ML-Assignment 4

2022-11-04

```
library(caret)
```

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

```
## Loading required package: lattice
```

```
library(tidyverse)
```

```
## ── Attaching packages
## ─────────────────────────────────────
## tidyverse 1.3.2 ──
```

```
## ✓ tibble  3.1.8      ✓ dplyr   1.0.10
## ✓ tidyr   1.2.1      ✓ stringr 1.4.1
## ✓ readr   2.1.3      ✓ forcats 0.5.2
## ✓ purrr   0.3.5
## ── Conflicts ──────────────────────────────────────── tidyverse_conflict
s() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ✖ purrr::lift()   masks caret::lift()
```

```
library(factoextra)
```

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

```
library(esquisse)
set.seed(123)
```

```
getwd()
```

```
## [1] "/Users/thupiliabhinav/Desktop/ML/ML- Assignment 4"
```

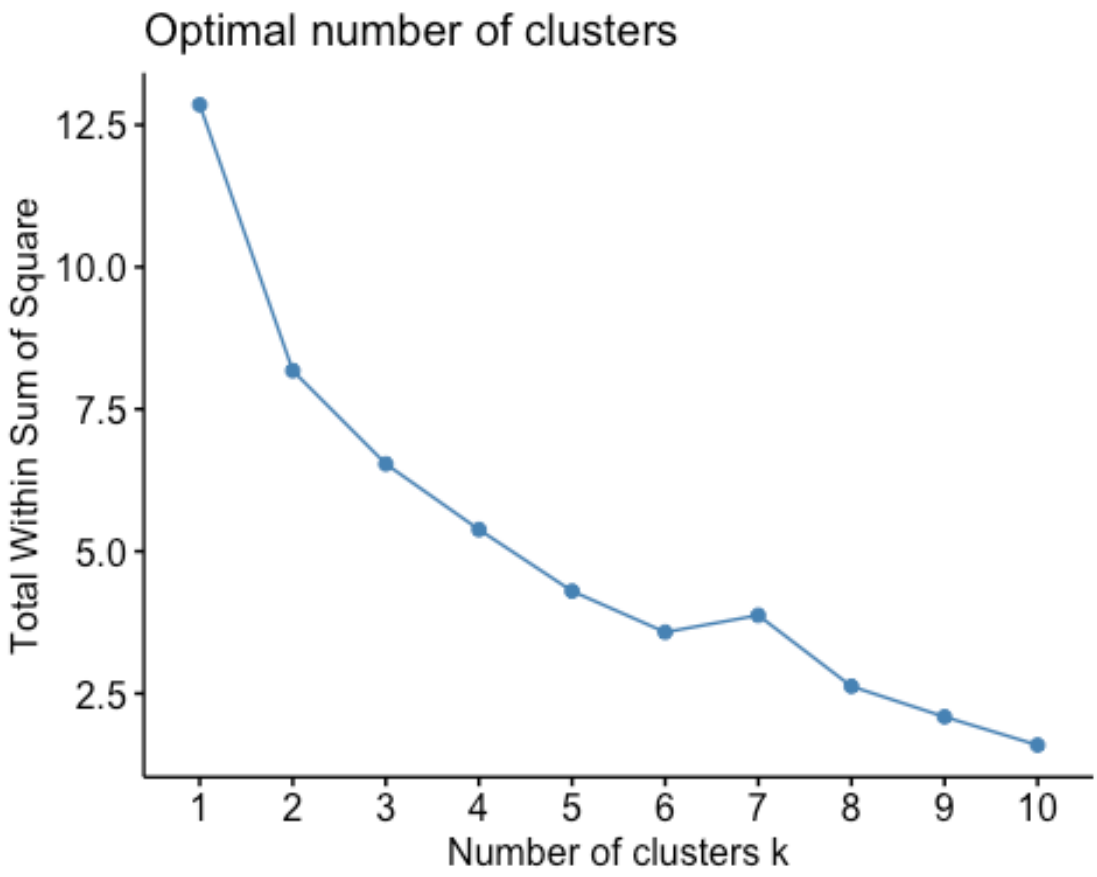
```
setwd("/Users/thupiliabhinav/Desktop/ML/ML- Assignment 4")
pharma <- read.csv("Pharmaceuticals.csv")
```

#a. Using only the numerical variables (1 to 9) to cluster the 21.

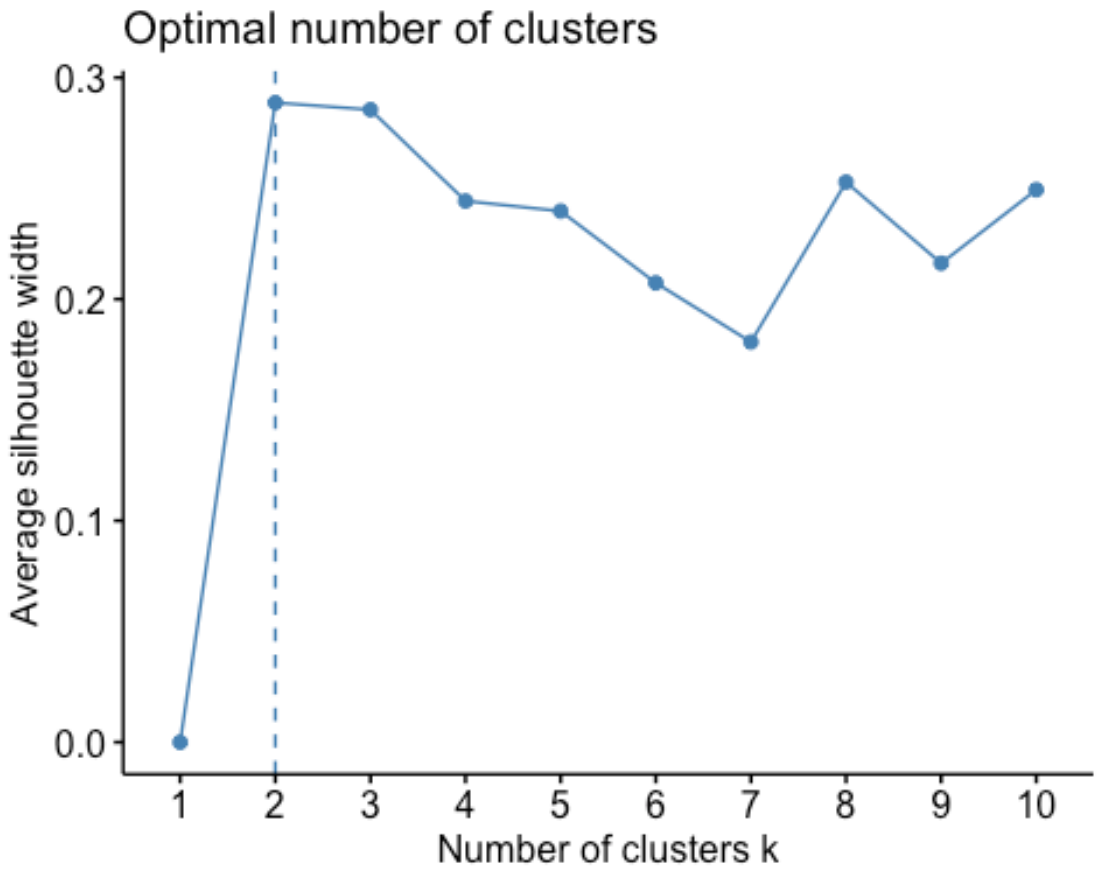
```
norm_mean<- pharma %>% select('Market_Cap', 'Beta', 'PE_Ratio', 'ROE', 'ROA',
'Asset_Turnover', 'Leverage', 'Rev_Growth', 'Net_Profit_Margin')

#Scaling the Data.
norm_train <- preProcess(norm_mean, method = "range")
norm_predict<-predict(norm_train, norm_mean)
```

```
fviz_nbclust(norm_predict, kmeans, method = "wss")
```




```
fviz_nbclust(norm_predict, kmeans, method = "silhouette")
```



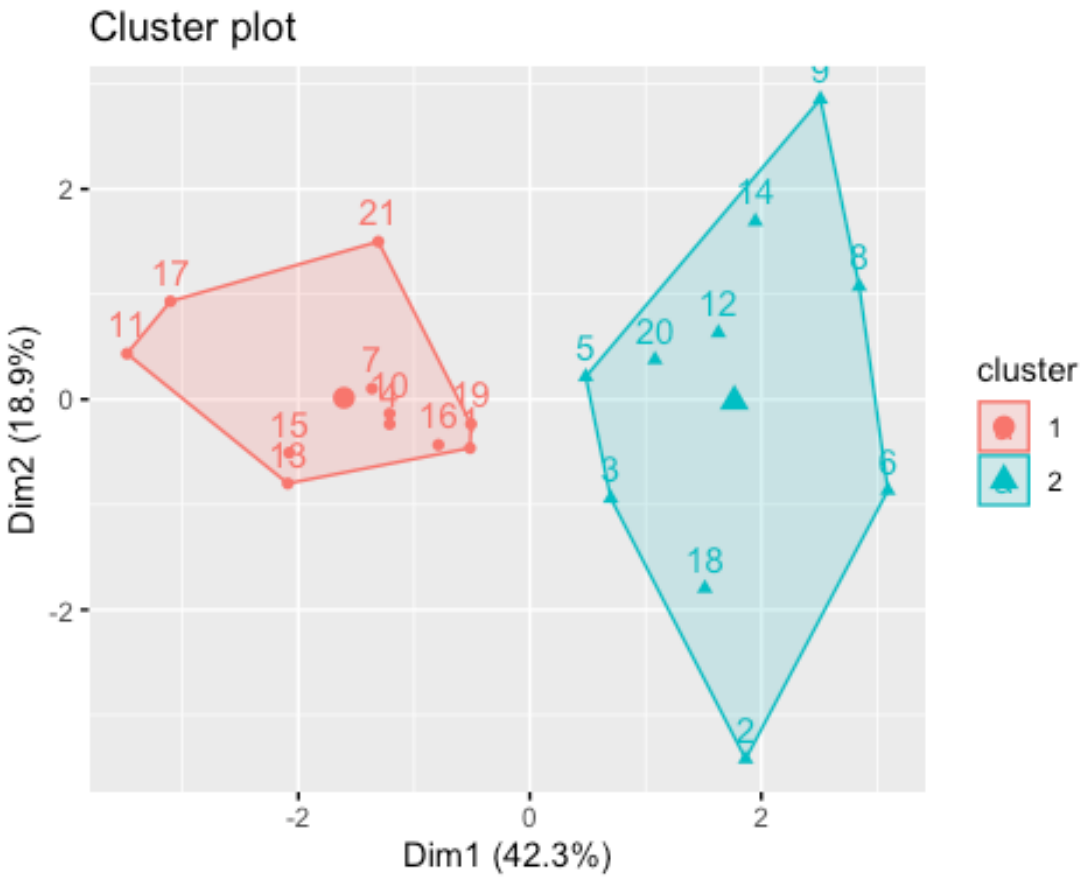
#From above we calculate kmeans optimal being k=2:

```
k_means_2 <- kmeans(norm_predict, centers = 2, nstart = 25)
k_means_2$centers
```

```
##   Market_Cap      Beta  PE_Ratio       ROE       ROA Asset_Turnover   Leve
rage
## 1 0.48580145 0.2727273 0.2199562 0.5389831 0.7171717         0.6250 0.0927
2209
## 2 0.06949161 0.4806452 0.3399240 0.1864407 0.2238095         0.3625 0.2484
3305
##   Rev_Growth Net_Profit_Margin
## 1  0.3567294         0.7673680
## 2  0.5368646         0.3567686
```

#Graphical representation of kmeans using cluster:

```
fviz_cluster(k_means_2, data= norm_mean)
```



#Grouping of clusters with original data:

```
k_cluster<- k_means_2$cluster
group_k <- cbind(pharma,k_cluster)
```

#Calculating mean for both clusters:

```
aggregate(group_k[,-c(1,2,12:14)],by=list(group_k$k_cluster),FUN="mean")
```

```
##   Group.1 Market_Cap      Beta PE_Ratio  ROE      ROA Asset_Turnover  Leve
rage
## 1       1   97.11364 0.4336364 20.95455 35.7 14.95455           0.80 0.325
4545
## 2       2   14.24300 0.6270000 30.42000 14.9  5.63000           0.59 0.872
0000
##   Rev_Growth Net_Profit_Margin k_cluster
## 1   10.16455          20.17273         1
## 2   16.89800          10.77000         2
```

#b). Interpret the clusters with respect to the numerical variables used in forming the clusters.

From above we can observe through clustering by "WSS" and "Silhouette" optimal K is 2 #Cluster 1- has companies with High-Market_Cap, PE_Ratio, ROE, ROA, and Net_Profit_Margin.

#Cluster 2- has companies with Low- Market_Cap, PE_Ratio, ROE, ROA and Net_Profit_Margin.

#Grouping of clusters with Original Data:

```
groupk2<-cbind(group_k, pharma$Location,pharma$Exchange)
```

#c). Is there a pattern in the clusters with respect to the numerical variables (10 to 12)? #With respect to numerical values to columns (10 to 12) are as follows:

#Analysis under column 10.Mediation_recommendation:
#Mediation_recommendation under cluster 1 consists-

a) hold recommendations-6 , b) buy recommendations-3,

c) sell recommendations - 2

##Mediation_recommendation under cluster 2 consists-

a) buy recommendations-5 , b) hold recommendations-3, c) sell recommendations - 2

#Analysis under columns 11.pharma$Location and 12.pharma$Exchange:
#Majority of pharma locations in cluster-1 and cluster-2 are US- based and for pharma exchange the majority is NYSE for both the clusters.

#**d).**Naming for each cluster using any or all of the variables in the dataset:
#Cluster 1- With majority mediation recommendations being held, this cluster is named "HOLD CLUSTER".

#Cluster 2- With majority mediation recommendations being bought, this cluster is named "BUY CLUSTER".