Name:Basam Pavani

Pin No: 21X05A6708

Branch:Data Science

College:Narasimha Reddy Engineering College

Project Title:

Analysis of prediction of "Mall_Customers.csv" of amrerican mall market called as phonix mall,Find out on the basis of requirments of dendrogram using scipy graphics library with the help of "scipy.cluster.hierarchy" to a ace the number of linkage of clustering to predict.

Problem Statement:

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

As a data science engineer find out ehich hierarchy cluster give maxmimun linkage in up coming future

TASK-1

With the help of scipy library import the library and import dataset

TASK-2

Using the dendrogram to find the optimal number of clusters

TASK-3

Create a hierarchy model and visualize the cluster with the help of matplot library

Hierarchical Clustering

Importing the libraries

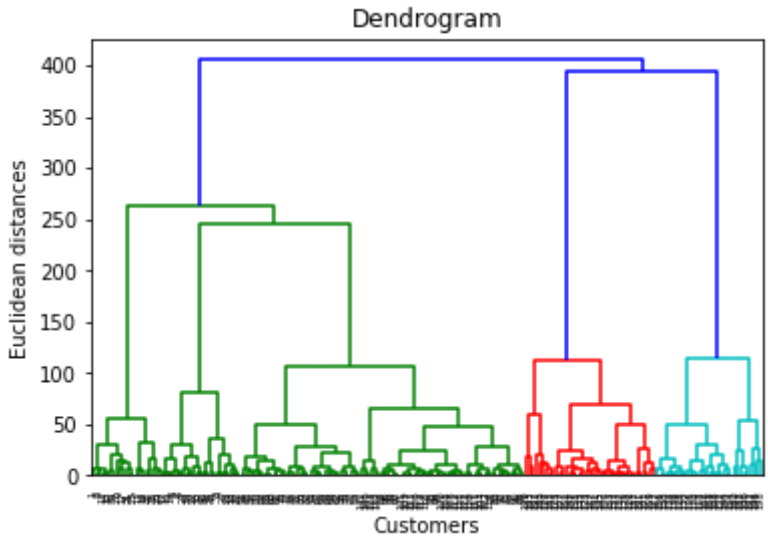
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

Importing the dataset

```
dataset = pd.read_csv('Mall_Customers.csv')
X = dataset.iloc[:, [3, 4]].values
```

Using the dendrogram to find 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.show()
```



Training the Hierarchical Clustering model on the dataset

```
from sklearn.cluster import AgglomerativeClustering
hc = AgglomerativeClustering(n_clusters = 5, affinity = 'euclidean', linkage = 'ward')
y_hc = hc.fit_predict(X)
```

Visualising the clusters

```
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()
```

Conclusion:

According to the model building as a engineer my prediction is cluster number is 3 has highest number of linkage

Insights:

1. Culster 1 contains ("red") which shows that unsupervised learning cluster has maximum eucliing distance from the centriod up to annual income approximate 139ks

2. cluster 2 contains("blue") which shows that unsupervised learning cluster has maximum euclilidng distance from centriod up to approximate 70 -80ks

3. cluster 3 contains ("green") which shows that unsupervised learning cluster has maximum euclilidng distance from centriod up to approximate 139ks

4. cluster 4 contains ("green") which shows that unsupervised learning cluster has maximum euclilidng distance from centriod up to approximate 40ks

5. cluster 5 contains ("green") which shows that unsupervised learning cluster has maximum euclilidng distance from centriod up to approximate 40ks