

Name : K.Nitheesh

RollNo : 21X05A6728

Branch : CSE(DS)

Project Title :

Analysis and Prediction of "Mall_Customers.csv" of American mall market called as phonix mall, to find out on the basis of client requirments of dendrograms using scipy graphics library with the help of "scipy.cluster.hierarchy", to ace the no linkage of clustering to predict.

Problem statement :

The American finance market client as per the rate of GDP of 2011 found as highest no of growth in their business market. As data science engineer find out which hierarcchy cluster use maximum linkage in upcomming future

TASK :

- 1 With the help of scipy library import the librarys and dataset
- 2 Using the dentogram to find the optimal no of clusters.
- 3 Create a hierarchy model and visualize the cluster with the help of matplotlib library.

Conclusion :

According to the model building as a engineer my prediction is cluster no 3 has given the highest no linkage.

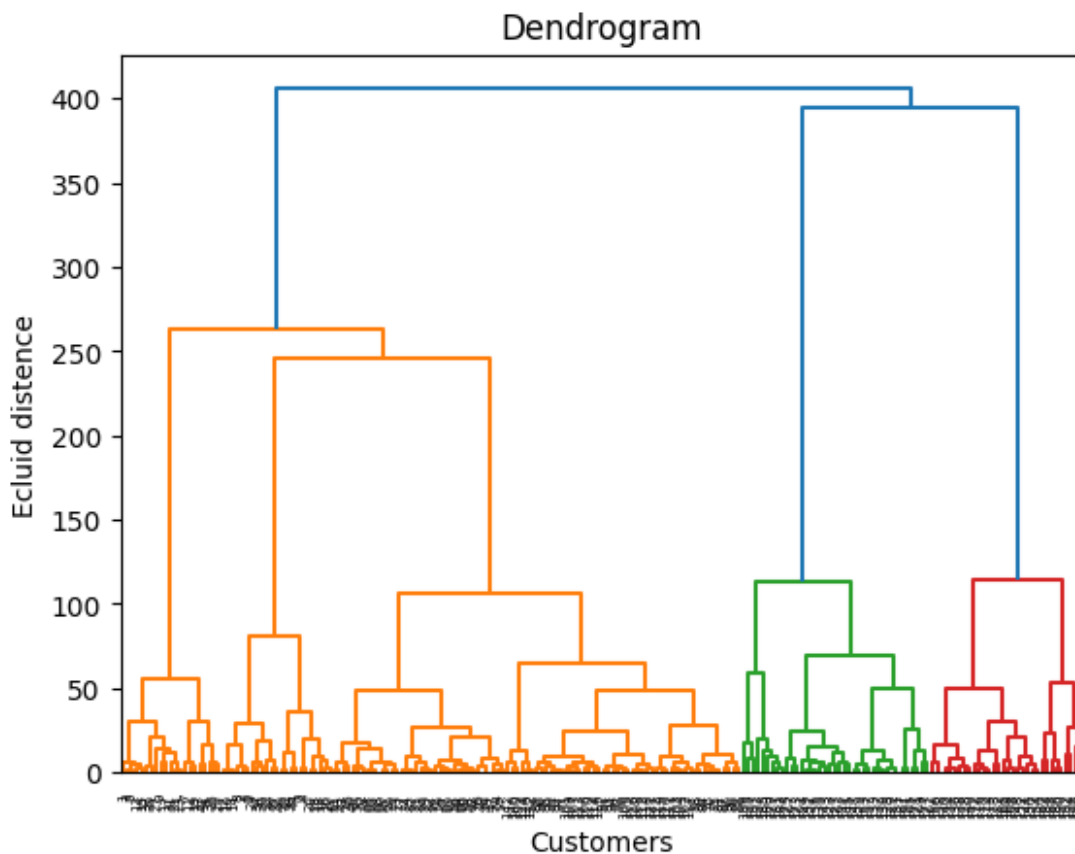
```
#Import the numpy, pandas , matplotlib, seaborn libery's
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

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

#import scipy cluster using attribute "from sklearn.cluster import
KMeans" as sch alias
```

```
#Using the dendrogram to find the optimal number of clusters
# Assign a variable as dendrogram and declares the
"sch.dendrogram(sch.linkage(X, method = 'ward'))"
```

```
import scipy.cluster.hierarchy as sch
dendrogram = sch.dendrogram(sch.linkage(X, method='ward'))
plt.title("Dendrogram")
plt.xlabel("Customers")
plt.ylabel("Ecluid distance")
plt.show()
```



```
# from "sklearn.cluster" attribute import "AgglomerativeClustering"
default argument.
from sklearn.cluster import AgglomerativeClustering

#Create a cluster for five or nth cluster which you want.
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(

#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()

```

