



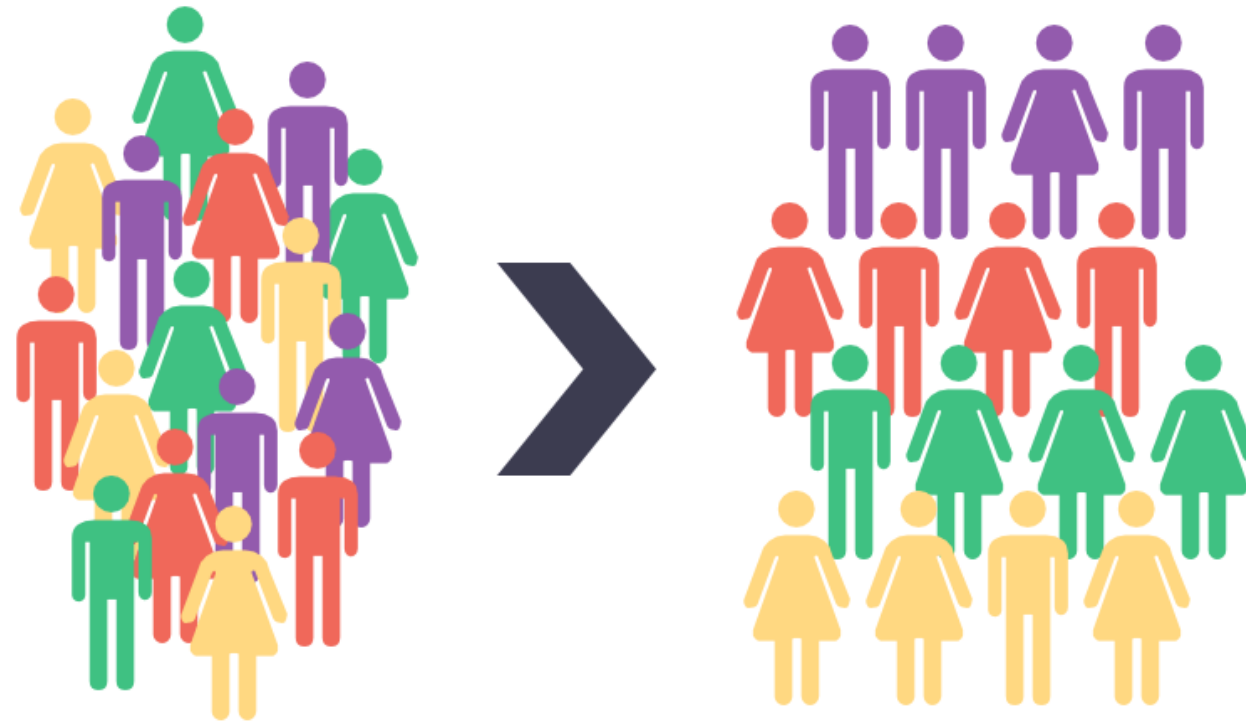
El problema  
de  
clusterización

# El problema de clusterización

---



¿Cómo encontrar segmentos en la data?

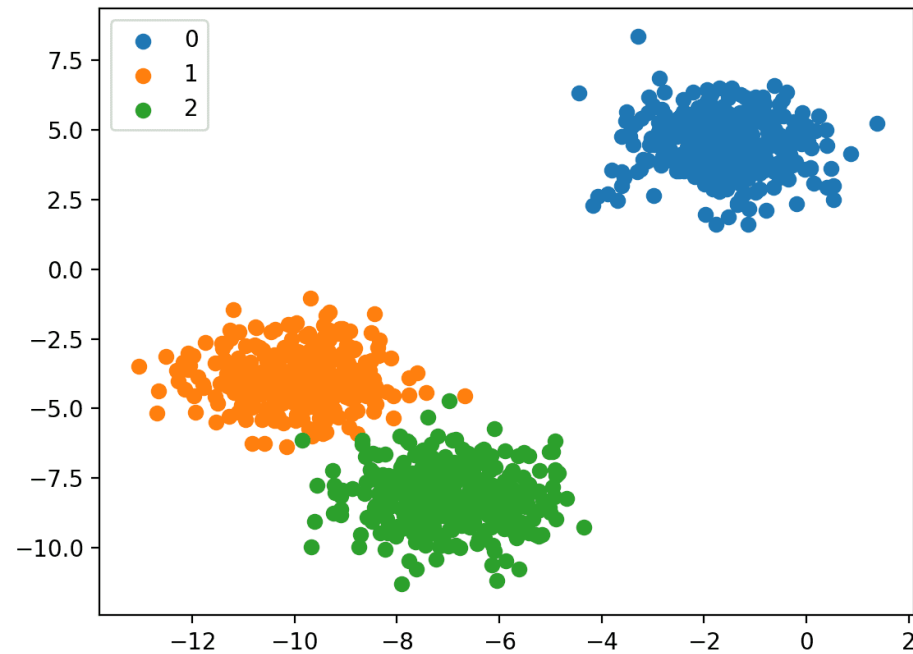


# El problema de clusterización

---



Los modelos de clusterización buscan encontrar segmentos en la data en función a diversas técnicas.

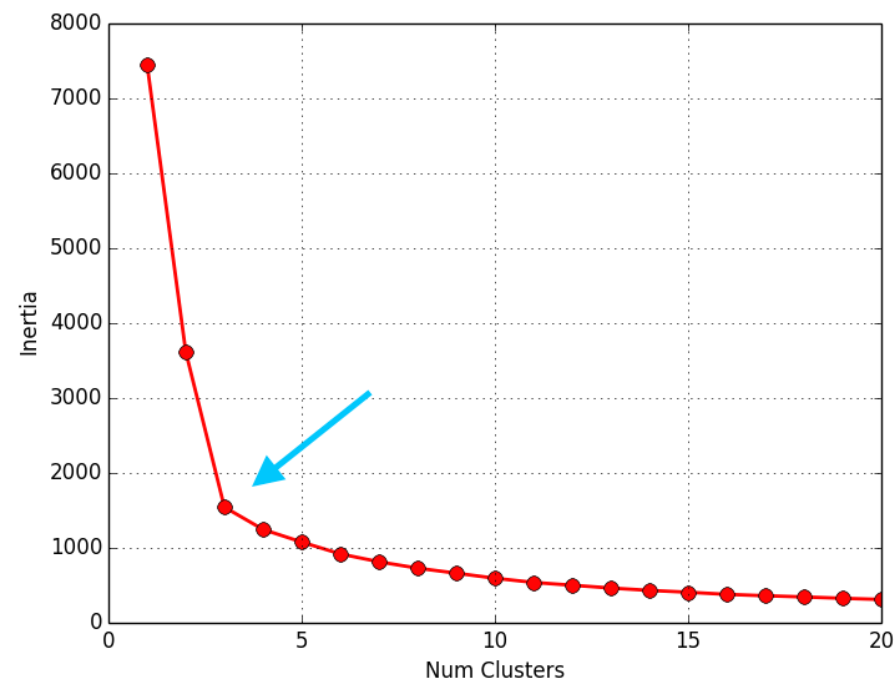


# El problema de clusterización



## ¿Cómo evaluamos el modelo?

Sino tenemos los datos reales, ¿cómo evaluamos el modelo?



# K-Means

---



Segmentar la data en K clusters basado en distancias



# Agglomerative Hierarchical Clustering

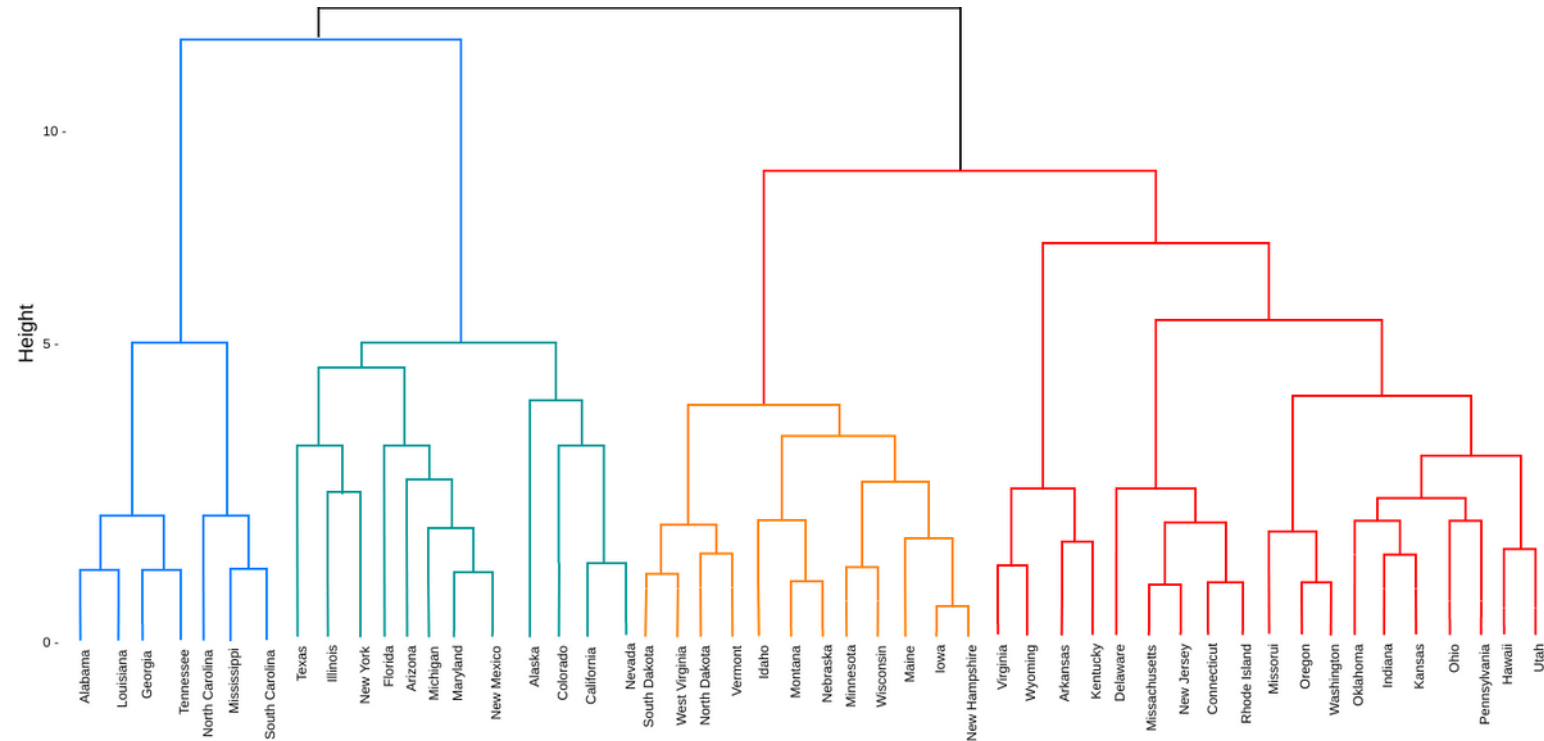


Generar N clusters

Unir los dos puntos más cercanos

Tomar los dos cluster más cercanos

Repetir hasta que solamente haya un cluster



# DBScan



Selecciona un punto  $p$ .

Agrupar todos los puntos a una distancia definida ( $\epsilon$ ) y que tengan un mínimo de elementos. Si el número de puntos es mayor al mínimo,  $p$  es un punto central, sino, itera con el siguiente punto.

Iterar hasta que todos los puntos sean clusterizados.

