

**Reconocimiento de patrones**

**Tarea 6**

Clustering 3

**Alumno:**

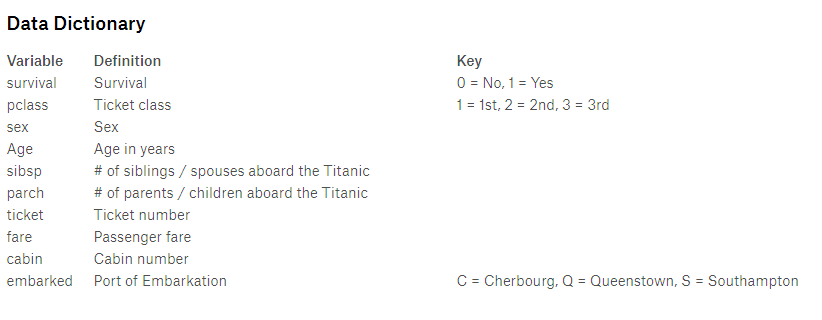
Pérez Rodríguez Raúl Francisco

**Octubre 2017**

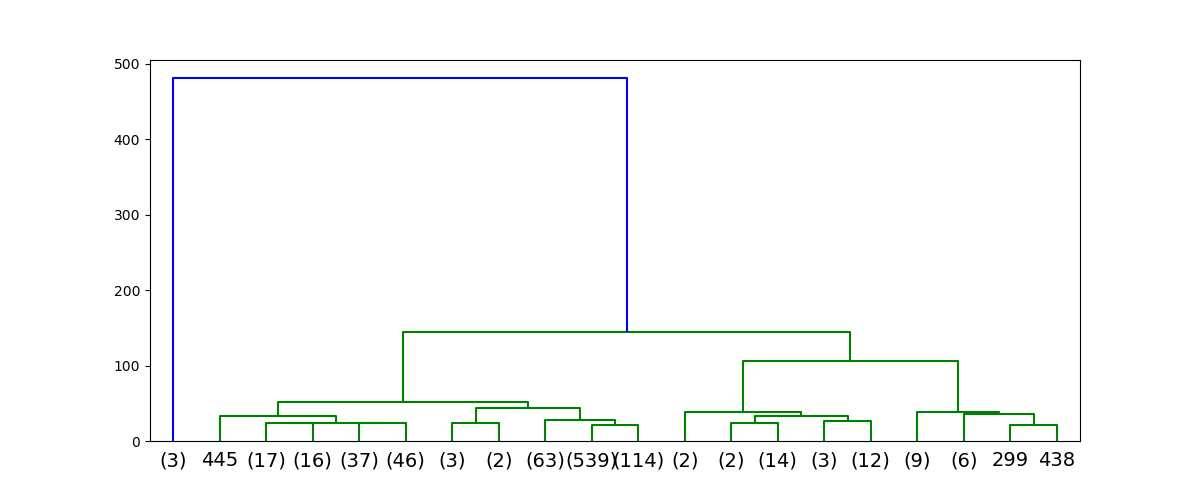
**Realice la identificación de clases por k-means de los datos de los sobrevivientes del Titanic.**

**Dendrogramas**

El conjunto de datos iniciales del Titanic, contienen 10 variables con 891 instancias. Eliminando las variables no numéricas, quedan las variables Survived ,Pclass, Age, SibSp, Parch, Fare.



Utilizando la técnica de dendrograma con el método de centroidy truncando para obtener los últimos 20 clusters obtenemos la siguiente gráfica.



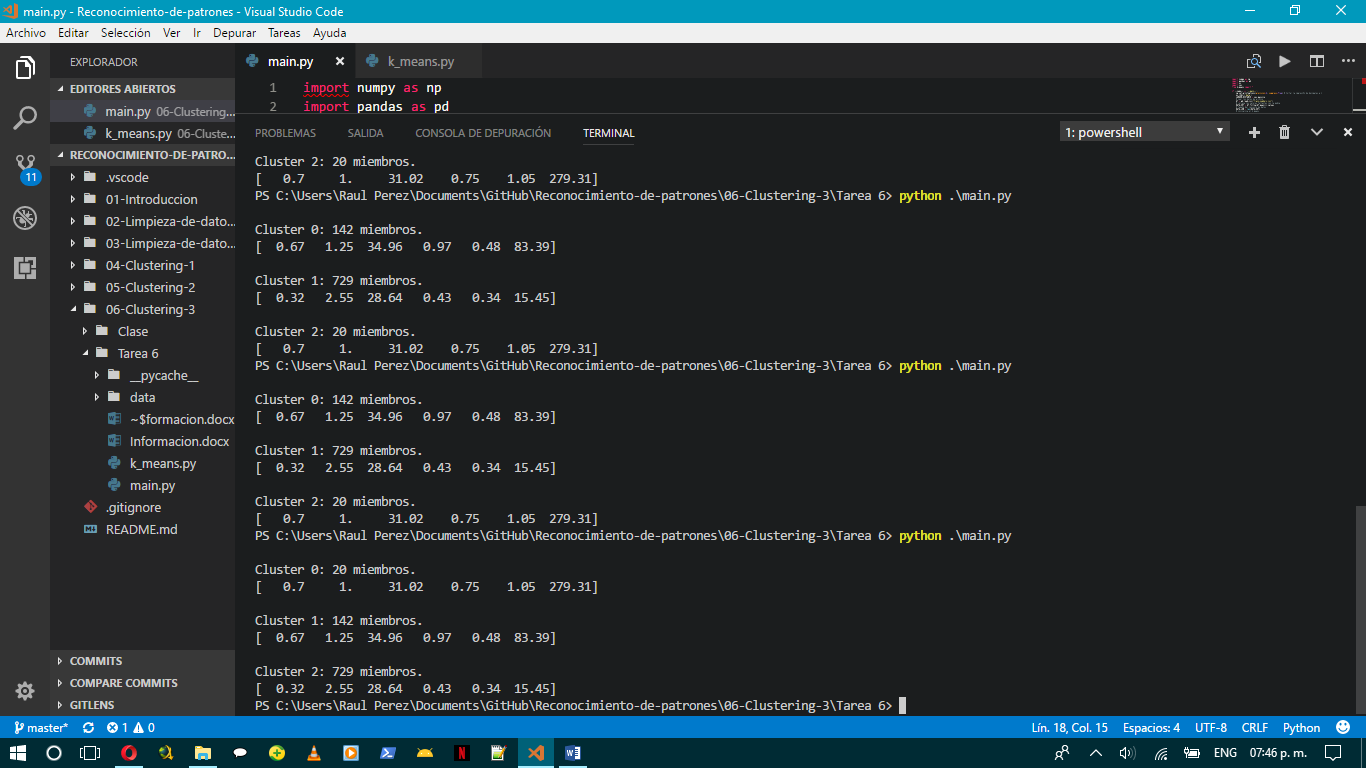
Analizando los últimos 20 clusters del dendrograma se pueden observar que puede haber entre 3 y 5 grupos.

**K-Means**

Utilizando el dendrograma como una primera prueba, se observaron la formación de entre 3 y 5 grupos de datos. Con estos resultados se utilizara el k-means con k igual a 3, 4 y 5 para observar la formación de los grupos con una selección aleatoria de los clusters iniciales.

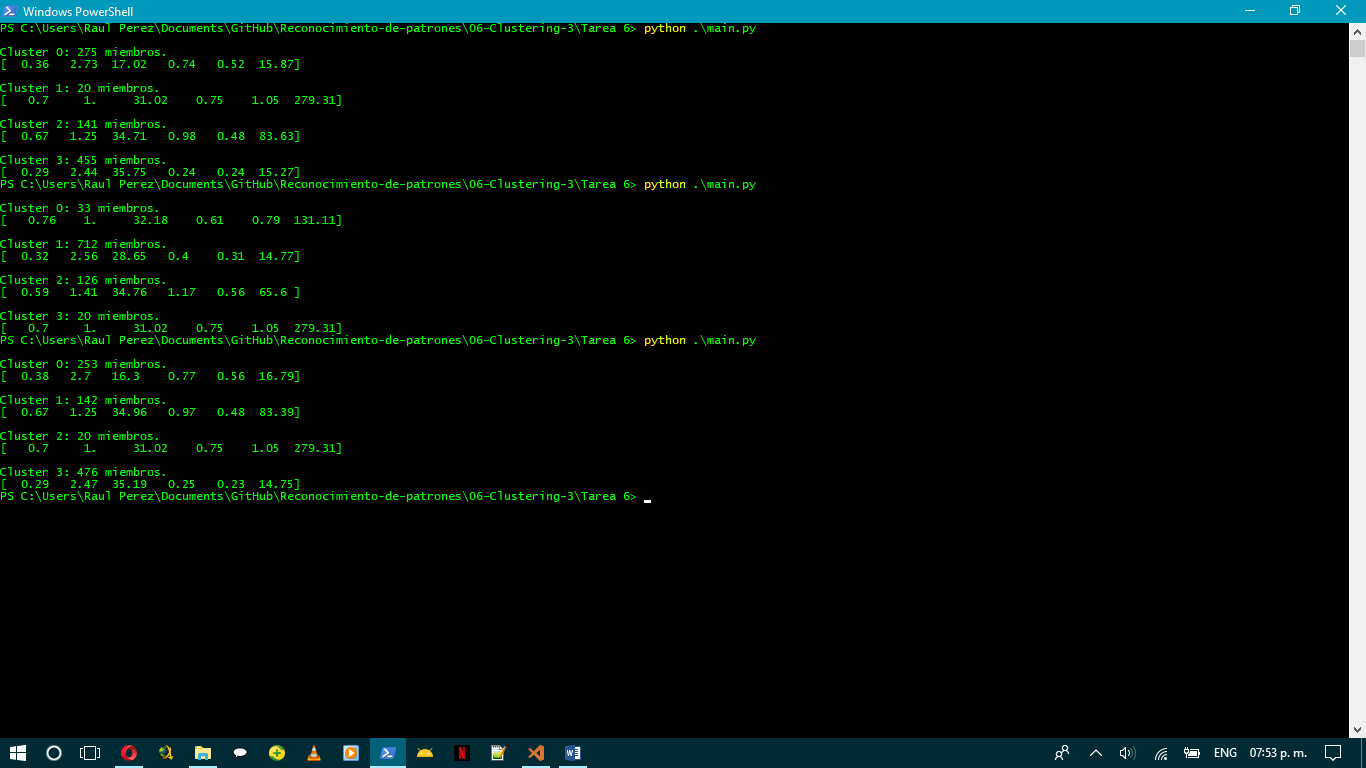
**Con k igual a 3**

Realizando 3 corridas con el k-means con k igual a 3, se obtuvieron los mismos clusters



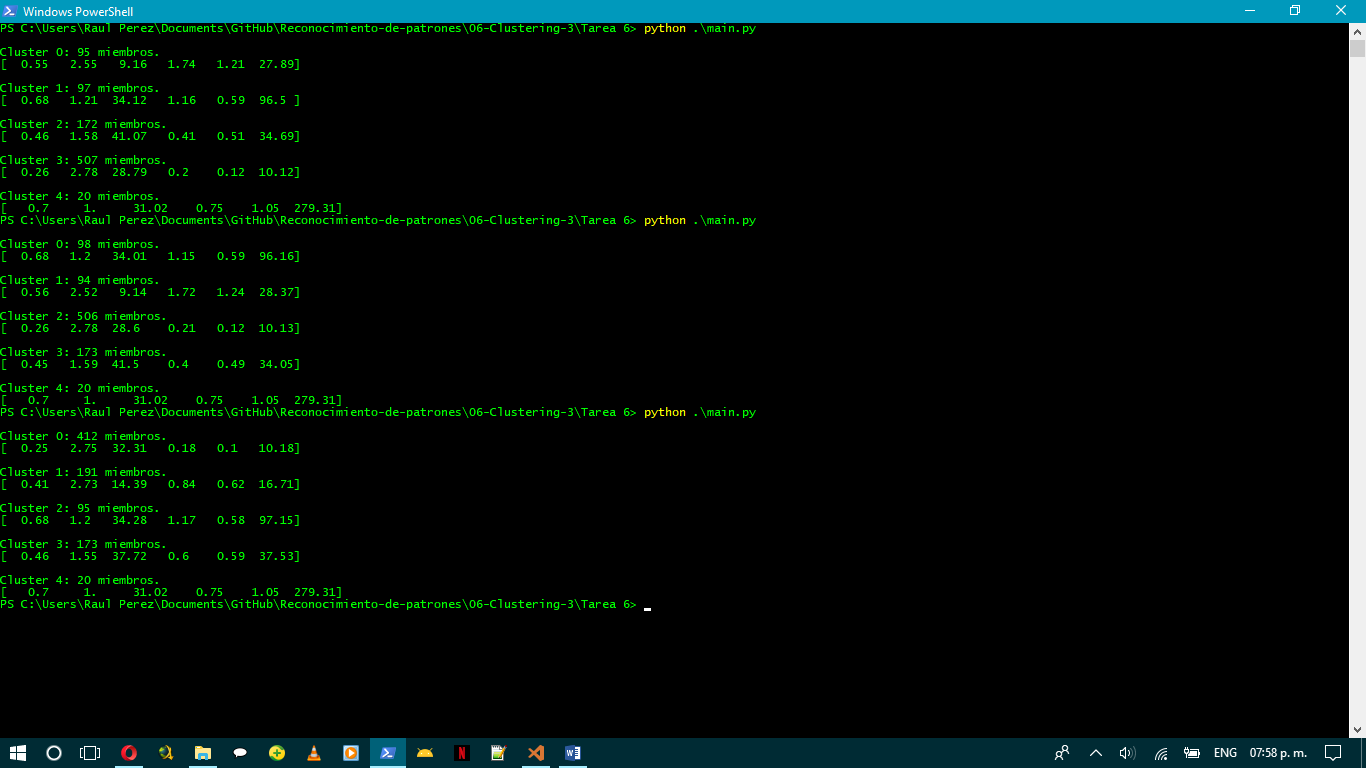
**Con k igual a 4**

Realizando 3 corridas con el k-means con k igual a 4, se obtuvieron clusters diferentes excepto el cluster con 20 elementos, el cual se repite con k igual a 3.



**Con k igual a 5**

Realizando 3 corridas con el k-means con k igual a 5, se obtuvieron clusters diferentes excepto el cluster con 20 elementos, el cual se repite con k igual a 3. Pero a comparación con k igual a 4, en este caso el número de miembros por cluster no es tan diferente.



**Conclusión**

Analizando los resultados al utilizar la técnica de k-means con los valores de k iguales a 3, 4 y 5. Se concluye que es más probable que haya 3 grupos de datos en el conjunto de datos.