Program 12

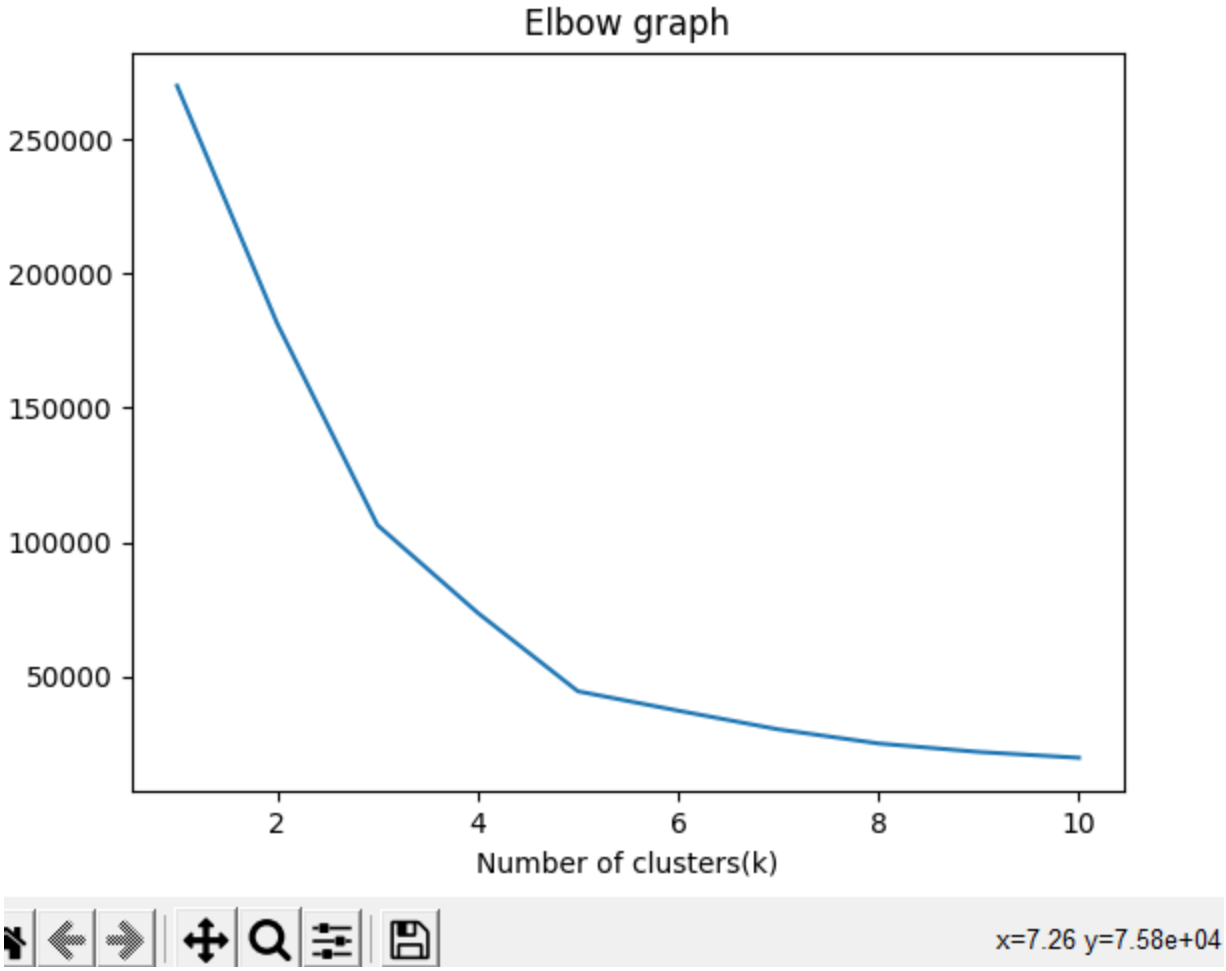
AIM: program to implement K-means clustering technique using any standard dataset available in public domain.

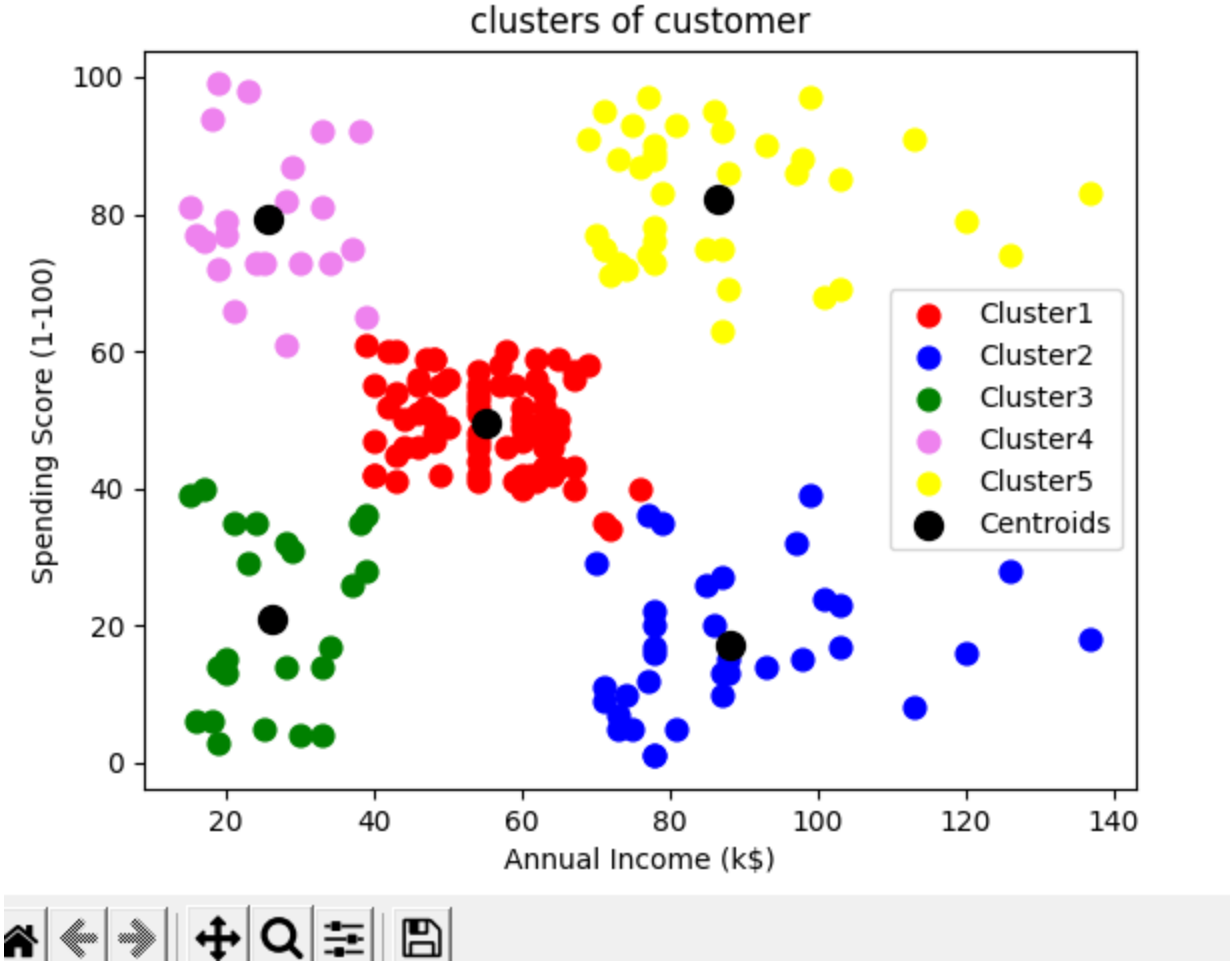
Source code:

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import sklearn
dataset = pd.read_csv('Mall_Customers.csv')
X = dataset.iloc[:, [3, 4]].values
print(X)

from sklearn.cluster import KMeans
wcss_list = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 42)
    kmeans.fit(X)
    wcss_list.append(kmeans.inertia_)
plt.plot(range(1, 11), wcss_list)
plt.title('Elbow graph')
plt.xlabel('Number of clusters(k)')
plt.ylabel('wcss_list')
plt.show()
kmeans = KMeans(n_clusters = 5, init = "k-means++", random_state = 42)
y_predict = kmeans.fit_predict(X)
print(y_predict)
plt.scatter(X[y_predict == 0, 0], X[y_predict== 0, 1], s = 60, c = 'red', label
= 'Cluster1')
plt.scatter(X[y_predict == 1, 0], X[y_predict == 1, 1], s = 60, c = 'blue',
label = 'Cluster2')
plt.scatter(X[y_predict == 2, 0], X[y_predict == 2, 1], s = 60, c = 'green',
label = 'Cluster3')
plt.scatter(X[y_predict == 3, 0], X[y_predict == 3, 1], s = 60, c = 'violet',
label = 'Cluster4')
plt.scatter(X[y_predict == 4, 0], X[y_predict == 4, 1], s = 60, c = 'yellow',
label = 'Cluster5')
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s =
100, c = 'black', label = 'Centroids')
plt.title('clusters of customer')
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score (1-100)')
plt.legend()
plt.show()
```

Output

```
C:\Users\ajcemca\PycharmProjects\kmeans\venv\Scripts\python.exe C:/Users/ajcemca/PycharmProjects/kmeans/main.py
[[ 15  39]
 [ 15  81]
 [ 16   6]
 [ 16  77]
 [ 17  40]
 [ 17  76]
 [ 18   6]
 [ 18  94]
 [ 19   3]
 [ 19  72]
 [ 19  14]
 [ 19  99]
 [ 20  15]
 [ 20  77]
 [ 20  13]
 [ 20  79]
 [ 21  35]
 [ 21  66]
 [ 23  29]
 [ 23  98]
 [ 24  35]
 [ 24  73]
 [ 25   5]
 [ 25  73]
 [101  68]
 [103  17]
 [103  85]
 [103  23]
 [103  69]
 [113   8]
 [113  91]
 [120  16]
 [120  79]
 [126  28]
 [126  74]
 [137  18]
 [137  83]]
[2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2
 3 2 3 2 3 2 0 2 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 4 1 4 0 4 1 4 1 4 0 4 1 4 1 4 1 4 1 4 0 4 1 4 1 4
 1 4 1 4 1 4 1 4 1 4 1 4 1 4 1 4 1 4 1 4 1 4 1 4 1 4 1 4 1 4 1 4 1 4 1 4 1
 4 1 4 1 4 1 4 1 4 1 4 1 4 1 4]

Process finished with exit code 0
```


Elbow graph
250000
200000
150000
100000
50000
2
4
6
8
10
Number of clusters(k)
x=7.26 y=7.58e+04


clusters of customer
Spending Score (1-100)
Annual Income (k$)
Cluster1
Cluster2
Cluster3
Cluster4
Cluster5
Centroids
0
20
40
60
80
100
20
40
60
80
100
120
140