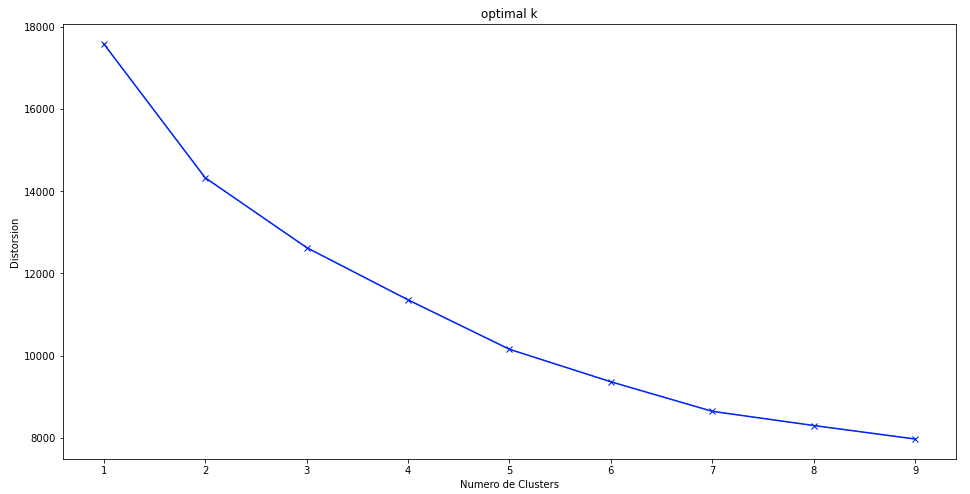
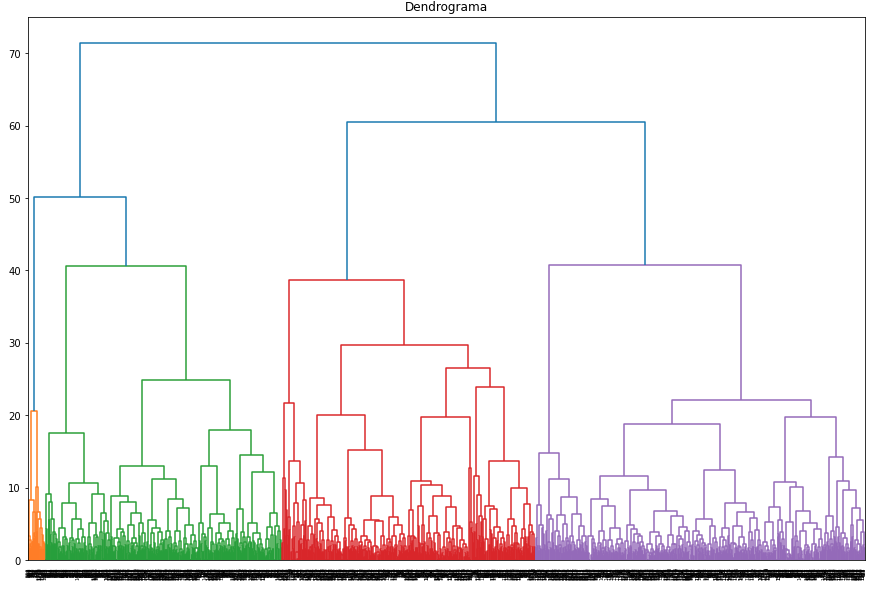
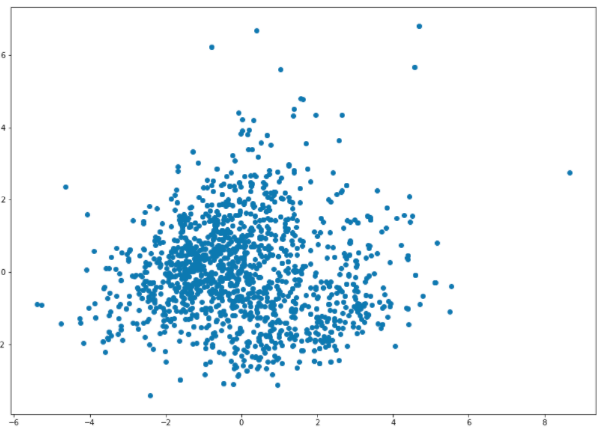
Comparativa de Resultados Entre técnica de K-Means y Agglomerative.



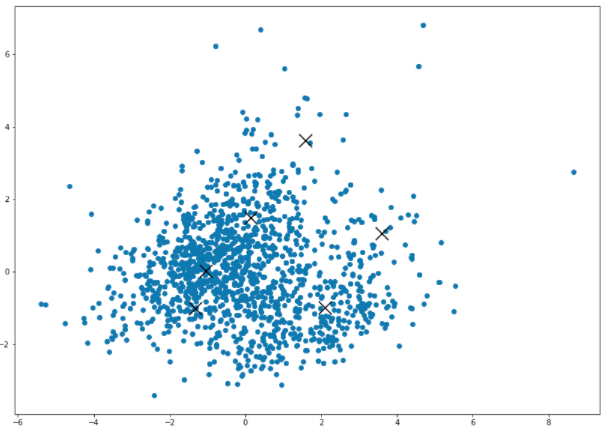


* Considerando el conjunto de datos y el objetivo definido bajo la técnica K-Means. Nuestro numero de Clusters lo teníamos definido. (6 clusters correspondientes a las 6 clasificaciones del vino).
* De igual manera se deja la representación grafica de la distorsión para hacer la comparación con la técnica Agglomerativa.
* La primera representación visual K-Means no se logra ver visualmente el punto de corte. Podríamos decir que fluctúa entre 5 y 6 clusters.
* A diferencia del agglomerativo que bajo mi interpretación detecta entre 4 y 5 clusters.

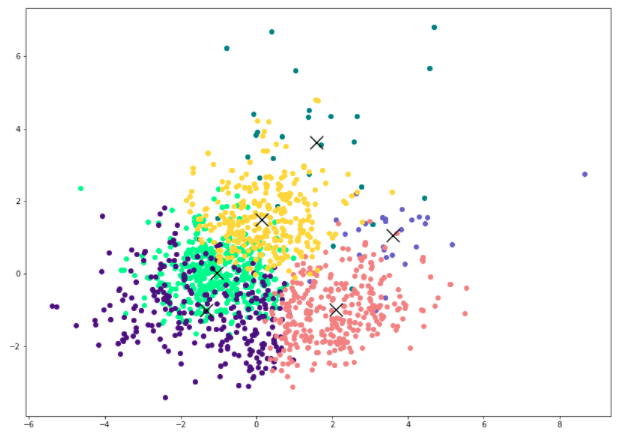
Agrupamiento y visualización K-Means



* Definiendo nuestra cantidad de Clusters, agrupamos utilizando el algoritmo de K-Means.
* Para obtener una representación visual, se reducen la cantidad de dimensiones. Aplicamos técnica PCA.
* Se pueden ver gráficamente en este diagrama todos los puntos de los datos, pero ahora solo en dos dimensiones.
* A simple vista se ve como un solo gran grupo de datos y algunos valores atípicos.

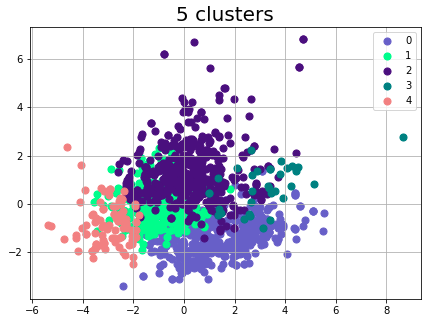


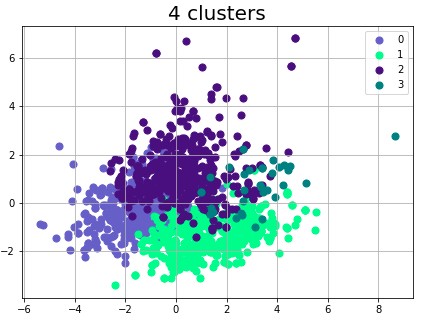
* Para poder visualizar los centroides de cada clusters, debemos reducir la dimensionalidad de estos al igual que en la comparativa anterior.
* Se reducen a dos dimensiones los centros de cada clusters.
* Se ven visualmente estos centroides, pero todavía es complejo identificar los agrupamientos.



* Agrupamos los clusters por color.
* Al ojo humano se logran identificar 6.

Agglomerativo





* Podemos observar que los clusters se encuentran demasiado juntos, esto se puede ver representado en el dendrograma.
* Es muy complejo poder visualizar de manera independiente cada agrupamiento.
* Considerando que el objetivo era el agrupamiento de las 6 clasificaciones de tipos de vino. No se logran apreciar de una manera sencilla.
* Con este método se interpretaron agrupamientos de 4 a 5 clusters.

Metricas

K-Means



Agglomerativo



* La precisión de los modelos no es del todo buena, un 60% no es considerado lo mejor…
* Esto puede ser debido a que quizás no existe una gran correlación entre las características y la calidad del vino
* En el contexto del dataset se especifica que las 11 variables corresponden a variables Fisicoquimicas. A diferencia de la calidad del vino que es una medida subjetiva. Depende netamente del catador de vino y su evaluación.

Conclusión

Según mi criterio, el algoritmo mas eficiente para agrupar las clasificaciones de vinos del conjunto de datos seleccionados corresponde a K-Means.

Fue el algoritmo que mas cercano estuvo a la cantidad de clusters (6 clasificaciones).

Mientras que el algoritmo aglomerativo predijo entre 4 y 5 clusters de clasificaciones (según mi interpretación).