
Estrategia De Accion



16 Agosto, 2024

El Problema

Diseñar una nueva estrategia de atracción para cerrar clientes de una segunda tarjeta de crédito. Teniendo en cuenta que el comportamiento del cliente es importante a la hora de entender el posible cierre de un negocio ustedes deciden diseñar un modelo de datos que permita establecer estos diferenciales para la estrategia



Data

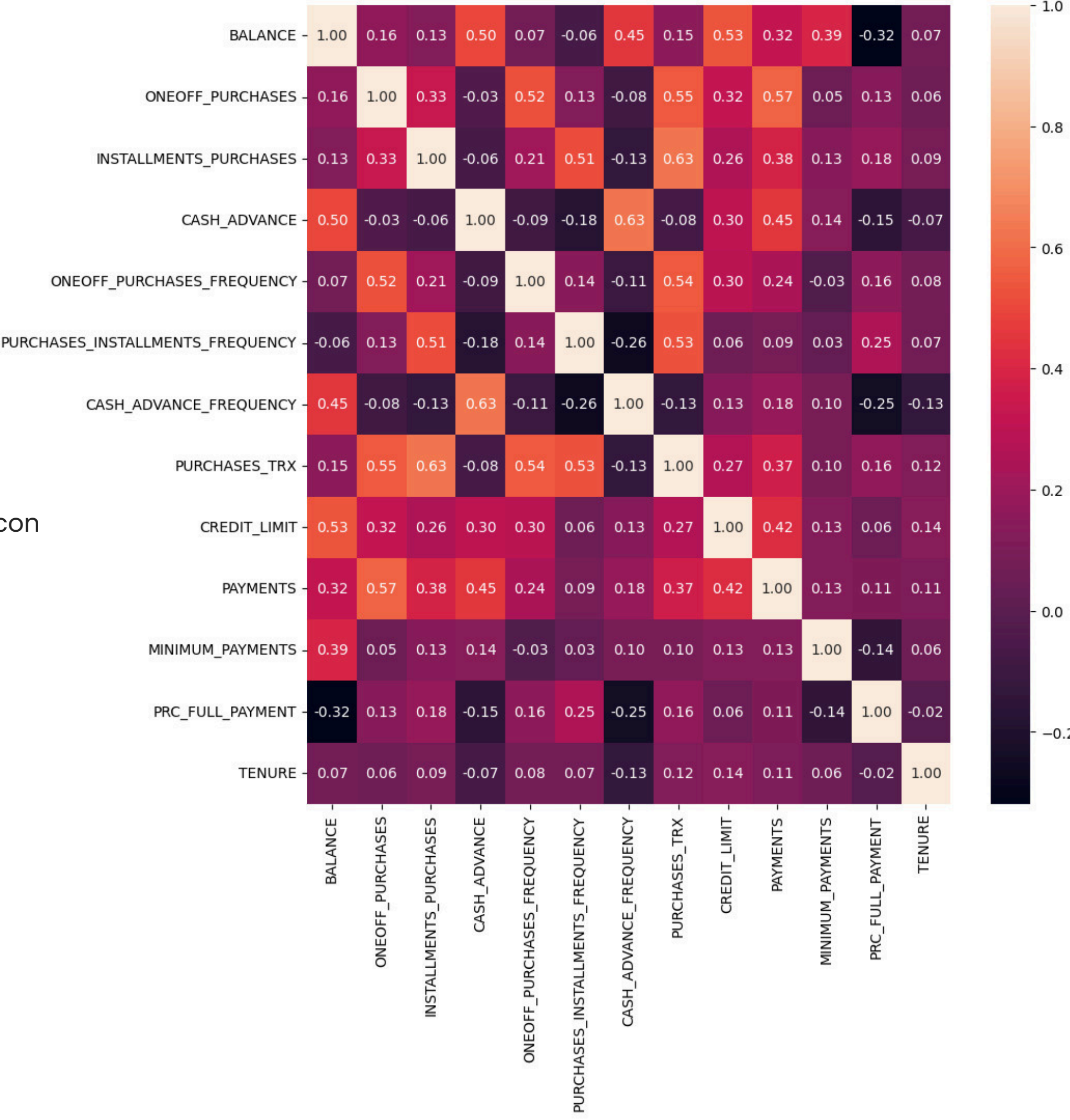
Se tienen datos de 8950 clientes los cuales tienen el conjunto de datos tiene el siguiente diccionario:

- CUST_ID: identificador del cliente
 - BALANCE: Monto restante en su cuenta de ahorros para hacer compras
 - BALANCE_FREQUENCY: Relación del riesgo del cliente (valor entre 0 y 1)
 - PURCHASES: Monto de compras realizado
 - ONEOFF_PURCHASES: Máximo de compras realizadas en una sola cuota
 - INSTALLMENTS_PURCHASES: Monto de compras realizadas a crédito
 - CASH_ADVANCE: Monto de avances en efectivo
 - PURCHASES_FREQUENCY: Calificación de la frecuencia de compras (valor entre 0 y 1)
 - ONEOFFPURCHASESFREQUENCY: Calificación de la frecuencia de compras a una sola cuota (valor entre 0 y 1)
 - PURCHASESINSTALLMENTSFREQUENCY: Calificación de la frecuencia de compras a crédito (valor entre 0 y 1)
 - CASHADVANCEFREQUENCY: Calificación de la frecuencia de avances (valor entre 0 y 1)
 - CASHADVANCETRX: Número de transacciones realizadas con avances
 - PURCHASES_TRX: Número de transacciones de compras hechas
 - CREDIT_LIMIT: Límite de crédito
 - PAYMENTS: Monto de pagos
 - MINIMUM_PAYMENTS: Monto mínimo de pagos
 - PRCFULLPAYMENT: Porcentaje de pagos completos hechos por el usuario
 - TENURE: Tenencia de tarjetas de crédito
-

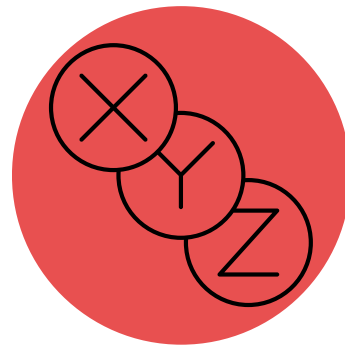
CORRELACION ENTRE VARIABLES

Entender la correlación entre variables es crucial al trabajar con modelos de agrupamiento (clustering).

- 1. Reducción de Redundancia
- 2. Mejora del Rendimiento del Modelo
- 3. Evitar la Multicolinealidad



Organizar los datos



IGNORAR VARIABLES

- Ignorar variables altamente correlacionadas (correlación directa fuerte y correlación inversa fuerte)
- Ignorar variables en la que todos los datos tiendan a un mismo valor

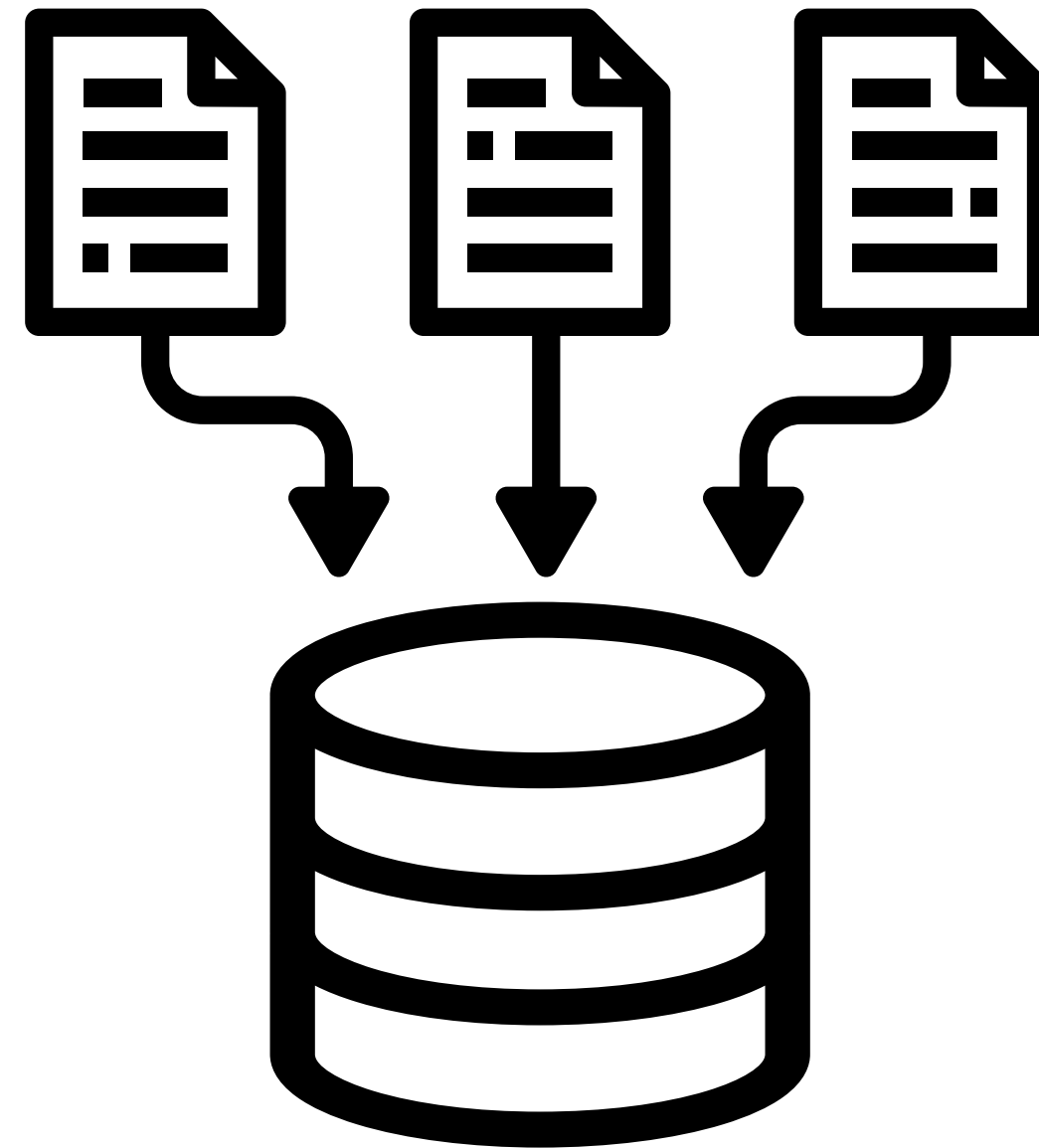


COMPLETAR DATOS FALTANTES

Ya que los datos faltantes son relativamente pocos comparados con la muestra total, completamos datos faltantes con la media de la variable

DATOS FINALES

- 0 BALANCE
- 1 ONEOFF_PURCHASES
- 2 INSTALLMENTS_PURCHASES
- 3 CASH_ADVANCE
- 4 ONEOFF_PURCHASES_FREQUENCY
- 5 PURCHASES_INSTALLMENTS_FREQUENCY
- 6 CASH_ADVANCE_FREQUENCY
- 7 PURCHASES_TRX
- 8 CREDIT_LIMIT
- 9 PAYMENTS
- 10 MINIMUM_PAYMENTS
- 11 PRC_FULL_PAYMENT
- 12 TENURE



SOLUCION



K-Means

K-means es una técnica de agrupamiento que se utiliza para organizar datos en grupos o "clusters".



Jerarquico

El agrupamiento jerárquico es otra técnica para organizar datos en grupos, pero lo hace de una manera que se parece más a construir un árbol genealógico o una pirámide de relaciones.

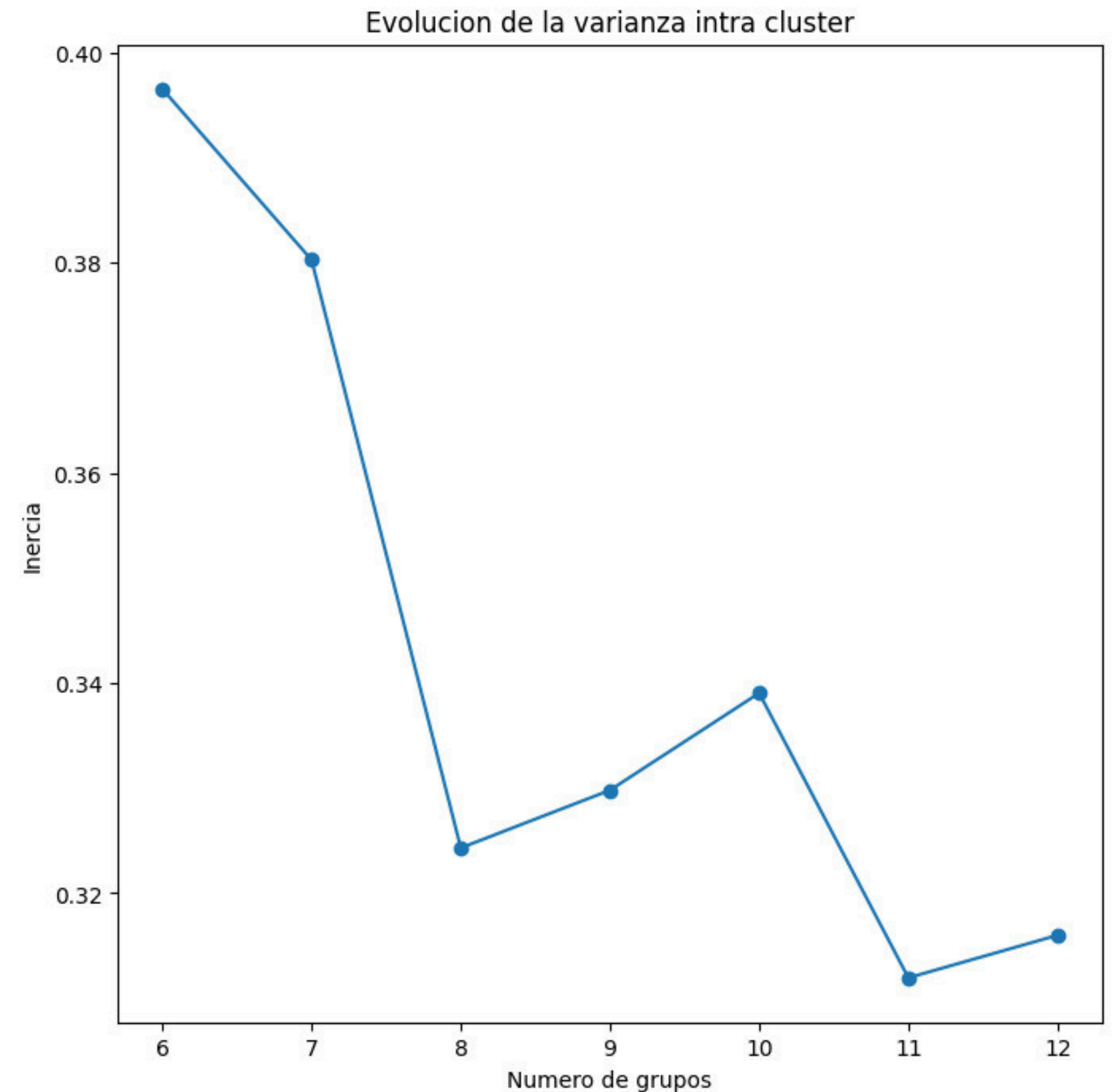


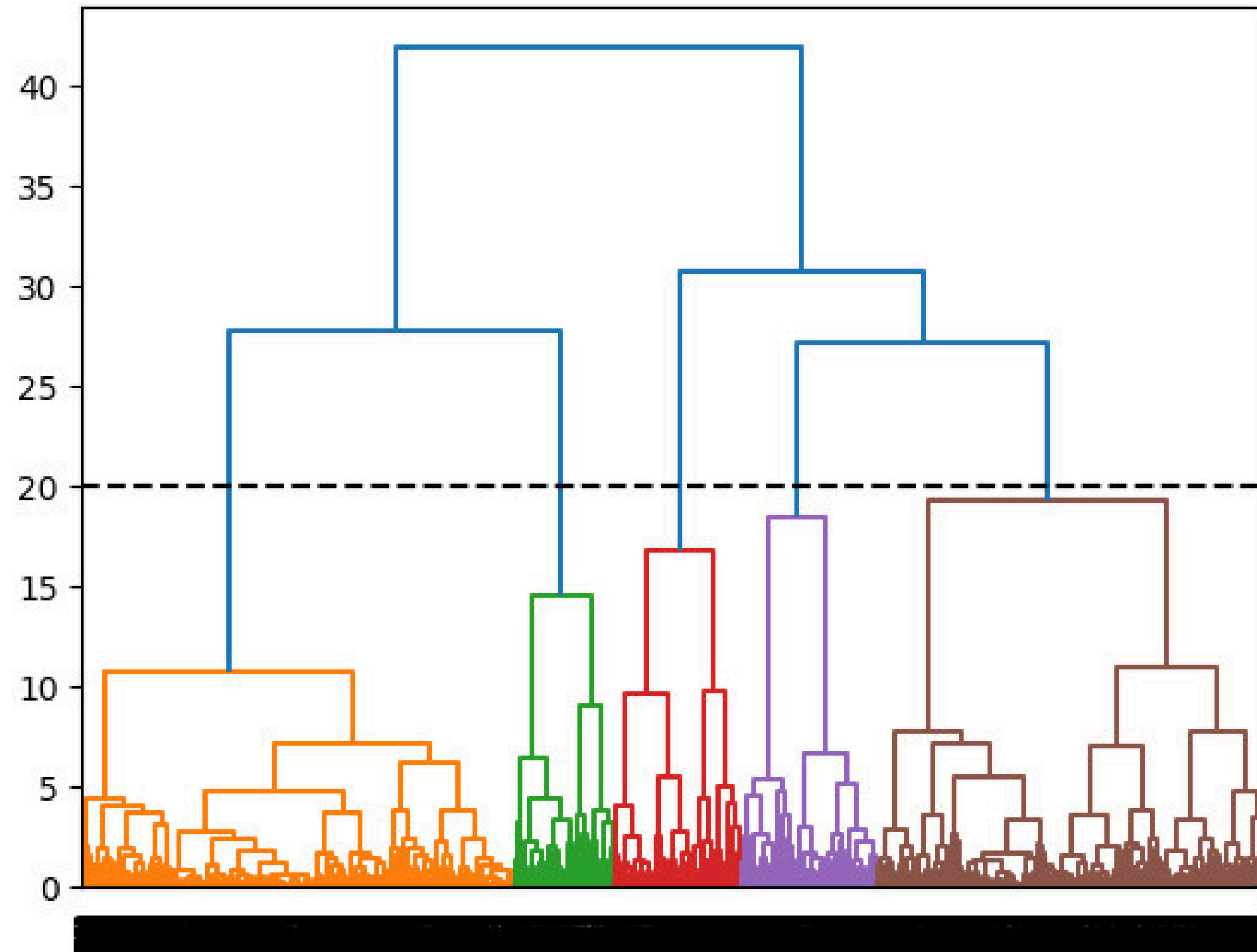
DB - SCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) es otro método para agrupar datos, pero funciona de manera diferente a K-means o el agrupamiento jerárquico. DBSCAN es una técnica de agrupamiento basada en la densidad de los datos.

K - Means

Utilizando el método de agrupamiento K - Means, se identificaron 6 grupos óptimos. Cada uno de estos grupos representa un segmento distinto dentro del conjunto de datos, permitiendo un análisis más detallado de las características y comportamientos de los elementos en cada categoría. Este agrupamiento facilita la personalización de estrategias y soluciones específicas para cada grupo, mejorando la eficiencia y efectividad de las decisiones tomadas.



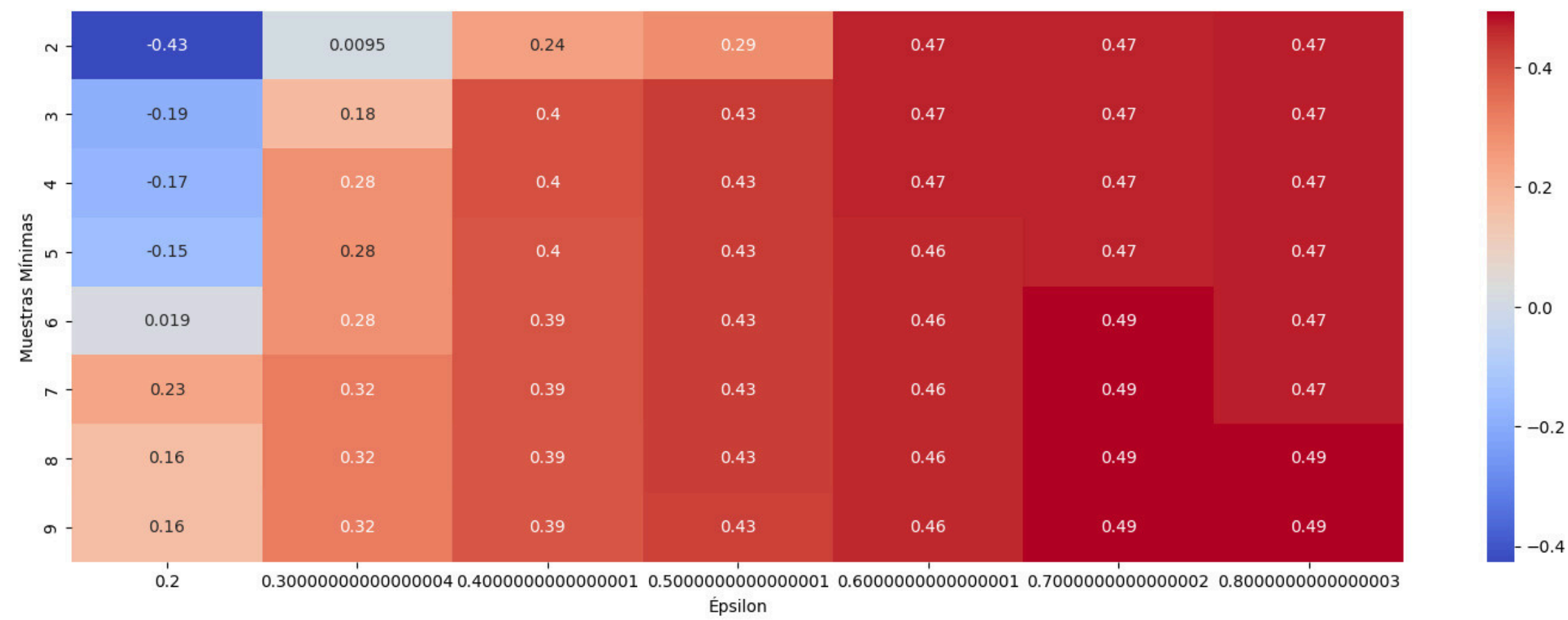


Jerarquico

Utilizando el método de agrupamiento Jerarquico se identificaron 5 grupos óptimos. Cada uno de estos grupos presentaba características distintas y relevantes que permitieron una mejor comprensión de los datos. Al analizar los patrones emergentes dentro de cada grupo, los investigadores pudieron extraer valiosas conclusiones sobre las relaciones y tendencias subyacentes.

DB – Scan

Por el método de agrupamiento DB – SCAN el modelo, utilizando nuestras mínimas de 6 y un épsilon de 0.7 se logro determinar que la mejor organización para los grupos es de 2. Esta configuración permitió identificar dos conglomerados bien definidos, maximizando la separación entre ellos y minimizando la variabilidad interna. Al analizar estos grupos, se observó una clara distinción en las características predominantes, lo cual nos brinda una visión más clara y detallada del comportamiento de los datos.

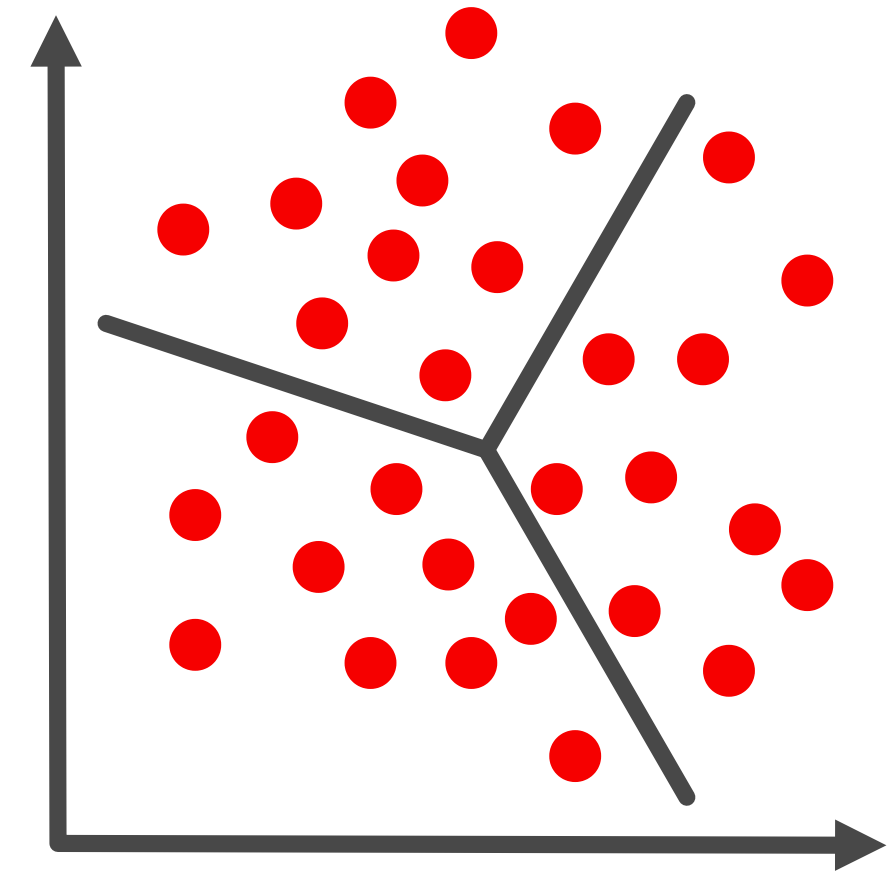


RESULTADOS – SILHOUETTE

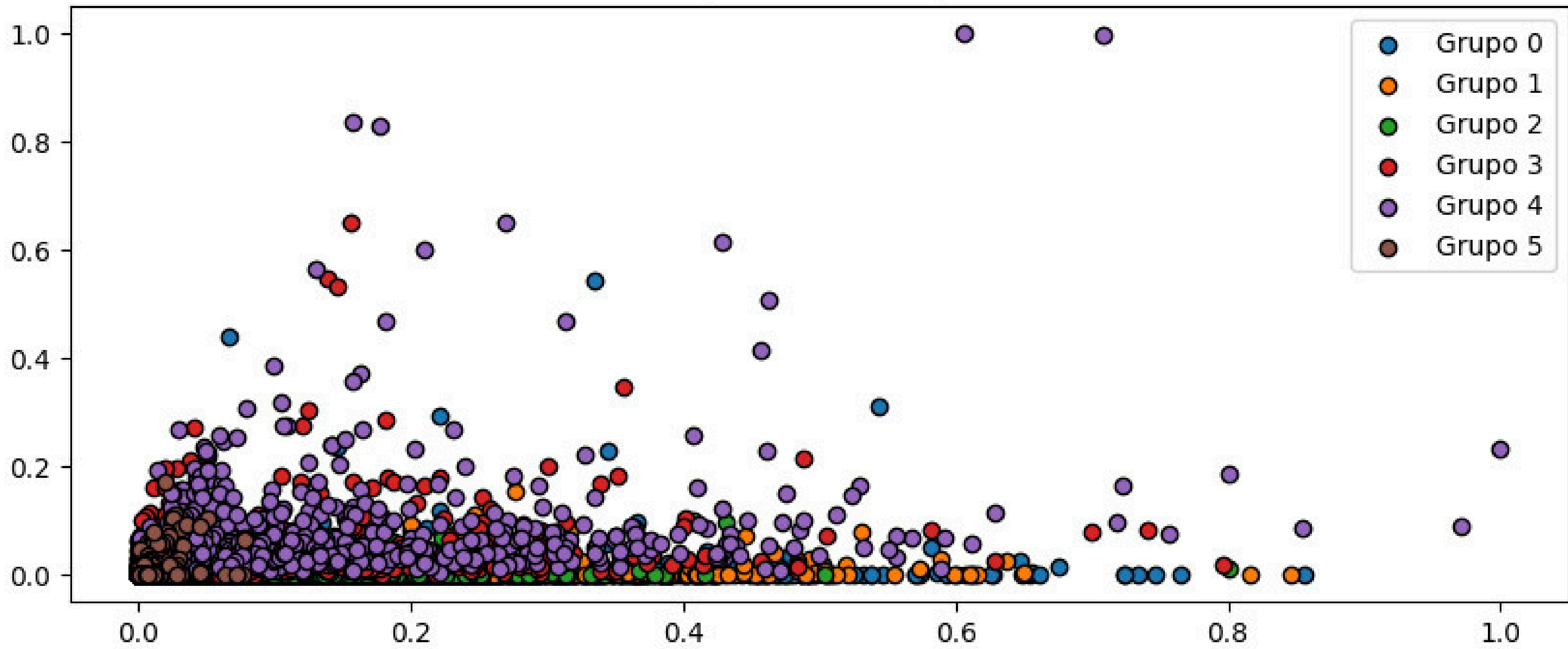
SILHOUETTE K-MEANS:
0.39655

SILHOUETTE
AGGLOMERATIVE:
0.29746

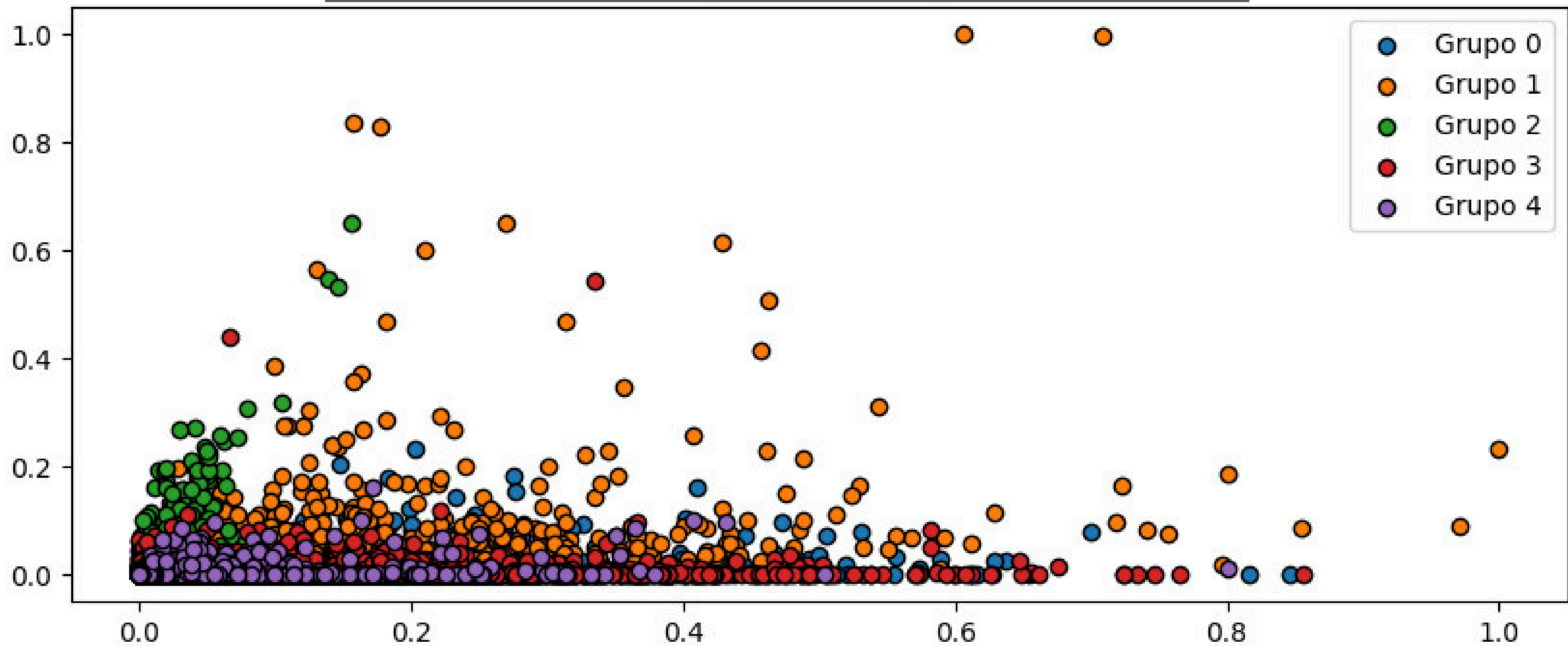
SILHOUETTE DBSCAN:
0.49412



GRUPOS OBTENIDOS POR K-MEANS

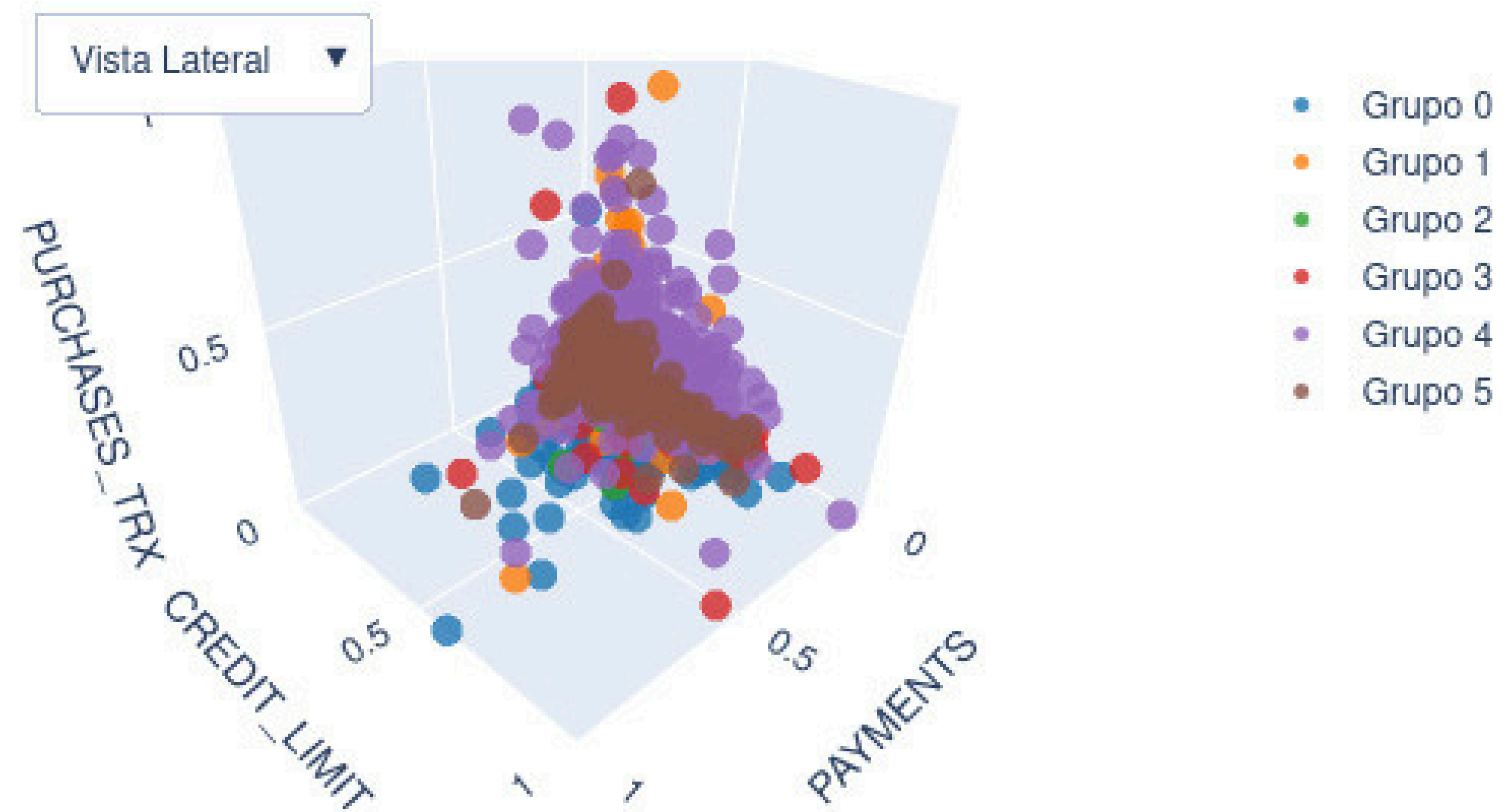


GRUPOS OBTENIDOS POR AGGLOMERATIVE



GRUPOS OBTENIDOS POR K-MEANS

Grupos K-means en 3D



GRUPOS K- MEANS

GRUPO	OBSERVACIONES	REPRESENTACION
0	4082	45.60%
1	1898	21.20%
2	612	6.83%
3	783	8.74%
4	708	7.91%
5	612	6.83%

GRUPOS JERARQUICO

GRUPO	OBSERVACIONES	REPRESENTACION
0	2938	32.82%
1	1019	11.38%
2	966	10.79%
3	3279	36.63%
4	748	8.85%

Transformación Inversa

BALANCE	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE
1812.177018	219.622882	71.657535	1349.180534
1481.278580	281.234421	774.020258	638.694340
1261.432634	268.456356	109.923611	1664.062558
1533.345571	1766.001416	148.149694	596.988660
2445.940804	2995.034738	1566.341513	689.743148
102.432040	237.134099	721.889688	77.687818

Conclusion

1. Se requieren de 6 grupos para optimizar los resultados
2. El mejor modelo es el K-Means
3. Perfil esperado

CASH_ADVANCE_FREQUENCY	PURCHASES_TRX	CREDIT_LIMIT	PAYMENTS	MINIMUM_PAYMENTS	PRC_FULL_PAYMENT	TENURE
0.182928	2.692477	4243.536195	1530.751707	915.912859	0.035007	11.876011
0.087879	21.727368	3970.063530	1474.545203	1234.158221	0.058628	11.847895
0.254136	4.305556	3011.973501	817.890383	507.405472	0.062971	7.325163
0.085314	22.205357	5789.090909	2193.908501	673.925300	0.242877	11.859694
0.092890	65.164074	7751.060820	4420.346789	931.114839	0.261100	11.915134
0.009182	15.321016	4044.290538	1290.121841	178.660257	0.817483	11.429561

BALANCE	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE
1812.177018	219.622882	71.657535	1349.180534
1481.278580	281.234421	774.020258	638.694340
1261.432634	268.456356	109.923611	1664.062558
1533.345571	1766.001416	148.149694	596.988660
2445.940804	2995.034738	1566.341513	689.743148
102.432040	237.134099	721.889688	77.687818

GRACIAS