FML FINAL PROJECT

PRAVALIKA

2022-12-17

##calling the required library

```
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 4.2.2
```

```
## Loading required package: ggplot2
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

Reading the csv file

```
Mall_Data<- read.csv("C:/Users/girne/Downloads/Mall_Customers.csv")
```

##printing the top portion data file

```
head(Mall_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
```

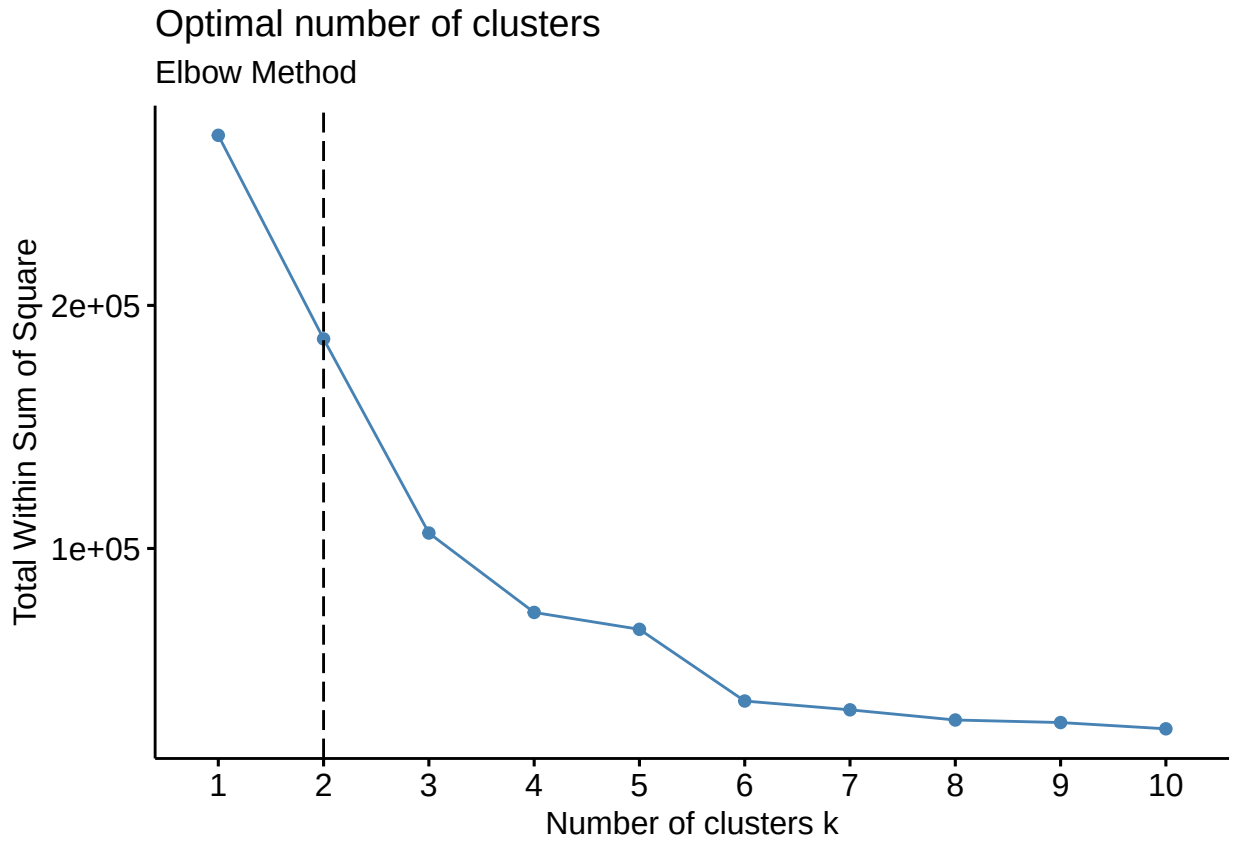
##Taking the quantitative variables in order to scale.

```
Mall_Data1<-Mall_Data[,4:5]
head(Mall_Data1)
```

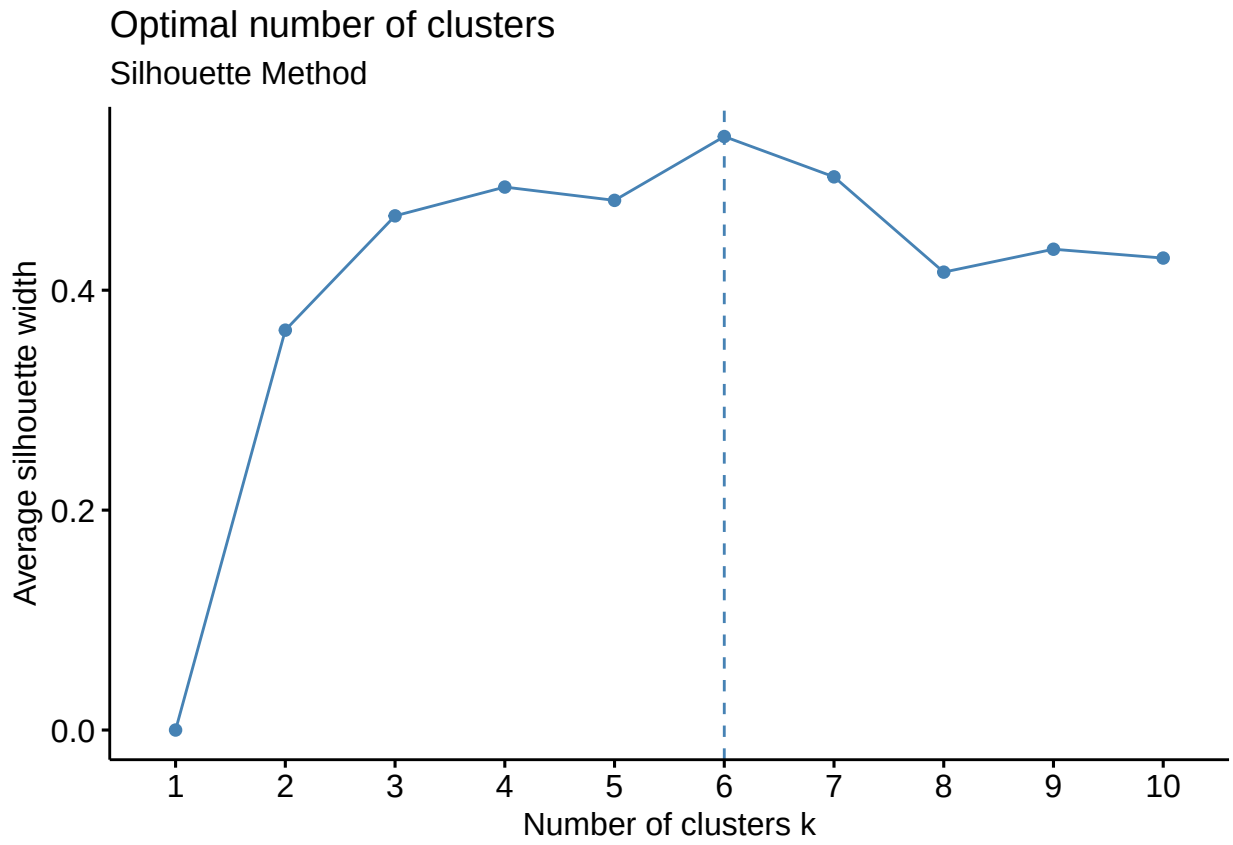
```
##   Annual.Income..k.. Spending.Score..1.100.
## 1                 15                     39
## 2                 15                     81
## 3                 16                      6
## 4                 16                     77
## 5                 17                     40
## 6                 17                     76
```

#Finding the value of K-means using unsupervised learning. Wanted to use the simplest, but most accurate method possible.

```
fviz_nbclust(Mall_Data1,kmeans,method="wss")+geom_vline(xintercept = 2,linetype= 5)+labs(subtitle = "El
```




```
fviz_nbclust(Mall_Data1,kmeans,method ="silhouette") + labs (subtitle = "Silhouette Method")
```




```
#Here, I will set the seed for kmeans.
set.seed(456)
k5<-kmeans(Mall_Data1, centers = 2, nstart = 50)
k5$centers
```

```
##   Annual.Income..k.. Spending.Score..1.100.
## 1           37.28889                50.28889
## 2           79.60000                50.12727
```

```
#Thus, K= 5, meaning that there will be 5 clusters.
```

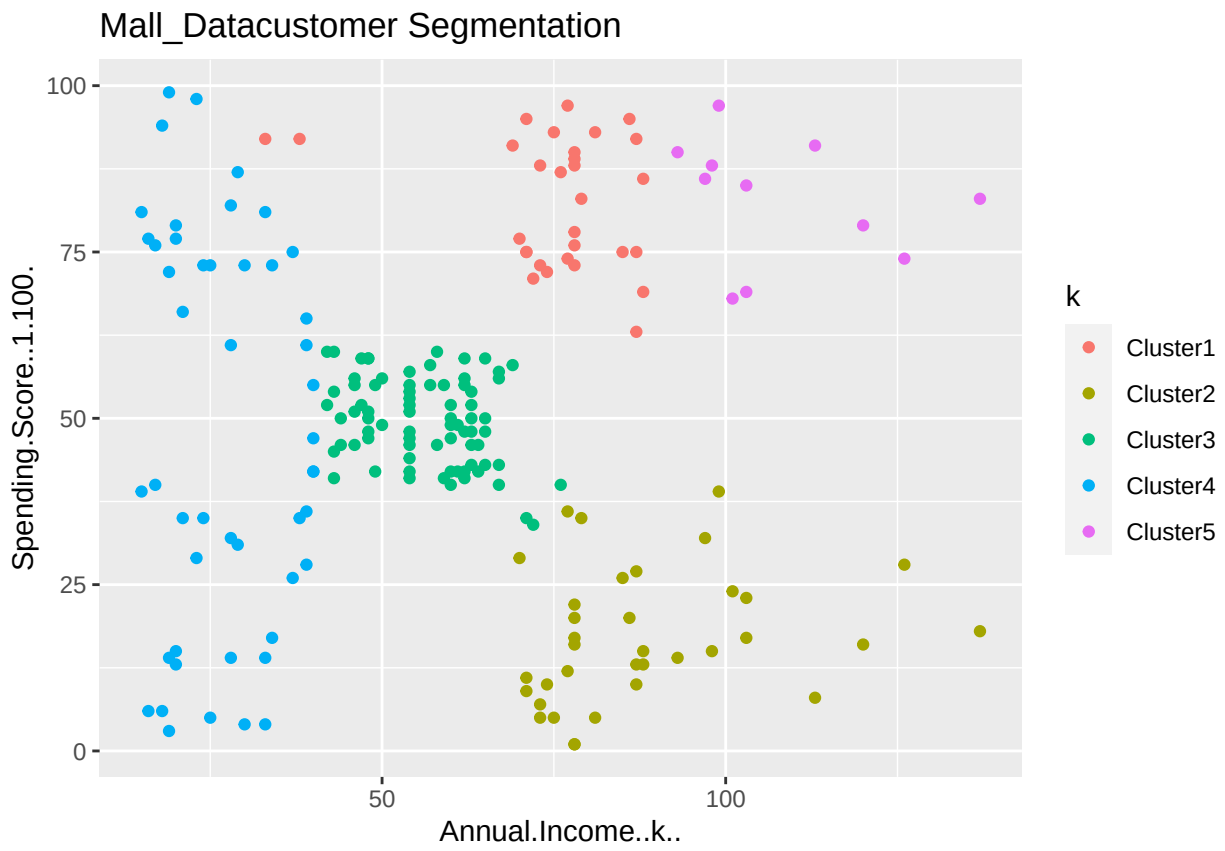
```
#Clustering the data from .csv file.
Mall_Dataclus<-kmeans(Mall_Data1,5)

Mall_Dataclus
```

```
## K-means clustering with 5 clusters of sizes 30, 35, 76, 48, 11
##
## Cluster means:
##   Annual.Income..k.. Spending.Score..1.100.
## 1           75.20000                82.56667
## 2           88.20000                17.11429
## 3           56.31579                49.52632
## 4           27.06250                47.70833
```

```
## 5          108.18182                82.72727
##
## Clustering vector:
##   [1] 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
##  [38] 4 4 4 4 1 4 4 4 4 4 4 4 4 4 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
##  [75] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
## [112] 3 3 3 3 3 3 3 3 3 3 3 1 2 1 3 1 2 1 2 1 3 1 2 1 2 1 2 1 2 1 3 1 2 1 2 1
## [149] 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 5 2 5 2 5 2
## [186] 5 2 5 2 5 2 5 2 5 2 5 2 5 2 5
##
## Within cluster sum of squares by cluster:
## [1]  7018.167 12511.143  8313.368 44258.729  2823.818
##  (between_SS / total_SS =  72.2 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
fviz_cluster(Mall_Dataclus,Mall_Data1)
```


```
#Now, I will visualize the clusters.
ggplot(Mall_Data1, aes(x = Annual.Income..k..,y = Spending.Score..1.100.)) +geom_point(stat = "identity
```




```
#Thus, the following can be concluded:
#Cluster 1 are Mall_Datacustomer who earn a medium annual income and have a medium annual spending rate
#Cluster 2 Mall_Datacustomer who have a high annual income and a low annual spending rate.
#Cluster 3 costumers who have low annual incomes and a high annual spending rates.
#Cluster 4 Mall_Datacustomer with high annual incomes and have high annual spending rates.
#Cluster 5 shows that Mall_Datacustomer with low annual incomes and low annual spending rates.
```