Customer Segmentation

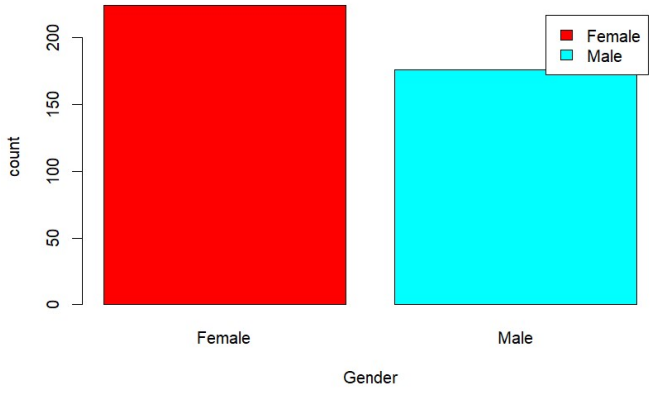
Customer segmentation means splitting customers into different groups. These groups are based on traits like age, buying habits, or preferences. The aim is to understand customers better and make marketing more relevant.

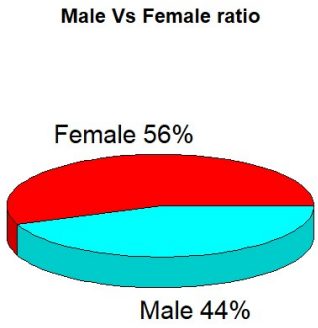
Project Objective:

In this project, we aim to implement value-based segmentation, which categorizes customers according to their economic value, thereby grouping those with similar value levels into distinct segments for targeted marketing efforts.

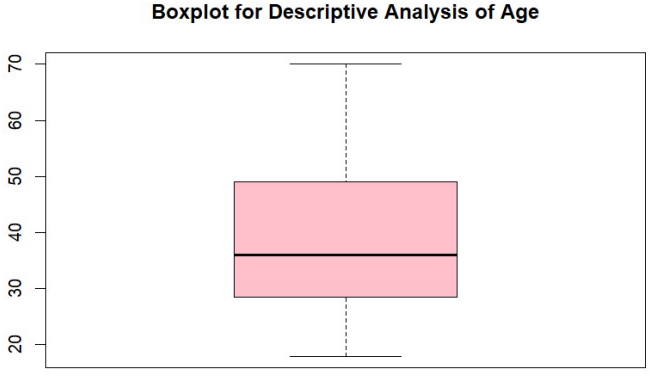
Results:

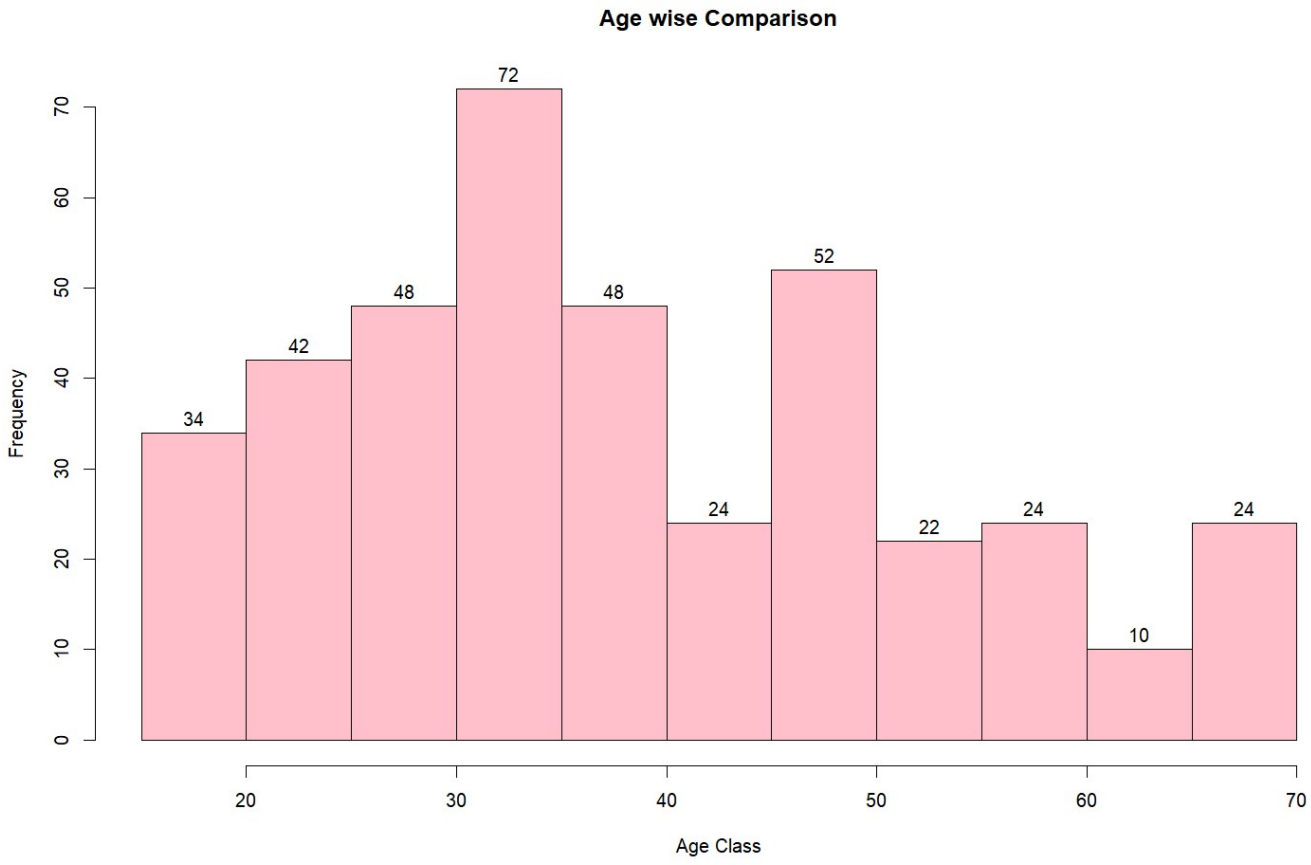
- We observe that the number of females spending rate is higher than the males.



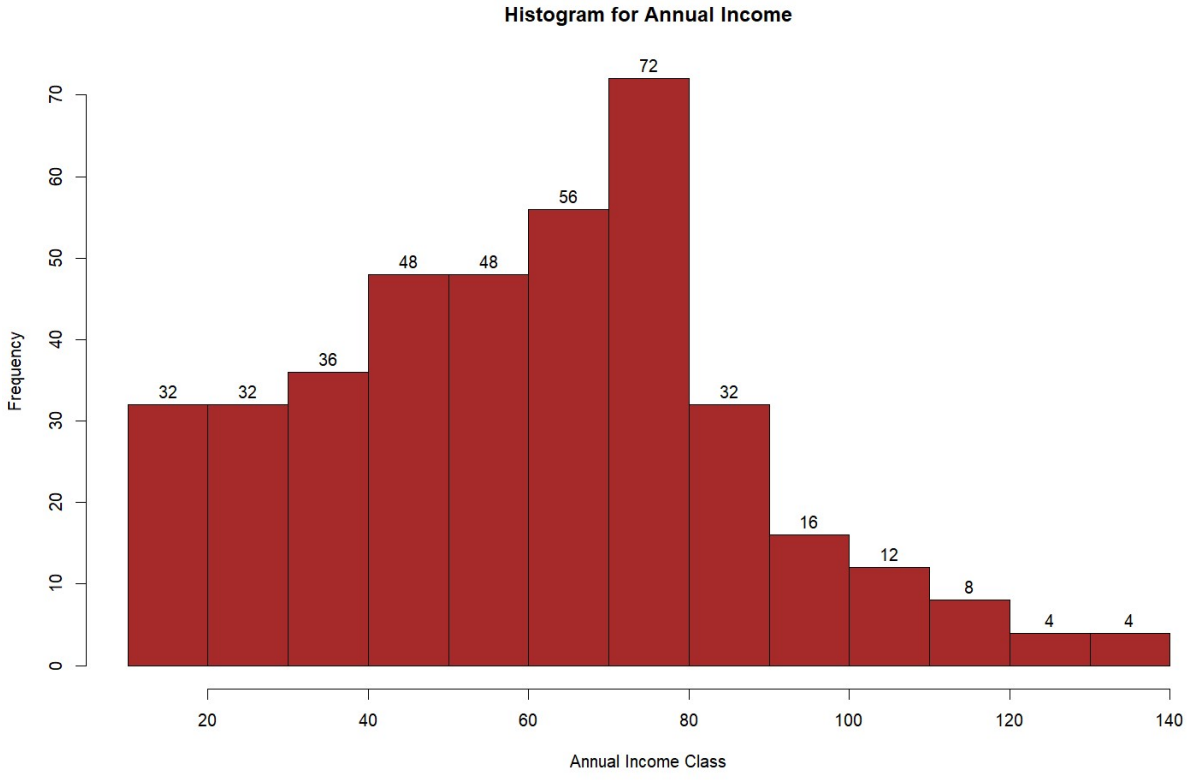


- Age group between 30-35 shows high frequency.

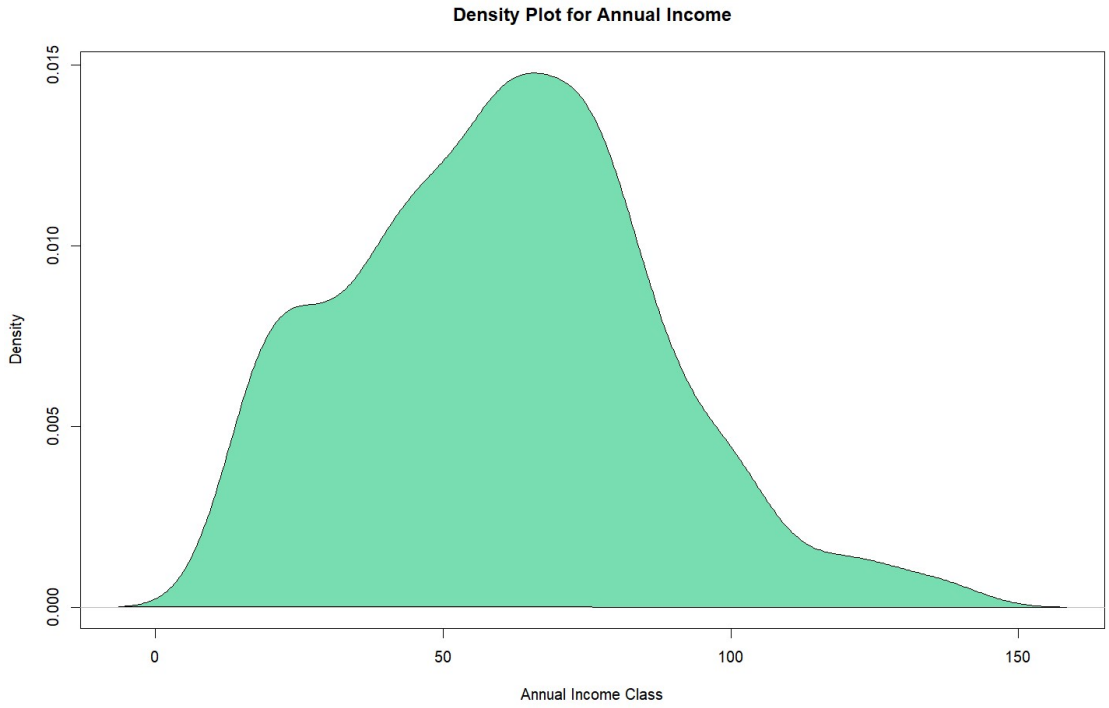




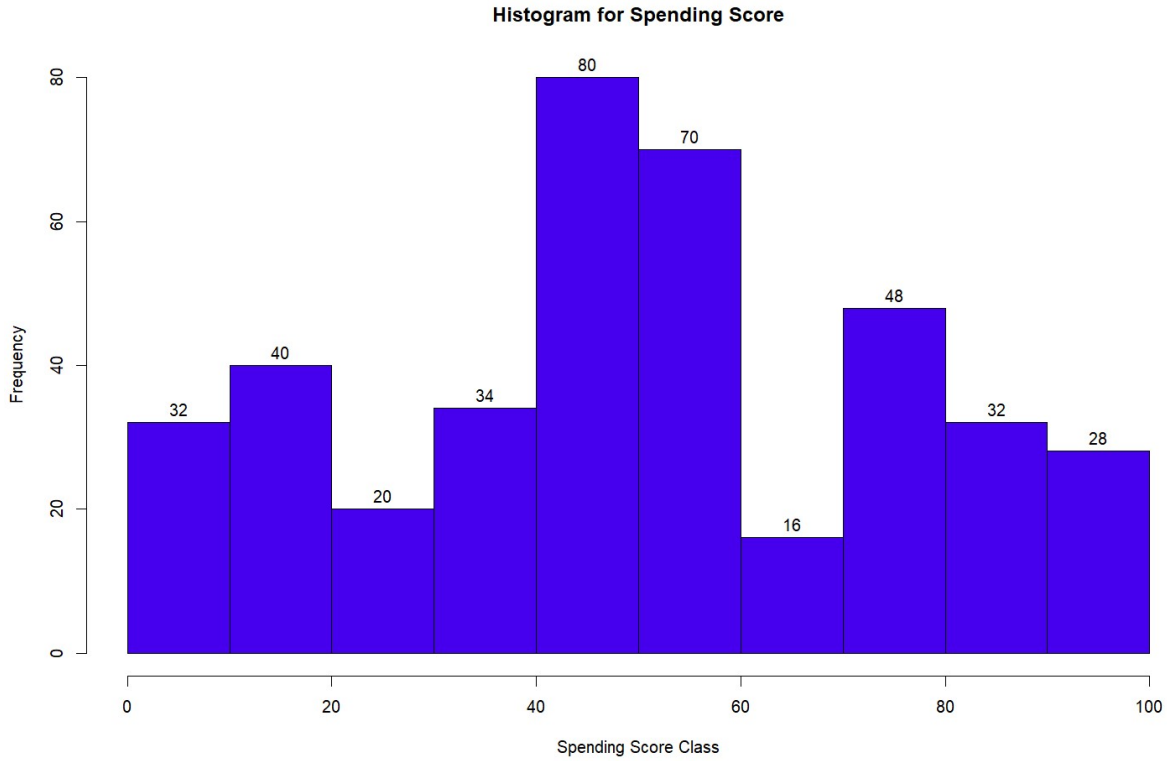
- Customers with an Annual Income of 70k$-80k$ shows higher frequency rate.



- In the Kernel Density Plot that we displayed above, we observe that the annual income has a normal distribution.

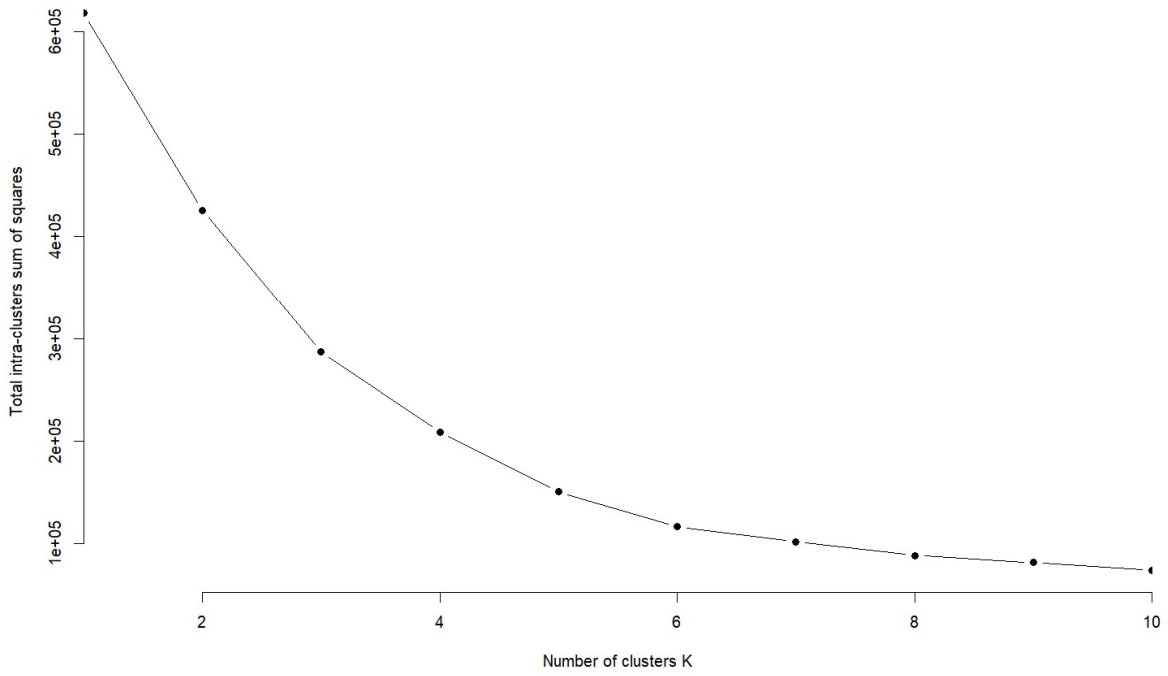


- From the histogram, we conclude that customers between class 40 and 50 have the highest frequency rate among all the classes.



K-MEAN ALGORITHM

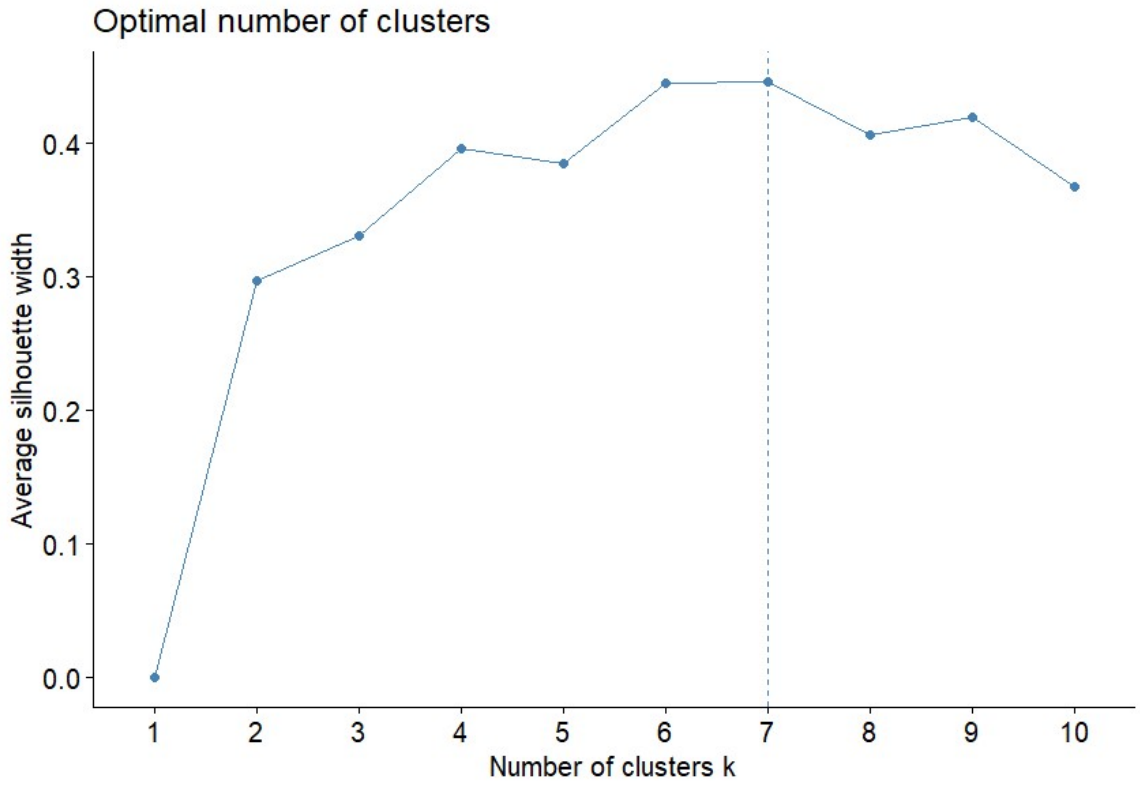
Elbow Method:



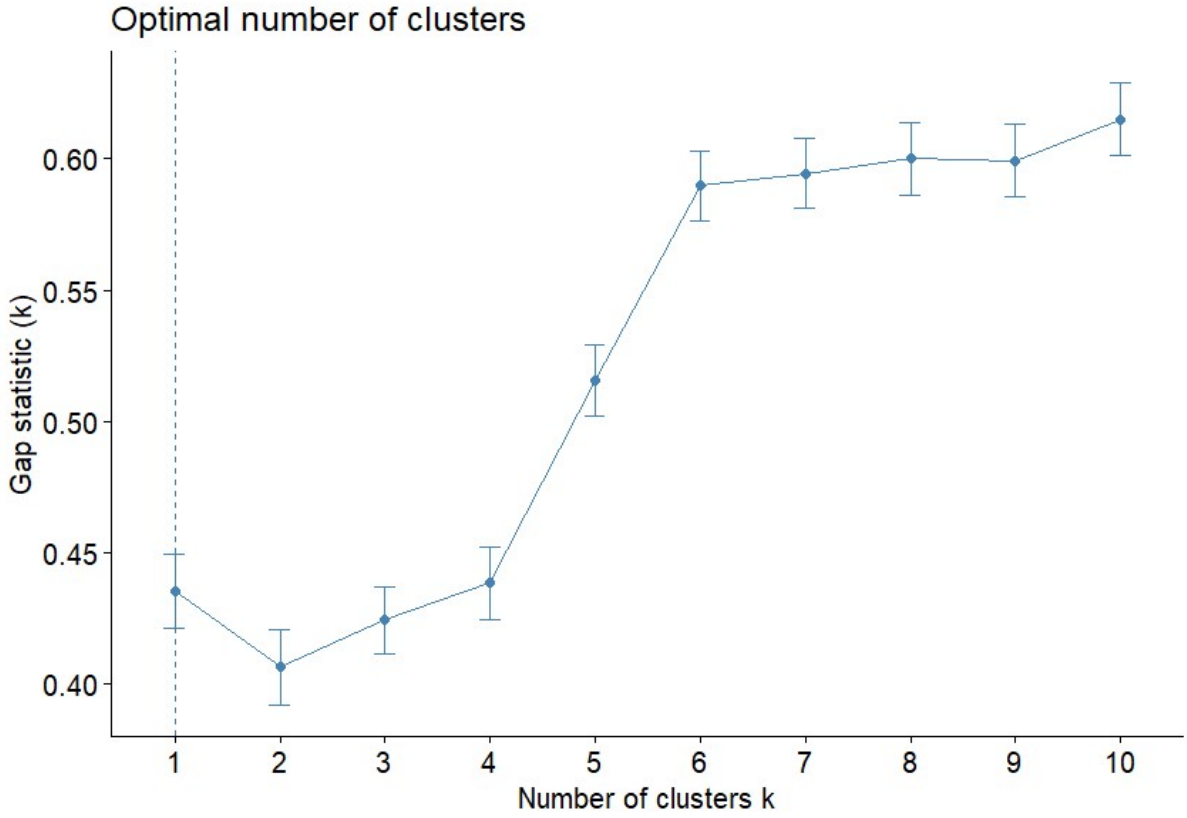
- we conclude that **6** is the appropriate number of clusters since it seems to be appearing at the bend in the elbow plot.

Average Silhouette Method:

- Using the silhouette function in the cluster package, we can compute the average silhouette width using the k-mean function. Here, **7** is the optimal cluster will possess highest average.



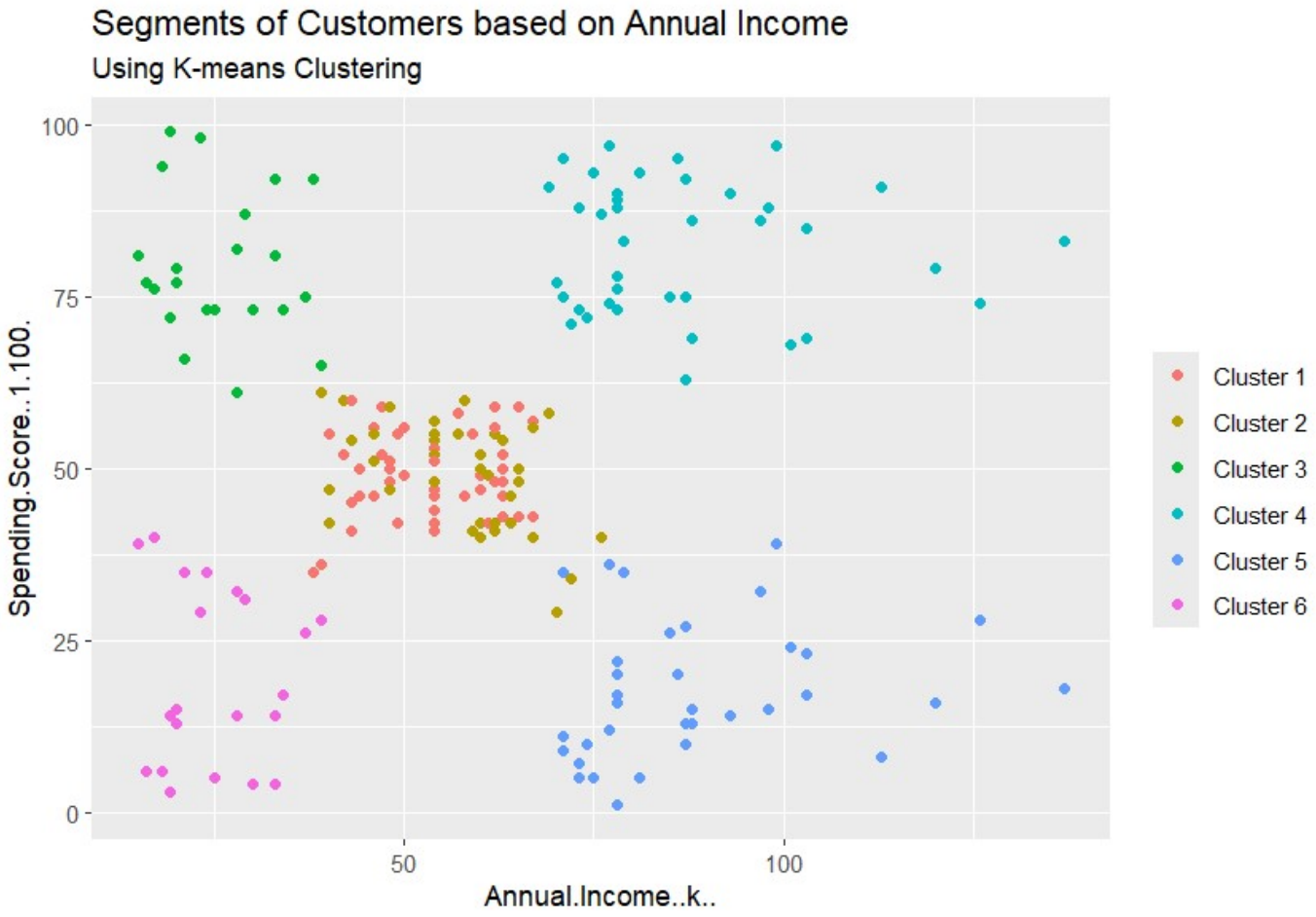
Gap Statistic Method:



- From the above results (Elbow Method, Average Silhouette Method and Gap Statistic Method), we choose **6** has optimal cluster.

From the above visualization, we observe that there is a distribution of 6 clusters as follows :

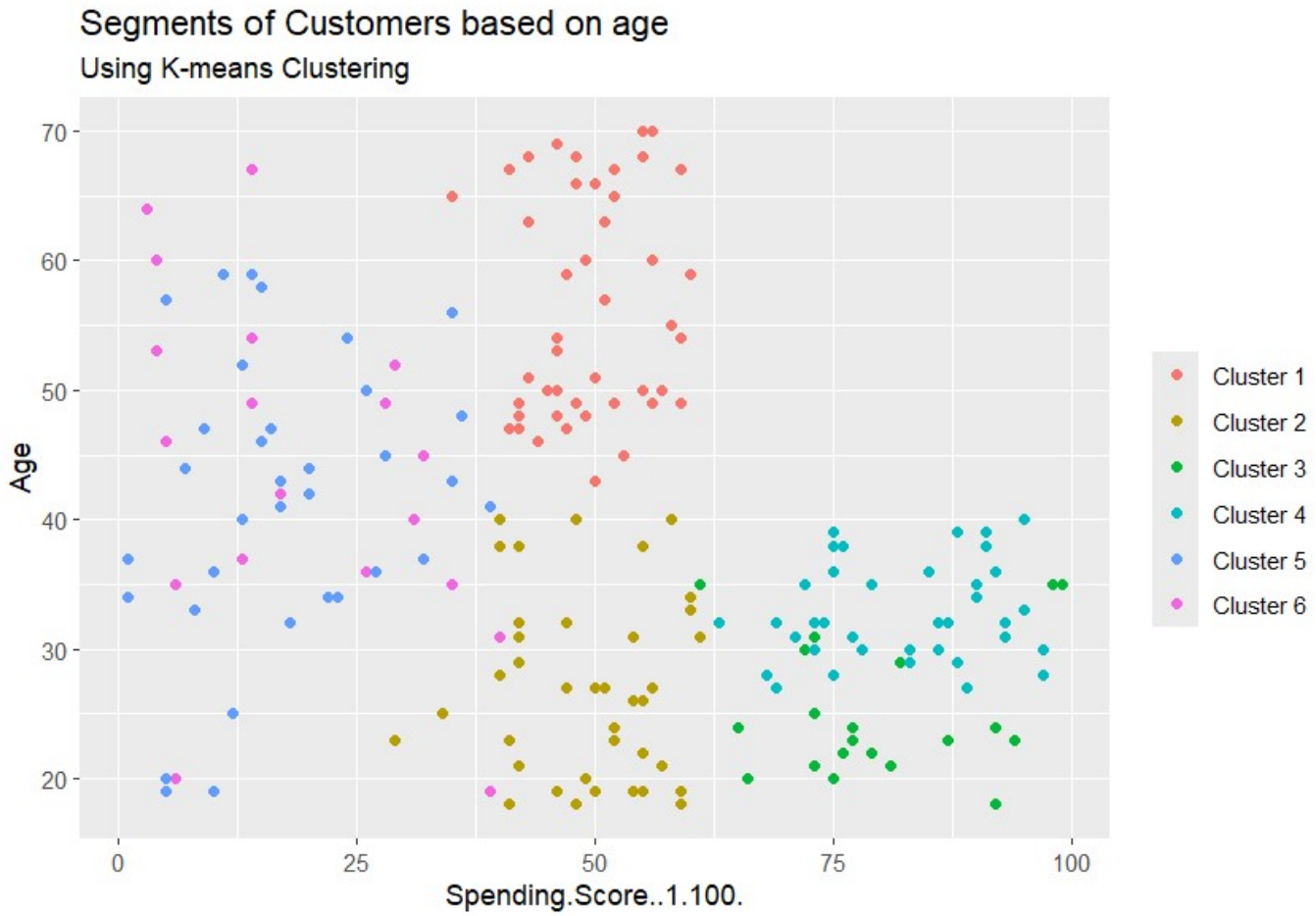
Cluster Analysis between Annual Income and Spending score



Analysis:

- **Cluster 1 and 2 -** These clusters represent the customer data with the medium income salary as well as the medium annual spend of salary.
- **Cluster 3 -** This cluster represents a low annual income but its high yearly expenditure.
- **Cluster 4 -** This cluster represents the customer data having a high annual income as well as a high annual spend.
- **Cluster 5 -** This cluster denotes a high annual income and low yearly spend.
- **Cluster 6 -** This cluster denotes the customer data with low annual income as well as low yearly spend of income.

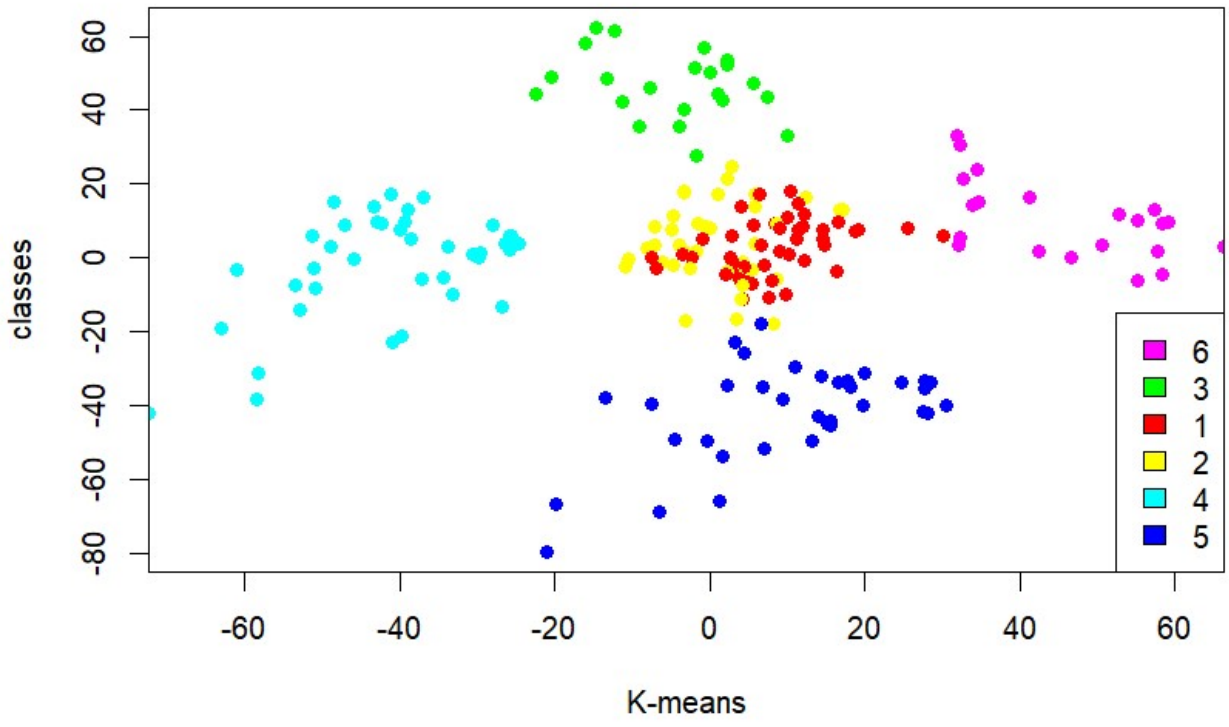
Cluster Analysis between Age and Spending score



Analysis:

- **Cluster 1 -** This clusters represents customers with an average annual income (above 25 - below 60), average spending score(above 25 and below 60) and ages above 40 years.
- **Cluster 2 -** This clusters represents customers with an average annual income (above 30 below 80), average spending score (above 25 and below 62) and ages of 40 and below.
- **Cluster 3 -** This cluster represents a low annual income but its high yearly expenditure with ages (below 20 – below 35).
- **Cluster 4 –** This cluster represents the customers having a high annual income as well as a high annual spend with an age between age 28 and 40.
- **Cluster 5 -** This cluster denotes a high annual income and low yearly spending score with with ages spreading throughout the spectrum.
- **Cluster 6 -** This cluster denotes the customers with low annual income as well as low yearly spend of income with ages spreading throughout the spectrum.

K – Means Clusters



- **Cluster 1 and 2** - These two clusters consist of customers with medium PCA1 and medium PCA2 score.
- **Cluster 3** - This comprises of customers with a high PCA2 and a medium annual spend of income.
- **Cluster 4 -** This cluster represents customers having a high PCA2 and a low PCA1.
- **Cluster** 5 - In this cluster, there are customers with a medium PCA1 and a low PCA2 score.
- **Cluster 6** - This cluster comprises of customers with a high PCA1 income and a high PCA2.