

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ingeniería en Electricidad y Computación

Agrupación de Instituciones Financieras en Ecuador Basada en el Uso
de Tarjetas de Crédito y Débito usando Técnicas de Aprendizaje No
Supervisado

PROYECTO DE TITULACIÓN

Previo la obtención del Título de:

Magister en Ciencias de Datos

Presentado por:

Alberto Alfredo Guime Vásquez

Kevin Andrés Zambrano Cortez

GUAYAQUIL - ECUADOR

Año: 2025

DEDICATORIA

A Giannella, mi amada esposa, por ser mi roca, mi inspiración y la motivación que guía mis días. Tu amor, paciencia y fe en mí han sido el motor que me impulsó a llegar hasta aquí.

A Sophia, mi pequeña luz. Eres mi mayor motivo de superación y mi mayor orgullo. Que este logro sea el primer paso hacia el futuro que sueño construir para ti. Con todo mi amor, para siempre.

A mi madre Lorena, por ser el pilar fundamental de mi vida, por sus sacrificios, su apoyo incondicional y su ejemplo de fortaleza y amor. Sin ti, este sueño nunca habría sido posible.

Y a Dios, por darme fuerza, sabiduría y guía en cada paso de este camino. A Él dedico este logro con gratitud infinita.

DEDICATORIA

A mis padres, Jaime y Elizabeth, cuyo amor, educación y ejemplo han sido la mayor fuente de inspiración para alcanzar mis metas. Gracias por inculcarme valores sólidos y una fe inquebrantable en Dios, que me han guiado en cada paso de este camino.

AGRADECIMIENTOS

Este proyecto no habría sido posible sin el valioso apoyo de la Escuela Superior Politécnica del Litoral (ESPOL) y la Facultad de Ingeniería en Electricidad y Computación (FIEC), instituciones que proporcionaron los recursos y el entorno académico necesarios para su desarrollo. Expresamos nuestro especial agradecimiento a nuestra profesora tutora, M.Sc. Karen Calva, por su guía experta y constante motivación, así como al Ph.D. José Córdova, coordinador de la maestría, por su liderazgo y dedicación al fortalecimiento académico. Su apoyo ha sido fundamental para alcanzar los objetivos de este trabajo.

DECLARACIÓN EXPRESA

Nosotros Alberto Alfredo Guime Vásquez y Kevin Andrés Zambrano Cortez acordamos y reconocemos que: La titularidad de los derechos patrimoniales de autor (derechos de autor) del proyecto de graduación corresponderá al autor o autores, sin perjuicio de lo cual la ESPOL recibe en este acto una licencia gratuita de plazo indefinido para el uso no comercial y comercial de la obra con facultad de sublicenciar, incluyendo la autorización para su divulgación, así como para la creación y uso de obras derivadas. En el caso de usos comerciales se respetará el porcentaje de participación en beneficios que corresponda a favor del autor o autores. Los estudiantes deberán procurar en cualquier caso de cesión de sus derechos patrimoniales incluir una cláusula en la cesión que proteja la vigencia de la licencia aquí concedida a la ESPOL.

La titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, secreto industrial, secreto empresarial, derechos patrimoniales de autor sobre software o información no divulgada que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada por nosotros durante el desarrollo del proyecto de graduación, pertenecerán de forma total, exclusiva e indivisible a la ESPOL, sin perjuicio del porcentaje que nos corresponda de los beneficios económicos que la ESPOL reciba por la explotación de nuestra innovación, de ser el caso.

En los casos donde la Oficina de Transferencia de Resultados de Investigación (OTRI) de la ESPOL comunique a los autores que existe una innovación potencialmente patentable sobre los resultados del proyecto de graduación, no se realizará publicación o divulgación alguna, sin la autorización expresa y previa de la ESPOL.

Guayaquil, 5 de marzo del 2025.

Alberto Alfredo Guime
Vásquez

Kevin Andrés
Zambrano Cortez

COMITÉ EVALUADOR

.....
M.Sc.Karen Priscilla Calva Yaguana

PROFESOR TUTOR

.....
María Isabel Mera Collantes, Ph.D.

PROFESOR EVALUADOR

RESUMEN

Este proyecto de titulación analiza el uso de tarjetas de crédito y débito en instituciones financieras de Ecuador mediante técnicas de aprendizaje no supervisado, con el objetivo de segmentar estas entidades según su comportamiento transaccional. Para ello, se emplearon algoritmos de clústering como K-Means y DBSCAN, permitiendo la identificación de patrones de uso y la clasificación de los bancos en distintos grupos.

Los datos utilizados provienen de la Superintendencia de Bancos del Ecuador y abarcan información transaccional detallada, la cual fue preprocesada mediante normalización, selección de características y reducción de dimensionalidad. Posteriormente, los modelos fueron evaluados con métricas como el coeficiente de silueta y el método del codo, lo que facilitó la identificación de clústeres representativos.

Los resultados revelan la existencia de cuatro segmentos principales: (1) instituciones especializadas en transacciones con tarjetas de débito, (2) bancos de baja actividad transaccional, (3) entidades líderes en crédito con un alto volumen de operaciones, y (4) bancos diversificados con un equilibrio entre crédito y débito. Esta segmentación proporciona información valiosa para la toma de decisiones estratégicas, optimización de productos financieros y gestión de riesgos en el sector bancario.

Se concluye que la aplicación de técnicas de clustering permite una mejor comprensión del ecosistema financiero ecuatoriano, facilitando el desarrollo de estrategias comerciales y regulatorias que promuevan la inclusión y la eficiencia en la industria bancaria. Asimismo, el análisis predictivo basado en estas agrupaciones podría contribuir a la formulación de políticas adaptadas a las tendencias emergentes en el mercado financiero.

Palabras clave: Clustering, Aprendizaje no supervisado, Finanzas, Segmentación, Tarjetas de crédito y débito.

ABSTRACT

This thesis project analyzes the use of credit and debit cards in Ecuadorian financial institutions using unsupervised learning techniques, aiming to segment these entities based on their transactional behavior. Clustering algorithms such as K-Means and DBSCAN were employed, enabling the identification of usage patterns and the classification of banks into distinct groups.

The data used comes from the Superintendency of Banks of Ecuador and includes detailed transactional information, which was preprocessed through normalization, feature selection, and dimensionality reduction. The models were then evaluated using metrics such as the silhouette coefficient and the elbow method, facilitating the identification of representative clusters.

The results reveal four main segments: (1) institutions specialized in debit card transactions, (2) banks with low transactional activity, (3) leading credit institutions with high transaction volumes, and (4) diversified banks maintaining a balance between credit and debit operations. This segmentation provides valuable insights for strategic decision-making, financial product optimization, and risk management within the banking sector.

The study concludes that clustering techniques offer a deeper understanding of the Ecuadorian financial ecosystem, supporting the development of commercial and regulatory strategies that promote inclusion and efficiency in the banking industry. Additionally, predictive analysis based on these groupings could contribute to the formulation of policies tailored to emerging trends in the financial market.

Keywords: *Clustering, Unsupervised learning, Finance, Segmentation, Credit and debit cards.*

COMITÉ EVALUADOR	vi
RESUMEN	vii
ABSTRACT	viii
ABREVIATURAS	xi
1 CAPÍTULO 1	11
1.1 Introducción	11
1.2 Descripción del Problema	11
1.3 Justificación del Problema.....	11
1.4 Objetivos.....	12
1.4.1 Objetivo General:.....	12
1.4.2 Objetivos Específicos:.....	12
1.5 Metodología:.....	12
1.5.1 Recopilación de Datos:.....	12
1.5.2 Preprocesamiento de Datos:	12
1.5.3 Implementación de Clústering:	13
1.5.4 Evaluación del Modelo:.....	13
1.5.5 Proyección de Clústers:.....	13
1.5.6 Documentación y Presentación:	13
1.6 Herramientas Y Librerías Utilizadas:.....	13
1.7 Set de datos.	14
1.7.1 Descripción.....	14
2 Capítulo 2	15
2.1 Preparación de los datos.....	18
2.2 Selección de características.....	18
2.3 Implementación del algoritmo k-means.....	19
2.4 Evaluación del modelo	21
2.5 Visualización de Resultados	23
3 Capítulo 3	25
3.1 Análisis Exploratorio de Datos.	25
3.2 Análisis de la Distribución de Transacciones por Tipo de Tarjeta.....	25
3.3 Distribución por Tipo de Pago.....	27

3.4	Monto Pagado por Segmento	28
3.5	Tendencia en uso de tarjetas de crédito y débito.....	30
3.6	Resultados del Clústering	32
3.7	Clústering con K-Means	32
3.8	Clústering con K-Means sin Banco Pichincha.....	39
3.9	Evaluación de la Calidad del Clústering.....	47
4	Capítulo 4.....	50
	Conclusiones	59
	Recomendaciones	61
	Bibliografía	63

ABREVIATURAS

PDS: Precision Dispatch System

API: Application Programming Interface

KPI: Key Performance Indicator

DBSCAN: Density-Based Spatial Clustering of Applications with Noise

K-Means: K-Means Clustering

EFT: Electronic Funds Transfer

SB: Superintendencia de Bancos

BI: Business Intelligence

POS: Point of Sale

ML: Machine Learning

DL: Deep Learning

CRM: Customer Relationship Management

CSV: Comma-Separated Values

ÍNDICE DE FIGURAS

Img 1 "Gráfico del análisis de silueta." Origen. (Ivan Ishchenko, Larysa Globa, Yurii Buhaienko, Andrii Liashenko, 2018)	21
Img 2 "Gráfico de la Curva del Codo mostrando la suma de las distancias al cuadrado para diferentes valores de k en un análisis de agrupamiento." Origen. (Ivan Ishchenko, Larysa Globa, Yurii Buhaienko, Andrii Liashenko, 2018)	22
Img 3 "Gráfico del criterio Gap Statistic." Origen. Origen. (Ivan Ishchenko, Larysa Globa, Yurii Buhaienko, Andrii Liashenko, 2018)	22
Img 4 "Grafico de K-means en plano bidimensional" Origen el autor.....	23
Img 5 "Gráfico de la Distribución de Distancias Dentro del Clúster." Origen el autor.	24
Img 6 "Gráfico de la Distribución de Numero de Transacciones por Marca de Tarjeta." Origen el autor.....	26
Img 7 "Gráfico de la Distribución de Numero de transacciones por canal de pago." Origen el autor.	27
Img 8 "Gráfico de la Distribución de Numero de Transacciones por tipo de Pago." Origen el autor.	28
Img 9 "Gráfico de la Distribución de Monto Pagado por Segmento." Origen el autor.	29
Img 10 "Gráfico de la tendencia Mensual de número de transacciones (Crédito y Débito)." Origen el autor.	30
Img 11 "Gráfico de la Tendencia Mensual de montos pagados (Crédito y Débito)." Origen el autor.	30
Img 12 "Gráfico del Codo." Origen el autor.....	32
Img 13 "Gráfico del ancho de silueta." Origen el autor.	33
Img 14 "Gráfico del ACP." Origen el autor.....	35
Img 15 "Gráfico del Ancho de Silueta." Origen el autor.....	37
Img 16 "Gráfico del ACP." Origen el autor.....	38
Img 17 "Gráfico del Ancho de Silueta." Origen el autor.....	40
Img 18"Gráfico del ACP." Origen el autor.....	41
Img 19 "Gráfico del Ancho de Silueta." Origen el autor.....	43
Img 20 "Gráfico del ACP." Origen el autor.....	45
Img 21 "EDA de Tarjetas de crédito." Origen el autor.	52
Img 22 "EDA de Tarjetas de crédito." Origen el autor.	53
Img 23 "EDA de Tarjetas de débito." Origen el autor.	54
Img 24 "EDA de Tarjetas de débito." Origen el autor.	55

Img 25 "Número de emisores por clúster." Origen el autor.	56
Img 26 "Razón Crédito/Débito entre clústeres." Origen el autor.....	57

ÍNDICE DE TABLAS

Tabla 1. Clústering con K-means=3	36
Tabla 2. Clústering con K-means=4	39
Tabla 3. Clústering con K-means=3 sin dato atípico	42
Tabla 4. Clústering con K-means=4 sin dato atípico	46
Tabla 5. Comparativa de valores de k-means.	48
Tabla 5. Cuadro de clasificación de bancos según su clúster con k-means=4	49
Tabla 6. Características por Clúster de Bancos según el Uso de Tarjetas de Crédito y Débito.....	51

1 CAPÍTULO 1

1.1 Introducción

El presente proyecto tiene como objetivo analizar y segmentar instituciones financieras en Ecuador en función de su comportamiento transaccional, utilizando técnicas de aprendizaje no supervisado. A través del uso de algoritmos de clústering como K-means y DBSCAN, se buscó identificar patrones en el uso de tarjetas de crédito y débito, permitiendo clasificar a los bancos en clústeres específicos. Este análisis proporciona información clave para la toma de decisiones estratégicas, optimización de productos financieros y asignación de recursos, contribuyendo al desarrollo del sector bancario en el país.

1.2 Descripción del Problema

El sector bancario ecuatoriano ha experimentado un crecimiento significativo en el uso de tarjetas de crédito y débito, generando grandes volúmenes de datos transaccionales. Sin embargo, la cantidad de datos disponibles no se ha aprovechado para identificar patrones y comportamientos en las instituciones financieras. Aquí es donde las técnicas de aprendizaje no supervisado, como el clústering, pueden ser de gran utilidad. Estas técnicas permiten agrupar instituciones financieras según similitudes en sus datos transaccionales sin necesidad de etiquetas predefinidas. Al aplicar algoritmos como K-means y DBSCAN, se pueden identificar patrones ocultos y segmentar a las instituciones financieras en grupos con comportamientos similares, proporcionando una visión más profunda y detallada que puede informar estrategias operativas y de marketing.

1.3 Justificación del Problema

El análisis de grandes volúmenes de datos transaccionales ofrece una oportunidad única para entender mejor el comportamiento de los usuarios y las estrategias operativas de las instituciones financieras. La utilización de técnicas de aprendizaje no supervisado, como el clústering, permitirá identificar grupos de bancos con patrones de comportamiento similares, proporcionando información valiosa para

mejorar la toma de decisiones estratégicas. Este estudio no solo beneficiará a las instituciones financieras, sino también a los reguladores y a los usuarios finales, al facilitar servicios financieros más adaptados y eficientes contribuyendo a la formulación de políticas que promuevan una mayor eficiencia y competitividad en el sector bancario ecuatoriano.

1.4 Objetivos

1.4.1 Objetivo General:

- Clasificar las instituciones financieras en Ecuador según sus patrones de uso de tarjetas de crédito y débito.

1.4.2 Objetivos Específicos:

- Implementar algoritmos de clústering como K-means y DBSCAN para identificar grupos de bancos con comportamientos similares.
- Analizar las características que diferencian a cada clúster de bancos.
- Evaluar la estabilidad y la interpretación de los clústers generados.

1.5 Metodología:

1.5.1 Recopilación de Datos:

- Los datos se obtuvieron de la Superintendencia de Bancos del Ecuador, de un archivo Excel titulado "Consumos Tarjetas consolidado MARZO 2024". Se puede acceder a los datos en el siguiente enlace: <https://www.superbancos.gob.ec/estadisticas/portalestudios/servicios-financieros/#01ONK3WL6477HFCFZ6VZAK7MAVEVU5KAFG> .
- Los datos incluyen información mensual detallada sobre transacciones realizadas con tarjetas de crédito y débito para varias instituciones financieras desde diciembre 2023 hasta noviembre 2024.

1.5.2 Preprocesamiento de Datos:

- Limpieza y normalización de datos para asegurar su calidad.

- Selección de características relevantes, como el número de cuentas con pagos, el número total de transacciones, el monto total facturado, el tique promedio (monto/número de transacciones), y la moda de la franquicia de la tarjeta (crédito o débito).

1.5.3 Implementación de Clústering:

- Ajuste de parámetros y evaluación inicial de los modelos.
- Aplicación de algoritmos de clústering (K-means) para formar grupos basados en patrones de número de transacciones y montos pagados.

1.5.4 Evaluación del Modelo:

- Evaluación de los clústers formados utilizando el ancho de silueta para determinar la calidad y estabilidad de los clústers.

1.5.5 Proyección de Clústers:

- Utilización de los datos de noviembre 2024 para proyectar posibles clústers en los meses siguientes.
- Análisis predictivo para identificar patrones futuros y tendencias.

1.5.6 Documentación y Presentación:

- Elaboración del informe final.
- Preparación de presentaciones y publicaciones académicas.
- Presentación de los resultados a las partes interesadas.

1.6 Herramientas Y Librerías Utilizadas:

- Python: Lenguaje principal para el análisis de datos y la implementación de los modelos de clústering.
- Pandas: Para la manipulación y análisis de los datos.
- NumPy: Para operaciones numéricas y matemáticas.
- Scikit-learn: Para la implementación de algoritmos de clústering (K-means, clústering jerárquico, DBSCAN, y Gaussian mixture).
- Matplotlib y Seaborn: Para la visualización de datos y resultados.

- PyCaret: Librería para la automatización de los procesos de machine learning, incluyendo la comparación de diferentes modelos de clústering y la evaluación de su desempeño.

1.7 Set de datos.

1.7.1 Descripción.

El set de datos utilizado en este estudio proviene de la Superintendencia de Bancos del Ecuador y contiene datos transaccionales relacionados con el uso de tarjetas de crédito y débito. El archivo proporcionado incluye información mensual sobre las transacciones de tarjetas para varias instituciones financieras en Ecuador, detallando el número de cuentas con pagos, el número de transacciones totales, el monto total facturado, y otras variables adicionales.

1.7.2 Estructura.

El set de datos incluye las siguientes columnas principales:

- Entidad Emisora: Nombre de la institución financiera emisora de las tarjetas.
- Tipo de Métrica:
 - Suma de Cuentas con pagos: Número de cuentas con pagos realizados.
 - Suma de Transacciones totales de pago: Número total de transacciones realizadas.
 - Suma de Monto total facturado: Monto total facturado por las transacciones.
- Tique Promedio: Monto total facturado dividido por el número total de transacciones.
- Moda de la Franquicia de la Tarjeta: Identificación de la franquicia más utilizada (crédito o débito).
- Datos mensuales que indican los valores para cada tipo de métrica en los meses especificados.
- Meses:
 - 2023: diciembre.
 - 2024: enero, febrero, marzo, abril, mayo, junio, julio, agosto, septiembre, octubre, noviembre.

2 Capítulo 2

El sector financiero en Ecuador ha crecido mucho en el uso de tarjetas de crédito y débito, lo que ha generado muchos datos sobre las transacciones y el comportamiento de los usuarios. Sin embargo, la diversidad y complejidad de estos datos presentan desafíos significativos para las instituciones financieras que buscan entender mejor el mercado y mejorar sus servicios. En este contexto, la agrupación de instituciones financieras basada en el uso de tarjetas de crédito y débito utilizando técnicas de aprendizaje no supervisado ofrece una solución innovadora. Estas técnicas, que incluyen algoritmos como k-means, clústering jerárquico y modelos de mezcla gaussiana, permiten identificar patrones y segmentar las instituciones en grupos homogéneos. Los resultados de esta agrupación pueden ser utilizados para optimizar estrategias de marketing, personalizar productos financieros, mejorar la gestión de riesgos y fomentar la colaboración entre instituciones. En última instancia, estos beneficios no solo mejoran la eficiencia operativa de las instituciones financieras, sino que también contribuyen a un mejor servicio al cliente y a un crecimiento más sostenible del sector financiero en Ecuador. (Superintendencia de Bancos del Ecuador, 2024)

En el marco regulatorio establecido por el "Código Orgánico Monetario y Financiero 2024", se definen con precisión las obligaciones de las entidades bancarias en relación con la entrega de información a la Superintendencia de Bancos. A continuación, se describe la información detallada de los datos recabados a utilizar en los reportes mensuales:

Año y Mes: Indican el periodo durante el cual se realizaron las transacciones. Esta temporalidad permite realizar análisis de tendencias y comportamientos de uso de las tarjetas en diferentes épocas del año, similar a las series históricas proporcionadas por la Superintendencia.

Emisor: Se refiere al banco o entidad financiera que emitió la tarjeta. La Superintendencia recopila y reporta datos de múltiples instituciones financieras, lo cual permite una comparación y análisis entre diferentes emisores.

Tipo de Tarjeta: Diferencia entre tarjetas de crédito y débito, lo cual es fundamental para entender el tipo de producto financiero utilizado en las transacciones. Estos datos son cruciales para los informes de la Superintendencia que clasifican el uso de diferentes tipos de tarjetas.

Marca de Tarjeta: Especifica la red o marca de la tarjeta (por ejemplo, Visa, MasterCard, etc.). La Superintendencia también clasifica los datos según la marca de la tarjeta para identificar la participación de mercado de cada red.

Clasificación y Segmento: Incluye información sobre el tipo de cliente según su score crediticio. La Superintendencia podría utilizar estas categorías para analizar el comportamiento de diferentes segmentos de clientes.

Ubicación de la Transacción: Indica si la transacción se realizó a nivel nacional o en el exterior. La Superintendencia reporta estadísticas sobre transacciones nacionales e internacionales, lo cual es útil para entender el alcance y uso global de las tarjetas emitidas en Ecuador.

Canal de Pago: Define dónde se realizó la transacción (por ejemplo, POS, Internet, Oficina). Esta información es relevante para la Superintendencia al analizar cómo y dónde los consumidores realizan sus pagos.

Tipo de Pago: Incluye detalles sobre la modalidad de pago (Corriente-Rotativo, Diferido sin intereses, etc.). La Superintendencia utiliza estos datos para comprender las preferencias de pago de los consumidores y la estructura del crédito en el país.

Número de Transacciones y Monto Pagado: Proveen un conteo y el valor total de las transacciones realizadas. La Superintendencia recopila y publica estos datos para ofrecer una visión cuantitativa del uso de tarjetas en el país, permitiendo un análisis del volumen de transacciones y el gasto asociado. (Corporación del Seguro de Depósitos (COSEDE), 2024)

2.1 Preparación de los datos

La preparación de los datos para el clústering con k-means comienza con la carga del dataset en un entorno de análisis, como Python. Luego, se realiza una limpieza de datos para manejar valores faltantes, eliminar duplicados y corregir errores. Las variables categóricas son transformadas a formato numérico mediante técnicas como One-Hot Encoding. (Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E., 2011)

A continuación, se procede a la selección de características relevantes para el análisis, como el número de transacciones y el monto pagado. Por último, los datos son escalados para asegurar que todas las características tengan la misma influencia en la formación de clústers, utilizando técnicas de normalización o estandarización. Con estos pasos, los datos estarán listos para la implementación del algoritmo de clústering.

2.2 Selección de características

La selección de características es un paso fundamental en la preparación de los datos para el clústering. Primero, se debe identificar las variables relevantes que tendrán un impacto significativo en la formación de clústers. En el contexto del dataset de tarjetas de crédito y débito, algunas características clave podrían ser el número de transacciones, el monto pagado, el año y el mes.

Una vez seleccionadas las características, es importante considerar la reducción de dimensionalidad si el número de características es muy alto, lo que puede simplificar el modelo y mejorar el rendimiento. Técnicas como PCA (Análisis de Componentes Principales) pueden ser útiles para reducir la dimensionalidad manteniendo la mayor cantidad de variabilidad posible en los datos. (Hastie, T., Tibshirani, R., & Friedman, J., 2009)

Luego, se procede al escalado de datos. Los algoritmos de clústering, como k-means, son sensibles a las diferentes escalas de las características. Por lo tanto, es crucial normalizar o estandarizar las características seleccionadas para asegurar que todas tengan el mismo peso en el análisis. La normalización transforma las características

para que estén en un rango uniforme, mientras que la estandarización ajusta los datos para que tengan una media de cero y una desviación estándar de uno. (Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E., 2011)

2.3 Implementación del algoritmo k-means

La implementación del algoritmo k-means comienza con la determinación del número óptimo de clústers (k). Para esto, se utiliza la técnica del ancho de silueta la cual es una medida que cuantifica qué tan similar es un punto a su propio clúster en comparación con otros clústers. Se calcula para cada punto en el dataset y proporciona un valor entre -1 y 1. (ROUSSEEUW, 1987)

- **Definición:** Para evaluar el desempeño de todo el clústering, se calcula el ancho de silueta promedio de todos los puntos:

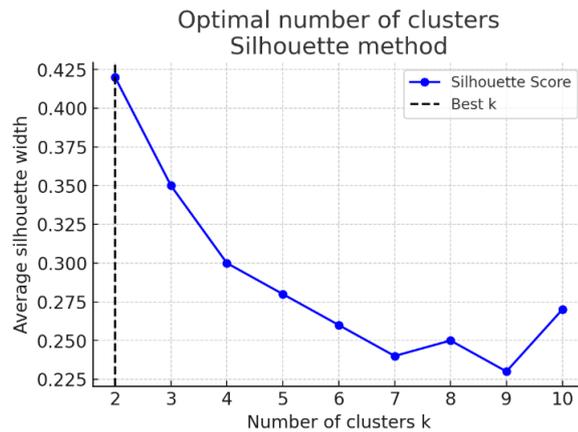
$$S = \frac{1}{n} \sum_{i=1}^n s(i)$$

Donde:

- n es el número total de puntos en el dataset.
 - s(i) es el coeficiente de silueta de cada punto.
- **Cálculo del coeficiente de silueta (s(i)):**

$$s(i) = \frac{b(i) - a(i)}{\max(a(1), b(i))}$$

- Si s(i) es cercano a 1, significa que el punto está bien agrupado dentro de su clúster.
 - Si s(i) es cercano a 0, significa que el punto está en el borde de dos clústers.
 - Si s(i) es negativo, significa que el punto podría estar en el clúster incorrecto.
- **Interpretación del Ancho de Silueta:**
 - Un valor promedio cercano a 1 indica clústers bien formados.
 - Un valor cercano a 0 sugiere clústers solapados o mal definidos.
 - Un valor negativo indica que los puntos están más cerca de otros clústers que del propio, lo que podría sugerir una mala asignación de clústers.



Img 1 "Gráfico del análisis de silueta." Origen. (Ivan Ishchenko, Larysa Globa, Yurii Buhaienko, Andrii Liashenko, 2018)

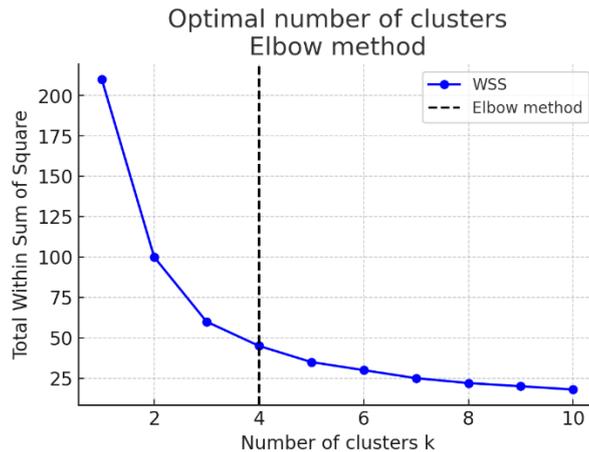
Una vez determinado k (Img 1), se procede a aplicar el algoritmo k-means. Este proceso implica inicializar k centroides aleatoriamente, asignar cada punto de datos al centroide más cercano, recalculando los centroides como la media de los puntos asignados y repetir estos pasos hasta que los centroides no cambien significativamente o se alcance el número máximo de iteraciones; durante este proceso, se monitorea la inercia (SSE) y el coeficiente de silueta para asegurar que los clústers sean compactos y bien definidos. Al finalizar, los puntos de datos estarán agrupados en k clústers, cada uno con su propio centroide, listos para ser analizados y evaluados más a fondo.

2.4 Evaluación del modelo

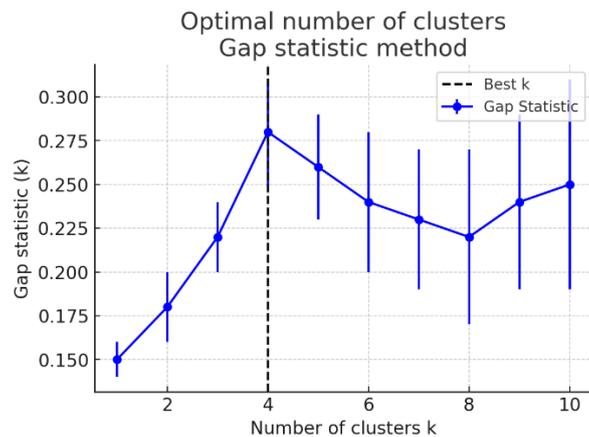
La evaluación del modelo de clústering comienza con la interpretación de los clústers formados. Esto implica analizar las características promedio de los puntos en cada clúster para entender qué representa cada grupo. Por ejemplo, un clúster puede representar clientes de alto valor si tiene un alto número de transacciones y un monto pagado elevado.

A continuación, se evalúa la calidad del clústering utilizando varias métricas adicionales tales como:

- El Método del Codo (Elbow Method), que grafica la suma de los errores cuadráticos (SSE) contra el número de clústers y busca el punto donde la disminución de SSE se vuelve menos pronunciada, (*Img 2*)
- Otra técnica es el criterio de Gap Statistic, el cual compara los resultados de clústering con un modelo nulo, lo que ayuda a evitar la selección de un número excesivo o insuficiente de clústers. (*Img 3*) (Ivan Ishchenko, Larysa Globa, Yurii Buhaienko, Andrii Liashenko, 2018)



Img 2 "Gráfico de la Curva del Codo mostrando la suma de las distancias al cuadrado para diferentes valores de k en un análisis de agrupamiento." Origen. (Ivan Ishchenko, Larysa Globa, Yurii Buhaienko, Andrii Liashenko, 2018)



Img 3 "Gráfico del criterio Gap Statistic." Origen. Origen. (Ivan Ishchenko, Larysa Globa, Yurii Buhaienko, Andrii Liashenko, 2018)

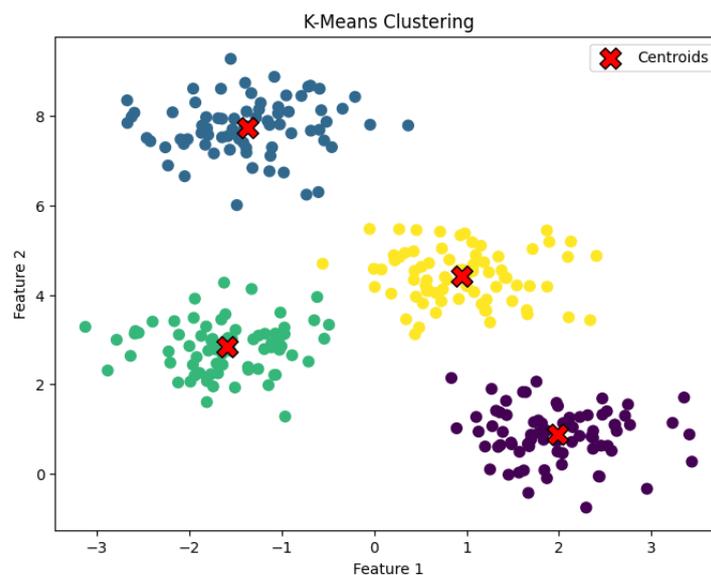
Además de estas métricas, es útil visualizar los clústers para una mejor interpretación. Gráficos de dispersión (scatter plots) pueden mostrar cómo se distribuyen los puntos en función de dos características seleccionadas, coloreados por clúster. Pair plots permiten ver las relaciones bivariadas entre múltiples características, y heatmaps

pueden mostrar la media de las características por clúster, facilitando la interpretación de las características dominantes en cada clúster. (Moosa, 2020)

Finalmente, se realiza una evaluación continua para ajustar el modelo según sea necesario, asegurando que los clústers sigan siendo relevantes y útiles para el análisis y toma de decisiones.

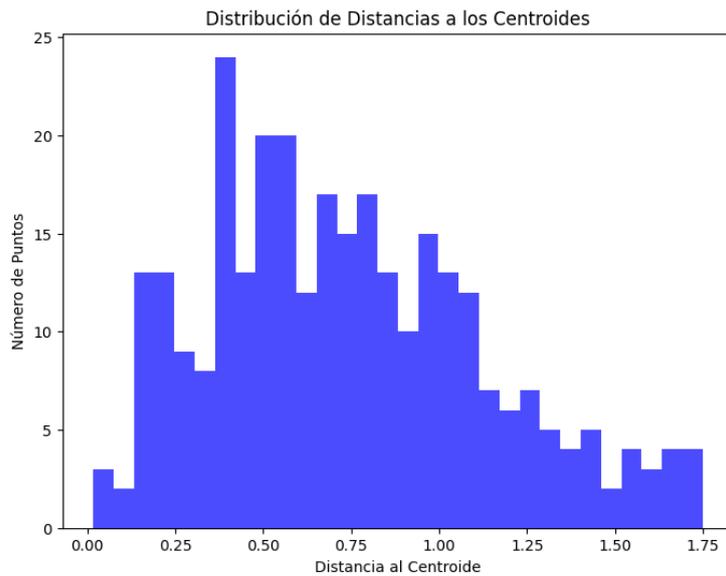
2.5 Visualización de Resultados

La visualización de los resultados del clústering comienza con la creación de gráficos formados en un plano bidimensional, que muestran cómo se distribuyen los puntos de datos en función de dos características seleccionadas, con cada punto coloreado según el clúster al que pertenece. Esto ayuda a visualizar la separación y compacidad de los clústers. (Img 4) (Moosa, 2020)



Img 4 "Gráfico de K-means en plano bidimensional" Origen el autor.

Para una vista más detallada de las características promedio de cada clúster, se puede utilizar un Gráfico de la Distribución de Distancias Dentro del Clúster el cual muestra la distribución de las distancias de los puntos a sus centroides, lo que ayuda a entender la dispersión dentro de los clústers. (Img 5) (Moosa, 2020)



Img 5 "Gráfico de la Distribución de Distancias Dentro del Clúster." Origen el autor.

Finalmente, es importante interpretar las visualizaciones en el contexto del problema de negocio. Esto implica identificar patrones significativos y traducirlos en insights accionables. Por ejemplo, los clústers pueden representar diferentes segmentos de clientes con comportamientos de gasto distintos, lo que puede guiar estrategias de marketing y toma de decisiones.

3 Capítulo 3

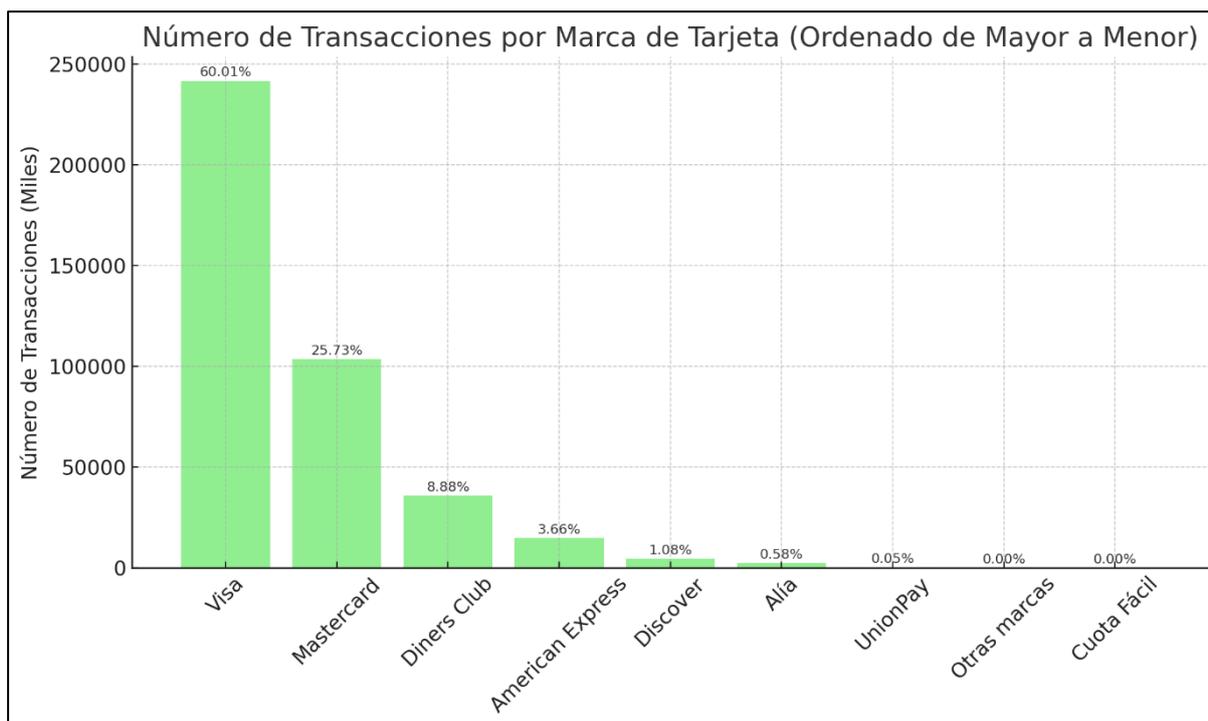
En esta sección se presentan el análisis exploratorio de datos, sus insights y los resultados obtenidos tras la aplicación de las técnicas de clústering y aprendizaje no supervisado a los datos transaccionales de las instituciones financieras en Ecuador. El objetivo fue identificar patrones de uso de tarjetas de crédito y débito, agrupar a las instituciones financieras en clústeres y analizar las características que diferencian a cada uno de estos grupos.

3.1 Análisis Exploratorio de Datos.

El objetivo del análisis es identificar los patrones de uso de tarjetas de crédito en diferentes segmentos de mercado, ubicaciones geográficas y canales de pago de un set de datos que va de octubre/2023 a septiembre/2024 en un único archivo .csv de los reportes financieros mensuales de las IFIs a la Superintendencia de Bancos del Ecuador. A partir de estos patrones, se busca proponer estrategias comerciales que permitan optimizar las ofertas y mejorar la experiencia del cliente. En este análisis se detectaron varias oportunidades para promover el crecimiento y la eficiencia del negocio en las áreas de pagos digitales, diferenciación por tipo de tarjeta y comportamiento de gasto de los clientes.

3.2 Análisis de la Distribución de Transacciones por Tipo de Tarjeta

El gráfico Img 6 muestra que Visa es, con diferencia, la marca de tarjeta más utilizada, acumulando más de 250 millones de transacciones, seguida por Mastercard, aunque con un número significativamente menor. Diners Club y American Express también tienen una presencia destacada, aunque mucho más modesta. Otras marcas como Cuota Fácil, Discover, UnionPay, y Alía tienen un número mucho menor de transacciones, lo que indica una menor penetración en el mercado. En resumen, Visa y Mastercard dominan el mercado, mientras que las demás marcas tienen una participación más limitada.

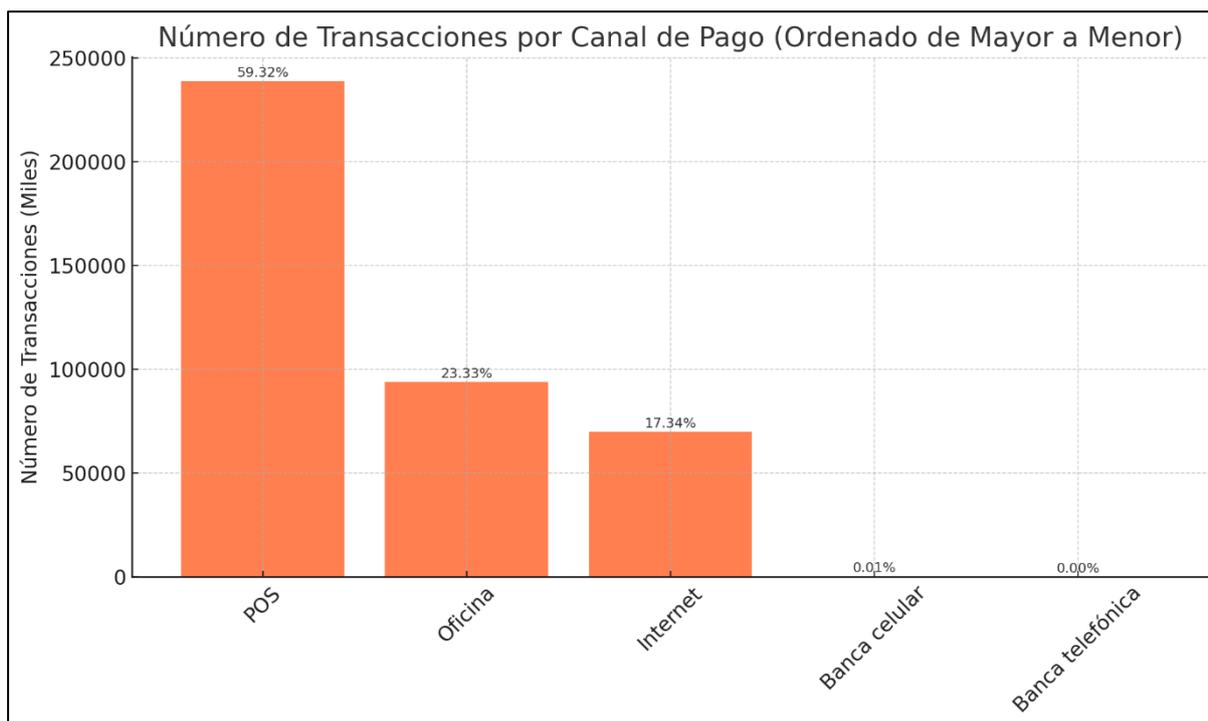


Img 6 "Gráfico de la Distribución de Numero de Transacciones por Marca de Tarjeta." Origen el autor.

La estrategia comercial que surge de este análisis debe enfocarse en la diversificación de los productos financieros. Las entidades bancarias podrían buscar promover el uso de tarjetas de débito mediante incentivos personalizados, tales como descuentos en comercios asociados, o recompensas por transacciones de débito frecuentes. Al mismo tiempo, se podrían desarrollar campañas de educación financiera para incentivar el uso responsable de las tarjetas de crédito.

3.2.1 Canales de Pago Preferidos

El gráfico Img 7 muestra el Número de Transacciones por Canal de Pago, destacando que el canal POS (puntos de venta) es el más utilizado con una amplia diferencia, superando los 250 millones de transacciones. Los canales Oficina e Internet siguen con un volumen considerable pero mucho menor, lo que indica que los usuarios prefieren realizar sus pagos de forma presencial en puntos de venta. Por otro lado, los canales de Banca telefónica y Banca celular tienen un uso casi inexistente, lo que sugiere que estas modalidades no son tan populares o no están tan ampliamente adoptadas por los clientes. En resumen, el canal POS domina las transacciones, mientras que otros métodos como Internet y Oficina tienen una participación menor, y la Banca celular y telefónica son marginales en su uso.

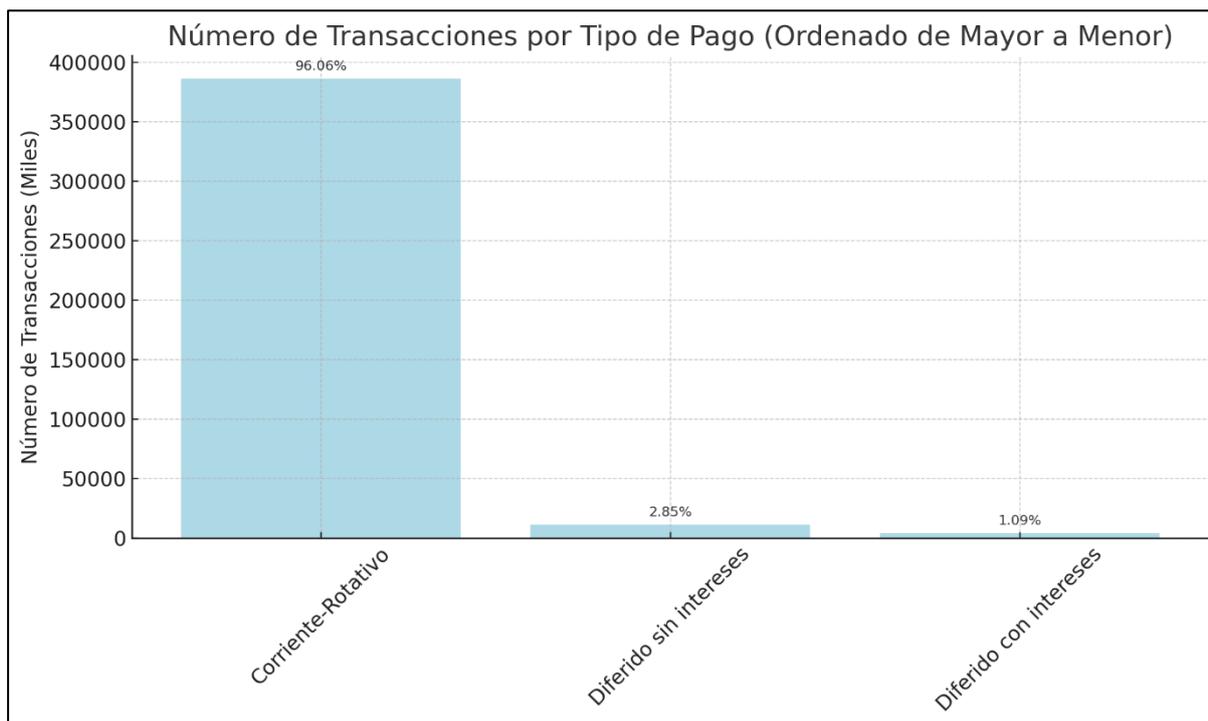


Img 7 "Gráfico de la Distribución de Numero de transacciones por canal de pago." Origen el autor.

La oportunidad aquí es clara: las empresas y los bancos deben invertir en la mejora continua de sus plataformas digitales, asegurando que la experiencia de usuario sea lo más fluida posible. El hecho de que la mayoría de los clientes prefiera estos canales de pago también pone de relieve la necesidad de integrar soluciones de pago más rápidas y seguras, como billeteras digitales, y opciones de pagos sin contacto.

3.3 Distribución por Tipo de Pago

El gráfico Img 8 muestra el Número de Transacciones por Tipo de Pago, destacando que el tipo de pago Corriente-Rotativo es el más utilizado de manera abrumadora, superando los 350 millones de transacciones. Los otros tipos de pago, como Diferido con intereses y Diferido sin intereses, tienen un uso muy marginal en comparación, con un volumen significativamente menor de transacciones. Esto sugiere que los usuarios prefieren realizar pagos con la modalidad Corriente-Rotativa, mientras que las opciones de pagos diferidos, ya sea con o sin intereses, no son tan populares o comunes.

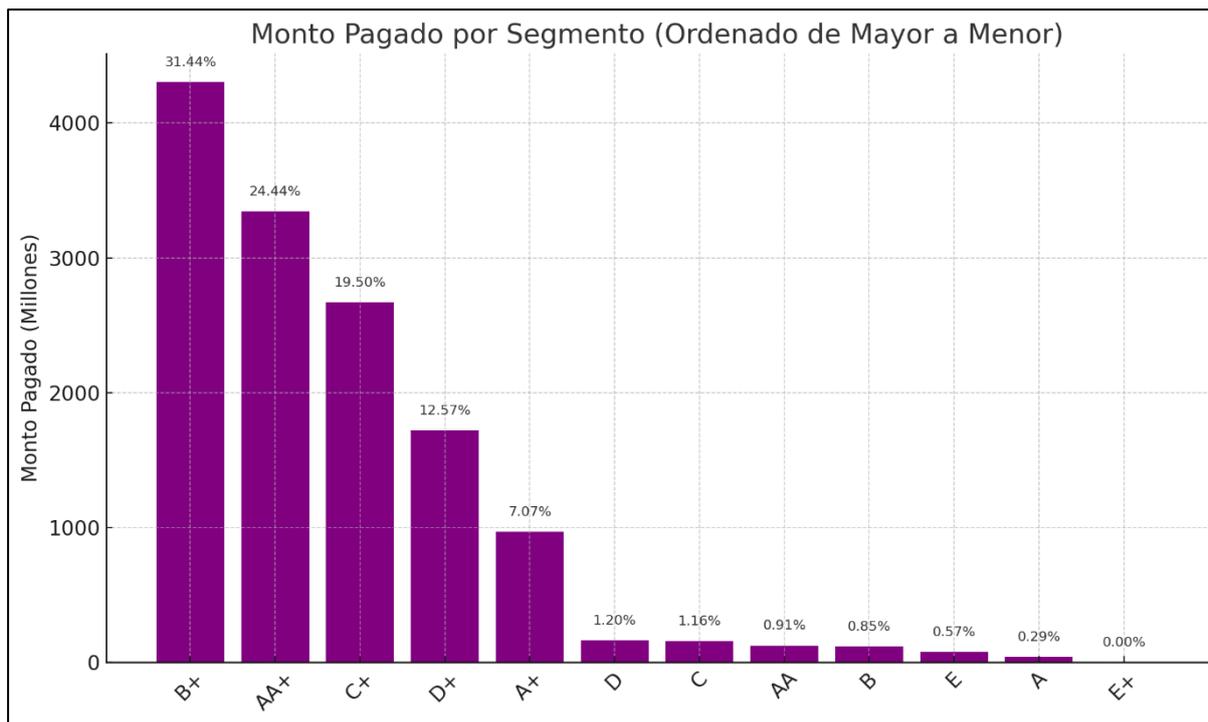


Img 8 "Gráfico de la Distribución de Numero de Transacciones por tipo de Pago." Origen el autor.

El Diferido sin intereses es también una opción atractiva para usuarios que prefieren distribuir sus compras sin costos adicionales, mientras que el Diferido con intereses sigue siendo popular entre quienes priorizan la flexibilidad de plazos más largos, aunque implique un mayor costo. Las instituciones pueden aprovechar este comportamiento al promover pagos sin intereses y ofrecer incentivos a quienes usen la modalidad rotativa, mientras gestionan adecuadamente el crédito con intereses para usuarios que buscan más flexibilidad.

3.4 Monto Pagado por Segmento

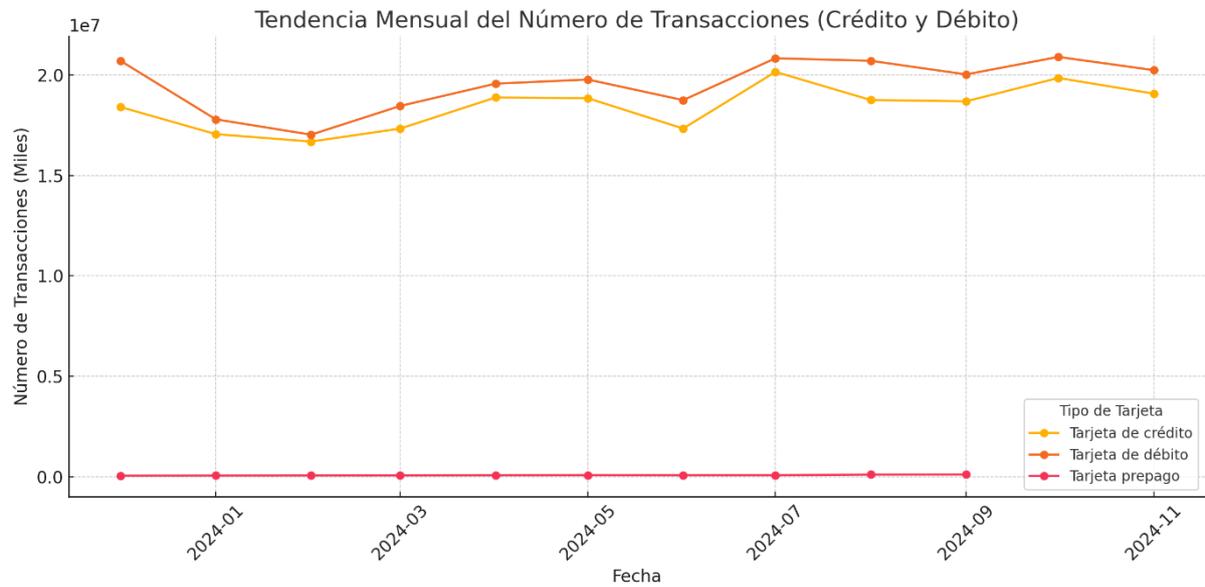
En el gráfico Img 9 se observa que los segmentos B+ y AA+ son los que tienen el mayor monto pagado, superando los 4 billones. Estos segmentos representan los clientes que realizan los pagos más altos, seguidos por los segmentos C y D+, que también tienen una participación importante, con montos pagados que superan los 2 billones. Los demás segmentos, como A, A+, B, D, y E, tienen un monto pagado considerablemente menor, lo que indica que la mayoría de las transacciones con montos altos provienen de los segmentos B+ y AA+.



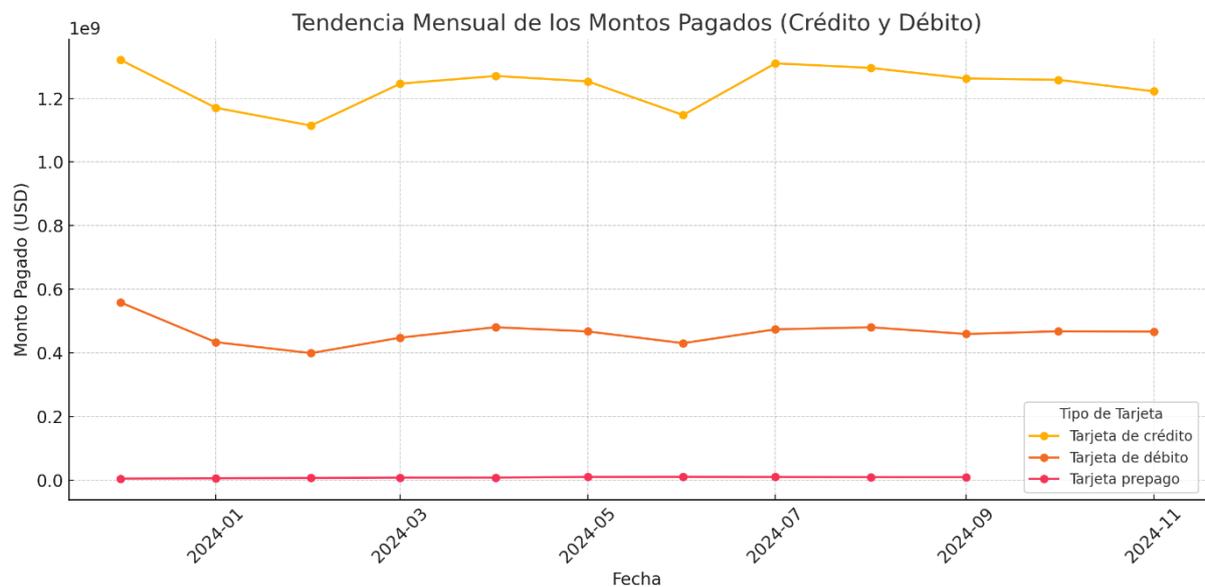
Img 9 "Gráfico de la Distribución de Monto Pagado por Segmento." Origen el autor.

Para maximizar el valor de estos clientes, sería relevante evaluar si ya se están implementando programas de fidelización que recompensen el comportamiento de compra premium con beneficios tangibles, como acceso prioritario a productos financieros, atención VIP o descuentos en bienes de lujo. En caso de que estas estrategias aún no estén en marcha o puedan optimizarse, su implementación o mejora podría fortalecer la relación con los clientes. Además, el uso de análisis predictivo permitiría anticipar sus necesidades y personalizar ofertas de manera más efectiva, brindando una mejor experiencia y fomentando el uso continuo de las tarjetas.

3.5 Tendencia en uso de tarjetas de crédito y débito.



Img 10 "Gráfico de la tendencia Mensual de número de transacciones (Crédito y Débito)." Origen el autor.



Img 11 "Gráfico de la Tendencia Mensual de montos pagados (Crédito y Débito)." Origen el autor.

En el gráfico Img 10 representa la tendencia mensual en el número de transacciones, diferenciando entre tarjetas de crédito y débito. Los datos indican que las

transacciones con tarjetas de débito mantienen un volumen más alto en comparación con las de tarjetas de crédito durante los meses analizados, lo que sugiere una preferencia o necesidad de los usuarios por utilizar sus fondos de manera directa en lugar de asumir créditos.

Observamos una variación estacional donde ciertos meses presentan un aumento significativo en el número de transacciones. Esto podría estar asociado a eventos económicos o culturales recurrentes, como temporadas de vacaciones o eventos comerciales. La estabilidad de las transacciones de débito sugiere un uso constante, mientras que las transacciones de crédito presentan picos más abruptos, posiblemente asociados a compras específicas o promociones.

En el gráfico lmg 11 muestra los montos mensuales pagados con tarjetas de crédito y débito en millones. Aquí, se observa que las tarjetas de crédito manejan montos considerablemente más altos en comparación con las de débito, lo cual es característico de este tipo de tarjeta, ya que permite realizar compras de mayor valor a crédito.

La curva de las tarjetas de crédito revela una mayor sensibilidad a variaciones mensuales, con incrementos marcados en ciertos períodos, que podrían estar relacionados con promociones o facilidades de pago en cuotas. En contraste, los montos pagados con tarjetas de débito tienden a ser menores y presentan una curva menos volátil, lo cual es consistente con su uso para compras cotidianas y de menor valor.

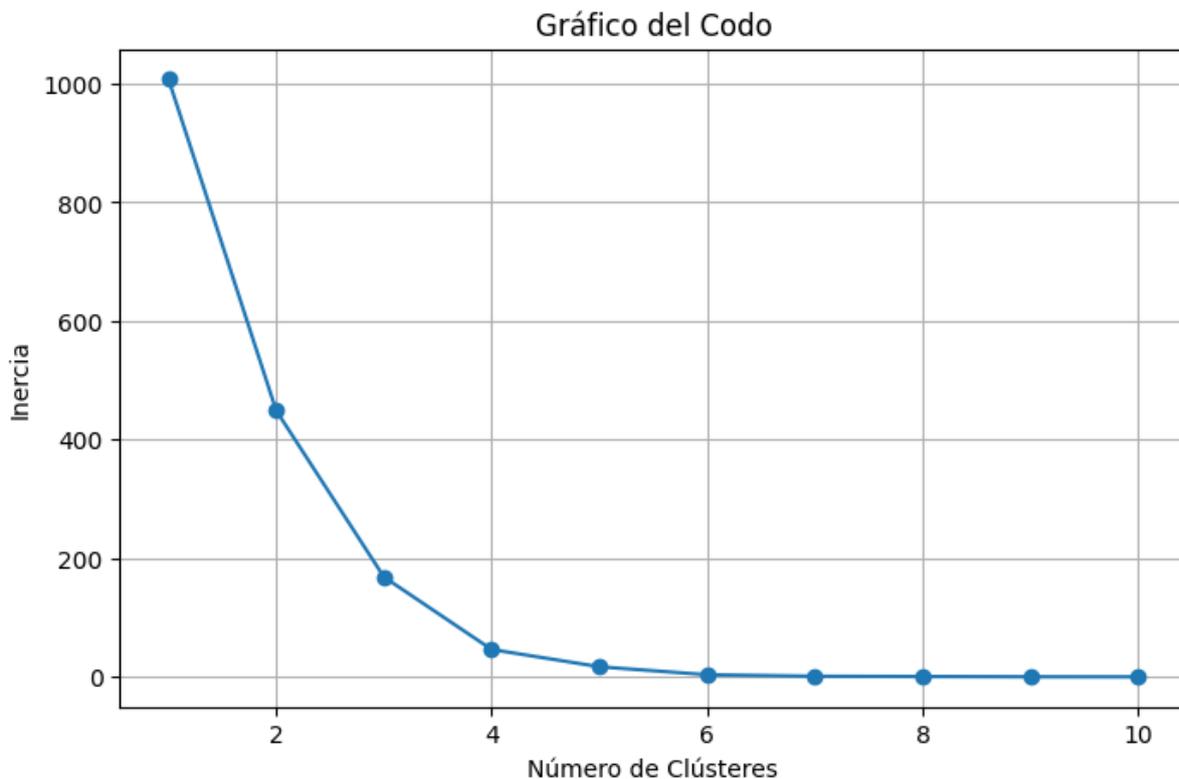
Podemos decir que en el análisis conjunto de ambos gráficos permite inferir que las tarjetas de crédito son preferidas para transacciones de mayor valor, mientras que las tarjetas de débito mantienen un uso regular para transacciones de menor importe. Esto refleja un comportamiento de consumo donde los usuarios administran sus gastos cotidianos mediante débito, reservando el crédito para compras más importantes o planificadas.

3.6 Resultados del Clústering

Tras el preprocesamiento de los datos provenientes de la Superintendencia de Bancos del Ecuador, donde únicamente se incluyeron los datos de transacciones con tarjetas de crédito y débito correspondientes al mes de septiembre, se analizaron utilizando K-Means por su capacidad de agrupar grandes volúmenes de datos de manera eficiente; se describirá el análisis con K-Means, usando un valor de $k=3$ y $k=4$, y los resultados obtenidos en términos de agrupación.

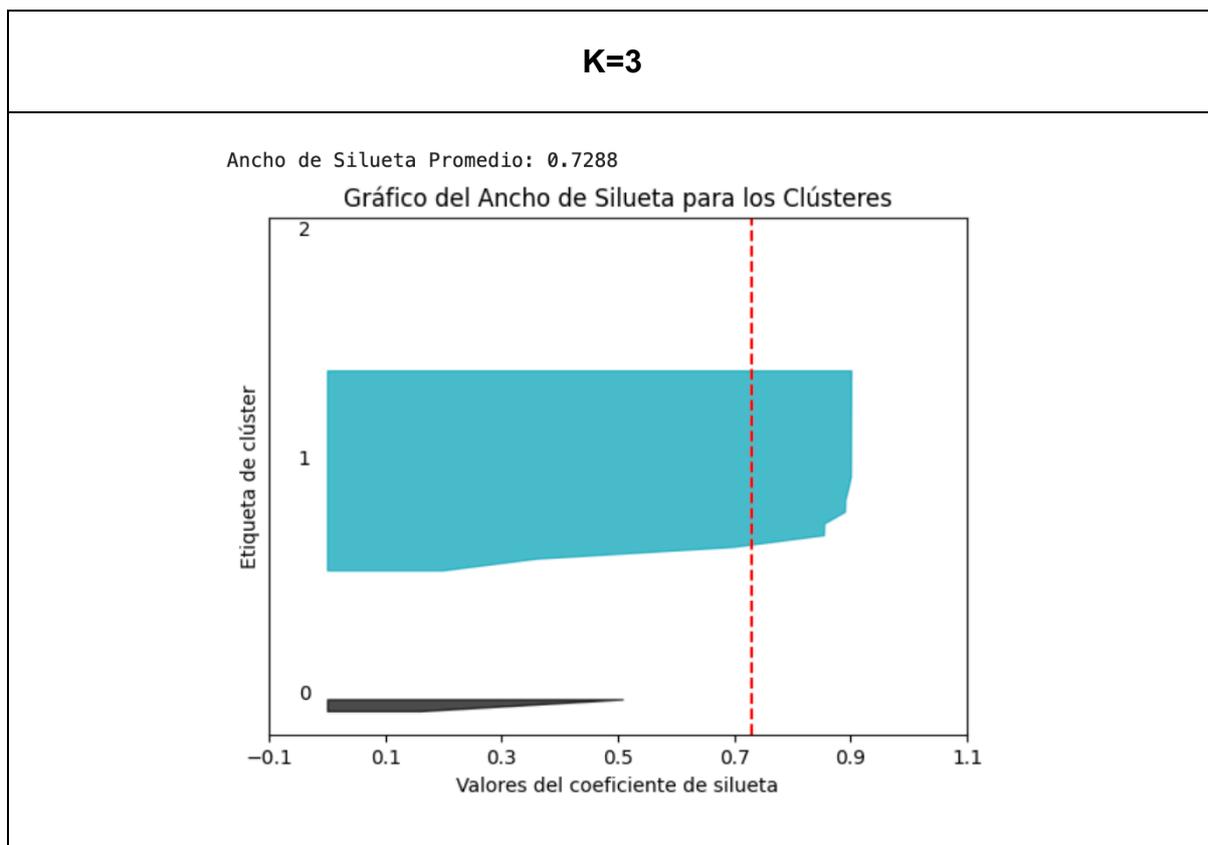
3.7 Clústering con K-Means

El algoritmo K-Means se implementó con $k=3$ y $k=4$, lo que significa que se formaron tres y cuatro grupos principales de instituciones financieras basados en patrones de uso de tarjetas de crédito y débito respectivamente. Este número de clústeres se determinó mediante el Gráfico del Codo y una evaluación práctica de la interpretación de los grupos. En este caso, el "codo" fue claramente visible entre $k=3$ y $k=4$ (véase *Img 12*), lo que indicaba que entre estos dos números de agrupaciones estaba la mejor calidad del agrupamiento.



Img 12 "Gráfico del Codo." Origen el autor.

El coeficiente de silueta se utiliza para evaluar la calidad de los clústeres en términos de cohesión (la cercanía entre los puntos dentro de un mismo clúster) y separación (la distancia entre clústeres diferentes). Valores más altos del coeficiente indican que los puntos están mejor agrupados en su propio clúster y separados de los otros.



Img 13 "Gráfico del ancho de silueta." Origen el autor.

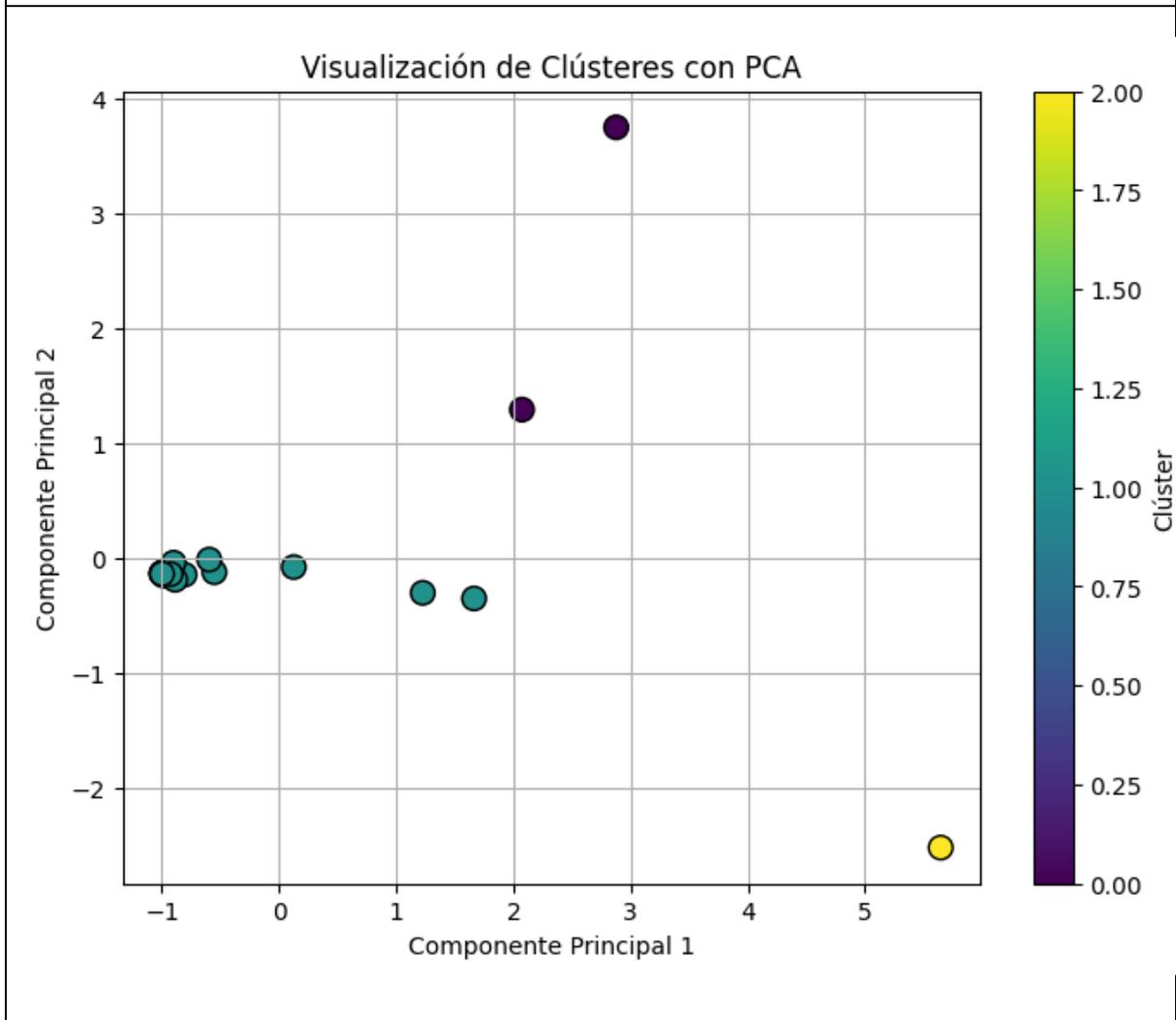
En el gráfico Img 13, La línea roja vertical indica que el promedio del coeficiente de silueta está alrededor de 0.75. Esto sugiere que, en general, los clústeres están moderadamente bien definidos, aunque hay margen para mejorar. El Clúster 0 tiene un coeficiente de silueta bajo, con algunos valores incluso negativos. Los puntos con valores negativos indican que estos elementos podrían estar más cerca de otro clúster que del suyo propio; esto sugiere que el clúster 0 no está bien cohesionado o que hay solapamiento con otro clúster.

El clúster 1 es dominante en términos de número de puntos. Tiene una distribución amplia del coeficiente de silueta, con valores que van desde cerca de 0 hasta 0.7; esto indica que hay cierta heterogeneidad dentro del clúster, con algunos puntos

menos bien cohesionados. El clúster restante tiene coeficientes de silueta consistentemente altos (cerca de 1). Esto sugiere que está bien definido y separado de los otros clústeres, su tamaño reducido podría reflejar un grupo de datos compacto y bien diferenciado.

En el gráfico Img 14 de dispersión, con $k=3$, se observa que el clúster 0 (en color verde) contiene la mayoría de las muestras, que están agrupadas muy cerca del origen. Esto indica que estas instituciones financieras tienen características bastante similares. Los clústeres 1 (en color morado) y 2 (en color amarillo) están más dispersos y se encuentran alejados del clúster 0, sugiriendo que representan grupos con características bastante distintas y menos comunes dentro del conjunto de datos (véase Tabla 1).

K=3



Img 14 "Gráfico del ACP." Origen el autor.

EMISOR	CLÚSTER
BanEcuador B.P	1
Banco Amazonas S.A	1
Banco Amibank S.A	1
Banco Bolivariano C.A	1
Banco Comercial de Manabi S.A	1
Banco D-miro S.A	1
Banco Desarrollo de los Pueblos S.A Codesarrollo	1
Banco Diners Club del Ecuador S.A	0
Banco General Ruminahui S.A	1
Banco Guayaquil S.A	1
Banco Internacional S.A	1
Banco Pichincha C.A	2
Banco Procredit S.A	1
Banco Solidario S.A	1
Banco Visionfund Ecuador S.A	1
Banco de Loja S.A	1
Banco de Machala S.A	1
Banco de la Produccion S.A Produbanco	1
Banco del Austro S.A	1
Banco del Litoral S.A	1
Banco del Pacifico S.A	0

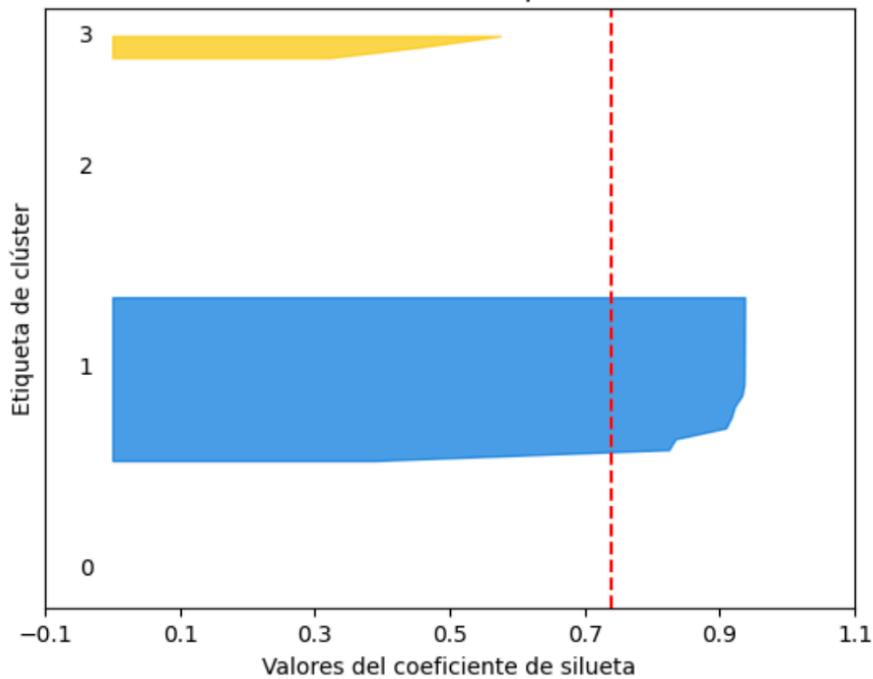
Tabla 1. Clústering con K-means=3

Similar al gráfico lmg 13 de ancho de silueta con $k=3$, en $k=4$ el promedio está cerca de 0.75. Esto indica que la calidad general del clústering sigue siendo moderada.

K=4

Ancho de Silueta Promedio: 0.7386

Gráfico del Ancho de Silueta para los Clústeres



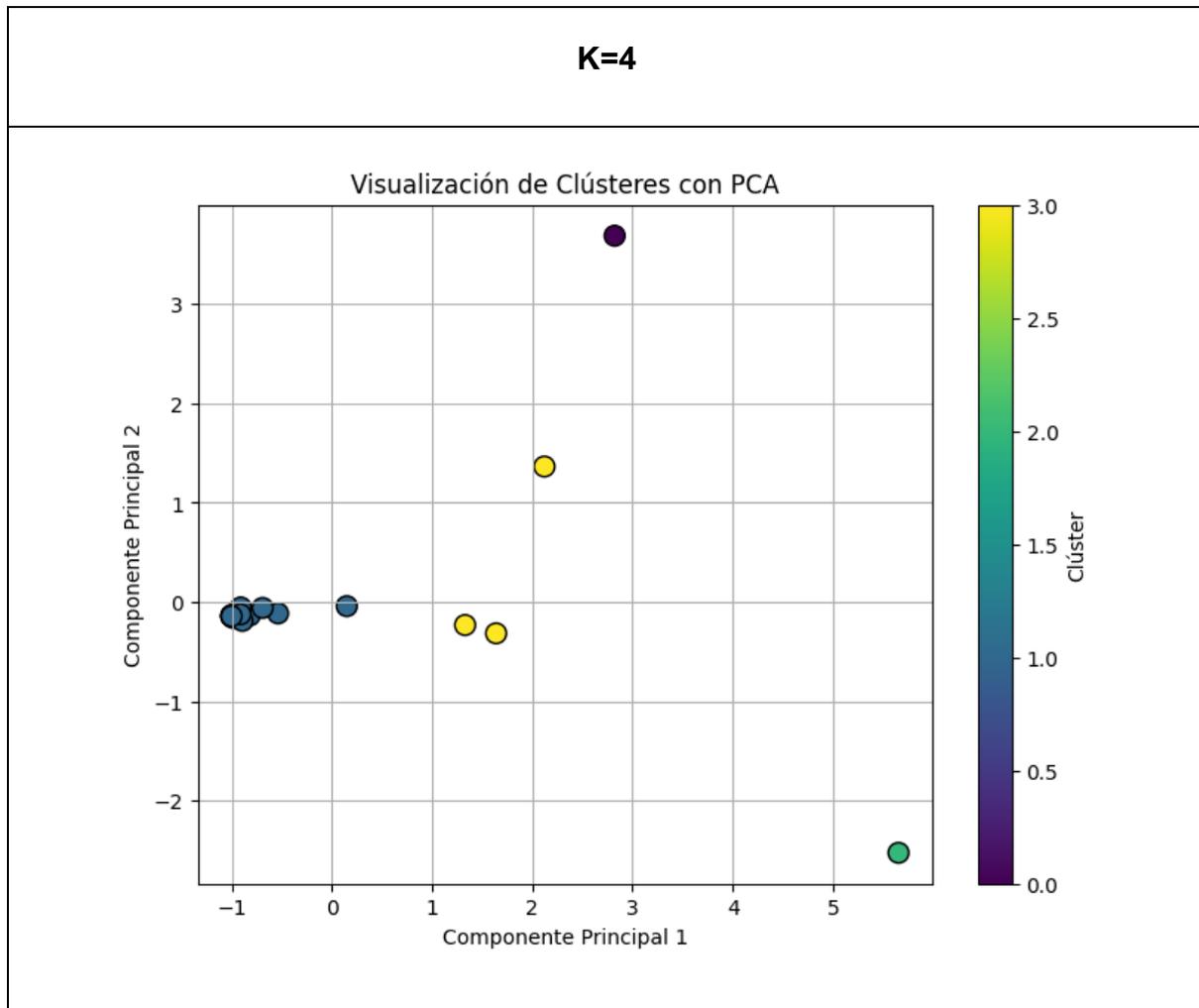
Img 15 "Gráfico del Ancho de Silueta." Origen el autor.

En el gráfico Img 15, el clúster 0 sigue mostrando coeficientes de silueta bajos, con algunos valores negativos. Esto confirma que persisten los problemas de cohesión y asignación de puntos; la forma del clúster en el gráfico parece ser ligeramente más uniforme que en el primer gráfico, lo que podría reflejar una mejora marginal en su cohesión.

En el clúster 1 nuevamente domina en tamaño y tiene una distribución amplia de coeficientes de silueta; sin embargo, en este gráfico parece mostrar una mayor proporción de puntos con valores más altos de coeficiente de silueta (más cercanos a 0.75 o superiores), lo que podría reflejar una ligera mejora en su cohesión.

Al igual que en el primer gráfico, el clúster 2 tiene coeficientes consistentemente altos, lo que confirma que está bien definido y separado. En este gráfico, la forma del clúster es más uniforme que en el primer gráfico, lo que indica una mejora en su estructura interna. Aunque el algoritmo generó 4 clústeres, el clúster 3 tiene una población muy

baja (incluso un solo punto) y los puntos asignados a este clúster no representen un grupo real en los datos.



Img 16 "Gráfico del ACP." Origen el autor.

En el gráfico Img 16, con $k=4$, la introducción de un clúster adicional (clúster 3 en color morado) permite una diferenciación más detallada. Este nuevo clúster está posicionado de manera aislada, lo que indica que las instituciones en este grupo presentan características claramente diferentes a las de los otros clústeres. El clúster 0 sigue siendo el grupo dominante y está concentrado cerca del origen, mientras que los clústeres 1, 2 y 3 están más dispersos, resaltando las diferencias entre sus miembros (véase tabla 2).

EMISOR	CLÚSTER
BanEcuador B.P	1
Banco Amazonas S.A	1
Banco Amibank S.A	1
Banco Bolivariano C.A	1
Banco Comercial de Manabi S.A	1
Banco D-miro S.A	1
Banco Desarrollo de los Pueblos S.A Codesarrollo	1
Banco Diners Club del Ecuador S.A	0
Banco General Ruminahui S.A	1
Banco Guayaquil S.A	3
Banco Internacional S.A	1
Banco Pichincha C.A	2
Banco Procredit S.A	1
Banco Solidario S.A	1
Banco Visionfund Ecuador S.A	1
Banco de Loja S.A	1
Banco de Machala S.A	1
Banco de la Produccion S.A Produbanco	3
Banco del Austro S.A	1
Banco del Litoral S.A	1
Banco del Pacifico S.A	3

Tabla 2. Clústering con K-means=4

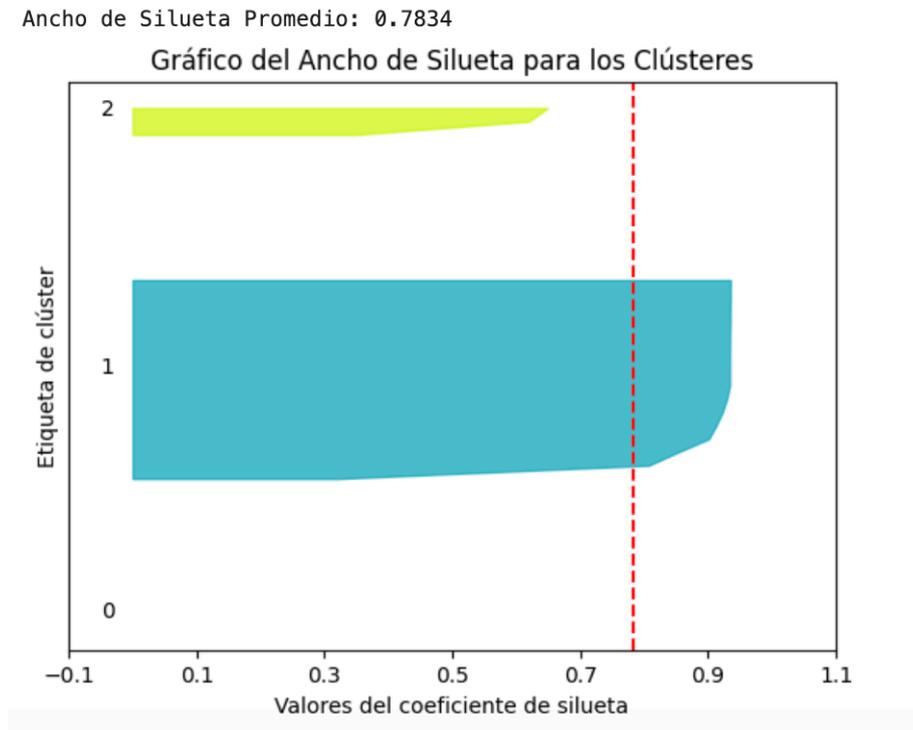
Comparando los gráficos de dispersión de ambos k (Img 14 vs. Img 16), el modelo con k=4 proporciona una segmentación más refinada, identificando un grupo adicional que parece tener características únicas. Esto sugiere que el modelo con k=4 puede capturar mejor las diferencias sutiles entre las instituciones financieras, especialmente si se busca entender comportamientos atípicos o nichos específicos. En cambio, el modelo con k=3 es más adecuado si el objetivo es una segmentación general sin tanto nivel de detalle.

3.8 Clústering con K-Means sin Banco Pichincha

Para contraponer el análisis previo, se decidió realizar el mismo proceso de análisis sin el Banco Pichincha que es el emisor de tarjetas con mayor número de transacciones y monto en el sistema financiero del Ecuador ya que al estar solo en un clúster se lo identifica como una anomalía o punto atípico.

- K=3

El gráfico Img 17 de ancho de silueta para k=3 presenta un promedio del coeficiente de silueta de 0.7919, lo que refleja que los clústeres generados están bien definidos, con una cohesión interna adecuada y una separación razonable entre ellos. Este valor, aunque no alcanza el ideal de 1, indica que la calidad del agrupamiento es aceptable para los propósitos del análisis.



Img 17 "Gráfico del Ancho de Silueta." Origen el autor.

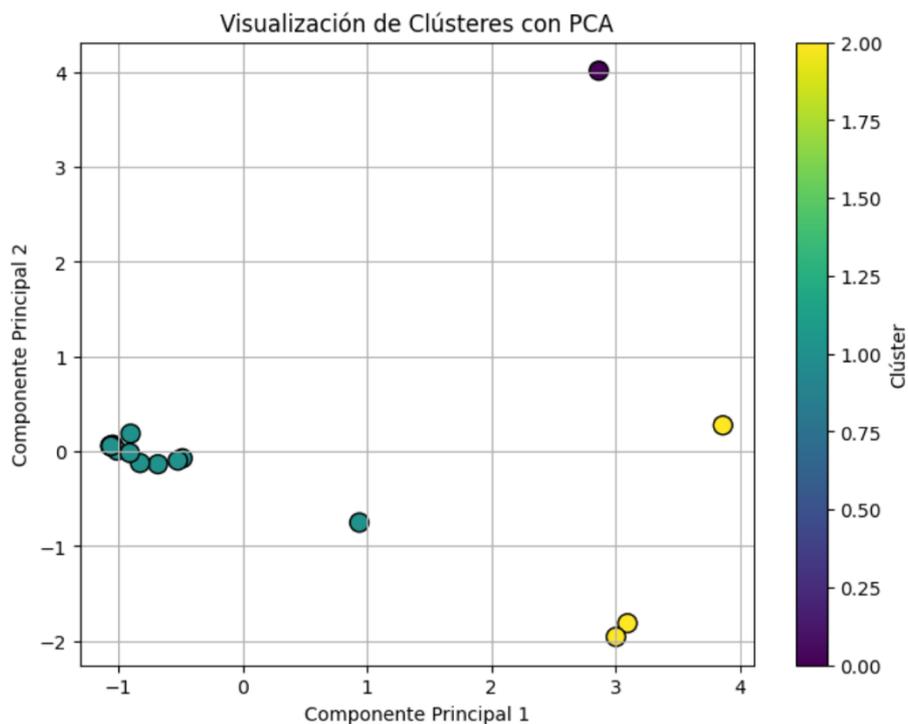
En cuanto a la estructura de los clústeres, se observa también en el gráfico Img 17 que el clúster 1, representado en color azul, es el grupo más grande y dominante dentro del modelo. Este clúster muestra una distribución uniforme del ancho de silueta con valores cercanos al promedio global, sin evidenciar valores negativos. Esto sugiere que los puntos dentro del clúster están correctamente asignados y no presentan solapamientos significativos con otros grupos.

Por otro lado, el clúster 2, representado en color amarillo, se caracteriza por ser más pequeño y compacto. Los valores de silueta de este clúster son altos, cercanos a 1, lo que indica que está excelentemente definido y separado del resto de los clústeres.

Además, la forma alargada y uniforme de este grupo sugiere que los datos agrupados comparten características muy similares.

La línea roja vertical en el gráfico indica el promedio del coeficiente de silueta (0.7919), y la mayoría de los puntos en ambos clústeres se encuentran por encima de este valor. Esto refuerza la percepción de que el modelo presenta una calidad general adecuada, con clústeres bien cohesionados y separados.

El gráfico de visualización de clústeres mediante Análisis de Componentes Principales (Img 18) permite observar la distribución de los datos en un espacio bidimensional, donde las componentes principales representan las dimensiones con mayor varianza explicada en el conjunto de datos. En este caso, se han identificado tres clústeres, representados por diferentes colores según la escala del lado derecho del gráfico.



Img 18 "Gráfico del ACP." Origen el autor.

El clúster 0, predominante en el gráfico, se concentra cerca del origen de los ejes principales, lo que indica que los datos asociados a este grupo comparten características similares y están estrechamente relacionados. Esta densidad refleja

que las instituciones financieras en este clúster tienen comportamientos homogéneos en términos de las variables analizadas.

El clúster 1 se posiciona en el cuadrante inferior derecho del gráfico, con una separación clara del clúster 0. Este grupo tiene una menor densidad y está formado por puntos más dispersos, lo que podría reflejar características distintivas de las instituciones financieras agrupadas en este clúster. La distancia respecto al origen y a otros clústeres indica que este grupo representa comportamientos más diferenciados dentro del dataset.

Por último, el clúster 2 se encuentra en la parte superior del gráfico, claramente separado de los otros dos clústeres. La posición aislada de este grupo sugiere que las instituciones financieras en este clúster poseen características atípicas o particulares que no se superponen con las de los demás grupos. La baja densidad de este clúster puede ser indicativa de un tamaño reducido del grupo o de la presencia de valores atípicos (véase Tabla 3).

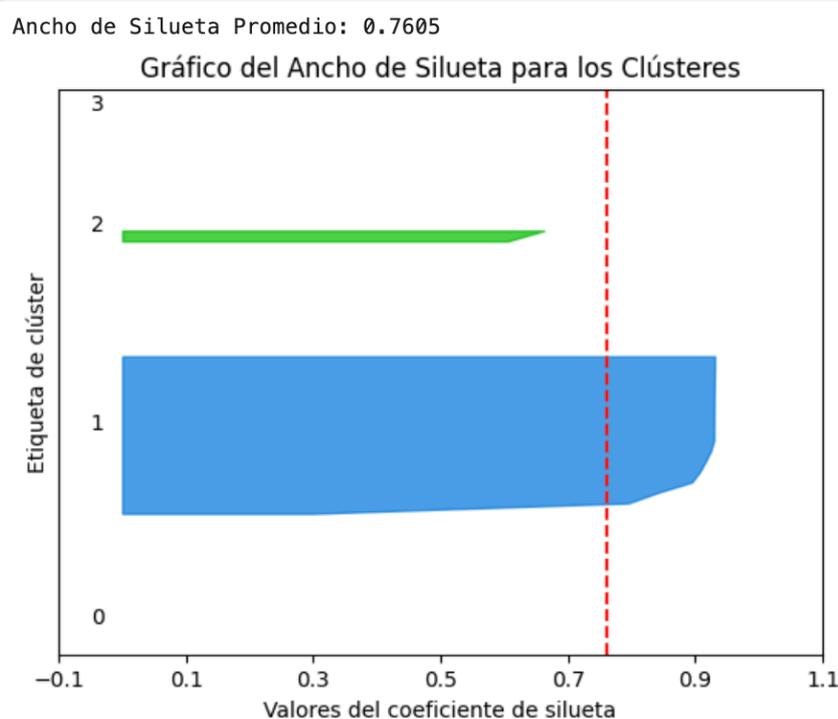
Emisor	CLÚSTER
BanEcuador B.P	1
Banco Amazonas S.A	1
Banco Amibank S.A	1
Banco Bolivariano C.A	1
Banco Comercial de Manabi S.A	1
Banco D-miro S.A	1
Banco Desarrollo de los Pueblos S.A Codesarrollo	1
Banco Diners Club del Ecuador S.A	0
Banco General Ruminahui S.A	1
Banco Guayaquil S.A	2
Banco Internacional S.A	1
Banco Pichincha C.A	x
Banco Procredit S.A	1
Banco Solidario S.A	1
Banco Visionfund Ecuador S.A	1
Banco de Loja S.A	1
Banco de Machala S.A	1
Banco de la Produccion S.A Produbanco	2
Banco del Austro S.A	1
Banco del Litoral S.A	1
Banco del Pacifico S.A	2

Tabla 3. Clústering con K-means=3 sin dato atípico

En general, la proyección mediante PCA destaca la clara separación entre los clústeres, lo que evidencia una adecuada capacidad del modelo para diferenciar las características de las instituciones financieras. Este análisis proporciona una representación visual del comportamiento de los clústeres y ayuda a validar la calidad del agrupamiento generado.

- K=4

El gráfico Img 19 de ancho de silueta muestra un promedio del coeficiente de silueta de 0.7605, lo que indica una calidad moderada en la agrupación. Este valor sugiere que los clústeres están razonablemente bien definidos, aunque existe margen para mejorar la cohesión interna y la separación entre ellos.



Img 19 "Gráfico del Ancho de Silueta." Origen el autor.

En el análisis de los clústeres, el clúster 1, representado en color azul, es el grupo más grande y dominante. Su ancho de silueta muestra una distribución amplia, con valores que oscilan entre negativos y cercanos al promedio global. La presencia de

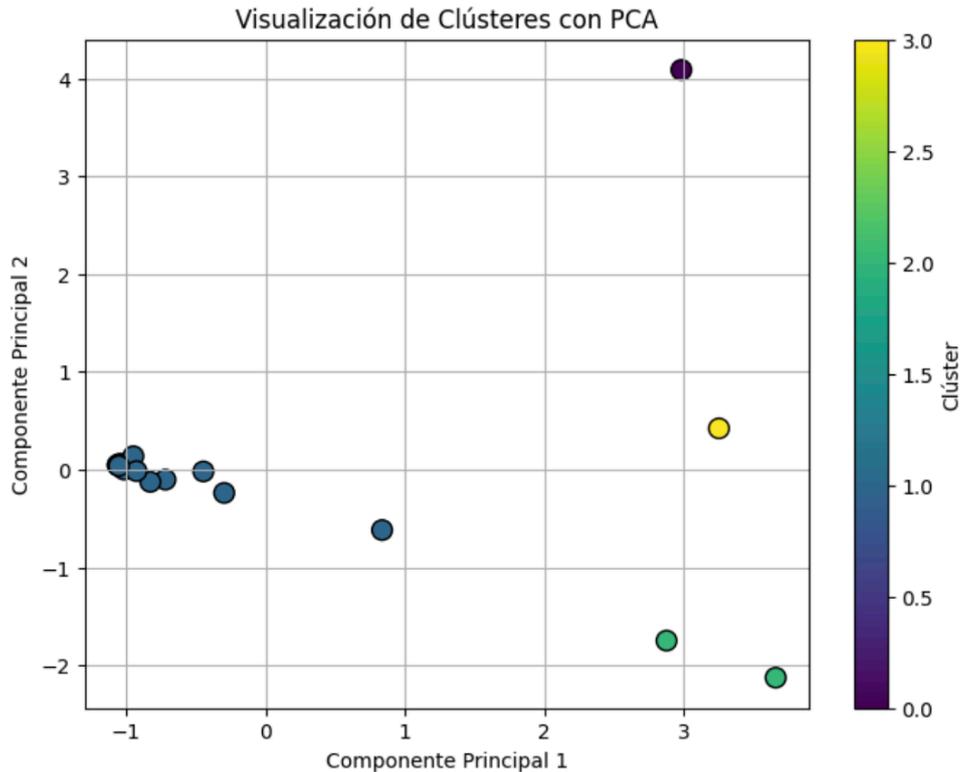
valores negativos indica que algunos puntos en este clúster podrían estar más cerca de otros clústeres que del propio, lo que refleja problemas de cohesión interna o solapamiento con otros grupos.

El clúster 2, en color verde, es más pequeño y presenta valores consistentemente altos de silueta, cercanos a 1. Esto sugiere que este clúster está bien definido y separado de los demás. La forma alargada y uniforme en el gráfico refuerza la idea de que los puntos dentro de este clúster comparten características muy similares, lo que contribuye a su clara definición.

La línea roja vertical marca el promedio del coeficiente de silueta (0.7605). Si bien la mayoría de los puntos del clúster 2 están por encima de este valor, en el caso del clúster 1 hay puntos por debajo del promedio, lo que indica cierta heterogeneidad y posibles desafíos en la calidad del agrupamiento en este grupo.

En general, el gráfico refleja una segmentación razonablemente adecuada, donde el clúster 2 está claramente definido y separado, mientras que el clúster 1 presenta áreas de mejora en términos de cohesión y asignación de puntos. Esto sugiere que, aunque el modelo captura patrones importantes, podrían implementarse ajustes adicionales para mejorar la calidad global del agrupamiento.

El gráfico de visualización de clústeres utilizando Análisis de Componentes Principales (Img 20) representa la proyección de los datos en un espacio bidimensional, con el objetivo de observar la distribución y separación de los clústeres generados. En este caso, se identifican cuatro clústeres, diferenciados por colores según la escala mostrada a la derecha del gráfico.



Img 20 "Gráfico del ACP." Origen el autor.

El clúster 0, el más denso y predominante, se concentra cerca del origen de los ejes principales. Esto indica que las instituciones financieras agrupadas en este clúster comparten características similares y presentan comportamientos homogéneos en términos de las variables analizadas. La proximidad entre los puntos de este clúster sugiere una buena cohesión interna.

El clúster 1, localizado en la parte inferior derecha del gráfico, se encuentra claramente separado del clúster 0. Este grupo tiene una menor densidad y sus puntos están distribuidos de manera más dispersa, lo que podría indicar características distintivas o variabilidad dentro de este clúster. Su posición lejana al origen sugiere que las instituciones en este grupo difieren significativamente de las pertenecientes al clúster 0.

El clúster 2, ubicado en la parte superior del gráfico, está aislado del resto de los clústeres. Su posición y separación clara indican que las instituciones en este grupo poseen características únicas y atípicas respecto a los otros clústeres. La baja

densidad y distribución limitada de puntos pueden sugerir que se trata de un grupo pequeño o incluso de valores atípicos.

Por último, el clúster 3 se encuentra en la parte inferior central del gráfico, con una separación evidente respecto a los demás clústeres. Su distribución sugiere que las instituciones agrupadas en este clúster tienen comportamientos diferenciados, aunque la dispersión limitada de puntos podría indicar una cohesión aceptable dentro del grupo (véase Tabla 4).

EMISOR	CLÚSTER
BanEcuador B.P	1
Banco Amazonas S.A	1
Banco Amibank S.A	1
Banco Bolivariano C.A	1
Banco Comercial de Manabi S.A	1
Banco D-miro S.A	1
Banco Desarrollo de los Pueblos S.A Codesarrollo	1
Banco Diners Club del Ecuador S.A	0
Banco General Ruminahui S.A	1
Banco Guayaquil S.A	2
Banco Internacional S.A	1
Banco Pichincha C.A	x
Banco Procredit S.A	1
Banco Solidario S.A	1
Banco Visionfund Ecuador S.A	1
Banco de Loja S.A	1
Banco de Machala S.A	1
Banco de la Produccion S.A Produbanco	2
Banco del Austro S.A	1
Banco del Litoral S.A	1
Banco del Pacifico S.A	3

Tabla 4. Clústering con K-means=4 sin dato atípico

En general, la visualización mediante PCA muestra una clara separación entre los clústeres, lo que respalda la capacidad del modelo para distinguir patrones y agrupar adecuadamente a las instituciones financieras según sus características. Este análisis

resalta la utilidad del modelo para identificar tanto grupos principales como comportamientos atípicos dentro del conjunto de datos.

3.9 Evaluación de la Calidad del Clustering

Comparando ambos modelos, el valor promedio del coeficiente de silueta es similar en ambos casos, lo que indica una cohesión moderada en los dos modelos. Sin embargo, con $k=4$ se observa un clúster pequeño (clúster 3) que podría representar un grupo de instituciones con características particulares y claramente diferenciadas del resto. Esto no ocurre en el modelo con $k=3$, donde los clústeres son más homogéneos. Aunque ambos modelos presentan una cohesión moderada, el modelo con $k=3$ muestra clústeres más equilibrados en tamaño, mientras que $k=4$ ofrece la posibilidad de identificar un grupo específico de características diferenciadas.

	K = 3	K=4
Ancho de silueta	Sin BP: 0.7834 Con BP: 0.7288	Sin BP: 0.7605 Con BP: 0.7386
Número de Clústeres	Tres clústeres: uno grande y dos pequeños.	Cuatro clústeres: uno grande y tres pequeños.
Distribución de Muestras	Clúster "0" concentra la mayor parte de los puntos, mientras que "1" y "2" tienen menos muestras.	Clúster "0" sigue siendo el más grande, mientras que "1", "2" y "3" tienen una distribución pequeña.
Cohesión	Clúster "0" tiene heterogeneidad interna; "2" está bien cohesionado; "1" muestra problemas leves.	Clúster "0" sigue teniendo mayor heterogeneidad; "3" muestra mayor cohesión que los demás pequeños.
Separación de Clústeres	Clúster "2" está bien separado; "1" y "0" muestran cierta proximidad, pero manejable.	"1" y "2" están bien separados; "3" tiene una proximidad mayor a "0", lo que podría generar solapamientos.
Identificación de Comportamientos	Permite identificar dos comportamientos minoritarios y uno general.	Permite mayor detalle en la identificación, dividiendo patrones del clúster grande en subconjuntos.
Utilidad	Adecuado para un análisis general de los datos con divisiones básicas.	Útil para una segmentación más detallada y el análisis de subgrupos en el clúster dominante.
Ventaja Principal	Simplicidad en la interpretación y manejo de los clústeres.	Mejor identificación de subcomportamientos en los datos y mayor resolución analítica.
Desventaja Principal	Los clústeres grandes tienen poca cohesión interna.	Posible solapamiento entre clústeres pequeños ("3" y "0").

Tabla 5. Comparativa de valores de k-means.

Consideramos que el modelo de agrupamiento con $k=4$ incluyendo Banco Pichincha es la opción más adecuada. Al analizar los resultados, se observa que en el clúster 0 se encuentra Diners, que opera exclusivamente con crédito, mientras que en el clúster 3 se ubica Banco Pichincha, el cual presenta el mayor número de transacciones y volumen en USD. Esta segmentación parece estar mejor definida y más alineada con las características del sistema financiero que creemos es muy estable en el Ecuador.

Además, en un segundo grupo quedarían instituciones relevantes como Banco Guayaquil, Produbanco y Banco del Pacífico, dejando en el clúster 1 a las demás entidades financieras. Por estos puntos creemos que la segmentación con $k=4$ refleja mejor la realidad del sistema financiero ecuatoriano (véase Tabla 5).

Entidad Financiera	Oct	Nov	Dic	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep
BanEcuador B.P	1	1	1	1	1	1	1	1	1	1	1	1
Banco Amazonas S.A	1	1	1	1	1	1	1	1	1	1	1	1
Banco Amibank S.A	1	1	1	1	1	1	1	1	1	1	1	1
Banco Bolivariano C.A	1	1	1	1	1	1	1	1	1	1	1	1
Banco Comercial de Manabi S.A	1	1	1	1	1	1	1	1	1	1	1	1
Banco D-miro S.A	1	1	1	1	1	1	1	1	1	1	1	1
Banco Desarrollo de los Pueblos S.A Codesarrollo	1	1	1	1	1	1	1	1	1	1	1	1
Banco Diners Club del Ecuador S.A	0	0	0	0	0	0	0	0	0	0	0	0
Banco General Ruminahui S.A	1	1	1	1	1	1	1	1	1	1	1	1
Banco Guayaquil S.A	3	3	3	3	3	3	3	3	3	3	3	3
Banco Internacional S.A	1	1	1	1	1	1	1	1	1	1	1	1
Banco Pichincha C.A	2	2	2	2	2	2	2	2	2	2	2	2
Banco Procredit S.A	1	1	1	1	1	1	1	1	1	1	1	1
Banco Solidario S.A	1	1	1	1	1	1	1	1	1	1	1	1
Banco Visionfund Ecuador S.A	1	1	1	1	1	1	1	1	1	1	1	1
Banco de Loja S.A	1	1	1	1	1	1	1	1	1	1	1	1
Banco de Machala S.A	1	1	1	1	1	1	1	1	1	1	1	1
Banco de la Produccion S.A Produbanco	3	3	3	3	3	3	3	3	3	3	3	3
Banco del Austro S.A	1	1	1	1	1	1	1	1	1	1	1	1
Banco del Litoral S.A	1	1	1	1	1	1	1	1	1	1	1	1
Banco del Pacifico S.A	3	3	3	3	0	3	3	3	3	3	3	3
Clúster 0	1	1	1	1	2	1	1	1	1	1	1	1
Clúster 1	16	16	16	16	16	16	16	16	16	16	16	16
Clúster 2	1	1	1	1	1	1	1	1	1	1	1	1
Clúster 3	3	3	3	3	2	3	3	3	3	3	3	3

Tabla 5. Cuadro de clasificación de bancos según su clúster con $k\text{-means}=4$

4 Capítulo 4

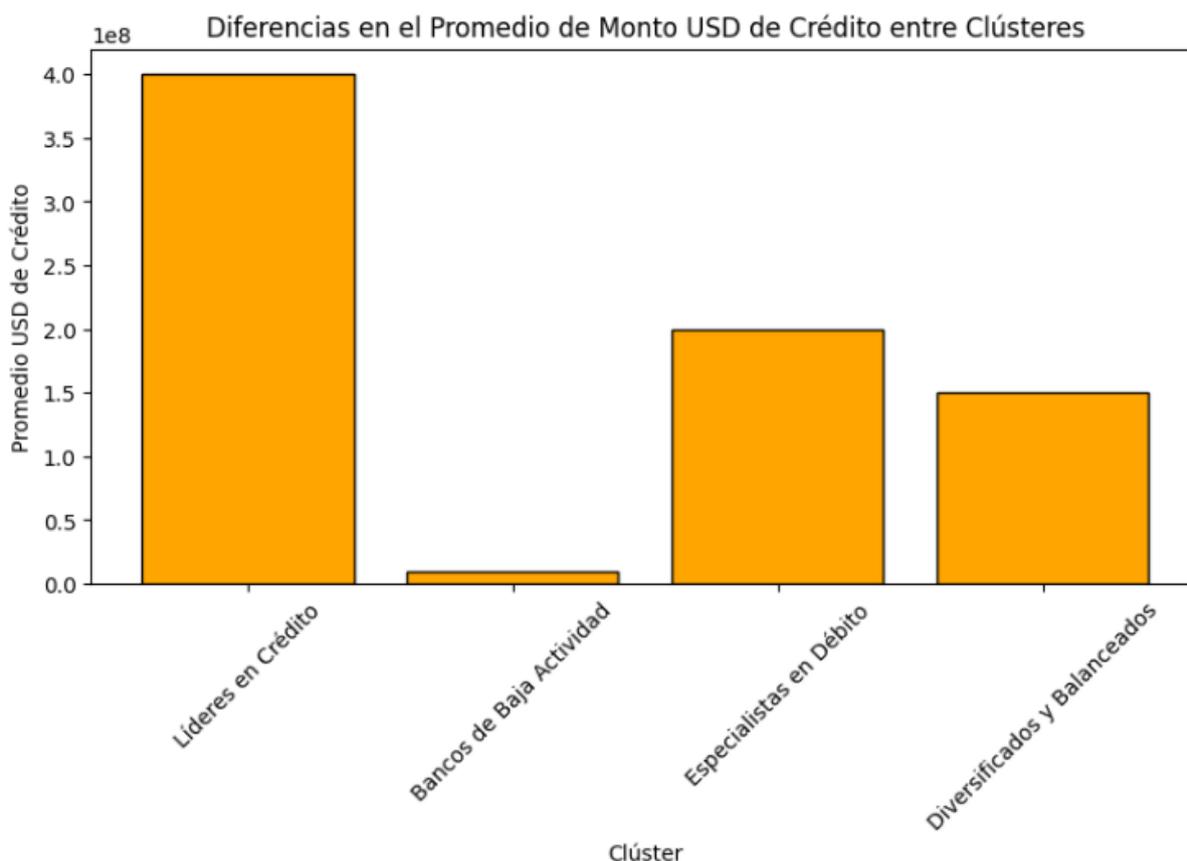
En este capítulo, realizamos la interpretación detallada de los resultados obtenidos a partir del análisis de agrupamiento mediante el algoritmo K-Means con $k=4$. Este modelo permitió identificar cuatro clústeres de instituciones financieras en Ecuador, los cuales presentan características distintivas en cuanto al uso de tarjetas de crédito y débito. La segmentación propuesta no solo proporciona una visión comprensible de los patrones transaccionales, sino que también facilita la formulación de estrategias orientadas a optimizar los servicios financieros.

Dada la información ya compartida, se presentan los resultados del análisis de segmentación realizado en el marco del proyecto. Este análisis tuvo como objetivo identificar patrones en el comportamiento transaccional de los bancos, permitiendo agruparlos en clústeres con características específicas. Cada clúster refleja diferencias clave en términos de transacciones, montos y proporciones de uso entre crédito y débito, brindando una perspectiva detallada sobre las dinámicas y estrategias de cada grupo (véase Tabla 6).

	Clúster_0 - Especialistas en Débito	Clúster_1 - Bancos de Baja Actividad	Clúster_2 - Líderes en Crédito	Clúster_3 - Diversificados y Balanceados
num_bancos	1.0	16.0	1.0	3.0
promedio_trx_credito	6146808.0	143803.5	3487232.0	2379740.67
std_trx_credito		270757.88		1140615.52
min_trx_credito	6146808.0	0.0	3487232.0	1581459.0
max_trx_credito	6146808.0	1020038.0	3487232.0	3686110.0
sum_trx_credito	6146808.0	2300856.0	3487232.0	7139222.0
promedio_usd_credito	404842306.15	8347755.42	214936749.48	156395529.77
std_usd_credito		15257084.52		61915648.15
min_usd_credito	404842306.15	0.0	214936749.48	110501847.66
max_usd_credito	404842306.15	56329732.52	214936749.48	226817363.64
sum_usd_credito	404842306.15	133564086.79	214936749.48	469186589.32
promedio_trx_debito	1277.0	151818.69	10191008.0	2540818.67
std_trx_debito		309007.82		728733.92
min_trx_debito	1277.0	188.0	10191008.0	1747151.0
max_trx_debito	1277.0	1214850.0	10191008.0	3179772.0
sum_trx_debito	1277.0	2429099.0	10191008.0	7622456.0
promedio_usd_debito	26116.67	3990420.88	225625630.74	59248083.62
std_usd_debito		7090801.0	0.49	
min_usd_debito	26116.67	3423.64	225625630.74	43776750.67
max_usd_debito	26116.67	26669718.05	225625630.74	73790385.48
sum_usd_debito	26116.67	63846734.08	225625630.74	177744250.87

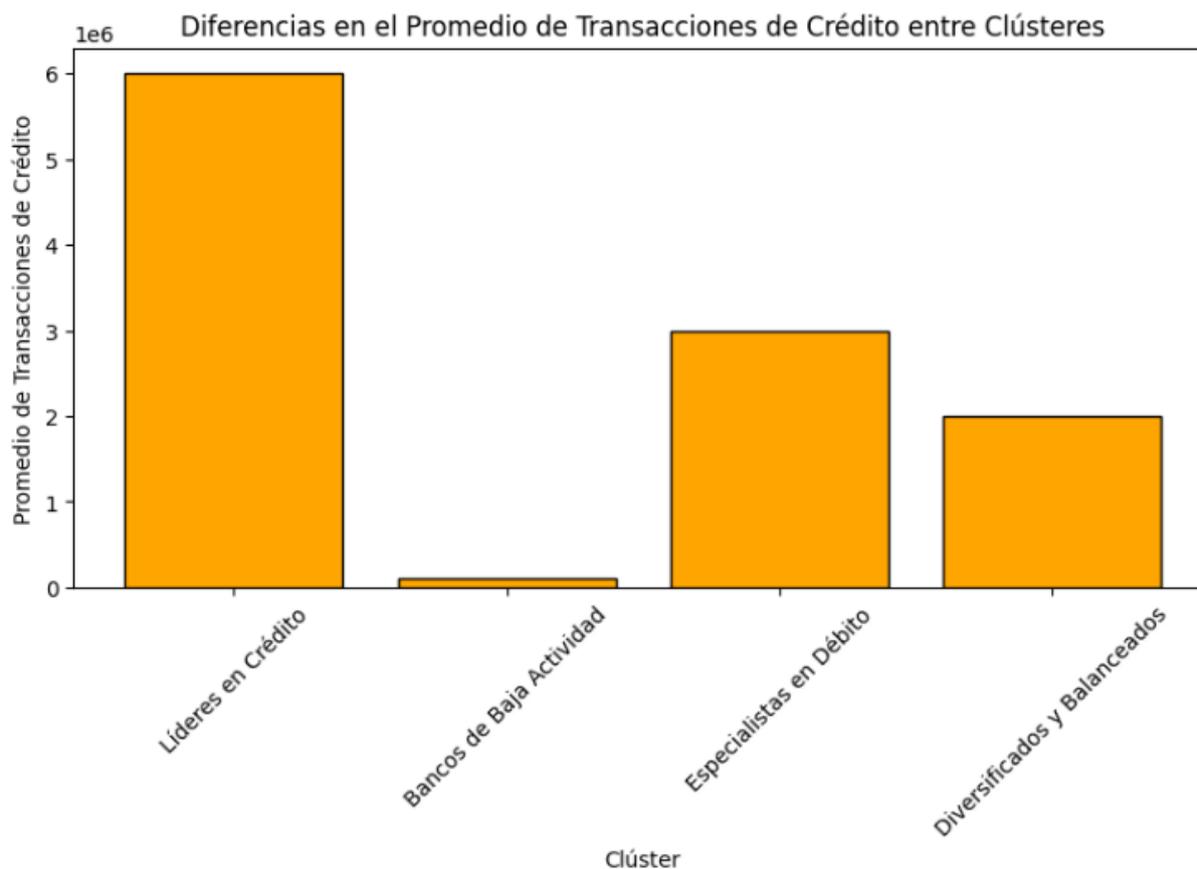
Tabla 6. Características por Clúster de Bancos según el Uso de Tarjetas de Crédito y Débito

Los gráficos presentados reflejan los principales resultados obtenidos del análisis de segmentación de bancos basado en su comportamiento transaccional. Estos resultados destacan diferencias significativas entre los clústeres identificados, como la proporción de transacciones de crédito y débito, el promedio de montos y transacciones, así como la distribución del número de bancos por clúster.



Img 21 "EDA de Tarjetas de crédito." Origen el autor.

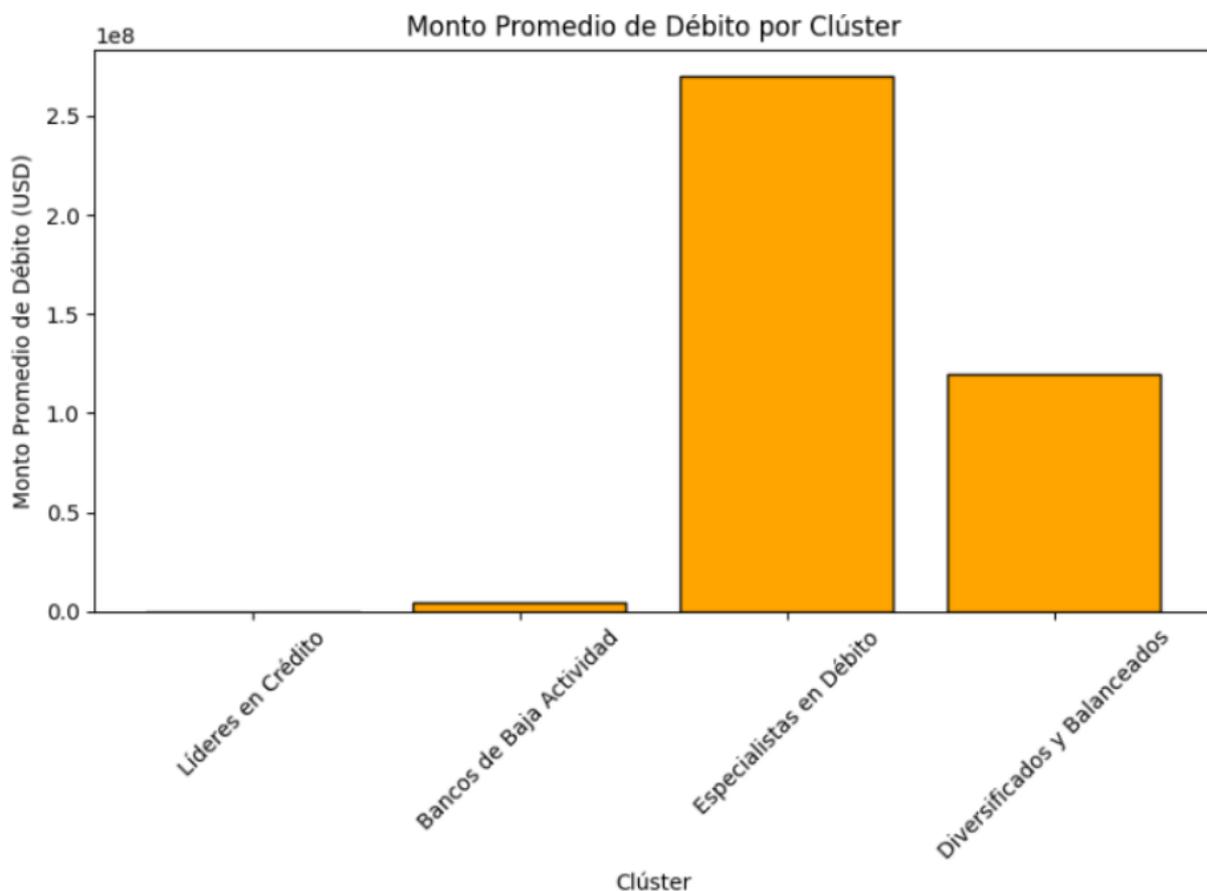
El gráfico Img 21 muestra las diferencias en el promedio del monto en dólares estadounidenses (USD) de transacciones con tarjetas de crédito entre los clústeres identificados. Se destaca que el clúster de "Líderes en Crédito" presenta el mayor monto promedio de transacciones de crédito, superando los 400 millones de dólares, lo que sugiere que estas instituciones tienen un enfoque predominante en el financiamiento a través de tarjetas de crédito. En contraste, los "Bancos de Baja Actividad" muestran valores considerablemente inferiores, indicando una participación mínima en el mercado crediticio. Los "Especialistas en Débito", a pesar de que su principal actividad está centrada en transacciones con tarjetas de débito, presentan un volumen de crédito intermedio, lo que sugiere que algunas instituciones dentro de este grupo también ofrecen líneas de crédito de manera significativa. Por su parte, el clúster de "Diversificados y Balanceados" mantiene una estrategia más equilibrada entre el uso de tarjetas de débito y crédito, con un monto promedio de crédito menor al de los líderes, pero superior al de los bancos de baja actividad.



Img 22 "EDA de Tarjetas de crédito." Origen el autor.

El gráfico Img 22 complementa este análisis al mostrar las diferencias en el promedio del número de transacciones de crédito entre los clústeres. Se observa que el clúster de "Líderes en Crédito" no solo tiene el mayor monto de transacciones, sino que también registra el mayor número de transacciones de crédito, superando los seis millones. Esto indica que estas instituciones manejan tanto un alto volumen de transacciones como montos elevados, consolidándose como las principales en el sector crediticio. En el otro extremo, los "Bancos de Baja Actividad" presentan un número marginal de transacciones, lo que refuerza la idea de su escasa participación en este segmento. Los "Especialistas en Débito" muestran un número considerable de transacciones de crédito, en línea con su monto promedio, lo que confirma que, aunque su enfoque principal es el débito, mantienen una presencia significativa en el mercado de crédito. Finalmente, los bancos "Diversificados y Balanceados" registran un número de transacciones de crédito menor que los especialistas en débito, pero superior al de los bancos de baja actividad, lo que refuerza su estrategia de equilibrio entre ambos tipos de transacciones.

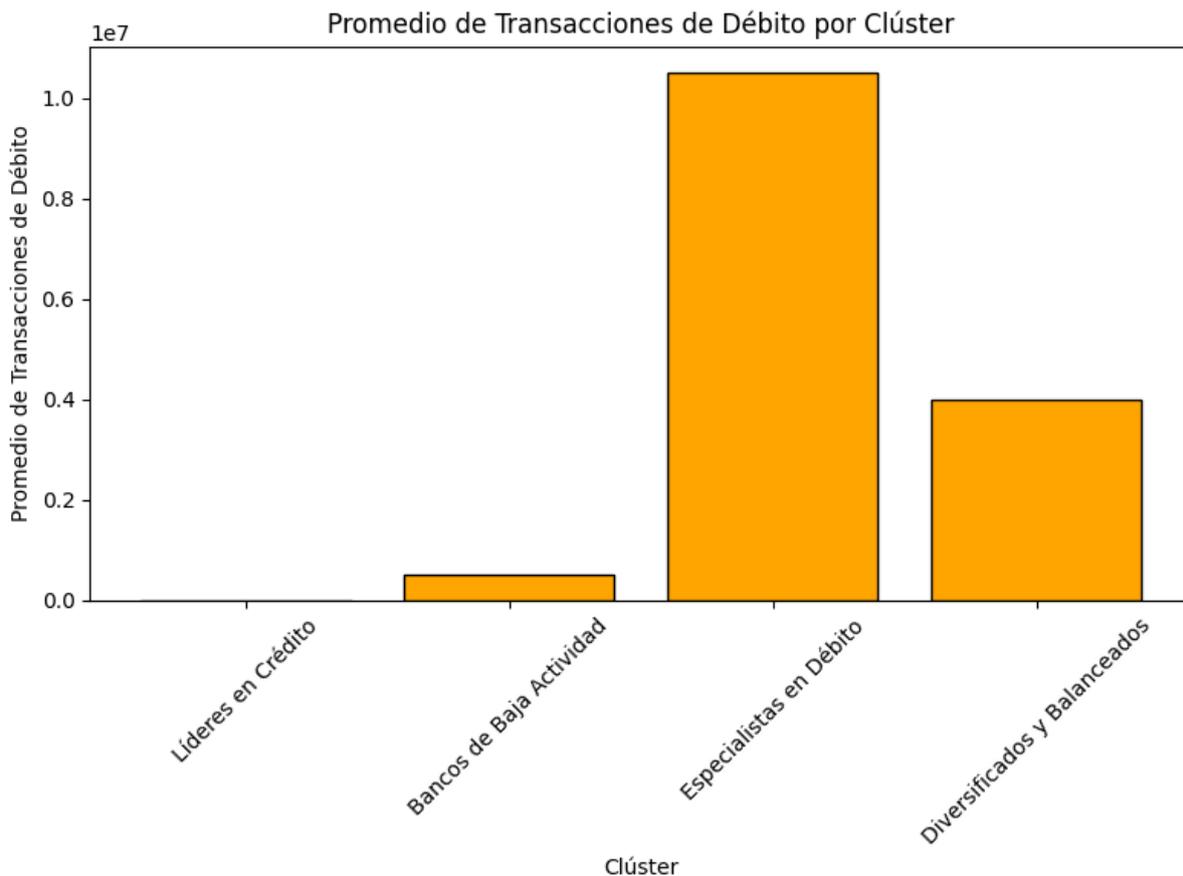
En conjunto, estos gráficos permiten identificar patrones claros en el comportamiento de los clústeres. Las instituciones financieras dentro del clúster de "Líderes en Crédito" pueden enfocar sus estrategias en fortalecer la fidelización de clientes y la expansión de su oferta crediticia. Los "Bancos de Baja Actividad" podrían considerar estrategias para incrementar su participación en este segmento, mientras que los "Especialistas en Débito" pueden evaluar si desean diversificar aún más sus productos financieros. Por su parte, las instituciones "Diversificadas y Balanceadas" tienen la oportunidad de consolidar su presencia tanto en crédito como en débito, optimizando su portafolio de servicios para atender mejor las necesidades de sus clientes.



Img 23 "EDA de Tarjetas de débito." Origen el autor.

El gráfico Img 23 muestra el monto promedio de transacciones con tarjetas de débito por clúster. Se observa que los "Especialistas en Débito" tienen el monto más alto, superando los 250 millones de dólares. Esto indica que las instituciones financieras en este clúster dependen fuertemente del uso de tarjetas de débito como su principal medio de transacción. En segundo lugar, el clúster de "Diversificados y Balanceados"

presenta un monto considerable, lo que sugiere que estas instituciones mantienen un equilibrio entre transacciones de débito y crédito. En contraste, los "Líderes en Crédito" muestran un uso insignificante de tarjetas de débito, lo que reafirma su enfoque en productos financieros basados en crédito. Finalmente, los "Bancos de Baja Actividad" tienen el menor monto promedio, lo que confirma su baja participación en el mercado transaccional.

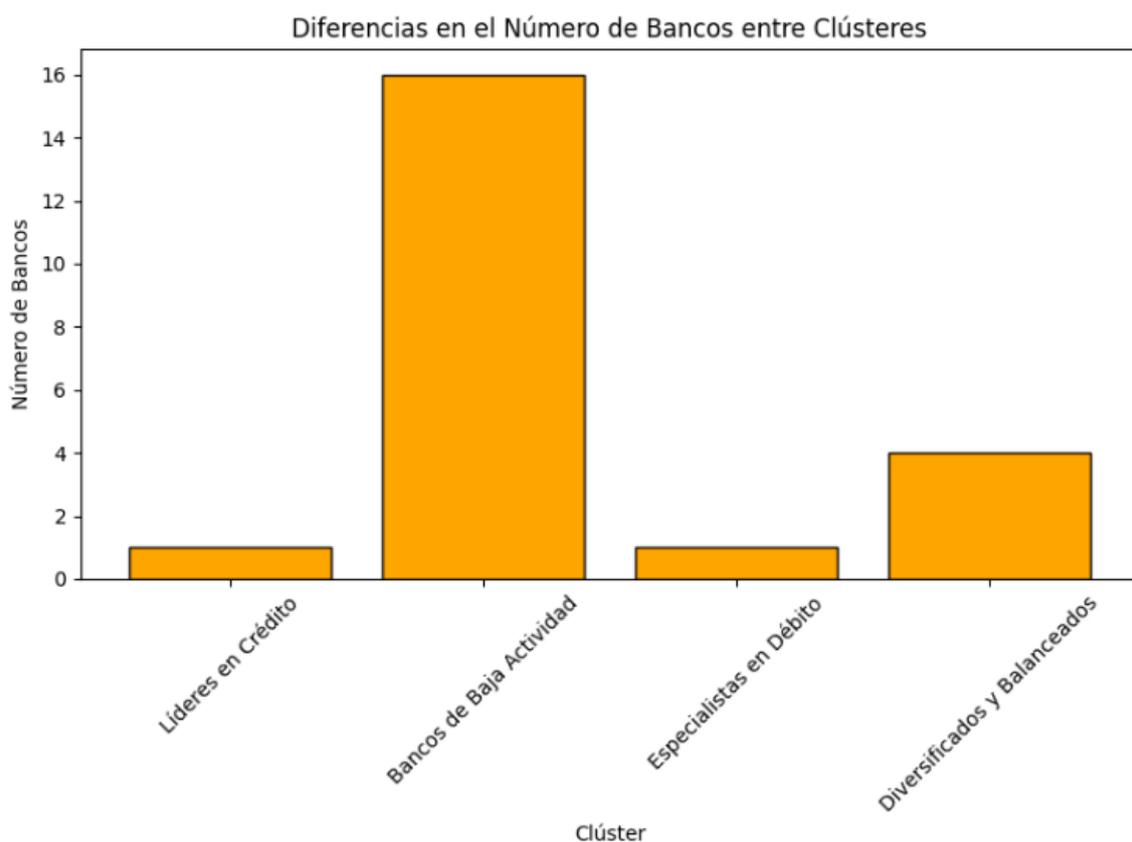


Img 24 "EDA de Tarjetas de débito." Origen el autor.

El gráfico Img 24 complementa este análisis al mostrar el número promedio de transacciones de débito por clúster. Se aprecia que los "Especialistas en Débito" no solo tienen el mayor monto transaccionado, sino también el mayor número de transacciones, superando los diez millones. Esto confirma que estas instituciones tienen un alto volumen operativo basado en pagos con débito. Por otro lado, los "Diversificados y Balanceados" presentan un número considerable de transacciones, aunque menor que el de los "Especialistas en Débito", lo que refuerza la idea de que combinan el uso de crédito y débito en su oferta de servicios. En contraste, los "Líderes en Crédito" muestran una cantidad mínima de transacciones con débito, lo

que sugiere que sus clientes dependen en gran medida de instrumentos financieros basados en crédito. Por su parte, los "Bancos de Baja Actividad" presentan cifras marginales, lo que indica una baja penetración en el mercado transaccional.

Estos resultados reflejan la clara segmentación de las instituciones financieras en cuanto al uso de tarjetas de débito. Los "Especialistas en Débito" pueden enfocarse en fortalecer la adopción de pagos digitales y ofrecer incentivos a clientes que utilizan este método con frecuencia. Los "Diversificados y Balanceados" pueden continuar con una estrategia dual, asegurando un crecimiento sostenible tanto en crédito como en débito. Los "Líderes en Crédito", por su parte, podrían considerar estrategias para diversificar su portafolio si desean incrementar su presencia en el mercado de débito. Finalmente, los "Bancos de Baja Actividad" podrían beneficiarse de campañas que incentiven el uso de tarjetas de débito, mejorando su competitividad en el sector.

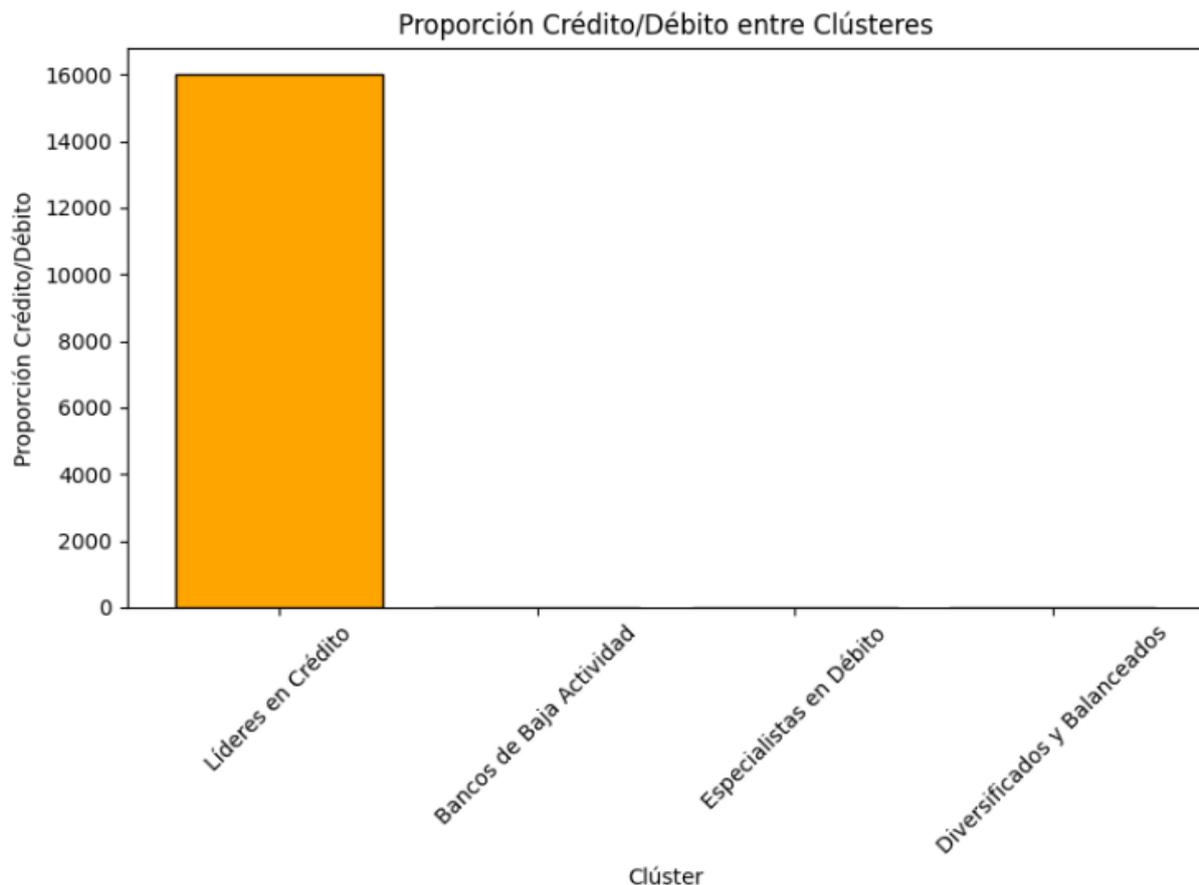


Img 25 "Número de emisores por clúster." Origen el autor.

El gráfico Img 25 muestra la distribución del número de bancos en cada clúster, destacando que la mayoría de las instituciones se agrupan en el clúster de "Bancos de Baja Actividad", con 16 emisores. Esto indica que muchas entidades financieras tienen una participación limitada en transacciones con tarjetas

En contraste, los "Líderes en Crédito" y "Especialistas en Débito" tienen pocos bancos, lo que sugiere que solo algunas instituciones dominan estos segmentos. El clúster de "Diversificados y Balanceados" tiene una presencia intermedia, reflejando estrategias mixtas entre crédito y débito.

Estos resultados sugieren que los "Bancos de Baja Actividad" podrían beneficiarse de estrategias de crecimiento para incrementar su presencia en el mercado de tarjetas, mientras que los bancos líderes pueden enfocarse en consolidar su especialización.



Img 26 "Razón Crédito/Débito entre clústeres." Origen el autor.

El gráfico Img 26 muestra la proporción entre transacciones de crédito y débito en cada clúster. Se observa que el clúster de "Líderes en Crédito" tiene una proporción extremadamente alta, lo que indica que estas instituciones se especializan casi exclusivamente en el uso de tarjetas de crédito. En contraste, los demás clústeres no presentan valores significativos, lo que sugiere que su uso de crédito es mucho más equilibrado con el débito o incluso dominado por este último.

Este resultado confirma la fuerte segmentación del mercado financiero, donde un grupo reducido de bancos se enfoca casi por completo en productos crediticios, mientras que el resto mantiene estrategias más diversificadas o centradas en débito. Las instituciones dentro del clúster de "Líderes en Crédito" pueden continuar consolidando su modelo de negocio basado en crédito, mientras que los demás clústeres podrían analizar oportunidades para optimizar su mix de productos financieros.

Conclusiones

Las conclusiones del análisis destacan diversos propósitos y aplicaciones relacionadas con la segmentación y gestión de bancos basada en su comportamiento transaccional. En primer lugar, la segmentación de clientes o bancos permite comprender cómo se agrupan los bancos según su comportamiento, lo que facilita el diseño de estrategias personalizadas para cada segmento. Esto incluye enfocar productos financieros específicos en bancos con alta actividad en crédito o débito y priorizar recursos en aquellos con mayor potencial de crecimiento o necesidades particulares.

La optimización de productos o servicios financieros surge como una oportunidad clave para desarrollar soluciones adaptadas a las necesidades de los bancos en cada clúster. Por ejemplo, se pueden ofrecer productos centrados en crédito para los Líderes en Crédito, diseñar estrategias enfocadas en débito para los Especialistas en Débito y equilibrar los portafolios de productos en bancos diversificados del clúster Diversificados y Balanceados.

En cuanto al benchmarking y monitoreo, se plantea la posibilidad de comparar el desempeño de bancos dentro y entre clústeres. Esto permite identificar líderes en cada segmento, monitorear cómo evolucionan los bancos entre clústeres a lo largo del tiempo y analizar patrones de desempeño que puedan ser replicados para fomentar estrategias exitosas en otros bancos.

La gestión de riesgos cobra relevancia al evaluar riesgos financieros asociados con los diferentes clústeres. Por ejemplo, los bancos con alta actividad en crédito, agrupados en el clúster Líderes en Crédito, podrían presentar mayores riesgos de incumplimiento, mientras que aquellos con baja actividad en ambos sectores, dentro del clúster Bancos de Baja Actividad, podrían indicar inestabilidad financiera o mercados menos activos.

La identificación de oportunidades de expansión es otra conclusión importante, ya que permite detectar bancos con potencial de crecimiento en áreas donde no están activos. Los bancos en el clúster Bancos de Baja Actividad, por ejemplo, podrían ser

candidatos para campañas de expansión en productos de crédito o débito, mientras que los bancos Diversificados y Balanceados podrían beneficiarse de estrategias de diversificación.

En términos de estrategias de marketing, se sugiere crear campañas adaptadas al perfil de cada clúster. Esto incluye promocionar tarjetas de débito en el clúster Especialistas en Débito, ofrecer beneficios para altos volúmenes de crédito en los Líderes en Crédito y diseñar programas de fidelización para los bancos dentro del clúster Diversificados y Balanceados.

Finalmente, en cuanto a la planificación y asignación de recursos, se recomienda priorizar inversiones y esfuerzos operativos en función de la actividad de los bancos. Esto incluye asignar más recursos de soporte técnico y ventas a bancos con alta actividad y concentrar los esfuerzos de integración tecnológica en clústeres con alta actividad en crédito o débito.

Recomendaciones

Las recomendaciones planteadas buscan enriquecer y contextualizar el análisis, proporcionando acciones concretas para que sea más accionable y alineado con los objetivos estratégicos de las organizaciones. En primer lugar, la contextualización del análisis se orienta a proporcionar un marco interpretativo más profundo. Esto implica agregar información sobre el tamaño, mercado objetivo y región de los bancos, así como explicar los factores subyacentes que influyen en los patrones observados en los clústeres, como políticas económicas, comportamiento de clientes o regulaciones. Además, es clave vincular el análisis con metas estratégicas como el aumento de la participación de mercado, la reducción de costos o la mejora de la rentabilidad.

La visualización avanzada se propone como una forma de facilitar la interpretación y comunicación de los resultados. Esto incluye el uso de gráficos interactivos que permitan explorar los datos de manera dinámica, la incorporación de mapas geográficos si los bancos están distribuidos en diferentes regiones, y la presentación de tendencias temporales para visualizar la evolución de los clústeres. Se recomienda también desarrollar un dashboard interactivo con herramientas como Power BI o Tableau para que equipos internos puedan aprovechar los hallazgos en tiempo real.

El análisis temporal es esencial para comprender la dinámica y evolución de los clústeres a lo largo del tiempo. Este enfoque incluye analizar las razones detrás de los cambios de clúster de los bancos, como el aumento de transacciones o variaciones económicas, así como incorporar predicciones para identificar posibles futuros cambios en la clasificación. También se sugiere identificar patrones relacionados con estaciones o eventos específicos que puedan influir en la actividad bancaria, como el fin de año o períodos de recesión económica.

Finalmente, la integración con datos adicionales es una recomendación clave para enriquecer el análisis. Esto implica agregar métricas de desempeño financiero como ingresos, ganancias y tasas de interés, así como datos demográficos de los clientes de cada banco. Además, es fundamental cruzar la información con datos de campañas de marketing o adopción de productos para entender mejor cómo estas

iniciativas impactan en la actividad transaccional y en la posición de los bancos dentro de los clústeres.

Bibliografía

- Corporación del Seguro de Depósitos (COSEDE). (2024). *Código Orgánico Monetario y Financiero*. Obtenido de <https://www.cosedec.gov.ec/wp-content/uploads/2024/03/Codigo-Organico-Monetario-y-Financiero-2024.pdf>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer. Obtenido de *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
- Ivan Ishchenko, L. G. (2018). Approach to determining the number of .
- Moosa, A. (2020). *PyCaret: An open-source, low-code machine learning library in Python*. *PyCaret Documentation*. . Obtenido de <https://pycaret.readthedocs.io/en/latest/api/>.html
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. . (2011). *Scikit-learn: Machine Learning in Python*. *Journal of Machine Learning Research*, 12, 2825-2830. . Obtenido de <https://scikit-learn.org/stable/modules/preprocessing.html#scaling-features-to-a-range>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). *Scikit-learn: Machine Learning in Python*. *Journal of Machine Learning Research*, 12, 2825-2830. Obtenido de <https://scikit-learn.org/stable/modules/generated/s>
- Superintendencia de Bancos del Ecuador. (sf de sf de 2024). *Servicios Financieros*. Obtenido de <https://www.superbancos.gob.ec/estadisticas/portalestudios/servicios-financieros/>