

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



**APLICACIONES DE LA  
CIENCIA DE DATOS PARA  
PROYECTOS DE LA REGIÓN  
NORTE DE AMÉRICA  
LATINA**

CASO  
QUE PARA OBTENER EL GRADO DE  
MAESTRO EN CIENCIA DE DATOS  
PRESENTA  
EDUARDO HIDALGO GARCÍA  
ASESOR  
DR. LEON BERDICHESVSKY ACOSTA

CIUDAD DE MÉXICO

2019

«Con fundamento en los artículos 21 y 27 de la Ley Federal del Derecho de Autor y como titular de los derechos moral y patrimonial de la obra titulada **“APLICACIONES DE LA CIENCIA DE DATOS PARA PROYECTOS DE LA REGIÓN NORTE DE AMÉRICA LATINA”**, otorgo de manera gratuita y permanente al Instituto Tecnológico Autónomo de México y a la Biblioteca Raúl Baillères Jr., la autorización para que fijen la obra en cualquier medio, incluido el electrónico, y la divulguen entre sus usuarios, profesores, estudiantes o terceras personas, sin que pueda percibir por tal divulgación una contraprestación».

5 DE MARZO DE 2019

---

EDUARDO HIDALGO GARCÍA

*Este documento lo dedico a todas las personas que de alguna forma me apoyaron durante el estudio de la maestría y la generación de este trabajo.*

# Agradecimientos

A mi madre por haberme apoyada en la lectura y corrección del documento, así como el apoyo brindado en proyectos relacionados con el fin de la maestría. A mi novia por su apoyo y soporte en el proyecto que representó el estudio de la maestría. A la empresa (Mastercard) y particularmente a las personas que trabajan en ella, en específico a aquellas que de alguna forma contribuyeron en la generación de los casos de estudios expuestos.

Además, agradescos la revisión y comentarios de los doctores :

1. Leon Berdichevsky Acosta
2. Juan Salvador Marmol Yahya
3. Juan Carlos Martinez Ovando

# Resumen

El presente documento resume la actividad del autor en el período que abarcó del primero de junio de 2016 al 31 de mayo de 2018 dentro del área de *Loyalty Analytics* en Mastercard para *Geo-North* (zona norte de américa latina). En el presente documento se presentan seis proyectos de gran relevancia para el área de lealtad cuya información ha sido enmascarada o anonimizada, por cuestiones de confidencialidad o por llegar a constituir un secreto industrial de Mastercard o de sus clientes. No obstante, aún bajo esta circunstancia, el documento muestra como la ciencia de datos permite combinar la aplicación de técnicas estadísticas avanzadas y la generación de algoritmos para el procesamiento de grandes cantidades de información para el modelado de cuestiones cotidianas en el negocio, con el objeto de generar más oportunidades y permitir una mejor toma de decisiones. Cada proyecto se expone en el mismo formato: Objetivo, Contexto, Herramientas de Ciencia de Datos y Lenguajes de Programación Utilizados e Impacto para el Negocio. Los proyectos aquí expuestos fueron desarrollados por el autor y son propiedad de Mastercard International Incorporated, se presentan en este documento exclusivamente para fines académicos.

# Índice general

<b>1. Metodología para el Análisis del <i>ROI</i> de campañas de comunicación</b>	<b>1</b>
1.1. Objetivo . . . . .	1
1.2. Contexto . . . . .	1
1.3. Estructura y Contenido . . . . .	2
1.4. Herramientas de Ciencia de Datos y Lenguajes de Programación Utilizados . . . . .	11
1.5. Impacto para el negocio . . . . .	11
<b>2. Modelo de Rentabilidad Recompensas Crédito</b>	<b>12</b>
2.1. Objetivo . . . . .	12
2.2. Contexto . . . . .	12
2.3. Estructura y Contenido . . . . .	14
2.4. Herramientas de Ciencia de Datos y Lenguajes de Programación Utilizados . . . . .	18
2.5. Impacto para el negocio . . . . .	19
<b>3. Modelo de Rentabilidad Recompensas Débito</b>	<b>20</b>
3.1. Objetivo . . . . .	20
3.2. Contexto . . . . .	20

3.3.	Estructura y Contenido . . . . .	22
3.4.	Herramientas de Ciencia de Datos y Lenguajes de Programación Utilizados . . . . .	26
3.5.	Impacto para el negocio . . . . .	27
<b>4.</b>	<b>Procesamiento de datos para reportes personalizados</b>	<b>28</b>
4.1.	Objetivo . . . . .	28
4.2.	Contexto . . . . .	28
4.3.	Estructura y Contenido . . . . .	29
4.4.	Herramientas de Ciencia de Datos y Lenguajes de Programación Utilizados . . . . .	29
4.5.	Impacto para el negocio . . . . .	30
<b>5.</b>	<b>Reporte Interactivo de <i>KPI's</i></b>	<b>31</b>
5.1.	Objetivo . . . . .	31
5.2.	Contexto . . . . .	31
5.3.	Estructura y Contenido . . . . .	32
5.4.	Herramientas de Ciencia de Datos y Lenguajes de Programación Utilizados . . . . .	40
5.5.	Impacto para el negocio . . . . .	41
<b>6.</b>	<b>Segmentaciones <i>ad hoc</i></b>	<b>42</b>
6.1.	Objetivo . . . . .	42
6.2.	Contexto . . . . .	42
6.3.	Estructura y Contenido . . . . .	43
6.4.	Herramientas de Ciencia de Datos y Lenguajes de Programación Utilizados . . . . .	45
6.5.	Impacto para el negocio . . . . .	45
<b>A.</b>	<b>Más estimaciones del ATET</b>	<b>46</b>

<b>B. Ciclo de la información</b>	<b>54</b>
B.0.1. Extracción de la información . . . . .	54
B.0.2. Toma de decisiones . . . . .	55
<b>Referencias</b>	<b>56</b>

# Capítulo 1

## Metodología para el Análisis del *ROI* de campañas de comunicación

### 1.1. Objetivo

Generar una metodología estandarizada que permita:

1. Conocer la rentabilidad de una campaña de comunicación en específico
2. Comparar la rentabilidad de forma fácil entre distintas campañas

### 1.2. Contexto

En el contexto del programa de recompensas, los clientes<sup>1</sup> tenían la necesidad de conocer cuál había sido el impacto de ciertos esfuerzos de

---

<sup>1</sup>Las instituciones dueñas del programa, principalmente bancos

comunicación que habían llevado a cabo para modificar el comportamiento de sus tarjetahabientes (ej: compras en ciertos comercios específicos, redención de algún canal en particular, entre otros). En este sentido, también tenían la necesidad de comparar entre distintas iniciativas y usar esta información para una mejor toma de decisiones. Dependiendo del contexto, el análisis para medir el impacto estuvo basado en técnicas que permitieran medir el efecto de tratamiento promedio en el grupo tratado (ATET por sus siglas en inglés) tales como:

1. Análisis de Diferencias en Diferencias junto con técnicas de matching y selección aleatorio de grupos
2. Generación de grupo de control estadísticamente comparable con modelos ARIMA (por ejemplo el que genera la librería de R *CausalImpact*)

Por último, para tener un mecanismo estandarizado para la comparación entre campañas se desarrolló un formato en Excel que permite una visualización estándar de la información al imprimirla en PDF.

### **1.3. Estructura y Contenido**

Para la presentación del proyecto se llevaron a cabo las siguientes etapas:

1. Estudio estadístico para estimar el impacto causal (*ATET*)
2. Llenado del formato desarrollado en Excel para imprimir en PDF

## Efecto Causal

Eduardo Hidalgo

June 12, 2016

### Análisis Campaña MTR Semana Santa 2018

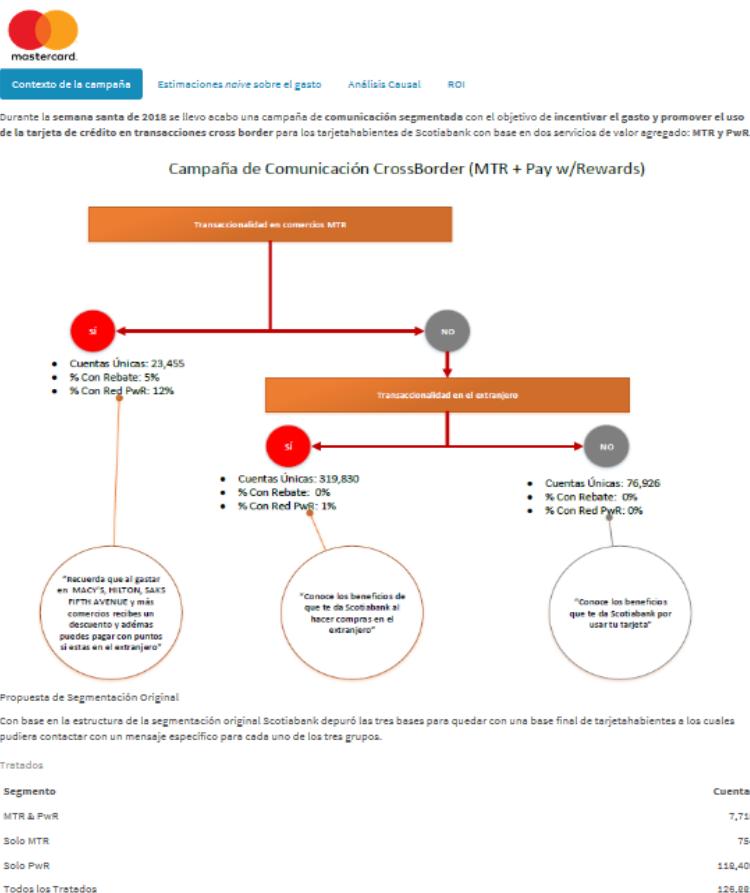


Figura 1.1: Planeación de la campaña

# Efecto Causal

*Eduardo Hidalgo*

*June 12, 2018*

## Análisis Campaña MTR Semana Santa 2018



Contexto de la campaña

Estimaciones *naïve* sobre el gasto

Análisis Causal

ROI

En un primer acercamiento se hace uso de la *aleatoriedad en la asignación de la campaña*. La *aleatoriedad en la asignación de la campaña* se basa en la *depuración* hecha a la *base original* propuesta, donde se eligió tarjetahabientes con características similares.

Estimaciones sin matching por propensity score en un período de 80 días

Tipo de Estimador	Impacto en Gasto Promedio por cuenta
A/B testing	12,442
Diferencia en Diferencias	17,798

El resultado anterior muestra el impacto sobre el *gasto por cuenta* durante el *periodo* de 80 días que van del **2018-03-20** al **2018-06-08**. Y, es el resultado de comparar el promedio del gasto por cuenta del grupo seleccionado para la campaña de MTR y que recibió la comunicación contra el promedio del gasto por cuenta del grupo seleccionado para la campaña de MTR y que *NO* recibió la comunicación.

Durante esos **80 días** se estima un impacto en el gasto promedio diario por cuenta de

Estimaciones por día sin matching por propensity score

Tipo de Estimador	Impacto en Gasto Promedio por cuenta
A/B testing	156
Diferencia en Diferencias	222

Figura 1.2: Estimaciones del efecto en el gasto (parte 1)

# Efecto Causal

Eduardo Hidalgo

June 12, 2018

## Análisis Campaña MTR Semana Santa 2018



Contexto de la campaña    Estimaciones naïve sobre el gasto    Análisis Causal    ROI

Para lograr un **análisis causal** del impacto de la campaña sobre el gasto promedio por cuenta es necesario **equilibrar** las diferencias pre-existentes antes del inicio de la campaña.

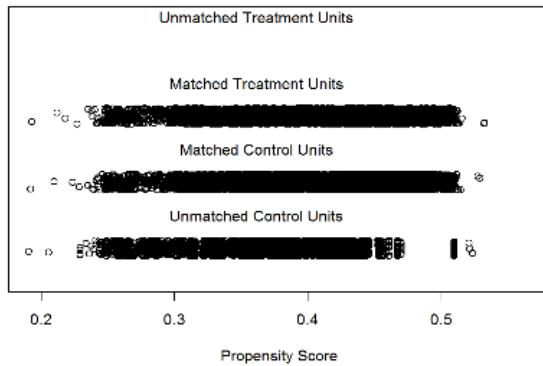
Comenzamos por entender el **balance promedio** de los covariantes dentro del grupo de tarjetahabientes seleccionados en la base original para la comunicación de MTR.

```
## Multivariate Imbalance Measure: LI=0.820
## Percentage of local common support: LCS=11.2%
##
## Univariate Imbalance Measures:
##
##          statistic      type    L1 min 25% 50% 75% max
## Edad      -0.9192585 (diff) 0.079827954  -1   -1   -1   -1   -6
## Tenure     -0.3885274 (diff) 0.054693656     0   0   0   0   0
## Uso        -0.3564533 (diff) 0.053987412     0   0   0   -1   0
## Redenciones -0.1982310 (diff) 0.066090270     0   0   0   0   186
## PROGRAM_NAM 371.0996503 (Chi2) 0.075594094 NA NA NA NA NA
## STATUS      0.9098817 (Chi2) 0.001028969 NA NA NA NA NA
## Trxs        -3.3697592 (diff) 0.037725563     0   -1   -3   -7  -662
```

Como se puede ver en el resultado anterior, la metodología *naïve* produce resultados imprecisos debido a que ya **existía una diferencia** en el promedio de los covariantes antes del inicio de la campaña. De hecho, el imbalance promedio de éstos (con base en si fueron comunicados o no lo fueron) es de **82%**.

Por lo anterior, se busca seleccionar dentro del grupo de control a aquellos tarjetahabientes que contaran con niveles promedio semejantes en los covariantes a los que presenta el grupo tratado. Para esto se ocupa el proceso de *matching* por *propensity score*.

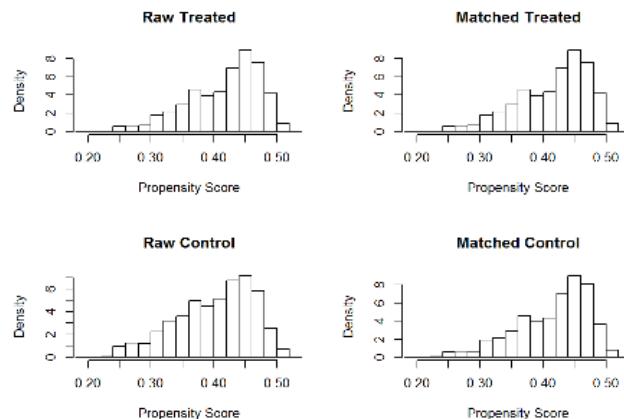
Figura 1.3: Prueba de Imbalance



La grafica muestra el resultado del *matcheo* donde se busca que el total de las unidades tratadas sean representadas por las unidades más semejantes del grupo de control.

Los métodos de **matching** pueden verse como ejercicios de estimación por etapas, pues en una primer etapa se ocupan los **covariables** seleccionados para estimar la probabilidad de ser tratado (`pscore`) y generar unos pesos para ponderar a cada una de las variables de interés al agrupar. En la segunda etapa, se utilizan los parámetros estimados en la etapa anterior para hacer el agrupamiento o selección de las unidades de la base original.

La documentación completa de la librería `MatchIt` que se utilizó para estos resultados se encuentra en:  
<https://cran.r-project.org/web/packages/MatchIt/MatchIt.pdf>



Podemos ver que después del proceso de *matcheo* ambos grupos tienen histogramas de distribución de la probabilidad de ser comunicados/tratados similares.

Figura 1.4: Generación de grupo control

```

## 
## Paired t-test
## 
## data: m.data0$GastoAnual[m.data0$treatment == 1] and m.data0$GastoAnual[m.data0$treatment == 0]
## t = 2.2245, df = 8176, p-value = 0.02614
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   2482.688 39313.165
## sample estimates:
## mean of the differences
##                      20897.93

```

Sin embargo, aún después de haber realizado un proceso de *matching* por propensity score no se logra reducir la diferencia en el promedio de gasto por cuenta entre el **grupo tratado** y el **grupo de control**. La diferencia promedio es de 20,920.75 con un intervalo de confianza de [2,512 – 39,329]. Dicha diferencia es estadísticamente significativa.

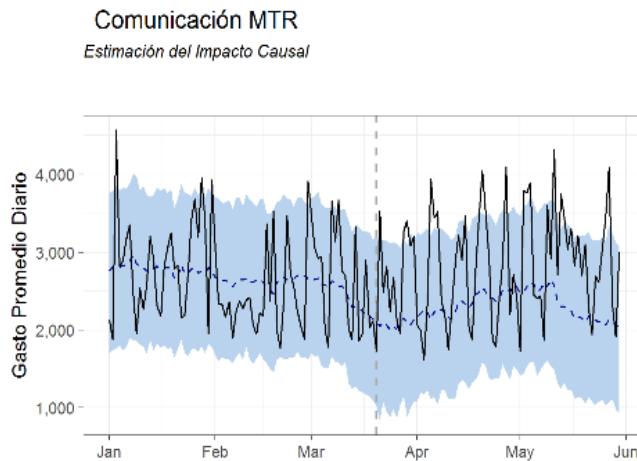
Es por lo anterior que se decidió utilizar la librería **CausalImpact**, la cual realiza inferencia causal a través de generar predicciones para el **contrafactual** (cómo hubiera sido el gasto promedio por cuenta diario en ese periodo de no haber recibido la comunicación).

La descripción completa de la mecánica de estimación así como de los supuestos para la predicción se encuentra en:  
<http://google.github.io/CausalImpact/>

En términos generales lo que la librería hace es asumir que la serie de tiempo de las unidades tratadas puede ser explicada por una serie de covariantes (en nuestro caso: **Edad**, **Tenure**, **Uso**, **Redenciones**, **PROGRAM\_NAM**, **STATUS**, **Trxs**) que no son afectados por la intervención cuyo efecto causal nos interesa medir.

Figura 1.5: Resultados del proceso de 'matcheo'

A continuación los resultados de la estimación para el gasto promedio diario por cuenta con este método estocástico (en inglés):



#### Analysis report [CausalImpact]

During the post-intervention period, the response variable had an average value of approx. 2.82K. By contrast, in the absence of an intervention, we would have expected an average response of 2.29K. The 95% interval of this counterfactual prediction is [1.98K, 2.58K]. Subtracting this prediction from the observed response yields an estimate of the causal effect the intervention had on the response variable. This effect is 0.53K with a 95% interval of [0.28K, 0.84K]. For a discussion of the significance of this effect, see below.

Summing up the individual data points during the post-intervention period (which can only sometimes be meaningfully interpreted), the response variable had an overall value of 200.22K. By contrast, had the intervention not taken place, we would have expected a sum of 162.59K. The 95% interval of this prediction is [140.30K, 181.87K].

The above results are given in terms of absolute numbers. In relative terms, the response variable showed an increase of +23%. The 95% interval of this percentage is [+11%, +37%].

This means that the positive effect observed during the intervention period is statistically significant and unlikely to be due to random fluctuations. It should be noted, however, that the question of whether this increase also bears substantive significance can only be answered by comparing the absolute effect (0.53K) to the original goal of the underlying intervention.

The probability of obtaining this effect by chance is very small (Bayesian one-sided tail-area probability  $p = 0.001$ ). This means the causal effect can be considered statistically significant.

Figura 1.6: Efecto de Tratamiento Promedio en el grupo tratado

# Efecto Causal

Eduardo Hidalgo

June 12, 2018

## Análisis Campaña MTR Semana Santa 2018



Contexto de la campaña    Estimaciones naïve sobre el gasto    Análisis Causal    **ROI**

En el análisis de la sección anterior, se estimó un impacto sobre el gasto promedio diario por cuenta de 530 mxn. Y, un agregado de 37,360 mxn.

A continuación se presentan las estimaciones para el ROI de esta iniciativa con base en:

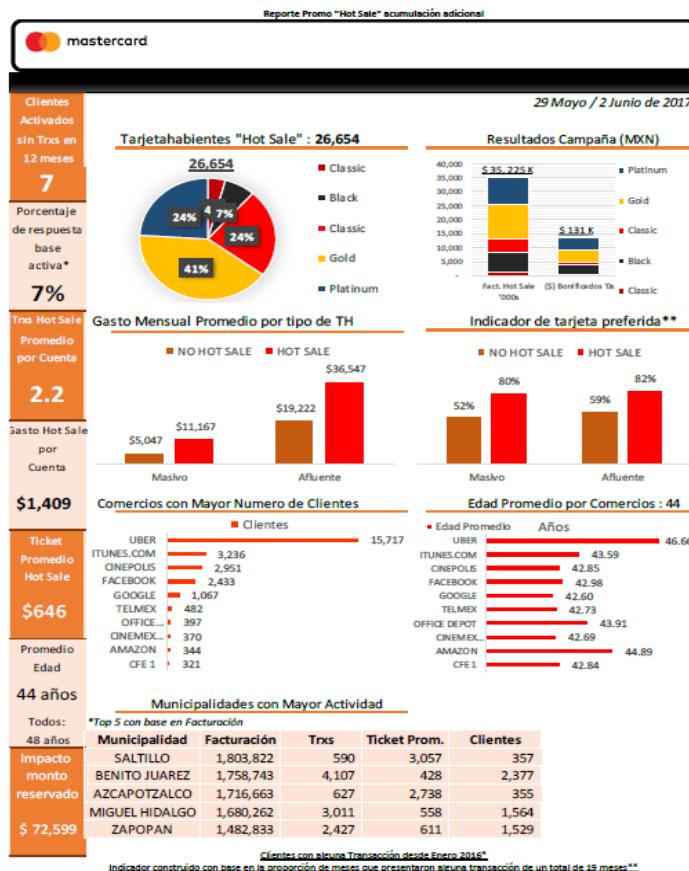
- Las estimaciones de la sección anterior para el **gasto promedio diario por cuenta**
- Supuestos sobre:
  - Costos del envío de la comunicación.- Se asume que el **costo total** del mensaje para comunicar la campaña fue de 1 mxn
  - Ingresos por Facturación incremental.- Se asume que la **tasa de intercambio** promedio es de 1.5%, la **tasa de interés** es del 2.3%, el **tiempo promedio de revolvencia** es de **6 meses** y el **porcentaje de revolvencia** es del **45%**

Estimaciones del ROI de la campaña x Cliente

Tipo de ROI	ROI
Por Dia por Cliente	40
Total x Cliente	2,900

En conclusión el **retorno a la inversión** para Scotiabank de la **campaña de comunicación de semana santa** fue de **40 mxn por dia por cliente**, para un **retorno a la inversión total por cliente** de la iniciativa de **2,900 mxn**

Figura 1.7: Rentabilidad de la campaña



## 1.4. Herramientas de Ciencia de Datos y Lenguajes de Programación Utilizados

- Estudio del *lift* en facturación
  - Software y Estructuras
    - Oracle
    - RStudio
  - Lenguajes de Programación
    - SQL
    - R<sup>2</sup>
- Modelaje del P&L del programa de recompensas
  - Software y Estructuras
    - Excel
  - Lenguajes de Programación
    - VBA
- Presentación y visualización de la información
  - Software y Estructuras
    - PowerPoint

## 1.5. Impacto para el negocio

Este proyecto permitió incentivar campañas de comunicación específicas para estimular el consumo o promover la redención en canales específicos.

---

<sup>2</sup>Para establecer la conexión entre Oracle y R se hace uso de la librería RJDBC[2]

## Capítulo 2

# Modelo de Rentabilidad Recompensas Crédito

### 2.1. Objetivo

Demostrar el valor agregado que generan los programas de recompensas en el sector de tarjetas de crédito. Lo anterior se logra al modificar la actitud del tarjetahabiente mediante un esquema de incentivos adecuado que modifique sus acciones hacia aquellas que le sean más rentables al banco.

### 2.2. Contexto

En el país, la idea de tener un programa de recompensas en tarjetas de crédito no es nueva. De hecho, la mayor parte de los jugadores dentro del mercado de tarjetas de crédito ya cuentan con uno de algún tipo (ej. cupones, descuentos, puntos). Sin embargo, para la mayor parte de ellos, los programas de recompensas son vistos como un *commodity*, o

elemento que deben tener dentro de la oferta de valor de su producto sólo porque la competencia también lo tiene. De hecho, consideran a estos programas como un costo más que impacta directamente al estado de resultados (P&L)<sup>1</sup> del portafolio de crédito sin generar ninguna clase de ingresos.

Con base en lo descrito en el párrafo anterior, el proyecto consistió en generar una historia. La cual, respaldada por datos, lograra sustentar la idea de que los programas de recompensas, para el portafolio de crédito, son una inversión. Y, el beneficio se puede mostrar al comparar el comportamiento que en términos de gasto tienen los siguientes grupos de tarjetahabientes:

1. Tarjetahabientes sin programa de recompensas
2. Tarjetahabientes con programa de recompensas pero sin redención
3. Tarjetahabientes con programa de recompensas y con redención

Además, mediante el uso de datos se mostró la veracidad de las siguientes hipótesis:

1. Los programas de recompensas no son iguales a lealtad. La lealtad consiste en todas las formas en las que una institución busca modificar el comportamiento de sus clientes y no exclusivamente de un programa de recompensas
2. El programa de recompensas y todos los acercamientos para estimular la lealtad del tarjetahabiente deberían de ser utilizados a lo largo de todo el ciclo de vida del tarjetahabiente.

---

<sup>1</sup>En el contexto de este análisis se usa el concepto P&L (*Profit & Loss*) de manera intercambiable con el de estado de resultados. Y, se define como un ejercicio contable que consiste en mostrar la estructura de costos e ingresos detalladamente.

## 2.3. Estructura y Contenido

Este proyecto se expuso al cliente en forma de presentación. Sin embargo, la generación del modelo de rentabilidad y la exposición al cliente se realizaron en tres etapas:

1. Estudio estadístico del impacto o *lift* por concepto del programa y/o la redención de puntos dentro del mismo
2. Modelo del estado de resultados (P&L) del cliente que incluyera el impacto que en términos de revenue tiene el *lift* estudiado en la primera etapa
3. Generación de una presentación ejecutiva que permitiera contar la historia antes mencionada

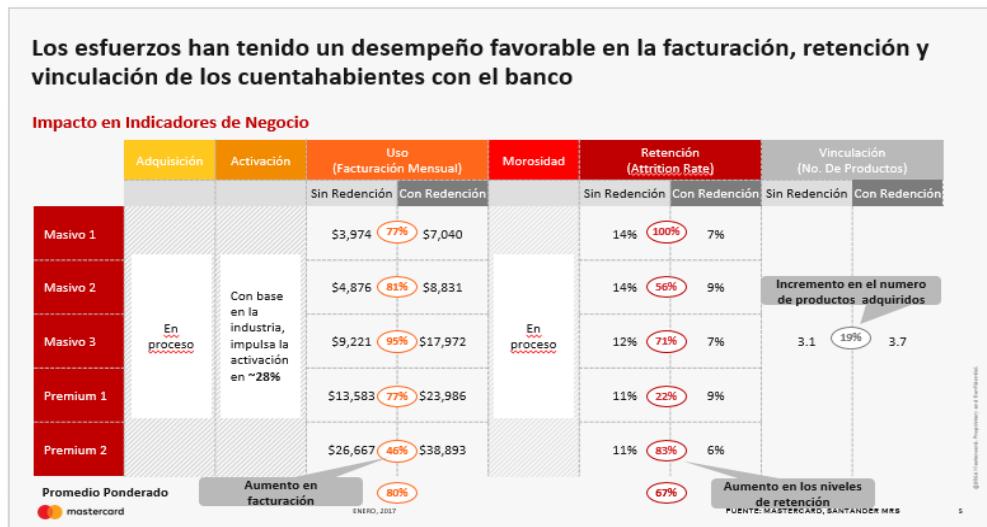


Figura 2.1: Impacto del programa en todo el ciclo de vida del tarjetahabiente

El ciclo de vida del tarjetahabiente lo definimos como todas las partes del proceso de interacción de un tarjetahabiente con su banco desde que adquiere el primer producto, a saber: Adquisición, Activación, Uso, Morosidad, Retención y Vinculación.

Para cada una de las etapas antes mencionadas se realizaron pruebas<sup>2</sup> para medir el impacto de alguno de los elementos de lealtad asociado a esa etapa.<sup>3</sup>

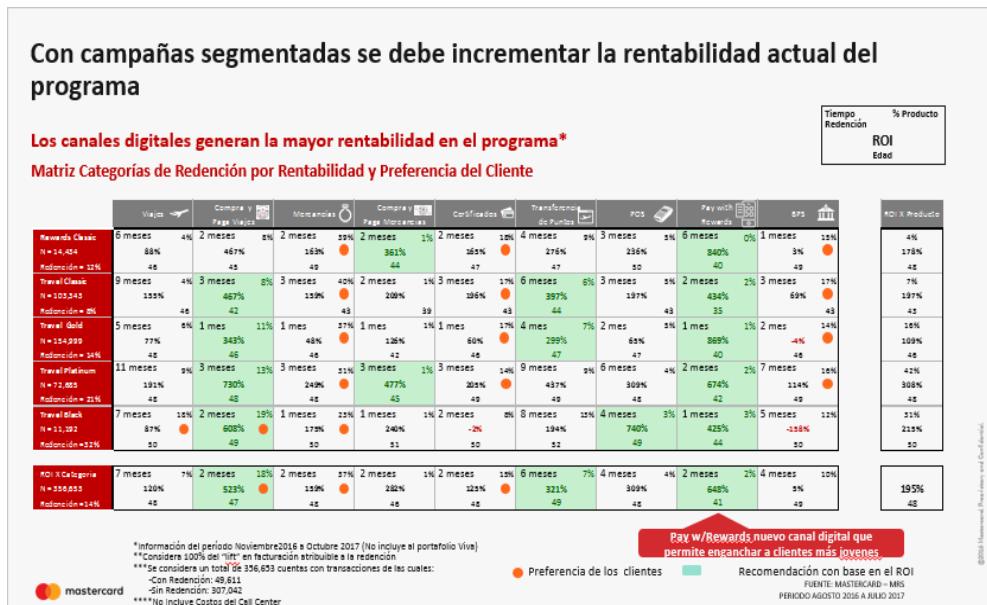


Figura 2.2: Rentabilidad por “lift” en gasto, por categoría de redención y producto

La estrategia recomendada es optimizar los ingresos incrementales.

<sup>2</sup>A | B testing y diferencia en diferencias

<sup>3</sup>Por ejemplo: durante la activación se comparó el tiempo promedio entre la activación del plástico y el primer uso dependiendo si el plástico cuenta con un programa de recompensas.

Para hacerlo es necesario incentivar la redención en las categorías que son más redituables para cada programa. En la tabla se muestra la rentabilidad por programa y categoría de redención. En color verde se muestran las interacciones de programa y categoría más rentables, con un círculo de color naranja se marca las tres categorías por programa con mayor número de redenciones.<sup>4</sup>

	2016 Completo	2017 Completo	2018 Completo	2019 Completo	2020 Completo
<b>Cuentas con Rewards</b>	<b>559,827</b>	<b>658,487</b>	<b>763,341</b>	<b>874,062</b>	<b>1,013,243</b>
<b>Cuentas vigentes Credito</b>	<b>500,935</b>	<b>557,025</b>	<b>659,755</b>	<b>764,679</b>	<b>875,745</b>
<b>Cuentas nuevas Credito</b>	<b>58,892</b>	<b>101,462</b>	<b>103,586</b>	<b>109,382</b>	<b>137,498</b>
Flujo Neto de Cuentas Nuevas Mensual	58,892	101,462	103,586	109,382	137,498
Sin Recompensas	-	-	-	-	-
Con Recompensas	58,892	101,462	103,586	109,382	137,498
<b>Facturación Total</b>	<b>22,565,328,751</b>	<b>28,774,398,335</b>	<b>34,819,030,273</b>	<b>41,656,421,936</b>	<b>49,271,637,759</b>
% de la Facturación Incremental	21%	21%	22%	23%	24%
Facturación Incremental	4,750,374,484	6,071,532,896	7,664,430,693	9,502,992,091	11,696,012,346
Saldo Incremental	2,137,668,518	2,732,189,803	3,448,993,812	4,276,346,441	5,263,205,556
<b>Gasto Promedio por Cuenta</b>	<b>40,308</b>	<b>43,698</b>	<b>45,614</b>	<b>47,658</b>	<b>48,628</b>
% Cuentas Redimiendo	2%	2%	2%	2%	2%
<b>Ingresos Incrementales</b>	<b>216,511,740</b>	<b>282,947,081</b>	<b>354,013,964</b>	<b>435,281,314</b>	<b>528,463,712</b>
<b>Costos del Programa</b>	<b>155,502,136</b>	<b>173,467,080</b>	<b>200,195,388</b>	<b>228,734,953</b>	<b>260,960,000</b>
<b>Ingresos Incrementales Totales</b>	<b>61,009,603</b>	<b>109,480,001</b>	<b>153,818,576</b>	<b>206,546,361</b>	<b>267,503,711</b>
<b>ROI</b>	<b>28%</b>	<b>39%</b>	<b>43%</b>	<b>47%</b>	<b>51%</b>
<b>Ratio Costos vs. Facturación Total</b>	<b>0.7%</b>	<b>0.6%</b>	<b>0.6%</b>	<b>0.5%</b>	<b>0.5%</b>
<b>Costo por Cuenta Enrolada</b>	<b>278</b>	<b>263</b>	<b>262</b>	<b>262</b>	<b>258</b>

Figura 2.3: Estado de resultados del programa de recompensas

El estado de resultados se construyó de tal forma que reflejara los insumos que dentro del contexto del programa generan tanto los ingresos incrementales como los costos del programa. Los indicadores más relevantes son el **ROI** y el ratio de costos versus facturación.<sup>5</sup>

<sup>4</sup>Para explotar la comunicación y maximizar las ganancias incrementales es necesario llevar los círculos color naranja a las casillas color verde

<sup>5</sup>En la industria se considera una buena practica que los costos del programa de recompensas no excedan el 1 % de la facturación total del portafolio.

## Actualmente los clientes redentores generan un *lift* incremental que permite cubrir el costo de la redención

### Rentabilidad por Cliente por Impacto en Facturación

	Valores Promedio del mercado	Tipos de Clientes	
		Con Programa Recompensas y Sin Redención	Con Programa Recompensas y Con Redención
<i>Facturación Promedio Anual</i>	\$74,322	\$74,322	\$74,322
<i>Lift de Facturación Anual por Uso**</i>	NA	\$67,849	\$67,849
<i>Lift de Facturación Anual por Ind. de Negocio***</i>	\$25,797	\$25,797	\$25,797
<i>Facturación Total</i>	\$100,119	\$100,119	\$107,988
<b>Ingresos Incrementales Anuales</b>			
Tasa de Interambio	1.5%	NA	\$1,018
Tasa de Interés // Pay down // % de revolución	2.3% // 6 meses // 45%	NA	\$4,122
<b>Ingresos Incrementales Totales</b>		NA	\$5,140
<b>Costos Incrementales Anuales</b>			
Costos del Programa	\$3.14 mxn/mes	\$39	\$39
Costos de la Redención	Variable	-	\$1,508
<b>Costos Incrementales Totales</b>		\$39	\$1,548
<b>Rentabilidad Promedio Anual</b>		NA	\$3,592
Rentabilidad (Ingresos-Costos)		NA	1.44x
ROI			
<b>Ingresos Incrementales Anuales</b>			
Ingreso por indicadores de Negocio	\$2,825	\$2,825	\$2,825
Ingreso Incremental Total	\$2,825	\$2,825	\$2,825
<b>Rentabilidad Promedio Anual</b>			
Rentabilidad (Ingresos-Costos)		\$2,785	\$6,398
ROI		1.34x	2.67x
Total			
<small>*Activación, Retención y Vinculación **Considera 100% del <i>lift</i> en facturación atribuible a la redención ***Incluye el efecto en activación (26% de la facturación anual de la base atribuible), retención (17% de la facturación anual de la base atribuible) y vinculación (19% de la facturación anual de la base)</small>			

FUENTE: MASTERCARD – MBS // BANCOO

©2020 Mastercard. Todos los derechos reservados.

Figura 2.4: Estado de resultados por cuenta del programa de recompensas

El estado de resultados se contruyó usando como ingreso exclusivamente el *lift* en facturación<sup>6</sup>. A nivel de cliente vimos que cuando la rentabilidad se compara contra los clientes que no tienen programa, los clientes con programa tienen 1.3 veces más<sup>7</sup> y los clientes activos en el programa presenta una rentabilidad 2.6 veces mayor.<sup>8</sup>

<sup>6</sup>En el anexo A se muestra un análisis de las distintas técnicas para estudiar este fenómeno.

<sup>7</sup>Considerando los costos del programa.

<sup>8</sup>Considerando los costos del programa más el costo de las mercancías que redimen.

## 2.4. Herramientas de Ciencia de Datos y Lenguajes de Programación Utilizados

La generación del archivo consistió en las tres etapas antes mencionadas:

- Estudio del *lift* en facturación
  - Software y Estructuras
    - Oracle
    - RStudio
  - Lenguajes de Programación
    - SQL
    - R<sup>9</sup>
- Modelaje del P&L del programa de recompensas
  - Software y Estructuras
    - Excel
  - Lenguajes de Programación
    - VBA
- Presentación y visualización de la información
  - Software y Estructuras
    - PowerPoint

---

<sup>9</sup>Para establecer la conexión entre Oracle y R se hace uso de la librería RJDBC[2]

## **2.5. Impacto para el negocio**

Este proyecto modificó la forma en la que los clientes perciben el valor que generan los servicios de lealtad. Además, permitió la implementación de un proyecto de consultoría en términos de lealtad para consolidar todos los esfuerzos, antes descentralizados, que tenía un cliente en particular para modificar el comportamiento de sus tarjetahabientes. El modelo representa una ventaja comparativa para plantear un primer acercamiento con un posible cliente.

## **Capítulo 3**

# **Modelo de Rentabilidad Recompensas Débito**

### **3.1. Objetivo**

Demostrar el valor agregado que generan los programas de recompensas en el sector de tarjetas de débito. La idea clave es la misma que con los productos de crédito, es decir, lograr modificar la actitud del tarjetahabiente mediante un esquema de incentivos adecuado hacia las actitudes que le son más rentables al banco.

### **3.2. Contexto**

Los ingresos generados para el banco en el mercado de tarjetas de débito son mucho menores, que los ingresos generados por las tarjetas de crédito.

De hecho, en este mercado nunca se había visto un programa de recompensas para los productos de débito. Así, el proyecto consistió en

generar una historia que, respaldada por datos, logrará contar la idea de que los programas de recompensas en los productos de débitos son una inversión en la cual el beneficio se puede mostrar al comparar el comportamiento que en términos de gasto y tenencia promedio de saldo tienen los siguientes grupos de tarjetahabientes:

1. Tarjetahabientes sin programa de recompensas
2. Tarjetahabientes con programa de recompensas pero sin redención
3. Tarjetahabientes con programa de recompensas y con redención

Sin embargo, para los productos de débito, dada la estructura de ingresos, se descubrió la existencia de un punto después del cual los beneficios por el cambio de comportamiento de los tarjetahabientes (ej. mayor facturación en TPV) son sobrepasados por los costos variables del programa (ej. costo de la mercancía redimida).<sup>1</sup>

Con base en los supuestos previamente descritos, el objetivo principal de este proyecto fue encontrar la combinación de *tasa de redención óptima* y costo promedio de lo redimido que maximizará los ingresos para el banco por concepto de una mayor facturación en TPV. Ya qué, los generadores de ingreso de este tipo de programas imponen un límite al número de clientes que pueden redimir de tal manera que continúe siendo reddituable para el banco.<sup>2</sup>

---

<sup>1</sup>Como si los ingresos incrementales pudieran ser descritos por una función cóncava cuyo punto máximo es el punto de mayores ingresos. En el eje horizontal se indica el número de clientes con redención. Despues, de este punto a mayor número de clientes con redención menores ingresos incrementales.

<sup>2</sup>En contraste, encontramos que en los programas de crédito la diferencia entre grupos permite al banco sostener que mientras más clientes redentores tenga, mayores serán sus ingresos provenientes de la facturación incremental.

Por último, es importante notar que la idea clave fue encontrar una correlación positiva entre la facturación en TPV y la tenencia de saldo en los productos de débito. A diferencia de los productos de crédito, el programa de recompensas en débito no busca incentivar una mayor facturación, más bien busca modificar la composición de gasto de ATM a TPV. En este sentido, se observó que cuando un tarjetahabiente comienza a tener más transacciones en la TPV aumenta su tenencia promedio de saldo, lo que se puede relacionar con que, para un mismo nivel de ingresos, va al cajero automático (ATM) con menor frecuencia.

### **3.3. Estructura y Contenido**

Este proyecto se expuso al cliente en forma de presentación. Sin embargo, la generación del modelo de rentabilidad y la exposición al cliente se realizaron en tres etapas:

1. Estudio estadístico del impacto o *lift* por concepto del programa y/o la redención de puntos dentro del mismo
2. Modelo del estado de resultados (P&L) del cliente que incluyera el impacto que en términos de revenue tiene el *lift* estudiado en la primera etapa
3. Generación de una presentación ejecutiva que permitiera contar la historia antes mencionada

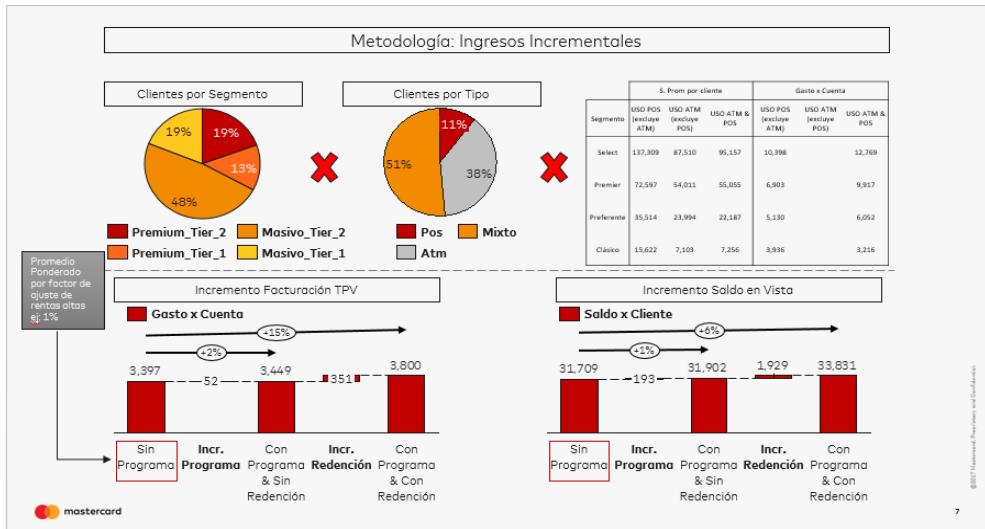


Figura 3.1: Impacto del programa sobre la proporción de gasto en TPV

El modelo controla por el segmento (variable utilizada por el banco para segmentar a sus tarjetahabientes) y el tipo de cliente.<sup>3</sup>

El análisis del comportamiento sobre el saldo promedio mensual se hizo con base en los resultados presentados por un grupo piloto. De la misma manera, se controla por el segmento y tipo de cliente.

<sup>3</sup>Esta variable categórica fue construida para separar a los tarjetahabientes con base en el número de transacciones promedio al mes realizadas en TPV.

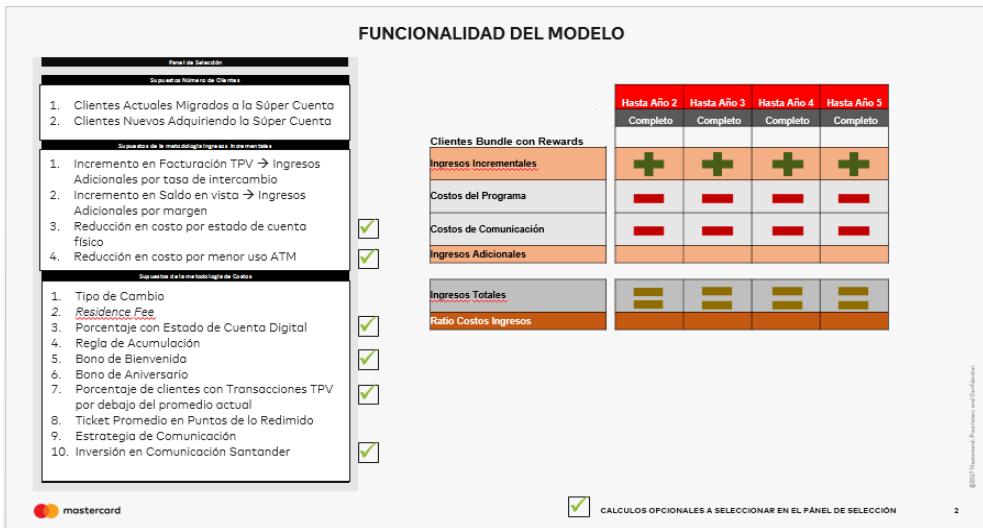


Figura 3.2: Modelo Financiero a cinco años

Además, esta construido para simular escenarios mediante diez selectores relacionados al costo, cuatro selectores relacionados con los ingresos incrementales y dos con el número de clientes a enrolar. Con base en estos selectores el modelo permite simular gran cantidad posibles escenarios.

En la **Figura 3.3**, se observan los resultados del grupo piloto cuando se comparan contra clientes de su mismo segmento, pero que no estaban enrolados en el programa de recompensas. La explicación de este fenómeno<sup>4</sup> es la siguiente: En México por lo regular se observa que los usuarios de productos de débito tienen un número transacciones promedio al mes igual a dos. Esto sucede, pues, la mayor parte retira su dinero cada vez que le es depositado. Por otro lado y a diferencia de los productos de crédito, con los productos de débito el

<sup>4</sup>Independiente de causalidad sobre el programa

## Nota sobre saldo en vista

Ejercicio con datos de MRS y Débito

SEGMENTO	Con Redención	Saldo Vista	Lift	Monto lift
Masivo_Tier_1	NO	5,098		
Masivo_Tier_2	NO	10,666		
Premium_Tier_1	NO	35,809		
Premium_Tier_2	NO	68,626		
Masivo_Tier_1	SI	7,688	151%	2,590
Masivo_Tier_2	SI	13,817	130%	3,152
Premium_Tier_1	SI	59,227	165%	23,418
Premium_Tier_2	SI	94,961	138%	26,335

Figura 3.3: Correlación entre transacciones en TPV y saldo en TDD

saldo disponible es igual a los ingresos recibidos. Por lo que aumentar el número de transacciones en TPV implica reducir (pues, para un mismo nivel de ingreso es forzosamente un *trade-off*) en la misma cantidad gastada el gasto en ATM.

### 3.4. Herramientas de Ciencia de Datos y Lenguajes de Programación Utilizados

- Estudio del *lift* en facturación
  - Software y Estructuras
    - Oracle
    - RStudio
  - Lenguajes de Programación
    - SQL
    - R<sup>5</sup>
- Modelaje del P&L del programa de recompensas
  - Software y Estructuras
    - Excel
  - Lenguajes de Programación
    - VBA
- Presentación y visualización de la información
  - Software y Estructuras
    - PowerPoint

---

<sup>5</sup>Para establecer la conexión entre Oracle y R se hace uso de la librería RJDBC[2]

### **3.5. Impacto para el negocio**

Las hipótesis y aseveraciones planteadas por este proyecto permitieron reforzar la venta del programa a nuevos clientes. En particular, se logró el lanzamiento del primer producto de débito con un programa de recompensas en México.

## Capítulo 4

# Procesamiento de datos para reportes personalizados

### 4.1. Objetivo

Generar un proceso (*tubería*) que facilitara el procesamiento de la información para la generación de reportes específicos (ej. reportar los movimientos de puntos por período, reportar el número de tarjetahabientes que cumplen con alguna meta de facturación en particular).

### 4.2. Contexto

En el contexto del programa de recompensas, para los clientes es importante la medición de un conjunto de *Key Performance Indicators* (“KPIs”) en específico. En algunos casos, para reportar a sus áreas de

finanzas o en otros para accionar mecanismos de comunicación hacia sus clientes. En este sentido, estos proyectos consistieron en utilizar Rstudio[10] y las librerías que permiten conexiones con bases de datos externas (por ejemplo JDBC[2]) para generar una *tubería* de códigos que con solo correrlo en el momento deseado generen el reporte en el formato en el cual el cliente lo deseé reportar.

### 4.3. Estructura y Contenido

Este proyecto se contruyó 100 % estructurado en Rstudio[10] y toma provecho de la gran cantidad de librerías que tiene R[11], en particular las que permiten conexiones con bases remotas, para crear una *tubería* de código que puede ser ejecutado por cualquier usuario para generar el reporte en el formato solicitado por el cliente.

### 4.4. Herramientas de Ciencia de Datos y Lenguajes de Programación Utilizados

- Software y Estructuras
  - Oracle
  - RStudio
- Lenguajes de Programación
  - SQL
  - R<sup>1</sup>

---

<sup>1</sup>Para establecer la conexión entre Oracle y R se hace uso de la librería RJDBC[2]

#### **4.5. Impacto para el negocio**

Este proyecto fue trabajado a solicitud de un cliente, el cual lo solicitó desde el proceso de implementación. En este sentido, el proyecto contribuyó a que el cliente en cuestión contará de forma “automatica” con la información que necesita para realizar actividades de conciliación.

# Capítulo 5

## Reporte Interactivo de *KPI's*

### 5.1. Objetivo

Proveer información relevante y accionable de forma oportuna sobre los indicadores del portafolio a lo largo de las siguientes dimensiones:

- Acumulación y uso de puntos recompensa.
- Número de cuentas enroladas.
- Uso y nivel de gasto de las cuentas enroladas en el programa, enfocado en el comparativo entre redentores vs. no redentores.
- Monitoreo de costos.

### 5.2. Contexto

En el área se necesitaba generar un archivo que consolidará la información de distintos reportes ya existentes. Además, se requería

que la información se presentase de forma ejecutiva y en un formato que fuera operable por las distintas áreas del banco.

### 5.3. Estructura y Contenido

El archivo está estructurado a nivel producto/programa y se compone de:

- Hoja de Inicio con gráfica interactiva de indicadores a través de todos los programas.
- Hoja(s) para la visualización y extracción de datos (una por programa y el agregado).
- Tabla general de datos.

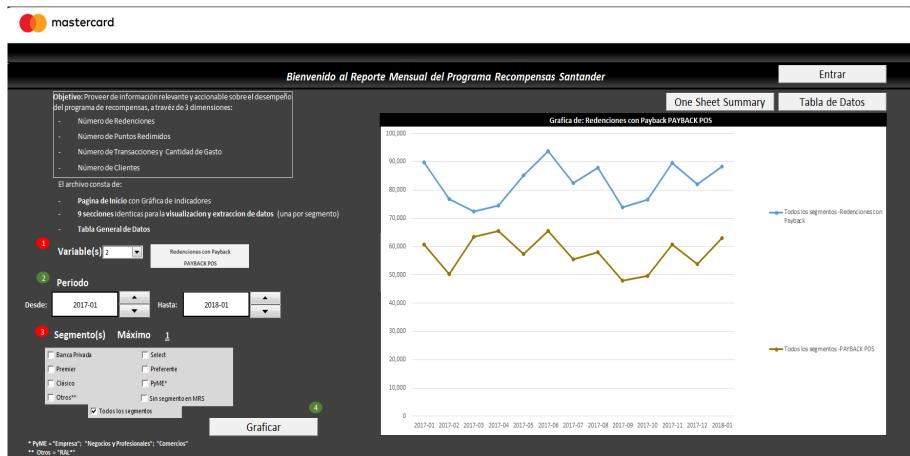


Figura 5.1: Hoja de Inicio

Esta es la *hoja de inicio del documento* contiene:

- Explicación sobre el contenido del documento.

- Estructuras de selección para la generación de una gráfica que permite comparar indicadores entre programas.
- Botones de navegación.

El botón **Entrar** lleva al usuario hasta los paneles donde se puede visualizar, interactivamente, la información de todos sus programas. El documento contiene tantas de estas páginas como programas tenga el documento (más una adicional con el consolidado de todos los programas).

El documento, además, cuenta con una hoja resumen de indicadores. El botón **One sheet summary** lleva al usuario a una página resumen por período. Ésta se construye con el objeto de monitorear de forma sencilla el comportamiento que en términos de gasto presentan los usuarios del programa.

La **Tabla de Datos** constituye la base de datos del documento. En ella se consolida la información mensual para cada uno de los distintos programas de un mismo banco. Está estructurada de la misma forma que la navegación del documento. La información de cada programa se encuentra en tablas independientes y la conexión entre ellas o *llave* es el período de análisis.



Figura 5.2: Dashboard Interactivo por programa

Cada programa tiene una "pestaña", la cual está estructurada en un esquema de 12 páneles. Para navegar entre programas, el usuario necesita presionar el nombre del programa del cual quiera vizualizar o extraer la información. Éstos se encuentran en la primer franja superior (por encima del botón de salida). Dentro de cada página, los primeros seis paneles muestran visualizaciones referentes a indicadores de uso y acumulación de puntos<sup>1</sup>. De izquierda a derecha se lee así:

- **Panel 1:** Gráfica de barras donde se muestra el número de puntos o redenciones para el mes seleccionado.
- **Panel 2:** Gráfica de pie donde se desglosa el porcentaje de puntos o redenciones para el mes seleccionado en la *rama* de **viajes**.
- **Panel 3:** Gráfica de pie donde se desglosa el porcentaje de puntos o redenciones para el mes seleccionado en la *rama* de **mercancías**.
- **Panel 4:** Gráfica de serie de tiempo donde se muestra el ciclo anual de las redenciones por *rama*<sup>2</sup>.
- **Panel 5:** Panel de tres paneles donde se muestra el consolidado de 12 meses de número de puntos o redenciones totales y por cada una de las *ramas*.
- **Panel 6:** Gráfica de serie de tiempo donde se muestra el ciclo anual de las redenciones por *canal*<sup>3</sup>.
- **Panel 7:** Gráfica de serie de tiempo donde se compara la evolución - para un periodo a seleccionar - del comparativo del

---

<sup>1</sup>En el centro de la página se encuentra un selector entre puntos o número de redenciones. Éste funciona para los primeros seis paneles.

<sup>2</sup>Mercancías, Viajes, TPV

<sup>3</sup>TPV, internet, call center.

número de transacciones por cliente entre grupos (redentores versus no redentores).

- **Panel 8:** Panel de tres paneles donde el comparativo del período de gasto y número de transacciones por cliente. El primero de los tres paneles muestra el comparativo año contra año del gasto o el número de transacciones por cliente. El segundo panel muestra el comparativo año contra año del gasto o el número de transacciones para clientes con al menos una redención. El tercer panel muestra el comparativo año contra año del gasto o el número de transacciones para clientes que nunca han presntado una redención.
- **Panel 9:** Gráfica de serie de tiempo donde se compara la evolución - para un periodo a seleccionar - del comparativo de gasto por cliente entre grupos (redentores versus no redentores).
- **Panel 10:** Gráfica de serie de tiempo donde se compara la evolución - para un periodo a seleccionar - del porcentaje de clientes con redención.
- **Panel 11:** Panel de tres paneles donde se observa el comparativo del período en términos de cuentas. El primero de los tres paneles muestra el comparativo año contra año de las cuentas enroladas en el programa. El segundo panel muestra el comparativo año contra año del número cuentas con transacciones. El tercer panel muestra el comparativo año contra año del número de cuentas con redención.
- **Panel 12:** Gráfica de serie de tiempo donde se compara la evolución - para un periodo a seleccionar - del número de clientes enrolados en el programa de recompensas.

La **Figura 5.3** es el resumen a una hoja del documento. Tiene la finalidad de monitorear el impacto en términos de ingresos incrementales generan los esfuerzos que buscan incentivar la redención en categorías específicas. Únicamente cuenta con dos listas de selección: período (en el centro del documento) y categoría de redención (en la gráfica de la evolución de redenciones).

En la franja izquierda del documento se localizan los siguientes indicadores:

1. Incremento año contra año del número de redenciones.
2. Incremento año contra año del número de puntos redimidos.
3. Comparación (mismo periodo) del gasto clientes con redención contra clientes sin redención.
4. Comparación (mismo periodo) del número de transacciones por cliente en clientes con redención contra clientes sin redención.
5. Ticket Promedio de lo redimido por *rama*.
6. Porcentaje de redenciones por *rama*.
7. Promedio de puntos acumulados por cliente en cada programa.

El cuerpo de la página está estructurado con base en seis paneles de información más una tabla con el **análisis de redención por segmento**. De izquierda a derecha el documento se lee de la siguiente manera:

- **Panel 1:** Gráfica de barras donde se muestra el porcentaje de redentores por programa.



2018-01 ▾

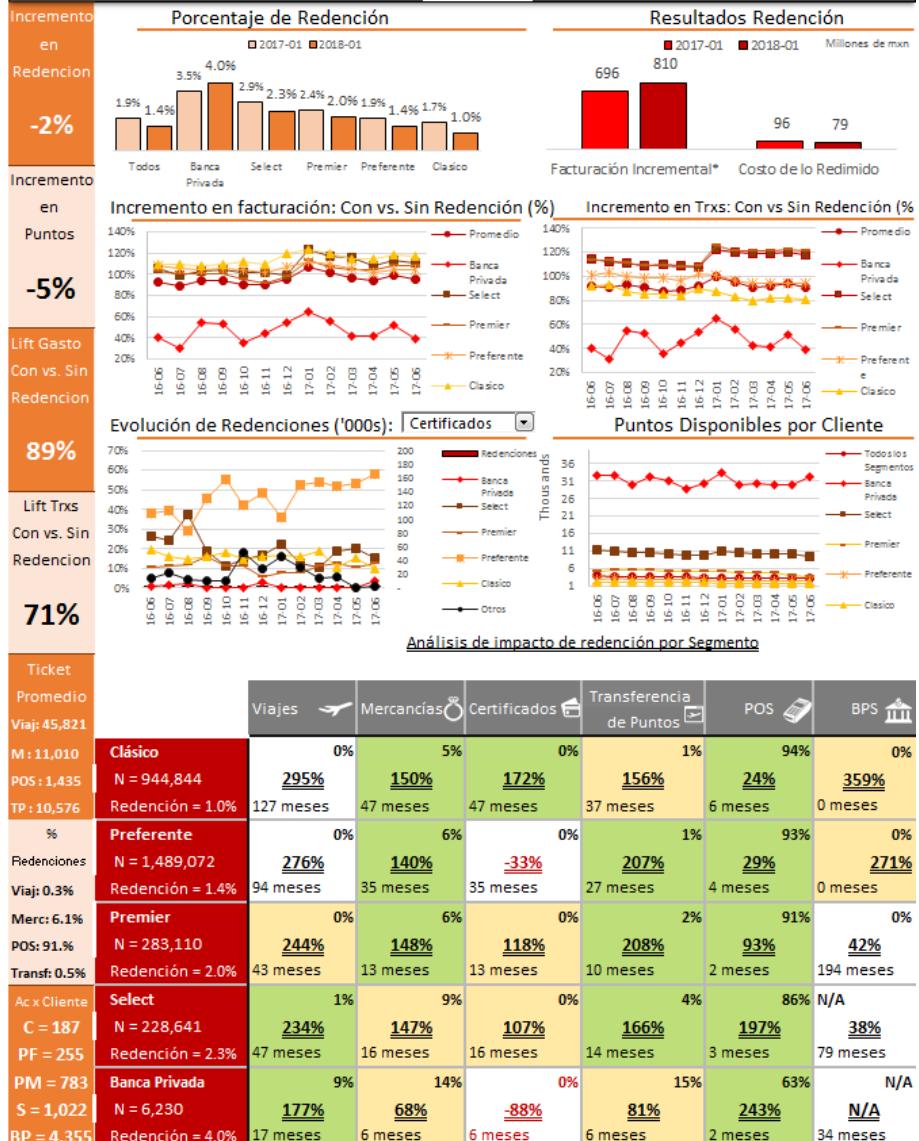


Figura 5.3: One Sheet Summary

- **Panel 2:** Gráfica de barras donde se compara el incremento en facturación con el costo de las mercancías redimidas en el mes seleccionado.
- **Panel 3:** Gráfica de serie de tiempo donde se muestra la evolución en el comparativo por programa del gasto entre grupos (con redención vs. sin redención).<sup>4</sup>
- **Panel 4:** Gráfica de serie de tiempo donde se muestra la evolución en el comparativo por programa del número de transacciones por cliente entre grupos (con redención vs. sin redención).<sup>5</sup>
- **Panel 5:** Gráfica de serie de tiempo donde se muestra la evolución del número de redenciones - así como la participación de cada programa - por categoría de redención.
- **Panel 6:** Gráfica de serie de tiempo donde se muestra la evolución del número de puntos disponibles por cliente de acuerdo al programa.<sup>6</sup>

Por último, en la tabla se muestra la rentabilidad por programa y categoría de redención. En color verde se muestran las interacciones de programa y categoría mas rentables, con un círculo de color naranja se

---

<sup>4</sup>Este tipo de análisis permite mostrar con números lo que en muchas ocasiones los empleados de los bancos ya conocen, en este caso vale la pena notar que los clientes de los programas *más caros* (por ejemplo: *platino* y *world elite*) son los que menos cambio en facturación tienen (y por lo tanto son los menos rentables).

<sup>5</sup>Al igual que el caso anterior, vale la pena notar que los clientes de los programas más caros son los que menos cambio en este indicador tienen (y por lo tanto son los menos rentables).

<sup>6</sup>Vale la pena contrastar la tenencia promedio de puntos que los clientes de los programas más caros tienen versus los que cambian su comportamiento cuando redimen.

marca las tres categorías por programa con mayor número de redenciones.<sup>7</sup>

## 5.4. Herramientas de Ciencia de Datos y Lenguajes de Programación Utilizados

La generación del archivo consistió en dos proyectos paralelos:

- Extracción y manipulación de la información, enfocado a una actualización automática basada en scripts
  - Software y Estructuras
    - Oracle
    - RStudio
  - Lenguajes de Programación
    - SQL
    - R<sup>8</sup>
- Presentación y visualización de la información
  - Software y Estructuras
    - Excel
  - Lenguajes de Programación
    - VBA

---

<sup>7</sup>Para explotar la comunicación y maximizar las ganancias incrementales es necesario por renglon aumentar el porcentaje de redención (en la esquina superior derecha de cada casilla) en las casillas color verde

<sup>8</sup>Para establecer la conexión entre Oracle y R se hace uso de la librería RJDBC[2]

## 5.5. Impacto para el negocio

El archivo por sus funcionalidades se volvió uno de los elementos de valor agregado que se le entrega a los clientes - bancos - que contratan los servicios del programa de recompensas. El documento se actualiza mediante un código R que hace la conexión con la base de datos para la extracción y posterior transformación de la información. Únicamente es necesario agregar la tabla generada a la sección de **Tabla de Datos** para que el documento se actualice.

Este proyecto fue el primero que desarrollé en *Analytics* y basicamente consistió en *poner la casa en orden* es decir acordar con el cliente el nivel de detalle que necesitaba conocer de su portafolio, así como del formato en el cual necesitaba tenerlo. En la empresa tenemos **Tableau**, sin embargo por cuestiones de data privacy (y el hecho de que no todos los bancos tienen licencia de **Tableau**) opté por realizarlo en Excel, un formato que se comparte actualmente con los clientes como parte de la oferta de valor del programa.

Entre otras cosas, este proyecto logró consolidar un proceso **ETL**<sup>9</sup> que permite - desde un mismo ambiente - trabajar en la construcción de una base de datos en formato *tidy*[4]. Además, permite estandarizar la extracción de información y generación de indicadores entre los distintos clientes.

---

<sup>9</sup>Proceso de extraer, transformar y cargar(por sus siglas en inglés) información.

# Capítulo 6

## Segmentaciones *ad hoc*

### 6.1. Objetivo

Generar segmentaciones que permitieran caracterizar a los tarjetahabientes de acuerdo a necesidades específicas con base en los demográficos e información de consumo disponible, por ejemplo:

1. Segmentaciones de acuerdo a preferencia de redención
2. Segmentación de clientes viajeros

### 6.2. Contexto

En el contexto del programa de recompensas, los clientes envían a nuestro sistema una gran cantidad de información, por ejemplo demográficos como la edad o el sexo. Además de la información de gasto al compartir información de cada transacción para que nuestro motor de acumulación de puntos pueda aplicarlos de acuerdo a las características de gasto señaladas por éstos. El objetivo de este

proyecto fue apalancar toda la información disponible para caracterizar grupos de tarjetahabientes y facilitar la venta de ciertos productos de valor agregado del área de lealtad o promover el lanzamiento de campañas segmentadas.

### 6.3. Estructura y Contenido

La generación de este proyecto se llevó a cabo en dos etapas:

1. Extracción y manipulación de la información
2. Generación de presentación ejecutiva

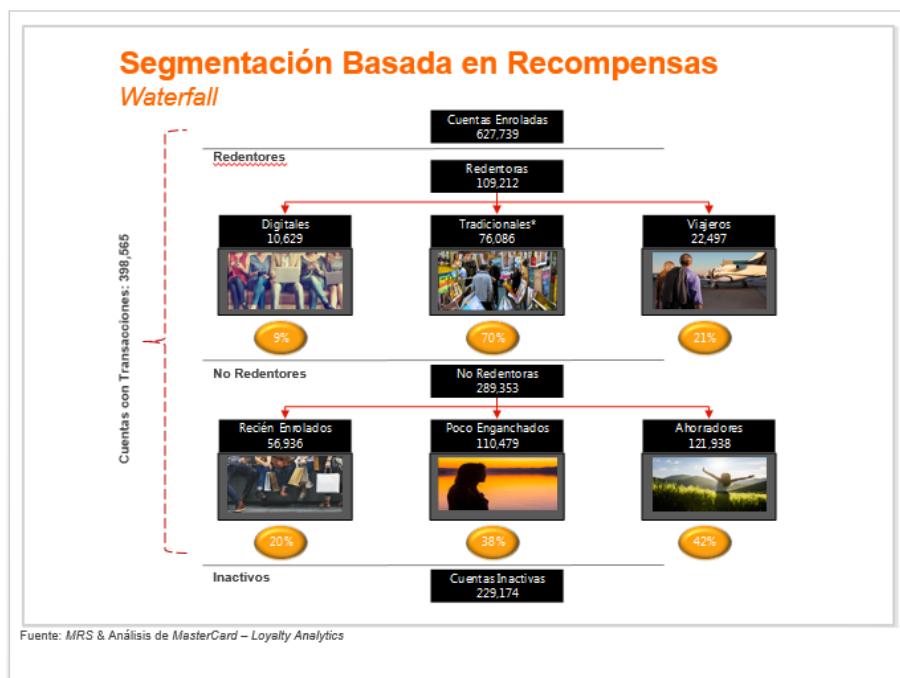


Figura 6.1: Esquema de Segmentación

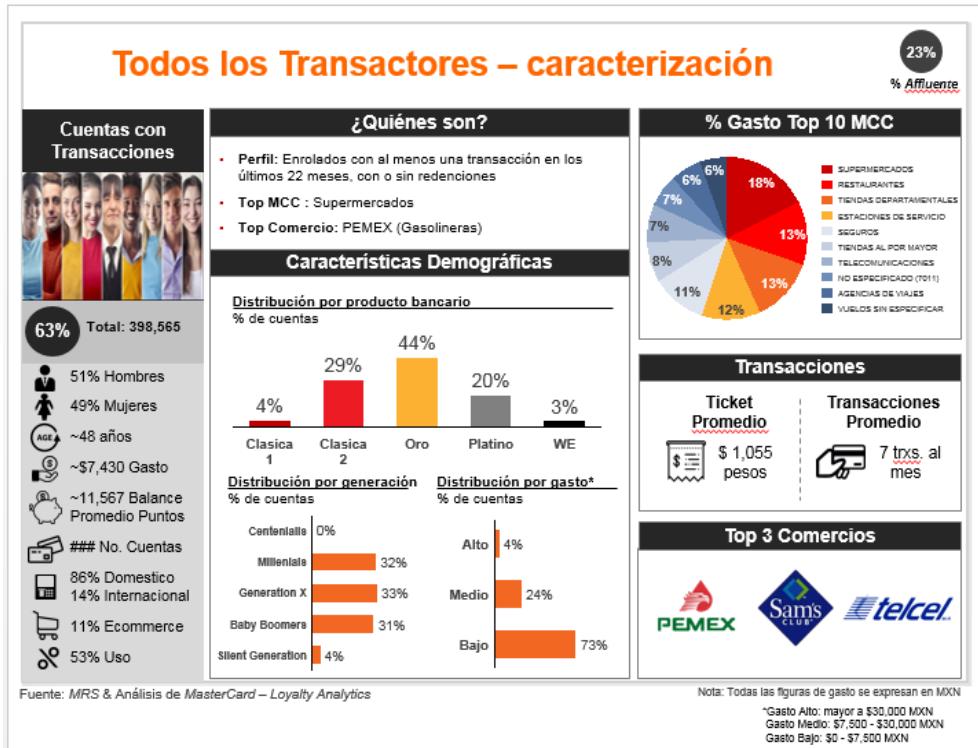


Figura 6.2: Caracterización Transacciones

## **6.4. Herramientas de Ciencia de Datos y Lenguajes de Programación Utilizados**

- Software y Estructuras
  - Oracle
  - RStudio
- Lenguajes de Programación
  - SQL
  - R<sup>1</sup>
- Presentación y visualización de la información
  - Software y Estructuras
    - PowerPoint

## **6.5. Impacto para el negocio**

Este proyecto aumentó el portafolio de servicios de valor agregado del área de lealtad y permitió la implementación de proyectos de consultoría para la generación de campañas focalizadas.

---

<sup>1</sup>Para establecer la conexión entre Oracle y R se hace uso de la librería RJDBC[2]

## Apéndice A

# Más estimaciones del ATET

Además de los análisis ya presentados, se realizaron muchos más con el objetivo de estudiar un posible efecto causal entre el evento de redención y la faturación de los tarjetahabientes. En este apéndice se presenta un estudio que resume la estimación de este efecto con base en tres metodologías.

El estudio considera tarjetahabientes en tres países de América Central: Guatemala, Nicaragua y Panamá. Después de extraer<sup>1</sup> y consolidar la información, se generó un grupo control sintético por el método Coursened Exact Matching (CEM)<sup>2</sup>. La construcción del grupo control se realizó con base en el grupo no tratado. Posterior a su construcción se realizaron tres estimaciones no paramétricas del efecto de tratamiento promedio en el grupo tratado (ATET).

---

<sup>1</sup>La información se extrajo con base en *cosechas*. Es decir, durante 12 meses se recopiló la información de gasto y número de transacciones de los tarjetahabientes. En cada mes, se observó dicha actividad seis meses antes y seis meses después.

<sup>2</sup>Para este objeto se utilizó la librería CEM[6]

# Propensity Score Matching (CEM)

A Statistical Methodology for Causal Inference

February 18, 2019  
Loyalty Analytics



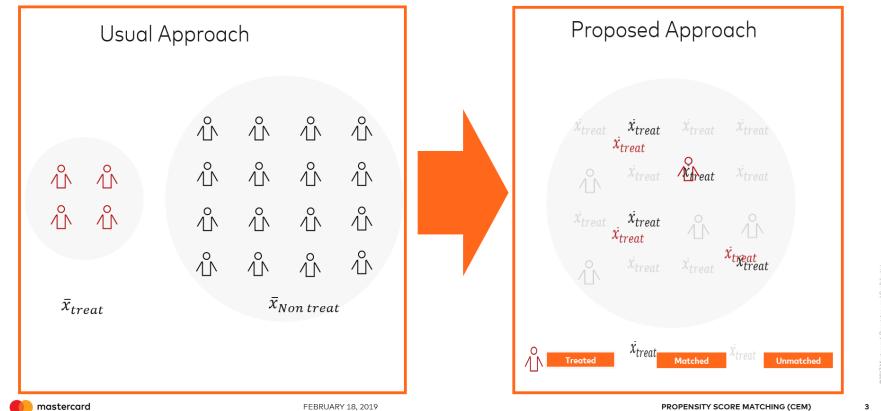
Figura A.1: Coursened Exact Matching

We have two group of clients, those who redeem at least one time vs those who do not redeem



Figura A.2: Tratados y No Tratados

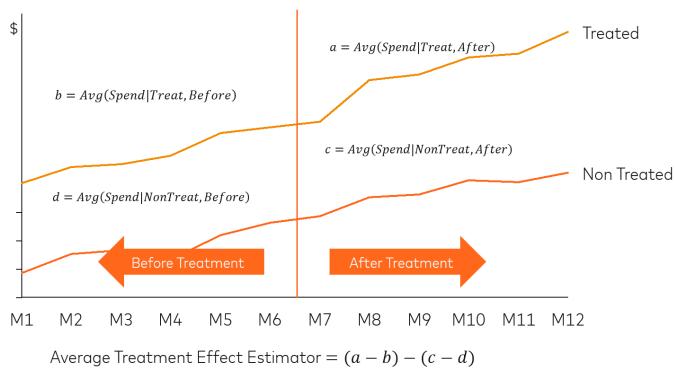
We match treated clients to selected untreated clients based on their proximity using the propensity score model: Coursened Exact Matching



©2019 Mastercard. Proprietary and Confidential

Figura A.3: Construcción grupo control

**Usual approach allows for different behaviors before treatment happens**



©2019 Mastercard. Proprietary and Confidential

Figura A.4: Enfoque antiguo en el negocio

**The propensity score allows us to identify the treated group and select a *matched* group based on similar behavior and/or characteristics**

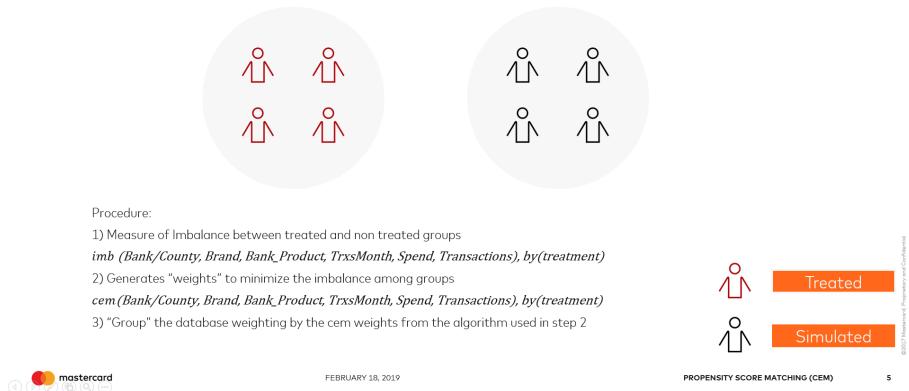


Figura A.5: Proceso de matching

**After that process ideally behavior between clients at time of redemption will better help us capture the change between the two groups of people**

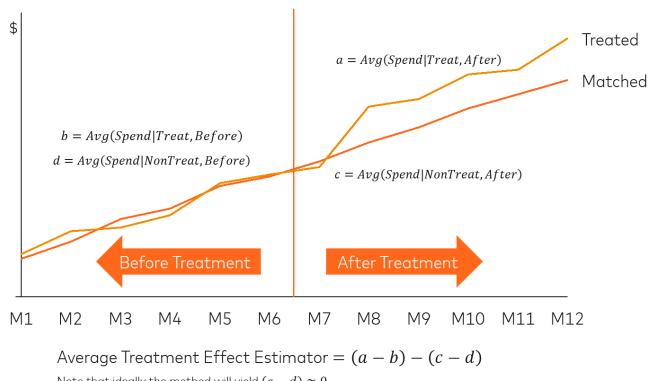
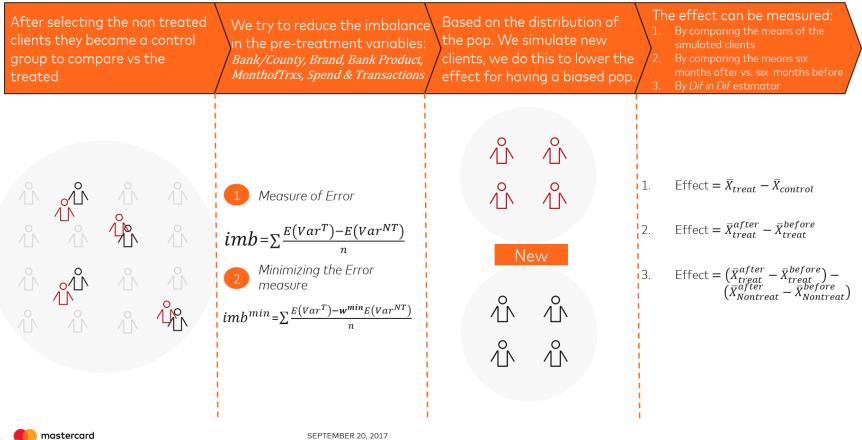


Figura A.6: Enfoque propuesto

## After computing the weights we analyze the effect by simulating new population



©2013 MasterCard Worldwide. Proprietary and Confidential.

8

Figura A.7: Resumen

## Structure of the data set used for the exercise

### Average Spend among groups

	Simulated Control Group (Non Redemptions)												Treated Group (Redemptions)															
	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre	Enero	Febrero	Marzo	Average	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre	Enero	Febrero	Marzo	Average						
Month 1	1,213	1,213	1,140	1,140	1,140	1,140	1,140	1,140	1,140	1,140	1,182	1,381	1,381	1,381	2,142	2,142	2,142	1,868	1,868	1,868	2,840	2,840	2,351	2,053	1,958			
Month 2	1,403	1,403	1,365	1,365	1,365	1,365	1,365	1,365	1,365	1,365	1,469	1,881	1,881	1,881	2,147	2,147	2,147	1,664	1,664	1,664	2,707	2,707	2,301	1,111	1,174			
Month 3	1,110	1,110	1,365	1,365	1,365	1,365	1,365	1,365	1,365	1,365	1,498	1,381	1,381	1,381	1,272	1,272	1,272	1,334	1,334	1,334	1,087	1,087	1,087	813	813	3,870	1,351	2,249
Month 4	1,365	1,365	1,140	1,140	1,140	1,140	1,140	1,140	1,140	1,140	1,498	1,381	1,381	1,381	1,260	1,260	1,260	1,309	1,309	1,309	2,577	2,577	1,785	1,225	1,669			
Month 5	1,140	1,140	1,509	1,509	1,509	1,509	1,509	1,509	1,509	1,509	1,418	1,382	1,382	1,382	1,272	1,272	1,272	1,317	1,317	1,317	1,544	1,544	1,544	1,027	1,027	1,035	1,735	2,104
Month 6	1,509	1,509	1,418	1,418	1,418	1,418	1,418	1,418	1,418	1,418	1,498	1,381	1,381	1,381	1,260	1,260	1,260	1,345	1,345	1,345	1,746	1,746	1,746	3,345	3,345	3,283	1,122	5,005
Month 7	1,509	1,509	1,418	1,418	1,418	1,418	1,418	1,418	1,418	1,418	1,498	1,381	1,381	1,381	1,272	1,272	1,272	1,317	1,317	1,317	1,746	1,746	1,746	3,283	3,283	2,830	1,222	2,299
Month 8	1,418	1,418	1,382	1,382	1,382	1,382	1,382	1,382	1,382	1,382	1,498	1,381	1,381	1,381	1,272	1,272	1,272	1,317	1,317	1,317	1,784	1,784	1,784	2,830	2,830	2,830	1,784	1,784
Month 9	1,382	1,382	1,499	1,499	1,499	1,499	1,499	1,499	1,499	1,499	1,498	1,381	1,381	1,381	1,272	1,272	1,272	1,317	1,317	1,317	1,784	1,784	1,784	3,111	3,111	3,111	1,951	3,713
Month 10	1,382	1,382	1,499	1,499	1,499	1,499	1,499	1,499	1,499	1,499	1,498	1,381	1,381	1,381	1,272	1,272	1,272	1,317	1,317	1,317	1,784	1,784	1,784	3,276	3,276	3,050	2,321	3,050
Month 11	1,260	1,260	1,321	1,321	1,321	1,321	1,321	1,321	1,321	1,321	1,385	1,385	1,385	1,385	1,260	1,260	1,260	1,347	1,347	1,347	1,662	1,662	1,662	2,351	2,351	3,006	5,393	5,499
Month 12	1,272	1,272	3,217	3,217	3,217	3,217	3,217	3,217	3,217	3,217	3,385	3,385	3,385	3,385	1,260	1,260	1,260	1,347	1,347	1,347	1,662	1,662	1,662	2,189	2,189	4,937	4,138	8,363

### Number of Accounts among groups

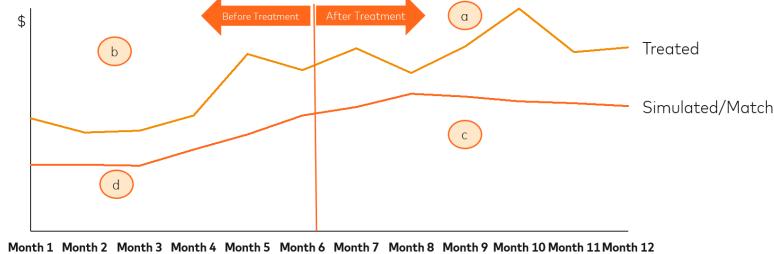
	Simulated Control Group (Non Redemptions)												Treated Group (Redemptions)																	
	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre	Enero	Febrero	Marzo	Average	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre	Enero	Febrero	Marzo	Average								
Month 1	699	704	703	537	990	1,084	1,106	1,124	1,102	1,091	990	8	12	5	8	8	5	18	20	25	12	1,213	1,213	1,213	1,213	1,213				
Month 2	704	703	937	990	1,084	1,106	1,124	1,102	1,091	1,094	990	8	12	7	8	10	7	4	17	23	25	12	1,213	1,213	1,213	1,213	1,213			
Month 3	703	937	990	1,084	1,106	1,124	1,124	1,124	1,102	1,091	990	9	12	7	9	10	6	4	18	22	27	12	1,213	1,213	1,213	1,213	1,213			
Month 4	937	990	1,084	1,106	1,124	1,124	1,124	1,124	1,102	1,091	990	9	12	7	9	10	6	4	18	22	27	12	1,213	1,213	1,213	1,213	1,213			
Month 5	990	1,084	1,106	1,124	1,124	1,124	1,124	1,124	1,102	1,091	990	9	11	7	6	10	6	6	15	20	23	11	1,213	1,213	1,213	1,213	1,213			
Month 6	1,084	1,106	1,124	1,102	1,091	1,094	932	1,149	1,094	1,094	1,098	1,078	858	1,065	10	13	4	7	11	6	5	17	20	22	12	1,213	1,213	1,213	1,213	1,213
Month 7	1,105	1,124	1,102	1,091	1,104	952	1,149	1,084	1,094	1,084	1,078	858	895	1,044	10	10	5	8	9	6	4	11	17	22	10	1,213	1,213	1,213	1,213	1,213
Month 8	1,124	1,102	1,091	1,104	952	1,149	1,084	1,078	1,094	1,084	1,078	858	895	1,044	9	12	7	8	10	4	6	13	22	22	11	1,213	1,213	1,213	1,213	1,213
Month 9	1,102	1,091	1,094	952	1,149	1,084	1,078	1,058	1,084	1,078	1,058	895	921	1,023	8	12	7	9	10	4	6	13	22	22	11	1,213	1,213	1,213	1,213	1,213
Month 10	1,091	1,094	1,094	952	1,149	1,084	1,078	1,058	1,084	1,078	1,058	895	921	1,023	9	12	7	8	10	4	6	13	22	22	11	1,213	1,213	1,213	1,213	1,213
Month 11	1,104	952	1,149	1,084	1,078	895	921	1,023	1,084	1,078	1,058	895	921	1,023	10	10	7	8	8	5	5	14	23	22	11	1,213	1,213	1,213	1,213	1,213
Month 12	952	1,149	1,084	1,078	895	921	1,023	1,084	1,078	1,058	895	921	1,023	1,023	8	10	6	7	9	5	5	15	20	17	10	1,213	1,213	1,213	1,213	1,213

©2013 MasterCard Worldwide. Proprietary and Confidential.

9

Figura A.8: Cosechas

We can see a higher spend on Client clients who had a redemption vs the spend on Client clients who did not redeem (over three different estimators)



#### Results:

1. Effect =  $\bar{X}_{treat} - \bar{X}_{control}$  = Treated – Simulated/Match = \$ 674 usd (36%)
2. Effect =  $\bar{X}_{treat}^{after} - \bar{X}_{treat}^{before}$  = (a) – (b) = \$ 771 usd (36%)
3. Effect =  $(\bar{X}_{treat}^{after} - \bar{X}_{treat}^{before}) - (\bar{X}_{NonTreat}^{after} - \bar{X}_{NonTreat}^{before})$  = (a) – (b) – ((c) – (d)) = \$ 136 usd (6%)

©2017 Mastercard. Proprietary and Confidential.



SEPTEMBER 20, 2017

10

Figura A.9: Resultados

#### Summary of three different estimation methods

- 1 Cardholders who redeem have a higher level of spending (on average 36% more than the portfolio average).

Average Spend Simulated Group	1,868
Average Spend Treated	2,542
Matched Dif Treated vs. Simulated:	674
% From Base (Simulated):	36%

- 2 When cardholders redeem they increase their level of spending even more (on average 36% more than the previous six months).

Average Spend 6 Months Before Treated	2,156
Average Spend 6 Months After Treated	2,927
Difference After vs. Before:	771
% Growth from before redemption:	36%

- 3 Due to the small amount of data, we do not have a statistically significant result. However, in a first approach an average causal effect of 6% was estimated on the avg monthly spend of the clients that redeem.

Temporal Change Simulated	635
Temporal Change Treated	771
Matched Dif Treated vs. Simulated:	136
% Growth Treated vs. Simulated	6%

©2017 Mastercard. Proprietary and Confidential.



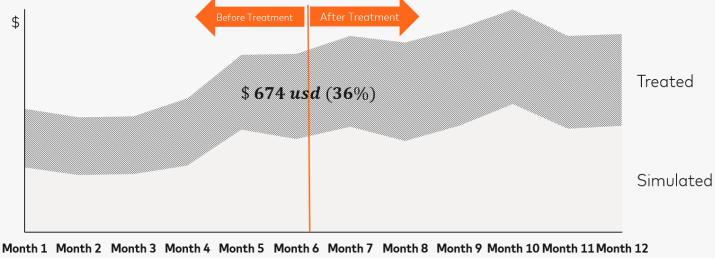
FEBRUARY 18, 2019

PROPENSITY SCORE MATCHING (CEM)

11

Figura A.10: Resumen de Resultados

**Cardholders who redeem have a higher level of spending (on average 36% more than the portfolio average)**



Month 1 Month 2 Month 3 Month 4 Month 5 Month 6 Month 7 Month 8 Month 9 Month 10 Month 11 Month 12

**Results:**

$$1. \text{ Effect} = \bar{x}_{treat} - \bar{x}_{control} = \text{Treated} - \text{Simulated}/\text{Match} = \$674 \text{ usd (36\%)}$$

$$2. \text{ Effect} = \bar{x}_{treat}^{after} - \bar{x}_{treat}^{before} = (a) - (b) = \$771 \text{ usd (36\%)}$$

$$3. \text{ Effect} = (\bar{x}_{treat}^{after} - \bar{x}_{treat}^{before}) - (\bar{x}_{Non treat}^{after} - \bar{x}_{Non treat}^{before}) = (a) - (b) - ((a) - (b)) = \$136 \text{ usd (6\%)}$$



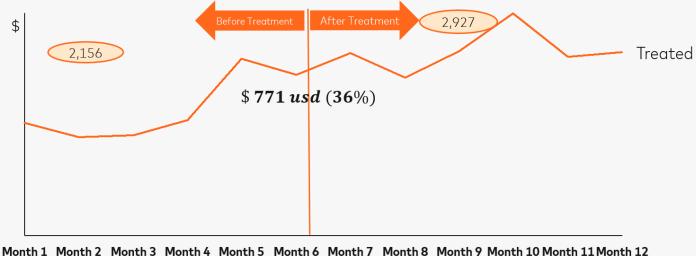
SEPTEMBER 20, 2017

©2017 Mastercard. Proprietary and Confidential.

13

Figura A.11: Tratados vs. No Tratados

**When cardholders redeem they increase their level of spending even more (on average 36% more than the previous six months).**



Month 1 Month 2 Month 3 Month 4 Month 5 Month 6 Month 7 Month 8 Month 9 Month 10 Month 11 Month 12

**Results:**

$$1. \text{ Effect} = \bar{x}_{treat} - \bar{x}_{control} = \text{Treated} - \text{Simulated}/\text{Match} = \$674 \text{ usd (36\%)}$$

$$2. \text{ Effect} = \bar{x}_{treat}^{after} - \bar{x}_{treat}^{before} = (a) - (b) = \$771 \text{ usd (36\%)}$$

$$3. \text{ Effect} = (\bar{x}_{treat}^{after} - \bar{x}_{treat}^{before}) - (\bar{x}_{Non treat}^{after} - \bar{x}_{Non treat}^{before}) = (a) - (b) - ((a) - (b)) = \$136 \text{ usd (6\%)}$$



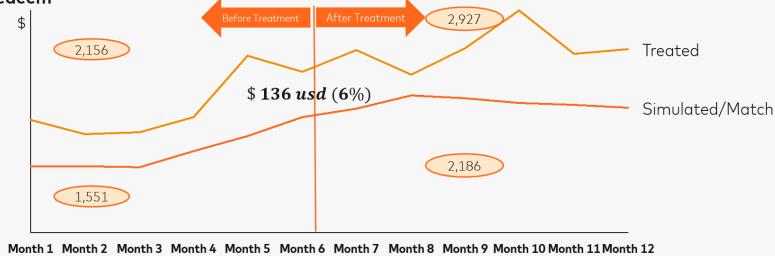
SEPTEMBER 20, 2017

©2017 Mastercard. Proprietary and Confidential.

14

Figura A.12: Tratados Despues vs. Antes

**Due to the small amount of data, we do not have a statistically significant result. However, in a first approach an average causal effect of 6% was estimated on avg. monthly spend of the clients that redeem**



**Results:**

$$1. \quad \text{Effect} = \bar{x}_{treat} - \bar{x}_{control} = \text{Treated} - \text{Simulated}/\text{Match} = \$ 674 \text{ usd (36\%)}$$

$$2. \quad \text{Effect} = \bar{x}_{treat}^{after} - \bar{x}_{treat}^{before} = \text{(a)} - \text{(b)} = \$ 771 \text{ usd (36\%)}$$

$$3. \quad \text{Effect} = (\bar{x}_{treat}^{after} - \bar{x}_{treat}^{before}) - (\bar{x}_{Non treat}^{after} - \bar{x}_{Non treat}^{before}) = \text{(a)} - \text{(b)} - \text{(c)} - \text{(d)} = \$ 136 \text{ usd (6\%)}$$



SEPTEMBER 20, 2017

©2013 MasterCard. All rights reserved. Confidential

15

Figura A.13: Diferencia temporal en Diferencia de grupos (*DifinDif*)

## Apéndice B

# Ciclo de la información

Este apéndice tiene como objetivo proporcionar mayor detalle sobre el proceso de uso de la información, con relación a:

- ¿Cómo se extraía la información?
- ¿En donde y cómo se hacían los análisis?
- ¿Cuál era el proceso para eficientar la toma de decisiones mediante el uso de los análisis?

### B.0.1. Extracción de la información

En el contexto del programa de recompensas. La información sobre los tarjetahabientes es enviada en archivos por lotes de forma diaria, o *en linea* mediante el proceso de autorización, compensación y liquidación para las transacciones de tarjetas de la marca. Luego, la información es almacenada de forma segura en nuestros sistemas. Y, en ese mismo ambiente es procesada para generar los análisis necesarios.

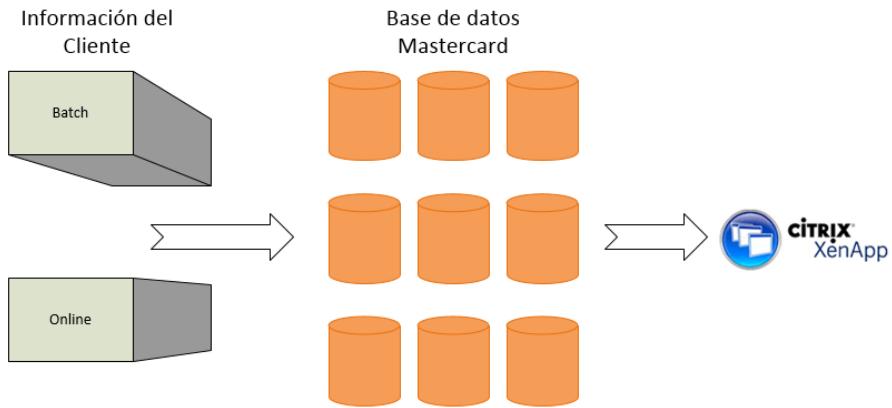


Figura B.1: Estrucura de extracción y manipulacion de los datos

### B.0.2. Toma de decisiones

Para entender el proceso de decisiones es necesario mencionar que el área estaba dividida, de acuerdo a sus funciones en: Operaciones e Implementaciones, Ventas y *Analytics*. La labor del último era proveer de toda la información relevante para poder mejorar la toma de decisiones. Es decir, el objetivo era hacer recomendaciones, basadas en datos, de una gran variedad de temas tanto comerciales como operativos.

# Referencias

ANGRIST, J. D., PISCHKE, J.-S. (2009). “Mostly harmless econometrics: an empiricist’s companion”. Princeton, Princeton University Press.

RJDBC, title = RJDBC: Provides access to databases through the JDBC interface, author = Simon Urbanek, year = 2014, note = R package version 0.2-5, url = <http://CRAN.R-project.org/package=RJDBC>,

tidyverse, title = tidyverse: Easily Install and Load the 'Tidyverse', author = Hadley Wickham, year = 2017, note = R package version 1.2.1, url = <https://CRAN.R-project.org/package=tidyverse>,

Paper tidy-data, author = Hadley Wickham, issue = 10, journal = The Journal of Statistical Software, selected = TRUE, title = Tidy data, url = <http://www.jstatsoft.org/v59/i10/>, volume = 59, year = 2014, bdsk-url-1 = <http://www.jstatsoft.org/v59/i10/>,

CausalImpact, title = Inferring causal impact using Bayesian structural time-series models, author = Kay H. Brodersen and Fabian Gallusser and Jim Koehler and Nicolas Remy and Steven L. Scott, journal = Annals of Applied Statistics, year = 2014, volume = 9, pages = 247–274, url = <http://research.google.com/pubs/pub41854.html>,

cem, title = cem: Coarsened Exact Matching, author = Iacus and Stefano M. and King and Gary and Porro and Giuseppe, year = 2018, note = R package version 1.1.19, url = <https://CRAN.R-project.org/package=cem>,

MatchIt, title = MatchIt: Nonparametric Preprocessing for Parametric Causal Inference, author = Daniel E. Ho and Kosuke Imai and Gary King and Elizabeth A. Stuart, journal = Journal of Statistical Software, year = 2011, volume = 42, number = 8, pages = 1–28, url = <http://www.jstatsoft.org/v42/i08/>,

Zelig, title = Zelig: Everyone's Statistical Software, author = Christine Choirat and James Honaker and Kosuke Imai and Gary King and Olivia Lau, year = 2018, note = Version 5.1.6, url = <http://zeligproject.org/>,

Zelig Article, title = Toward A Common Framework for Statistical Analysis and Development, author = Kosuke Imai and Gary King and Olivia Lau, journal = Journal of Computational Graphics and Statistics, volume = 17, number = 4, year = 2008, pages = 892-913, url = <http://j.mp/msE15c>,

Rstudio, title = RStudio: Integrated Development Environment for R, author = RStudio Team, organization = RStudio, Inc., address = Boston, MA, year = 2015, url = <http://www.rstudio.com/>,

R, title = R: A Language and Environment for Statistical Computing, author = R Core Team, organization = R Foundation for Statistical Computing, address = Vienna, Austria, year = YEAR, url = <https://www.R-project.org>