

Fecha:  
Septiembre 2025

# MODELOS DE AGRUPACIÓN PARA CARACTERIZAR VARIEDADES DE CAFÉ A PARTIR DE ATRIBUTOS MORFOLÓGICOS DE GRANOS

Realizado por:  
Maria Alejandra Londoño, David  
Almanza

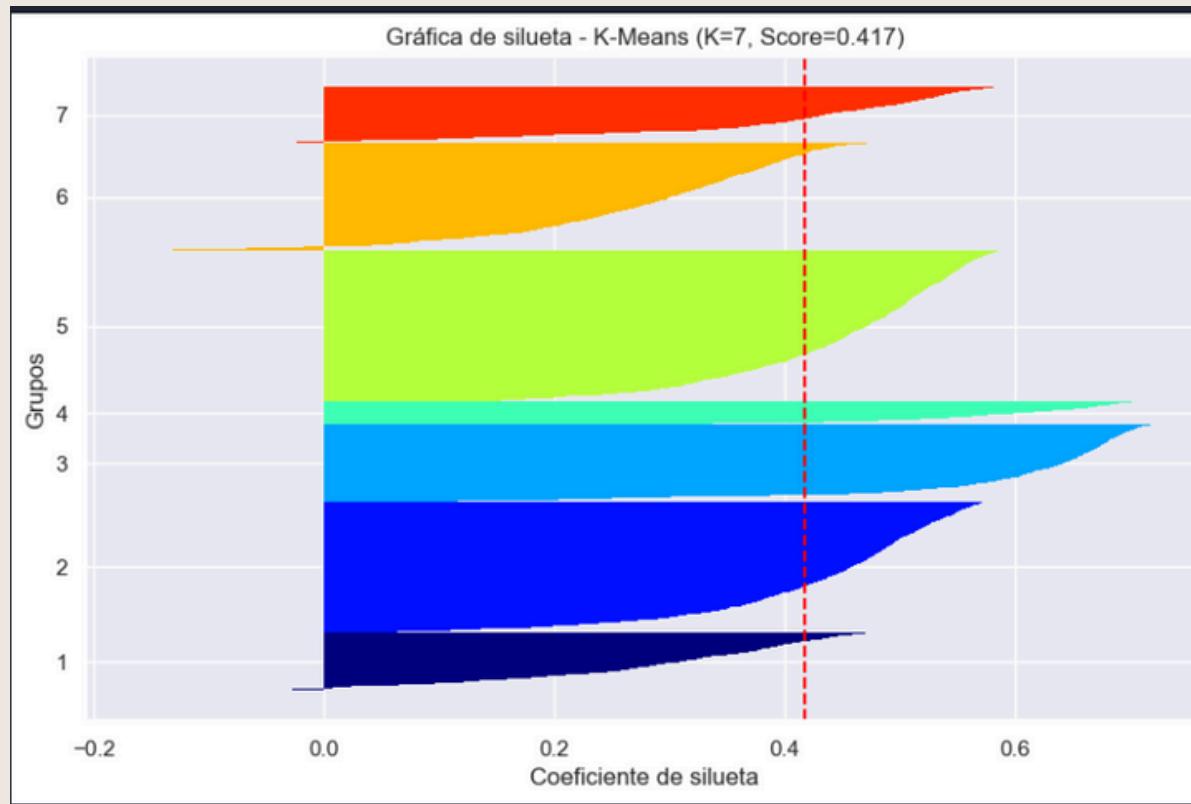
# OBJETIVOS

- Aplicar técnicas de clustering para identificar patrones morfológicos en los granos de café.
- Comparar tres algoritmos de agrupación para caracterizar variedades de café.
- Optimizar hiperparámetros para mejorar la calidad de los modelos.
- Comunicar hallazgos útiles para control de calidad, mercado y trazabilidad.

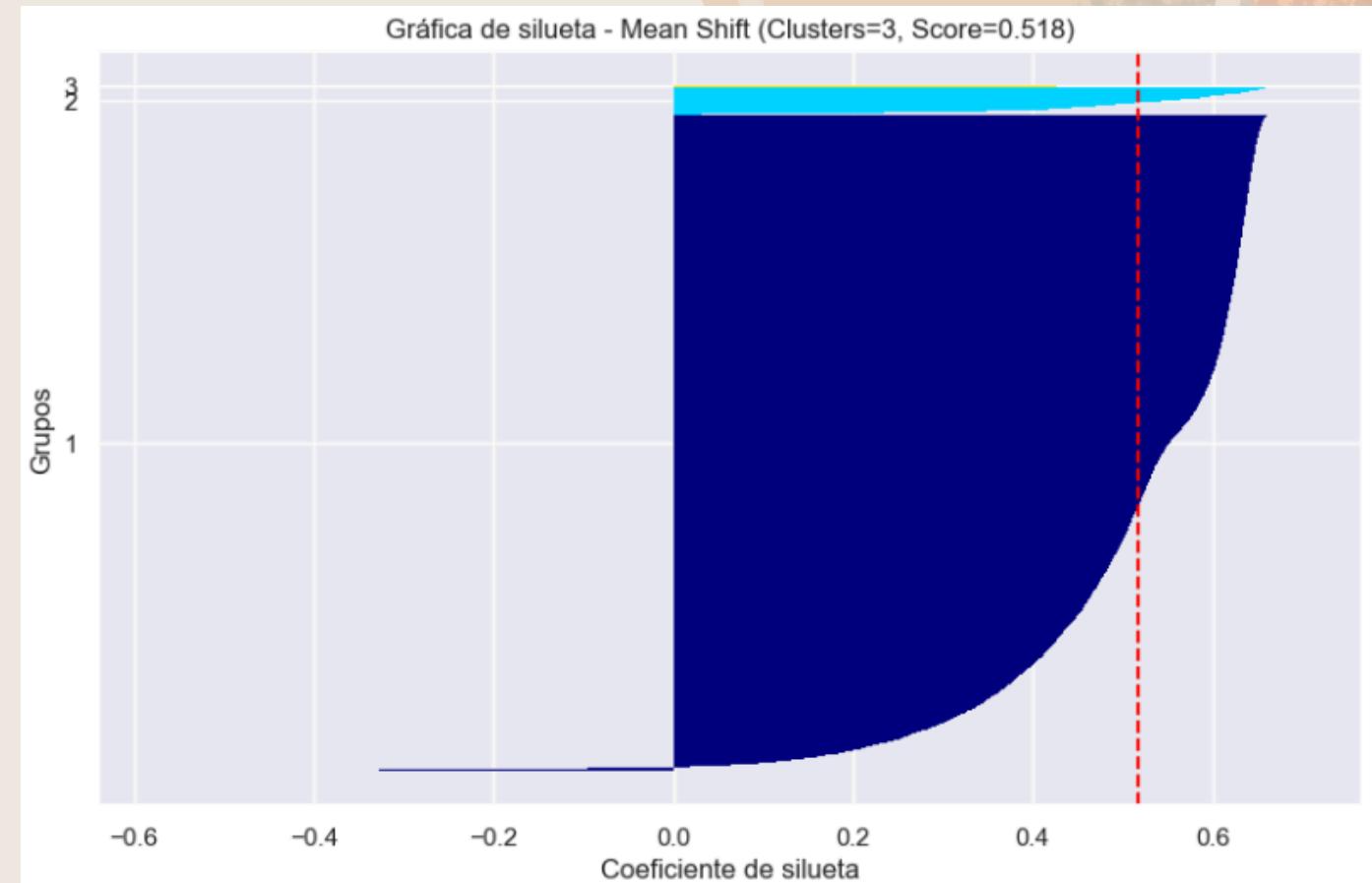
# MODELOS UTILIZADOS

- K-Means
- Mean Shift
- DBSCAN

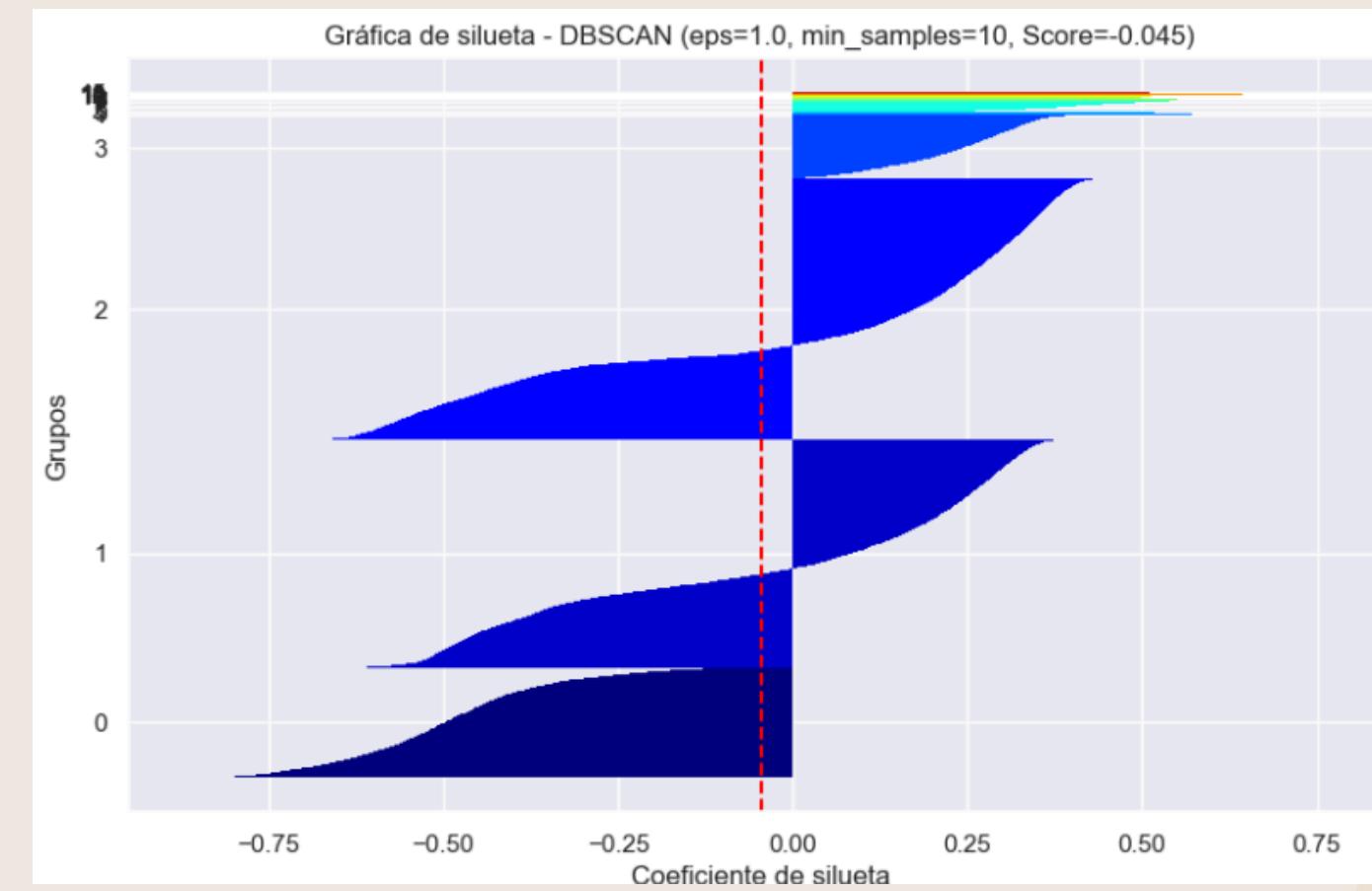
# MODELO K-MEANS



# MODELO MEAN SHIFT



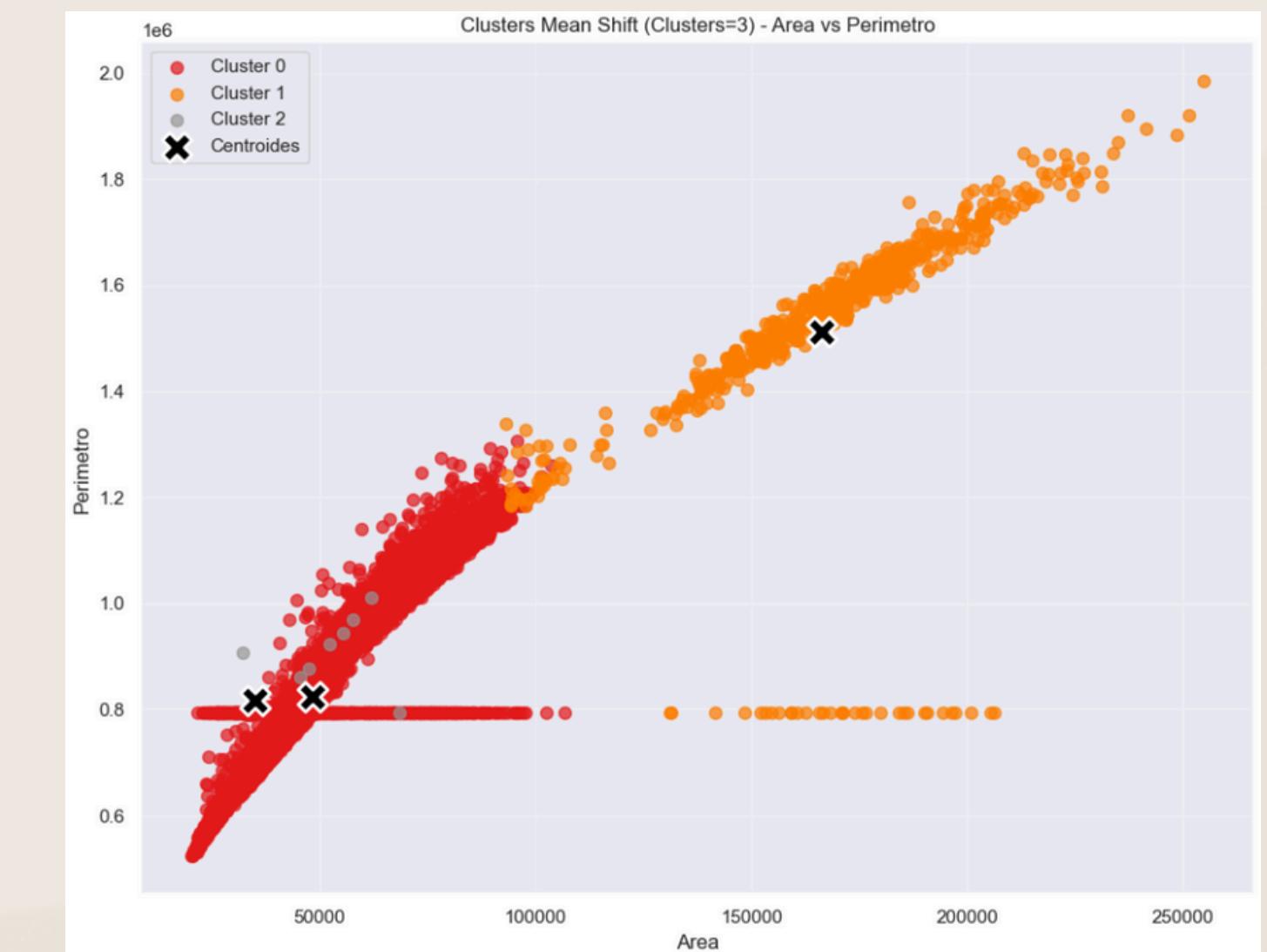
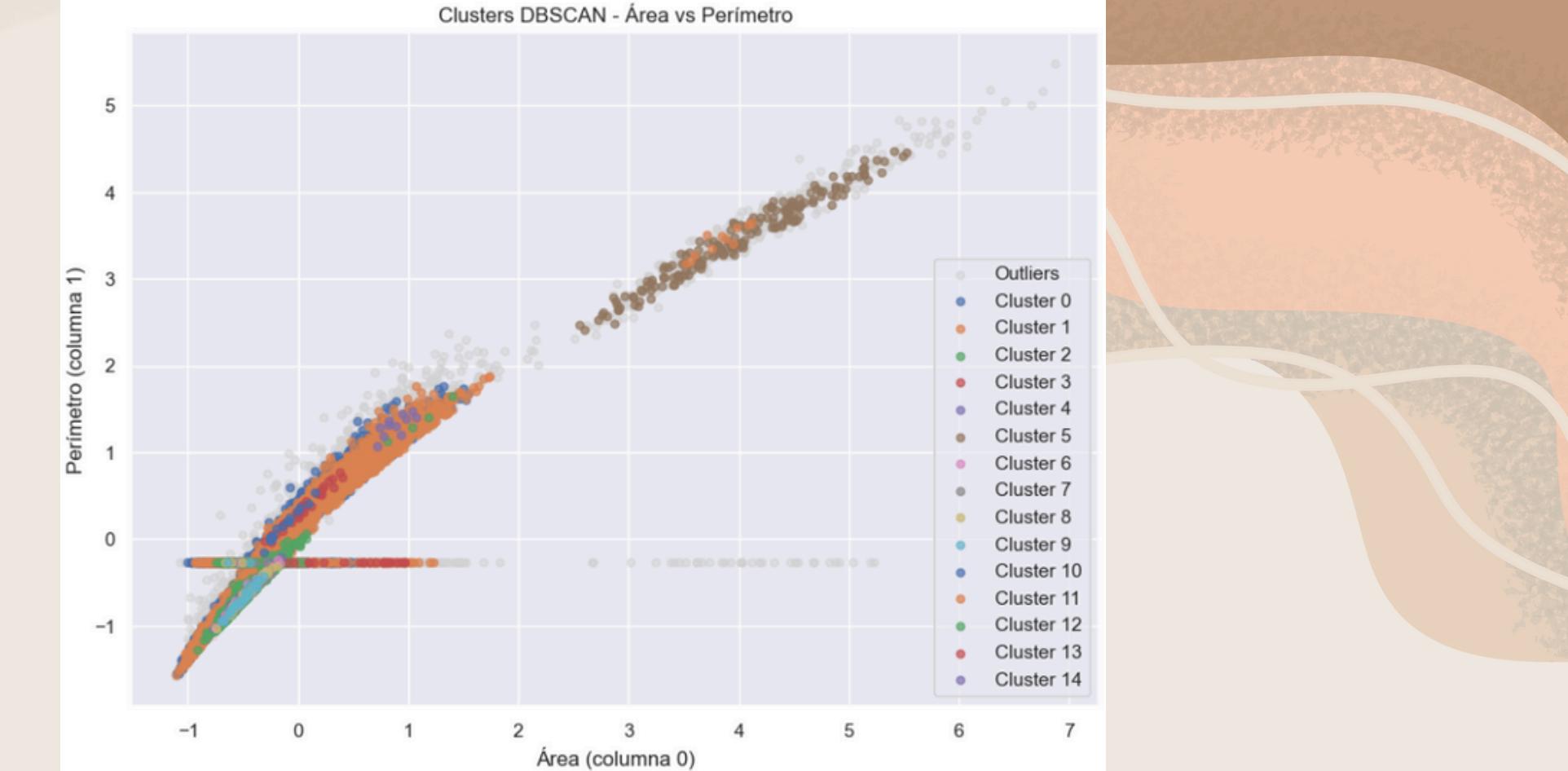
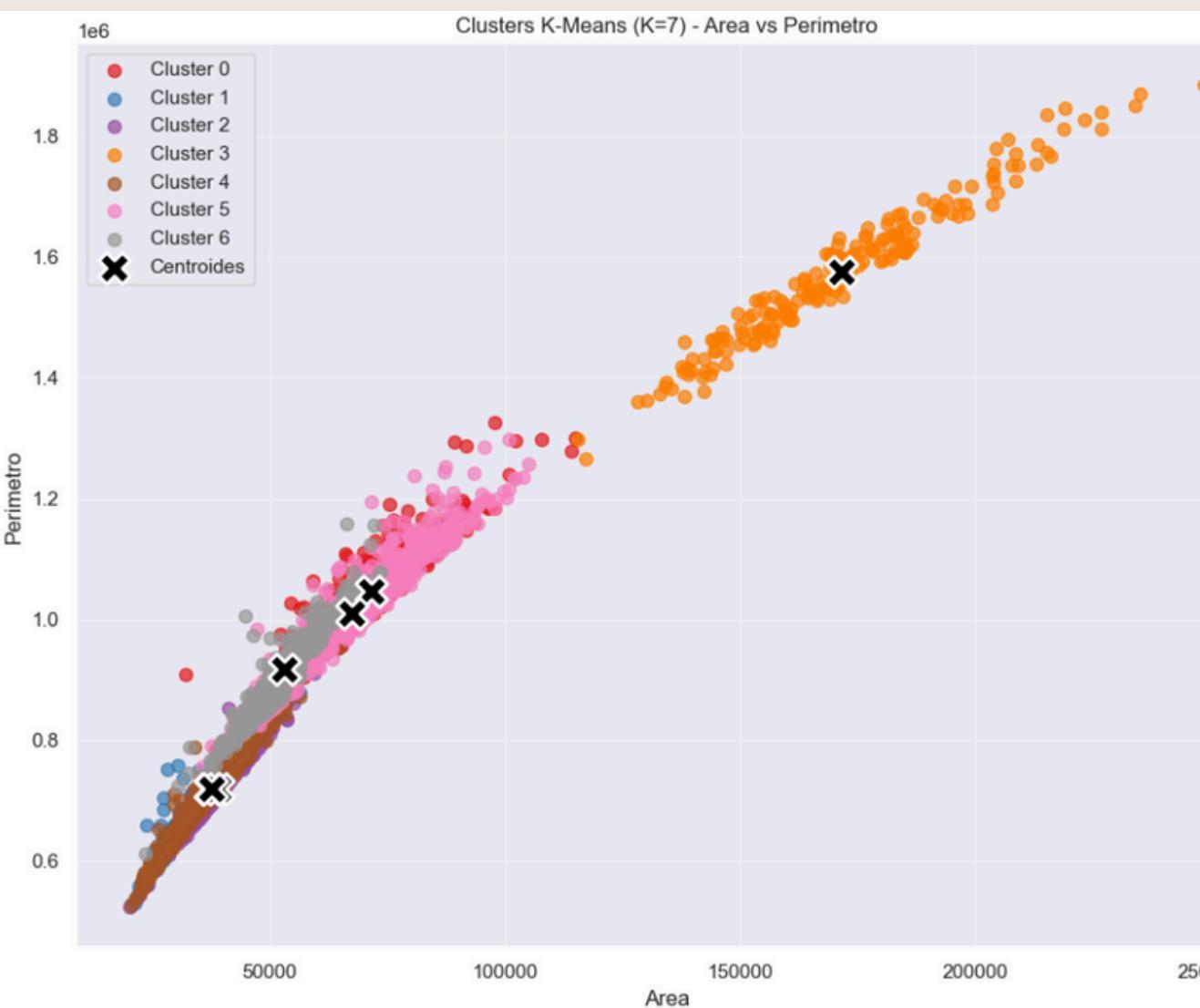
# MODELO DBSCAN



# CONCLUSIONES

- Las gráficas muestran el desempeño de tres algoritmos de clustering aplicados a los granos de café. En el caso de K-Means con  $K=7$ , el coeficiente de silueta promedio es 0.417, lo que indica una separación moderada entre grupos. Aunque hay cierta coherencia, algunos clústeres presentan solapamiento y observaciones cercanas a cero, lo que evidencia que no todos los grupos están claramente definidos.
- Por otro lado, Mean Shift con 3 clústeres alcanza un coeficiente de silueta superior (0.518), reflejando una mejor compactación y separación de los grupos. La mayoría de las muestras están bien asignadas, lo que sugiere que este método logra capturar estructuras naturales más estables en los datos, aunque con menor número de grupos que K-Means.
- En contraste, DBSCAN con los parámetros usados ( $\text{eps}=1.0$ ,  $\text{min\_samples}=10$ ) obtiene un coeficiente de silueta negativo (-0.045), lo que indica un desempeño deficiente. Los clústeres resultan difusos, con asignaciones incorrectas y ruido mal identificado, lo que sugiere que este algoritmo no es adecuado en estas condiciones o que requiere un ajuste más cuidadoso de hiperparámetros.

# VISUALIZACIÓN DE CLUSTERES



# CONCLUSIONES

- En el caso de K-Means con K=7, se observa una división más detallada de los datos, donde los clústeres siguen la tendencia lineal entre área y perímetro. Sin embargo, existe cierto solapamiento entre varios grupos (en la parte baja y media de la gráfica), lo que indica que algunos clústeres podrían no estar claramente diferenciados, aunque el método logra capturar variaciones graduales en la morfología de los granos.
- El modelo de Mean Shift con 3 clústeres genera una partición más simple y compacta, con grupos bien diferenciados en torno a las regiones principales de la relación área-perímetro. Esto refuerza la idea de que este algoritmo ofrece mayor estabilidad y separabilidad, aunque sacrifica granularidad, lo que puede ser positivo si se buscan categorías generales y robustas en lugar de una clasificación muy fragmentada.
- En contraste, DBSCAN produce una alta fragmentación con numerosos clústeres y gran cantidad de outliers, lo que dificulta una interpretación clara. Aunque es útil para detectar puntos atípicos, en este caso parece sobreajustar la distribución, generando clústeres muy pequeños y dispersos. Esto sugiere que los parámetros elegidos no fueron óptimos para capturar la estructura natural de los datos en la relación área-perímetro.

# ANÁLISIS CUANTITATIVO

- Se recomienda priorizar el uso de métricas intrínsecas como el coeficiente de silueta para comparar la calidad de los modelos. En este caso, Mean Shift (0.518) superó a K-Means (0.417) y a DBSCAN (-0.045), lo que indica que sus agrupaciones presentan mayor cohesión interna y mejor separación entre clústeres.
- DBSCAN, en las condiciones actuales, debe descartarse o reajustarse, ya que el puntaje negativo refleja una baja calidad de los grupos. Se podría explorar un ajuste fino de hiperparámetros (eps y min\_samples) antes de descartarlo totalmente, dado que este algoritmo es valioso para detectar outliers.
- En términos de número de grupos, K-Means con 7 clústeres ofrece mayor granularidad, lo que puede ser útil para análisis detallados; sin embargo, se debe evaluar si esa fragmentación añade valor real o solo complica la interpretación.

# ANÁLISIS CUALITATIVO

- Los grupos generados deben interpretarse en relación con los atributos morfológicos de los granos (área, perímetro, forma, etc.). En este sentido, Mean Shift con 3 clústeres facilita la interpretación, pues ofrece categorías claras y estables que podrían relacionarse con variedades de café o con rangos de calidad de grano.
- La empresa puede utilizar los clusters de K-Means para una clasificación más detallada cuando se requiera estandarizar procesos internos (ej. selección automática en planta), mientras que los grupos más generales de Mean Shift pueden ser útiles en estrategias de mercado (ej. mezclas comerciales diferenciadas por morfología).
- Es fundamental complementar este análisis con expertos en café dentro de la organización: validar si los clústeres identificados corresponden a categorías reconocidas en la práctica cafetera. Si los grupos no resultan comprensibles para quienes conocen el producto, habría que ajustar atributos, algoritmos o escalado de datos hasta lograr una caracterización alineada con los objetivos de trazabilidad, calidad e innovación.



**Muchas  
GRACIAS**