

Importing necessary libraries

```
In [1]: import pandas as pd
        from sklearn.cluster import KMeans
        from matplotlib import pyplot as plt
        from sklearn.metrics import silhouette_score
```

Reading and Processing Data

```
In [2]: df = pd.read_csv(r"C:\CSV files\Mall_Customers.csv");
```

```
In [3]: df
```

Out[3]:

	CustomerID	Genre	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
...
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

200 rows × 5 columns

```
In [4]: df.isnull().sum()
```

```
Out[4]: CustomerID      0
        Genre          0
        Age            0
        Annual Income (k$)  0
        Spending Score (1-100)  0
        dtype: int64
```

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

Out[5]:

	CustomerID	Genre	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

In [6]: `df.tail()`

Out[6]:

	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

In [7]: `x = df.iloc[:, [3,4]].values`

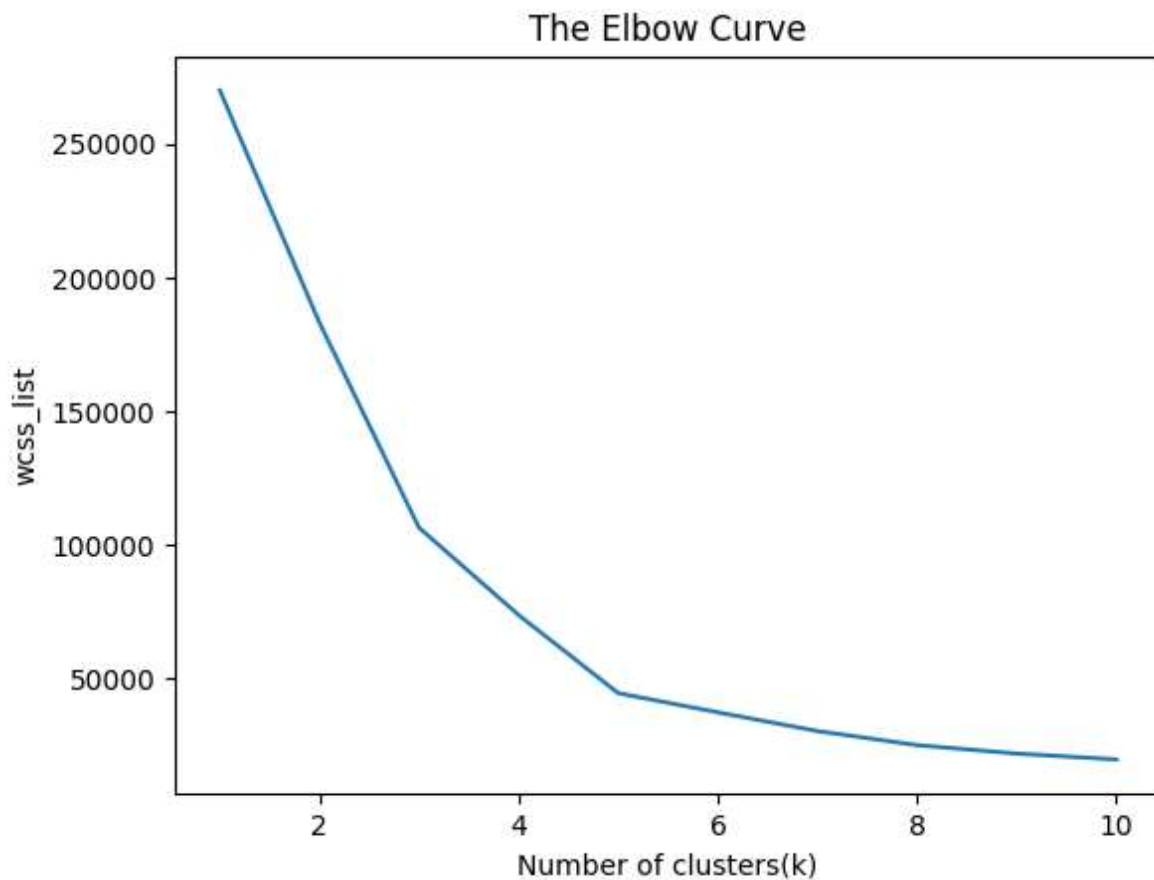
In [8]: `x`

```
[ 97,  86],
[ 97,  86],
[ 98,  15],
[ 98,  88],
[ 99,  39],
[ 99,  97],
[101,  24],
[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]], dtype=int64)
```

In [9]: `wcss_list = []`

The Elbow Curve

```
In [10]: for i in range(1, 11):  
          kmeans = KMeans(n_clusters=i, init="k-means++", random_state=55)  
          kmeans.fit(x)  
          wcss_list.append(kmeans.inertia_)  
plt.plot(range(1, 11), wcss_list)  
plt.title("The Elbow Curve")  
plt.xlabel("Number of clusters(k)")  
plt.ylabel("wcss_list")  
plt.show()
```



```
In [11]: kmeans = KMeans(n_clusters=5, init="k-means++", random_state=55)  
y_pred = kmeans.fit_predict(x)
```

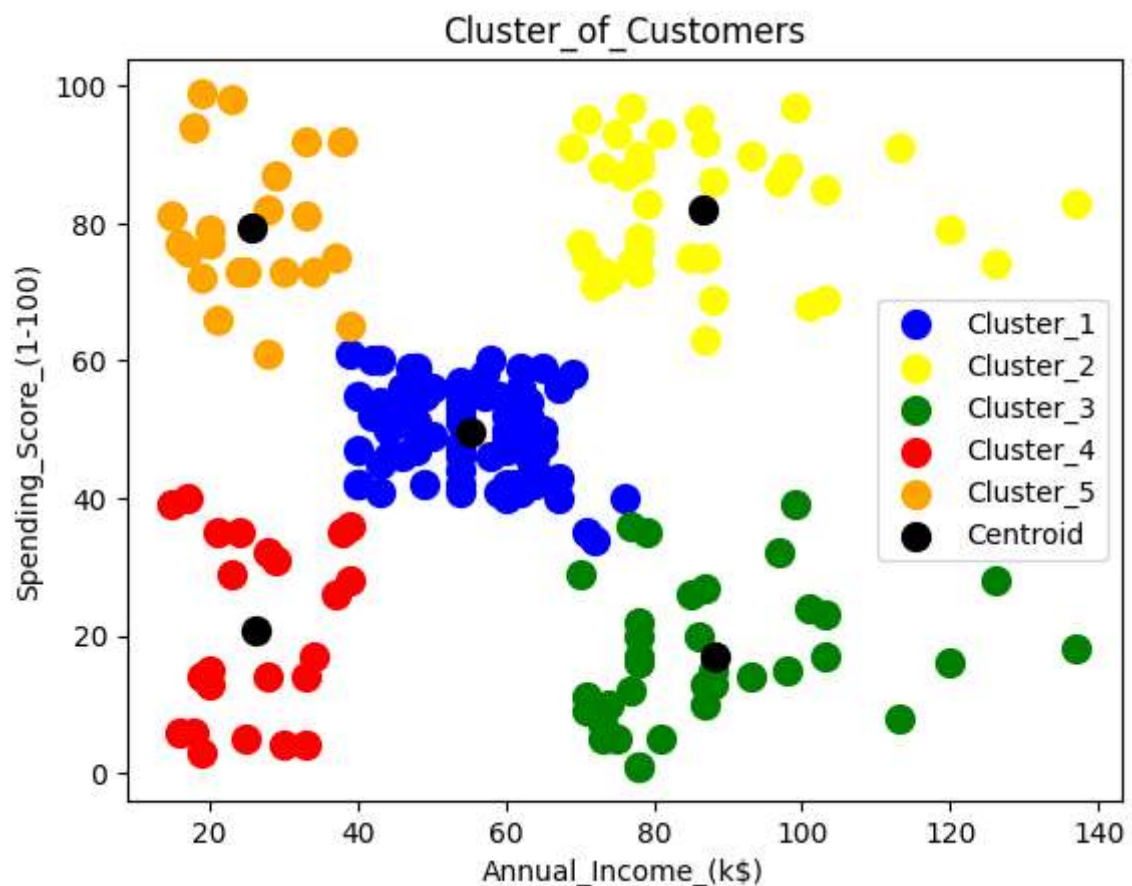
```
In [12]: y_pred
```

```
Out[12]: array([[3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4,
                3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 0,
                3, 4, 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, 1, 2, 1, 0, 1, 2, 1, 2, 1,
                0, 1, 2, 1, 2, 1, 2, 1, 2, 1, 0, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1,
                2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1,
                2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1,
                2, 1])
```

```
In [13]: plt.scatter(x[y_pred == 0, 0], x[y_pred == 0, 1], s=100, c="blue", label="Cluster 0")
plt.scatter(x[y_pred == 1, 0], x[y_pred == 1, 1], s=100, c="yellow", label="Cluster 1")
plt.scatter(x[y_pred == 2, 0], x[y_pred == 2, 1], s=100, c="green", label="Cluster 2")
plt.scatter(x[y_pred == 3, 0], x[y_pred == 3, 1], s=100, c="red", label="Cluster 3")
plt.scatter(x[y_pred == 4, 0], x[y_pred == 4, 1], s=100, c="orange", label="Cluster 4")

plt.scatter(kmeans.cluster_centers_[0, 0], kmeans.cluster_centers_[0, 1], s=100, c="black", label="Cluster 0 Center")

plt.title("Cluster of Customers")
plt.xlabel("Annual_Income_(k$)")
plt.ylabel("Spending_Score_(1-100)")
plt.legend()
plt.show()
```



```
In [14]: silhouette_score(x, y_pred)
```

```
Out[14]: 0.553931997444648
```

```
In [ ]:
```