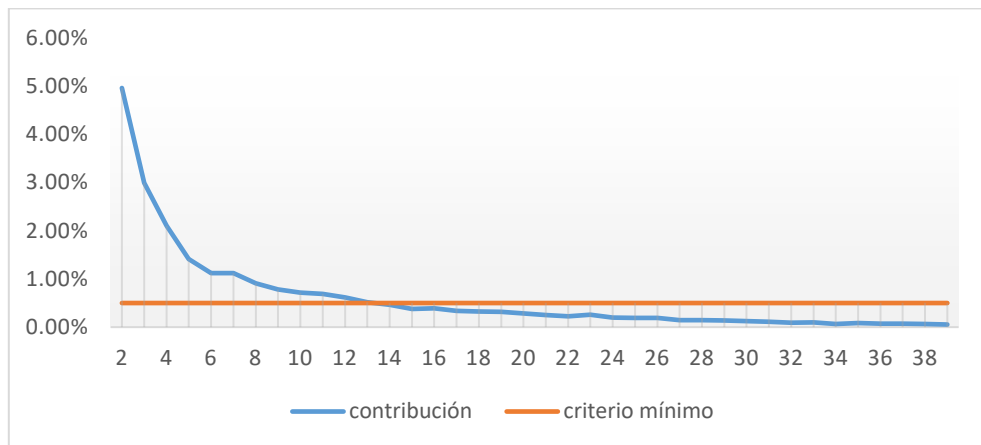
Segmento	Número de variables a incluir
<b>1.1</b>	36
<b>1.2</b>	39
<b>3.1</b>	42
<b>3.2</b>	54
<b>5.1</b>	38
<b>5.2</b>	41

Para reducir las variables se realizó un análisis del aporte a la logverosimilitud. El resultado del análisis determinó la cantidad de variables a incorporar para el modelo de cada segmento. Los gráficos de contribución a la logverosimilitud se muestran a continuación.

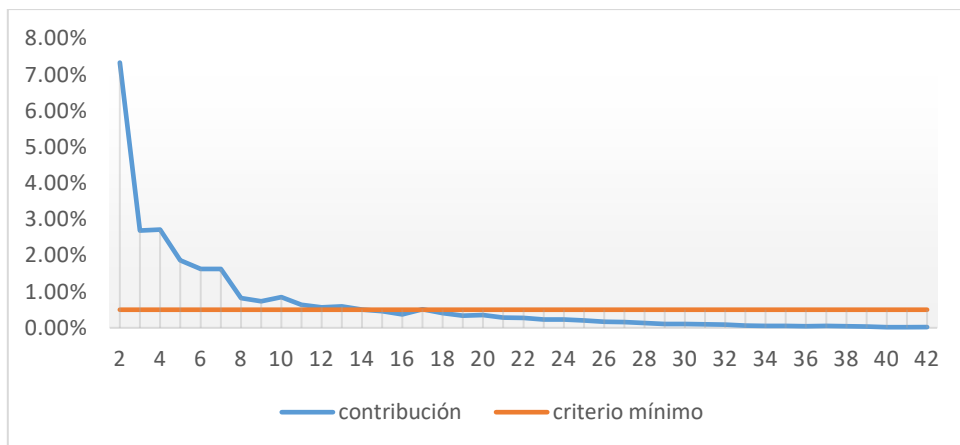
- Segmento 1.1



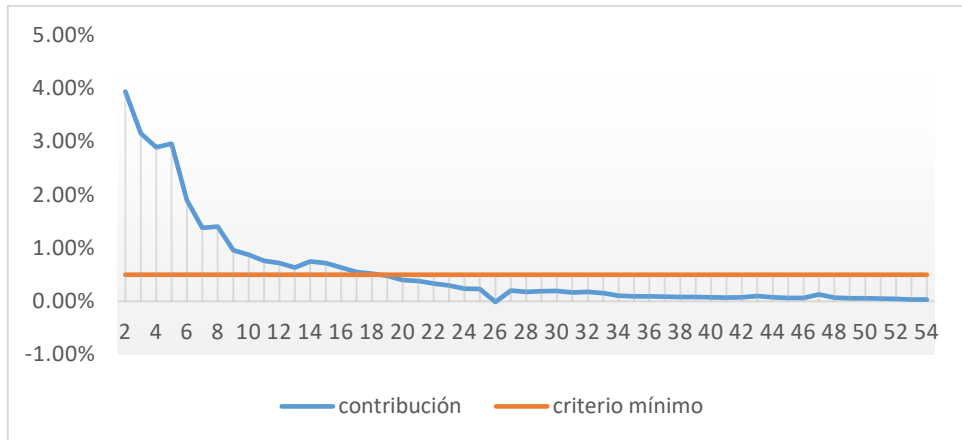
- Segmento 1.2



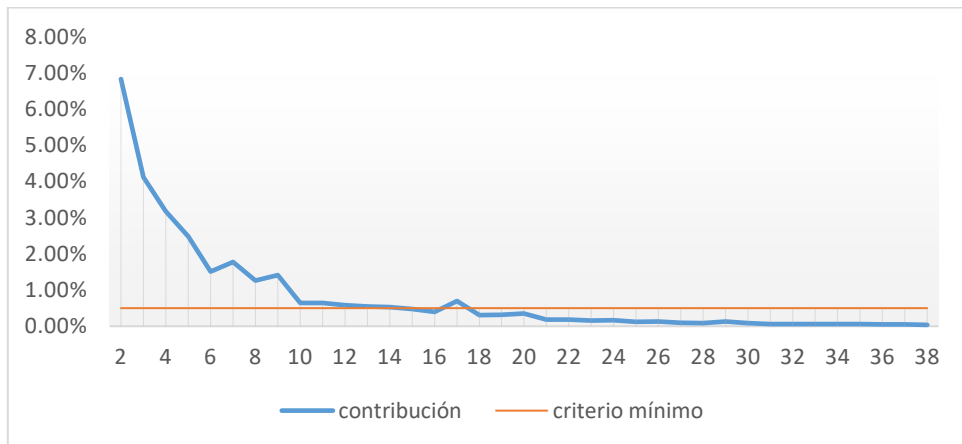
- Segmento 3.1



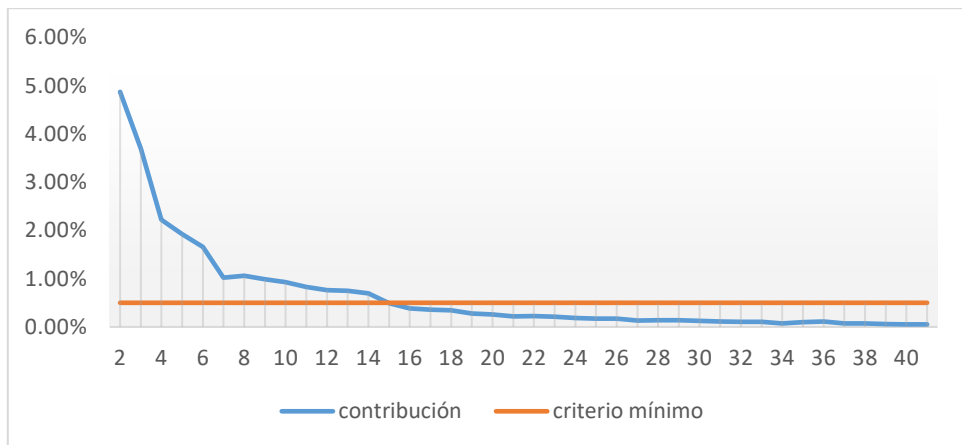
- Segmento 3.2



- Segmento 5.1



- Segmento 5.2





## Anexo 7: Detalle de pasos para modelamiento de modelos de la propuesta N°3

En primer lugar, para cada segmento se realizó el análisis univariado para las 1.446 variables. A continuación, se observa el número de variables tramificadas para cada uno de los segmentos.

Segmento	Número de variables tramificadas
1	1.032
2	1.109
3	1.121
4	1.266
5	1.120
6	1.141
7	715

Con las variables ya tramificadas, se procedió con cálculo del valor WOE para cada tramo de cada variable. Luego, para cada variable se calculó su *Information Value*. Posteriormente, para cada segmento se elaboró el ranking de *Information Value*. Del ranking construido, se descartó las variables que posean un IV menor a 2, pues significa que su poder de discriminancia es bajo. En la siguiente tabla se aprecia el número de variables para cada segmento luego de la eliminación por IV.

Segmento	Número de variables posterior a eliminación
1	935
2	1.047
3	1.092
4	1.183
5	1.112
6	1.054
7	715

El siguiente paso que se realizó fue el análisis de correlaciones, con el fin de eliminar las variables que sean redundantes para el modelo. El resultado de este análisis redujo la cantidad de variables en cada segmento, como se observa en la siguiente tabla.

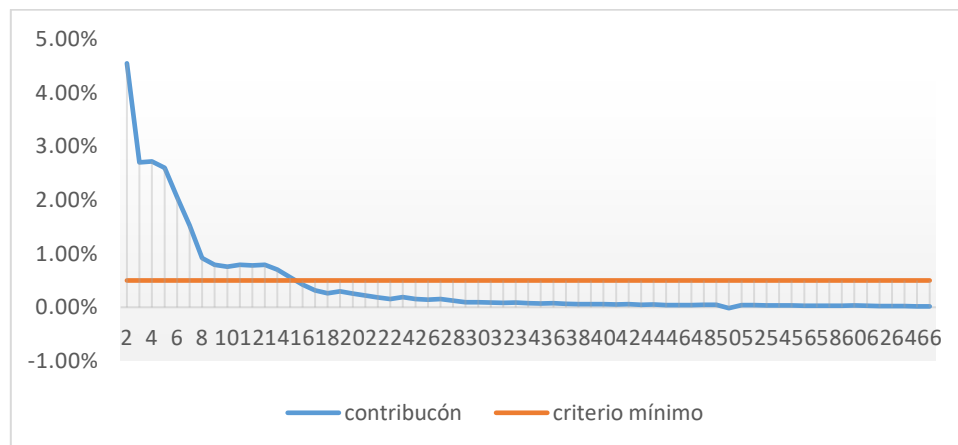
Segmento	Número de variables posterior a análisis de correlación
1	129
2	118
3	94
4	73
5	95
6	144
7	90

Luego, se realizó la regresión logística para cada segmento. El número de variables que se obtuvo como resultado para cada segmento se aprecia a continuación

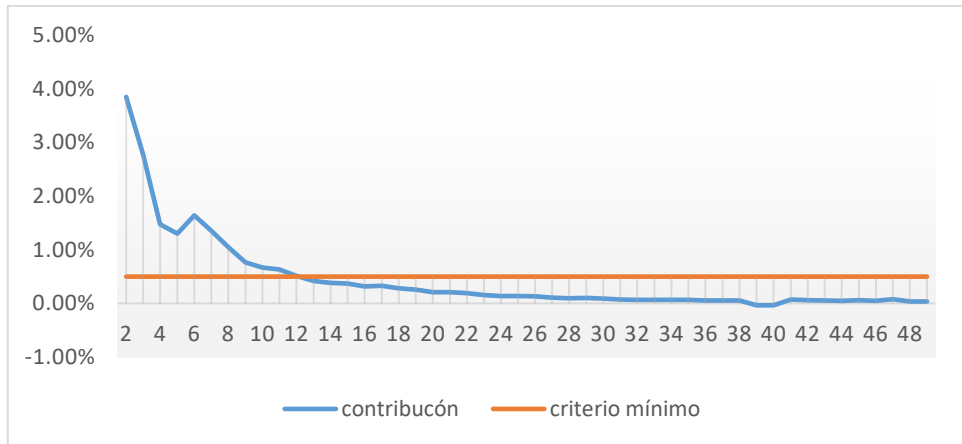
Segmento	Número de variables a incluir
<b>1</b>	66
<b>2</b>	49
<b>3</b>	49
<b>4</b>	34
<b>5</b>	42
<b>6</b>	57
<b>7</b>	41

Se observa en la tabla anterior, que es una gran cantidad de variables para cada segmento, por lo cual sería un modelo difícil de comprender e interpretar. Para reducir las variables se realizó un análisis del aporte a la logverosimilitud, como se explica en el capítulo 3.5.4. El resultado del análisis determinó la cantidad de variables a incorporar para el modelo de cada segmento. Los gráficos de la contribución a la logverosimilitud se observan a continuación.

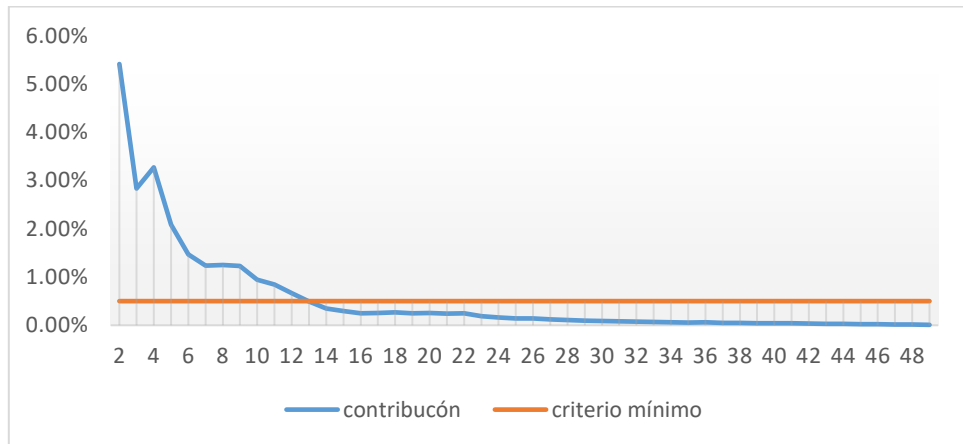
- Contribución a logverosimilitud, segmento 1



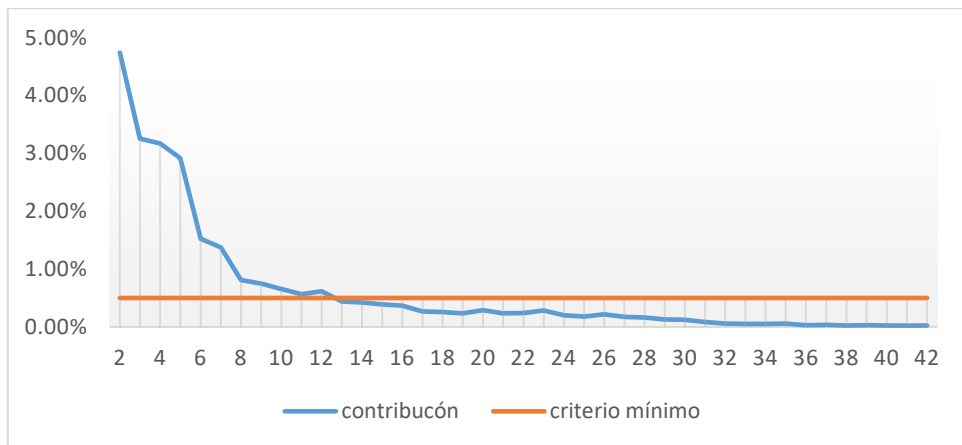
- Contribución a logverosimilitud, segmento 2



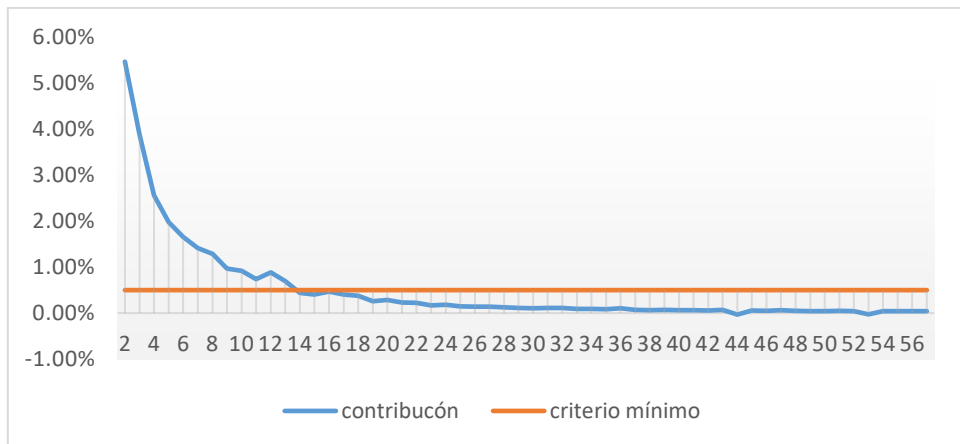
- Contribución a logverosimilitud, segmento 3



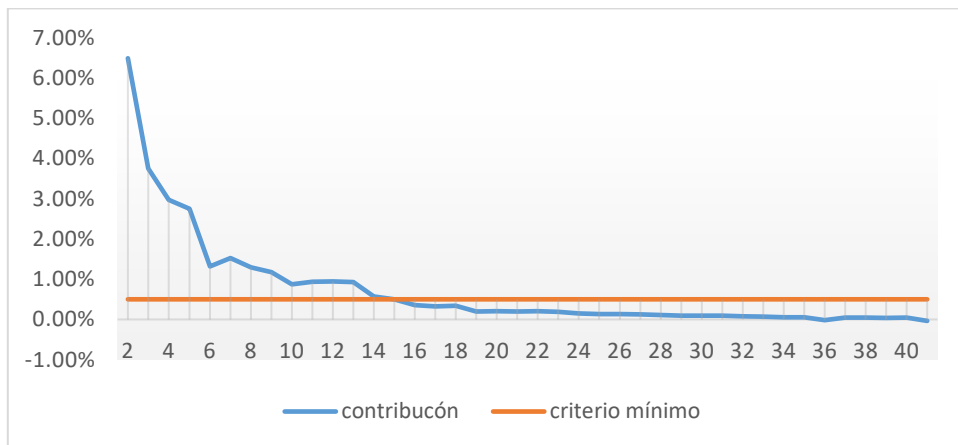
- Contribución a logverosimilitud, segmento 5



- Contribución a logverosimilitud, segmento 6

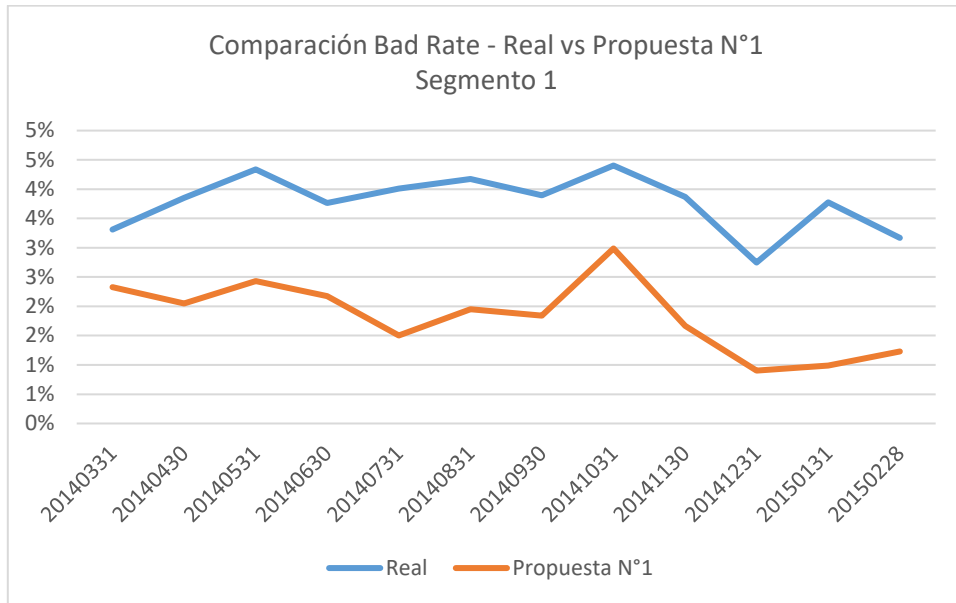


- Contribución a logverosimilitud, segmento 7

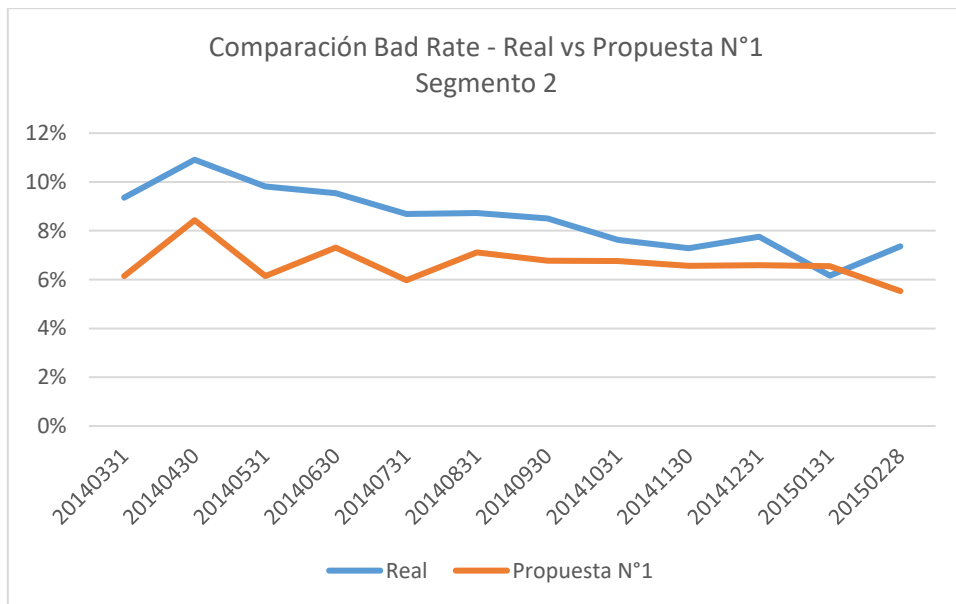


## Anexo 8: Backtest de Bad Rate de propuesta N°1

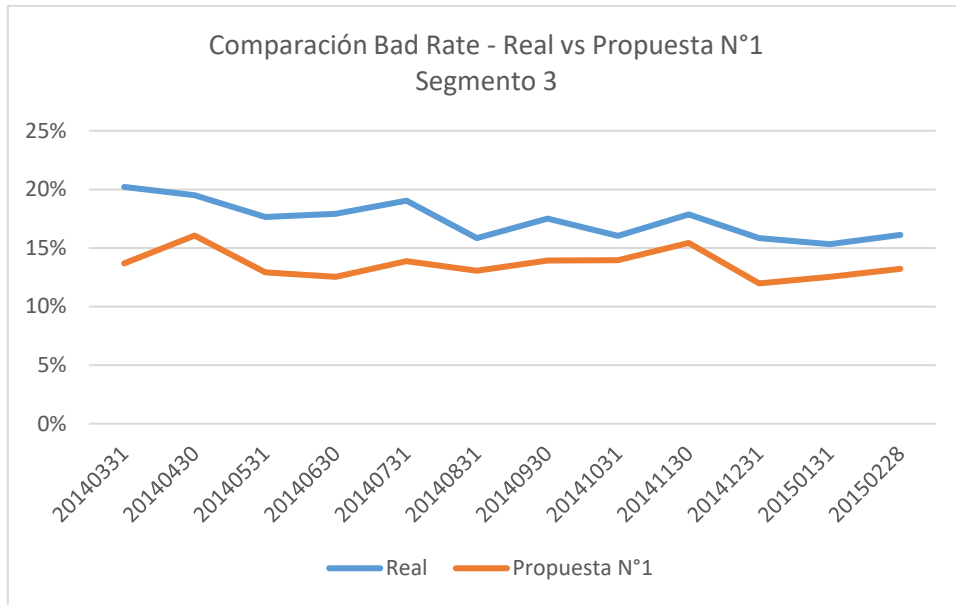
- Segmento 1



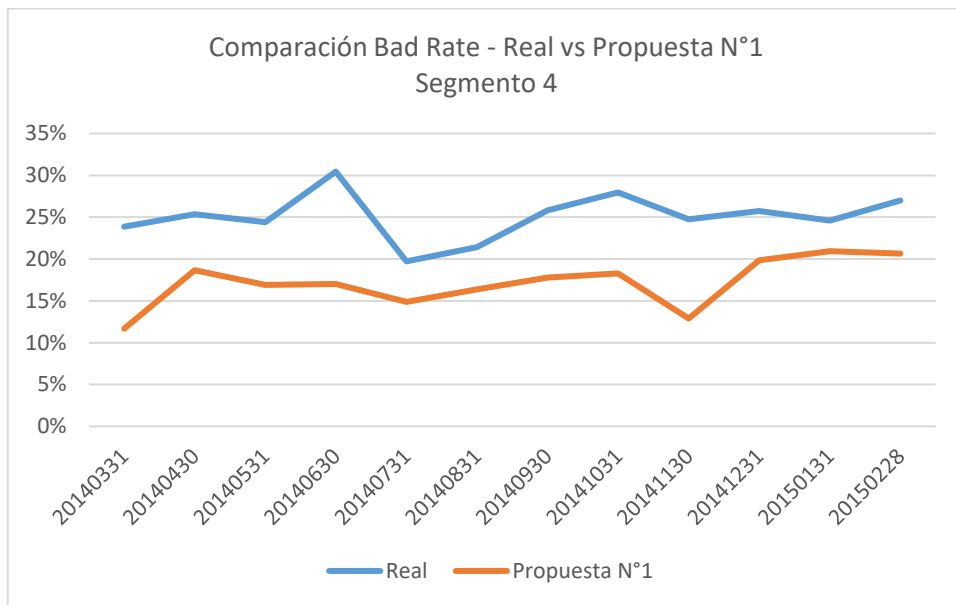
- Segmento 2



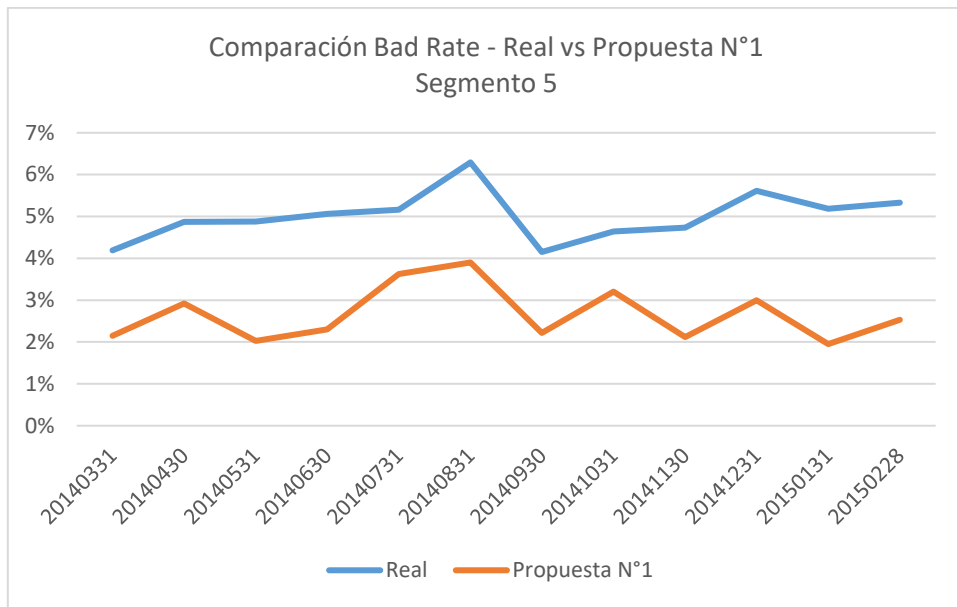
- Segmento 3



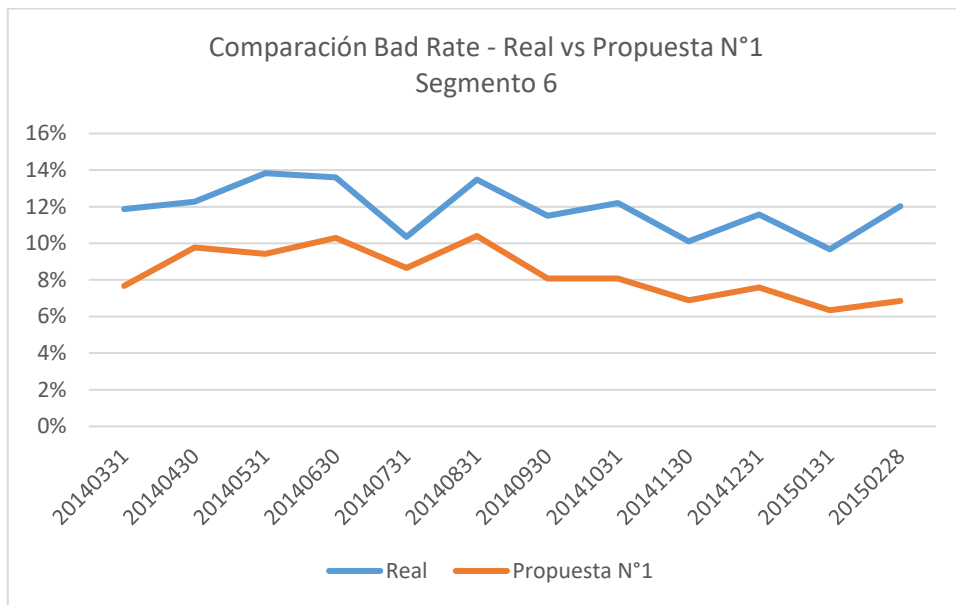
- Segmento 4



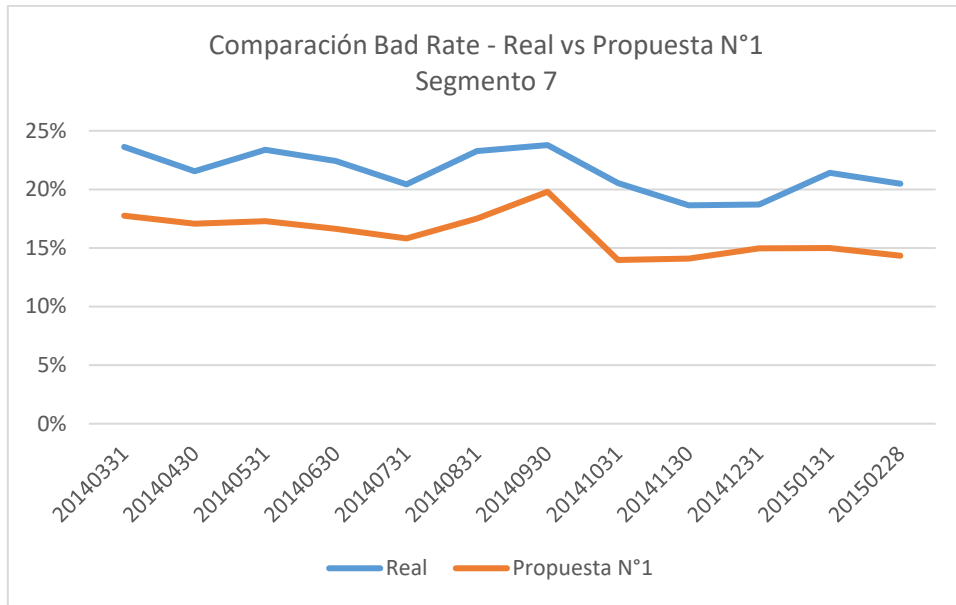
- Segmento 5



- Segmento 6

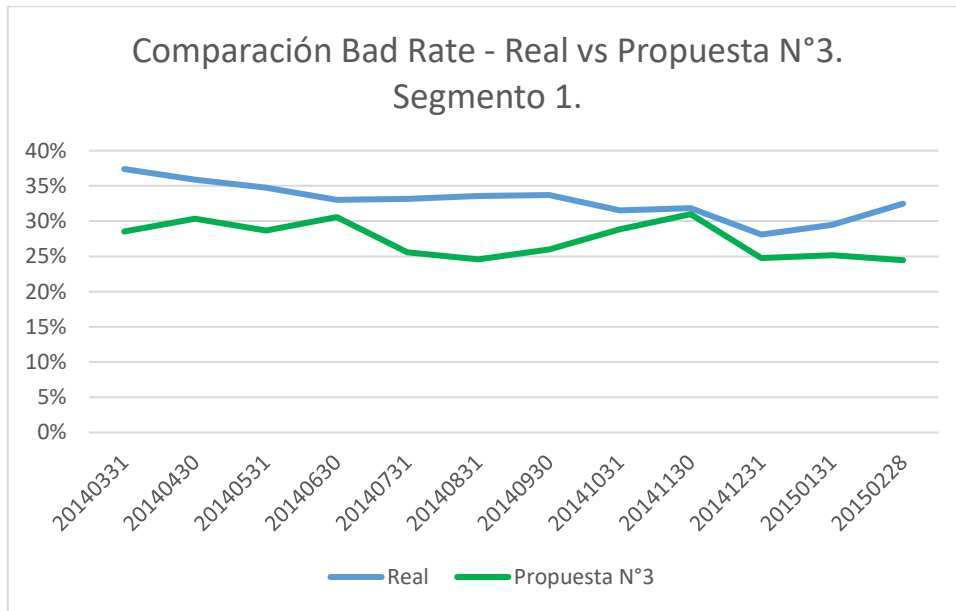


- Segmento 7



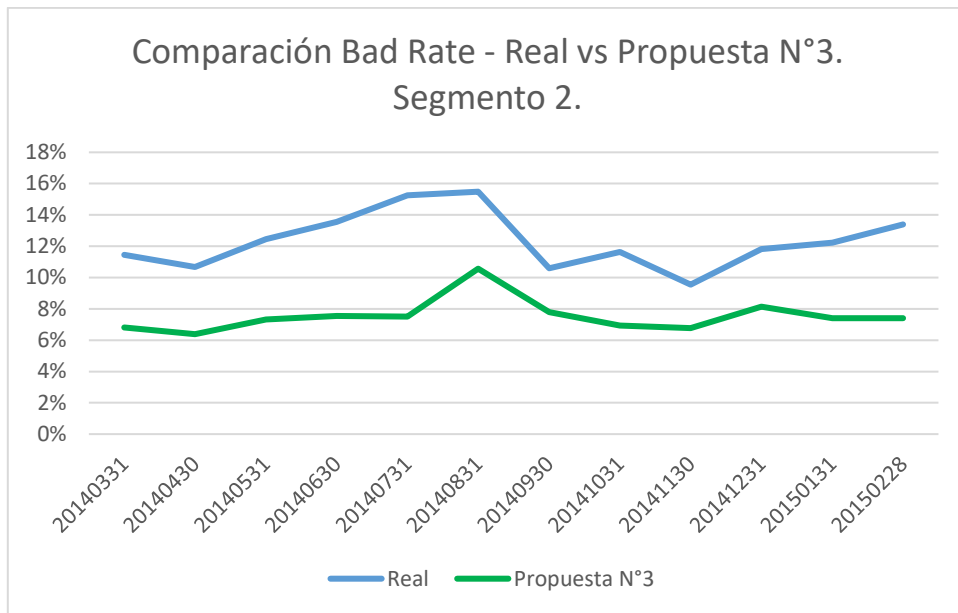
### Anexo 9: Backtest de Bad Rate de propuesta N°3

- Segmento 1

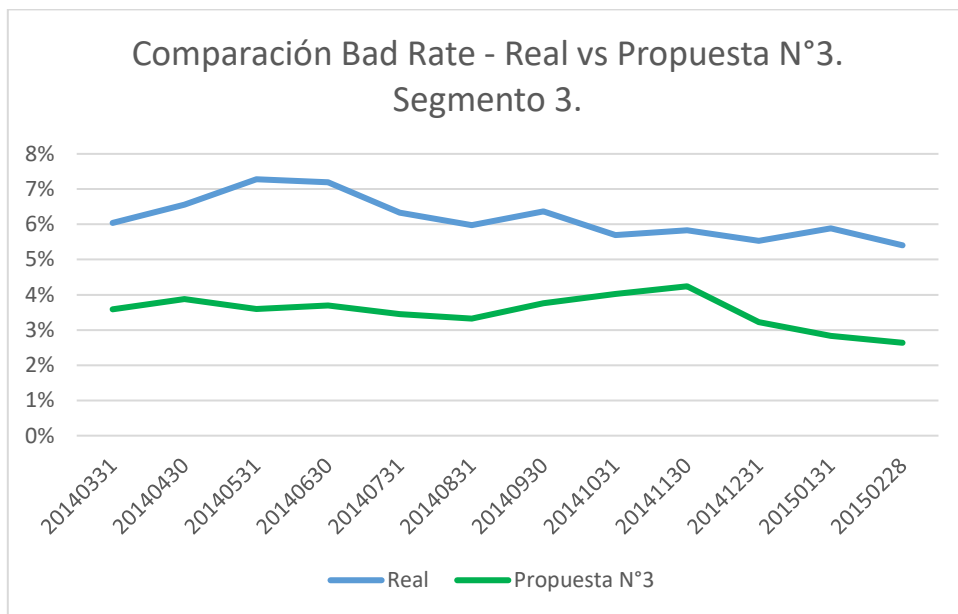




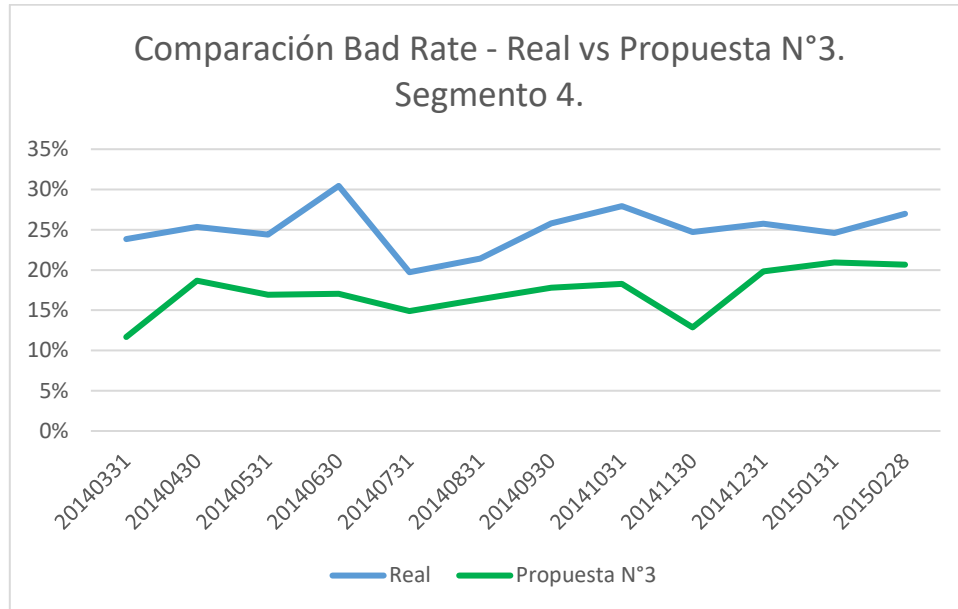
- Segmento 2



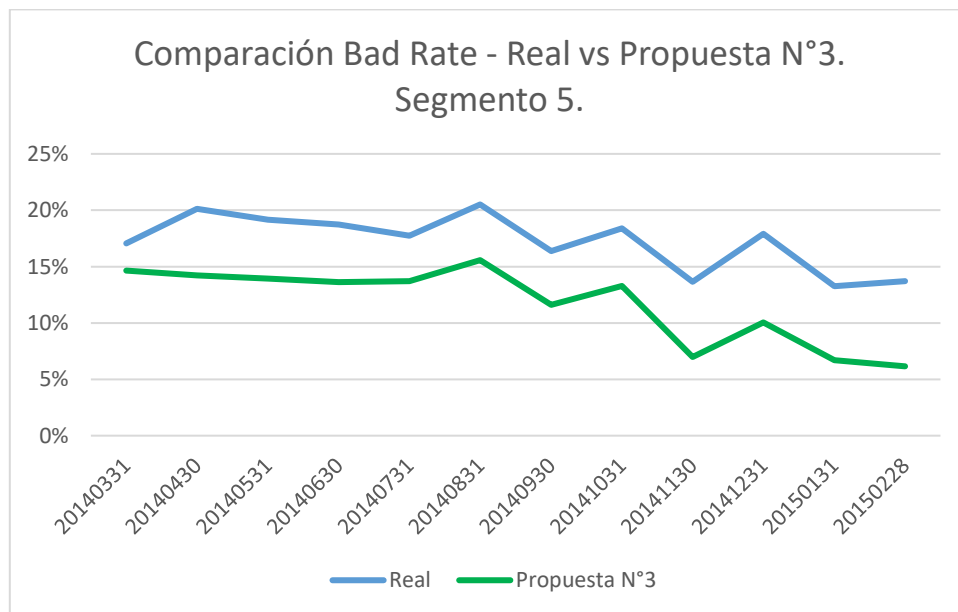
- Segmento 3



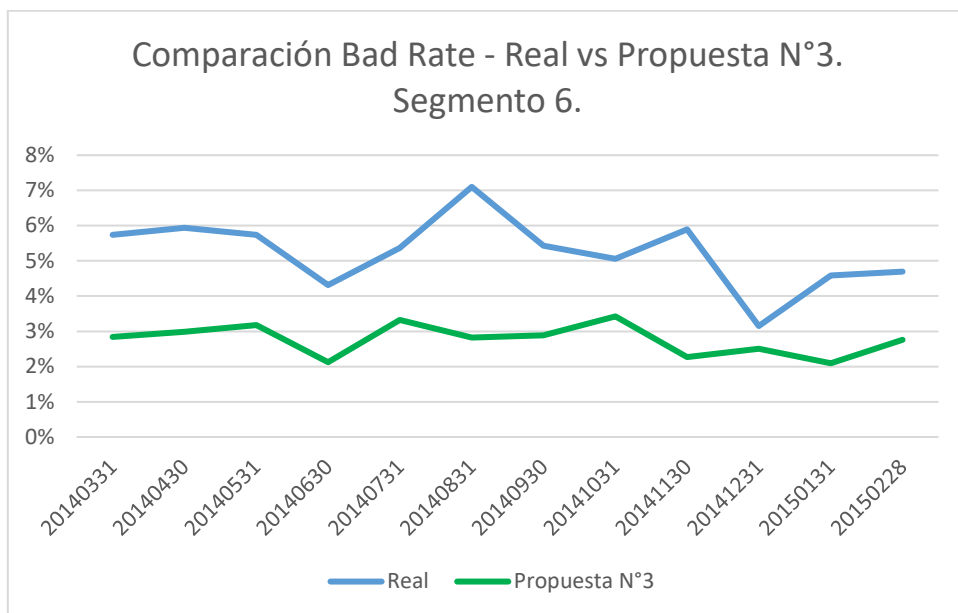
- Segmento 4



- Segmento 5



- Segmento 6



- Segmento 7

