



Universidad de los Andes

SEGMENTACIÓN DE CLIENTES PARA FINANZASALPES

Laboratorio 2 - Agrupación



Grupo 19



Agenda

1 Contexto

- Introducción a la segmentación de clientes en FinanzasAlpes
- Objetivo del análisis

2 Perfil de los Clientes

- Datos analizados y su relevancia para la segmentación

3 Estrategia de Solución

- Metodología CRISP-ML aplicada
- Algoritmos seleccionados y su justificación

4 Resultados y Comparación de Modelos

- Evaluación de cada algoritmo
- Visualización y descripción de los clusters obtenidos

5 Conclusiones y Recomendaciones

- Principales hallazgos
- Aplicación de la segmentación en estrategias comerciales



Contexto

📌 **FinanzasAlpes** es una empresa financiera especializada en la emisión de tarjetas de crédito y soluciones de pago, enfocada en ofrecer productos personalizados según los hábitos de consumo de sus clientes.

◆ RETOS ACTUALES:

- Identificar patrones de consumo en sus clientes.
- Optimizar estrategias de marketing y fidelización.
- Desarrollar productos financieros adaptados a las necesidades de cada segmento.

💡 ¿POR QUÉ ES RELEVANTE LA SEGMENTACIÓN?

La empresa busca implementar una estrategia de segmentación de clientes basada en técnicas de aprendizaje no supervisado para personalizar su oferta y mejorar la retención de usuarios.

Análisis de Datos

DESCRIPCIÓN DEL CONJUNTO DE DATOS

-  Total de registros: 8,950 clientes
-  Total de variables: 18 atributos

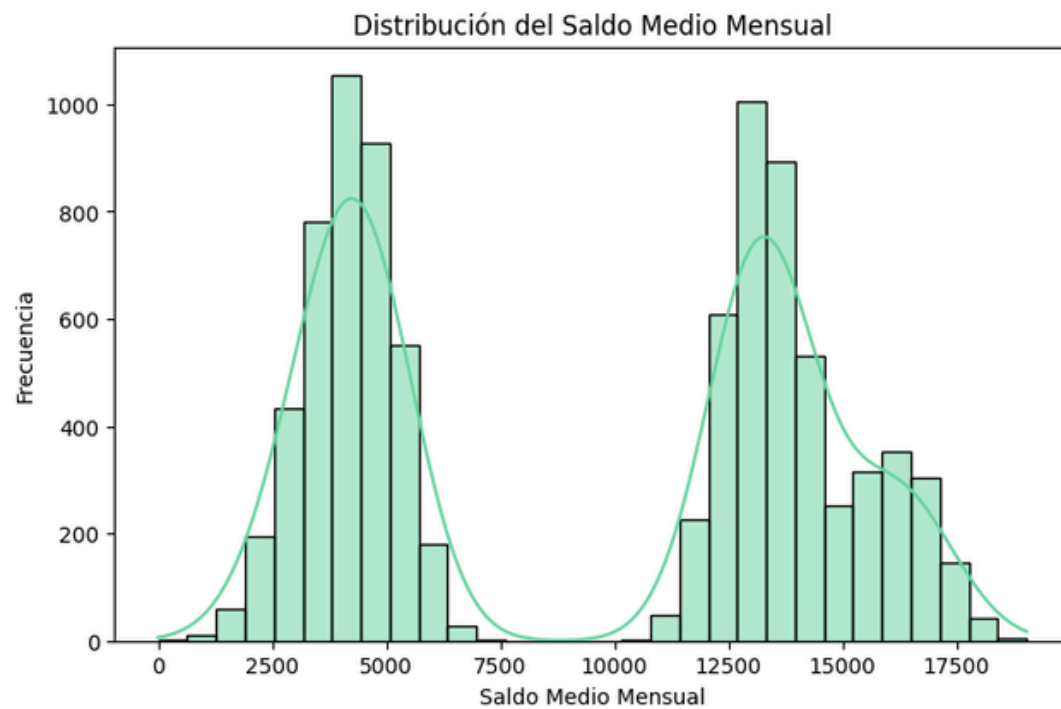
OBJETIVO DEL ANÁLISIS

- ✓ Identificar patrones en los clientes según sus hábitos de consumo y pagos.
- ✓ Segmentar clientes para personalizar estrategias comerciales y de crédito.

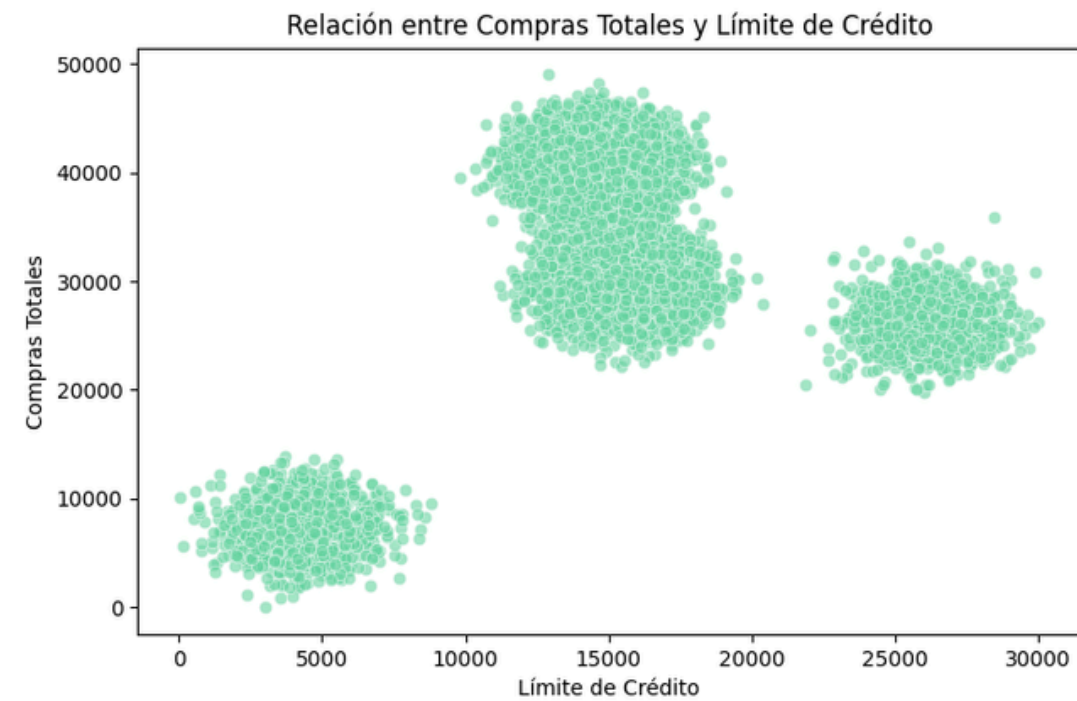
PRINCIPALES VARIABLES

- ✓ Comportamiento financiero: Saldo medio, pagos, pagos mínimos y límite de crédito.
- ✓ Patrones de compra: Importe total y frecuencia de compras (puntuales, a plazos y en efectivo).
- ✓ Historial y fidelidad: Antigüedad del cliente y frecuencia de actualización del saldo.

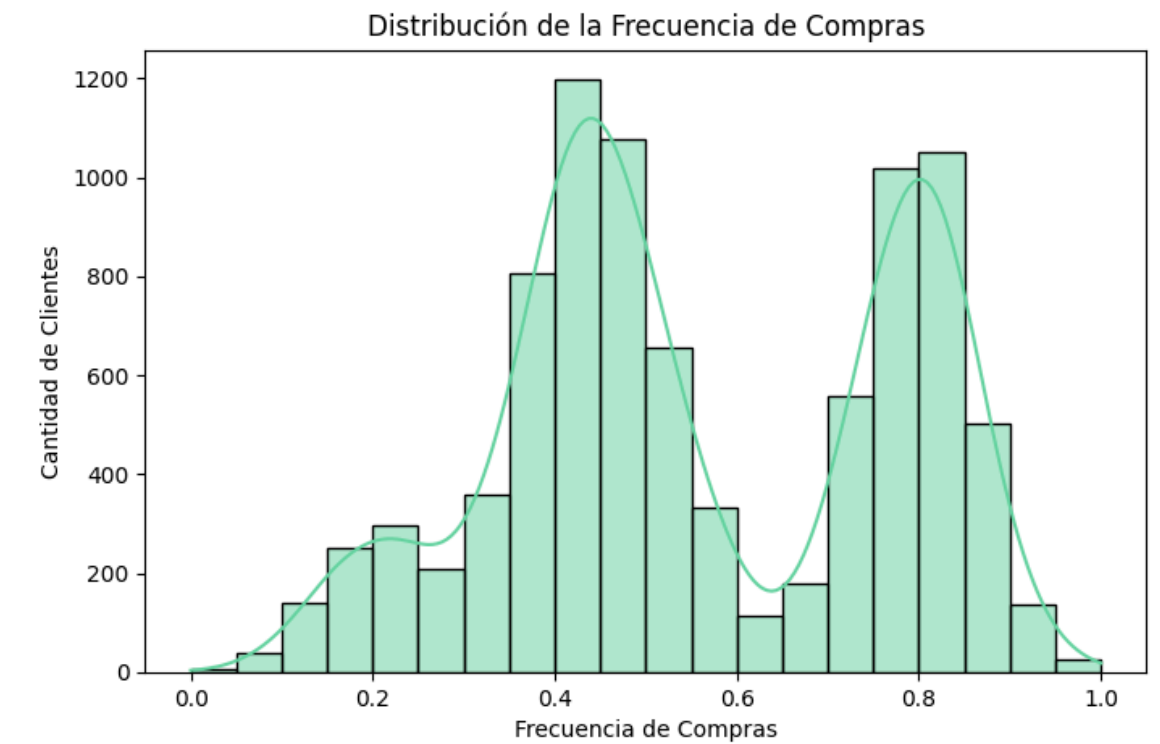
Graficas relevantes



- La mayoría de los clientes tienen un saldo medio mensual en **dos rangos** bien definidos, lo que sugiere la presencia de dos grupos distintos con diferentes patrones de saldo.

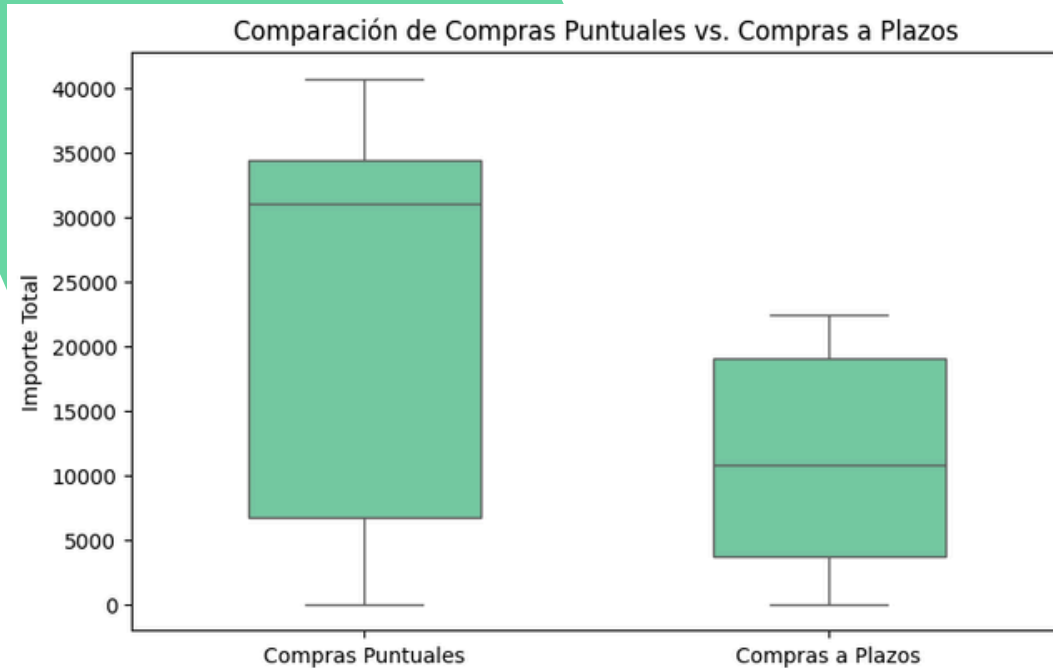


- Los clientes se agrupan en tres niveles de límite de crédito y compras totales. Esto indica que los clientes con **mayor crédito suelen gastar más**, pero hay variabilidad dentro de cada grupo.

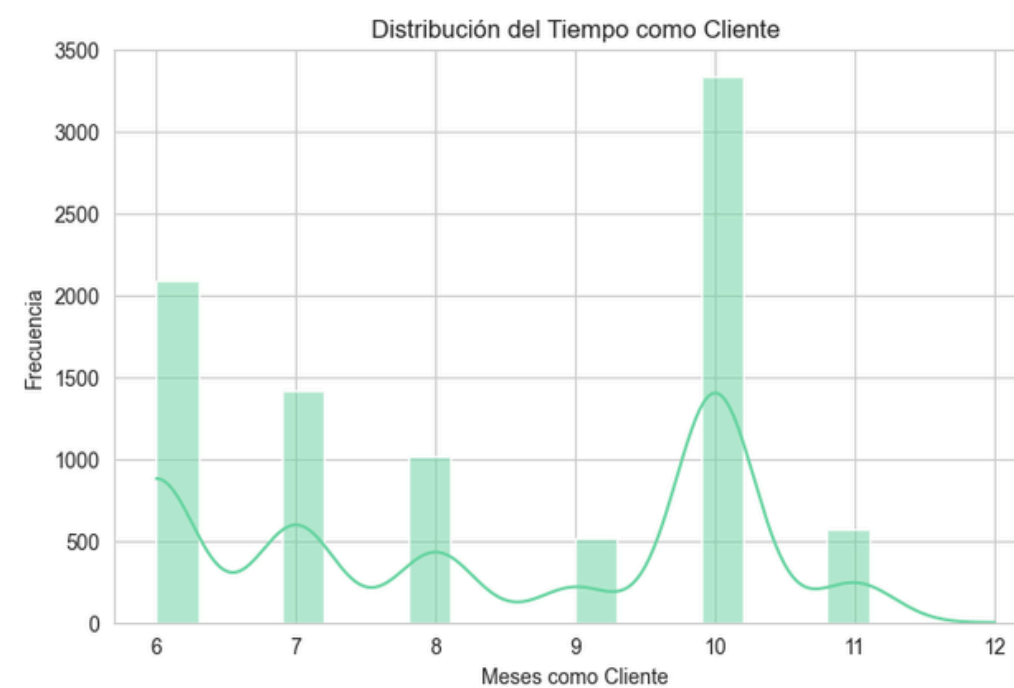


- Hay dos tipos de clientes: unos que compran con poca frecuencia y otros que lo hacen muy seguido. Esto indica que hay **diferencias en los hábitos de compra**, lo que podría ayudar a crear estrategias para que los clientes compren más seguido.

Graficas relevantes



- Los clientes que hacen compras puntuales gastan mucho más en total que aquellos que compran a plazos. Esto sugiere que quienes pagan de una vez **suelen hacer compras más grandes**, mientras que los que prefieren pagar en cuotas gastan menos en cada compra.

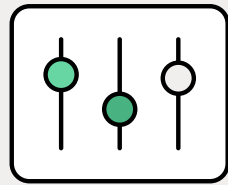


- Los clientes se concentran en 6 y 10 meses de antigüedad, con un pico en 10 meses, posiblemente por campañas promocionales. La baja en 9 y 11 meses sugiere **posible deserción**.



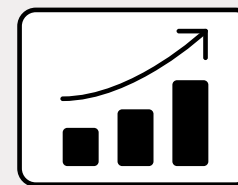
- Se observan tres grupos bien definidos, lo que indica que los clientes tienden a agruparse en ciertos niveles de saldo y pago. Esto sugiere que existen **diferentes perfiles de clientes**: algunos con saldos y pagos bajos, otros con valores medios y un grupo con valores altos.

Análisis de Calidad de Datos



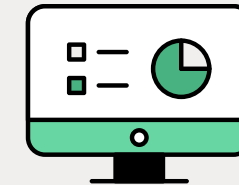
✓ COMPLETITUD:

- Hay algunos datos faltantes en "Pagos Mínimos" (3.5%) y "Límite de Crédito" (0.01%).
- Se pueden eliminar, completar con un valor específico o usar un promedio.
- Las demás variables están completas.



✓ UNICIDAD:

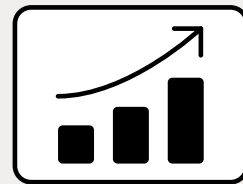
No hay datos repetidos.



VALIDEZ:

- La mayoría de los datos están dentro de los valores esperados.
- La Frecuencia de los avances en efectivo tiene un 33.96% de valores fuera del rango correcto.
- Se consultó con un experto y se decidió corregir esos valores.

Análisis de Calidad de Datos



✓ CONSISTENCIA:

La Frecuencia de los avances en efectivo es la única variable con datos inconsistentes, ya corregidos.



⊘ VARIABLE DESCARTADA:

La variable "ID" no aporta información útil y será eliminada.



Procesamiento del conjunto

✓ Ajuste de valores extremos:

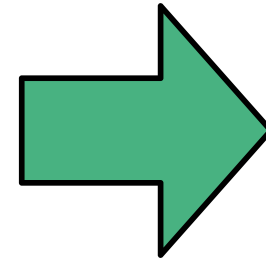
- Para evitar problemas en los cálculos, los datos se ajustan a una escala estándar.

✓ Conversión de datos:

- Algunos algoritmos solo funcionan con números.
- Se transformó la variable del número de meses como cliente en varias columnas numéricas para facilitar el análisis.

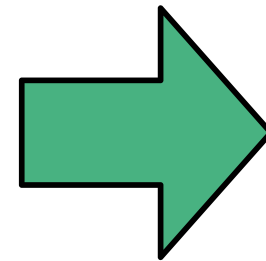


Modelos



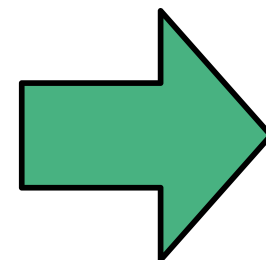
K-MEANS

* Natalia Villegas Calderon - 202113370



AGGLOMERATIVE CLUSTERING

* Juan Martin Vasquez - 202113314

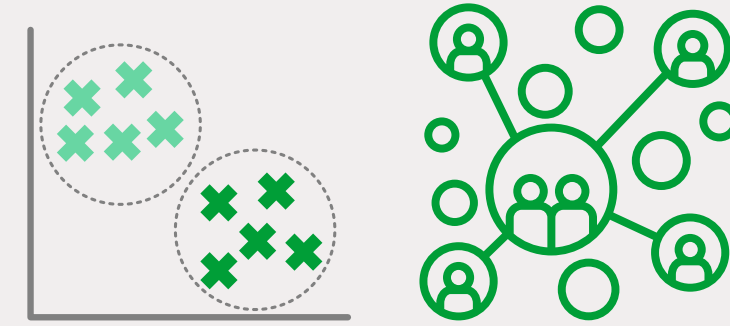


GAUSSIAN MIXTURE MODEL (GMM)

- Carol Sofia Florido Castro - 202111430

K-Means

- Algoritmo de agrupación que segmenta clientes con características similares.
- Permite personalizar productos y servicios según el comportamiento del cliente.
- Facilita campañas de marketing dirigidas y mejora la experiencia del usuario.



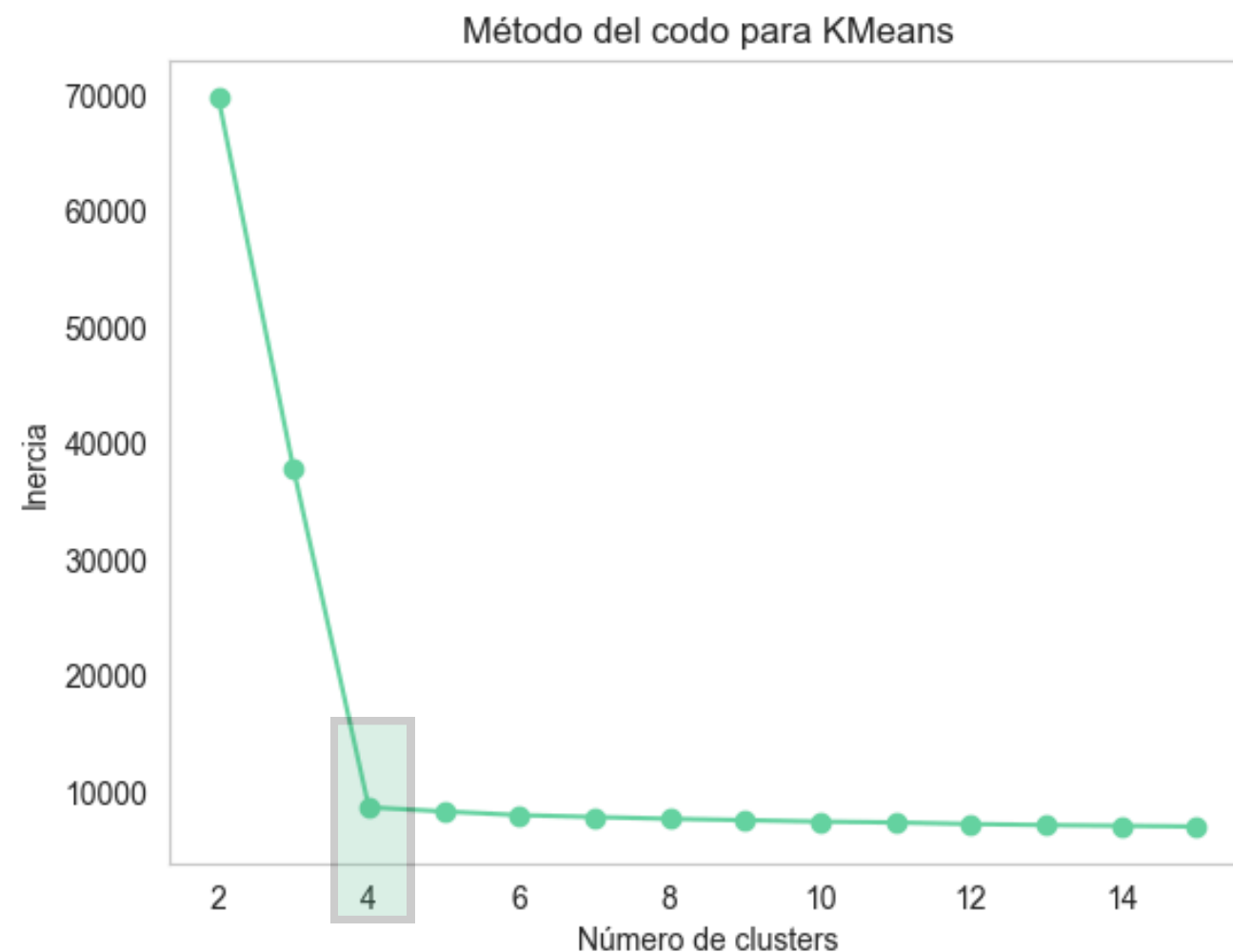
¿Qué se necesita determinar?

- Número **óptimo de clusters (k)** para agrupar a los clientes correctamente.
- Se utiliza **método del codo** y **método de la silueta** para elegir el mejor valor de k.

Heurísticas para determinar número de clusters

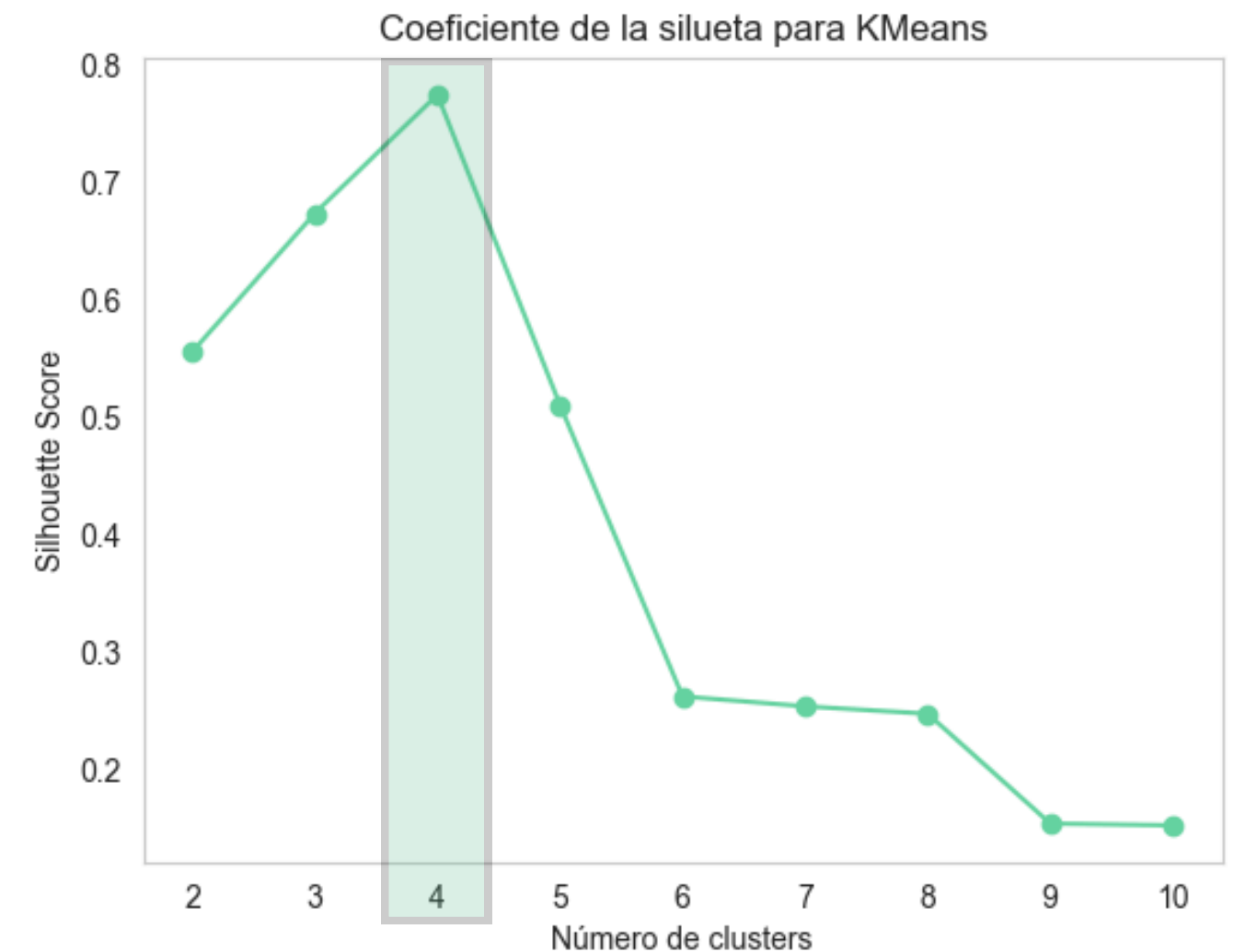
✓ MÉTODO DEL CODO:

Grafica la inercia (variabilidad dentro de los clusters) en función del número de grupos. El punto óptimo es donde la inercia deja de disminuir significativamente, formando un "codo" en la curva.



✓ MÉTODO DEL COEFICIENTE DE SILUETA:

Evalúa la calidad de los clusters según qué tan bien agrupados y separados están los puntos. Se elige k donde el coeficiente de silueta (de -1 a 1) es más alto.



Implementación de K-Means

📌 Configuración del modelo

- ✓ Número de clusters (k): 4
- ✓ Inicialización: "auto" (optimiza el proceso de inicialización)
- ✓ Aleatoriedad controlada (para reproducibilidad)

📝 Desempeño y resultados

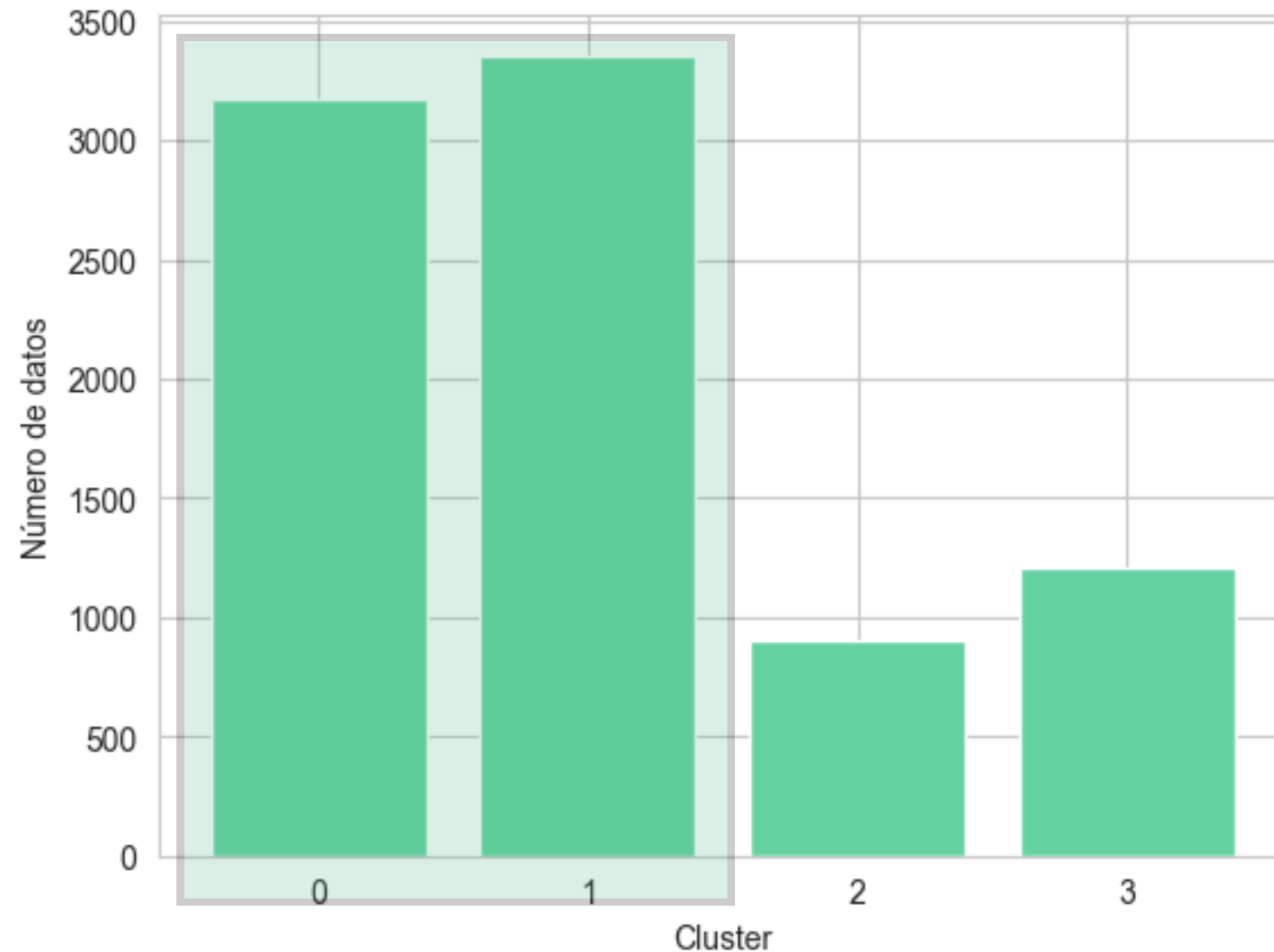
COEFICIENTE DE SILUETA

0.7739

Indica una segmentación bien definida, con clientes claramente agrupados y bien separados entre clusters.

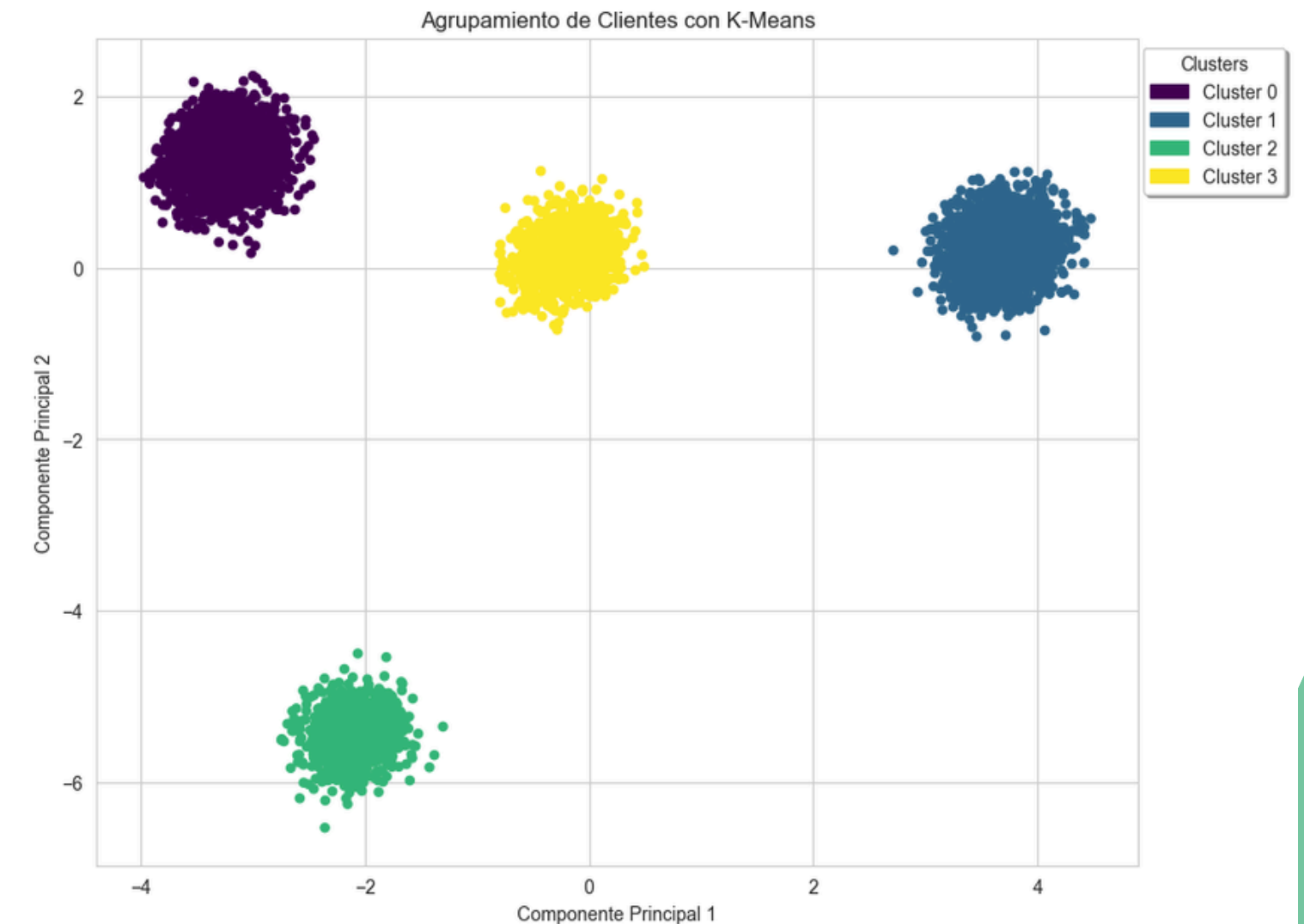
Distribución de los Clusters

La distribución de los clusters muestra que los grupos 0 y 1 tienen la mayor cantidad de clientes, mientras que el cluster 2 es el más pequeño, lo que sugiere una segmentación desigual pero bien definida.



Visualización de los Clusters

Confirma que los clusters están claramente diferenciados, con una estructura bien definida, lo que sugiere que los clientes dentro de cada grupo comparten características similares.



Agglomerative Clustering

- Se aplicó un modelo de Agrupamiento Aglomerativo (Agglomerative Clustering) para identificar grupos ocultos dentro de los datos. Este método de clustering jerárquico permitió descubrir patrones naturales en los datos, formando grupos mediante la fusión progresiva de observaciones similares, clasificando los datos de manera determinística según su similitud.

¿Mejores hiperparámetros del modelo?

- Se evaluaron diferentes valores de número de clusters, métodos de enlace (linkage) y métricas de distancia, asegurando que el método ward solo se combinara con la distancia euclidiana mediante un gridsearch.
- Para cada combinación, se entrenó un modelo y se calcularon los coeficientes de silueta, permitiendo medir la cohesión y separación de los clusters.

#Clusters	linkage	metrica	silhouette_score
4	complete	l2	0.773856
4	complete	l1	0.773856
...			
7	complete	manhattan	0.143534
7	complete	l1	0.143534

Implementación de Agglomerative Clustering

Configuración del modelo

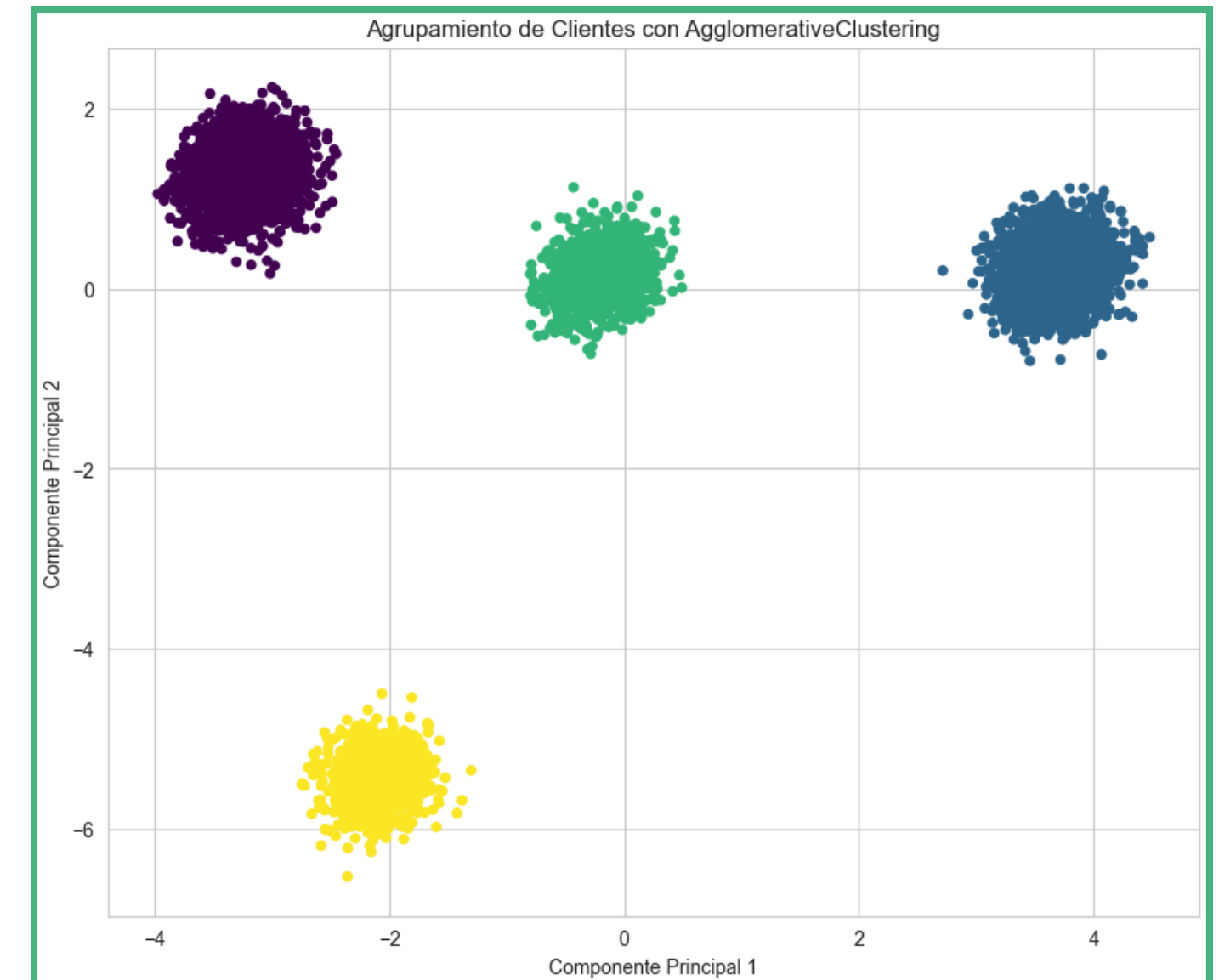
- **Número de clusters (k):** 4
- **linkage:** complete (mayor distancia entre cualquier par de puntos)
- **metric:** l2 (norma distancia Euclidiana mide la distancia entre 2 puntos)

Coeficiente de Silueta

0.77

Visualización de los Clusters:

Después de reducir los datos a 2 dimensiones con PCA, se observa que el algoritmo de clustering logró una segmentación clara, con cuatro grupos bien definidos y sin solapamiento. Esto indica que el modelo capturó correctamente las diferencias entre los clusters, validando su efectividad en la identificación de patrones en los datos. Lo cual se confirma mediante el coeficiente de silueta.



Gaussian Mixture Model (GMM)

- Se aplicó un modelo llamado GMM (Gaussian Mixture Model) para identificar grupos ocultos dentro de los datos. Este modelo permitió descubrir patrones naturales sin necesidad de definir de antemano cuántos grupos buscar, asignando a cada dato una probabilidad de pertenecer a cada grupo en lugar de una clasificación rígida.

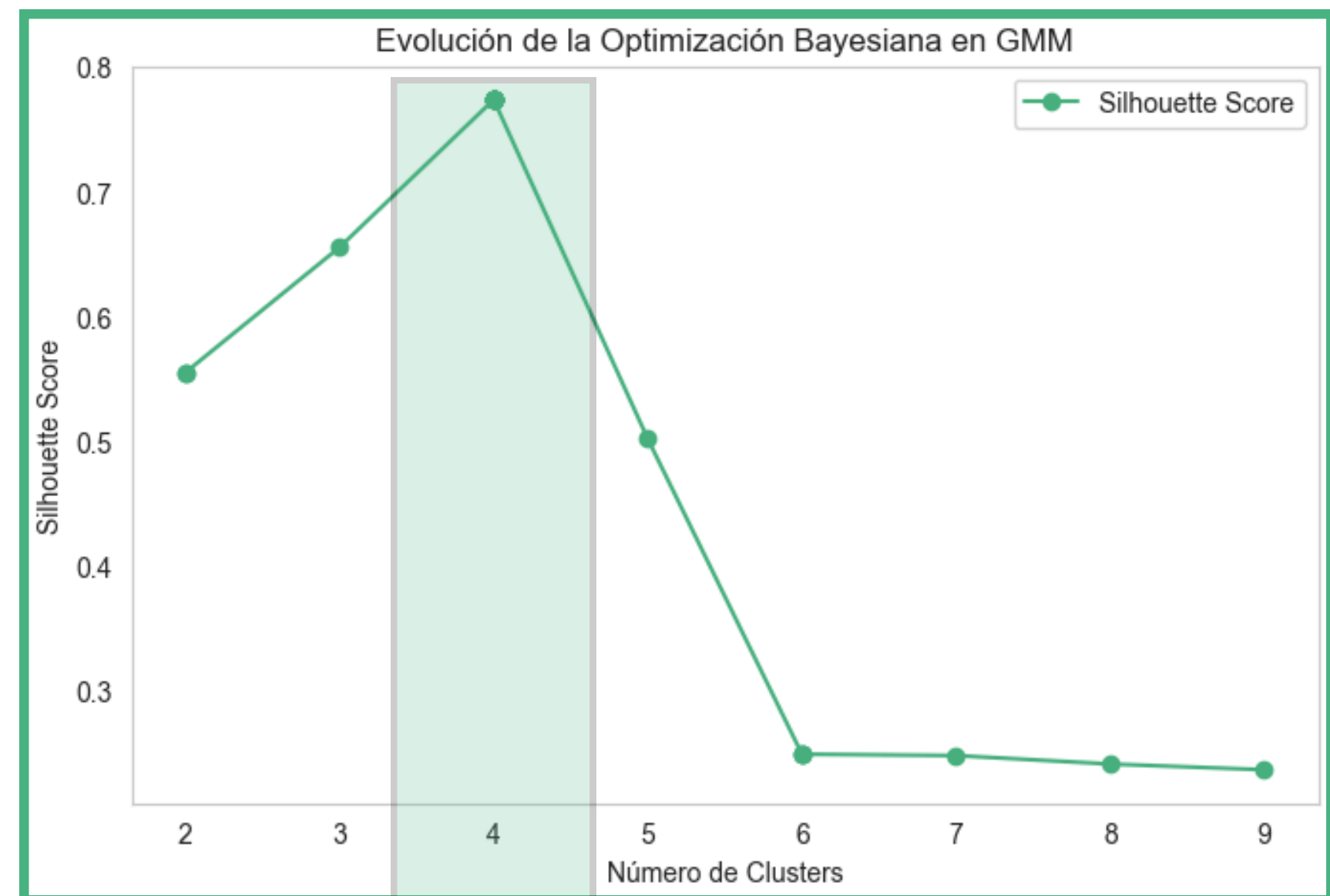
¿Cuántos grupos hay en los datos?

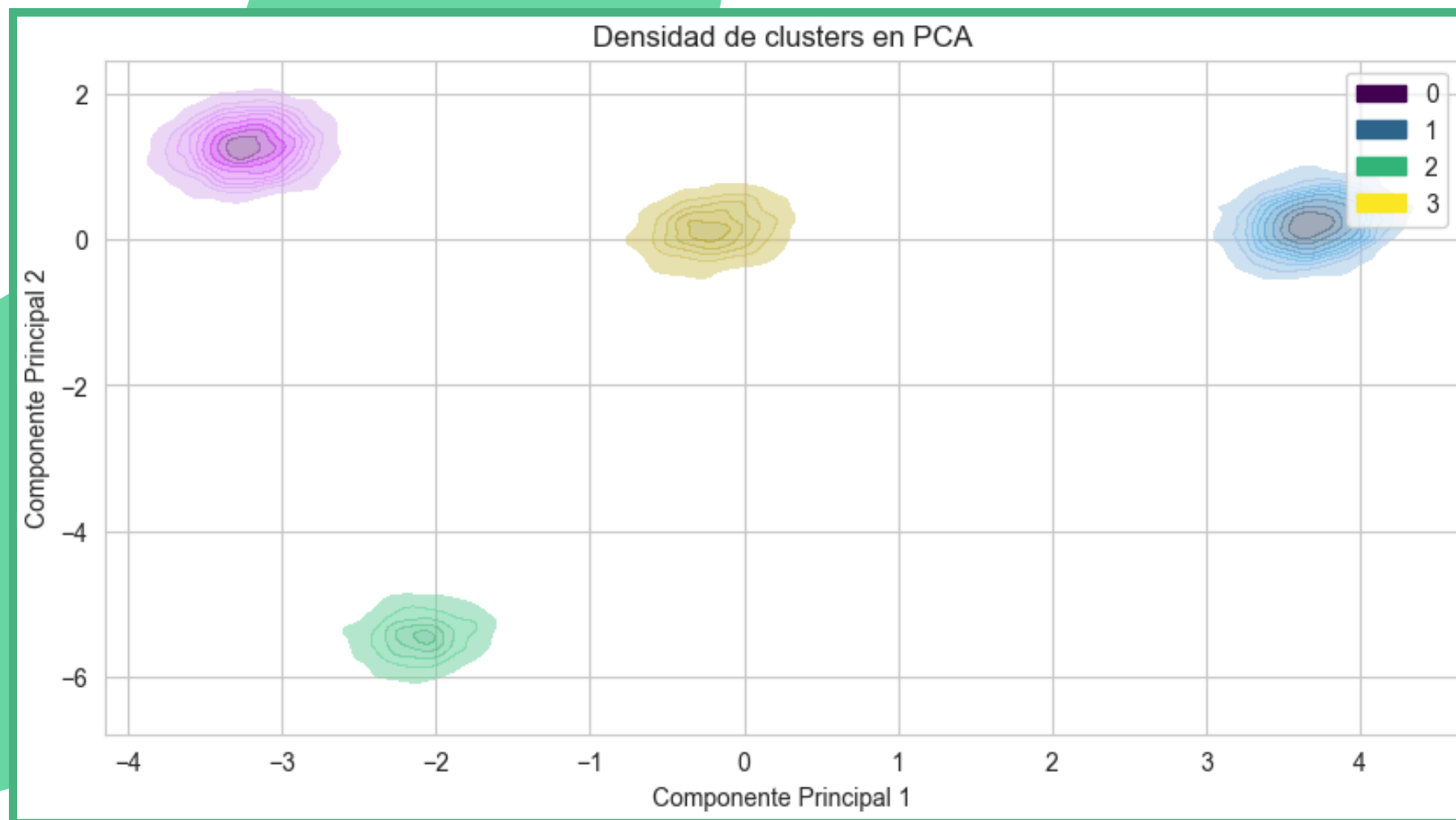
Para descubrirlo, se probaron diferentes opciones y se midió qué tan bien estaban separados los grupos.

🔍 Técnica utilizada: Se usó una estrategia de comparación para encontrar la mejor cantidad de grupos.

✅ Resultado: La mejor división se logró con 4 grupos, ya que permitía una separación clara y bien definida.

● Esto nos ayuda a entender mejor los datos y encontrar patrones ocultos.





📌 ¿Qué muestra esta gráfica?

- Representa los grupos encontrados en los datos después de aplicar el modelo.
- Se utilizó una técnica para reducir la información a dos dimensiones y facilitar su interpretación.

📌 ¿Qué observamos?

- Los colores representan diferentes grupos o clusters.
- Las zonas más oscuras indican mayor concentración de datos.
- Los clusters están bien separados, lo que indica que el modelo identificó patrones claros.

¿Qué se encontró?

- ◆ 4 grupos bien definidos que representan diferentes categorías dentro de la información analizada.
 - ◆ Los datos se agrupan de forma natural, con ciertas categorías más compactas y otras más dispersas.
 - ◆ Se confirma que nuestro modelo logró una segmentación precisa y útil.

Coeficiente de Silueta - GMM

0.77

Coeficiente de Silueta:

- Mide qué tan bien están agrupados los datos.
- Un valor cercano a 1 significa que los grupos están bien definidos y separados.
- Con GMM y 4 clusters, se obtuvo 0.77, indicando una segmentación clara y efectiva.

Comparación de Modelos



Coeficiente de Silueta

- K-means: 0.773856
- Agglomerative Clustering: 0.773856
- GMM: 0.773856



Clusters y su distribución

- Los tres modelos (K-Means, GMM y Agglomerative Clustering) generan segmentaciones muy similares.
- K-Means y GMM producen los mismos grupos, mostrando patrones consistentes.
- Agglomerative Clustering intercambia dos clusters, debido a su forma distinta de agrupar los datos.
- Los clusters 0 y 1 contienen la mayoría de los clientes, mientras que los otros dos representan segmentos más exclusivos.

Cluster	KMeans	Agglomerative	GMM
0	3,171	3,171	3,171
1	3,356	3,356	3,356
2	903	1,206	903
3	1,206	903	1,206

- Adicionalmente, la visualización con PCA muestra que los tres modelos de clustering (K-Means, GMM y Agglomerative Clustering) generan los mismos grupos, confirmando una segmentación consistente y estable sin importar el algoritmo utilizado.

El grupo escoge para realizar el análisis de caracterización de los clusters el modelo de:

K-Means

Se recomienda K-Means por su simplicidad, eficiencia y facilidad de interpretación, lo que permite una segmentación clara y rápida de los clientes. Además, ofrece resultados consistentes, alineados con otros métodos más complejos, pero con menor costo computacional y mayor facilidad para el análisis de los clusters.

Análisis de los clústers

✓ Cluster 0: Clientes Moderados y Responsables

✓ Datos desescalados:

- Saldo promedio: 4,451.13
- Compras altas y mayormente puntuales (35,140.27).
- Alto porcentaje de pagos completos (83.75%).

✓ Conclusión:

Grupo de clientes con manejo financiero estable y preferencia por compras de contado.

➔ **Acción recomendada:** Mantener su lealtad con beneficios en crédito y recompensas.

✓ Cluster 1: Alto Endeudamiento y Uso de Crédito

✓ Datos desescalados:

- Saldo alto (13,252.47) y alto uso de avances en efectivo (37,231.77).
- Gran parte de sus compras son a plazos (19,462.25).
- Bajo porcentaje de pagos completos (36.37%), indicando posible dependencia del crédito.

✓ Conclusión:

Clientes que tienen alto endeudamiento y riesgo financiero.

➔ **Acción recomendada:** Ofrecer planes de refinanciamiento y asesoría financiera.

✓ Cluster 2: Clientes de Alto Crédito y Buen Manejo Financiero

✓ Datos desescalados:

- Saldo más bajo (3,105.91) pero con altas compras puntuales (31,054.68).
- Bajo uso de avances en efectivo (9,355.10) y 72.33% de pagos completos.
- Mayor límite de crédito (284,805.09).

✓ Conclusión:

Clientes de alto crédito, buen historial financiero y manejo responsable del crédito.

➔ **Acción recomendada:** Ofertas premium y programas de fidelización.

✓ Cluster 3: Alto Saldo y Riesgo Financiero

✓ Datos desescalados:

- Saldo más alto (16,285.11) pero compras bajas (7,392.76).
- Prefieren compras a plazos (31,408.67) y tienen bajo porcentaje de pagos completos (14.95%).
- Uso intermedio de avances en efectivo (21,763.98).

✓ Conclusión:

Representan clientes con alto saldo, mayor uso de crédito y mayor riesgo de impago.

➔ **Acción recomendada:** Estrategias para mejorar sus pagos o reducir su dependencia del crédito.

Análisis de los clústers

La gráfica de radar muestra las diferencias clave entre los cuatro grupos de clientes, permitiendo identificar patrones en su comportamiento financiero y de uso de la tarjeta de crédito.

Características de los clústers: con los valores promedio

Cluster 1

Altos valores en saldo y pagos mínimos, lo que indica que estos clientes mantienen deudas elevadas y solo realizan pagos parciales de sus obligaciones. Además, tienen una alta frecuencia de compras, lo que sugiere que usan activamente el crédito, pero pueden estar en riesgo de sobreendeudamiento.

Cluster 0

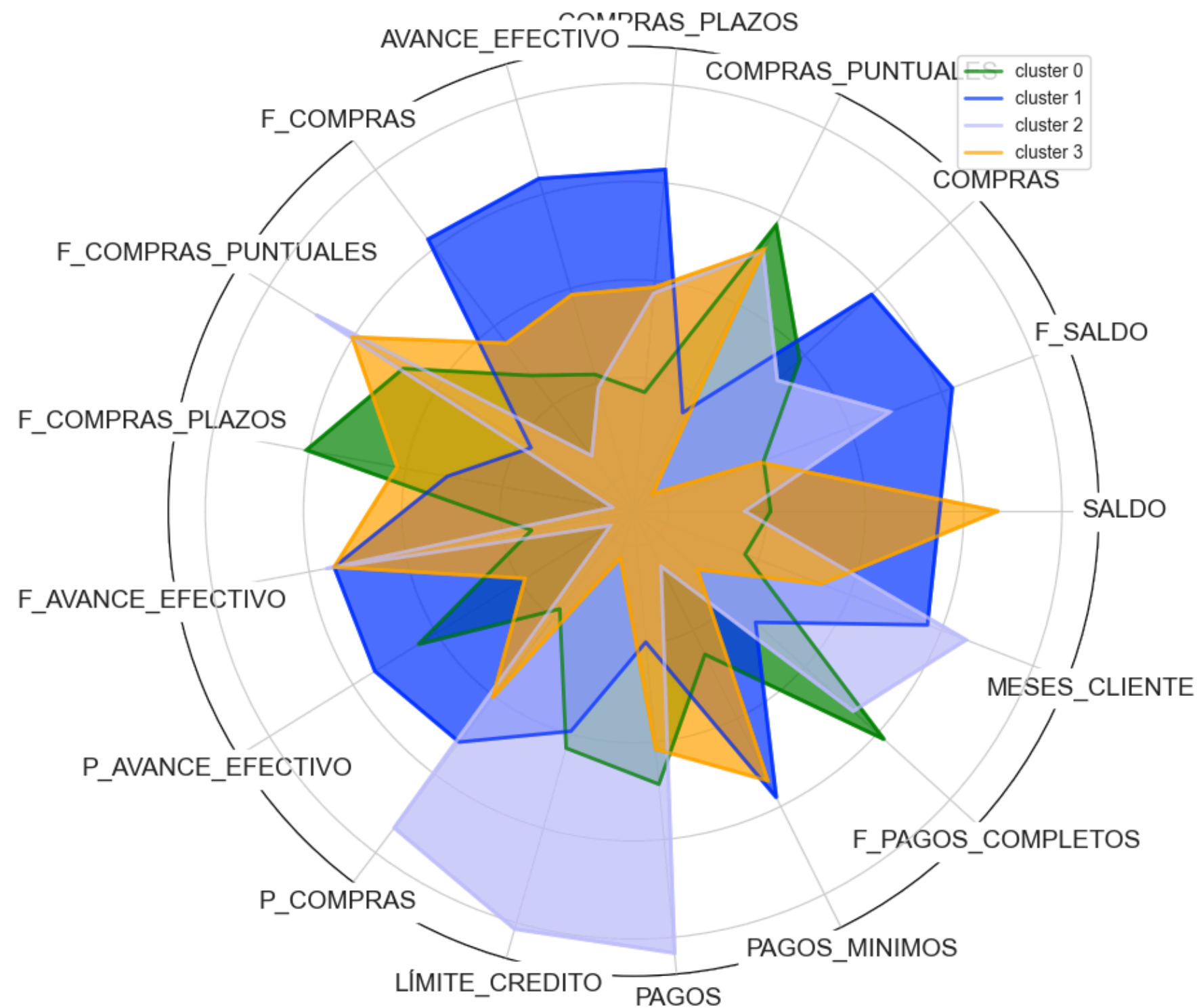
Valores moderados en la mayoría de las variables, con una ligera tendencia hacia compras a plazos y compras puntuales. Presentan un saldo estable y realizan pagos de manera regular, lo que indica que son usuarios responsables del crédito sin depender excesivamente del financiamiento.

Cluster 2

Altos valores en límite de crédito y pagos, lo que indica clientes con una buena capacidad de pago y un manejo responsable de sus finanzas. También presentan una tendencia a realizar pagos completos, lo que sugiere que no dependen excesivamente del crédito revolvente.

Cluster 3

Niveles moderados de saldo y compras, con un comportamiento equilibrado entre compras a plazos, avances en efectivo y pagos completos. Indica clientes que usan su tarjeta de forma activa pero sin llegar a altos niveles de endeudamiento.



Conclusión

Tras aplicar y comparar tres modelos de agrupación (K-Means, GMM y Agglomerative Clustering), se seleccionó K-Means por su simplicidad, eficiencia y facilidad de interpretación, permitiendo una segmentación clara de los clientes según su comportamiento de compra. Los cuatro clusters identificados muestran diferencias clave en el uso del crédito, hábitos de pago y nivel de endeudamiento, lo que proporciona información valiosa para la toma de decisiones estratégicas.

Recomendaciones:

- **Cluster 1 (Alto endeudamiento):** Ofrecer planes de refinanciamiento y asesoría financiera para fomentar pagos completos.
- **Cluster 2 (Clientes de alto crédito):** Incentivar productos premium y programas de fidelización exclusivos.
- **Cluster 3 (Usuarios activos pero equilibrados):** Ofrecer promociones para aumentar la lealtad sin incrementar el riesgo de impago.
- **Cluster 0 (Clientes moderados y responsables):** Mantener beneficios de crédito controlado y mejorar su experiencia con descuentos o recompensas.

✓ **Optimización de Campañas de Marketing:** Segmentar las estrategias de comunicación y promoción según las características de cada grupo, enfocando los esfuerzos en los clientes con mayor potencial de crecimiento o riesgo financiero.

✓ **Monitoreo y Actualización de la Segmentación:** Revisar periódicamente los clusters y actualizar las estrategias para adaptarse a cambios en el comportamiento de los clientes y tendencias del mercado.

Con esta segmentación, FinanzasAlpes podrá optimizar la personalización de sus servicios, reducir riesgos financieros y mejorar la rentabilidad a través de estrategias dirigidas, fortaleciendo su relación con los clientes y aumentando su fidelización.