

hierarchical-clustering-1

August 28, 2023

0.1 Name : B.Rishitha

0.2 Roll No : 21X05A6710

0.3 Branch : Data Science

0.4 College : Narasimha Reddy Engineering College UGC Autonomous

1 Project Title :

1.0.1 Analysis and prediction of “Mall_Customers.csv” of American mall Market called as phonix mall, find out on the basis of clients requirements of dendrogram using scipy graphics library with the help of “scipy.cluster.hierarchy”, to ace the no. of linkage of clustering to predict.

2 Problem Statement :

2.0.1 The American finance market clients as per the rate of GDP of 2011 found as highest number of growth in their business market.

2.0.2 As data science engineer find out which hierarchy cluster gives maximum linkage in upcoming future.

3 Task-1 :

3.0.1 With the help of numpy library import the library and import the dataset.

4 Task-2 :

4.0.1 Usin the dendrogram to find the optimum number of clusters.

5 Task-3 :

5.0.1 Create a hierarchy model and visualize the cluster with the help of matplotlib library.

6 importing the libraries

```
[1]: #Import the numpy, pandas , matplotlib, seaborn libery's
import numpy as np
import pandas as pd
```

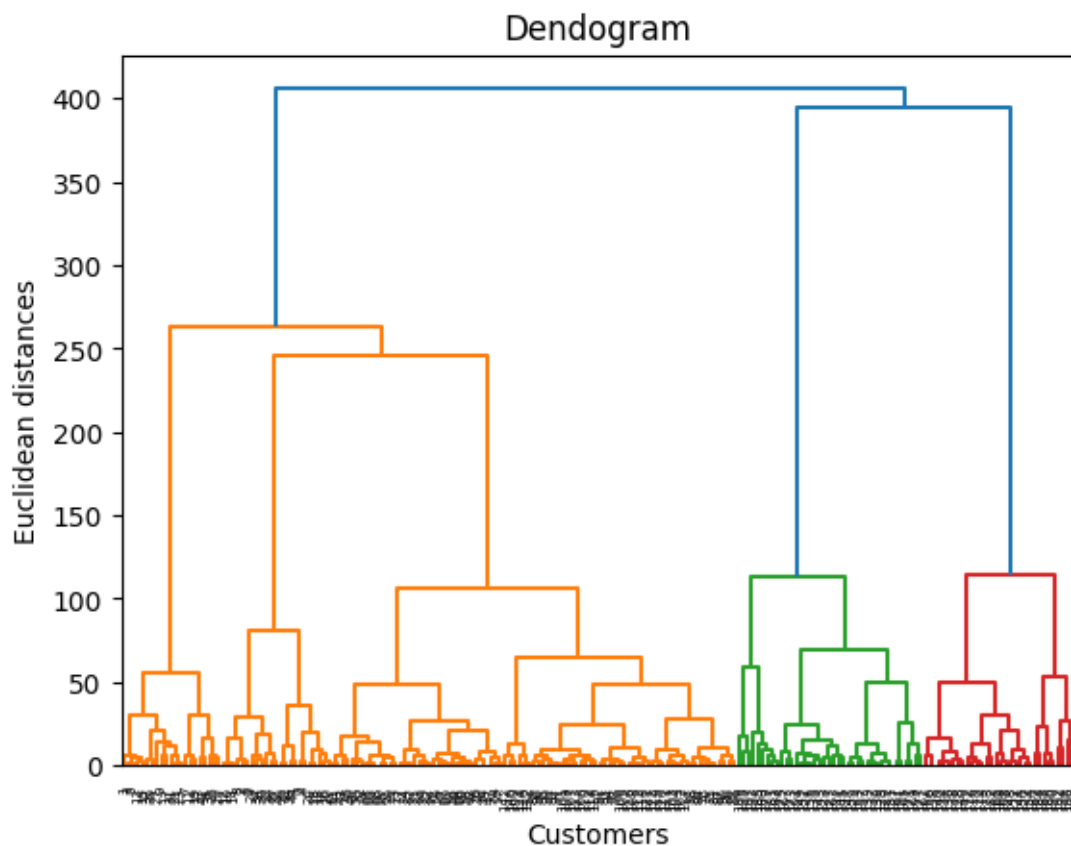
```
import matplotlib.pyplot as plt
import seaborn as sns
```

7 loading the dataset

```
[2]: #Assign variable name "dataset" and the input variable as "X" including select
      ↪ all the row and index columns which you want [column_index, Column_index].
dataset = pd.read_csv("Mall_Customers.csv")
x = dataset.iloc[:,[3,4]].values
```

8 using the dendrogram to find the optimal number of clusters

```
[3]: #import scipy cluster using attribute "scipy.cluster.hierarchy" as sch alias
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.show()
```



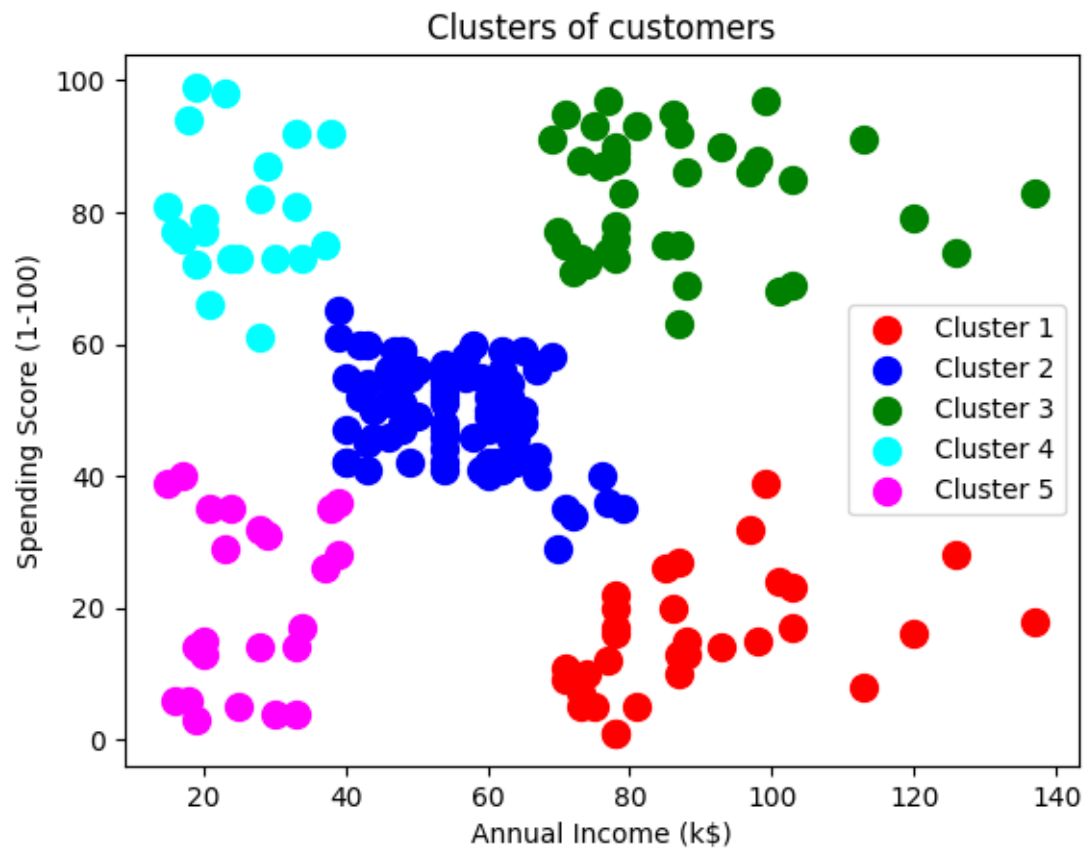
9 Training the Hierarchical Clustering model on the dataset

```
[4]: #Create a cluster for five or nth cluster which you want.
from sklearn.cluster import AgglomerativeClustering
hc = AgglomerativeClustering(n_clusters = 5, affinity = 'euclidean', linkage =
    ↪ 'ward')
y_hc = hc.fit_predict(x)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_agglomerative.py:983:
FutureWarning: Attribute `affinity` was deprecated in version 1.2 and will be
removed in 1.4. Use `metric` instead
  warnings.warn(
```

10 Visualising the clusters

```
[5]: #Plot the scatter plot for scatter visualization.
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()
```



11 Conclusion :

- 11.0.1 According to the model building as the engineer my prediction is cluster number 3 as give highest number of linkage.

12 Insights :

- 12.0.1 Cluster 1 contains(red) color which shows that unsupervised learning cluster has maximum euclidean distance from the centroid upto annual income approximate 139ks.
- 12.0.2 Cluster 2 contains(blue) color which shows that unsupervised learning cluster has maximum euclidean distance from the centroid upto annual income approximately 79-80ks.
- 12.0.3 Cluster 3 contains(green) color which shows that unsupervised learning cluster has maximum euclidean distance from the centroid upto annual income approximately 139ks.
- 12.0.4 Cluster 4 contains(blue) color which shows that unsupervised learning cluster has maximum euclidean distance from the centroid upto annual income approximately 39-40ks.
- 12.0.5 Cluster 5 contains(violet) color which shows that unsupervised learning cluster has maximum euclidean distance from the centroid upto annual income approximately 40-41ks.