Customer Segmentation using Data Science

TEAM MEMBER: SAKTHIBALA A

NM ID : aut22leai03

Introduction:

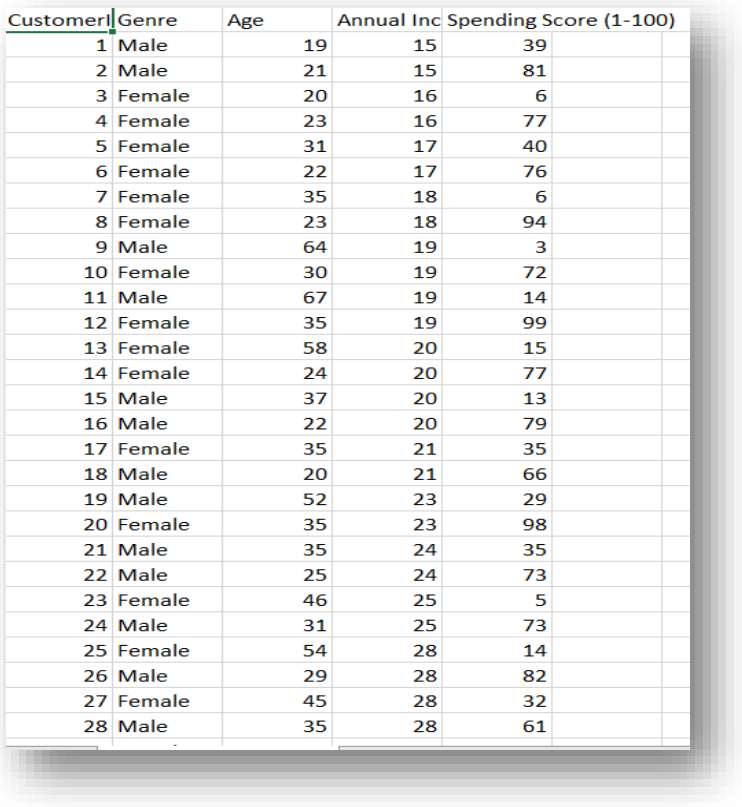
- It is the process of grouping customers according to how and why they are buy products.
- The problem is to implement data science techniques to segment customers based on their behavior, preferences, and demographic attributes.
- The goal is to enable businesses to personalize marketing strategies and enhance customer satisfaction.
- This project involves data collection, data preprocessing, feature engineering, clustering algorithms, visualization, and interpretation of results.
- The main goal for the customer segmentation using data science is to divide the customer base into distinct groups based on similar characteristics.
- This segment will helpful for many Business purpose.

Project Phase 2:

Consider incorporating dimensionality reduction techniques like PCA or t-SNE to visualize high-dimensional customer data and discover underlying patterns.

Dataset:

Dataset link :(https://www.kaggle.com/datasets/akram24/mall-customers)



CustomerI	Genre	Age	Annual Inc	Spending Score (1-100)
1	Male	19	15	39
2	Male	21	15	81
3	Female	20	16	6
4	Female	23	16	77
5	Female	31	17	40
6	Female	22	17	76
7	Female	35	18	6
8	Female	23	18	94
9	Male	64	19	3
10	Female	30	19	72
11	Male	67	19	14
12	Female	35	19	99
13	Female	58	20	15
14	Female	24	20	77
15	Male	37	20	13
16	Male	22	20	79
17	Female	35	21	35
18	Male	20	21	66
19	Male	52	23	29
20	Female	35	23	98
21	Male	35	24	35
22	Male	25	24	73
23	Female	46	25	5
24	Male	31	25	73
25	Female	54	28	14
26	Male	29	28	82
27	Female	45	28	32
28	Male	35	28	61

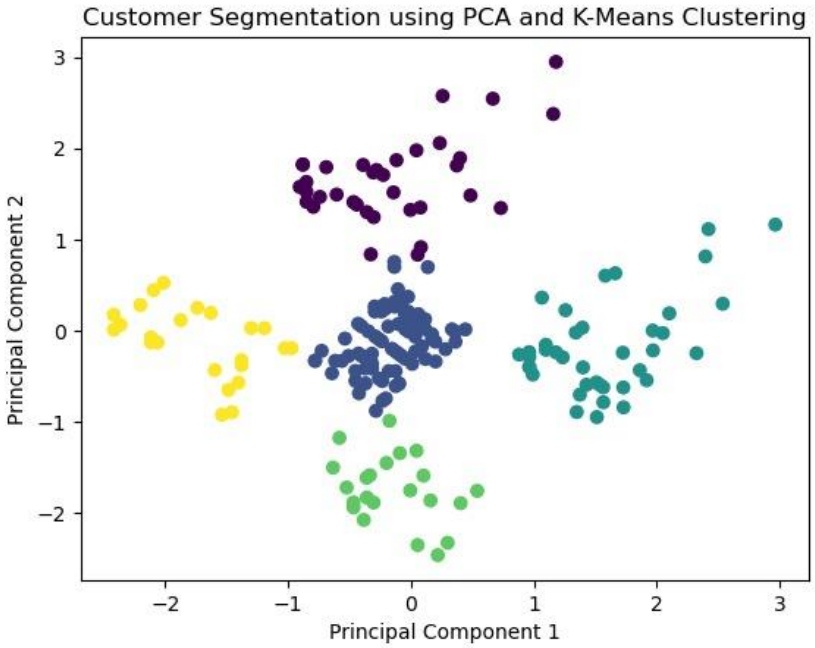
Program :

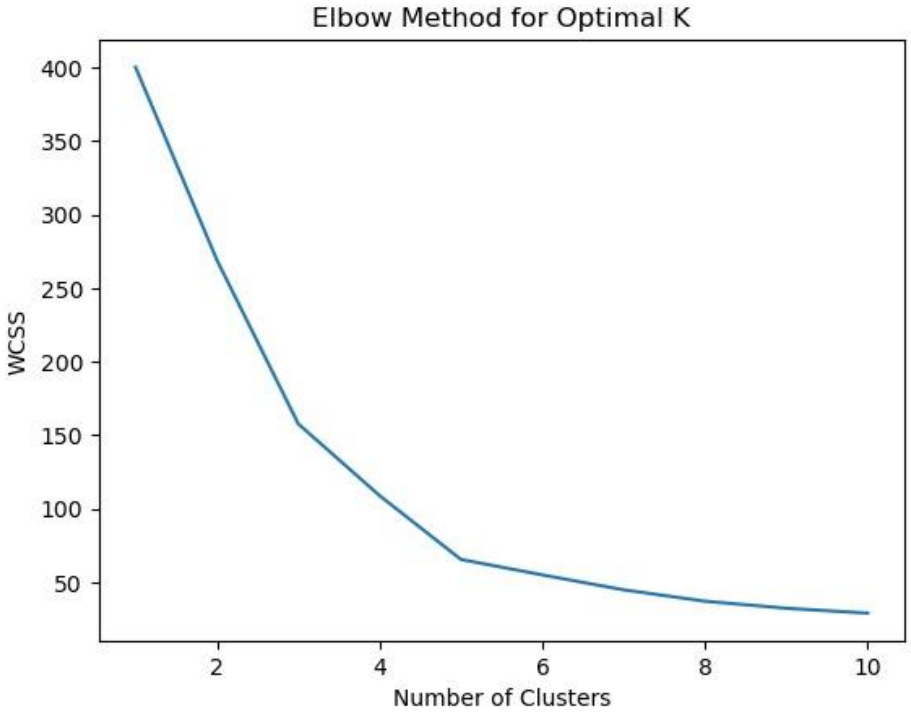
```
import pandas as pd

import matplotlib.pyplot as plt

from sklearn.preprocessing import StandardScaler

from sklearn.decomposition import PCA

from sklearn.cluster import KMeans

# Load the mall customer dataset

data = pd.read_csv("C:\\Users\\Student\\Downloads\\archive (1)\\Mall_Customers.csv")

# Select relevant features for segmentation

X = data[['Annual Income (k$)', 'Spending Score (1-100)']]

# Standardize the data
```

```
scaler = StandardScaler()

X_scaled = scaler.fit_transform(X)

# Perform PCA to reduce dimensionality

pca = PCA(n_components=2)

X_pca = pca.fit_transform(X_scaled)

# Determine the optimal number of clusters using the "elbow method"

wcss = []

for i in range(1, 11):

   kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300, n_init=10,
random_state=0,)

   kmeans.fit(X_pca)

   wcss.append(kmeans.inertia_)

# Plot the Within-Cluster-Sum-of-Squares (WCSS) to find the elbow point

plt.plot(range(1, 11), wcss)

plt.xlabel('Number of Clusters')

plt.ylabel('WCSS')

plt.title('Elbow Method for Optimal K')

plt.show()

# Based on the plot, choose the optimal number of clusters

optimal_k = 5

# Perform K-Means clustering with the chosen number of clusters

kmeans = KMeans(n_clusters=optimal_k, init='k-means++', max_iter=300, n_init=10,
random_state=0)

cluster_labels = kmeans.fit_predict(X_pca)

# Add the cluster labels to the original dataset

data['Cluster'] = cluster_labels

# Visualize the clusters

plt.scatter(X_pca[:, 0], X_pca[:, 1], c=data['Cluster'], cmap='viridis')

plt.xlabel('Principal Component 1')

plt.ylabel('Principal Component 2')
```

```
plt.title('Customer Segmentation using PCA and K-Means Clustering')
plt.show()
```

Output :





conclusion :

This code allows you to segment mall customers into distinct clusters based on their annual income and spending score. The number of clusters (5 in this case) can be adjusted based on business needs and insights gained from the analysis. These clusters can be further analyzed and interpreted to tailor marketing strategies, promotions, or services to each customer group, improving the overall shopping experience and potentially increasing sales.