Customer Segmentation System Using K-Means

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```
df = pd.read_csv('Mall_Customers.csv')
```

In [3]:
```
df.head(10)
```

Out[3]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
5	6	Female	22	17	76
6	7	Female	35	18	6
7	8	Female	23	18	94
8	9	Male	64	19	3
9	10	Female	30	19	72

In [4]:
```
df.shape
```

Out[4]:
```
(200, 5)
```

In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 5 columns):
 #   Column                  Non-Null Count  Dtype 
---  ------                  --------------  ----- 
 0   CustomerID              200 non-null    int64 
 1   Gender                  200 non-null    object
 2   Age                     200 non-null    int64 
 3   Annual Income (k$)      200 non-null    int64 
 4   Spending Score (1-100)  200 non-null    int64 
dtypes: int64(4), object(1)
memory usage: 7.9+ KB
```

In [6]:
```
X = df.iloc[:, [3, 4]].values
```

In [16]:
```
X
```

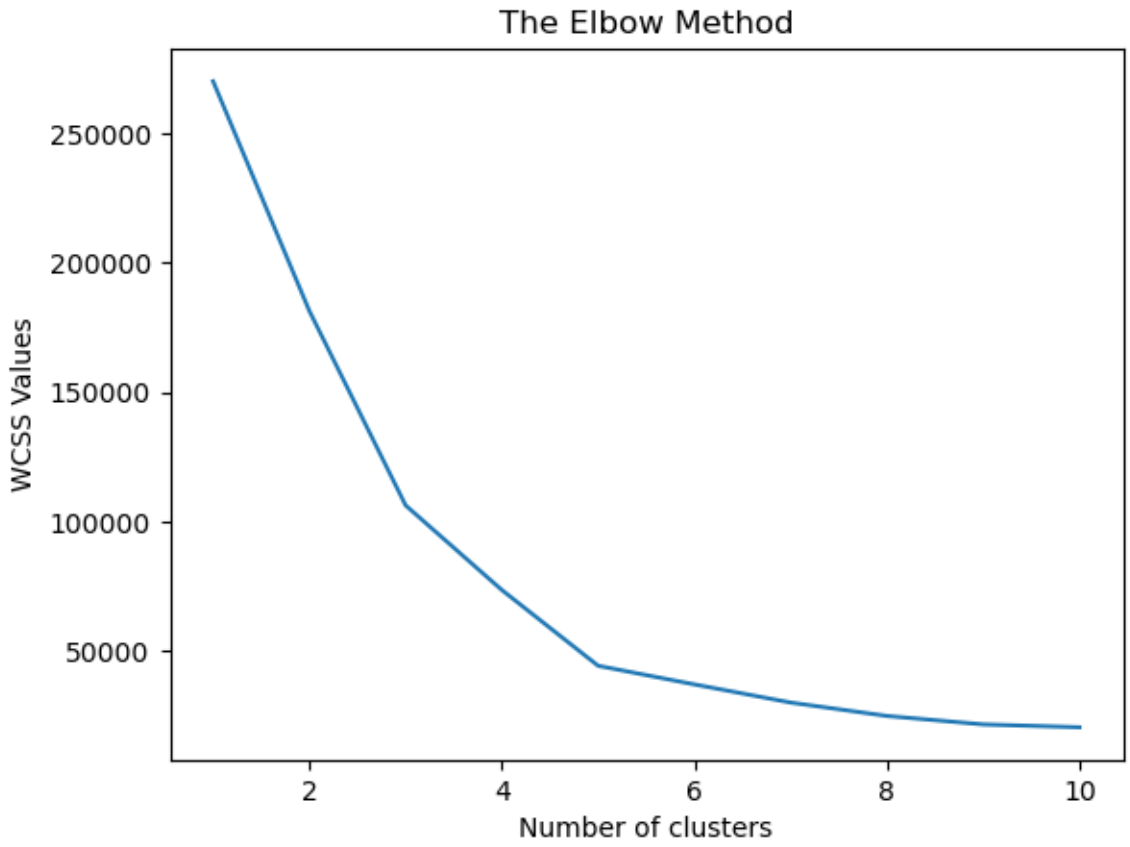
Out[16]:
```
array([[ 15,  39],
       [ 15,  81],
       [ 16,   6],
       [ 16,  77],
       [ 17,  40],
       [ 17,  76],
       [ 18,   6],
       [ 18,  94],
       [ 19,   3],
       [ 19,  72],
       [ 19,  14],
       [ 19,  99],
       [ 20,  15],
       [ 20,  77],
       [ 20,  13],
       [ 20,  79],
       [ 21,  35],
       [ 21,  66],
       [ 23,  29],
```

Perform Elbow Method to find Optimal No. of Clusters

In [8]:
```
from sklearn.cluster import KMeans
wcss = []
```

In [9]:
```
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', random_state=0)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
```

In [17]:
```
plt.plot(range(1, 11), wcss)
plt.title("The Elbow Method")
plt.xlabel('Number of clusters')
plt.ylabel('WCSS Values')
plt.show()
```



Training a model using Unsupervised Learning Algorithm

In [13]:
```
kmeansmodel = KMeans(n_clusters=5, init='k-means++', random_state=0)
```

In [14]:
```
y_kmeans = kmeansmodel.fit_predict(X)
```

In [18]:
```
plt.scatter(X[y_kmeans == 0,0], X[y_kmeans == 0,1], s= 80, c = "red", label=
plt.scatter(X[y_kmeans == 1,0], X[y_kmeans == 1,1], s= 80, c = "blue", label
plt.scatter(X[y_kmeans == 2,0], X[y_kmeans == 2,1], s= 80, c = "yellow", lab
plt.scatter(X[y_kmeans == 3,0], X[y_kmeans == 3,1], s= 80, c = "cyan", label
plt.scatter(X[y_kmeans == 4,0], X[y_kmeans == 4,1], s= 80, c = "black", labe
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s=
plt.title('Clusters of Customers')
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score (1-100)')
plt.legend()
plt.show()
```

You can find this project on **Github.** (https://github.com/Vyas-Rishabh/Customer_Segmentation_System_Using_K-Means)