Análisis de grupos de clientes para banco colombiano usando clustering por K-means

Taller 1 Individual

María Fernanda Izquierdo Aparicio

# BUSINESS UNDERSTANDING

## Background

El mercado de las tarjetas de crédito (TCs) en Colombia ha evolucionado de manera significativa en las últimas décadas, posicionando este medio de pago como parte importante en las finanzas de los ciudadanos. Esto se puede ver en el incremento del total de tarjetas, de acuerdo a Euromonitor International (2023) desde el 2012 al 2022, el número de tarjetas de crédito en el país pasó de 7.6 millones a 16.5 millones, lo que indica un aumento de aproximadamente 115%. Este crecimiento se ha caracterizado, en parte, por la evolución que este sector bancario ha tenido con el fin de ofrecer una amplia gama de opciones a los consumidores, desde tarjetas tradicionales hasta tarjetas con recompensas y beneficios específicos. Aun así, los competidores no han cambiado radicalmente, siendo Tuya S.A., Bancolombia y Colpatria los principales en el mercado. En los últimos años, se ha visto la incursión de NeoBancos y compañías FinTech en el mercado, quienes han traído un poco más de dinamismo a la industria al ofrecer opciones de aprobación más amigables con el consumidor, servicio personalizado a través de sus apps y diferentes beneficios como cashback o no tener cuota de manejo. Teniendo este contexto, a continuación se presenta un análisis PESTLE del mercado de tarjetas de crédito en Colombia:

***Político:***

Colombia ha tenido una reorientación de las políticas económicas de las instituciones financieras, lo que ha llevado a una transformación integral en la infraestructura, así como un incremento en la oferta y la demanda de tarjetas de crédito (Palomino-Martínez, 2019). Esta reorientación ha sido impulsada por el Estado, el cual ha generado “políticas (…) dirigidas a incluir a más ciudadanos en el sistema financiero” (Palomino-Martínez, 2019). Esto ha dado como resultado la introducción de diversas normatividades en los últimos 20 años que han beneficiado a los consumidores, por ejemplo regulación en la tasa de usura[[1]](#footnote-1), eliminación de penalizaciones bancarias por pronto pago[[2]](#footnote-2) y mayor facilidad para acceder a créditos[[3]](#footnote-3).

***Económico:***

En el contexto económico se tiene que, por un lado, luego de un alza generalizada en tasas de interés, durante y después de la pandemia, por causales macroeconómicas externas, en el 2023 “las tasas de casi todas las entidades que ofrecen (…) tarjetas de crédito pasaron de 40% a entre 20% y 25%” (López, 2023) Esto se traduce en una disminución de los costos de préstamos de dinero usando tarjetas de crédito. Por otro lado, el país sigue enfrentando una inflación sostenida lo cual frena el consumo de los hogares debido a que su poder adquisitivo se reduce y dejan de adquirir bienes y servicios con sus tarjetas de crédito. Así mismo, este panorama ha desencadenado que muchas personas cancelen sus tarjetas de crédito pues muchos han tomado una posición “más juiciosa y precavida con respecto a su nivel de endeudamiento” dada las condiciones actuales (Zapata Quinchía, 2022)

***Social:***

En temas sociales, la Banca de Oportunidades encontró en 2022 que con las tarjetas de crédito (TCs) “se realizaron 447 millones de operaciones: 93 % compras y 7,0 % avances, alcanzando 27,9 operaciones por tarjeta en promedio en 2022”. Esto quiere decir que la mayoría de colombianos usaron sus TCs mayoritariamente para consumo. Con este contexto se destaca que los colombiano han cambiado cómo usan las tarjetas de crédito, pasando de ser un medio de financiación para bienes durables a ser un medio de pago planificado para servicios públicos o matrículas debido a que en la actualidad varias empresas de financiamiento tiene estímulos por usar TCs que han hecho que éstas “pasen de ser un mecanismo para comprar (…) a hacer parte de la vida cotidiana de los colombianos” (Olaya, 2019). Para profundizar, el presidente de Fenalco en el 2019 dijo que entre los gastos más pagados con TCs están las vacaciones, la educación, los impuestos y los servicios. (Olaya, 2019).

***Tecnológico:***

En cuanto a la tecnología, se tiene que parte de la masificación de las tarjetas de crédito en el país se ha dado porque “los medios digitales han ganado terreno entre los colombianos” siendo que en el 2021 el “52,7 % (de las transacciones) fueron por el móvil, 20,6 % por Internet (… y) 6,7 % por datáfonos” (Semana, 2022). De forma más específica, en el 2022, La República reportaba que los datáfonos se consolidaban como uno de los canales más utilizados por el comercio y, para el final de 2021, el sistema financiero contaba con 919.751 datafonos; 235.526 más que en 2020 (Morales Arévalo, 2022). Sobre este canal, vale la pena mencionar que parte de este incremento se debe a empresas Fintech que “han democratizado los pagos, permitiendo que más comerciantes y emprendedores accedan a un datáfono móvil” (Mora Aguilar, 2023). En el país los proveedores de datafonos más importantes son, Credibanco, Redeban, Visionamos, seguidos por BOLD y BBVA’S Open Pay (Mora Aguilar, 2023).

***Legal:***

En temas legales respecto a las tarjetas de crédito, se tienen que considerar varios ángulos. El primero es que la Superintendencia Financiera de Colombia (SFC) es la entidad encargada de regular la Tasa de Usura la cual es el “valor máximo de los intereses (…) que puede cobrar un organismo a los agentes de la economía” (Banco de la República, 2023). Ésta se actualiza mensualmente y las compañías de financiación no pueden cobrar intereses mayores a esta tasa. Por otro lado, la SFC, como ente supervisor del sistema financiero colombiano, ha desarrollado varios esfuerzos legales para la protección al consumidor financiero como lo son “establecer cláusulas y control a prácticas abusivas, crear la figura del Defensor del Consumidor Financiero (DCF) e implementar el Sistema de Atención al Consumidor Financiero (SAC)” (Asobancaria, 2022). En contraste, las entidades financieras que expiden una tarjeta de crédito bajo las franquicias de Visa o Mastercard también tienen unos reglamentos a los que acogerse como lo son el *Visa’s Product & Service Rules* (2023) y el *Master Card’s Reglamento operativo* (2023), documentos que establecen las condiciones para poder ofrecer una tarjeta de crédito de éstas franquicias y los beneficios otorgados a los clientes finales. Finalmente, desde la perspectiva de los comercios que usan datáfono o alguna de las pasarelas de pago para usar TCs, éstos deben acogerse al reglamento operativo de los proveedores de éstas plataformas de pago para que los consumidores puedan hacer uso de TCs en sus establecimientos.

Teniendo el contexto del mercado según PESTEL, se revisan ahora las 5 Fuerzas de Porter:

***Clientes:***

Los clientes en este contexto son los usuarios de las tarjetas de crédito, quienes tienen cierto poder de negociación al tener la opción de escoger a que tarjeta de crédito e institución financiera aplicar, aunque no tienen garantizada la aprobación de la misma. Y en contraste, los clientes también cuentan con relativa facilidad para cancelarlas (El Tiempo, 2021). Ahora, la entrada de más competidores al mercado de las TCs, los cambios en las tasas de interés y los beneficios dinámicos que ofrecen las entidades financieras, hacen que el panorama actual sea que los consumidores tienen cierto poder para cambiar entre tarjetas en función de los beneficios, tarifas y tasas de interés.

***Proveedores:***

En la industria de tarjetas de crédito, los proveedores tiene un importante poder de negociación. Por un lado las redes de tarjetas de crédito, franquicias como Visa y MasterCard, tienen poder de negociación con las entidades financieras que ofrecen las TCs, ya que determinan las tarifas y condiciones de uso y otorgamiento de las mismas. De la misma forma, proveedores de datáfonos y pasarelas de pago son también importantes pues éstos son los que ofrecen los canales con los que el cliente va a poder hacer uso de su TC. Sin estos canales, el usuario no puede hacer uso de su TC y debe recurrir a otros medios de pago.

***Productos Sustitutos:***

Si se entienden las tarjetas de crédito como una forma de pago, la industria se enfrenta a las tarjetas de débito y billeteras digitales como producto sustituto, ya que ofrecen comodidad y rapidez sin tener que acumular deuda para pagar. En contraste, si las TCs son entendidas como una forma de financiamiento, en la actualidad, hay varios productos financieros que resuelven esta necesidad, como lo son: los créditos de consumo (líneas de crédito) o una nueva modalidad que es el *Peer-to-Peer lending* (préstamos entre persona particulares)propuesto por la plataforma Ikiwi en Colombia. En general, estas alternativas de financiamiento pueden ser atractivas para el consumidor al ofrecer tasas más bajas que una tarjeta de crédito (las cuales generalmente están cerca de la tasa de usura) y planes de pagos más flexibles.

***Competidores Existentes:***

En la industria colombiana de TCs, los principales competidores son bancos ya establecidos, como Bancolombia, Colpatria, Davivienda y Falabella. También se encuentra Tuya SA, que, aunque no es un banco tradicional, es el principal expedidor de tarjetas de crédito por medio de *cobranding* con tiendas de grandes superficies (Euromonitor, 2023). Cabe mencionar que, en los últimos años, la presencia de NeoBancos y compañías Fintech, como NuBank y Rappi, también han generado mayor dinamismo en el mercado con propuestas de beneficios diferentes a los bancos tradicionales y que proponen un valor agregado al cliente.

***Nuevos Competidores:***

En temas de nuevos competidores, entrar al mercado para expedir tarjetas de crédito en Colombia es un proceso riguroso ya que la compañía de financiamiento debe ser avalada por la Superintendencia Financiera[[4]](#footnote-4). Para conseguir este aval, la empresa debe cumplir unos requisitos mínimos de capital, gobierno corporativo, evaluación de riesgos y prácticas de prevención de lavado de activos, entre otros. Sin este aval, no sería legal su negocio, por lo cual la entrada al mercado tiene barreras significativas de entrada.

Teniendo en cuenta el contexto macro al que se enfrenta la industria y los componentes de la misma, se hace ahora un análisis de la posición sobre la cual se trabajará este proyecto. Se considera que el banco que está interesado en desarrollar estrategias para diferentes grupos de clientes, de acuerdo a sus ocasiones de consumo con tarjetas de crédito, es un banco ya establecido en el mercado colombiano. Además, es uno de los competidores más prominentes del mercado que cuenta con una amplia trayectoria en el sistema financiero y ha podido desarrollar su marca, así como una importante base de clientes. A continuación se presenta un análisis DOFA de la entidad:

***Fortalezas:***

Entre las fortalezas que tiene el banco es que, al ser uno de los más grandes y reconocidos del país, cuenta con una amplia presencia a lo largo de las regiones colombianas, lo cual genera que sea accesible a diferentes consumidores y a una base de clientes amplia. Cuenta además con un portafolio de tarjetas de crédito grande, el cual ofrece algunos beneficios a los titulares de la tarjeta, cubriendo diferentes necesidades de los clientes. De la misma manera, cuenta con alianzas estratégicas las cuales permiten que los clientes puedan acceder a promociones especiales.

***Oportunidades:***

En cuanto a oportunidades, se ha identificado que, dado el contexto del mercado de TCs, el cual recientemente ha tenido un crecimiento importante, el banco tiene oportunidad de seguir posicionándose y ganar un poco más de *market share.* Además, tiene la posibilidad de mejorar su convergencia y retención de clientes por canales digitales, ya que siendo un banco tradicional colombiano, su fuerte es convergencia en canales presenciales. Con esto, el banco también tiene la oportunidad de mejorar el servicio al cliente. Finalmente, aunque el banco ofrece beneficios tradicionales con sus tarjetas de crédito, que son: tener cobertura en viajes, casillero virtual y acumulación de millas con algunas aerolíneas, tiene la posibilidad de ampliar su oferta a la luz de que los nuevos competidores del mercado están ofreciendo beneficios más atractivos y que la teoría propone que beneficios más dinámicos pueden llegar a ser óptimos para el cliente también (Jalbert, Stewart & Martin, 2010).

***Amenazas:***

Sobre las amenazas a las que este banco se enfrenta, cabe resaltar los posibles cambios en los hábitos de consumo de los colombianos. Por un lado, la inflación sostenida y altas tasas de interés han influenciado a una parte de los consumidores a cancelar sus tarjetas de crédito. De otro lado, el interés de los consumidores en experiencias personalizadas de servicio al cliente hacen que los mismos puedan buscar soluciones diferentes a las tarjetas de crédito de este banco. En paralelo, se evidencia que el marcado ha visto a nuevos actores entrar con propuestas de valor diferentes, sobre todo en beneficios novedosos como *cashback*, cuota de manejo cero, *personal banker*, entre otros, que están atrayendo a un segmento importante de clientes. El desafío en este caso es competir por los nuevos clientes que llegan por primera vez al mercado y la retención de los mismos. Teniendo en cuenta la oportunidad digital, también existe una amenaza cibernética que puede comprometer la información sensible de los clientes y dañar la reputación del banco. Y como último punto, está la regulación gubernamental que puede llegar a afectar la forma en la que este banco opera y ofrece productos (por ejemplo: cambios en la regla de provisión de cartera o requerimientos de fondeo/liquidez).

***Debilidades:***

Sobre las debilidades del banco, la primera es el riesgo crediticio que el banco enfrenta por el préstamo de éste dinero. Esto es un riesgo significativo debido al entorno económico en la actualidad. De otro lado, las tarjetas de crédito de este banco tienen tasas de interés altas que pueden llegar a ser disuasorias para clientes existentes o para potenciales clientes. De hecho, en promedio, se encuentra que mensualmente se cancelan alrededor del 3% al 5% de las tarjetas activas, es decir, que la tasa de retención mensual es solamente del 95% al 97%[[5]](#footnote-5). Además, el banco al tener una amplia base de clientes también se enfrenta a un volumen de procesos altos lo cual impacta el servicio al cliente y los tiempos de aprobación de productos. Finalmente, al tener un portafolio amplio de TCs, que incluyen las franquías de Visa y MasterCard, el banco se enfrenta a costos relativamente altos por parte de las mismas o cambios en las condiciones de los contratos, lo cual puede generar traslado de costos para los clientes.

## Business goal

El objetivo de negocio es incrementar la tasa de retención del clientes (*Costumer Rentention Rate*) a partir de la implementación de estrategias comerciales para grupos de clientes específicos de acuerdo a los clústeres generados.

## Business success criteria

Incrementar la tasa de retención del cliente mensual entre un 1% y 3%.

## Determine Data mining goals

Desarrollar un número adecuado de clústeres coherentes, usando la metodología de K-means que cumplan con estándares de robustez.

## Data mining success criteria:

Que la medida de robustez de los clústeres sea mayor o igual 0.75.

# DATA UNDERSTANDING

## Describe data

La base de datos de éste banco contiene información de las transacciones de consumo en un mes de 47871 tarjetahabientes, quienes tienen tarjetas de crédito de tres franquicias: Visa, MasterCard y una tercera propia del banco. La base de datos contiene 26 columnas que presentan: identificador del cliente, información sobre las transacciones (total, promedio de valor consumido, valor mínimo y máximo de transacción, y desviación estándar por transacción). Además, tiene información, en porcentajes, sobre el uso que cada tarjetahabiente le da a su tarjeta, por ejemplo, uso por franquicia y territorio de la compra, uso en el día, y uso por días de semana. El *dataset* finaliza con una variable que describe el sitio de consumo más frecuente del cliente con su tarjeta de crédito.

## Explore data

Tanto los análisis univariado como multivariado se realizaron sobre la base de datos entera. De éste análisis se debe mencionar que hay 24 variables escalares y 2 categóricas. Solamente se analizan 23 variables escalares ya que *CLIENTES* es un identificador único que no aporta al desarrollo del proyecto. De las variables escalares se rescata:

* La mayoría de las distribuciones de las variables numéricas que se tienen presentan sesgos positivos evidenciados por coeficientes de sesgo mayores a cero. Solo hay 3 variables con sesgo negativo que son *porcentaje\_nacional\_total*, *porcentaje\_tarde* y *porcentaje\_mastercard\_nacional*. Ninguna variable es totalmente simétrica.
* En cuanto a las curtosis, la mayoría de éstas son mayores a cero indicando que éstas variables tienen una poca concentración de valores alrededor de la media y tiene colas pesadas, indicando presencia de valores atípicos. Sólo 4 variables escalares tiene curtosis menor a cero lo que indica una cantidad moderada de valores en la cola, pero la mayoría de los valores se encuentran cerca de la media. Estas variables con curtosis negativa son *porcentaje\_mastercard\_nacional*, *porcentaje\_tarde*, *porcentaje\_visa\_nacional* y *porcentaje\_manana*.
* De las variables categóricas, la primera es grupo de cliente, ésta sólo tiene 5 grupos, todos nombradas a partir de una letra. El grupo A concentra más del 85% de las observaciones en total. Grupos B, C, D y E tiene un bajo conteo de observaciones. La segunda variable categórica es el sitio más frecuente de consumo, la cual tiene 109 niveles distintos con diversos conteos de observaciones en cada uno. Es importante mencionar que los niveles de “Supermercados/Tiendas Express” y “Almacén Por Departamento Con Supermercado” son los únicos que tienen más del 10% del total de las observaciones (19.23% y 12. 97% respectivamente). El resto de los 107 niveles acumulan el 67.8% de las observaciones, y de hecho, 91 niveles tienen menos de un 1% del total. Esto indica que usar las 109 categorías en el análisis es complicado debido a su distribución. Finalmente, el uso de variables categóricas impide tener buen sentido de “distancia”, implicando que medir la similitud entre una categoría y otra es difícil.

Para el análisis multivariado se revisa la correlación, la cual indica que:

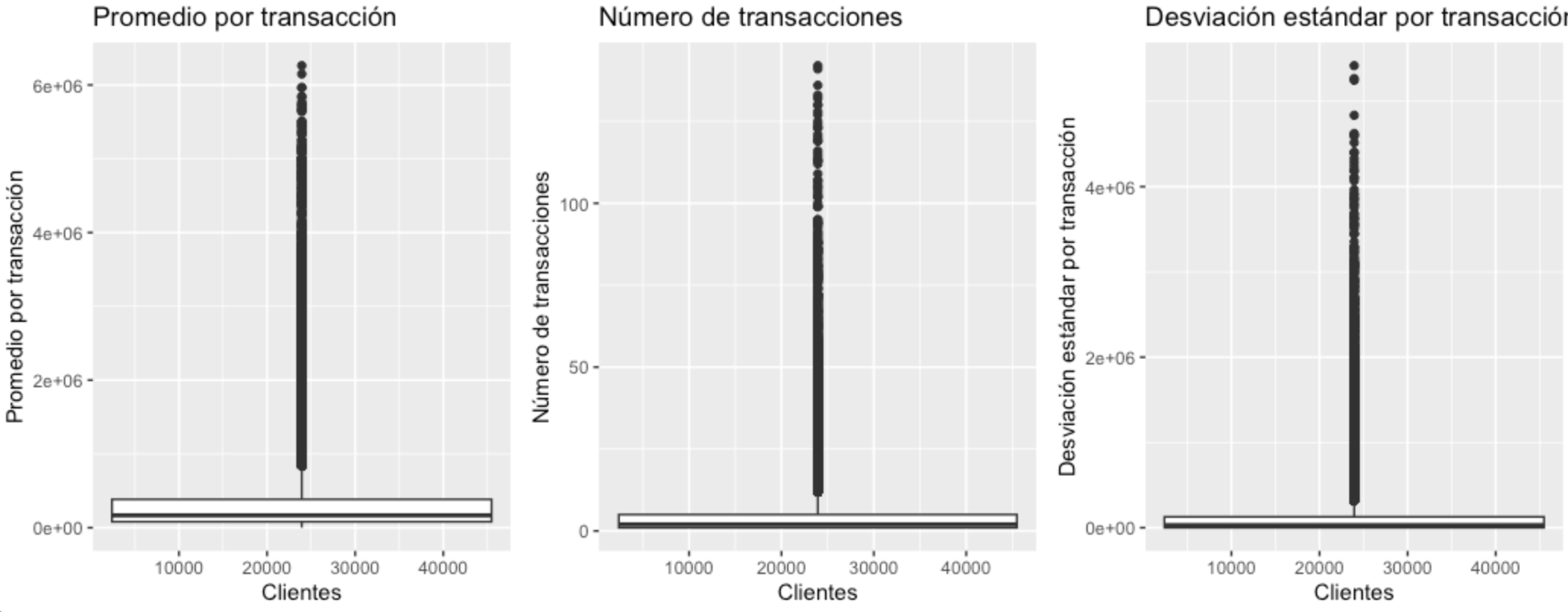
* Hay una alta correlación lineal positiva entre las variables promedio por transacción, valor mínimo, valor máximo, y desviación estándar por transacción. Esto se debe a que todas se relacionan a los valores pagados por el cliente en sus ocasiones de consumo.
* También hay correlaciones negativas entre los porcentajes de uso de las tarjetas Visa, MasterCard y de Otra Franquicia. Esto se puede dar debido a que si un usuario usa una tarjeta de una franquicia, probablemente no va a usar otra tarjeta de una franquicia diferente en la misma transacción.
* De la misma forma, hay correlaciones lineales negativas entre porcentajes de uso nacional (ya sea a nivel de franquicia o nivel total) con porcentajes de uso internacional. De manera complementaria, hay correlación positiva entre porcentajes nacional/internacional a nivel de franquicia con el porcentaje de uso total nacional/internacional.
* Finalmente, hay correlaciones negativas entre los porcentajes de uso durante diferentes momentos del día, y similar, también hay correlaciones negativas entre los porcentajes de uso durante diferentes días de la semana.

Es importante tener en cuenta estas correlaciones, debido a que si las variables se usan de forma individual, pueden ser redundantes en el uso del algoritmo de *K-means*, distorsionando la geometría de los clústeres y la interpretación de los mismos.

## Verify Data Quality

De la revisión de la data se destaca que:

* Ninguna de las variables tiene datos nulos.
* Las variables que indican porcentajes de uso tienen un alto porcentaje de ceros. Casos importantes a mencionar son: *porcentaje\_otrafranquicia\_internacional*, *Porcentaje\_otrafranquicia\_nacional*,  *porcentaje\_mastercard\_internacional*, *porcentaje\_visa\_internacional*, *Porcentaje\_internacional\_total,* , *porcentaje\_noche*, y los porcentajes de uso con respecto a los días de la semana tienen más del 60% de sus observaciones en cero. Esto puede sesgar el análisis si se usan las variables por si solas.
* En los datos referentes a gasto en transacciones y número de las mismas, se observa el 70% de las observaciones se concentra en valores muy bajos, mientras que el 25% restante dispersa en valores relativamente altos, evidenciando una cantidad importante de atípicos. Esto se puede dar debido a que la base contiene una amplia gama de diferentes clientes con diferentes hábitos de consumo, muchos de ellos con un consumo promedio pero también hay clientes con gastos mayores que, dentro del contexto de un banco, no se pueden ser ignorados. Hacer caso omiso de la existencia de éstos atípicos puede impactar los centroides de los clústeres y por ende, sesgar los grupos resultantes.



# DATA PREPARATION

## Clean and Construct data

Teniendo en cuenta los análisis univariado y multivariado junto con la revisión de la calidad de la data, se eliminaron algunas variables, otras fueron reemplazadas y se crearon variables nuevas a partir de variables originales para la construcción del nuevo *dataset* para la identificación de clusters. A continuación se indican los pasos realizados para este proceso:

### Eliminación inicial de variables:

|  |  |
| --- | --- |
| **Variables eliminadas** | **Justificación** |
| CLIENTE | Es un identificador numérico único para cada cliente que si se llega a utilizar puede distorsionar el análisis. |
| grupo\_de\_cliente | Es una variable categórica de la cual no se tiene información al respecto por lo cual no es de uso en el proyecto. |
| transaccion\_minima  transaccion\_maxima desviacion\_estandar\_por\_transaccion | Son variables con alta correlación con promedio de transacción por lo que pueden causar redundancia, distorsión y e impactar el resultados de la generación de clústeres. |
| porcentaje\_manana  porcentaje\_tarde  porcentaje\_noche | Debido al contexto del banco, se decide que la ocasión de consumo se centrará en qué días de las semana el cliente consume y no el momento del día. De la misma manera se eliminan los riesgos de las correlaciones que estas variables presentan entre ellas mimas. |

### Variables reemplazadas

La variable *Sitio\_consumo\_masfrecuente* se reemplazó por la variable *Consumo\_consolidado*. Debido a que la variable tiene 109 niveles y varios con poca cantidad de clientes. Con el fin de facilitar el análisis, el reemplazo fue agrupar los sitios usando la clasificación de REDEBAN para Categorías Mercantes[[6]](#footnote-6). Se elimina del *dataset* la variable *Sitio\_consumo\_masfrecuente.*

### Creación de nuevas variables

Con el fin de poder medir el consumo de los clientes, qué tarjetas usan, por cuánto las usan y si las usan el fin de semana, se crean las siguientes variables. Esto se hace con el fin de reducir la redundancia que presentan las variables medidas en porcentajes y poder asignar un monto de dinero para que las mismas sean comparables. Se agrega una sexta variable del total de tarjetas usadas por el cliente para identificar los clientes que tienen 1, 2 o 3 tarjetas y cómo las usan de acuerdo a las variables creadas.

|  |  |  |
| --- | --- | --- |
| **Variable nueva** | **Nombre Variable Nueva** | **Cálculo Variable Nueva** |
| Monto total | promedio\_total | promedio\_por\_transaccion \* Numero\_de\_transacciones |
| Total consumo Visa | visa | (porcentaje\_visa\_nacional + porcentaje\_visa\_internacional) \* promedio\_total |
| Total consumo MasterCard | mastercard | (porcentaje\_mastercard\_nacional + porcentaje\_mastercard\_internacional) \* promedio\_total |
| Total consumo nacional | total\_nacional | porcentaje\_nacional\_total \* promedio\_total |
| Total consumo fin de semana | fin\_de\_semana | (porcDOMINGO + porcSABADO) \* promedio\_total |
| Total tarjetas por cliente | tc\_usadas | Primero se calculan variables binarias que indican si se han utilizado cada tipo de franquicia (Visa, MasterCard y otras franquicias) y a partir de éstas, crea una variable adicional que representa la cantidad total de tarjetas de crédito usadas. |

### Eliminación de variables usadas por la creación de nuevas variables

|  |  |
| --- | --- |
| **Variables eliminadas** | **Justificación** |
| porcentaje\_nacional\_total, porcentaje\_internacional\_total | Ya se capturan en total de consumo nacional. Se elimina el porcentaje internacional debido a que tiene correlación. |
| porcLUNES,  porcMARTES  porcMIERCOLES  porcJUEVES  porcVIERNES,  porcSABADO  porcDOMINGO | Se eliminan porque tienen correlación negativa con la variable de fin de semana. |
| porcentaje\_visa\_nacional  porcentaje\_visa\_internacional  porcentaje\_mastercard\_nacional  porcentaje\_mastercard\_internacional  Porcentaje\_otrafranquicia\_nacional  porcentaje\_otrafranquicia\_internacional | Se eliminan ya que están consideradas en las variables de total consumo Visa y Mastercard.  Y se eliminan los porcentajes de la otra franquicia por correlación. |
| Numero\_de\_transacciones  promedio\_por\_transaccion  promedio\_total | Se eliminan por correlación con la variables creadas. |

### Separación de datos atípicos

Se procedió a la eliminación de 5072 valores atípicos del *dataset* utilizando como referencia la variable *promedio\_total*. Esta acción fue realizada con el propósito de obtener un conjunto de datos más apropiado para la aplicación del algoritmo de K-means. Los valores atípicos fueron identificados y removidos mediante la función de R denominada "remove\_outliers", los cuales conformarán un clúster aparte y particular. Es importante mencionar que este clúster de datos atípicos será tenido en cuenta al momento de diseñar estrategias de negocio, ya que su singularidad podría proporcionar información relevante para el banco.

La función "remove\_outliers" fue empleada para eliminar valores extremos, calculando los cuartiles y el rango intercuartílico (IQR) de la variable en cuestión, y marcando como valores faltantes (NA) aquellos que se encontraban por debajo del cuartil inferior restado de 1.5 veces el IQR, así como los que excedían el cuartil superior sumado a 1.5 veces el IQR. Se removieron aquellos valores marcados como NA del *dataset* pero fueron guardados aparte. Cabe mencionar que éstos valores ya tienen aplicadas las variables nuevas que se mencionaron anteriormente.

### Aplicación de logaritmos y estandarización de los datos

Finalmente, para generar un *dataset* que cumpla los estándares de K-means se aplicó la escala logarítmica a las variables seleccionadas y luego, los valores fueron estandarizados, de manera que las escalas de las variables sean similares y evitar que una variable domine sobre las demás en función de la escala.

## Dataset description

La base de datos final, que va a ser usada para identificar el número de clústeres óptimo, contiene información de consumo de un mes de 42799 tarjetahabientes. La base de datos contiene 5 columnas que presentan: 1) promedio de gasto total consumido con la tarjeta Visa, 2) promedio de gasto total consumido con la tarjeta MasterCard, 3) promedio del gasto total en comercios nacionales, 4) promedio del gasto total durante los fines de semana de un mes y 5) número total de franquicias utilizadas por el cliente.

# MODELING

## Select modeling techniques:

La elección de utilizar el algoritmo K-means para agrupar a los clientes se basa en tres consideraciones claves. La primera es la facilidad de determinar previamente el número óptimo de clústeres. Al contar con una gran cantidad de observaciones (más de 42.000), resulta más manejable emplear K-means, ya que el método jerárquico puede resultar complicado al tratar de comparar todas estas observaciones individualmente y decidir posteriormente el número de clústeres. La segunda consideración es la naturaleza escalar de las variables seleccionadas, lo que favorece la aplicación de la distancia euclidiana, que es la base del algoritmo K-means para asignar puntos a clústeres de manera eficiente. La tercera es que se ha realizado una preparación exhaustiva de los datos para asegurar resultados robustos. La transformación logarítmica y estandarización de las variables numéricas permiten homogeneizar las escalas y mejorar la precisión del algoritmo. Asimismo, la eliminación de outliers contribuye a evitar sesgos en la elección de los centros de los clústeres y a mejorar la calidad de las agrupaciones resultantes. Estas etapas de preprocesamiento preparan los datos de manera óptima para el método K-means y aseguran que el algoritmo opere en un espacio más controlado y representativo. En conjunto, estas consideraciones respaldan la elección de K-means como el enfoque más adecuado para la tarea de agrupar a los clientes de manera eficiente.

## Generate test design:

### Testeo con y sin outliers

El testeo con outliers, aunque produce resultados parecidos en una evaluación de 4 clústeres, presenta correlaciones más altas, que como se ha dicho anteriormente, pueden distorsionar los clústeres, por lo cual se elige hacer el proceso de clustering sin atípicos.

|  |  |
| --- | --- |
| **Clúster de 4 grupos Con Atípicos** | **Clúster de 4 grupos Sin Atípicos** |
|  |  |

### Comparación de número de clústeres

Teniendo claro que la remoción de valores atípicos impacta positivamente el análisis, se procede a definir el número de clústeres. Se evaluaron 3 criterios para la selección: tamaño de cada clúster, robustez de cada grupo y su visual. Se eligieron probar 2, 3 y 4 clústeres de acuerdo a lo sugerido por el diagrama de codo. De los resultados se destaca la selección de tres clústeres porque el tamaño de los mismos se encuentra relativamente mejor balanceado (similar cantidad de observaciones en cada grupo) y los resultados de la robustez son muy buenos, lo que indica que la selección de 3 clústeres es coherente y que, si hay pequeñas modificaciones en los datos, los resultados vararían muy poco.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **K=2** | **K=3** | **K=4** |
| **Tamaño de clústeres** | Clúster 1: 12.721  Clúster 2: 30.078 | Clúster 1: 17.909  Clúster 2: 12.438  Clúster 3: 12.452 | Clúster 1: 11.548  Clúster 2: 12.340  Clúster 3: 14.86  Clúster 4: 17.425 |
| **Robustez de clústeres** | Clúster 1: 0,8816454  Clúster 2: 0,8834124 | Clúster 1: 0,9981112  Clúster 2: 0,9347489  Clúster 3: 0,9370055 | Clúster 1: 0,8242105  Clúster 2: 0,4683553  Clúster 3: 0,4751701  Clúster 4: 0,4820576 |
| **Heatmap** |  |  |  |

## Build model:

Se escoge entonces un algoritmo de clúster usando K-means donde el numero elegido de clústeres es tres. Estos clústeres están basados en el promedio de gasto total consumido con las tarjetas Visa y MasterCard (no se incluye otra franquicia para evitar correlaciones en las variables), el promedio del gasto total en comercios nacionales (no se incluye el promedio de gasto total internacional para evitar correlación), el promedio del gasto total durante los fines de semana en un mes (siguiendo la lógica no se incluye gasto entre semana para evitar correlaciones) y el número total de tarjetas utilizadas por el cliente, variables que fueron estandarizadas previo a correr el algoritmo.

Teniendo en cuenta la remoción de atípicos, se debe aclarar que, aunque con el algoritmo se obtienen tres clústeres, el grupo de atípicos se conforma como su propio clúster y es por ende que, en total, el proyecto tendrá cuatro clústeres.

# EVALUATION:

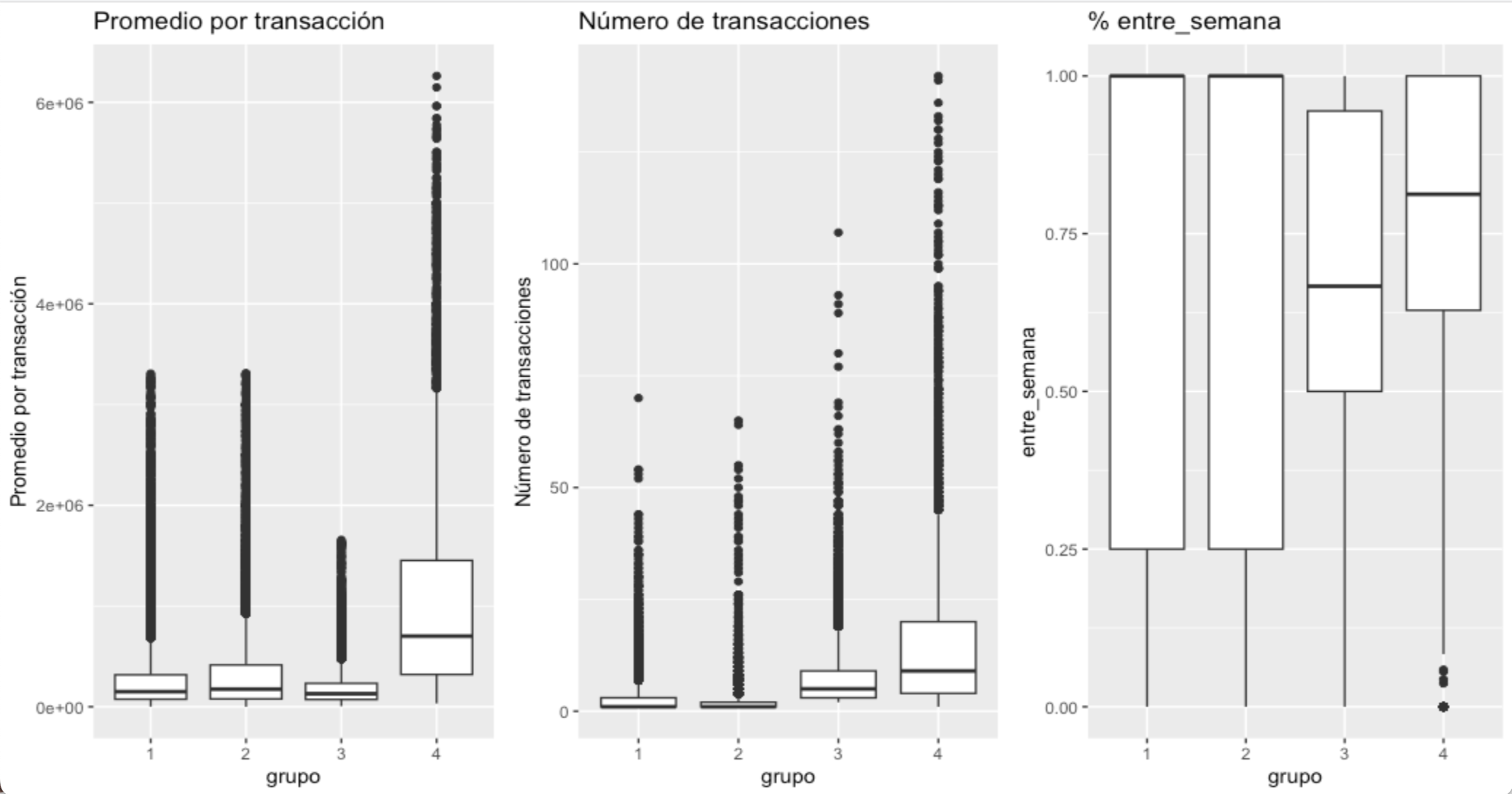
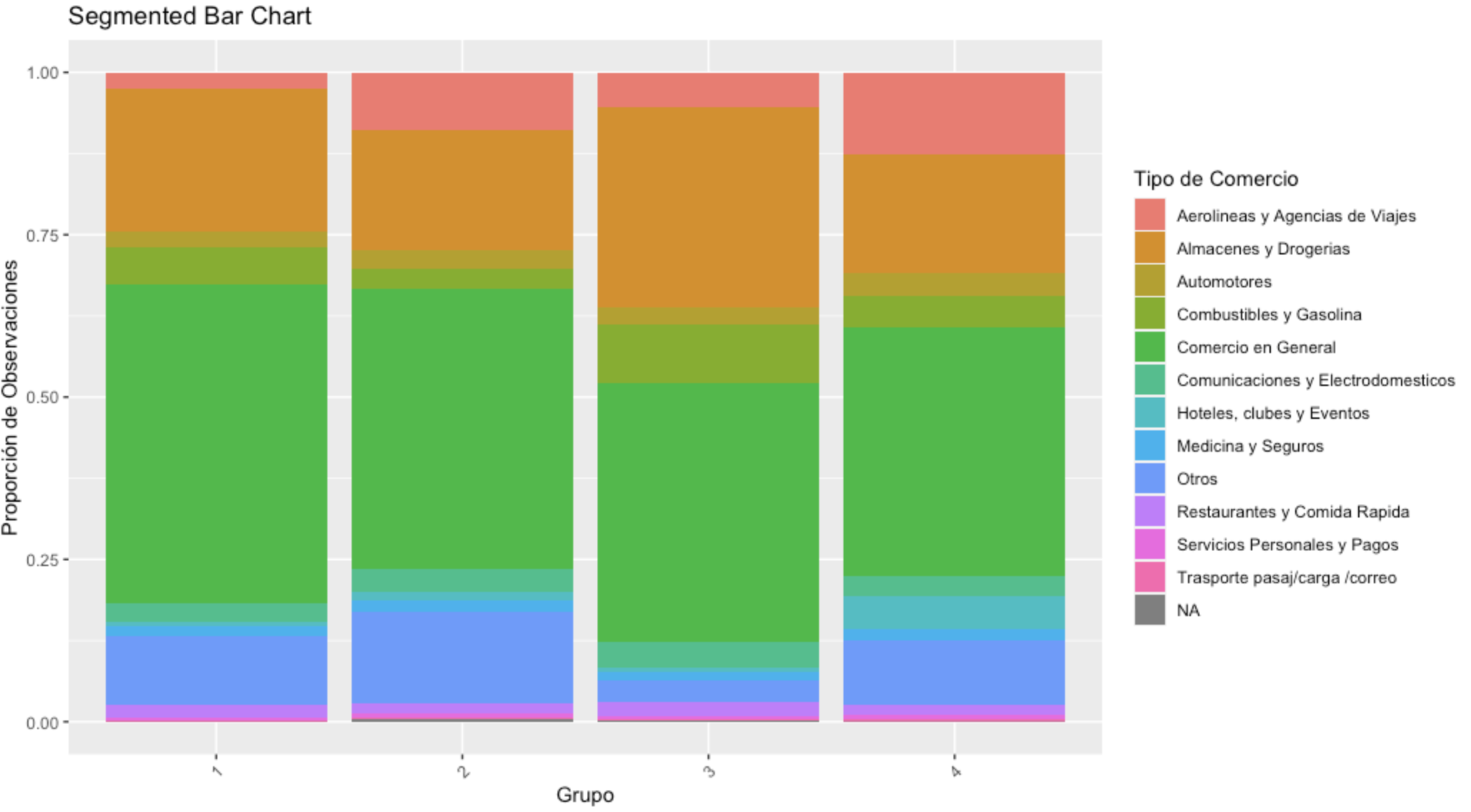
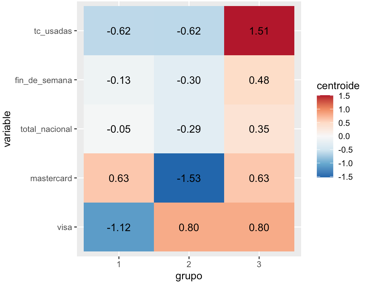
## Assess model:

Teniendo en cuenta que el criterio de éxito de modelo se había planteado como: que la medida de robustez de los clústeres sea mayor o igual a 0.75, y habiendo seleccionado un numero de clústeres tal que cada clúster tenga medida de robustez por encima de 0.9 (como se muestra en los apartados anteriores), se puede concluir que el modelo cumple el criterio técnico.

Ahora, con respeto al negocio, una vez identificadas las observaciones correspondiente a cada clúster (incluyendo el 4 grupo de atípicos) y cruzando con la base original, se puede hacer el siguiente perfilamiento de clientes:

### Perfilamiento de clústeres

|  |  |  |  |
| --- | --- | --- | --- |
| **Clúster 1** | **Clúster 2** | **Clúster 3** | **Clúster 4** |
| Clientes con uso importante de la franquicia de MasterCard, Su consumo se enfoca a comercios nacionales sobre todo en entre semana y en comercios de almacenes, droguerías y comercios generales. | Clientes con uso importante de la franquicia de Visa. Su consumo refleja pocas transacciones con un valor intermedio respecto a los clústeres 1 y 3. Su consumo, aunque bajo, se realiza entre semana, comercios de almacenes, droguerías, comercios generales y también en comercios de otro tipo. | Clientes con mayor número de tarjetas y gasto característico en fin de semana. Su consumo es un poco más bajo que el clúster 1 pero hace más transacciones que éste, lo que indica que son varios gastos al mes de poco valor. | Clientes con mayor número y valor de transacciones. Representan un segmento importante de gasto nacional como internacional. Se destaca que además de comercios tradicionales, también tiene un gasto relevante en aerolíneas, agencias de viajes y hoteles/clubs sociales. |



## Produce final report:

Retomando el objetivo de negocio plateado que es “incrementar la tasa de retención del clientes a partir de la implementación de estrategias comerciales para grupos de clientes específicos de acuerdo a los clústeres generados” se plantean las siguientes estrategias comerciales. Estas estrategias se concentran en 1) la *diferenciación* de los productos y servicios del banco y 2) *enfocándose* en dos grupos de clientes particulares.

### Estrategia y análisis CBA para el clúster 3

* ***Estrategia:* Cero cuota de manejo al realizar un número determinado de transacciones.**

Esta estrategia toma en cuenta que los clientes del clúster 3 tienen varias tarjetas de crédito y que, dadas las condiciones actuales, se encuentran pagando cuota de manejo sobre las mismas. Se evidencia entonces que es un grupo que tiene potencial para gastar más dinero con sus tarjetas. Razón por la cual, aprovechando que tienen un número alto de transacciones de poco valor, se puede incentivar su gasto implementado la condonación de la cuota de manejo.

* **Costo:** Al condonar la cuota de manejo se puede perder parte de los ingresos generados por estos pagos y el costo administrativo para el banco se incrementa.
* **Beneficio:** Se entra a competir con un beneficio adicional en el mercado, que, juntándolo con el *goodwill* de marca del banco, puede atraer a clientes que están incursionando por primera vez en el mercado de TCs e incentivas transacciones de mayor valor para los clientes actuales.
* **Mecanismo de evaluación:** Evaluar si el interés total generado por estas transacciones es mayor a la cuota de manejo de la tarjeta de crédito que se condona.

### Estrategia y análisis CBA para el clúster 4

* **Estrategia: Cashback y servicio personalizado para clientes con montos altos de consumo al mes.**

Teniendo en cuenta que este banco tradicional colombiano no tiene los mismos beneficios que sus competidores, se propone la implementación de 2 beneficios nuevos para este grupo de clientes. El primero es retornar a los clientes un pequeño porcentaje del dinero que gastan en ciertos comercios, como hoteles y aerolíneas/agencias de viaje. Esto puede ser posible partiendo de la base que el banco, por su posicionamiento en el mercado, tiene alianzas estratégicas con este tipo de comercios. El sistema funcionaría de la siguiente forma: por cada transacción que hace un cliente con su tarjeta en un comercio aliado, el comercio paga un porcentaje al banco. De este porcentaje se destina una parte para devolverle al cliente. Estos reembolsos pueden llegar a ser un excelente incentivo para que los clientes conserven su tarjeta de crédito y que potenciales clientes nuevos, con un nivel parecido de gastos, también abran soliciten tarjetas al banco. Y siendo que este es un segmento con un gasto importante, también debe tener un servicio especializado. Una de las oportunidades del banco es precisamente su servicio al cliente, y por el cual otros competidores del mercado se diferencian. Razón por la cual se propone que, complementario al cashback, se pueda aprovechar los nuevos canales digitales del banco para ofrecer asistencia en las transacciones de los clientes, con el fin de mostrar la capacidad que el banco tiene de ofrecer experiencias personalizadas.

* **Costos:** Debido a que el cashback se obtendría de las ganancias que el banco comparte con los comercios aliados, estas ganancias evidentemente se reducirían, así como el ingreso total por tarjeta. Además, la asistencia por canales digitales puede conllevar costos administrativos que también reducen el ingreso.
* **Beneficios:** Incentivar que los clientes permanezcan más tiempo con su tarjeta de crédito y promover el uso de ésta en comercios aliados (mayor gasto). Ambas cosas pueden generar que el mercado reconozca los nuevos beneficios del banco.
* **Mecanismo de evaluación:** Evaluar si el interés total generado por estas transacciones es mayor a suma de la cantidad de dinero retornado al cliente y los gastos administrativos del canal digital.

### Potenciales siguientes pasos

Para ambas estrategias es importante considerar los siguientes pasos:

* **Análisis de comportamiento de transacciones:** Realizar un análisis más detallado del comportamiento de transacciones de los clientes. Esto permitirá determinar el umbral adecuado para la condonación de la cuota de manejo o porcentaje de cashback y garantizar que la estrategia sea atractiva para los clientes.
* **Modelo de ingresos y costos:** Crear un modelo financiero que calcule los ingresos proyectados por las transacciones adicionales generadas por las estrategias en comparación con la posible pérdida de ingresos debido a la cuota de manejo condonada o al monto de dinero retornado. También, es necesario estimar los costos adicionales administrativos asociados con la implementación de esta estrategia en los canales digitales.
* **Análisis competitivo:** Realizar un análisis comparativo de otras ofertas en el mercado que compiten con las estrategias propuestas. Evaluar cómo se posicionan éstas en relación con las ofertas de la competencia y cómo pueden diferenciarse.
* **Análisis de aliados estratégicos:** Investigar y analizar las relaciones actuales del banco con comercios aliados, como hoteles y aerolíneas. Evaluar la viabilidad y el compromiso de estos aliados para participar en el sistema de cashback propuesto y si el porcentaje de ganancias compartidas es atractivo para ellos.
* **Evaluación del servicio personalizado**: Realizar una encuesta o investigación cualitativa para comprender las necesidades específicas de los clientes con montos altos de consumo al mes. Evaluar cómo se podría brindar un servicio personalizado a través de canales digitales y cómo esto afectaría la satisfacción del cliente y su lealtad.

# REFERENCIAS

Asobancaria. (2022). Protección Al Consumidor Financiero: Desarrollo Regulatorio. Asobancaria.com. <https://asobancaria.com/ws/semanas-economicas/1322-BE.pdf>

Banco de la República (2023). Tasa de usura. Banco de la República. <https://www.banrep.gov.co/es/glosario/tasa-usura>

El Tiempo. (2021, diciembre 14). Tarjeta de crédito: lo que tiene que saber si quiere cancelarla. El Tiempo. <https://www.eltiempo.com/economia/finanzas-personales/tarjeta-de-credito-como-cancelarlas-facilmente-638903>

Euromonitor International. (2023). Analysis of Credit Cards In Colombia analysis. Euromonitor International. <https://www.portal.euromonitor.com/?zKKgHpiPAQ85DBlRnEWmYmtHjj9nH8Ote0Kw2ZW8jOB424edK3OxjA%3d%3d>

Jalbert, T., Stewart, J. D., & Martin, D. (2010). The value of credit card benefits. *Financial Services Review, 19*(3), 227-244. Recuperado de <https://www.proquest.com/scholarly-journals/value-credit-card-benefits/docview/865926982/se-2>

López, J. (2023, junio 14). “Tasas de interés de tarjetas de crédito pasaron de estar en 40% a entre 20% y 25%”. Diario La República. Recuperado el 25 de agosto de 2023, de <https://www.larepublica.co/finanzas/las-tasas-de-las-entidades-en-tarjetas-de-credito-pasaron-de-40-a-entre-20-y-25-3636061>

MasterCard (2022). Reglamento operativo. Mastercard. Recuperado el 25 de agosto de 2023 de <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjKw63ckfiAAxXitoQIHS9RCOkQFnoECCQQAQ&url=https%3A%2F%2Fwww.mastercard.com.co%2Fcontent%2Fdam%2Fpublic%2Fmastercardcom%2Flac%2Fco%2Fhome%2Fcolombia-administradora%2Fmcca-reglamento-operativo-v7.pdf&usg=AOvVaw1GHosIBvdEo8_NwlN9V1zj&opi=89978449>

Mora Aguilar, K. (2023, mayo 20). Las empresas que están tras los 1,1 millones de datáfonos que operan en el mercado. Diario La República. Recuperado el 25 de agosto de 2023, de <https://www.larepublica.co/finanzas/las-empresas-que-estan-tras-los-1-1-millones-de-datafonos-que-operan-en-el-mercado-3619765>

Morales Arévalo, N. (2022, mayo 4). Redeban, Cajero y Bold, entre las opciones para que comercios adquieran datáfonos. Diario La República. Recuperado el 25 de agosto de 2023, de <https://www.larepublica.co/finanzas/redeban-cajero-y-bold-entre-las-opciones-para-que-comercios-adquieran-datafonos-3356562>

Olaya, M. (2019). ¿En qué usan los colombianos las tarjetas de crédito?. RCN Radio. Recuperado el 25 de agosto de 2023, de <https://www.rcnradio.com/economia/en-que-usan-los-colombianos-las-tarjetas-de-credito>

Palomino-Martínez, M. F. (2019). Tarjetas de crédito en Colombia: leyes, instituciones y efectos. Ciencia, Economía y Negocios, 3(1), 59-86. Recuperado el 25 de agosto de 2023, de <https://revistas.intec.edu.do/index.php/ciene/article/view/1511/2114>

REDEBAN (2023). Actividades consolidadas por MCC. REDEBAN. Recuperado de: <https://www.rbmcolombia.com/wps/wcm/connect/d4308449-d657-42f0-8a6d-6b4e5c8bfb79/Actividades+Consolidadas+MCC.pdf?MOD=AJPERES&CACHEID=ROOTWORKSPACE-d4308449-d657-42f0-8a6d-6b4e5c8bfb79-loblW1Q>

SEMANA. (2022, febrero 8). ¿Cuánto usaron los colombianos las tarjetas de crédito en 2021?. Revista Semana. Recuperado el 25 de agosto de 2023, de <https://www.semana.com/finanzas/credito/articulo/cuanto-usaron-los-colombianos-las-tarjetas-de-credito-en-2021/202202/#>

Superintendencia Financiera de Colombia. (2023). Reporte de Inclusión Financiera 2022. Superintendencia Financiera de Colombia. Recuperado el 25 de agosto de 2023, de <https://www.bancadelasoportunidades.gov.co/sites/default/files/2023-07/RIF2022%2018072023.pdf>

Visa Inc (2023). Visa Core Rules and Visa Product and Service Rules. Visa Inc. Recuperado el 25 de agosto de 2023, de <https://usa.visa.com/content/dam/VCOM/download/about-visa/visa-rules-public.pdf>

Zapata Quinchía, A. (2022, noviembre 25). Colombianos les están huyendo a las deudas con tarjeta de crédito. ElColombiano.com. Recuperado el 25 de agosto de 2023, de <https://www.elcolombiano.com/negocios/colombianos-les-estan-huyendo-a-las-deudas-con-tarjeta-OM19377786>

1. Ministerio de Hacienda y Crédito Público, decreto 519 de 2007 [↑](#footnote-ref-1)
2. Congreso de Colombia, Ley 1555 de 2012 [↑](#footnote-ref-2)
3. Congreso de Colombia, Ley 1676 de 2013 [↑](#footnote-ref-3)
4. Congreso de la República, Ley 510 de 1999 y Decreto 2555 de 2010 [↑](#footnote-ref-4)
5. Dato tomado del artículo del diario La Republica *En junio se cancelaron cerca de 420.800 tarjetas de crédito, 81% más que hace un año.* 4 de Octubre de 2022 [↑](#footnote-ref-5)
6. La red REDEBAN tiene a disposición en su página web los sitios de consumo consolidados por Actividad. Esta categorización, con leves modificaciones, fue la que se usó para agrupar la variable *Sitio\_consumo\_masfrecuente.* [↑](#footnote-ref-6)