Name: Harsh Patel

Roll NO: A42

Experiment No:8

Aim: Write a python program to implement K mean clustering in python

Code :

```
import numpy as np

import matplotlib.pyplot as plt

import pandas as pd


# Import Mall Dataset

dataset = pd.read_csv('3_Clustering/Mall_Customers.csv')


# Extracting features

X = dataset.iloc[:, [3, 4]].values


# Use Elbow Method to determine optimal number of clusters

from sklearn.cluster import KMeans

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)

    wcss.append(kmeans.inertia_)


# Plotting the Elbow Method graph

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

plt.title("ELBOW METHOD")

plt.xlabel("Number of Clusters")

plt.ylabel("Within-Cluster Sum of Squares (WCSS)")
```

```python
plt.show()


# Applying K-means to Mall
kmeans = KMeans(n_clusters=5, init='k-means++', max_iter=300, n_init=10,
random_state=0)
y_kmeans = kmeans.fit_predict(X)


# Visualising the Clusters
plt.scatter(X[y_kmeans==0, 0], X[y_kmeans==0, 1], s= 100, c='red', label='Cluster 1')
plt.scatter(X[y_kmeans==1, 0], X[y_kmeans==1, 1], s= 100, c='blue', label='Cluster 2')
plt.scatter(X[y_kmeans==2, 0], X[y_kmeans==2, 1], s= 100, c='yellow', label='Cluster 3')
plt.scatter(X[y_kmeans==3, 0], X[y_kmeans==3, 1], s= 100, c='green', label='Cluster 4')
plt.scatter(X[y_kmeans==4, 0], X[y_kmeans==4, 1], s= 100, c='cyan', label='Cluster 5')
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s=300, c='red',
label='Centroids')
plt.title('Clusters of customers')
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score (1-100)')
plt.legend()
plt.show()
```

OUTPUT:

ELBOW METHOD

Clusters of customers