

Customer Segmentation

April 29, 2023

1 Customer Segmentation using K-means Clustering

```
[1]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
%matplotlib inline
from warnings import filterwarnings
filterwarnings('ignore')
```

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

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

```
[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

1.1 for customer segmentation we will mainly focus on the income and the spending score ,

```
[5]: df.columns
```

```
[5]: Index(['CustomerID', 'Gender', 'Age', 'Annual Income (k$)',
        'Spending Score (1-100)'],
        dtype='object')
```

```
[38]: X = df[['Annual Income (k$)',
        'Spending Score (1-100)']]
```

```
[39]: X
```

```
[39]:
```

	Annual Income (k\$)	Spending Score (1-100)
0	15	39
1	15	81

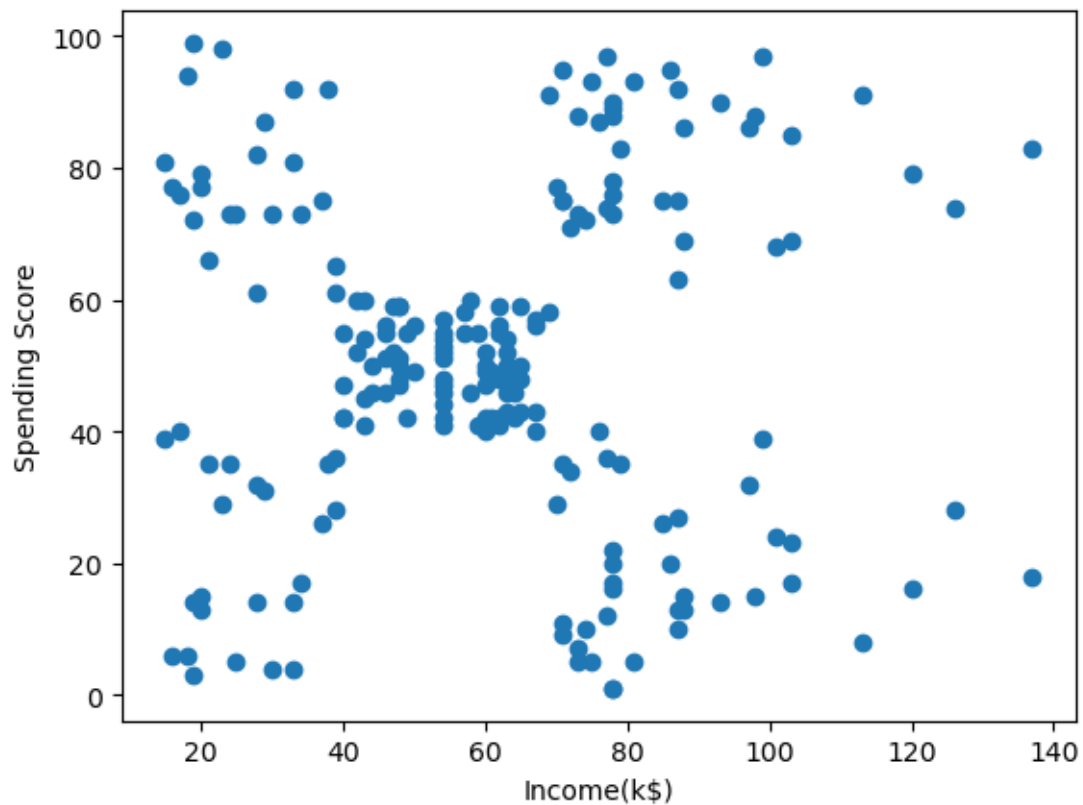
2	16	6
3	16	77
4	17	40
..
195	120	79
196	126	28
197	126	74
198	137	18
199	137	83

[200 rows x 2 columns]

1.2 lets visualize the scatter plot

```
[9]: plt.scatter(x=Income,y=Expenditure)
plt.xlabel("Income(k$)")
plt.ylabel('Spending Score')
```

```
[9]: Text(0, 0.5, 'Spending Score')
```



1.3 we will use K Means++ as our data in numerical.

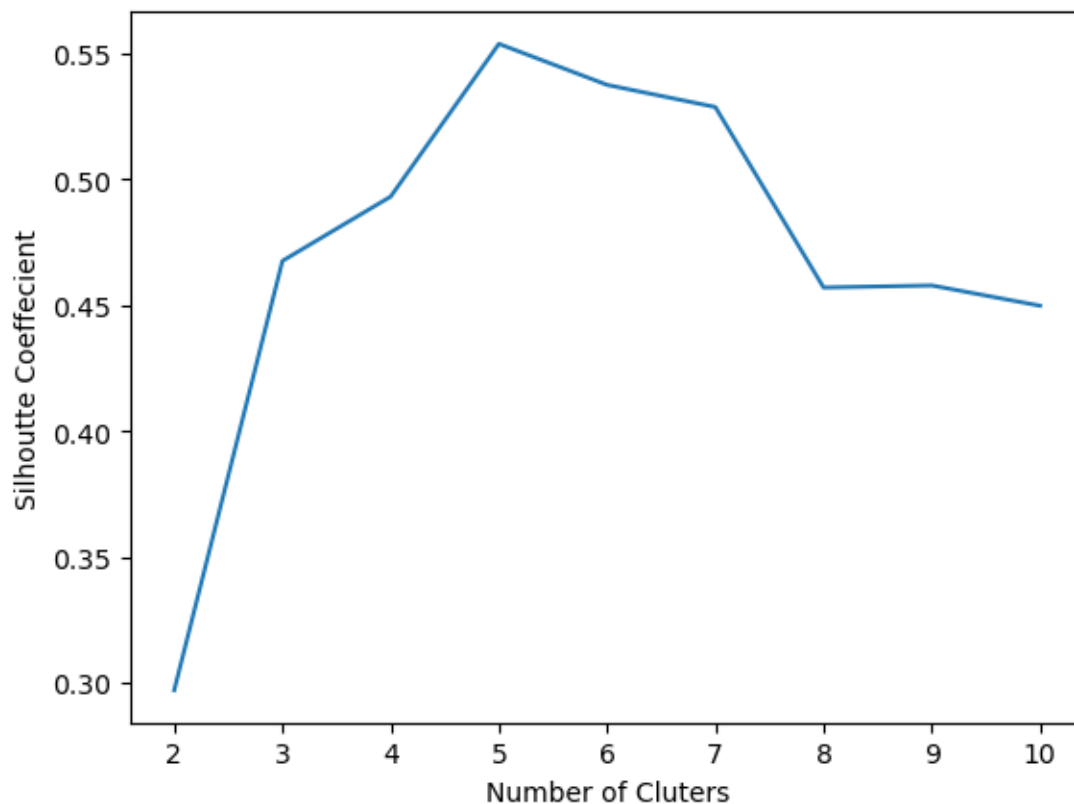
```
[11]: from sklearn.cluster import KMeans
```

```
[40]: X = X.to_numpy()
```

```
[53]: ## Silhoutte score  
from sklearn.metrics import silhouette_score
```

```
[56]: silhouette_coefficients=[]  
for k in range(2,11):  
    kmeans=KMeans(n_clusters=k,init="k-means++")  
    kmeans.fit(X)  
    score=silhouette_score(X,kmeans.labels_)  
    silhouette_coefficients.append(score)
```

```
[57]: ## plotting silhouette score  
plt.plot(range(2,11),silhouette_coefficients)  
plt.xticks(range(2,11))  
plt.xlabel("Number of Cluters")  
plt.ylabel("Silhoutte Coeffecient")  
plt.show()
```



1.4 as you can see the silhoeutte score is the highest for 5 clusters .So we will select n_cluster parameter as 5

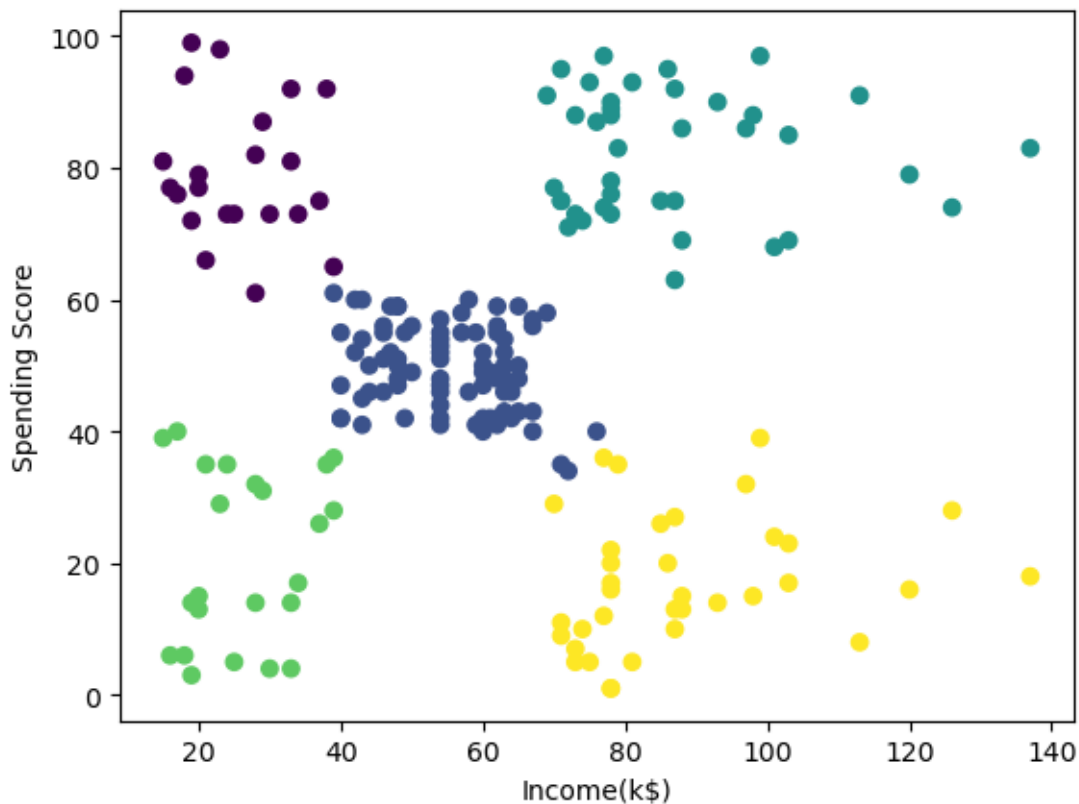
```
[ ]:
```

```
[47]: kmeans = KMeans(n_clusters=5,init='k-means++')
```

```
[48]: y=kmeans.fit_predict(X)
```

```
[58]: plt.scatter(X[:,0],X[:,1],c=y)
plt.xlabel("Income(k$)")
plt.ylabel('Spending Score')
```

```
[58]: Text(0, 0.5, 'Spending Score')
```



1.5 as you can see we have clustered the datapoints .

1.6 Now lets conclude somethings ->

- Light Green : people with less income and less expenses.

- Purple : People with less income bt high expenditures
- Blue : People with average income and Average Expenditures .
- Yellow : People with high income but less expenses .
- Deep Green : people with high income and high expenditures

1.7 In Total we have 5 categories of people .

1.8 based on the product a company launches e-commerce websites can recommend products to the people .

[]: