

Clustering Herarchy

Importo las librerias

```
In [15]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import preprocessing
import scipy.cluster.hierarchy as shc
from sklearn.cluster import AgglomerativeClustering
```

Carga de Datos

```
In [4]: df = df = pd.read_csv('./Wholesale customers data.csv')
df=df.dropna()
df.sample(10)
```

```
Out[4]:
```

	Channel	Region	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicassen
52	2	3	40721	3916	5876	532	2587	1278
124	1	3	36050	1642	2961	4787	500	1621
311	1	2	29635	2335	8280	3046	371	117
291	1	3	6022	3354	3261	2507	212	686
355	1	3	190	727	2012	245	184	127
207	2	1	2541	4737	6089	2946	5316	120
143	1	3	10405	1596	1096	8425	399	318
164	2	3	5224	7603	8584	2540	3674	238
182	1	3	694	8533	10518	443	6907	156
99	1	3	9658	2182	1909	5639	215	323

Datos de Muestra

```
In [7]: indices=[182,99,250]
df = df.drop(['Region', 'Channel'],axis=1)
muestra = pd.DataFrame(df.loc[indices],columns=df.keys()).reset_index(drop=True)
```

Procesamiento de Datos

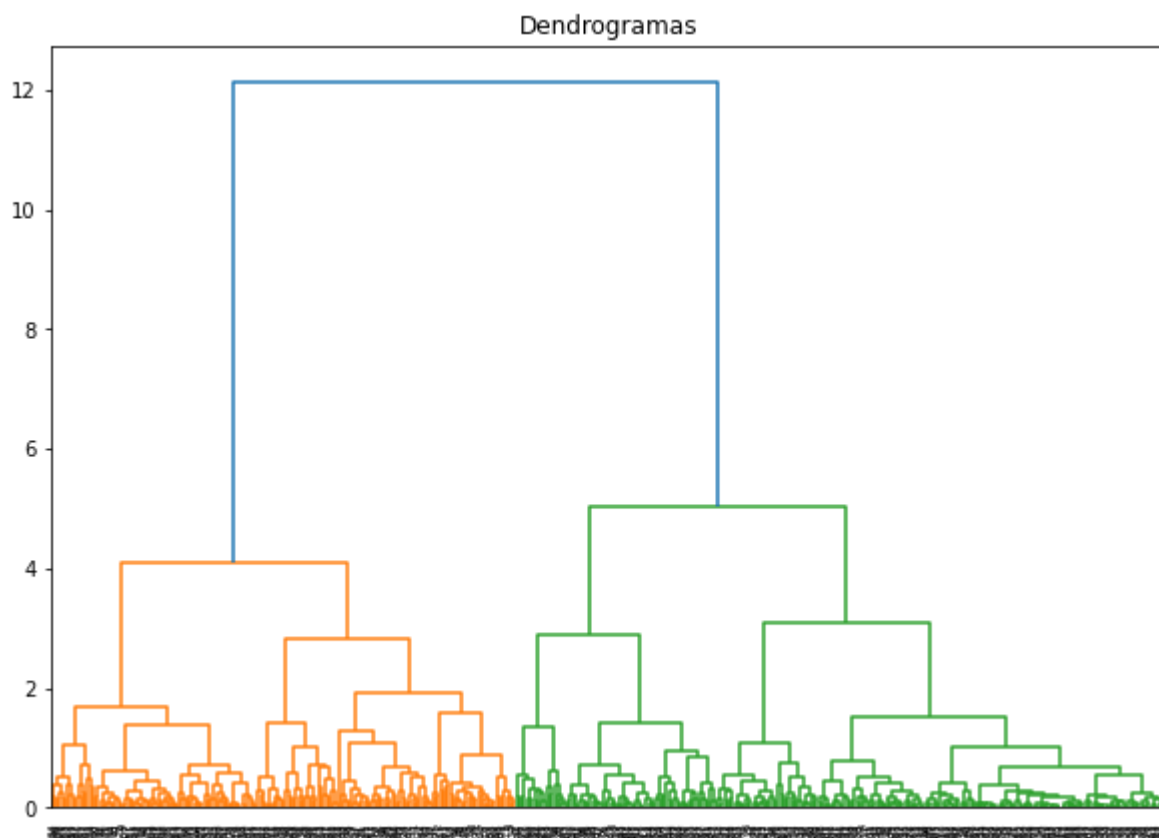
```
In [8]: df_escalada= preprocessing.Normalizer().fit_transform(df)
muestra_escalada= preprocessing.Normalizer().fit_transform(muestra)
```

Analisis y Grafica para obtener el numero de Clusters

```
In [9]: x= df_escalada
```

```
In [14]: plt.figure(figsize=(10,7))
```

```
plt.title('Dendrogramas')
dendrograma= shc.dendrogram(shc.linkage(x,method='ward'))
```



Aplicacion del algoritmo

```
In [19]: algoritmo= AgglomerativeClustering(n_clusters=3,affinity='euclidean',linkage='ward')
          algoritmo.fit(x)
          pred=algoritmo.fit_predict(x)
```

Verificamos con las muestras

```
In [20]: muestra_prediccion= algoritmo.fit_predict(muestra_escalada)
```

```
In [21]: for i, pred in enumerate(muestra_prediccion):
          print('Muestra',i,' se encuentra en el cluster: ',pred)
```

```
Muestra 0 se encuentra en el cluster: 1
Muestra 1 se encuentra en el cluster: 2
Muestra 2 se encuentra en el cluster: 0
```

```
In [ ]:
```