

**PRACTICAL NO.8****8. Write a program in python for K-Means Clustering.**

```
import matplotlib.pyplot as plt
```

```
import pandas as pd
```

```
dataset = pd.read_csv('Customers.csv')
```

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

```
from sklearn.cluster import KMeans
```

```
wcss = []
```

```
for i in range(1, 11):
```

```
    kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 42)
```

```
    kmeans.fit(X)
```

```
    wcss.append(kmeans.inertia_)
```

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

```
plt.title('The Elbow Method')
```

```
plt.xlabel('Number of clusters')
```

```
plt.ylabel('WCSS')
```

```
plt.show()
```

```
kmeans = KMeans(n_clusters = 5, init = 'k-means++', random_state = 42)
```

```
y_kmeans = kmeans.fit_predict(X)
```

```
plt.scatter(X[y_kmeans == 0, 0], X[y_kmeans == 0, 1], s = 100, c = 'pink', label = 'Cluster 1')
```

```
plt.scatter(X[y_kmeans == 1, 0], X[y_kmeans == 1, 1], s = 100, c = 'brown', label = 'Cluster 2')
```

```
plt.scatter(X[y_kmeans == 2, 0], X[y_kmeans == 2, 1], s = 100, c = 'gray', label = 'Cluster 3')
```

```
plt.scatter(X[y_kmeans == 3, 0], X[y_kmeans == 3, 1], s = 100, c = 'indigo', label = 'Cluster 4')
```

```
plt.scatter(X[y_kmeans == 4, 0], X[y_kmeans == 4, 1], s = 100, c = 'green', label = 'Cluster 5')

plt.scatter(kmeans.cluster_centers_[0], kmeans.cluster_centers_[1], s = 300, c = 'yellow', label =
'Centroids')

plt.title('Clusters of customers')

plt.xlabel('Annual Income')

plt.ylabel('Spending Score (1-100)')

plt.legend()

plt.show()
```

Create a Customers.csv file in Microsoft Excel and under that file enter the following data:-

CustomerID	Genre	Age	AnnualIncome	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	17	72

### Output:-

