hierarchical-clustering-2

August 28, 2023

0.1 Name: P.Akanksha

0.2 Pin No: 21X05A6741

0.3 class : 4th year

0.4 Branch : DataScience

0.5 Project title:

0.5.1 Analysis and prediction of small custerms.csv file of mall market called as phonix small, find out on the basis of clients requriments of dendogram using scipy graphics library with the help of "scipy.cluster.hierarchy", to ace the No of linkage of clustering to predict.

0.6 Problem Statement :

0.6.1 The american market clients as per the rate of GDP of 2011 found as highest no of growth in there business market

0.6.2 As a DataScience engineer find out which hierarcy cluster give maximum linkage in upcomming future.

1 Hierarchical Clustering

1.1 Importing the libraries

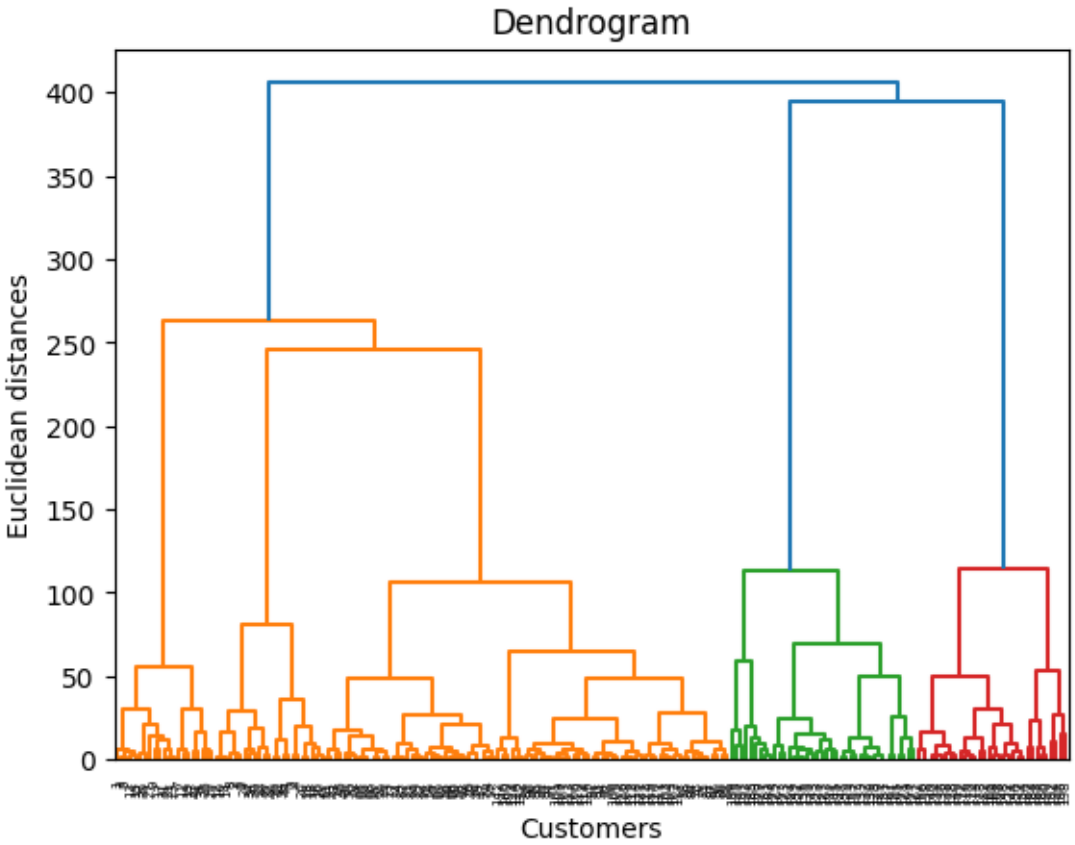
[ ]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

1.2 Importing the dataset

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

1.3 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()
```



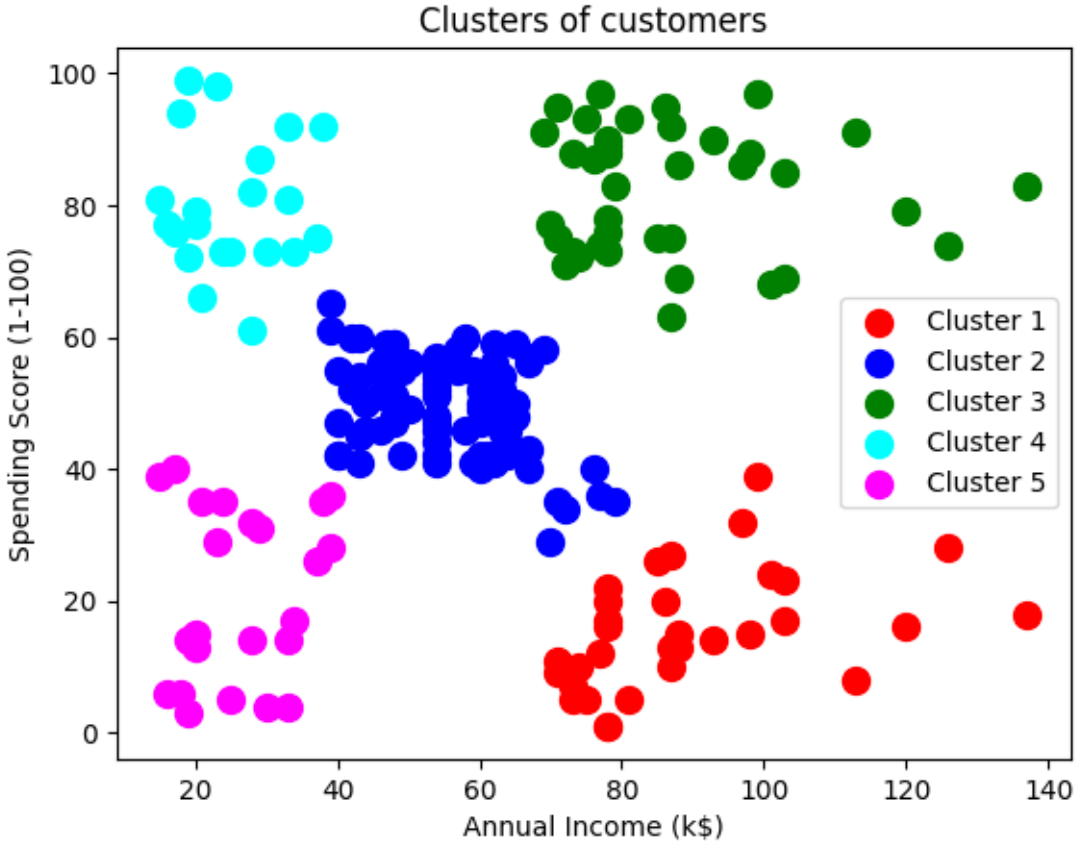
1.4 Training the Hierarchical Clustering model on the dataset

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

```
/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(
```

1.5 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: ### Accoriding to the model building as a enginner my prediction cluster no 3 hightest no of linkage

[ ]: