Market-Sectionalization

Nirmal Sai Swaroop Janapaneedi

24/04/2020

Introduction

Market Sectionalization also referred to as “Customer Segmentation” is one of the most important applications of unsupervised learning. Using clustering techniques, companies can identify the several segments of customers allowing them to target the potential user base. In this machine learning project, we will make use of K-means clustering which is the essential algorithm for clustering unlabeled dataset. Before ahead in this project, learn what actually customer segmentation is.

What is Market Sectionalization/Customer Segmentation ?

Market Segmentation is the process of division of customer base into several groups of individuals that share a similarity in different ways that are relevant to marketing such as gender, age, interests, and miscellaneous spending habits.

Companies that deploy customer segmentation are under the notion that every customer has different requirements and require a specific marketing effort to address them appropriately. Companies aim to gain a deeper approach of the customer they are targeting. Therefore, their aim has to be specific and should be tailored to address the requirements of each and every individual customer.

Furthermore, through the data collected, companies can gain a deeper understanding of customer preferences as well as the requirements for discovering valuable segments that would reap them maximum profit.

This way, they can strategize their marketing techniques more efficiently and minimize the possibility of risk to their investment.

The technique of customer segmentation is dependent on several key differentiators that divide customers into groups to be targeted. Data related to demographics, geography, economic status as well as behavioral patterns play a crucial role in determining the company direction towards addressing the various segments.

Importing Dataset

```
##     CustomerID Gender Age Annual.Income..k.. Spending.Score..1.100.
## 1            1   Male  19                 15                     39
## 2            2   Male  21                 15                     81
## 3            3 Female  20                 16                      6
## 4            4 Female  23                 16                     77
## 5            5 Female  31                 17                     40
## 6            6 Female  22                 17                     76
```

```
## 7            7 Female  35                  18                      6
## 8            8 Female  23                  18                     94
## 9            9   Male  64                  19                      3
## 10          10 Female  30                  19                     72
## 11          11   Male  67                  19                     14
## 12          12 Female  35                  19                     99
## 13          13 Female  58                  20                     15
## 14          14 Female  24                  20                     77
## 15          15   Male  37                  20                     13
## 16          16   Male  22                  20                     79
## 17          17 Female  35                  21                     35
## 18          18   Male  20                  21                     66
## 19          19   Male  52                  23                     29
## 20          20 Female  35                  23                     98
## 21          21   Male  35                  24                     35
## 22          22   Male  25                  24                     73
## 23          23 Female  46                  25                      5
## 24          24   Male  31                  25                     73
## 25          25 Female  54                  28                     14
## 26          26   Male  29                  28                     82
## 27          27 Female  45                  28                     32
## 28          28   Male  35                  28                     61
## 29          29 Female  40                  29                     31
## 30          30 Female  23                  29                     87
## 31          31   Male  60                  30                      4
## 32          32 Female  21                  30                     73
## 33          33   Male  53                  33                      4
## 34          34   Male  18                  33                     92
## 35          35 Female  49                  33                     14
## 36          36 Female  21                  33                     81
## 37          37 Female  42                  34                     17
## 38          38 Female  30                  34                     73
## 39          39 Female  36                  37                     26
## 40          40 Female  20                  37                     75
## 41          41 Female  65                  38                     35
## 42          42   Male  24                  38                     92
## 43          43   Male  48                  39                     36
## 44          44 Female  31                  39                     61
## 45          45 Female  49                  39                     28
## 46          46 Female  24                  39                     65
## 47          47 Female  50                  40                     55
## 48          48 Female  27                  40                     47
## 49          49 Female  29                  40                     42
## 50          50 Female  31                  40                     42
## 51          51 Female  49                  42                     52
## 52          52   Male  33                  42                     60
## 53          53 Female  31                  43                     54
## 54          54   Male  59                  43                     60
## 55          55 Female  50                  43                     45
## 56          56   Male  47                  43                     41
```

```
## 57           57 Female  51                  44                       50
## 58           58   Male  69                  44                       46
## 59           59 Female  27                  46                       51
## 60           60   Male  53                  46                       46
## 61           61   Male  70                  46                       56
## 62           62   Male  19                  46                       55
## 63           63 Female  67                  47                       52
## 64           64 Female  54                  47                       59
## 65           65   Male  63                  48                       51
## 66           66   Male  18                  48                       59
## 67           67 Female  43                  48                       50
## 68           68 Female  68                  48                       48
## 69           69   Male  19                  48                       59
## 70           70 Female  32                  48                       47
## 71           71   Male  70                  49                       55
## 72           72 Female  47                  49                       42
## 73           73 Female  60                  50                       49
## 74           74 Female  60                  50                       56
## 75           75   Male  59                  54                       47
## 76           76   Male  26                  54                       54
## 77           77 Female  45                  54                       53
## 78           78   Male  40                  54                       48
## 79           79 Female  23                  54                       52
## 80           80 Female  49                  54                       42
## 81           81   Male  57                  54                       51
## 82           82   Male  38                  54                       55
## 83           83   Male  67                  54                       41
## 84           84 Female  46                  54                       44
## 85           85 Female  21                  54                       57
## 86           86   Male  48                  54                       46
## 87           87 Female  55                  57                       58
## 88           88 Female  22                  57                       55
## 89           89 Female  34                  58                       60
## 90           90 Female  50                  58                       46
## 91           91 Female  68                  59                       55
## 92           92   Male  18                  59                       41
## 93           93   Male  48                  60                       49
## 94           94 Female  40                  60                       40
## 95           95 Female  32                  60                       42
## 96           96   Male  24                  60                       52
## 97           97 Female  47                  60                       47
## 98           98 Female  27                  60                       50
## 99           99   Male  48                  61                       42
## 100         100   Male  20                  61                       49
## 101         101 Female  23                  62                       41
## 102         102 Female  49                  62                       48
## 103         103   Male  67                  62                       59
## 104         104   Male  26                  62                       55
## 105         105   Male  49                  62                       56
## 106         106 Female  21                  62                       42
```

```
## 107          107 Female  66                                    63                                                      50
## 108          108   Male  54                                    63                                                      46
## 109          109   Male  68                                    63                                                      43
## 110          110   Male  66                                    63                                                      48
## 111          111   Male  65                                    63                                                      52
## 112          112 Female  19                                    63                                                      54
## 113          113 Female  38                                    64                                                      42
## 114          114   Male  19                                    64                                                      46
## 115          115 Female  18                                    65                                                      48
## 116          116 Female  19                                    65                                                      50
## 117          117 Female  63                                    65                                                      43
## 118          118 Female  49                                    65                                                      59
## 119          119 Female  51                                    67                                                      43
## 120          120 Female  50                                    67                                                      57
## 121          121   Male  27                                    67                                                      56
## 122          122 Female  38                                    67                                                      40
## 123          123 Female  40                                    69                                                      58
## 124          124   Male  39                                    69                                                      91
## 125          125 Female  23                                    70                                                      29
## 126          126 Female  31                                    70                                                      77
## 127          127   Male  43                                    71                                                      35
## 128          128   Male  40                                    71                                                      95
## 129          129   Male  59                                    71                                                      11
## 130          130   Male  38                                    71                                                      75
## 131          131   Male  47                                    71                                                       9
## 132          132   Male  39                                    71                                                      75
## 133          133 Female  25                                    72                                                      34
## 134          134 Female  31                                    72                                                      71
## 135          135   Male  20                                    73                                                       5
## 136          136 Female  29                                    73                                                      88
## 137          137 Female  44                                    73                                                       7
## 138          138   Male  32                                    73                                                      73
## 139          139   Male  19                                    74                                                      10
## 140          140 Female  35                                    74                                                      72
## 141          141 Female  57                                    75                                                       5
## 142          142   Male  32                                    75                                                      93
## 143          143 Female  28                                    76                                                      40
## 144          144 Female  32                                    76                                                      87
## 145          145   Male  25                                    77                                                      12
## 146          146   Male  28                                    77                                                      97
## 147          147   Male  48                                    77                                                      36
## 148          148 Female  32                                    77                                                      74
## 149          149 Female  34                                    78                                                      22
## 150          150   Male  34                                    78                                                      90
## 151          151   Male  43                                    78                                                      17
## 152          152   Male  39                                    78                                                      88
## 153          153 Female  44                                    78                                                      20
## 154          154 Female  38                                    78                                                      76
## 155          155 Female  47                                    78                                                      16
## 156          156 Female  27                                    78                                                      89
```

```
## 157         157   Male  37                  78                       1
## 158         158 Female  30                  78                      78
## 159         159   Male  34                  78                       1
## 160         160 Female  30                  78                      73
## 161         161 Female  56                  79                      35
## 162         162 Female  29                  79                      83
## 163         163   Male  19                  81                       5
## 164         164 Female  31                  81                      93
## 165         165   Male  50                  85                      26
## 166         166 Female  36                  85                      75
## 167         167   Male  42                  86                      20
## 168         168 Female  33                  86                      95
## 169         169 Female  36                  87                      27
## 170         170   Male  32                  87                      63
## 171         171   Male  40                  87                      13
## 172         172   Male  28                  87                      75
## 173         173   Male  36                  87                      10
## 174         174   Male  36                  87                      92
## 175         175 Female  52                  88                      13
## 176         176 Female  30                  88                      86
## 177         177   Male  58                  88                      15
## 178         178   Male  27                  88                      69
## 179         179   Male  59                  93                      14
## 180         180   Male  35                  93                      90
## 181         181 Female  37                  97                      32
## 182         182 Female  32                  97                      86
## 183         183   Male  46                  98                      15
## 184         184 Female  29                  98                      88
## 185         185 Female  41                  99                      39
## 186         186   Male  30                  99                      97
## 187         187 Female  54                 101                      24
## 188         188   Male  28                 101                      68
## 189         189 Female  41                 103                      17
## 190         190 Female  36                 103                      85
## 191         191 Female  34                 103                      23
## 192         192 Female  32                 103                      69
## 193         193   Male  33                 113                       8
## 194         194 Female  38                 113                      91
## 195         195 Female  47                 120                      16
## 196         196 Female  35                 120                      79
## 197         197 Female  45                 126                      28
## 198         198   Male  32                 126                      74
## 199         199   Male  32                 137                      18
## 200         200   Male  30                 137                      83
```

Required list of packages

```
## package 'tidyverse' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Dell\AppData\Local\Temp\RtmpeS45l6\downloaded_packages
```

Loading the above Library

```
library(plotrix)
library(ggplot2)
library(purrr)
library(cluster)
library(grid)
library(gridExtra)
library(NbClust)
library(factoextra)
library(dplyr)
```

Data Insight

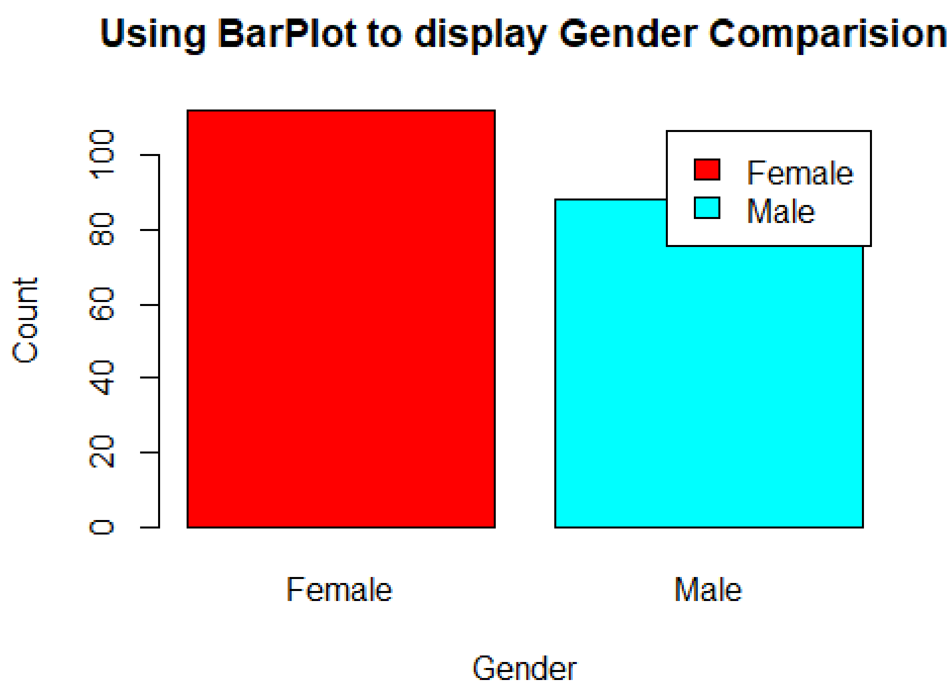
```
str(customer_data)

## 'data.frame':    200 obs. of  5 variables:
##  $ CustomerID            : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Gender                : Factor w/ 2 levels "Female","Male": 2 2 1 1 1 1
1 1 2 1 ...
##  $ Age                   : int  19 21 20 23 31 22 35 23 64 30 ...
##  $ Annual.Income..k..    : int  15 15 16 16 17 17 18 18 19 19 ...
##  $ Spending.Score..1.100.: int  39 81 6 77 40 76 6 94 3 72 ...

summary(customer_data)

##    CustomerID        Gender         Age        Annual.Income..k..
##  Min.   :  1.00   Female:112   Min.   :18.00   Min.   : 15.00
##  1st Qu.: 50.75   Male  : 88   1st Qu.:28.75   1st Qu.: 41.50
##  Median :100.50                Median :36.00   Median : 61.50
##  Mean   :100.50                Mean   :38.85   Mean   : 60.56
##  3rd Qu.:150.25                3rd Qu.:49.00   3rd Qu.: 78.00
##  Max.   :200.00                Max.   :70.00   Max.   :137.00
##  Spending.Score..1.100.
##  Min.   : 1.00
##  1st Qu.:34.75
##  Median :50.00
##  Mean   :50.20
##  3rd Qu.:73.00
##  Max.   :99.00

head(customer_data)

##   CustomerID Gender Age Annual.Income..k.. Spending.Score..1.100.
## 1          1   Male  19                 15                     39
## 2          2   Male  21                 15                     81
## 3          3 Female  20                 16                      6
## 4          4 Female  23                 16                     77
## 5          5 Female  31                 17                     40
## 6          6 Female  22                 17                     76
```

Data Visualisation

1.Gender

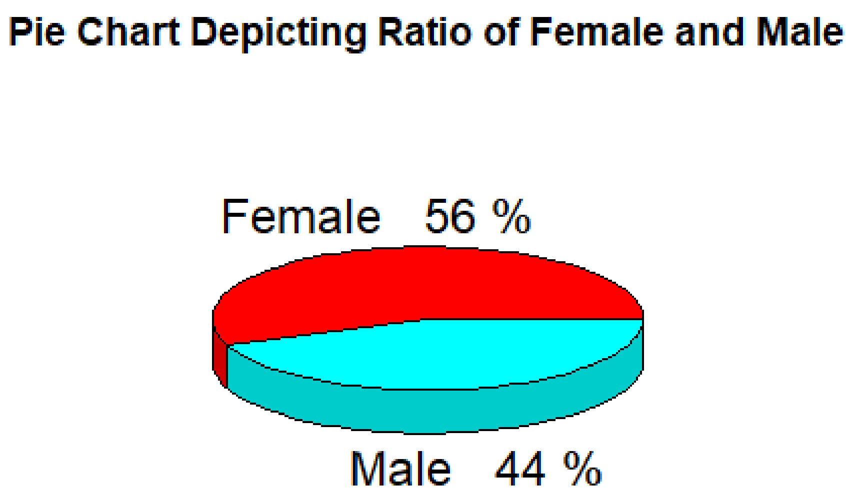
In this, we will create a barplot and a piechart to show the gender distribution across our customer_data dataset.

```
a=table(customer_data$Gender)
barplot(a,main="Using BarPlot to display Gender Comparision",
       ylab="Count",
       xlab="Gender",
       col=rainbow(2),
       legend=rownames(a)) # Bar plot Visualization
```



We observe that the number of females is higher than the males. Now, let us visualize a pie chart to observe the ratio of male and female distribution.

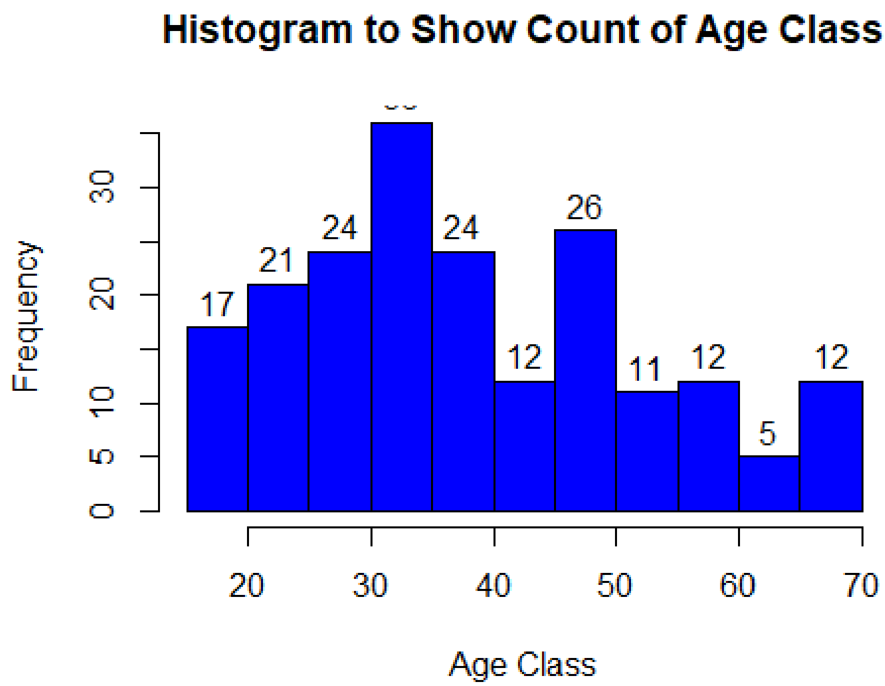
```
pct = round(a/sum(a) * 100)
lbs = paste(c("Female","Male")," ",pct,"%",sep=" ")
pie3D(a,labels = lbs,
      main="Pie Chart Depicting Ratio of Female and Male") # Pie chart
visualization
```



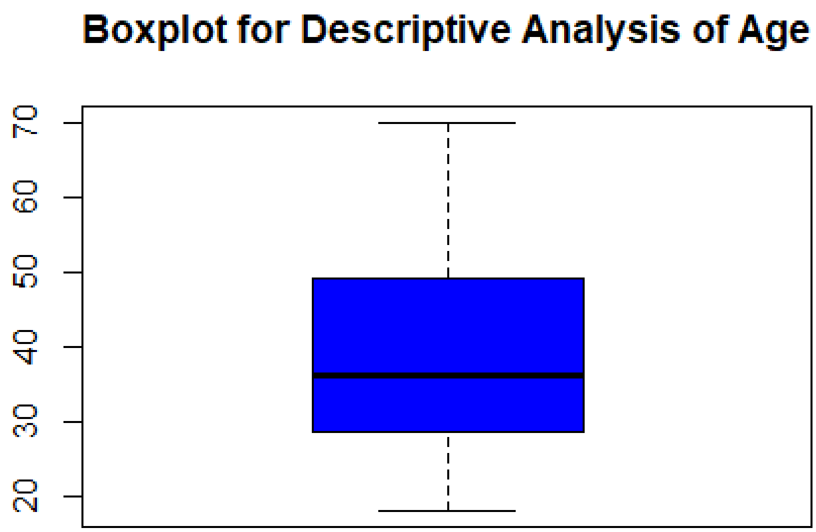
We conclude that the percentage of females is 56%, whereas the percentage of male in the customer dataset is 44%.

2.Age

```
hist(customer_data$Age,
     col = "blue",
     main = "Histogram to Show Count of Age Class",
     xlab = "Age Class",
     ylab = "Frequency",
     labels = TRUE) # Visualization through Histogram
```



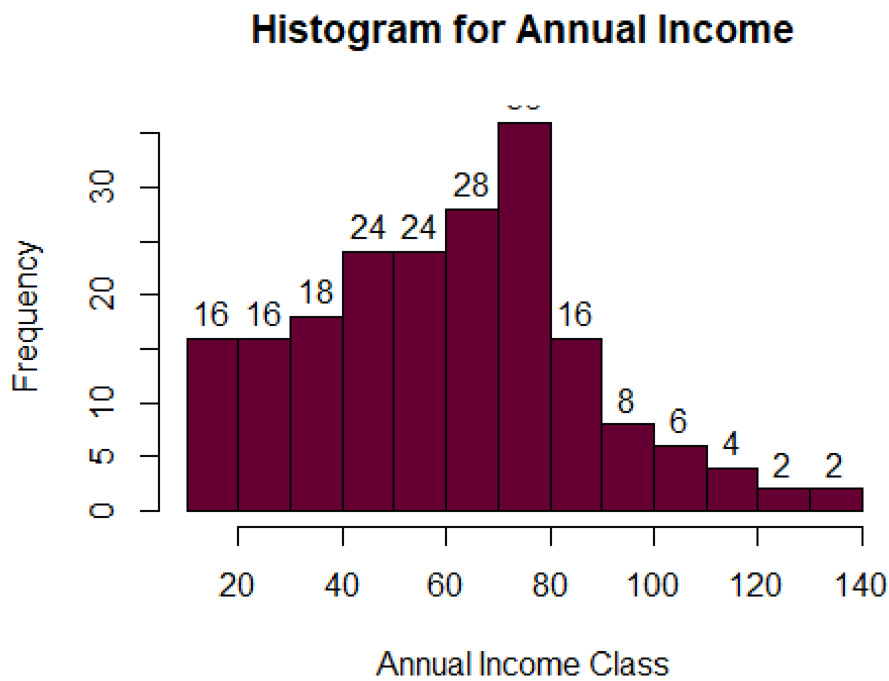

```
boxplot(customer_data$Age,
        col = "blue",
        main = "Boxplot for Descriptive Analysis of Age") # Boxplot
Visualization
```



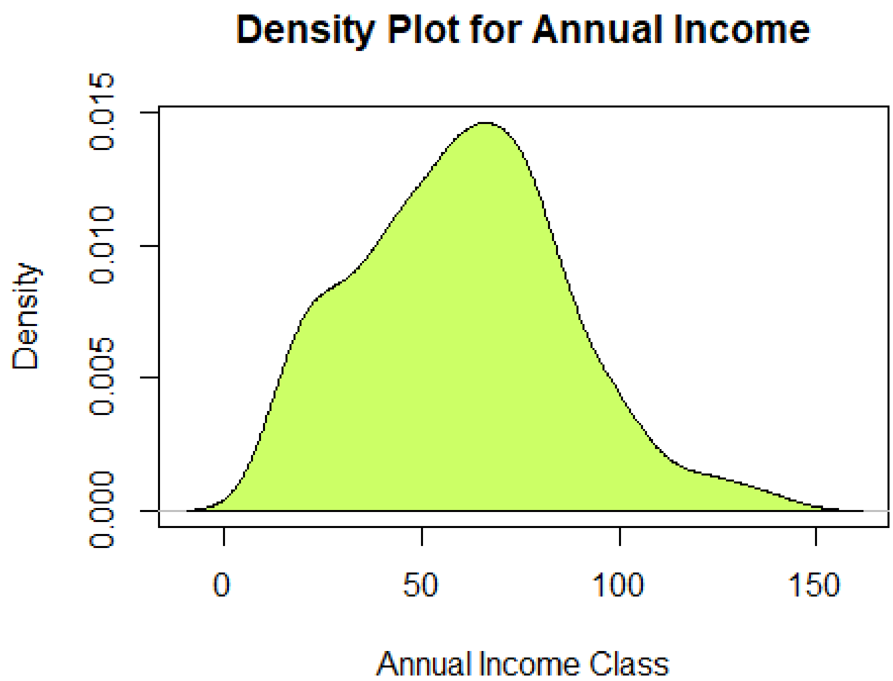
We conclude that the maximum customer ages are between 30 and 35. The minimum age of customers is 18, whereas, the maximum age is 70.

3.Annual Income of the Customers

```
hist(customer_data$Annual.Income..k..,
     col = "#660033",
     main = "Histogram for Annual Income",
     xlab = "Annual Income Class",
     ylab = "Frequency",
     labels = TRUE) # Visualization through Histogram
```



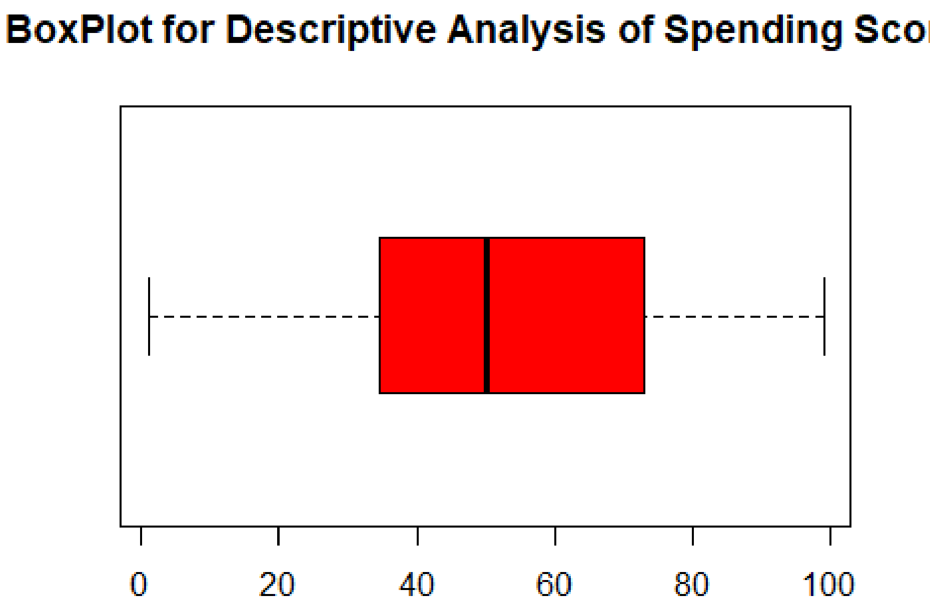

```r
plot(density(customer_data$Annual.Income..k..),
     col = "yellow",
     main = "Density Plot for Annual Income",
     xlab = "Annual Income Class",
     ylab = "Density")
polygon(density(customer_data$Annual.Income..k..),
        col = "#ccff66") # Visualization through Density plot
```



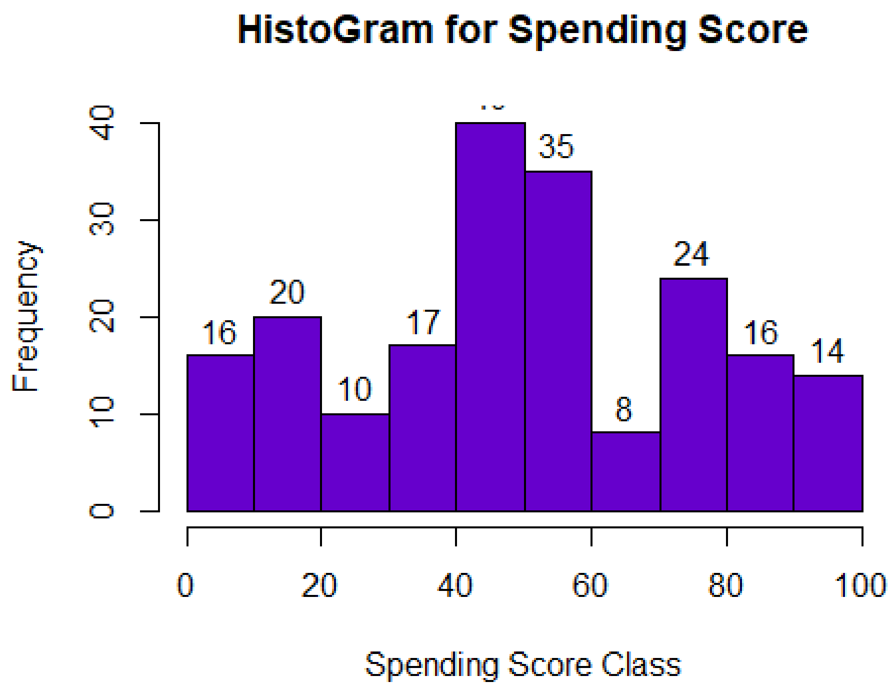
From the above descriptive analysis, we conclude that the minimum annual income of the customers is 15 and the maximum income is 137. People earning an average income of 70 have the highest frequency count in our histogram distribution. The average salary of all the customers is 60.56. In the Density Plot that we displayed above, we observe that the annual income has a normal distribution.

4.Spending score of the Customers

```
boxplot(customer_data$Spending.Score..1.100.,
        horizontal = TRUE,
        col = "red",
        main = "BoxPlot for Descriptive Analysis of Spending Score") #
Boxplot Visualization
```




```r
hist(customer_data$Spending.Score..1.100.,
     main = "HistoGram for Spending Score",
     xlab = "Spending Score Class",
     ylab = "Frequency",
     col = "#6600cc",
     labels = TRUE) # Visualization through Histogram
```



The minimum spending score is 1, maximum is 99 and the average is 50.20. We can see Descriptive Analysis of Spending Score is that Min is 1, Max is 99 and avg. is 50.20. From the histogram, we conclude that customers between class 40 and 50 have the highest spending score among all the classes.

K-means Algorithm

We specify the number of clusters that we need to create. The algorithm selects k objects at random from the dataset. This object is the initial cluster or mean. The closest centroid obtains the assignment of a new observation. We base this assignment on the Euclidean Distance between object and the centroid. k clusters in the data points update the centroid through calculation of the new mean values present in all the data points of the cluster. The kth cluster's centroid has a length of p that contains means of all variables for observations in the k-th cluster. We denote the number of variables with p. Iterative minimization of the total within the sum of squares. Then through the iterative minimization of the total sum of the square, the assignment stop wavering when we achieve maximum iteration. The default value is 10 that the R software uses for the maximum iterations.

While working with clusters, you need to specify the number of clusters to use. You would like to utilize the optimal number of clusters. To help you in determining the optimal clusters, there are three popular methods –

1.Elbow method 2.Silhouette method 3.Gap statistical method

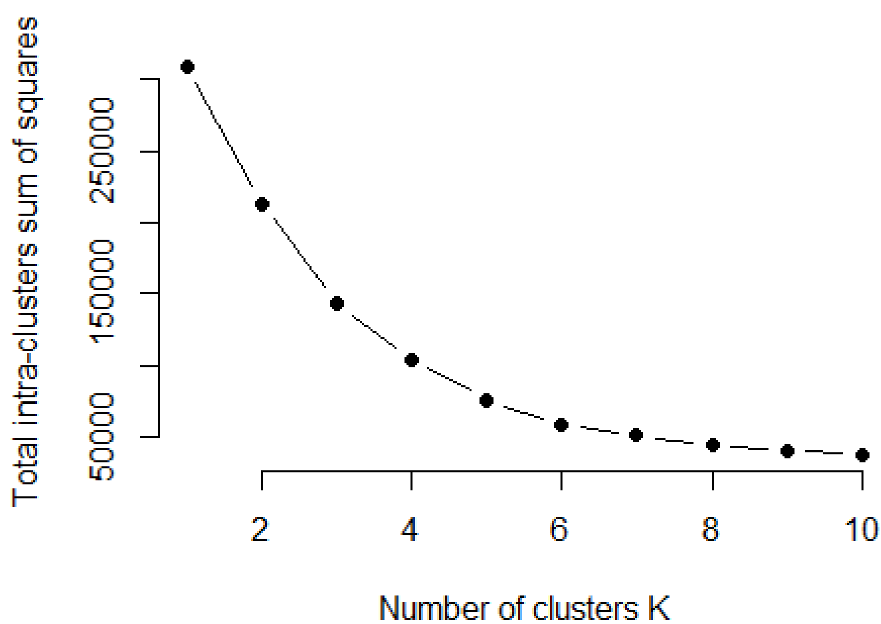
1.Elbow Method

The main goal behind cluster partitioning methods like k-means is to define the clusters such that the intra-cluster variation stays minimum.

minimize(sum W(Ck)), k=1…k

Where Ck represents the kth cluster and W(Ck) denotes the intra-cluster variation. With the measurement of the total intra-cluster variation, one can evaluate the compactness of the clustering boundary.

```r
set.seed(123)
# function to calculate total intra-cluster sum of square
iss <- function(k) {
  kmeans(customer_data[,3:5],k,iter.max=100,nstart=100,algorithm="Lloyd"
)$tot.withinss
}
k.values <- 1:10
iss_values <- map_dbl(k.values, iss)
plot(k.values, iss_values,
     type = "b", pch = 19, frame = FALSE,
     xlab = "Number of clusters K",
     ylab = "Total intra-clusters sum of squares")
```



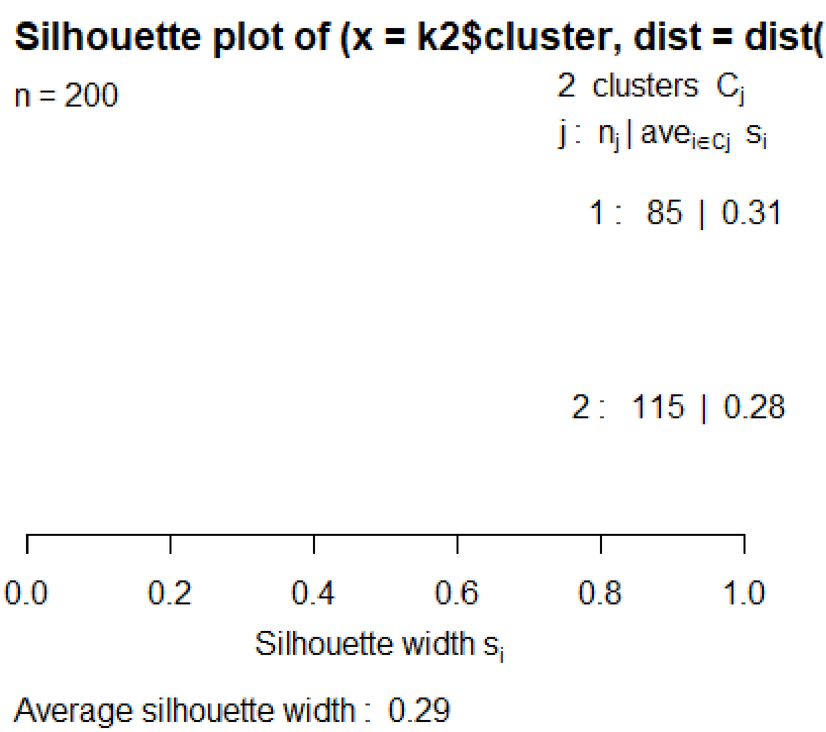
We conclude that 4 is the appropriate number of clusters since it seems to be appearing at the bend in the elbow plot.

2.Average Silhouette Method

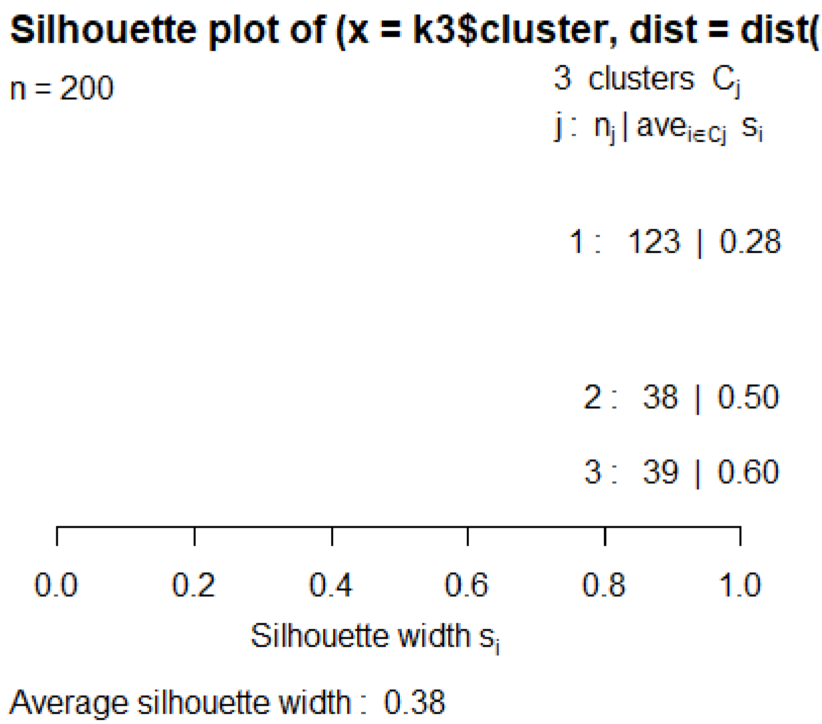
With the help of the average silhouette method, we can measure the quality of our clustering operation. With this, we can determine how well within the cluster is the data object. If we obtain a high average silhouette width, it means that we have good clustering. The average silhouette method calculates the mean of silhouette observations for different k values. With the optimal number of k clusters, one can maximize the average silhouette over significant values for k clusters.

Using the silhouette function in the cluster package, we can compute the average silhouette width using the kmean function.

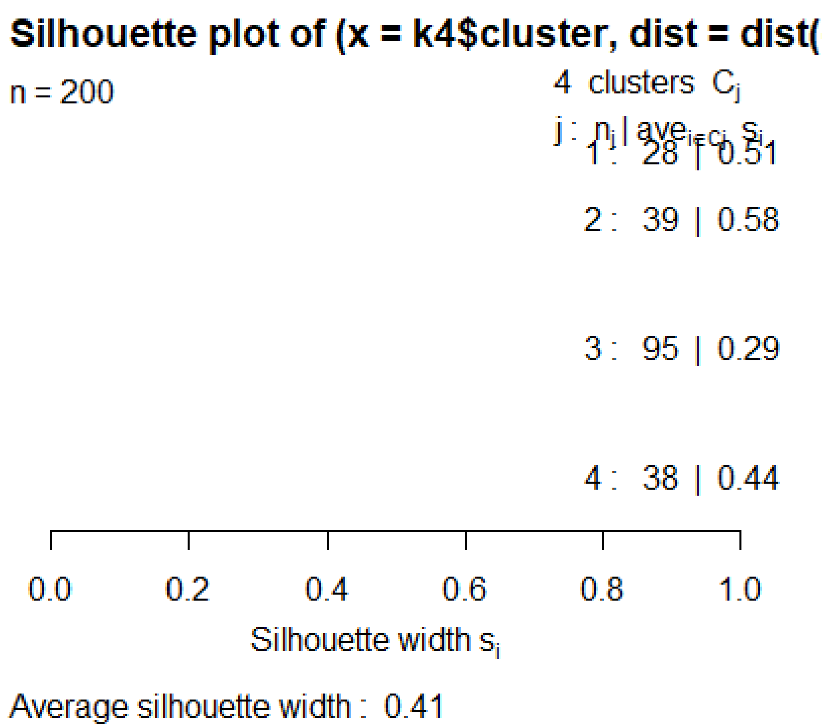
```
k2<-kmeans(customer_data[,3:5],2,iter.max = 100,nstart = 50,algorithm
="Lloyd")
s2<-plot(silhouette(k2$cluster,dist(customer_data[,3:5],"euclidean")))
```



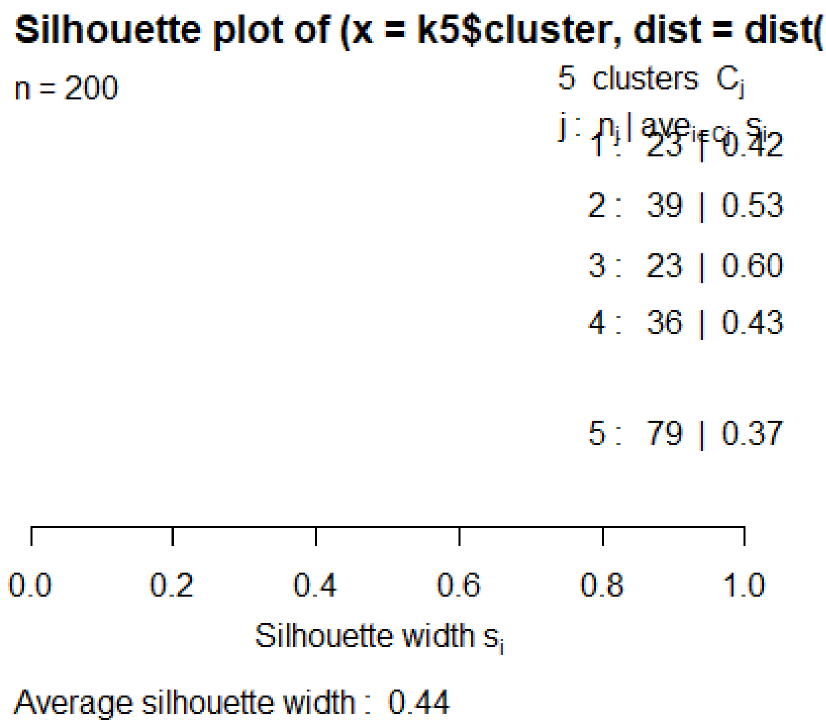

```
k3<-kmeans(customer_data[,3:5],3,iter.max = 100,nstart = 50,algorithm =
"Lloyd")
s3<-plot(silhouette(k3$cluster,dist(customer_data[,3:5],"euclidean")))
```



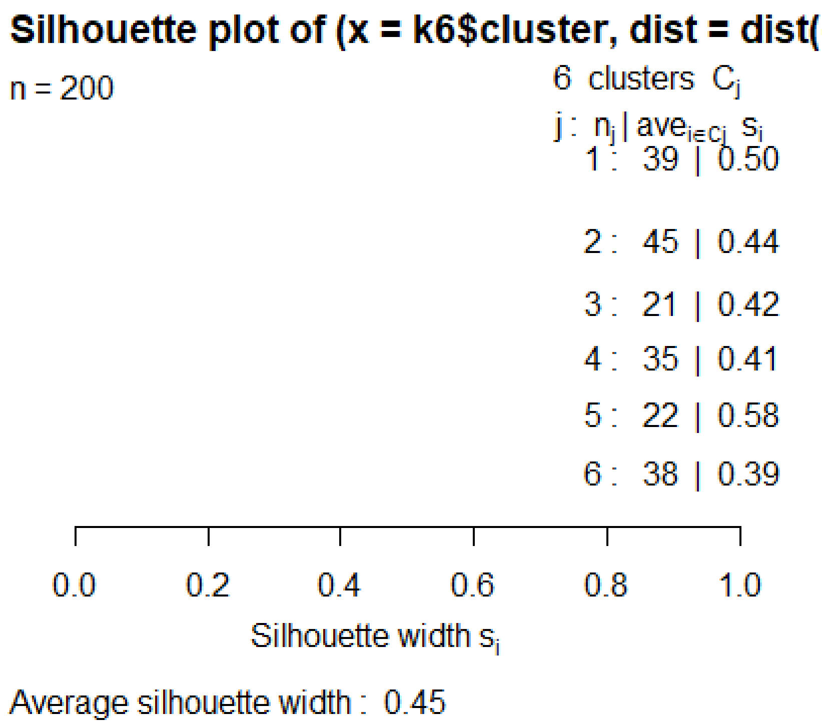

```
k4<-kmeans(customer_data[,3:5],4,iter.max = 100,nstart = 50,algorithm =
"Lloyd")
s4<-plot(silhouette(k4$cluster,dist(customer_data[,3:5],"euclidean")))
```



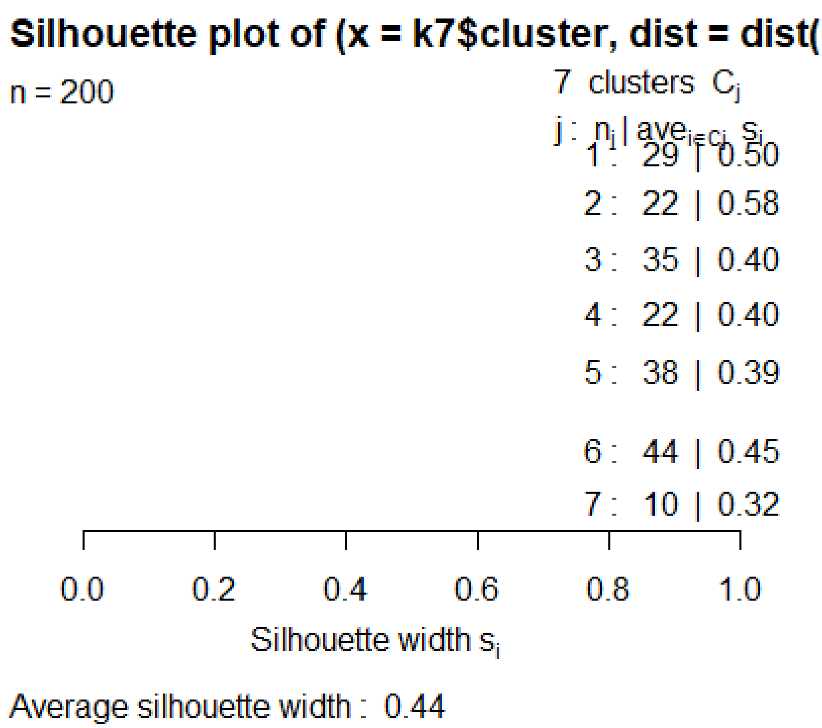
```r
k5<-kmeans(customer_data[,3:5],5,iter.max = 100,nstart = 50,algorithm =
"Lloyd")
s5<-plot(silhouette(k5$cluster,dist(customer_data[,3:5],"euclidean")))
```



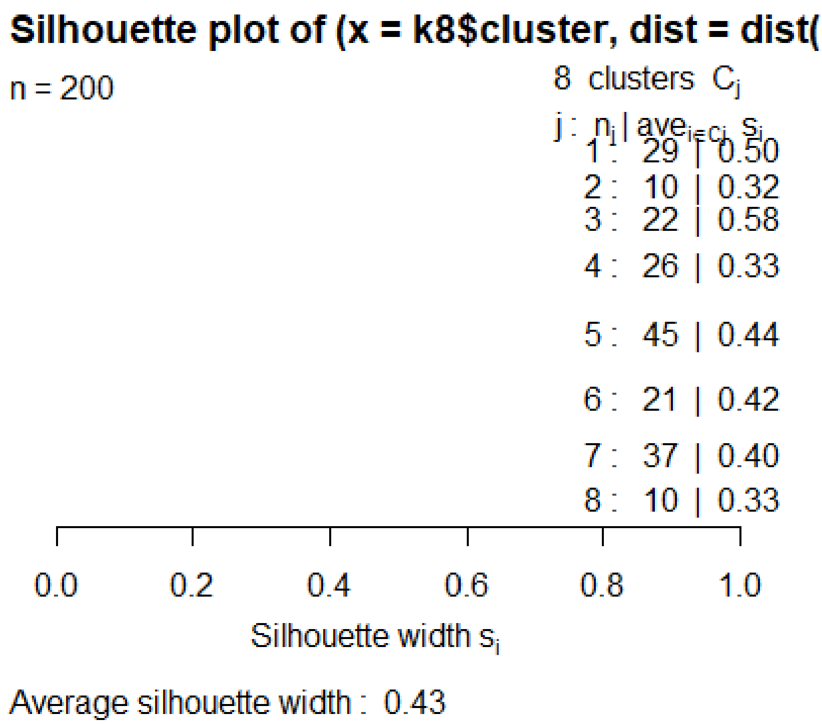

```r
k6<-kmeans(customer_data[,3:5],6,iter.max = 100,nstart = 50,algorithm =
"Lloyd")
s6<-plot(silhouette(k6$cluster,dist(customer_data[,3:5],"euclidean")))
```



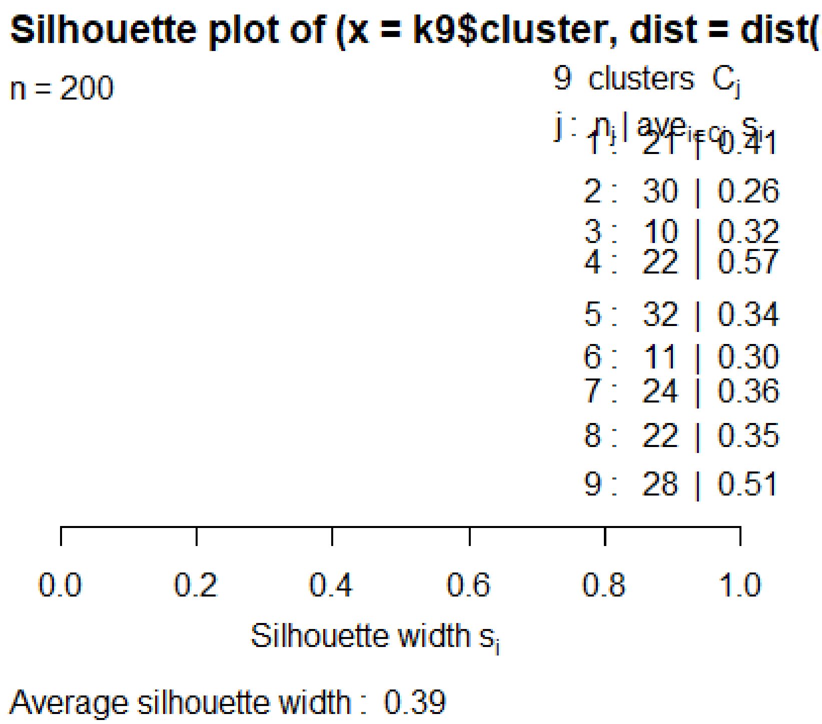

```
k7<-kmeans(customer_data[,3:5],7,iter.max = 100,nstart = 50,algorithm =
"Lloyd")
s7<-plot(silhouette(k7$cluster,dist(customer_data[,3:5],"euclidean")))
```



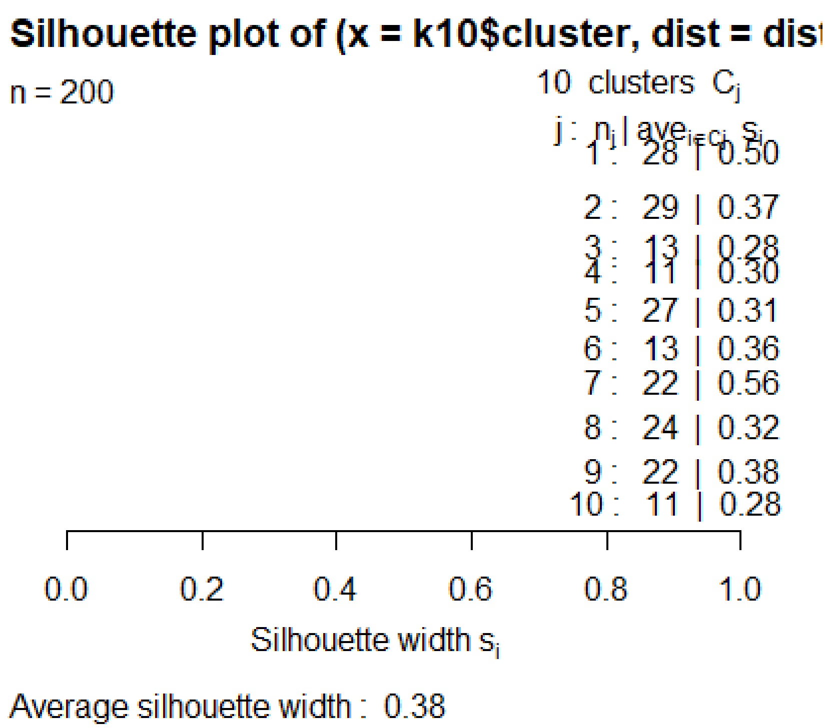
```r
k8<-kmeans(customer_data[,3:5],8,iter.max = 100,nstart = 50,algorithm = "Lloyd")
s8<-plot(silhouette(k8$cluster,dist(customer_data[,3:5],"euclidean")))
```




```r
k9<-kmeans(customer_data[,3:5],9,iter.max = 100,nstart = 50,algorithm = "Lloyd")
s9<-plot(silhouette(k9$cluster,dist(customer_data[,3:5],"euclidean")))
```

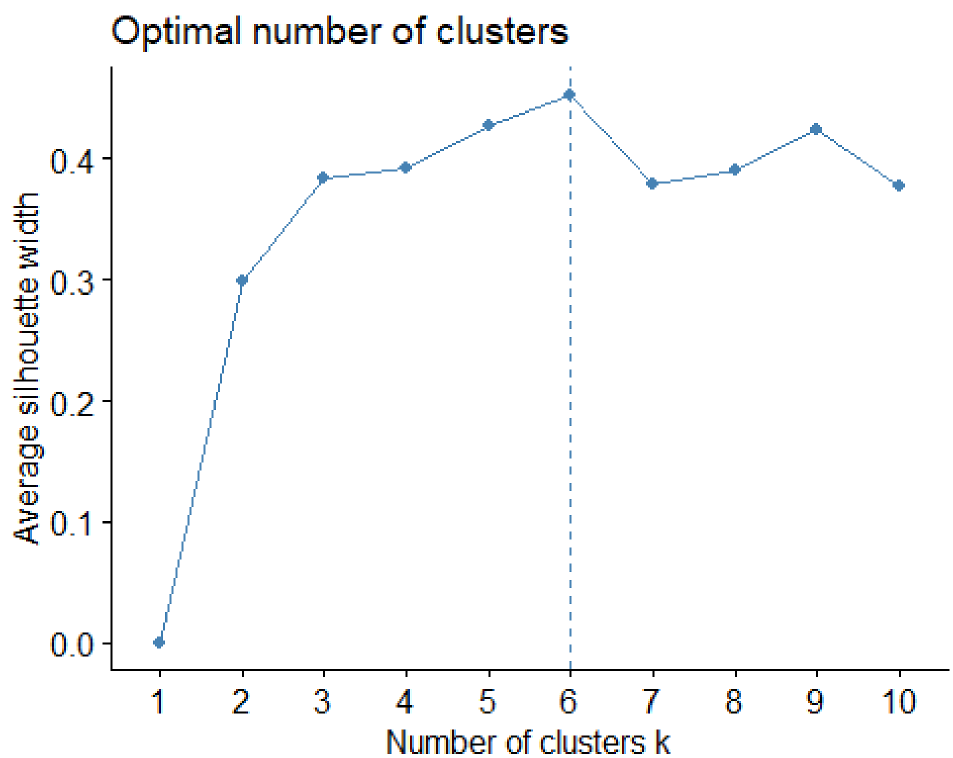



```
k10<-kmeans(customer_data[,3:5],10,iter.max = 100,nstart = 50,algorithm =
"Lloyd")
s10<-plot(silhouette(k10$cluster,dist(customer_data[,3:5],"euclidean")))
```



We will use fviz_nbclust() function to determine and visualize the optimal number of clusters

```
fviz_nbclust(customer_data[,3:5], kmeans, method = "silhouette")
```

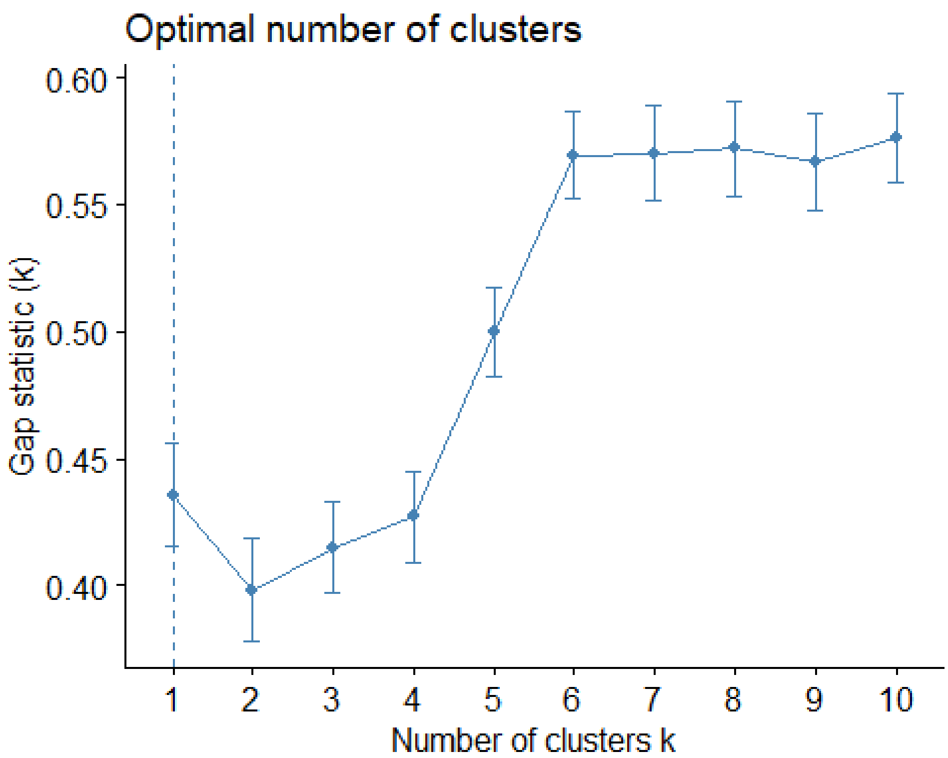


3.Gap Statistic Method

We can use this method to any of the clustering method like K-means. Using the gap statistic, one can compare the total intracluster variation for different values of k along with their expected values under the null reference distribution of data. With the help of Monte Carlo simulations, one can produce the sample dataset. For each variable in the dataset, we can calculate the range between min(xi) and max (xj) through which we can produce values uniformly from interval lower bound to upper bound.

For computing the gap statistics method we can utilize the clusGap function for providing gap statistic as well as standard error for a given output.

```
set.seed(125)
stat_gap <- clusGap(customer_data[,3:5], FUN = kmeans, nstart = 25,
            K.max = 10, B = 50)
fviz_gap_stat(stat_gap)
```



We will be taking k = 6 as our optimal cluster.

```
k6<-kmeans(customer_data[,3:5],6,iter.max=100,nstart=50,algorithm="Lloyd")
k6

## K-means clustering with 6 clusters of sizes 45, 22, 21, 38, 35, 39
##
## Cluster means:
##        Age Annual.Income..k.. Spending.Score..1.100.
## 1 56.15556           53.37778               49.08889
## 2 25.27273           25.72727               79.36364
## 3 44.14286           25.14286               19.52381
## 4 27.00000           56.65789               49.13158
## 5 41.68571           88.22857               17.28571
## 6 32.69231           86.53846               82.12821
##
## Clustering vector:
##   [1] 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2
3 2 3
##  [38] 2 3 2 1 2 1 4 3 2 1 4 4 4 1 4 4 1 1 1 1 1 4 1 1 4 1 1 1 4 1 1 4 4 1
1 1 1
##  [75] 1 4 1 4 4 1 1 4 1 1 4 1 1 4 4 1 1 4 1 4 4 4 1 4 1 4 4 1 1 4 1 4 1 1
1 1 1
## [112] 4 4 4 4 4 1 1 1 1 4 4 4 6 4 6 5 6 5 6 5 6 4 6 5 6 5 6 5 6 5 6 4 6 5
6 5 6
## [149] 5 6 5 6 5 6 5 6 5 6 5 6 5 6 5 6 5 6 5 6 5 6 5 6 5 6 5 6 5 6 5 6 5 6
```

```
5 6 5
## [186] 6 5 6 5 6 5 6 5 6 5 6 5 6 5 6
##
## Within cluster sum of squares by cluster:
## [1]  8062.133  4099.818  7732.381  7742.895 16690.857 13972.359
##  (between_SS / total_SS =  81.1 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
"tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

Final Prediction using Clusters

Visualizing the Clustering Results using the First Two Principle Components

```r
pcclust = prcomp(customer_data[,3:5],scale=FALSE) #principal component
analysis
summary(pcclust)
```

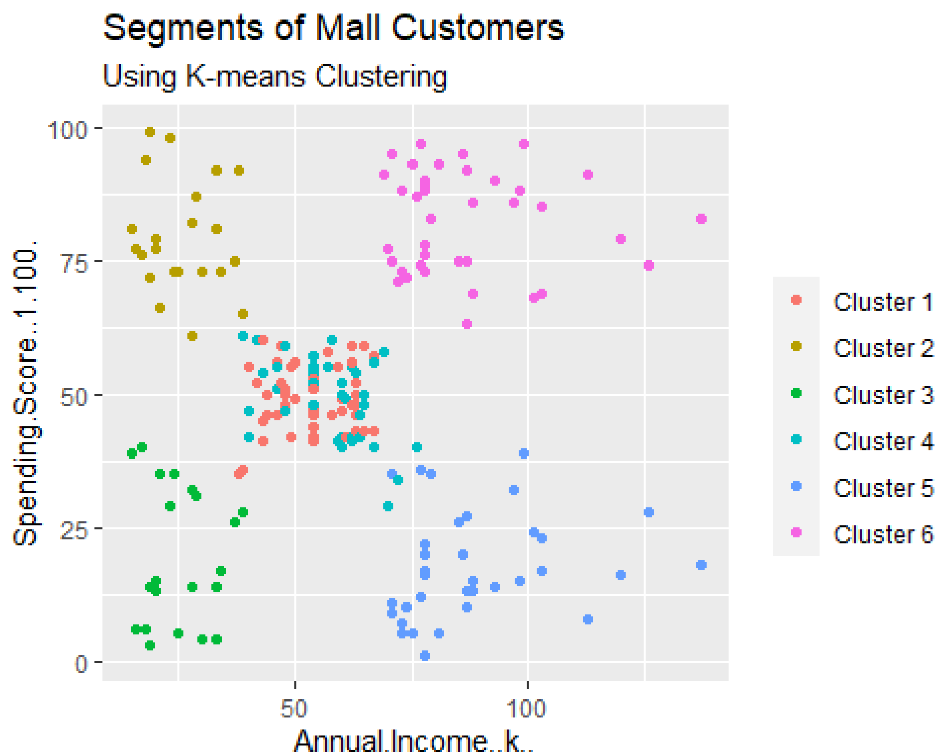
```
## Importance of components:
##                            PC1     PC2     PC3
## Standard deviation     26.4625 26.1597 12.9317
## Proportion of Variance  0.4512  0.4410  0.1078
## Cumulative Proportion   0.4512  0.8922  1.0000
```

```r
pcclust$rotation[,1:2]
```

```
##                               PC1        PC2
## Age                     0.1889742 -0.1309652
## Annual.Income..k..     -0.5886410 -0.8083757
## Spending.Score..1.100. -0.7859965  0.5739136
```

Visualising the cluster(Annual Income)

```r
set.seed(1)
ggplot(customer_data, aes(x = Annual.Income..k.., y =
Spending.Score..1.100.)) +
  geom_point(stat = "identity", aes(color = as.factor(k6$cluster))) +
  scale_color_discrete(name = " ",
                       breaks = c("1", "2", "3", "4", "5","6"),
                       labels = c("Cluster 1", "Cluster 2", "Cluster 3",
"Cluster 4", "Cluster 5","Cluster 6")) +
  ggtitle("Segments of Mall Customers", subtitle = "Using K-means
Clustering")
```



Outcome:

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

1.Cluster 6 and 4 – These clusters represent the customer_data with the medium income salary as well as the medium annual spend of salary.

2.Cluster 1 – This cluster represents the customer_data having a high annual income as well as a high annual spend.

3.Cluster 3 – This cluster denotes the customer_data with low annual income as well as low yearly spend of income.

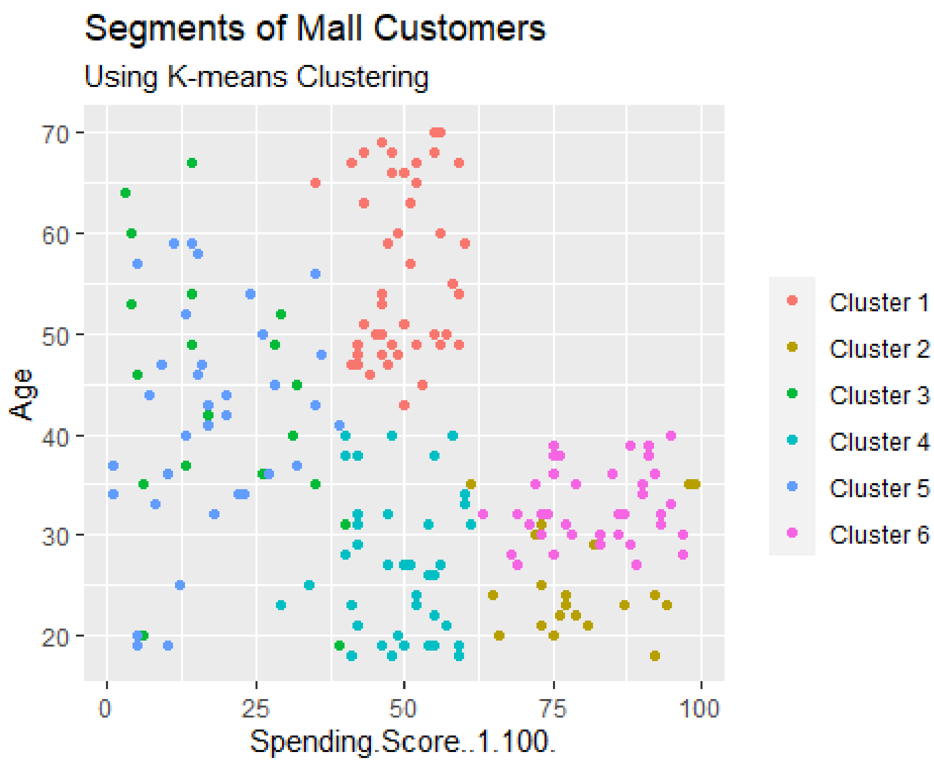
4.Cluster 2 – This cluster denotes a high annual income and low yearly spend.

5.Cluster 5 – This cluster represents a low annual income but its high yearly expenditure.

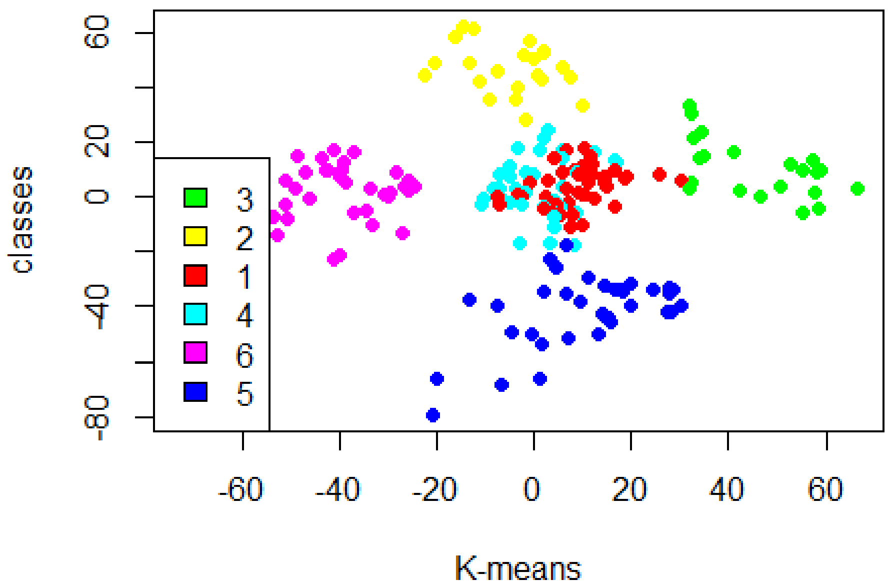
Building clusters(spending score)

```
ggplot(customer_data, aes(x = Spending.Score..1.100., y =Age)) +
  geom_point(stat = "identity", aes(color = as.factor(k6$cluster))) +
  scale_color_discrete(name = " ",
                       breaks = c("1", "2", "3", "4", "5","6"),
                       labels = c("Cluster 1", "Cluster 2", "Cluster 3",
```

```
"Cluster 4", "Cluster 5","Cluster 6")) +
  ggtitle("Segments of Mall Customers", subtitle = "Using K-means
Clustering")
```




```
kCols = function(vec){cols = rainbow (length (unique (vec)))
return (cols[as.numeric(as.factor(vec))])}
digCluster<-k6$cluster; dignm<-as.character(digCluster); # K-means clusters
plot(pcclust$x[,1:2], col = kCols(digCluster),pch = 19,xlab = "K-means",ylab
= "classes")
legend("bottomleft",unique(dignm),fill = unique(kCols(digCluster)))
```



Outcome :

From the above two data visualisations we conclude that;

1.Cluster 4 and 1 – These two clusters consist of customers with medium PCA1 and medium PCA2 score.

2.Cluster 6 – This cluster represents customers having a high PCA2 and a low PCA1.

3.Cluster 5 – In this cluster, there are customers with a medium PCA1 and a low PCA2 score.

4.Cluster 3 – This cluster comprises of customers with a high PCA1 income and a high PCA2.

5.Cluster 2 – This comprises of customers with a high PCA2 and a medium annual spend of income.

Conclusion

With the help of clustering, we can understand the variables much better, prompting us to take careful decisions. With the identification of customers, companies can release products and services that target customers based on several parameters like income, age, spending patterns, etc.

Furthermore, more complex patterns like product reviews are taken into consideration for better segmentation.

In this data science project, we went through the market segmentation model.

We developed this using a class of machine learning known as unsupervised learning. Specifically, we made use of a clustering algorithm called K-means clustering.

We analyzed and visualized the data and then proceeded to implement our algorithm.

Hope you have enjoyed exploring and analyzing my capstone project(Market Sectionalization), using Machine Learning approach in R.

THANK YOU