Experiment-14

Aim: Write a program of cluster analysis using simple K-means algorithm python programming Language.

Objectives:

Cluster analysis: Cluster analysis is a statistical method for processing data. It works by organizing items into groups, or clusters on the basis of how closely associated they are.

K-means algorithm: K-means is a simple two steps clustering process. The first step is cluster assignment and the second one is the move centroid step. However, this unsupervised algorithm can easily create, implement and handle massive datasets.

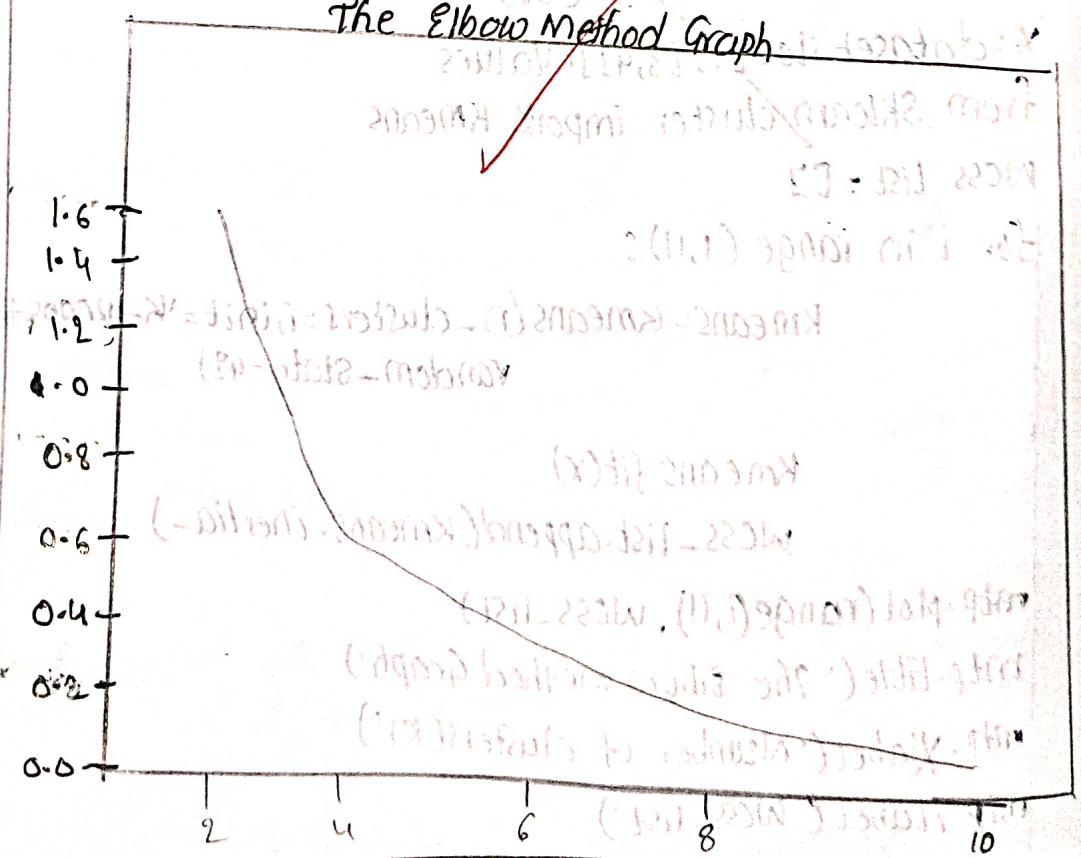
K-means Algorithm using python progeamming

```
import numpy as nm
import matplotlib.pyplot as mtp
import pandas as pd
dataset = pd.read_csv('/content/Sample_data/mall_customers.
                       csv')
X = dataset.iloc[:, [3,4]].values
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_)
mtp.plot(range(i,11), wcss_list)
mtp.title('The Elbow method Graph')
mtp.Xlabel('Number of clusters(K)')
mtp.Ylabel('wcss_list')
```

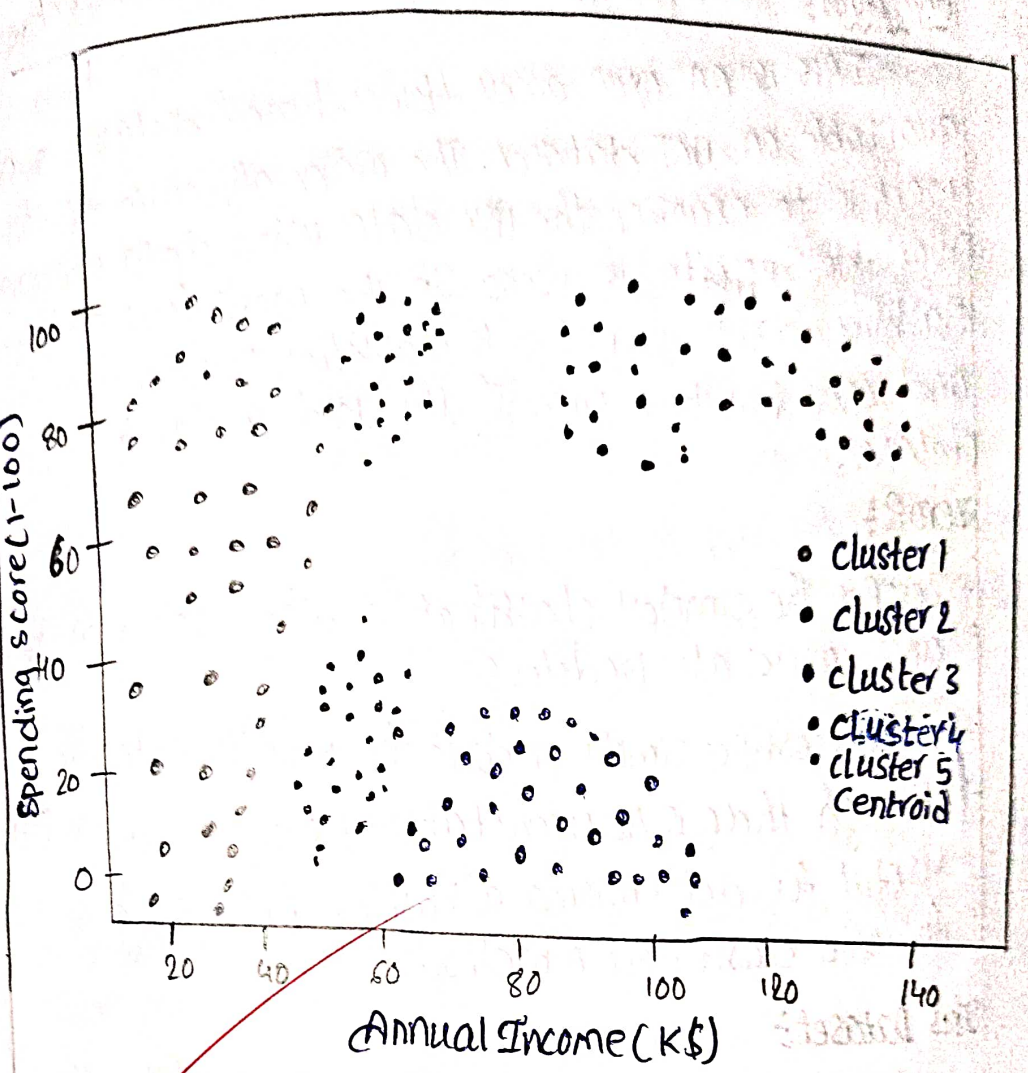
```
mtp.show()
Kmeans = Kmeans(n_clusters=5, init='k-means++', random_state
        =42)
Y_predict = Kmeans.fit_predict(x)
mtp.scatter(x[Y_predict=0,0], x[Y_predict=0,1], s=100, c='blue
        label='cluster1')
mtp.scatter(x[Y_predict=1,0], x[Y_predict=1,1], s=100, c='green
        label='cluster2')
mtp.scatter(x[Y_predict=2,0], x[Y_predict=2,1], s=100, c='red'
        label='cluster3')
mtp.scatter(Kmeans.cluster_centers_[:,0], Kmeans.cluster_
        Centers_[:,1], s=300, c='yellow', label='centroid
mtp.title('clusters of customers')
mtp.xlabel('Annual Income(K$)')
mtp.ylabel('spending score(1-100)')
mtp.legend()
mtp.show()
```

Output:



Number of clusters(K)

clusters of customers



$\frac{E}{21/10/22}$