

Cluster Analysis for a Colombian Bank's Promotional Strategy

Andrés Pérez Narváez

I. Entendimiento del negocio

Determinar objetivos de negocio

Antecedentes:

En Colombia, el uso de tarjetas de crédito ha experimentado una transformación notoria, pasando de ser un medio de pago ocasional a convertirse en una herramienta financiera omnipresente. A medida que la economía del país ha evolucionado, estas tarjetas se han integrado en la vida diaria de los colombianos, facilitando transacciones y ofreciendo flexibilidad en el manejo de sus finanzas personales.

A medida que el mercado de tarjetas de crédito se vuelve más competitivo, los bancos se encuentran desafiados a personalizar sus ofertas para diferentes segmentos de clientes. La fidelización de los usuarios a través de programas de recompensas, tasas de interés atractivas y servicios adicionales se convierte en una estrategia clave para retener y atraer a nuevos clientes en un entorno financiero dinámico. A continuación, se hace una valoración estratégica del entorno y de la compañía, Para examinar cada diagrama, se sugiere hacer clic en la descripción adjunta a la figura. En el caso del diagrama SWOT, se ha generalizado el contexto interno de la banca en Colombia, ya que no se dispone de información sobre el banco específico al que pertenece la base de datos.



[1. Valoración del entorno- PESTLE](#)

[Fig](#)



[Fig 2. Valoración del entorno-Fuerza de Porter](#)



[Fig 3. Análisis interno y externo – SWOT](#)

Objetivo empresarial:

1. Fomentar la lealtad del cliente y reducir la tasa de cancelación de tarjetas mediante programas de recompensas, ofertas personalizadas y servicios exclusivos.
2. Aumentar el número de tarjetahabientes atrayendo nuevos clientes a través de campañas de marketing dirigidas

Criterios de éxito empresarial:

1. Incrementar la retención de clientes.

Criterio de Éxito: Reducción de la tasa de cancelación de tarjetas en un 5% en los próximos 12 meses.

Indicadores Clave de Desempeño (KPIs):

- Porcentaje de retención de clientes.
- Incremento en el uso continuo de tarjetas por parte de los clientes existentes.

2. Expandir la Base de Clientes.

Criterio de Éxito: Aumento del 5% en la adquisición de nuevos tarjetahabientes en el próximo trimestre.

Indicadores Clave de Desempeño (KPIs):

- Tasa de crecimiento de la base de clientes.

Determinar los objetivos de la minería de datos.

Objetivo de la minería de datos:

- Implementar algoritmos de clusterización, como K-means, sobre los datos normalizados y reducidos por PCA.
- Desarrollar perfiles detallados para cada cluster, identificando características específicas de comportamiento transaccional que afectan la retención.
- Asignar etiquetas de cluster a los nuevos clientes, clasificándolos en grupos con perfiles similares a los clientes existentes de alto rendimiento.

Criterios de éxito de la minería de datos:

- Implementación de Algoritmos de Clusterización.
 - Lograr una reducción significativa en el valor de WCSS respecto a los clusters generados, indicando una cohesión interna fuerte de los grupos.
- Desarrollo de Perfiles Detallados
 - Obtener visualizaciones claras y comprensibles que destaquen las diferencias y similitudes entre clusters, facilitando la interpretación de perfiles.
- Asignación de Etiquetas a Nuevos Clientes
 - Garantizar que nuevos clientes sean clasificados en clusters de manera consistente y significativa, reflejando similitudes transaccionales relevantes.

II. Comprensión de datos

Describir datos: Se dispone de una base de datos con detalles sobre el consumo de 47,871 usuarios de tarjetas de crédito de un banco colombiano. Cada usuario tiene asociadas 26 variables que describen sus hábitos de consumo. A continuación, se presenta un resumen de estas variables.

Variables	Tipo de variable
Porcentaje de uso de TC por franquicia, ya sea Visa, MasterCard, otra o propia del banco, y si es para compras nacionales o internacionales	Decimal - Float64
Porcentaje de uso de TC por día y momento del día, ya sea día, tarde o noche.	Decimal - Float64
Grupo de cliente y sitio de consumo más frecuente	Objeto
Id y número de transacciones al mes	Entero - int64
Estadísticos de transacciones, mínimos, máximos, y desviaciones.	Decimal - Float64

Exploración de datos: El análisis descriptivo se aplicó a la base de datos completa. Es crucial entender la capacidad de gasto de los usuarios para ofrecer promociones adecuadas. El gráfico del promedio de consumo muestra que la mayoría de las transacciones son inferiores a 1 millón de pesos, posiblemente debido a limitaciones en el límite de la tarjeta y la capacidad de pago, como se observa en la Figura 4.

En la Figura 5, se detallan los momentos del día en que se realizan transacciones con tarjetas de crédito. Predomina la tarde, seguida por la mañana con menor actividad. Aunque las transacciones son menores en la noche, se registra un uso significativo a través de outliers, posiblemente atribuible a trabajadores que gestionan sus obligaciones financieras en ese horario. En la Figura 6, se evidencia que la mayoría de las transacciones corresponden a consumo nacional. Por otro lado, en la Figura 7, se destaca que los fines de semana representan el 46% de las transacciones semanales utilizando tarjetas de crédito.

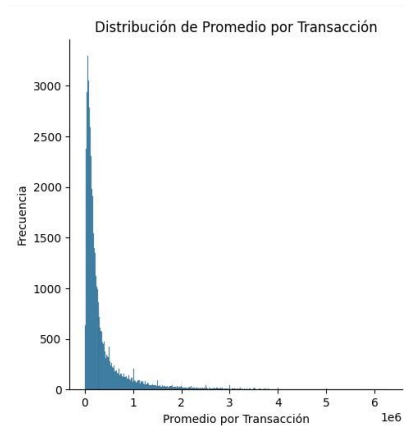


Fig 4. Promedio por transacción

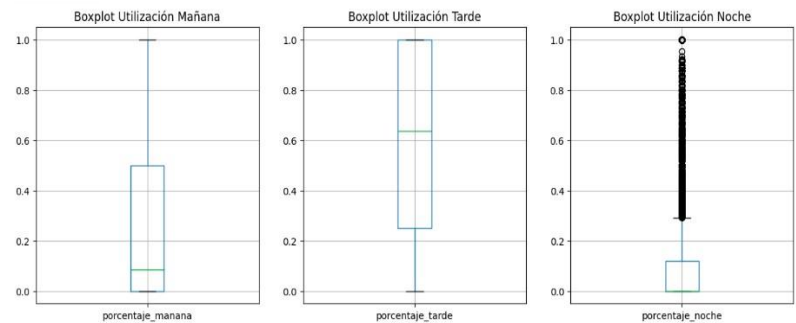


Fig 5. Utilización de TC por momento del día.

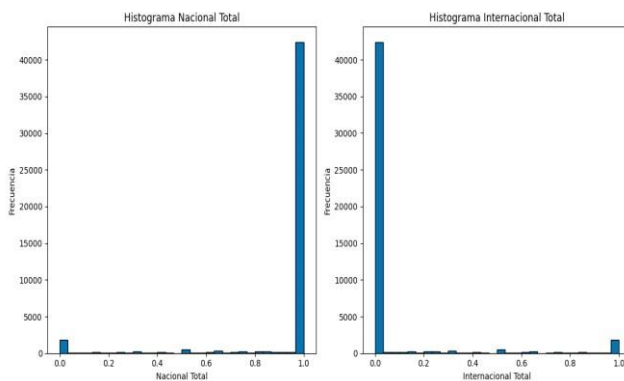


Fig 6. Uso nacional o internacional de TC

Distribución del Promedio de Uso de TC por Día de la Semana

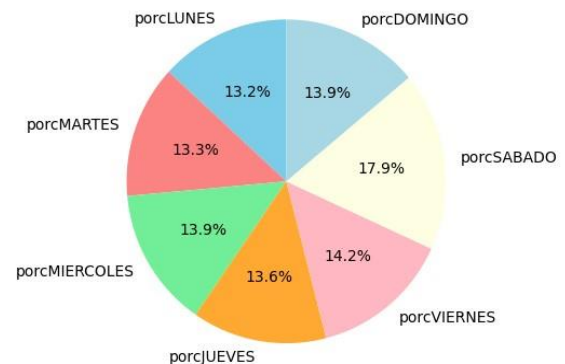


Fig 7. Utilización por día de la semana.

Verificar la calidad de los datos: Se constató la integridad de los registros para cada usuario, sin observar datos faltantes en la base de datos. Aunque el porcentaje de tarjetas de crédito en la noche presentaba valores atípicos, se retuvo por su abundancia, ya que podrían proporcionar perspectivas valiosas en análisis posteriores.

III. Preparación de datos

Limpieza de datos: Se excluyeron de la base de datos las columnas innecesarias para el estudio, buscando optimizar el modelo de clustering. Se eliminó "Sitio_consumo_masfrecuente" debido a la abundancia de datos y 'grupo_de_cliente' por falta de información relevante. También se descartaron 'transaccion_minima' y 'transaccion_maxima', junto con el "ID" que no aporta información relevante.

Tras un análisis de correlación, se eliminaron variables con alta correlación positiva para simplificar el modelo. Se excluyeron “porcentaje_visa_internacional”, “porcentaje_mastercard_internacional” y “porcentaje_otrafranquicia_internacional”, ya que están explicadas por “porcentaje_internacional_total”.

Construir datos: Inicialmente, se observaron los estadísticos descriptivos de cada columna y se identificó que los valores presentaban variaciones significativas entre ellas. Para garantizar una comparación equitativa entre las variables durante el análisis, se optó por normalizar los datos, asegurando que estén dentro del mismo rango. Este procedimiento se realizó utilizando la técnica de StandardScaler, que ajusta la escala y distribución de los datos. Como resultado, los datos transformados ahora comparten una escala común, lo que facilita la identificación de patrones y la evaluación de la importancia relativa de cada variable en análisis posteriores.

Descripción del conjunto de datos:

Finalmente, disponemos de un conjunto de datos con 47,871 entradas organizadas en 18 columnas. Se eliminaron 8 columnas del conjunto de datos original, y se procedió a normalizar la base de datos. Esta normalización simplifica el proceso de clustering de clientes, facilitando la extracción de insights valiosos que pueden ser utilizados por el banco.

IV. Modelado y Evaluación

Seleccionar técnicas de modelado:

La segmentación de clientes se llevará a cabo mediante el algoritmo K-means, que buscará identificar agrupaciones de clientes con características y comportamientos similares al comparar distancias entre datos. Previo a la aplicación de K-means, se realizará un análisis de Componentes Principales (PCA) para reducir la dimensionalidad del conjunto de datos.

Posteriormente, se aplicará el método del codo para determinar el número óptimo de clusters, considerando su eficiencia computacional. La medida de distancia empleada será la suma de los cuadrados dentro de cada grupo (WCSS), calculada mediante la distancia euclidiana al cuadrado entre cada punto y el centroide de su respectivo cluster. La minimización del valor de WCSS indicará que los puntos dentro de cada cluster son más similares, permitiendo que los clusters estén más distantes entre sí. Este enfoque facilitará una agrupación precisa de los clientes.

Generar diseño de pruebas: Una vez identificado que 10 componentes explican el 80.2% de la base de datos a través de PCA, se utilizarán varias herramientas para determinar la cantidad óptima de clústeres. Inicialmente, se visualizará cómo disminuye el valor de WSS a medida que aumentan los clústeres, permitiendo observar gráficamente el punto donde la tasa de cambio es menor. Además, se verificará este valor utilizando el método de la silueta, una medida que evalúa la coherencia de los clústeres.

Contrucción del modelo: En el método del codo, se observa que el valor de WCSS disminuye hasta alcanzar su mínimo con 10 clústeres. Sin embargo, crear 10 clústeres puede resultar en fragmentación excesiva. Al utilizar el coeficiente de silueta, figura 9, para determinar la cantidad óptima de clústeres de manera más eficiente, obtenemos un coeficiente de 0.2182. Según la gráfica correspondiente, este valor indica una separación moderada entre los clústeres, sugiriendo la elección de 7 u 9 clústeres como opciones más adecuadas. En este caso, se opta por 7 clústeres para lograr una fragmentación más efectiva de clientes.

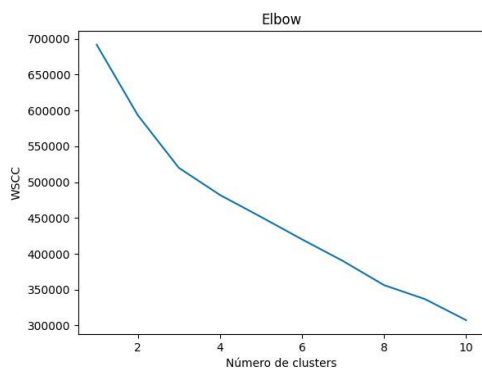


Fig 8. Método del codo.

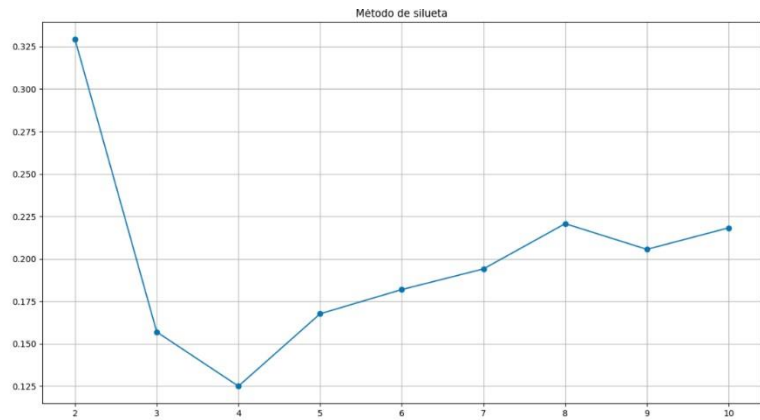


Fig 9. Método de la silueta

Luego se emplea el algoritmo de agrupamiento K-Means para clasificar los datos transformados por PCA en siete clusters distintos. Primero, se crea un modelo de K-Means con la especificación de siete clusters y una semilla aleatoria para garantizar la reproducibilidad. Luego, el modelo se ajusta a los datos transformados por PCA, asignando a cada punto de datos una etiqueta de cluster correspondiente. Finalmente, se imprime la asignación de etiquetas de cluster para cada punto de datos, lo que permite identificar la pertenencia de cada observación a uno de los siete grupos identificados por el modelo de K-Means.

Evaluar modelo:

Para evaluar el modelo es fundamental comprender a qué se refieren los 10 componentes que explican el 80.2% de la base de datos. Con este propósito, se ha generado la siguiente gráfica, fig 10, que detalla las variables más significativas en cada uno de estos componentes.

Fila/Componente	Variables Representativas
0	['porcentaje_internacional_total', 'porcentaje_noche', 'desviacion_estandar_por_transaccion']
1	['porcentaje_visa_nacional', 'porcentaje_tarde', 'promedio_por_transaccion']
2	['porcentaje_tarde', 'porcentaje_internacional_total', 'porcDOMINGO']
3	['promedio_por_transaccion', 'desviacion_estandar_por_transaccion', 'porcentaje_mastercard_nacional']
4	['porcSABADO', 'promedio_por_transaccion', 'desviacion_estandar_por_transaccion']
5	['porcVIERNES', 'porcJUEVES', 'Porcentaje_otrafranquicia_nacional']
6	['porcLUNES', 'porcMARTES', 'porcDOMINGO']
7	['porcJUEVES', 'porcDOMINGO', 'porcMARTES']
8	['porcLUNES', 'porcJUEVES', 'porcentaje_noche']
9	['porcDOMINGO', 'desviacion_estandar_por_transaccion', 'promedio_por_transaccion']

Fig 9. PCA – Variables representativas.

Inicialmente, se identifica la cantidad de tarjetahabientes por clúster, evidenciado en la Figura 10, donde cada clúster abarca una proporción significativa de la base de datos. Es importante recordar que el modelo explica al menos el 80,2% de la varianza en los datos. La puntuación de silueta, aproximadamente 0.22, indica que las observaciones dentro de los clusters están bien emparejadas y separadas de otros clusters. En términos generales, esta puntuación sugiere una calidad aceptable en la agrupación realizada por el modelo de k-means, destacando su cohesión intra-cluster y separación inter-cluster.

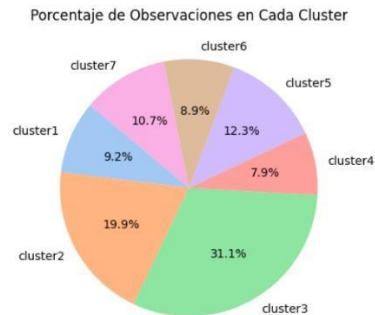


Fig 10. Usuarios por clúster.

V. Implementación

El modelo Kmeans posibilita la segmentación de los clientes de tarjetas de crédito en 7 grupos, basándose en sus patrones de consumo. A través de PCA, se logra explicar las variables que guardan una relación directa con sus hábitos de consumo. A continuación, se proporciona una descripción detallada de cada cluster, acompañada de una estrategia sugerida para alcanzar los objetivos de negocio mencionados al inicio del documento.

- A. Clúster 1: Este cluster engloba al 9,2% del total de clientes registrados. El 83,8% de estos clientes realiza sus transacciones principalmente los días lunes, todas ellas de carácter nacional y durante las horas de la tarde. En promedio, efectúan alrededor de 3 transacciones al mes, con valores medios de aproximadamente \$400,000. Cabe destacar que la autoría de la franquicia no presenta diferencias significativas.
- B. Clúster 2: Este clúster representa el 19,9% de toda la base de datos. La mayoría de las transacciones en este grupo son los días martes, jueves y viernes, de carácter nacional. Aproximadamente el 57% de estas transacciones se efectúan en la tarde, y el 85% de los clientes utiliza la franquicia Visa. En promedio, este segmento realiza alrededor de 4 transacciones mensuales, con valores promedio de \$410,000.
- C. Clúster 3: Este clúster es el más significativo, comprende la mayoría de los usuarios de tarjetas de crédito, representando el 31.3% del total. Se distinguen por mantener hábitos de uso uniformes, sin una marcada preferencia por un día específico de transacción, salvo una ligera inclinación hacia los días martes, jueves y viernes. En este grupo, el 57 % de las transacciones se realizan en la tarde, y un considerable 98 % son nacionales. Los usuarios de este clúster realizan, en promedio, alrededor de 8 transacciones al mes, siendo la franquicia Mastercard la preferida en la mayoría de los casos.
- D. Cluster 4: Este cluster, que comprende el 7.9% de la base de datos, se caracteriza por usuarios que realizan, en promedio, 8 transacciones al mes, con un valor promedio mensual por transacción de \$700,000. La mayoría de estas transacciones son de carácter internacional, representando el 80%, y se llevan a cabo principalmente en la tarde o noche. No hay una preferencia destacada por ningún día de la semana; todos los días presentan un uso regular de la tarjeta de crédito en este grupo.
- E. Cluster 5: Este cluster, que abarca el 12.3% del total de clientes, se caracteriza por el 88.7% de las transacciones realizadas los días sábados. La totalidad de estas transacciones son de carácter nacional, sin distinción en cuanto a la franquicia utilizada. Además, el 60% de las transacciones en este grupo tienen lugar en la tarde.

- F. Clúster 6: Este cluster, que representa el 8.9% de los datos, se distingue por realizar el 87.3% de las transacciones los días miércoles. Predominantemente, estas transacciones son de carácter nacional y se llevan a cabo en horario de la tarde.
- G. Clúster 7: Este cluster, que representa el 10.7% de los datos, destaca por realizar el 84% de sus transacciones los días domingo. Mayoritariamente, estas transacciones son de carácter nacional y se efectúan en horario vespertino.

A partir de los anterior se generan las siguientes estrategias para alinear los hábitos de consumo de cada clúster con los objetivos empresariales de fomentar la lealtad del cliente y aumentar la base de clientes.

A) Cluster 1:

- a. Promoción de puntos triples por cada transacción realizada los lunes por la tarde.
- b. Ofertas exclusivas en establecimientos nacionales los días lunes.

B) Cluster 2

- a. Descuentos especiales para transacciones realizadas los martes, jueves y viernes con tarjetas Visa.
- b. Programa de recompensas con bonificaciones adicionales al utilizar la franquicia Visa.

C) Cluster 3:

- a. Programa de lealtad premium para usuarios de Mastercard con beneficios exclusivos.
- b. Ofertas exclusivas en viajes y entretenimiento para transacciones realizadas en cualquier día.

D) Cluster 4:

- a. Programa de millas para usuarios que realicen transacciones internacionales.
- b. Descuentos especiales en establecimientos internacionales durante las noches.

E) Cluster 5:

- a. Ofertas especiales y descuentos en comercios nacionales los sábados.
- b. Participación en sorteos exclusivos al realizar transacciones los sábados por la tarde.

F) Cluster 6:

- a. Descuentos express en transacciones realizadas los miércoles por la tarde.
- b. Programa de recompensas con bonificaciones adicionales para usuarios activos los miércoles.

G) Cluster 7:

- a. Descuentos en spas y centros de bienestar para transacciones realizadas los domingos.
- b. Programa de puntos dobles para compras en establecimientos nacionales los domingos.

Para realizar un análisis costo/beneficio de las estrategias propuestas, podemos utilizar métricas clave que se alineen con los objetivos del negocio y las 4Ps del marketing (Producto, Precio, Plaza y Promoción). A continuación, proporciono un enfoque para evaluar el rendimiento de las estrategias, considerando diferentes aspectos dentro de cada P:

Producto:

- Beneficio: Aumento en la satisfacción del cliente debido a experiencias personalizadas.
- Costo: Implementación y mantenimiento de programas de recompensas y ofertas exclusivas.
- Beneficio: Incremento en el valor percibido por el cliente a través de descuentos y recompensas.
- Costo: Descuentos ofrecidos y posibles pérdidas de ingresos.

Plaza (Distribución):

- Beneficio: Mayor participación en transacciones a través de alianzas estratégicas con comercios asociados.
- Costo: Posibles inversiones en acuerdos de colaboración y descuentos para comercios. Promoción:
- Beneficio: Aumento en la retención y atracción de clientes mediante promociones específicas para cada clúster.
- Costo: Gastos en campañas de marketing y promociones.

Respecto a los mecanismos para evaluar, se propone lo siguiente recursos:

- Encuestas y Feedback de Clientes: Obtener comentarios directos sobre la percepción de las estrategias y su impacto.
- Seguimiento de Transacciones: Analizar el aumento en el uso de tarjetas y la adquisición de nuevos clientes.
- Análisis de Retorno de Inversión (ROI): Calcular el rendimiento financiero en comparación con los costos incurridos.

Simulación de Potenciales Objetivos y Resultados:

- A) Objetivo: Reducción de la tasa de cancelación de tarjetas en un 5% en 12 meses.
 - a. Simulación: Realizar análisis históricos para determinar la correlación entre las estrategias y la retención de clientes.
- B) Objetivo: Aumento del 5% en la adquisición de nuevos tarjetahabientes en el próximo trimestre.
 - a. Simulación: Ejecutar campañas de marketing específicas para medir el impacto en la adquisición de nuevos clientes.
- C) Objetivo: Incremento en el uso continuo de tarjetas de los clientes existentes.
 - a. Simulación: Implementar programas de recompensas y analizar la frecuencia de transacciones antes y después de la implementación.

Para dar seguimiento y profundizar en las estrategias propuestas, es crucial realizar un monitoreo constante del rendimiento, centrándose en métricas clave como la retención de clientes, la adquisición de nuevos tarjetahabientes y el aumento en el uso continuo de las tarjetas de crédito. Este análisis detallado a lo largo del tiempo permitirá ajustes en tiempo real para maximizar la efectividad de las estrategias.

Adicionalmente, se aconseja realizar encuestas y estudios de satisfacción del cliente para obtener retroalimentación directa sobre la percepción de ofertas personalizadas y programas de recompensas. Esta información cualitativa, reveladora de aspectos no capturados por datos cuantitativos, enriquecerá la comprensión de la experiencia del cliente.

Un paso adicional implica un análisis de segmentación más detallado dentro de cada clúster, identificando subgrupos o patrones específicos para ajustar estrategias según necesidades particulares. La evaluación financiera se centra en un ROI detallado, comparando beneficios generados con costos asociados. Un ROI positivo indicará el valor real generado por las estrategias. Finalmente, explorar tecnologías innovadoras como análisis predictivo o inteligencia artificial puede potenciar aún más la personalización de ofertas y anticipar necesidades futuras. Estos pasos garantizarán una optimización constante de las estrategias, alineándolas con los objetivos de negocio.