```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

dataset = pd.read\_excel('Mall\_Customers.xlsx')
dataset

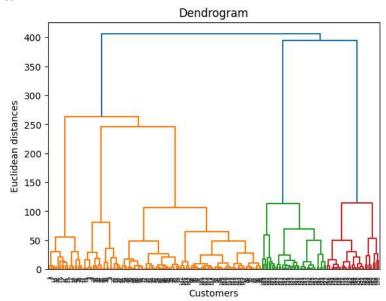
<del></del>		CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
	0	1	Male	19	15	39
	1	2	Male	21	15	81
	2	3	Female	20	16	6
	3	4	Female	23	16	77
	4	5	Female	31	17	40
			•••			
	195	196	Female	35	120	79
	196	197	Female	45	126	28
	197	198	Male	32	126	74
	198	199	Male	32	137	18
	199	200	Male	30	137	83
:	200 rows × 5 columns					

x =dataset.iloc[:,[3,4]].values

```
# In kmeans clustrting were were fining the optimal number of clusters ,
```

```
import scipy.cluster.hierarchy as sch
dendrogram = sch.dendrogram(sch.linkage(x,method = 'ward'))
plt.title('Dendrogram')
plt.xlabel('Customers')
plt.ylabel('Euclidean distances')
plt.plot()
```

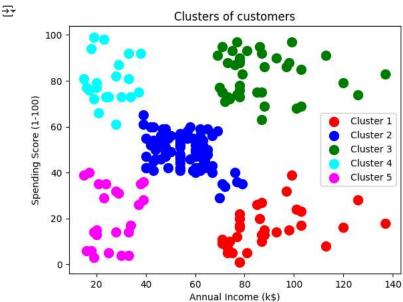
<del>\_</del> []



from sklearn.cluster import AgglomerativeClustering

<sup>#</sup> but in hirarchical clusering we don't need to find this , instade we will use Dendrogram

```
hc = AgglomerativeClustering(n_clusters = 5, linkage='ward')
y_hc = hc.fit_predict(x)
print(y_hc)
  0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2\ 0\ 2
                 202020202020202021
plt.scatter(x[y_hc == 0, 0], x[y_hc == 0, 1], s = 100, c = 'red', label = 'Cluster 1')
plt.scatter(x[y_hc == 1, 0], x[y_hc == 1, 1], s = 100, c = 'blue', label = 'Cluster 2')
plt.scatter(x[y_hc == 2, 0], x[y_hc == 2, 1], s = 100, c = 'green', label = 'Cluster 3')
plt.scatter(x[y_hc == 3, 0], x[y_hc == 3, 1], s = 100, c = 'cyan', label = 'Cluster 4')
plt.scatter(x[y_hc == 4, 0], x[y_hc == 4, 1], s = 100, c = 'magenta', label = 'Cluster 5')
plt.title('Clusters of customers')
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score (1-100)')
plt.legend()
plt.show()
```



Start coding or generate with AI.