Program No:12

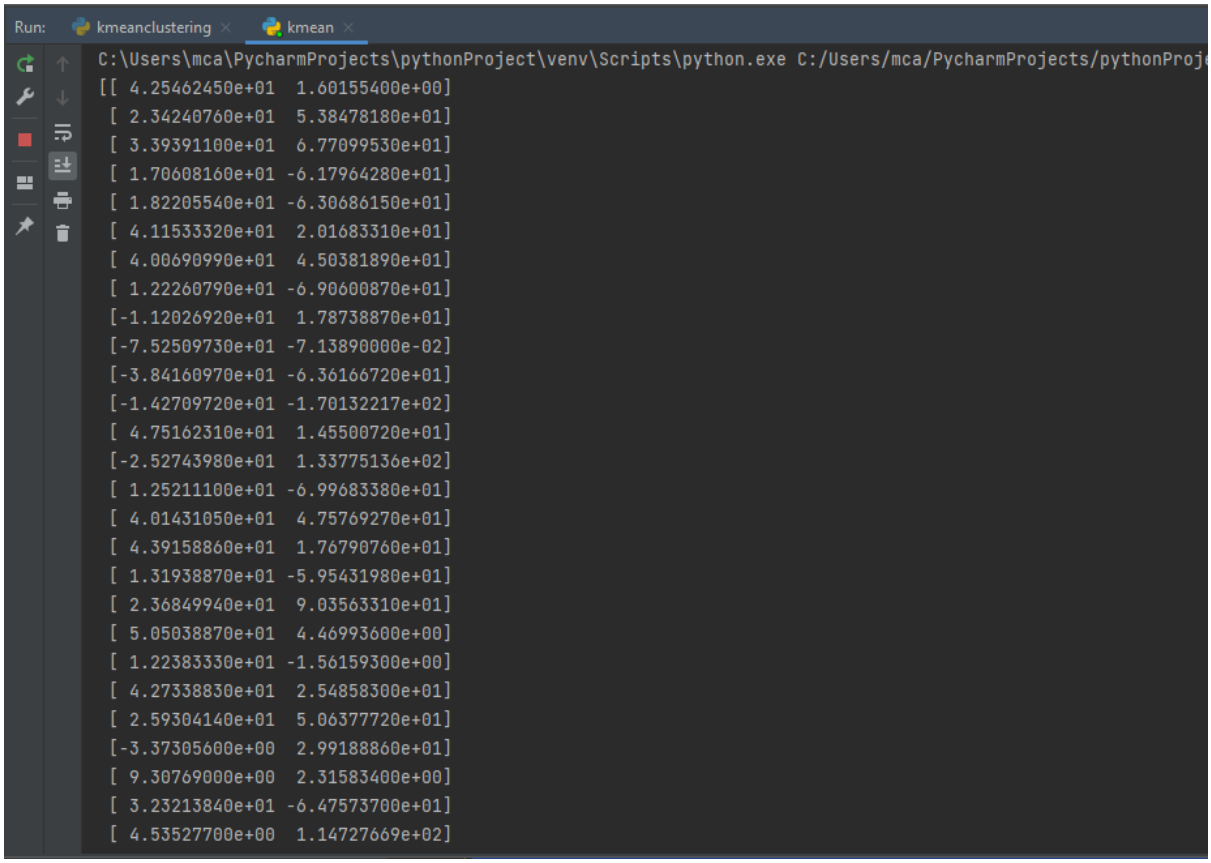
Aim:Program to implement k-means clustering techniques using any standard dataset available in the public domain.

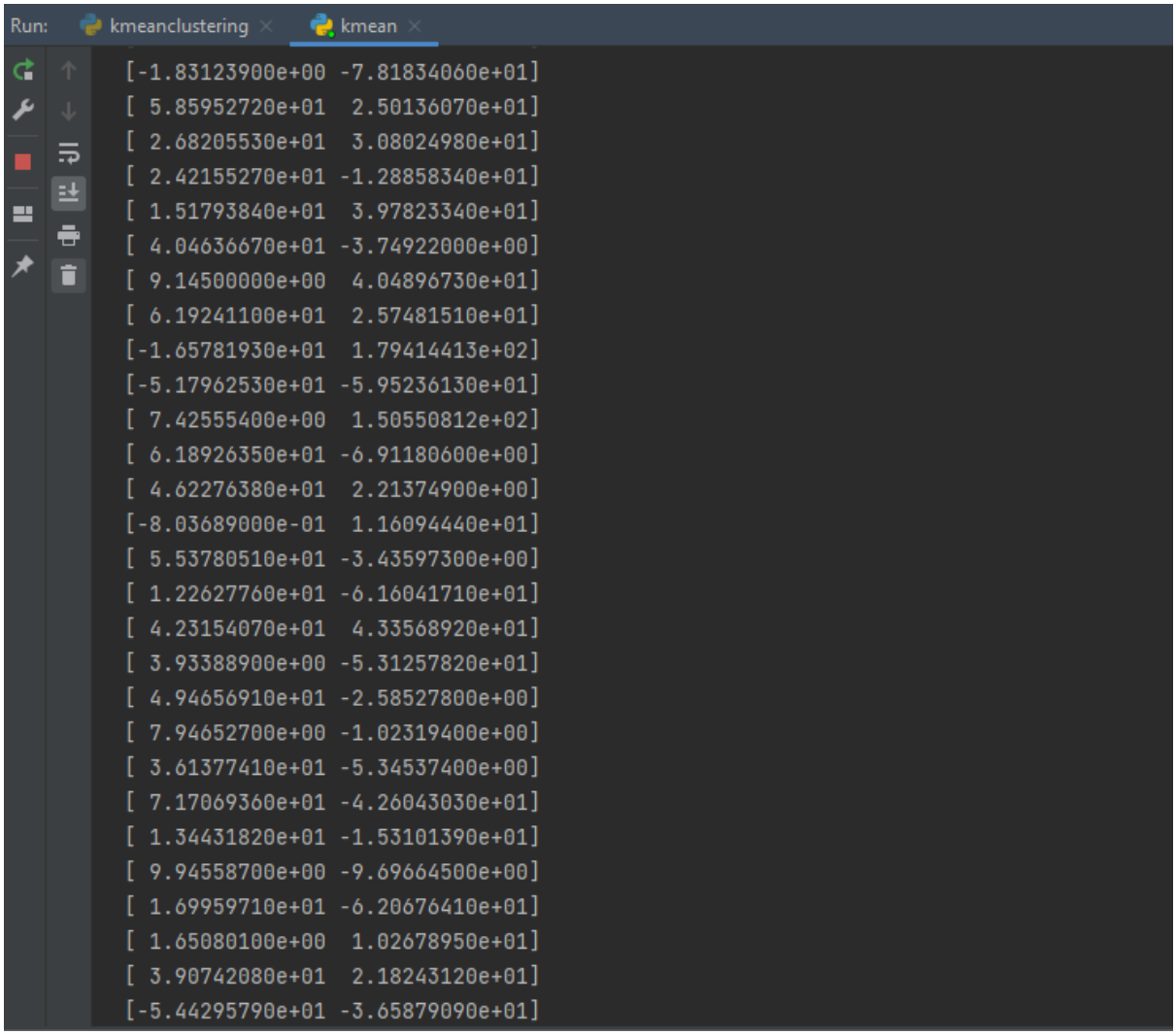
Program:

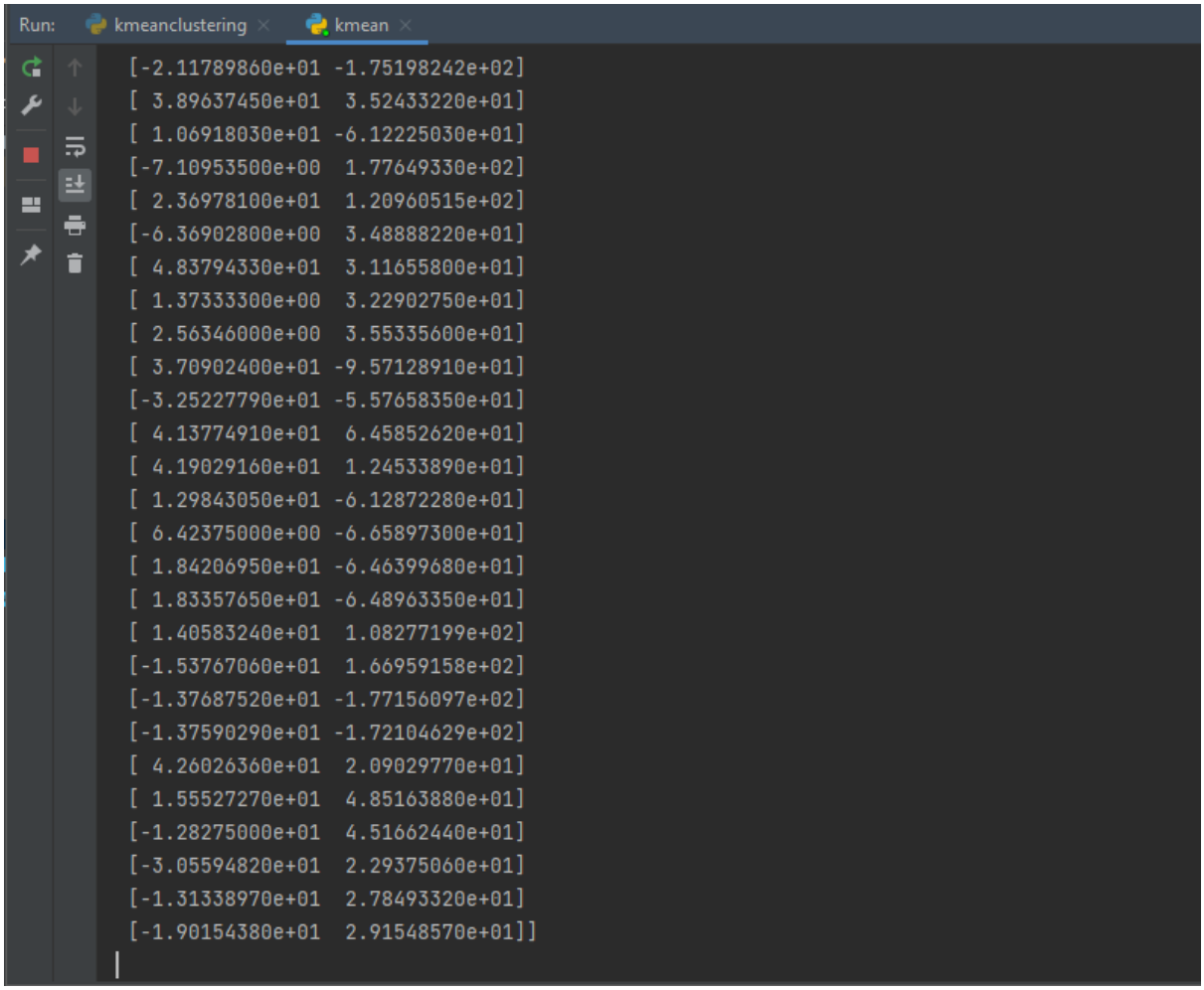
```
import numpy as np
import matplotlib.pyplot as mtp
import pandas as pd
dataset=pd.read_csv('world_country_and_usa_stat
es_latitude_and_longitude_values.csv')
x=dataset.iloc[:,[1,2]].values
print(x)
from sklearn.cluster import KMeans
wcss_list = []
for i in range(1, 11):
          kmeans = KMeans(n_clusters=i,
init='k-means++')
          kmeans.fit(x)
          wcss_list.append(kmeans.inertia_)
mtp.plot(range(1,11), wcss_list)
mtp.title('The elbow method Graph')
mtp.xlabel('Number of clusters (k)')
mtp.ylabel('wcss_list')
mtp.show()
kmeans = KMeans(n_clusters=3,init='k-
means++',random_state=42)
y_predict=kmeans.fit_predict(x)
print(y_predict)
mtp.scatter(x[y_predict == 0,0], x[y_predict
==0,1], s=100, c='blue', label='Cluster0')
mtp.scatter(x[y_predict == 1,0], x[y_predict
==1,1], s=100, c='green', label= 'Cluster1')
mtp.scatter(x[y_predict == 2,0], x[y_predict
==2,1], s=100, c='red', label= 'Cluster2')
mtp.scatter(x[y_predict == 3,0], x[y_predict
==3,1], s=100, c='cyan', label= 'Cluster3')
mtp.scatter(x[y_predict == 4,0], x[y_predict
==4,1], s=100, c='magenta', label= 'Cluster4')
```

```
mtp.scatter(kmeans.cluster_centers_[:,0],kmeans
.cluster_centers_[:,1], s = 300,)
mtp.title('clusters of customers')
mtp.xlabel('longtitude')
mtp.ylabel('latitude')
mtp.legend()
mtp.show()
```

Output:




Run:
kmeanclustering
kmean
[-1.83123900e+00 -7.81834060e+01]
[5.85952720e+01 2.50136070e+01]
[2.68205530e+01 3.08024980e+01]
[2.42155270e+01 -1.28858340e+01]
[1.51793840e+01 3.97823340e+01]
[4.04636670e+01 -3.74922000e+00]
[9.14500000e+00 4.04896730e+01]
[6.19241100e+01 2.57481510e+01]
[-1.65781930e+01 1.79414413e+02]
[-5.17962530e+01 -5.95236130e+01]
[7.42555400e+00 1.50550812e+02]
[6.18926350e+01 -6.91180600e+00]
[4.62276380e+01 2.21374900e+00]
[-8.03689000e-01 1.16094440e+01]
[5.53780510e+01 -3.43597300e+00]
[1.22627760e+01 -6.16041710e+01]
[4.23154070e+01 4.33568920e+01]
[3.93388900e+00 -5.31257820e+01]
[4.94656910e+01 -2.58527800e+00]
[7.94652700e+00 -1.02319400e+00]
[3.61377410e+01 -5.34537400e+00]
[7.17069360e+01 -4.26043030e+01]
[1.34431820e+01 -1.53101390e+01]
[9.94558700e+00 -9.69664500e+00]
[1.69959710e+01 -6.20676410e+01]
[1.65080100e+00 1.02678950e+01]
[3.90742080e+01 2.18243120e+01]
[-5.44295790e+01 -3.65879090e+01]


Run:
kmeanclustering
kmean
[-2.11789860e+01 -1.75198242e+02]
[3.89637450e+01 3.52433220e+01]
[1.06918030e+01 -6.12225030e+01]
[-7.10953500e+00 1.77649330e+02]
[2.36978100e+01 1.20960515e+02]
[-6.36902800e+00 3.48888220e+01]
[4.83794330e+01 3.11655800e+01]
[1.37333300e+00 3.22902750e+01]
[2.56346000e+00 3.55335600e+01]
[3.70902400e+01 -9.57128910e+01]
[-3.25227790e+01 -5.57658350e+01]
[4.13774910e+01 6.45852620e+01]
[4.19029160e+01 1.24533890e+01]
[1.29843050e+01 -6.12872280e+01]
[6.42375000e+00 -6.65897300e+01]
[1.84206950e+01 -6.46399680e+01]
[1.83357650e+01 -6.48963350e+01]
[1.40583240e+01 1.08277199e+02]
[-1.53767060e+01 1.66959158e+02]
[-1.37687520e+01 -1.77156097e+02]
[-1.37590290e+01 -1.72104629e+02]
[4.26026360e+01 2.09029770e+01]
[1.55527270e+01 4.85163880e+01]
[-1.28275000e+01 4.51662440e+01]
[-3.05594820e+01 2.29375060e+01]
[-1.31338970e+01 2.78493320e+01]
[-1.90154380e+01 2.91548570e+01]]

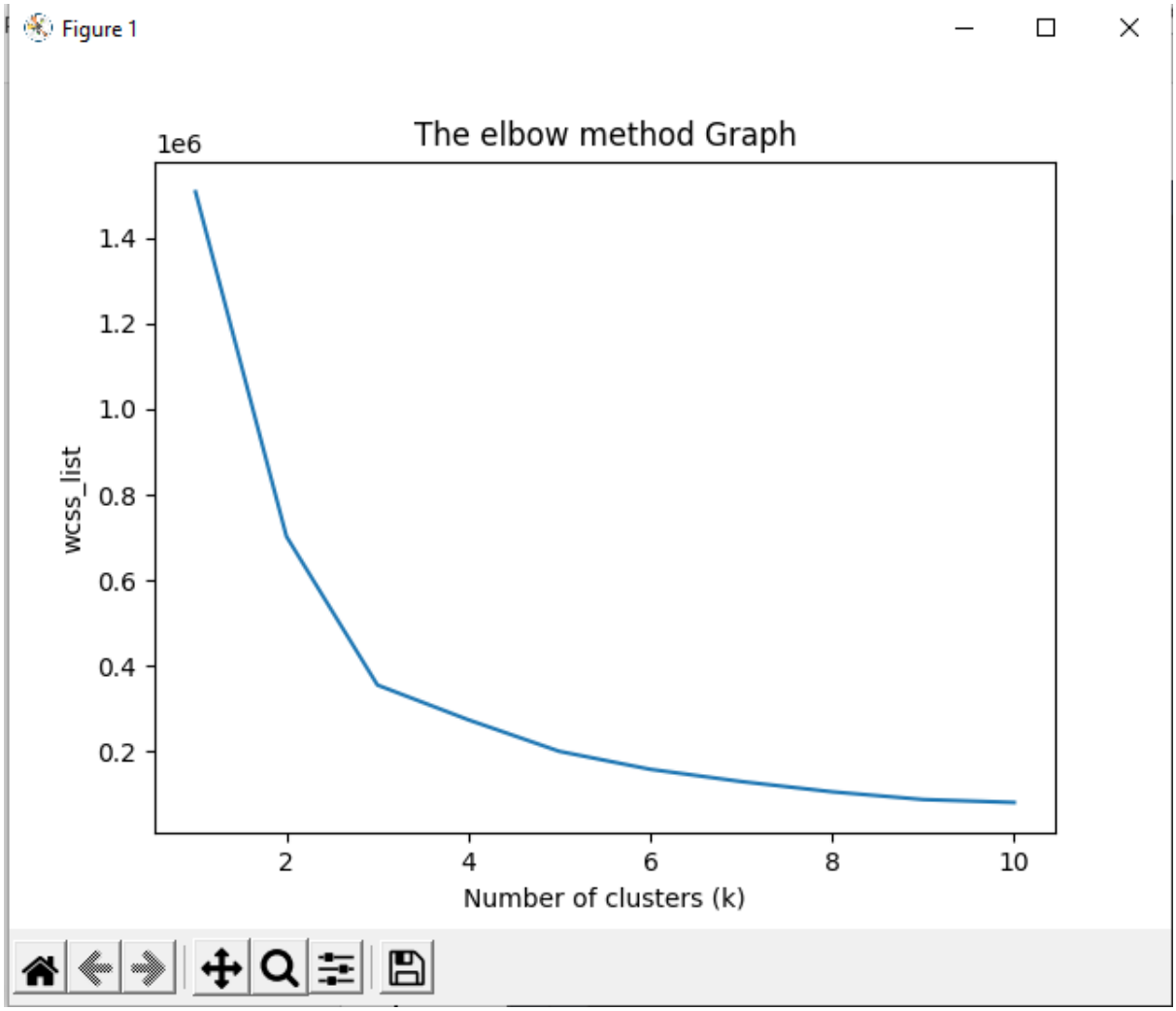


Figure 1
The elbow method Graph
1e6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
wcss_list
2
4
6
8
10
Number of clusters (k)

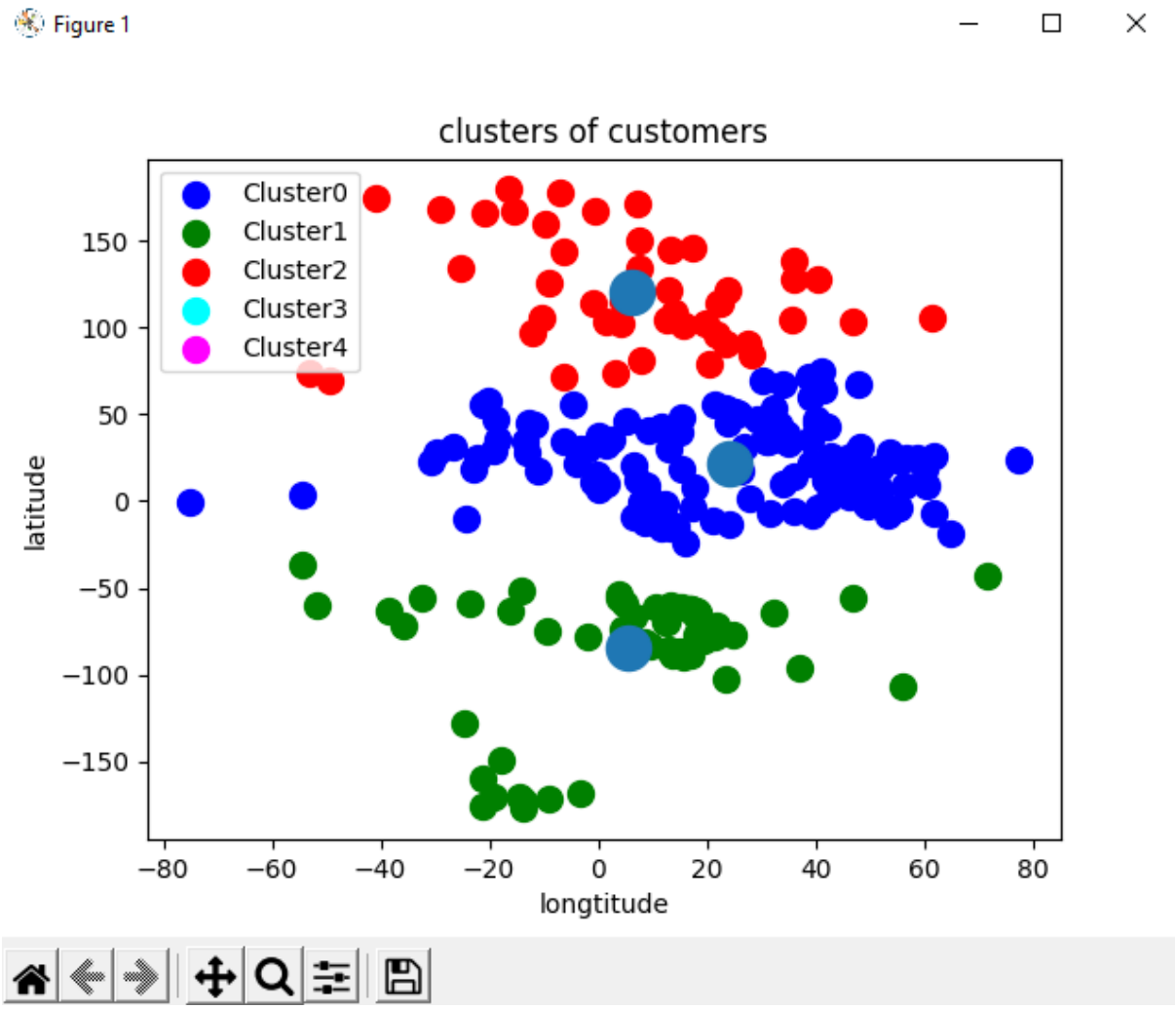


Figure 1
clusters of customers
Cluster0
Cluster1
Cluster2
Cluster3
Cluster4
latitude
longtitude
150
100
50
0
−50
−100
−150
−80
−60
−40
−20
0
20
40
60
80