Assignment_4

Tejasvini Mavuleti

3/19/2022

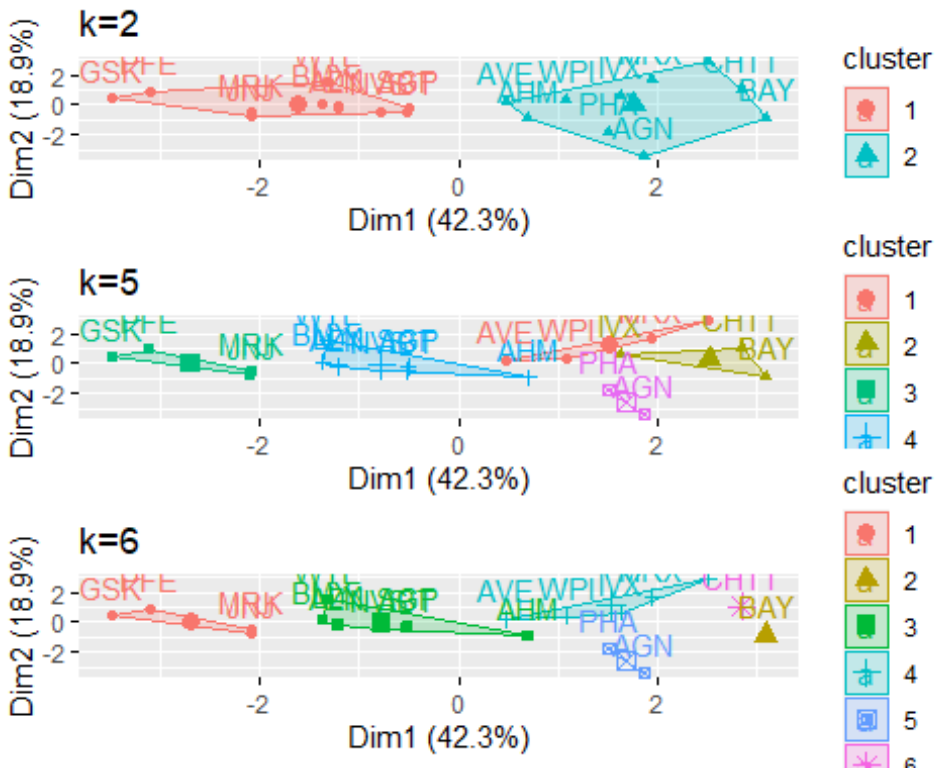
```
Pharmaceuticals <- read.csv("C:/Users/mavul/Downloads/Pharmaceuticals.csv")
View(Pharmaceuticals)

getwd()

## [1] "C:/Users/mavul/OneDrive/Documents/Assignment_4"

setwd("C:/Users/mavul/OneDrive/Documents")
Pharmaceuticaldata<- read.csv("C:/Users/mavul/Downloads/Pharmaceuticals.csv")
str(Pharmaceuticaldata)

## 'data.frame':    21 obs. of  14 variables:
##  $ Symbol               : chr  "ABT" "AGN" "AHM" "AZN" ...
##  $ Name                 : chr  "Abbott Laboratories" "Allergan, Inc."
"Amersham plc" "AstraZeneca PLC" ...
##  $ Market_Cap           : num  68.44 7.58 6.3 67.63 47.16 ...
##  $ Beta                 : num  0.32 0.41 0.46 0.52 0.32 1.11 0.5 0.85 1.08
0.18 ...
##  $ PE_Ratio             : num  24.7 82.5 20.7 21.5 20.1 27.9 13.9 26 3.6
27.9 ...
##  $ ROE                  : num  26.4 12.9 14.9 27.4 21.8 3.9 34.8 24.1 15.1
31 ...
##  $ ROA                  : num  11.8 5.5 7.8 15.4 7.5 1.4 15.1 4.3 5.1 13.5
...
##  $ Asset_Turnover       : num  0.7 0.9 0.9 0.9 0.6 0.6 0.9 0.6 0.3 0.6 ...
##  $ Leverage             : num  0.42 0.6 0.27 0 0.34 0 0.57 3.51 1.07 0.53
...
##  $ Rev_Growth           : num  7.54 9.16 7.05 15 26.81 ...
##  $ Net_Profit_Margin    : num  16.1 5.5 11.2 18 12.9 2.6 20.6 7.5 13.3
23.4 ...
##  $ Median_Recommendation: chr  "Moderate Buy" "Moderate Buy" "Strong Buy"
"Moderate Sell" ...
##  $ Location             : chr  "US" "CANADA" "UK" "UK" ...
##  $ Exchange             : chr  "NYSE" "NYSE" "NYSE" "NYSE" ...

# Calling libraries

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse
1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(factoextra)

## Warning: package 'factoextra' was built under R version 4.1.3

library(cluster)
library(ggplot2)
library(gridExtra)

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine

# a) Use only the numerical variables (1 to 9) to cluster the 21 firms

Pharmaceuticaldata <- na.omit(Pharmaceuticaldata)
Pharmaceuticaldata

##    Symbol                                Name Market_Cap Beta PE_Ratio  ROE
ROA
## 1     ABT                Abbott Laboratories      68.44 0.32     24.7 26.4
11.8
## 2     AGN                     Allergan, Inc.       7.58 0.41     82.5 12.9
5.5
## 3     AHM                       Amersham plc       6.30 0.46     20.7 14.9
7.8
## 4     AZN                    AstraZeneca PLC      67.63 0.52     21.5 27.4
15.4
## 5     AVE                            Aventis      47.16 0.32     20.1 21.8
7.5
## 6     BAY                           Bayer AG      16.90 1.11     27.9  3.9
1.4
## 7     BMY       Bristol-Myers Squibb Company      51.33 0.50     13.9 34.8
15.1
## 8    CHTT                         Chattem, Inc       0.41 0.85     26.0 24.1
4.3
## 9     ELN              Elan Corporation, plc       0.78 1.08      3.6 15.1
5.1
## 10    LLY              Eli Lilly and Company      73.84 0.18     27.9 31.0
13.5
## 11    GSK                GlaxoSmithKline plc     122.11 0.35     18.0 62.9
20.3
```

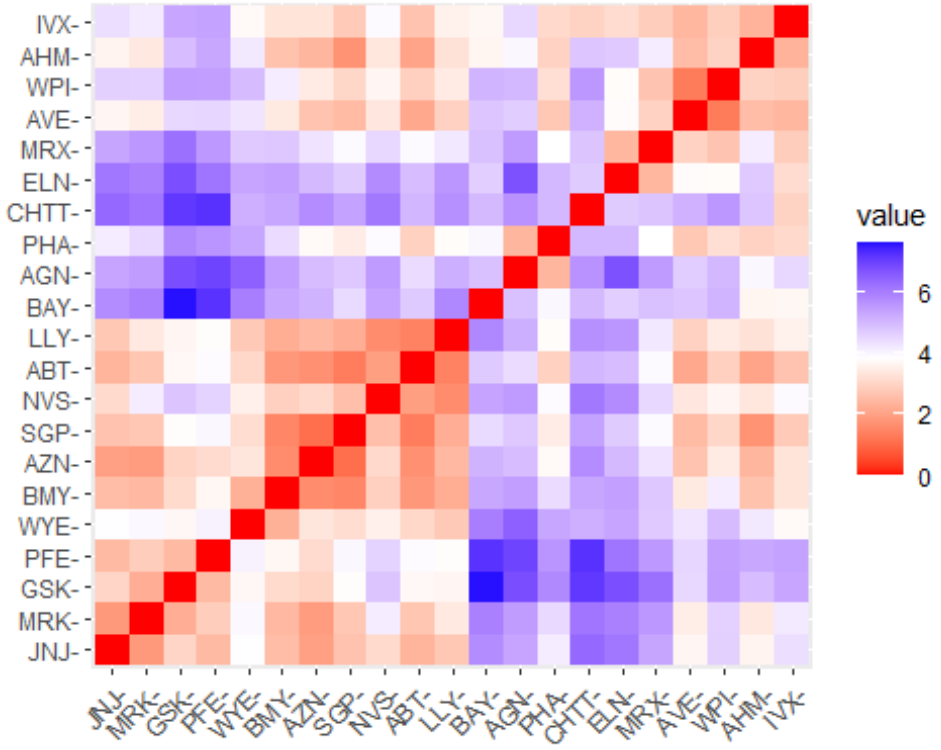
```
## 12    IVX                   IVAX Corporation       2.60 0.65     19.9 21.4
6.8
## 13    JNJ                  Johnson & Johnson     173.93 0.46     28.4 28.6
16.3
## 14    MRX Medicis Pharmaceutical Corporation       1.20 0.75     28.6 11.2
5.4
## 15    MRK                  Merck & Co., Inc.     132.56 0.46     18.9 40.6
15.0
## 16    NVS                        Novartis AG      96.65 0.19     21.6 17.9
11.2
## 17    PFE                         Pfizer Inc     199.47 0.65     23.6 45.6
19.2
## 18    PHA              Pharmacia Corporation      56.24 0.40     56.5 13.5
5.7
## 19    SGP        Schering-Plough Corporation      34.10 0.51     18.9 22.6
13.3
## 20    WPI       Watson Pharmaceuticals, Inc.       3.26 0.24     18.4 10.2
6.8
## 21    WYE                              Wyeth      48.19 0.63     13.1 54.9
13.4
##    Asset_Turnover Leverage Rev_Growth Net_Profit_Margin
Median_Recommendation
## 1             0.7     0.42       7.54              16.1          Moderate
Buy
## 2             0.9     0.60       9.16               5.5          Moderate
Buy
## 3             0.9     0.27       7.05              11.2            Strong
Buy
## 4             0.9     0.00      15.00              18.0         Moderate
Sell
## 5             0.6     0.34      26.81              12.9          Moderate
Buy
## 6             0.6     0.00      -3.17               2.6
Hold
## 7             0.9     0.57       2.70              20.6         Moderate
Sell
## 8             0.6     3.51       6.38               7.5          Moderate
Buy
## 9             0.3     1.07      34.21              13.3         Moderate
Sell
## 10            0.6     0.53       6.21              23.4
Hold
## 11            1.0     0.34      21.87              21.1
Hold
## 12            0.6     1.45      13.99              11.0
Hold
## 13            0.9     0.10       9.37              17.9          Moderate
Buy
## 14            0.3     0.93      30.37              21.3          Moderate
Buy
```

```
## 15            1.1     0.28      17.35              14.1
Hold
## 16            0.5     0.06      -2.69              22.4
Hold
## 17            0.8     0.16      25.54              25.2          Moderate
Buy
## 18            0.6     0.35      15.00               7.3
Hold
## 19            0.8     0.00       8.56              17.6
Hold
## 20            0.5     0.20      29.18              15.1         Moderate
Sell
## 21            0.6     1.12       0.36              25.5
Hold
##       Location Exchange
## 1           US     NYSE
## 2       CANADA     NYSE
## 3           UK     NYSE
## 4           UK     NYSE
## 5       FRANCE     NYSE
## 6      GERMANY     NYSE
## 7           US     NYSE
## 8           US   NASDAQ
## 9      IRELAND     NYSE
## 10          US     NYSE
## 11          UK     NYSE
## 12          US     AMEX
## 13          US     NYSE
## 14          US     NYSE
## 15          US     NYSE
## 16 SWITZERLAND     NYSE
## 17          US     NYSE
## 18          US     NYSE
## 19          US     NYSE
## 20          US     NYSE
## 21          US     NYSE

row.names(Pharmaceuticaldata)<- Pharmaceuticaldata[,1]
Pharmadata<- Pharmaceuticaldata[, 3:11]
head(Pharmadata)

##     Market_Cap Beta PE_Ratio  ROE  ROA Asset_Turnover Leverage Rev_Growth
## ABT      68.44 0.32     24.7 26.4 11.8            0.7     0.42       7.54
## AGN       7.58 0.41     82.5 12.9  5.5            0.9     0.60       9.16
## AHM       6.30 0.46     20.7 14.9  7.8            0.9     0.27       7.05
## AZN      67.63 0.52     21.5 27.4 15.4            0.9     0.00      15.00
## AVE      47.16 0.32     20.1 21.8  7.5            0.6     0.34      26.81
## BAY      16.90 1.11     27.9  3.9  1.4            0.6     0.00      -3.17
##     Net_Profit_Margin
## ABT              16.1
```

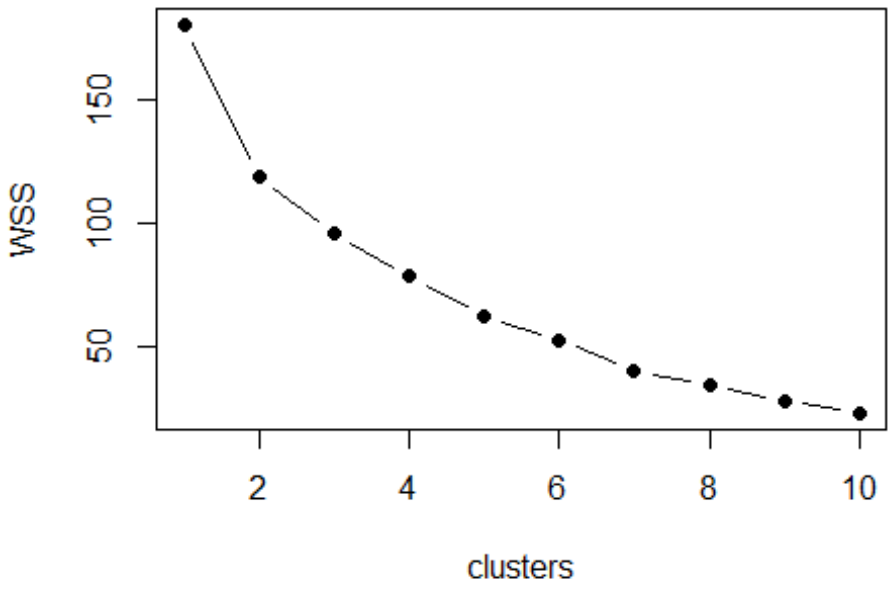
```
## AGN               5.5
## AHM              11.2
## AZN              18.0
## AVE              12.9
## BAY               2.6

# Scaling the Pharmadata using the scale function

Pharma_scale <- scale(Pharmadata)
head(Pharma_scale)

##     Market_Cap        Beta    PE_Ratio         ROE        ROA
Asset_Turnover
## ABT  0.1840960 -0.80125356 -0.04671323  0.04009035  0.2416121
0.0000000
## AGN -0.8544181 -0.45070513  3.49706911 -0.85483986 -0.9422871
0.9225312
## AHM -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700
0.9225312
## AZN  0.1702742 -0.02225704 -0.24290879  0.10638147  0.9181259
0.9225312
## AVE -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.5664461     -
0.4612656
## BAY -0.6953818  2.27578267  0.14948233 -1.45146000 -1.7127612     -
0.4612656
##       Leverage Rev_Growth Net_Profit_Margin
## ABT -0.2120979 -0.5277675        0.06168225
## AGN  0.0182843 -0.3811391       -1.55366706
## AHM -0.4040831 -0.5721181       -0.68503583
## AZN -0.7496565  0.1474473        0.35122600
## AVE -0.3144900  1.2163867       -0.42597037
## BAY -0.7496565 -1.4971443       -1.99560225

# Computing K-means clustering and using multiple values of K and examine the
difference

km1 <- kmeans(Pharma_scale, centers = 2, nstart = 30)
km2<- kmeans(Pharma_scale, centers = 5, nstart = 30)
km3<- kmeans(Pharma_scale, centers = 6, nstart = 30)
Plot1<-fviz_cluster(km1, data = Pharma_scale)+ggtitle("k=2")
plot2<-fviz_cluster(km2, data = Pharma_scale)+ggtitle("k=5")
plot3<-fviz_cluster(km3, data = Pharma_scale)+ggtitle("k=6")
grid.arrange(Plot1,plot2,plot3, nrow = 3)
```




```
distance<- dist(Pharma_scale, method = "euclidean")
fviz_dist(distance)
```



```r
# Computing and ploting wss for k = 1 to k = 10 and extracting wss for 2-15 clusters
# The location of a elbow in the plot is considered as an indicator of the number of clusters k =5

set.seed(64060)
wss<- function(k){kmeans(Pharma_scale, k, nstart =10)$tot.withinss}
k.values<- 1:10
wss_clusters<- map_dbl(k.values, wss)
plot(k.values, wss_clusters, type="b", pch = 16, frame = TRUE,
xlab="clusters", ylab="WSS")
```

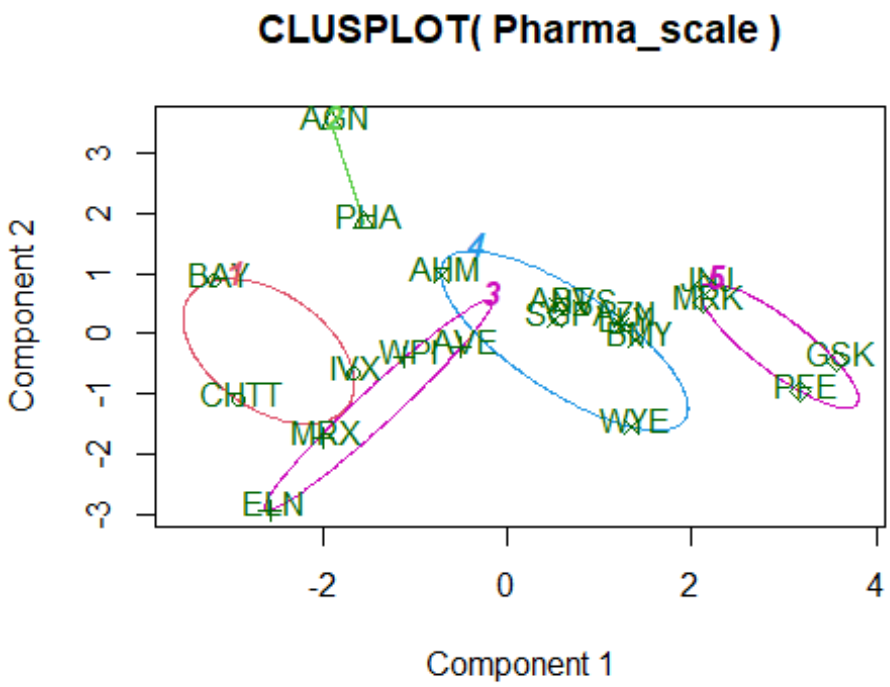



```r
# Final analysis and extracting results using 5 clusters
set.seed(64060)
final_Cluster<- kmeans(Pharma_scale, 5, nstart = 25)
print(final_Cluster)
```

```
## K-means clustering with 5 clusters of sizes 3, 2, 4, 8, 4
##
## Cluster means:
##    Market_Cap       Beta    PE_Ratio        ROE        ROA Asset_Turnover
## 1 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478     -0.4612656
## 2 -0.43925134 -0.4701800  2.70002464 -0.8349525 -0.9234951      0.2306328
## 3 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428     -1.2684804
## 4 -0.03142211 -0.4360989 -0.31724852  0.1950459  0.4083915      0.1729746
## 5  1.69558112 -0.1780563 -0.19845823  1.2349879  1.3503431      1.1531640
##      Leverage Rev_Growth Net_Profit_Margin
```

```
## 1  1.36644699 -0.6912914      -1.320000179
## 2 -0.14170336 -0.1168459      -1.416514761
## 3  0.06308085  1.5180158      -0.006893899
## 4 -0.27449312 -0.7041516       0.556954446
## 5 -0.46807818  0.4671788       0.591242521
##
## Clustering vector:
##  ABT  AGN  AHM  AZN  AVE  BAY  BMY CHTT  ELN  LLY  GSK  IVX  JNJ  MRX  MRK
NVS
##    4    2    4    4    3    1    4    1    3    4    5    1    5    3    5
4
##  PFE  PHA  SGP  WPI  WYE
##    5    2    4    3    4
##
## Within cluster sum of squares by cluster:
## [1] 15.595925  2.803505 12.791257 21.879320  9.284424
##  (between_SS / total_SS =  65.4 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
"tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
clusplot(Pharma_scale,final_Cluster$cluster, color = TRUE, labels = 2,lines =
0)
```



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

Cluster 1 - BAY,IVX,CHTT

Cluster 2 - AGN,PHA

Cluster 3 - ELN,AVE,WPI,MRX

Cluster 4 - BMY,WYE,AHM,ABT,NVS,AZN,LLY

Cluster 5 - JNJ,MRK,GSK,PFE

Pharma_Cluster <- Pharmaceuticaldata[,c(12,13,14)]%>% mutate(clusters = final_Cluster$cluster)%>% arrange(clusters, ascending = TRUE) Pharma_Cluster

c)Is there a pattern in the clusters with respect to the numerical variables (10 to 12)?

plot1<-ggplot(Pharma_Cluster, mapping = aes(factor(clusters), fill=Median_Recommendation))+geom_bar(position = 'dodge')+labs(x ='No of clusters')

plot2<- ggplot(Pharma_Cluster, mapping = aes(factor(clusters),fill = Location))+geom_bar(position = 'dodge')+labs(x ='No of clusters')

plot3<- ggplot(Pharma_Cluster, mapping = aes(factor(clusters),fill = Exchange))+geom_bar(position = 'dodge')+labs(x ='No of clusters') grid.arrange(plot1, plot2, plot3)

As per graph, Cluster 1- has the highest Beta , leverage and lowest market_cap, ROE, ROA, leverage, rev_growth, net_profit_margin

Cluster 2- has the highest PE_ratio

Cluster 3- has the highest rev_growth and the lowest PE_ratio, asset_turnover

Cluster 4- has the highest PE_ratio

Cluster 5- has the highest net_profit_margin and the lowest Beta

Therefore, clusters 1,3 and 5 have the most moderate buying recommendation and the clusters 2 and 4 have hold recommendation

d)Provide an appropriate name for each cluster using any or all of the variables in the dataset.

Cluster 1 - is the hold cluster

Cluster 2 - is the hold-buy cluster

Cluster 3 - is the buy-sell cluster

Cluster 4 - is the strong buy-sell-hold cluster

Cluster 5 - is the hold-buy cluster