PROJECT RUNNING PROCEDURE

PROCESS

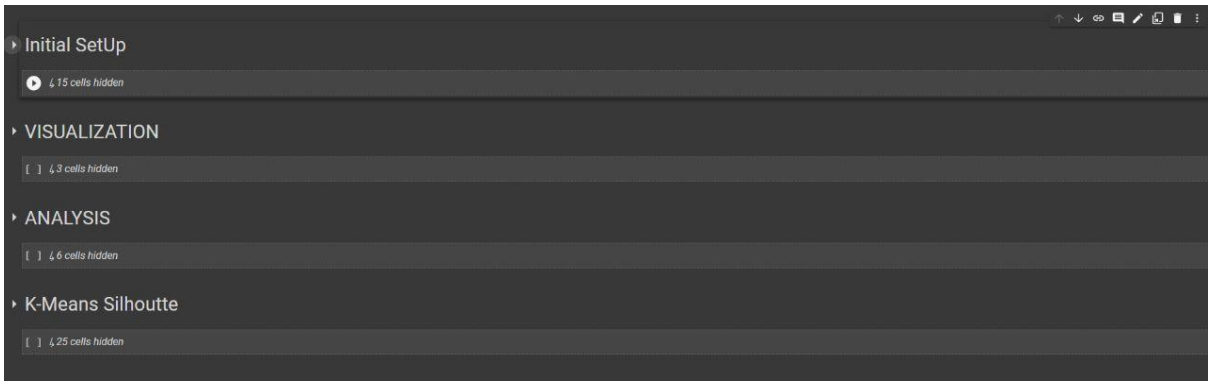
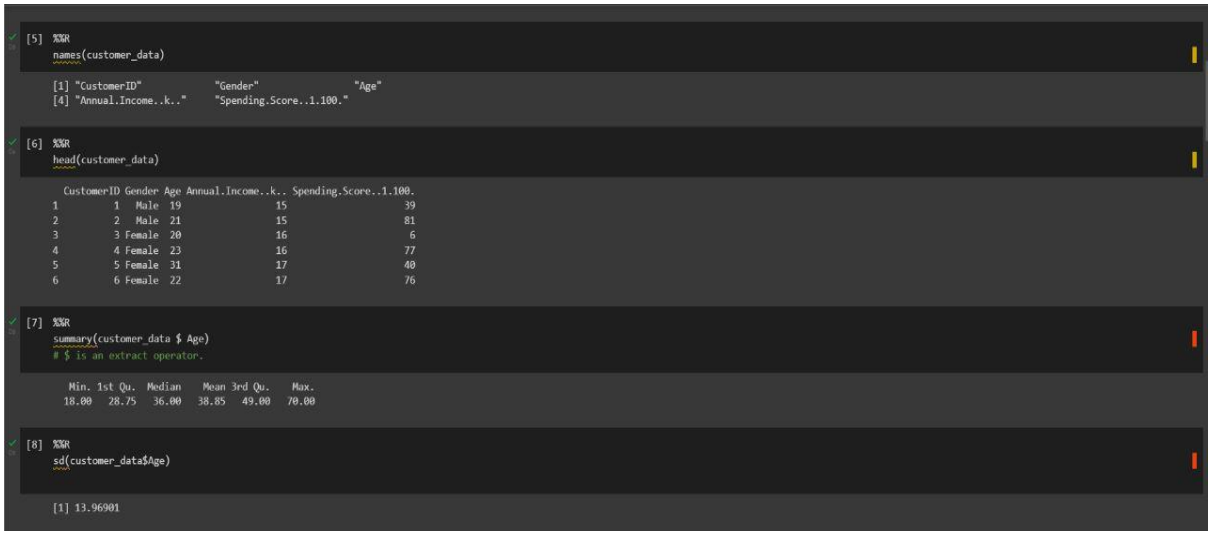


Figure: 1

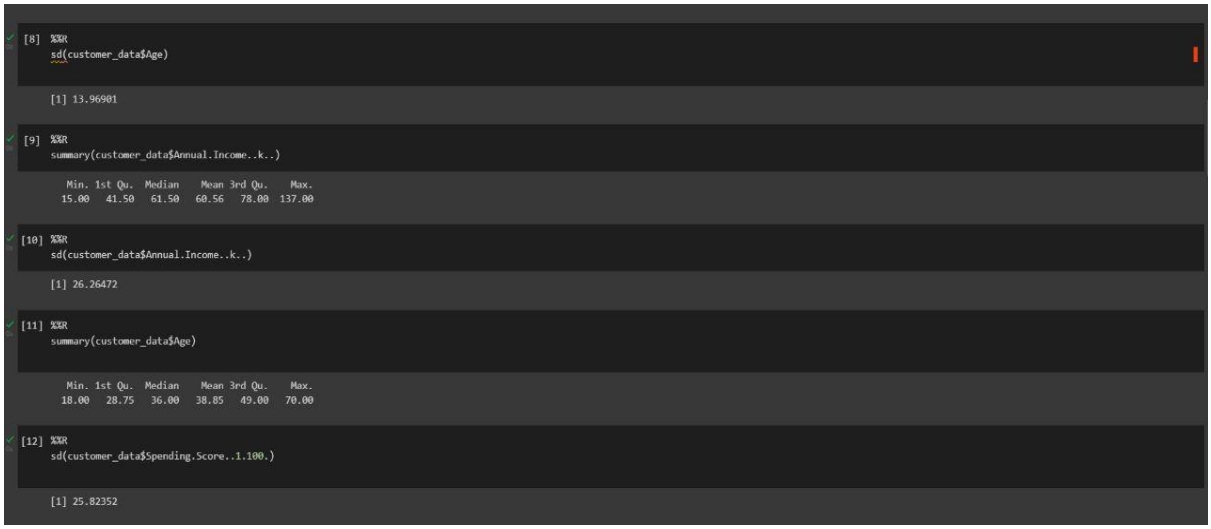
INITIAL SETUP

Figure: 2



```
[5] %%R
    names(customer_data)

    [1] "CustomerID"             "Gender"                 "Age"
    [4] "Annual.Income..k.."     "Spending.Score..1.100."

[6] %%R
    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

[7] %%R
    summary(customer_data $ Age)
    # $ is an extract operator.

       Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      18.00   28.75   36.00   38.85   49.00   70.00

[8] %%R
    sd(customer_data$Age)

    [1] 13.96901
```

Figure: 3



```
[8] %%R
    sd(customer_data$Age)

    [1] 13.96901

[9] %%R
    summary(customer_data$Annual.Income..k..)

       Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      15.00   41.50   61.50   60.56   78.00  137.00

[10] %%R
    sd(customer_data$Annual.Income..k..)

    [1] 26.26472

[11] %%R
    summary(customer_data$Age)

       Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      18.00   28.75   36.00   38.85   49.00   70.00

[12] %%R
    sd(customer_data$Spending.Score..1.100.)

    [1] 25.82352
```

Figure: 4

```
%%R
a=table(customer_data$Gender)
barplot(a,main="Using BarPlot to display Gender Comparision",
       ylab="Count",
       xlab="Gender",
       col=c('green', 'blue'),
       legend=rownames(a))
```

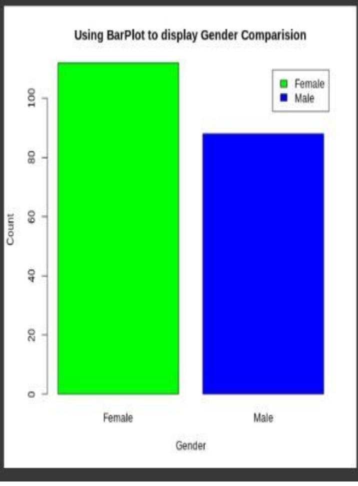


Figure: 5

```
%%R
install.packages("plotrix")
```

```
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: trying URL 'https://cran.rstudio.com/src/contrib/plotrix_3.8-2.tar.gz'

WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: Content type 'application/x-gzip'
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]:  length 318217 bytes (310 KB)

WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
```

Figure: 6

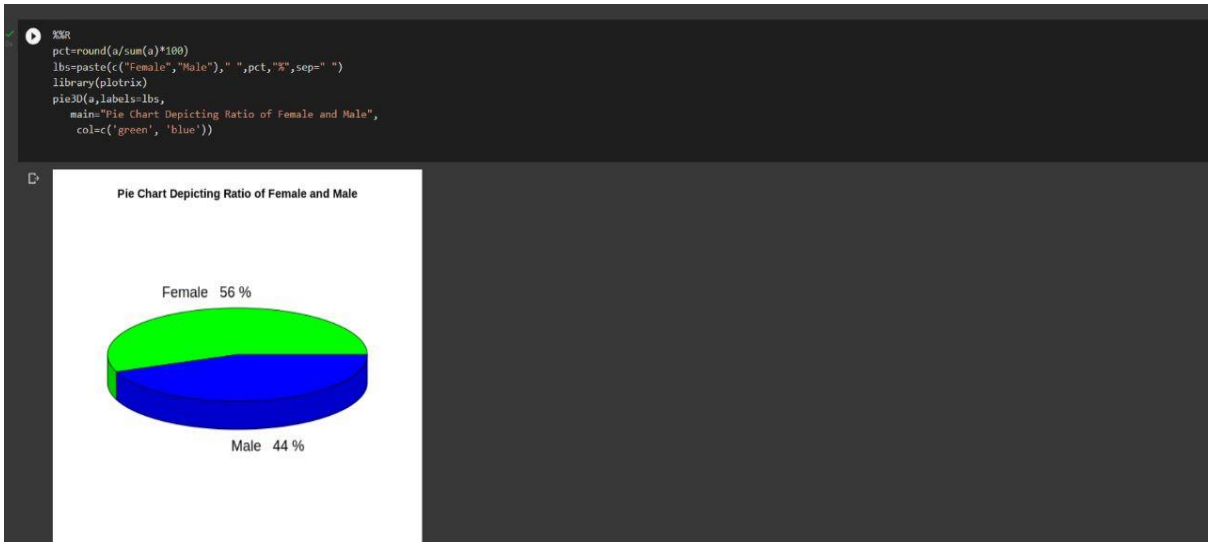


Figure: 7

VISUALIZATION

Figure: 8

Figure: 9

ANALYSIS

Figure: 10

```
%%R
hist(customer_data$Age,
     col="Blue",
     main="Histogram to Show Count of Age Class",
     xlab="Age Class",
     ylab="Frequency",
     labels=TRUE)
```

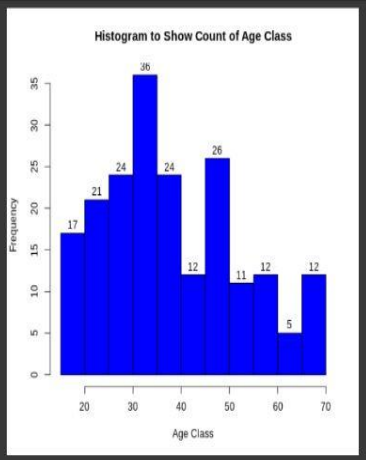


Figure: 11

```
%%R
boxplot(customer_data$Age,
        col="yellow",
        main="Boxplot for Descriptive Analysis of Age")
```

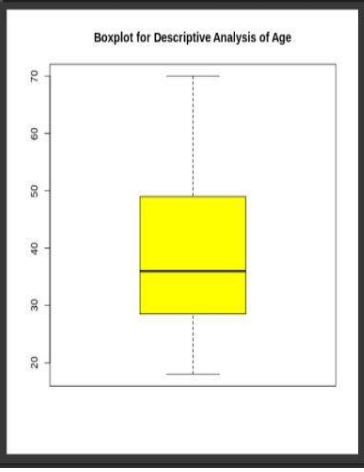


Figure: 12

```
%%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")
```

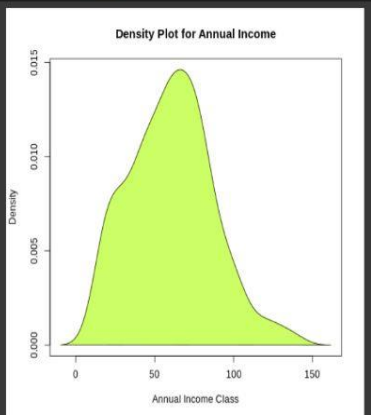


Figure: 13

```
%%R
hist(customer_data$Annual.Income..k..,
  col="cyan",
  main="Histogram for Annual Income",
  xlab="Annual Income Class",
  ylab="Frequency",
  labels=TRUE)
```

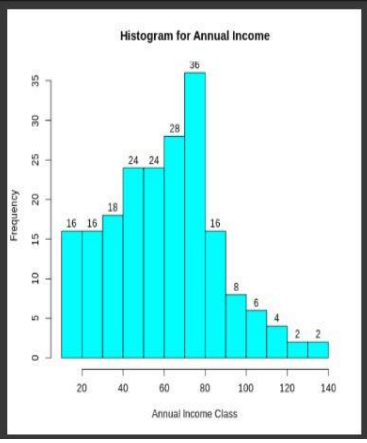


Figure: 14

```
[23] %%R
     summary(customer_data$Spending.Score..1.100.)

        Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
        1.00   34.75   50.00   50.20   73.00   99.00

     %%R
     boxplot(customer_data$Spending.Score..1.100.,
       horizontal=TRUE,
       col="YELLOW",
       main="BoxPlot for Descriptive Analysis of Spending Score")
```

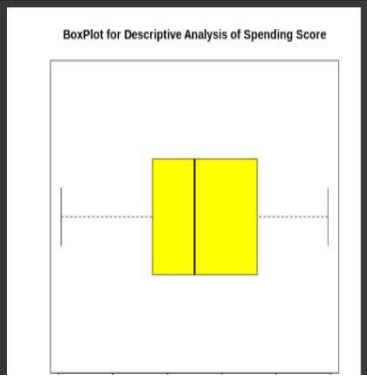


Figure: 15

```
%%R
hist(customer_data$Spending.Score..1.100.,
    main="HistoGram for Spending Score",
    xlab="Spending Score Class",
    ylab="Frequency",
    col="cyan",
    labels=TRUE)
```

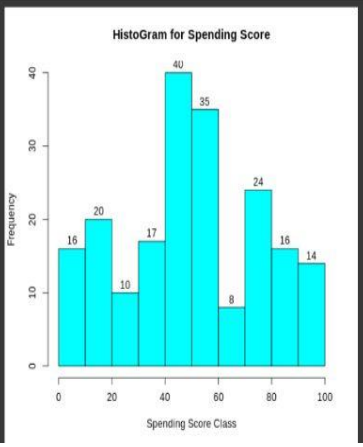


Figure: 16

```
[26] %%R
     install.packages("purrr")

     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
     WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
```

Figure: 17

ELBOW METHOD

```
plot(k.values, iss_values,
     type="b", pch = 19, frame = FALSE,
     xlab="Number of clusters K",
     ylab="Total intra-clusters sum of squares")
```

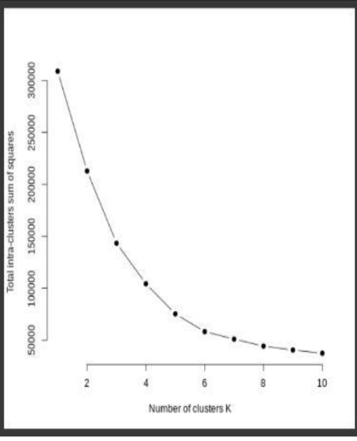


Figure: 18

SILHOUETTE GRAPHS

```
%%R
library(cluster)
library(gridExtra)
library(grid)

k2<- kmeans(customer_data[,3:4],2,iter.max=100,nstart=50,algorithm="Lloyd")
s2<- plot(silhouette(k2 $ cluster,dist(customer_data[,3:5],"euclidean")))
```

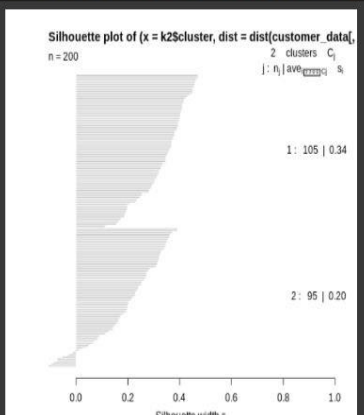


Figure: 19

```
%%R
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")))
```

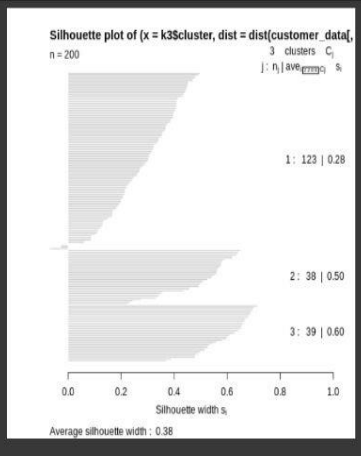


Figure: 20

Figure: 21

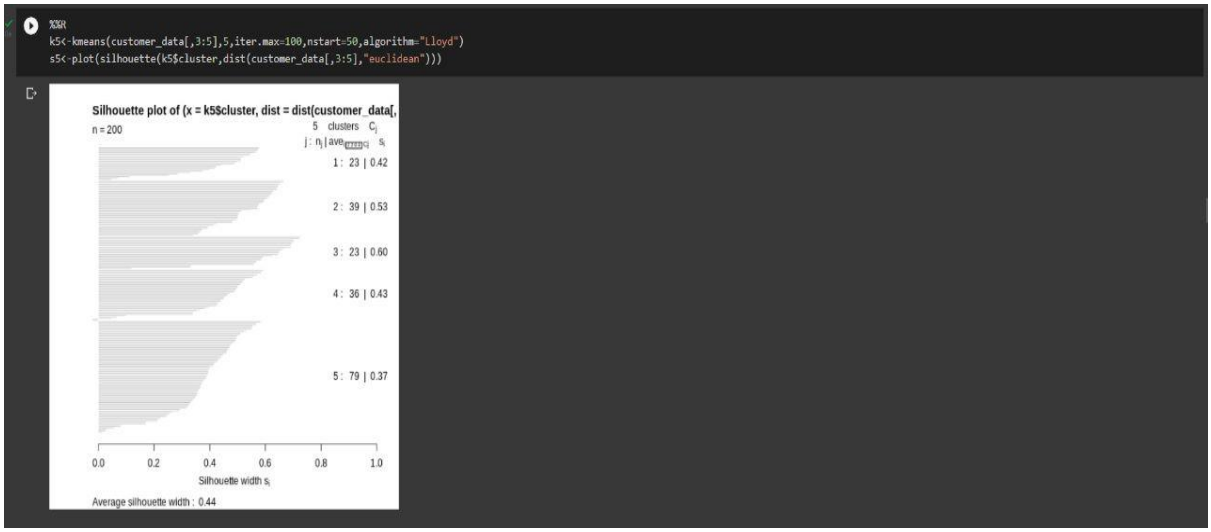


Figure: 22

Figure: 23

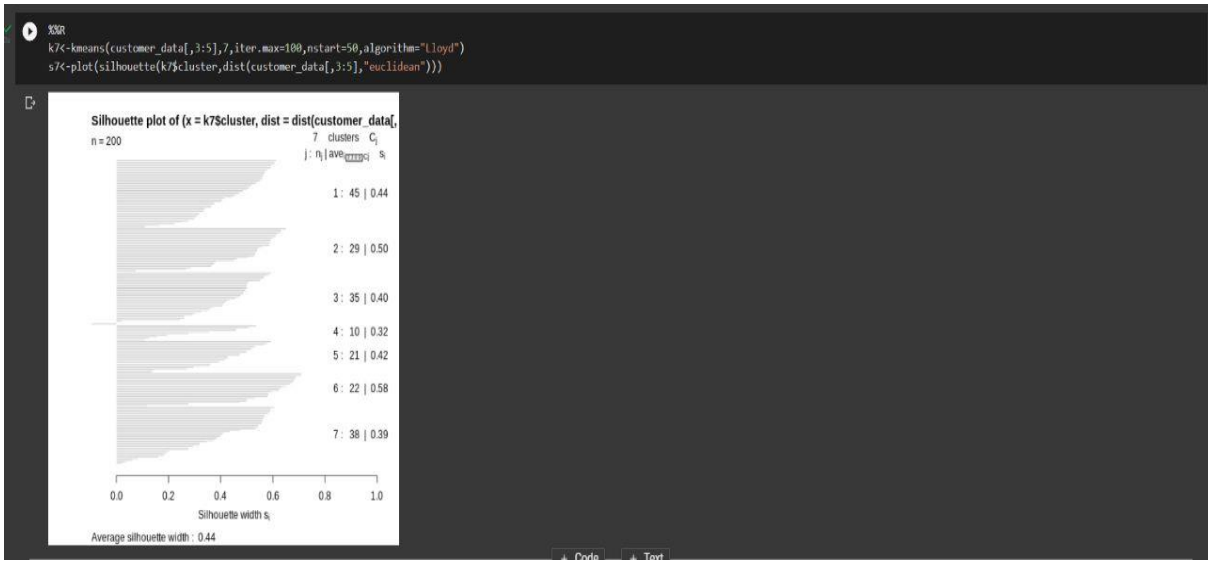


Figure: 24

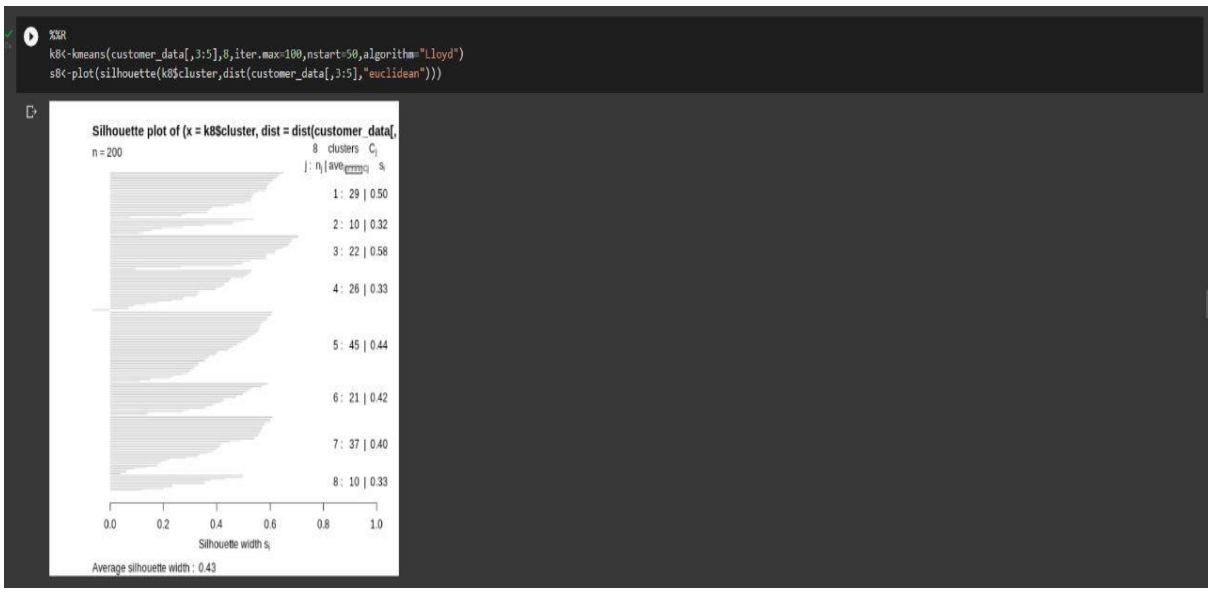


Figure: 25

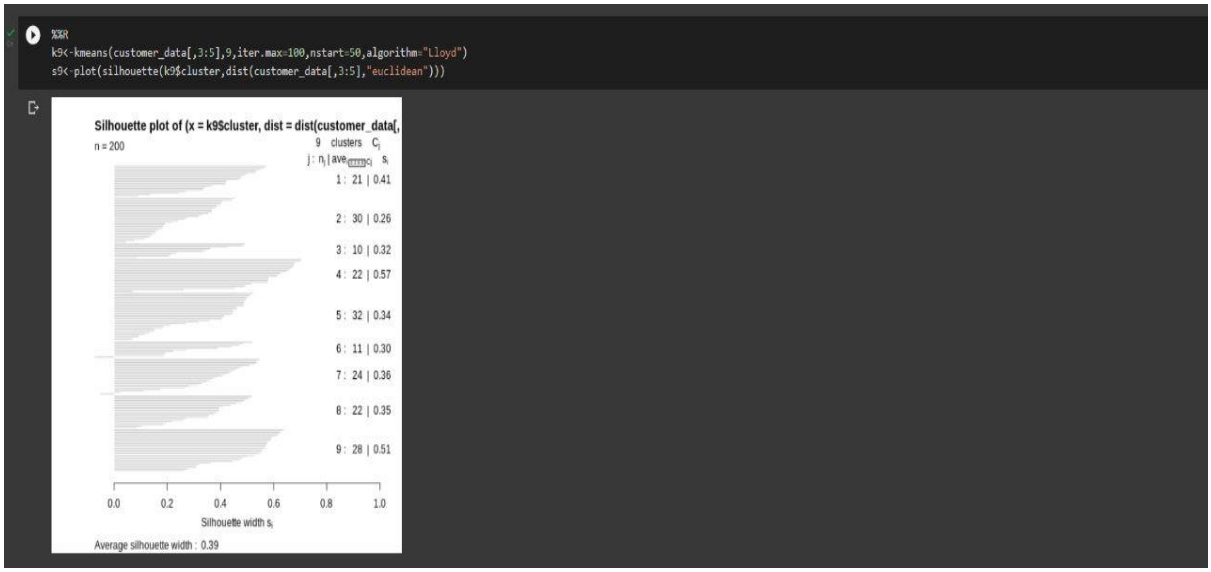


Figure: 26

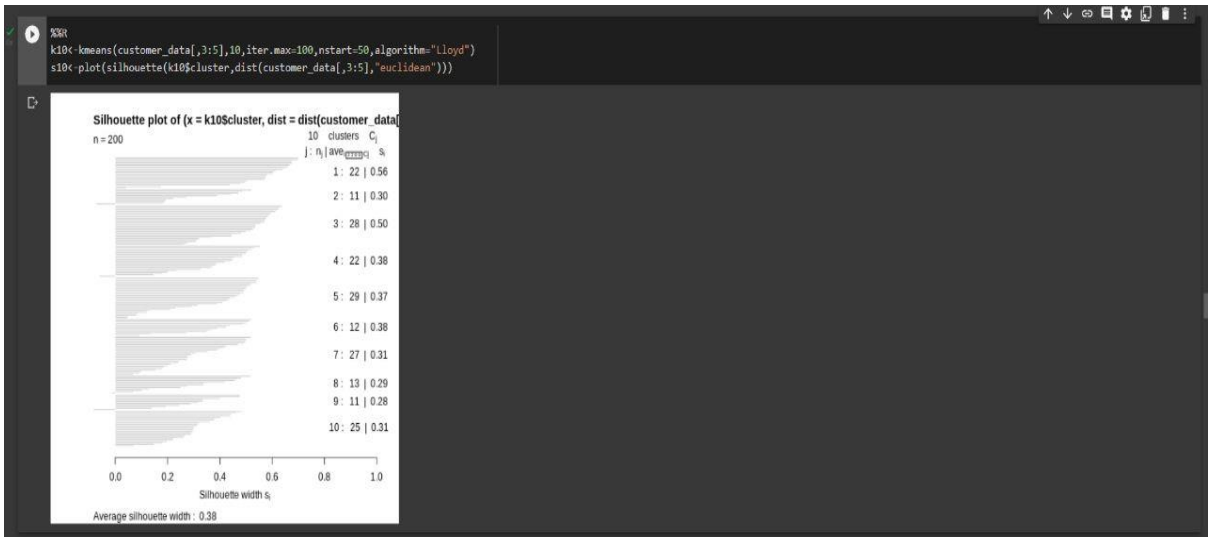


Figure: 27

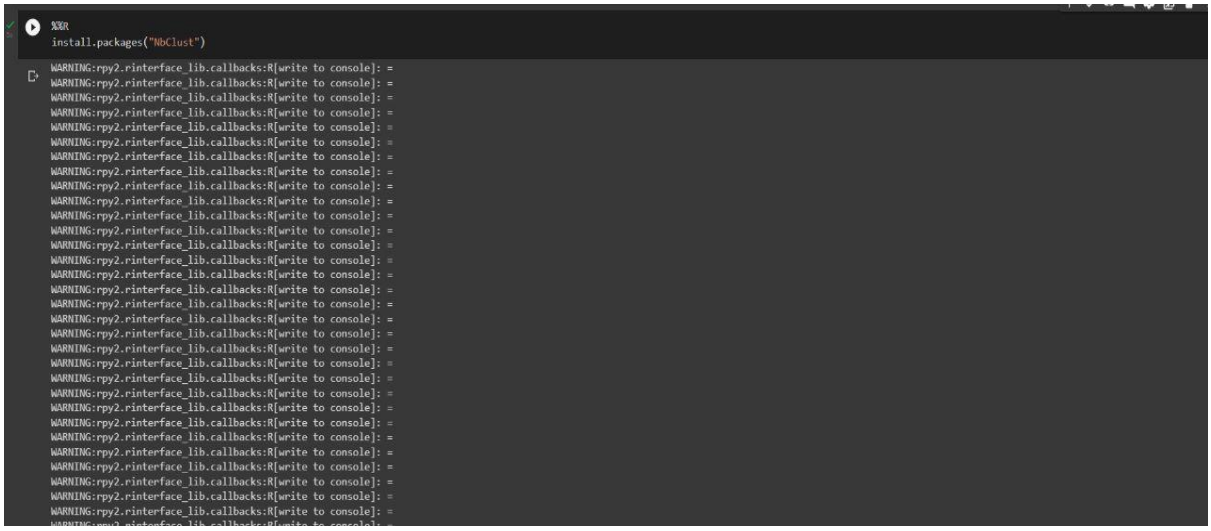


Figure: 28

```
%%R
install.packages("factoextra")
```

```
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]:  length 1386480 bytes (1.3 MB)

WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: =
```

Figure: 29

```
%%R
library(NbClust)
library(factoextra)

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

```
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: Loading required package: ggplot2

WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

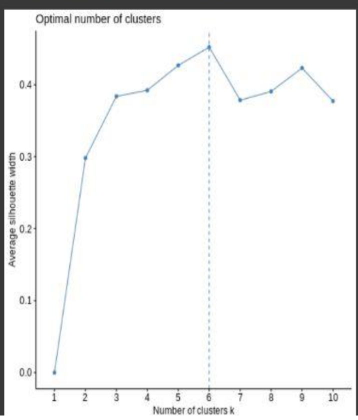


Figure: 30

GAP STATISTIC

Gap Statistic Method

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

```
Clustering k = 1,2,..., K.max (= 10): .. done
Bootstrapping, b = 1,2,..., B (= 50)  [one "." per sample]:
.................................................. 50
```

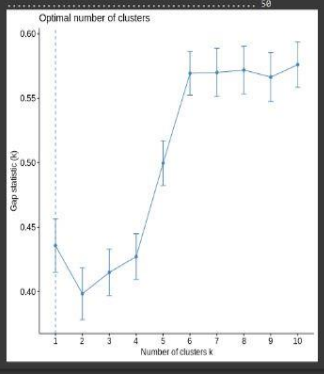


Figure: 31

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

```
K-means clustering with 6 clusters of sizes 45, 21, 35, 39, 38, 22

Cluster means:
      Age Annual.Income..k.. Spending.Score..1.100.
1 56.15556          53.37778               49.08889
2 44.14286          25.14286               19.52381
3 41.68571          88.22857               17.28571
4 32.69231          86.53846               82.12821
5 27.00000          56.65789               49.13158
6 25.27273          25.72727               79.36364

Clustering vector:
  [1] 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2
 [38] 6 2 6 1 6 1 5 2 6 1 5 5 5 1 5 5 1 1 1 1 1 5 1 1 5 1 1 5 1 1 5 5 1 1 1 1 1
 [75] 1 5 1 5 5 1 1 5 1 1 5 1 1 5 5 1 1 5 1 5 5 5 1 5 1 5 5 1 1 5 1 5 1 1 1 1 1
[112] 5 5 5 5 5 1 1 1 1 5 5 5 4 5 4 3 4 3 4 3 4 5 4 3 4 3 4 3 4 3 4 5 4 3 4 3 4
[149] 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3
[186] 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4

Within cluster sum of squares by cluster:
[1]  8062.133  7732.381 16690.857 13972.359  7742.895  4099.818
 (between_SS / total_SS =  81.1 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"
```

Figure: 32

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

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
```

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

```
K-means clustering with 6 clusters of sizes 45, 21, 35, 39, 38, 22

Cluster means:
       Age Annual.Income..k.. Spending.Score..1.100.
1 56.15556           53.37778               49.08889
2 44.14286           25.14286               19.52381
3 41.68571           88.22857               17.28571
4 32.69231           86.53846               82.12821
5 27.00000           56.65789               49.13158
6 25.27273           25.72727               79.36364

Clustering vector:
  [1] 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2 6 2
 [38] 6 2 6 1 6 1 5 2 6 1 5 5 5 1 5 5 1 1 1 1 1 5 1 1 5 1 1 1 5 1 1 5 5 1 1 1 1
 [75] 1 5 1 5 5 1 1 5 1 1 5 1 1 5 5 1 1 5 1 5 5 5 1 5 1 5 5 1 1 5 1 5 1 1 1 1 1
[112] 5 5 5 5 5 1 1 1 1 5 5 5 4 5 4 3 4 3 4 3 4 5 4 3 4 3 4 3 4 3 4 5 4 3 4 3 4
[149] 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3
[186] 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4

Within cluster sum of squares by cluster:
[1]  8062.133  7732.381 16690.857 13972.359  7742.895  4099.818
 (between_SS / total_SS =  81.1 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
```

Figure: 33

SCATTER PLOT

```
%%R
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")
```

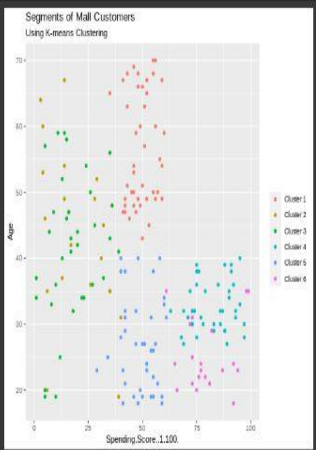


Figure: 34

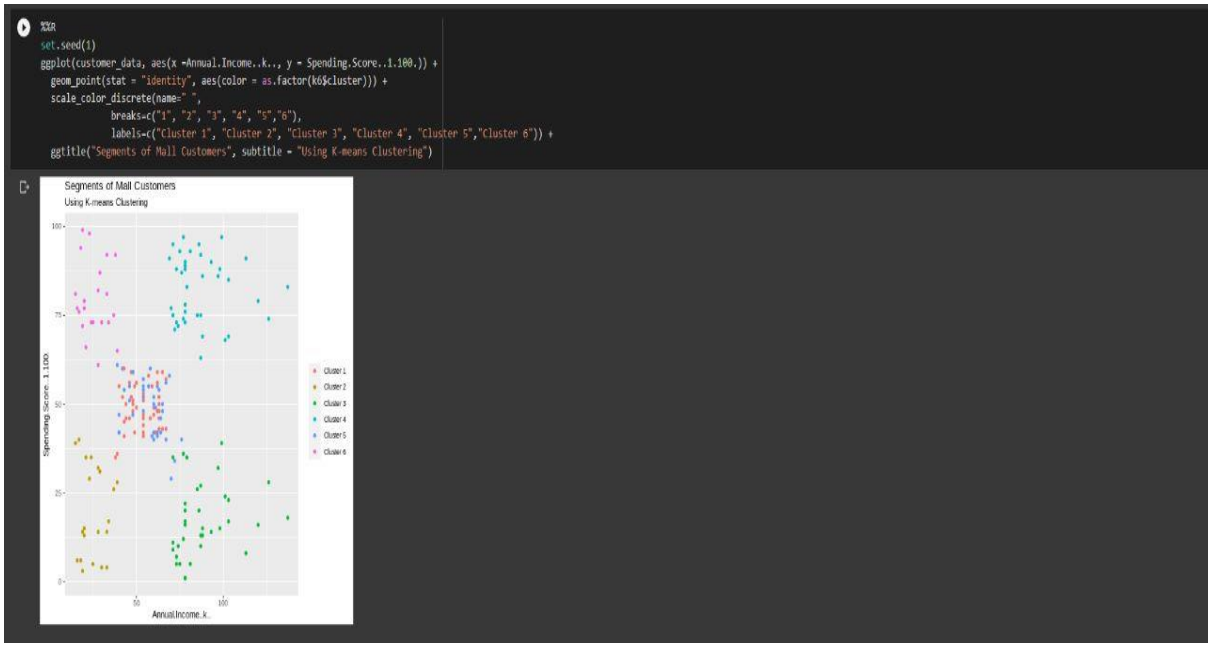


Figure: 35

RESULT

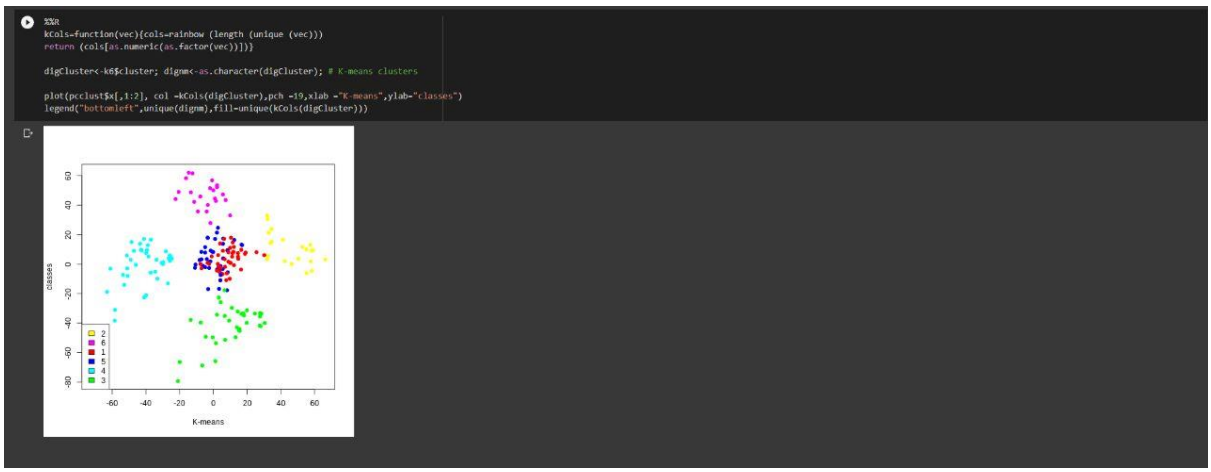


Figure: 36