

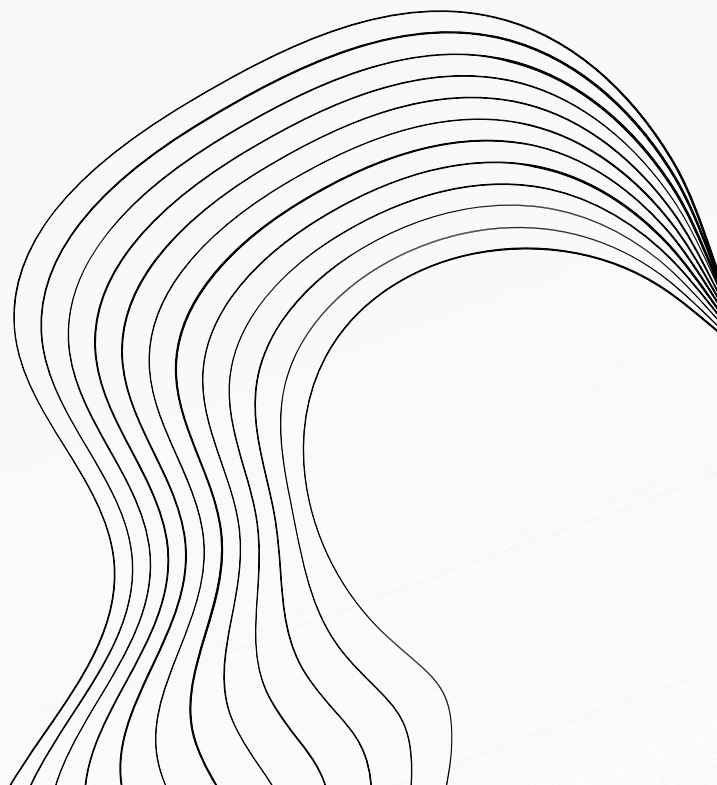
APRENDIZAJE AUTOMÁTICO PARA LA GESTIÓN DE DATOS MASIVOS

BY DR. PAULO LÓPEZ MEYE

02/07/24

TAREA 07

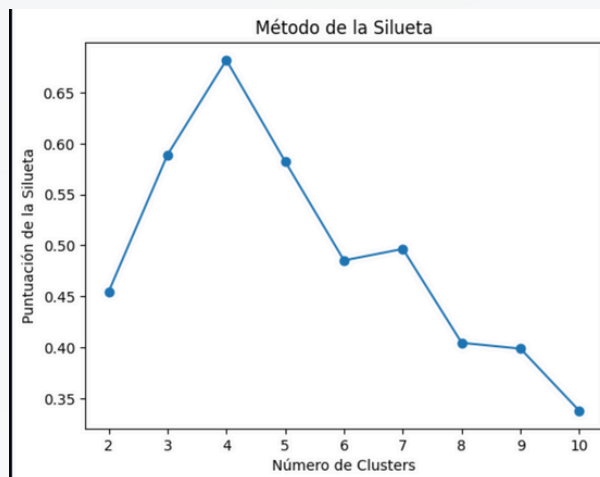
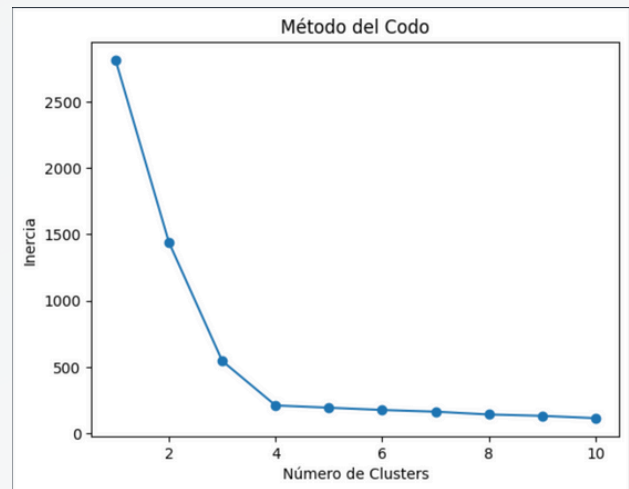
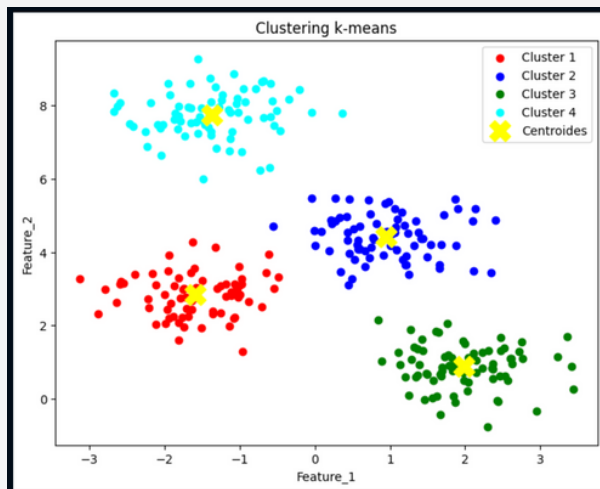
**ISAAC
MENCHACA**



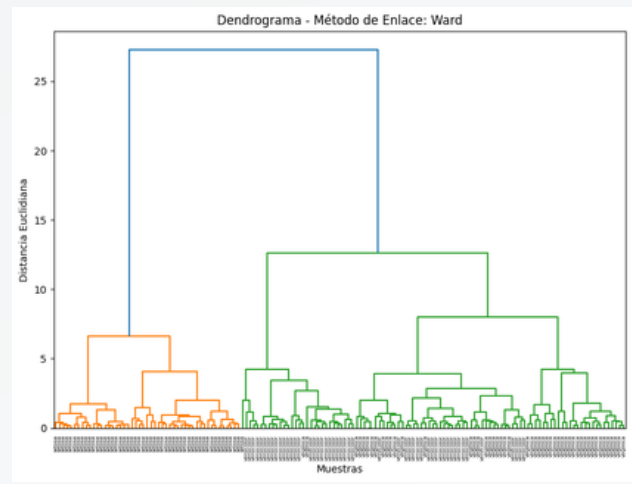
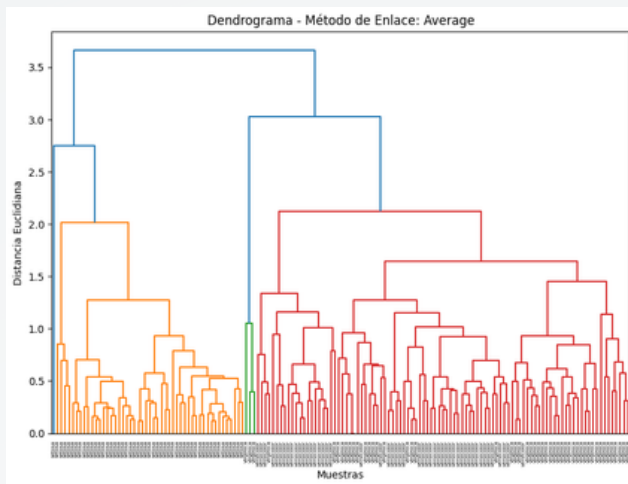
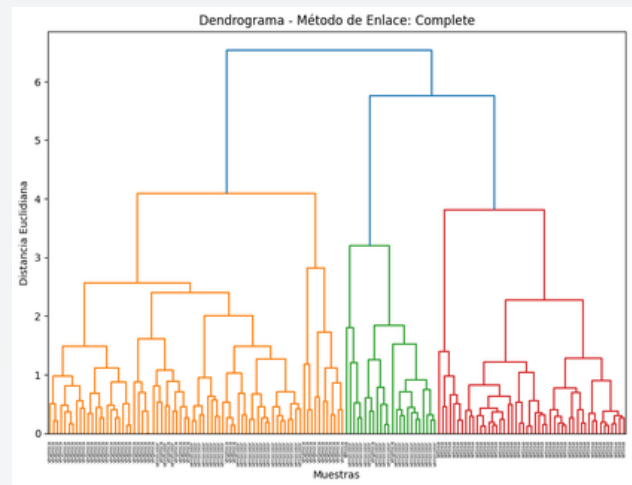
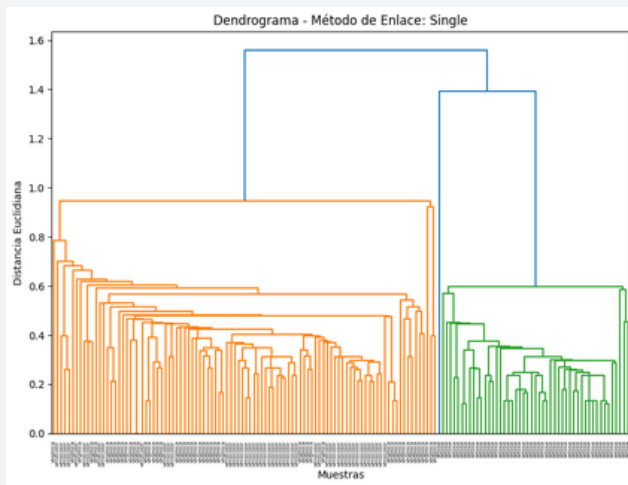
CLUSTERING

1. Ejecutar los 3 ejemplos Python de clase:

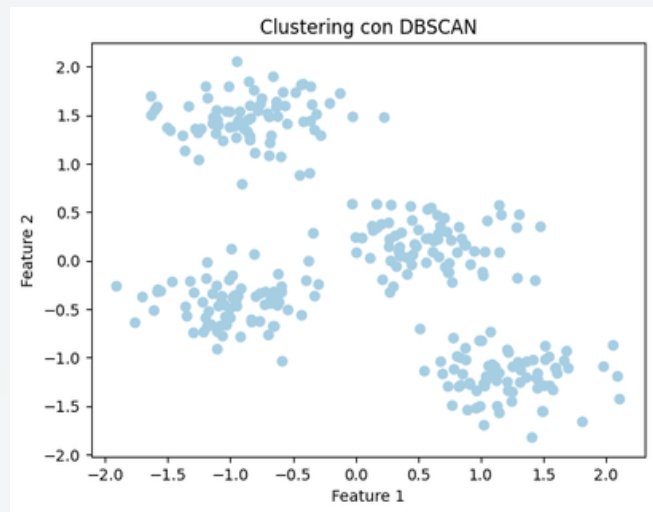
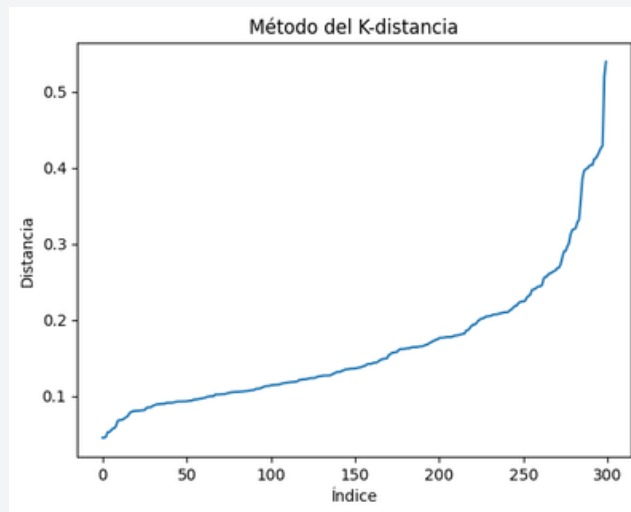
k-means



Hierarchical



DBSCAN



2. Responder las siguientes cuestiones:

- ¿Qué es el problema de la “maldición de la dimensionalidad”?
 - La "maldición de la dimensionalidad" es un término que se refiere a los problemas que surgen cuando se analizan y organizan datos en espacios de alta dimensión (muchas características o variables).
 - Para mitigar estos problemas, se utilizan técnicas como la reducción de dimensionalidad (por ejemplo, Análisis de Componentes Principales - PCA, Análisis de Discriminante Lineal - LDA), selección de características, y regularización en los modelos de aprendizaje. Estas técnicas ayudan a simplificar el modelo, mejorar su interpretabilidad y rendimiento.
- ¿Que es el PCA y para que se utiliza en clustering?
 - El Análisis de Componentes Principales (PCA, por sus siglas en inglés) es una técnica de reducción de dimensionalidad que se utiliza para transformar un conjunto de variables posiblemente correlacionadas en un conjunto más pequeño de variables no correlacionadas, llamadas componentes principales. El objetivo principal del PCA es identificar las direcciones (componentes principales) en las que los datos varían más, y proyectar los datos en un subespacio de menor dimensión manteniendo la mayor cantidad posible de varianza original.