K-Means Clustering

Submitted By – Kartik Nagras

Dataset Explanation

Dataset link - https://github.com/SteffiPeTaffy/machineLearningAZ/blob/master/Machine%20Learning%20A-Z%20Template%20Folder/Part%204%20-%20Clustering/Section%2024%20-%20K-Means%20Clustering/Mall_Customers.csv

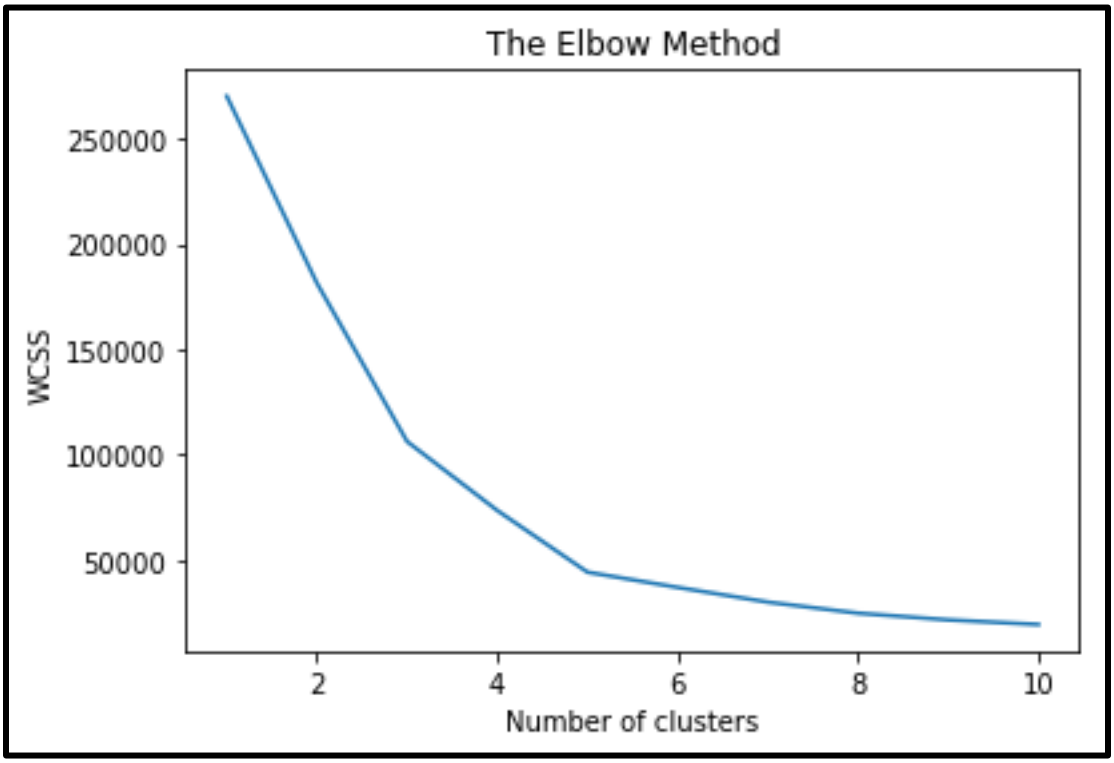
The data is about the Mall Customers which includes information about their **Spending Scores (1-100)** in the Mall with additional customer attributes like **Customer ID, Customer Age**, **Genre(Gender)**, **Annual Income.**

```
In [47]: dataset.head()
Out[47]:
   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
```

Business side of Customer clustering

So, the real question is how can we use this in our business? Based on the information from the clusters we can decide which strategy to run, what our target should be. There is also a possibility to conduct an survey on the potential new strategy for one or two segments of customers. Based on that feedback we can decide whether the new strategy is good for that customer segment or not, even before the strategy is released.

Using the Elbow method to find the optimal number of Clusters

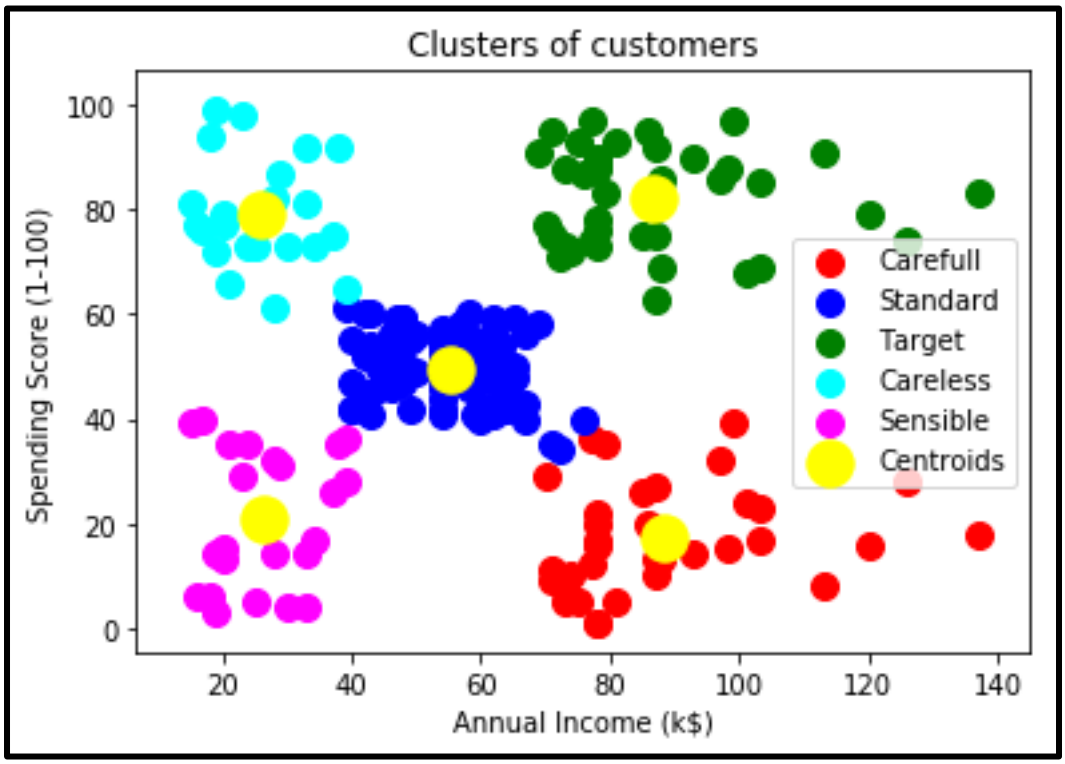


Optimal Number of Clusters =5

Fitting K-means Model to our Dataset.

```
# Fitting K-Means to the dataset
kmeans = KMeans(n_clusters = 5, init = 'k-means++', random_state = 0)
y_kmeans = kmeans.fit_predict(X)
```

Visualizing Our Cluster.



Analysing Our Clusters.

Cluster Categories	Annual Income	Spending Score
Careful Customers	High	Low
Standard Customers	Average	Average
Target Customers	High	High
Careless Customers	Low	High
Sensible Customers	Low	Low