

UT5 TA2

Ejercicio 2

Se utiliza Select attributes para dejar fuera los atributos no deseados a ser considerados en el modelo.

No se observan outliers pero sí datos faltantes, por lo que se decide realizar imputación de datos faltantes utilizando knn.

La normalización no sería necesaria dado que todas las escalas varían dentro del mismo rango dado que son resultados de pruebas pero como la prueba de admisión tiene como máximo 100 y las pruebas 10 sí debe normalizarse para igualar escalas de atributos.

Modelos

Configuraciones de los parámetros:

Para **DBSCAN** se utiliza un epsilon que varía desde 0,001 a 2 y mínima cantidad de puntos de 2 a 15 sin presentar variaciones en la distancia promedio y la cantidad de clusters (2 clusters, cluster_1 y residual)

Distancia promedio entre clusters: -1871.924

Para **k-means** un k de 20 y cantidad máxima de runs de 100.

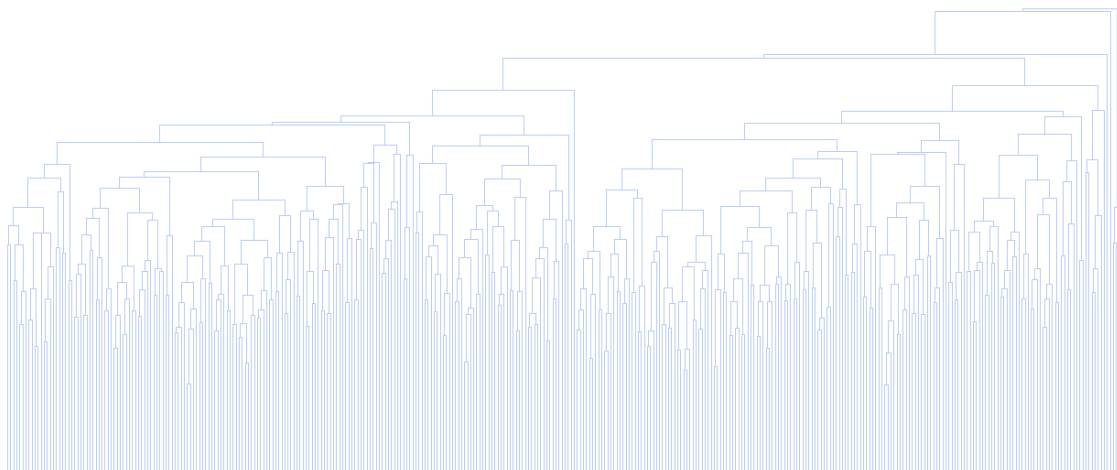
Distancia promedio de clusters: -70.916 obteniendo 19 clusters

El **Clustering aglomerativo** utiliza la distancia euclideana

Hierarchical Cluster Model Agglomerative

Número de clusters: 731

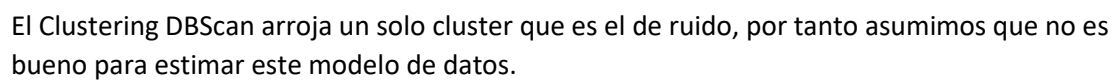
Número de items: 366



El **Clustering top-down** tiene máxima profundidad de 5 y cantidad de hojas de 3.

Número de clusters :63

Número de items: 366



El Clustering DBScan arroja un solo cluster que es el de ruido, por tanto asumimos que no es bueno para estimar este modelo de datos.