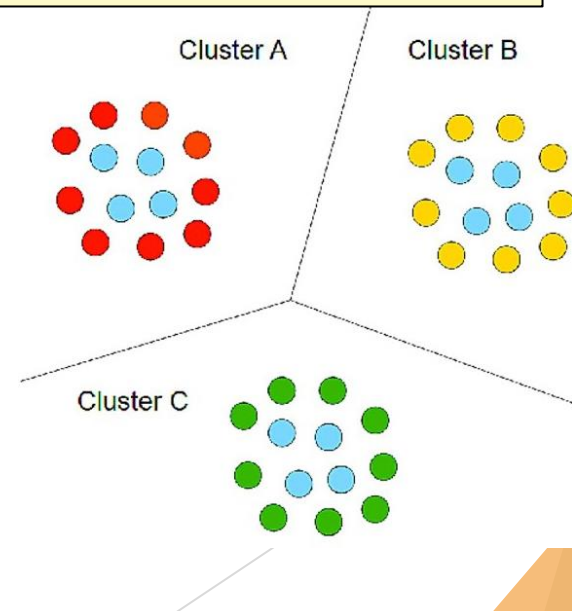


Customer Segmentation using K Means Clustering

Machine Learning Projects



Introduction

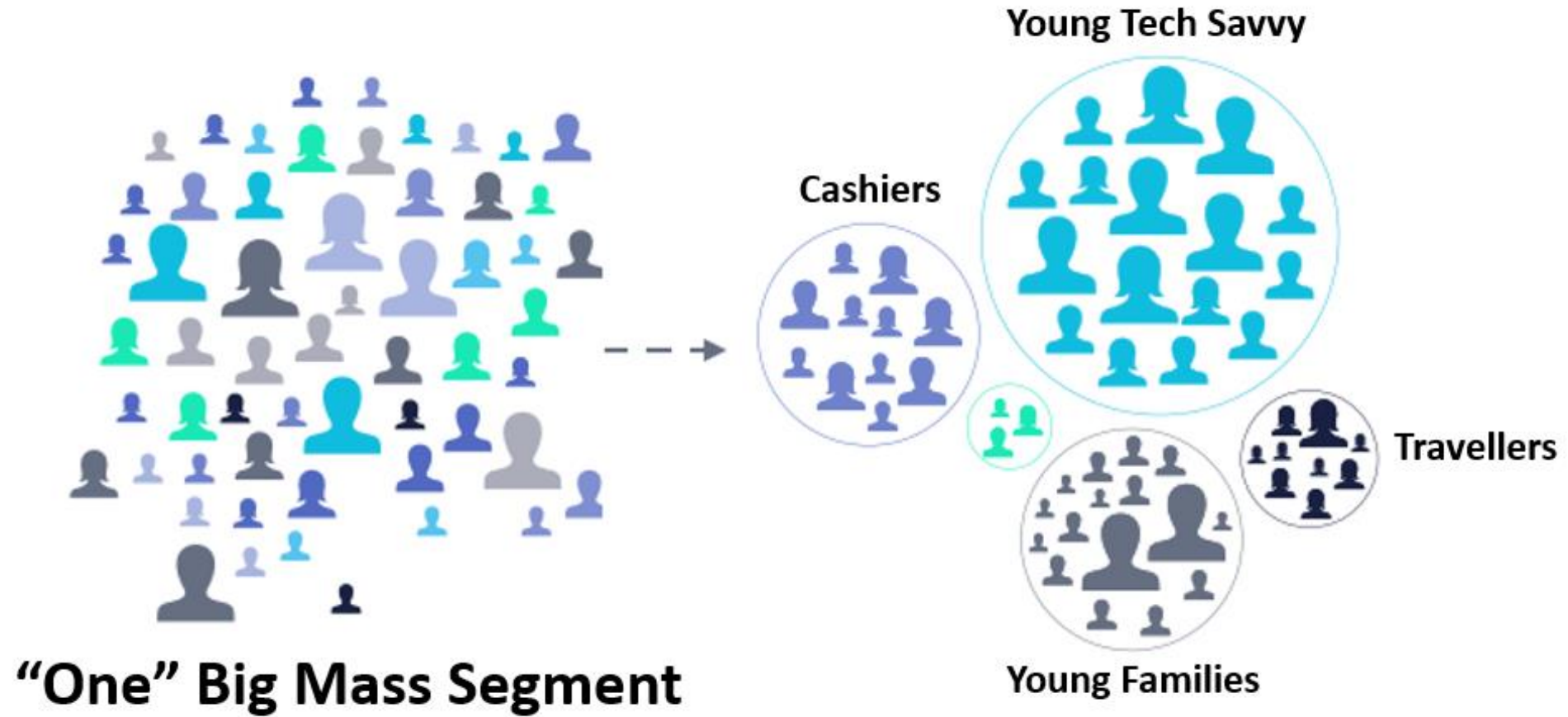
- Customer Segmentation is the subdivision of a market into discrete customer groups that share similar characteristics. Customer Segmentation can be a powerful means to identify unsatisfied customer needs.
- Customer Segmentation is an unsupervised method of targeting the customers in order to increase sales and market goods in a better way
- With the help of the algorithms, we can easily visualize the data and can get a segmentation of each customer so that we can target the customers in the better way

Advantages of Customer Segmentation:

- Determine appropriate product pricing.
- Develop customized marketing campaigns.
- Design an optimal distribution strategy.
- Choose specific product features for deployment.
- Prioritize new product development efforts.



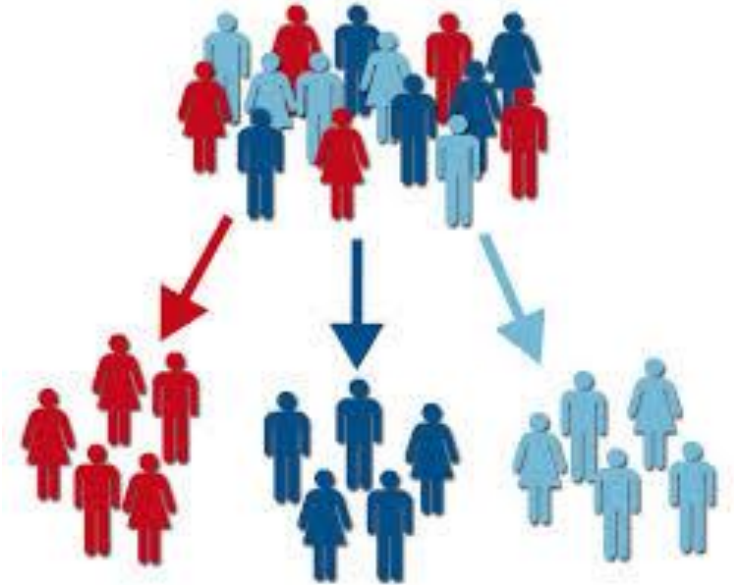
Clustering



Problem Statement



amazon



1. Getting Customer Data
2. Data Preprocessing
3. Data Analysis
4. Finding Number of Cluster
5. Importing Cluster Model
6. Visualiazing Clusters



Work Flow



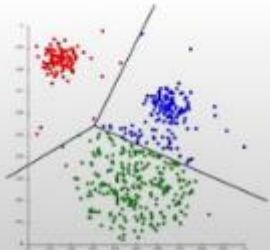
Customer
Data



Data pre processing



Data Analysis



Visualizing the Clusters



K-Means Clustering

$$WCSS = \sum_{i \in n} (X_i - Y_i)^2$$

Optimum number of Clusters

Analysing Data

Printing first 15 row of Customer Data:-

	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
10	11	Male	67	19	14
11	12	Female	35	19	99
12	13	Female	58	20	15
13	14	Female	24	20	77
14	15	Male	37	20	13

Data Preprocessing

Getting information about null values of data:-

```
<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
```

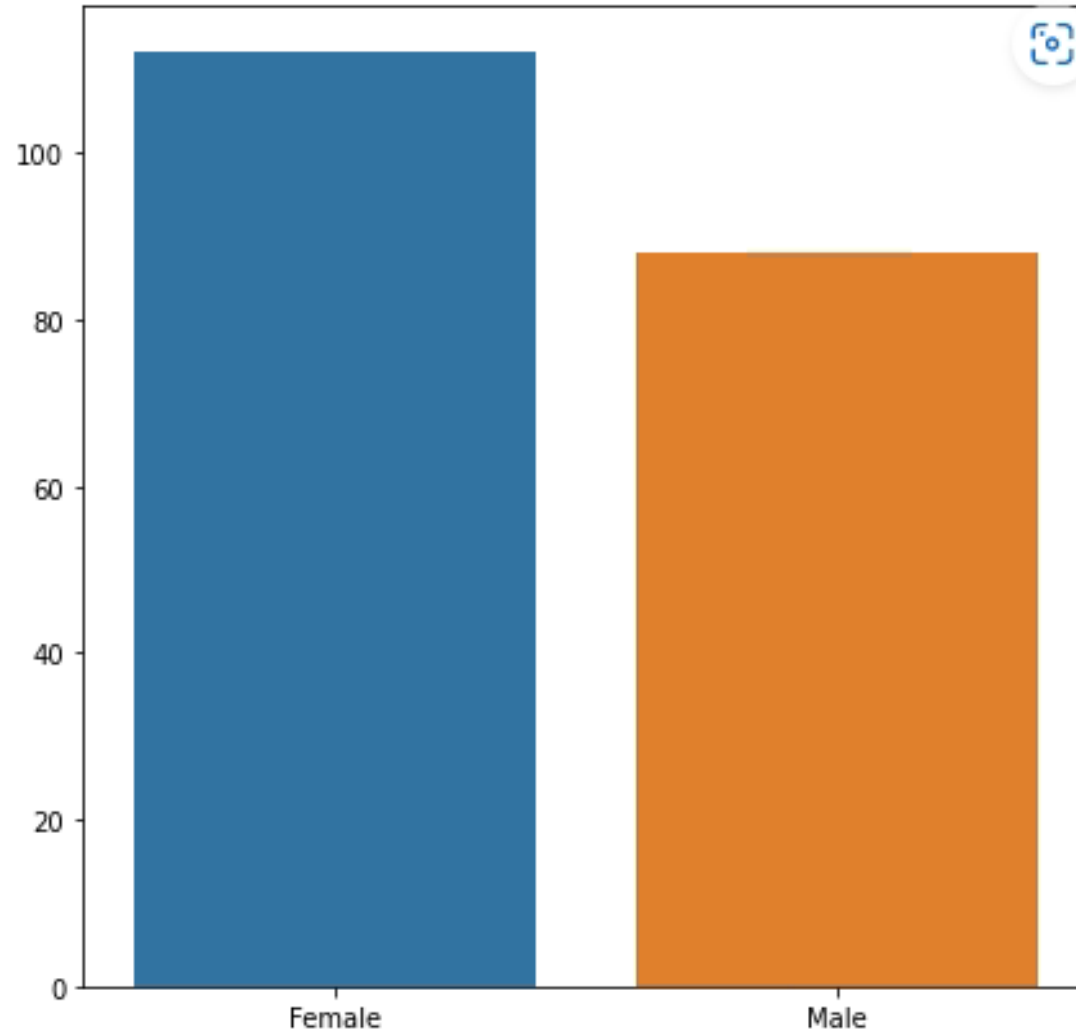
Getting Sum of null values of data:-

```
customerData.isnull().sum()

CustomerID      0
Gender          0
Age             0
Annual Income (k$)  0
Spending Score (1-100)  0
dtype: int64
```

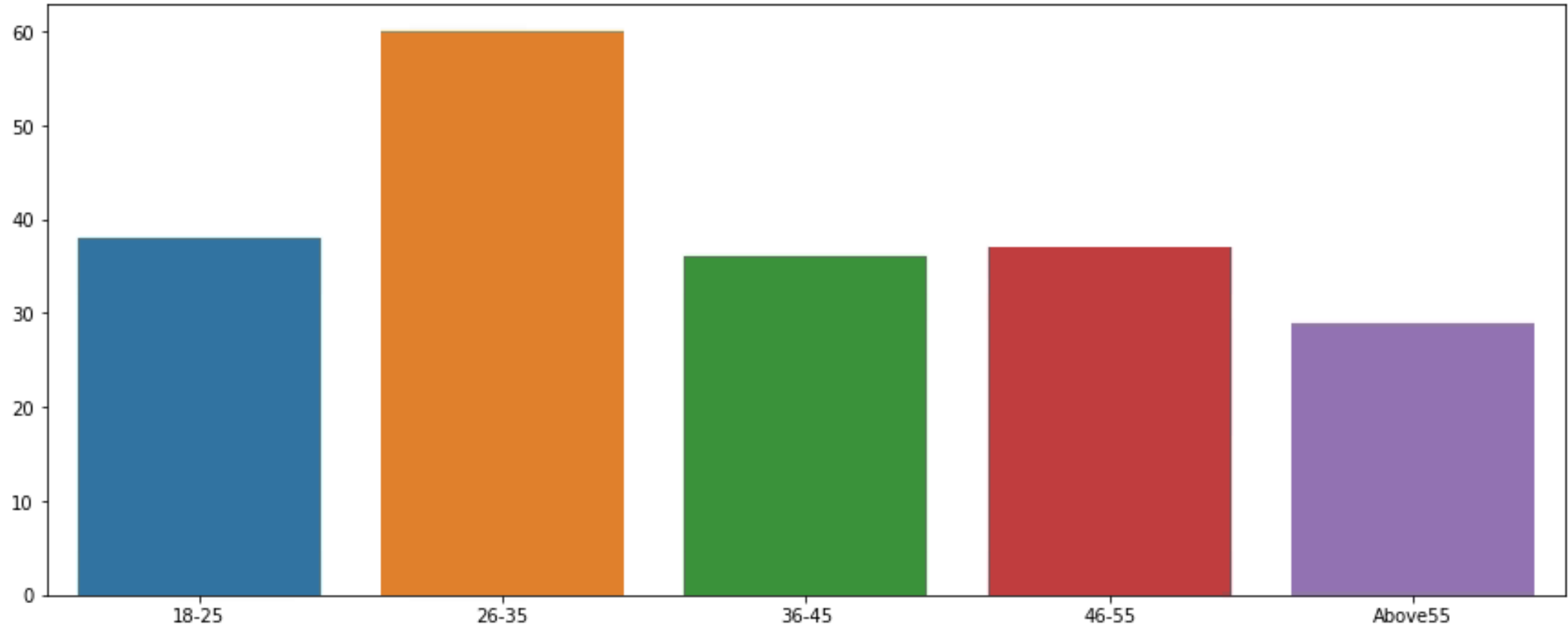

Visualizing Data Set

Bar Plot (Genders vs No of Customers):-



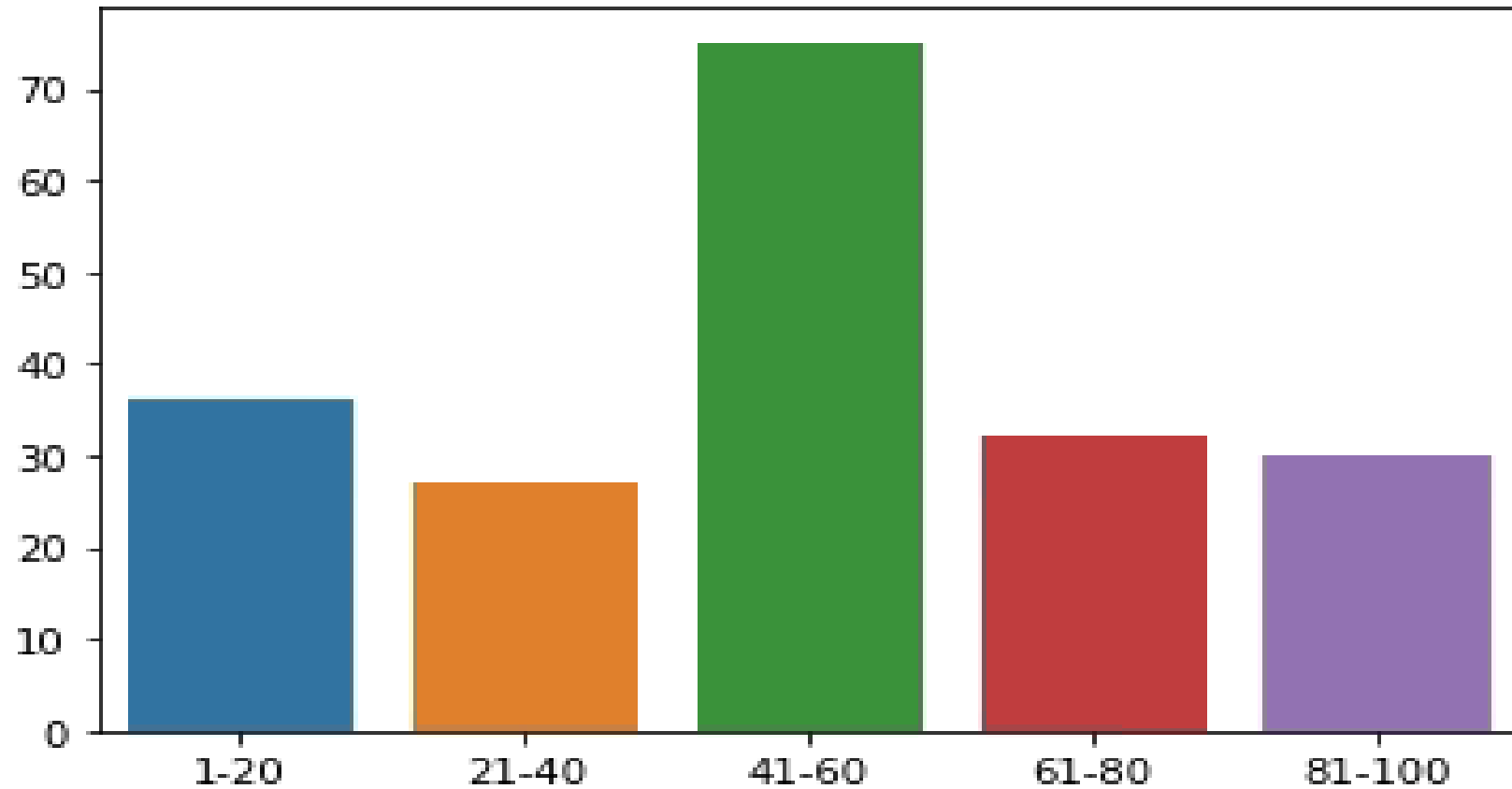
Visualizing Data Set

Bar Plot (Age vs No of Customers):-



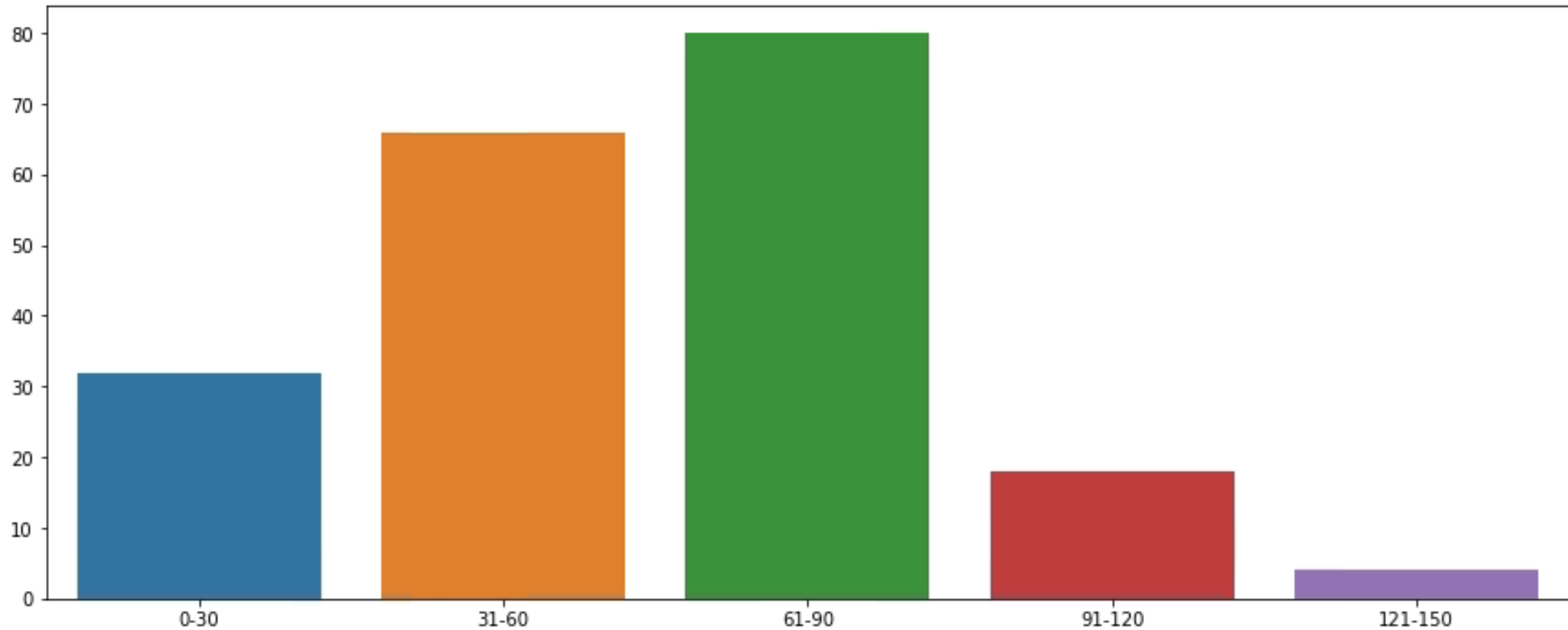
Visualizing Data Set

Bar Plot (Spending Score vs No of Customers):-



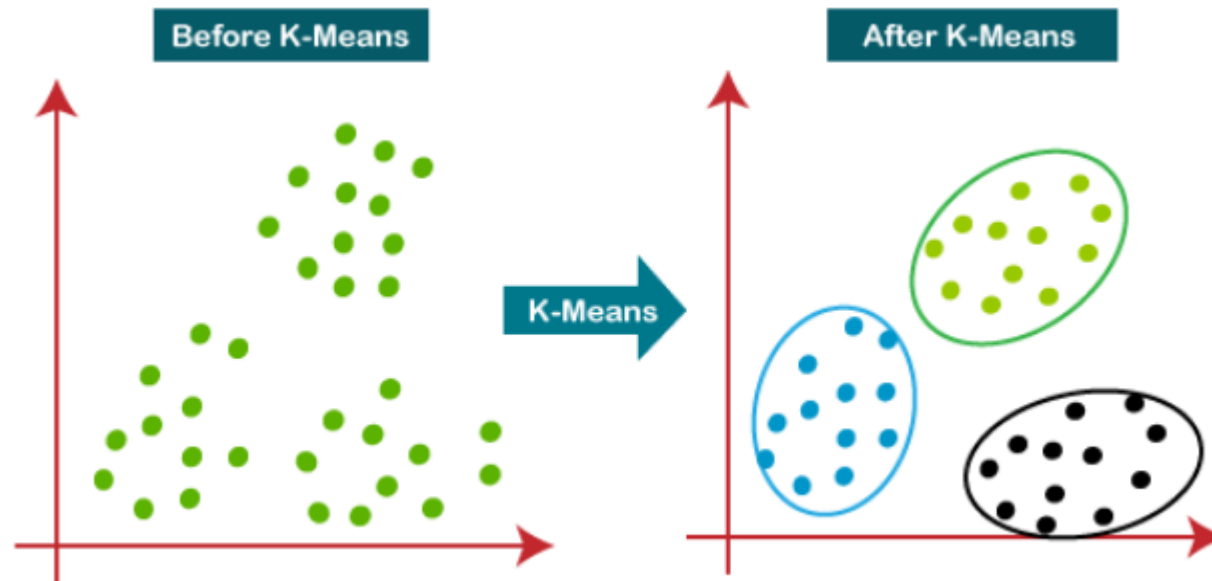
Visualizing Data Set

Bar Plot (Annual Income vs No of Customers):-



K Means Clustering Model

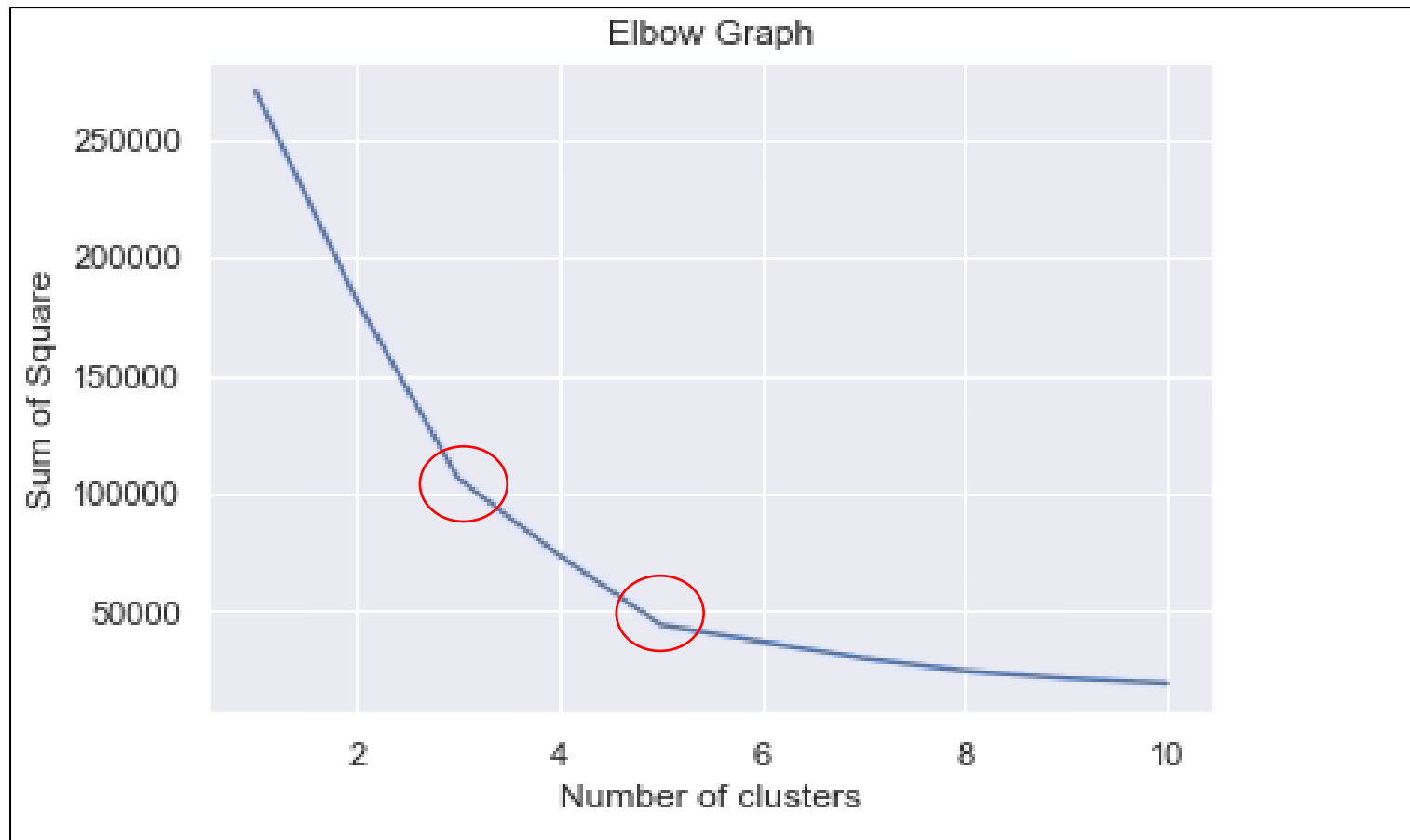
K-Means Clustering is an Unsupervised Learning Algorithm, which groups the unlabeled dataset into different clusters with similarity.



K Means Clustering Model

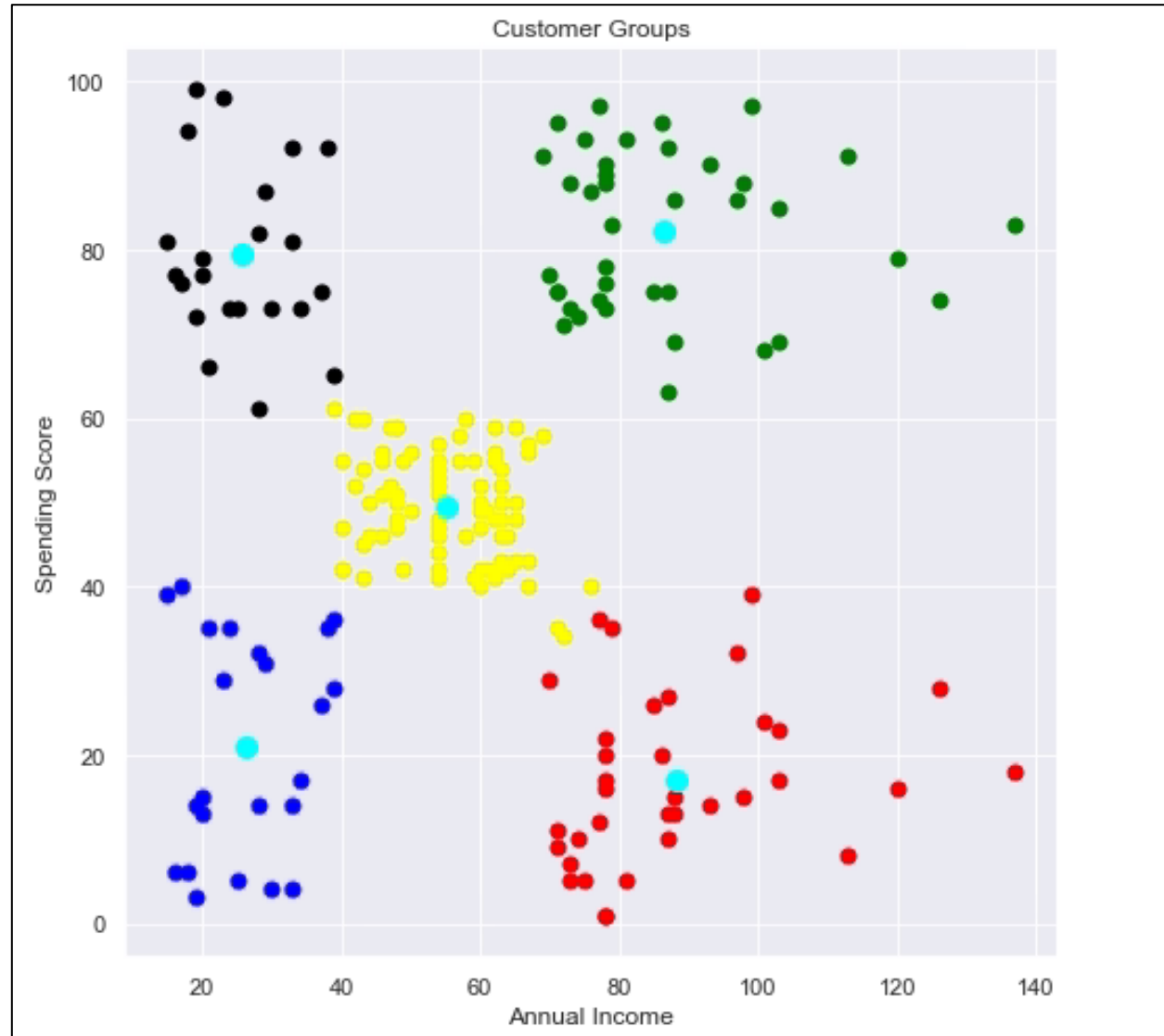
Using ELBOW method to find no of clusters:-

$$WCSS = \sum_{i \in n} (X_i - Y_i)^2$$



Visualizing the Clusters

Scatter plot of Data After Clustering customer into the groups:-



Thank You