FML Assignment4

Priyanka Jonnala

2023-11-12

Loading the required packages

library(flexclust)

## Warning: package 'flexclust' was built under R version 4.3.2

## Loading required package: grid

## Loading required package: lattice

## Loading required package: modeltools

## Loading required package: stats4

library(cluster)

## Warning: package 'cluster' was built under R version 4.3.2

library(ggcorrplot)

## Warning: package 'ggcorrplot' was built under R version 4.3.2

## Loading required package: ggplot2

Importing the dataset

Pharmaceuticals <- read.csv("C:\\Users\\priya\\OneDrive\\Desktop\\FML\\Assignment 4\\Pharmaceuticals.csv")  
head(Pharmaceuticals)

## Symbol Name Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover  
## 1 ABT Abbott Laboratories 68.44 0.32 24.7 26.4 11.8 0.7  
## 2 AGN Allergan, Inc. 7.58 0.41 82.5 12.9 5.5 0.9  
## 3 AHM Amersham plc 6.30 0.46 20.7 14.9 7.8 0.9  
## 4 AZN AstraZeneca PLC 67.63 0.52 21.5 27.4 15.4 0.9  
## 5 AVE Aventis 47.16 0.32 20.1 21.8 7.5 0.6  
## 6 BAY Bayer AG 16.90 1.11 27.9 3.9 1.4 0.6  
## Leverage Rev\_Growth Net\_Profit\_Margin Median\_Recommendation Location Exchange  
## 1 0.42 7.54 16.1 Moderate Buy US NYSE  
## 2 0.60 9.16 5.5 Moderate Buy CANADA NYSE  
## 3 0.27 7.05 11.2 Strong Buy UK NYSE  
## 4 0.00 15.00 18.0 Moderate Sell UK NYSE  
## 5 0.34 26.81 12.9 Moderate Buy FRANCE NYSE  
## 6 0.00 -3.17 2.6 Hold GERMANY NYSE

Choose columns 3 to 11 and store the resulting data frame in Pharma1

Pharma <- na.omit(Pharmaceuticals)  
Pharma

## 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

Pharma1 <- Pharma[,3:11]  
# Displaying the top six rows of Pharma1 using head function  
head(Pharma1)

## Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover Leverage Rev\_Growth  
## 1 68.44 0.32 24.7 26.4 11.8 0.7 0.42 7.54  
## 2 7.58 0.41 82.5 12.9 5.5 0.9 0.60 9.16  
## 3 6.30 0.46 20.7 14.9 7.8 0.9 0.27 7.05  
## 4 67.63 0.52 21.5 27.4 15.4 0.9 0.00 15.00  
## 5 47.16 0.32 20.1 21.8 7.5 0.6 0.34 26.81  
## 6 16.90 1.11 27.9 3.9 1.4 0.6 0.00 -3.17  
## Net\_Profit\_Margin  
## 1 16.1  
## 2 5.5  
## 3 11.2  
## 4 18.0  
## 5 12.9  
## 6 2.6

# Printing summary statistics for Pharma1  
summary(Pharma1)

## Market\_Cap Beta PE\_Ratio ROE   
## Min. : 0.41 Min. :0.1800 Min. : 3.60 Min. : 3.9   
## 1st Qu.: 6.30 1st Qu.:0.3500 1st Qu.:18.90 1st Qu.:14.9   
## Median : 48.19 Median :0.4600 Median :21.50 Median :22.6   
## Mean : 57.65 Mean :0.5257 Mean :25.46 Mean :25.8   
## 3rd Qu.: 73.84 3rd Qu.:0.6500 3rd Qu.:27.90 3rd Qu.:31.0   
## Max. :199.47 Max. :1.1100 Max. :82.50 Max. :62.9   
## ROA Asset\_Turnover Leverage Rev\_Growth   
## Min. : 1.40 Min. :0.3 Min. :0.0000 Min. :-3.17   
## 1st Qu.: 5.70 1st Qu.:0.6 1st Qu.:0.1600 1st Qu.: 6.38   
## Median :11.20 Median :0.6 Median :0.3400 Median : 9.37   
## Mean :10.51 Mean :0.7 Mean :0.5857 Mean :13.37   
## 3rd Qu.:15.00 3rd Qu.:0.9 3rd Qu.:0.6000 3rd Qu.:21.87   
## Max. :20.30 Max. :1.1 Max. :3.5100 Max. :34.21   
## Net\_Profit\_Margin  
## Min. : 2.6   
## 1st Qu.:11.2   
## Median :16.1   
## Mean :15.7   
## 3rd Qu.:21.1   
## Max. :25.5

library(factoextra)

## Warning: package 'factoextra' was built under R version 4.3.2

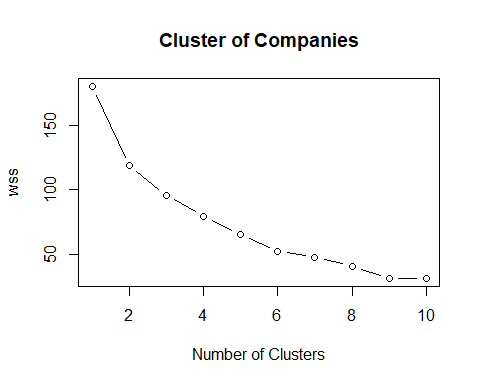
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

Normalizing the data with scale method

Pharma2 <- scale(Pharma1)  
# Set row names to match the first column of the original Pharma data  
row.names(Pharma2) <- Pharma[,1]  
# Calculate the distance matrix using get\_dist  
distance <- get\_dist(Pharma2)  
# Visualize the distance matrix using fviz\_dist  
fviz\_dist(distance)



# Set the random seed for reproducibility  
set.seed(10)  
# Use a for loop to calculate the within-cluster sum of squares (wss) for 1 to 10 clusters  
wss <- vector()  
for(i in 1:10) wss[i] <- sum(kmeans(Pharma2,i)$withinss)  
# Visualize the wss values using a line plot  
plot(1:10, wss , type = "b" , main = paste('Cluster of Companies') , xlab =  
"Number of Clusters", ylab="wss")



# Print the wss values for each number of clusters  
wss

## [1] 180.00000 118.56934 95.99420 79.21748 65.61035 52.67476 47.66961  
## [8] 41.12605 31.81763 31.57252

Silhouette Approach

# Use the fviz\_nbclust function to determine the optimal number of clusters using the silhouette method  
fviz\_nbclust(Pharma2, kmeans, method = "silhouette")

 This demonstrates that five clusters are the optimum number. Using the k-means method to create a 5 cluster.

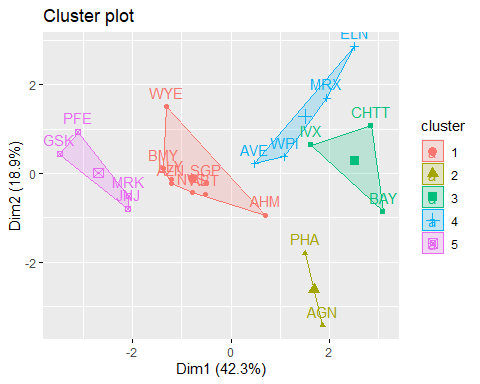
# Use the kmeans function to create 5 clusters and visualize the results using the fviz\_cluster function  
set.seed(21)  
clusterx <- kmeans(Pharma2, centers = 5, nstart = 25) # k = 5, number of restarts = 25  
clusterx$centers

## Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover  
## 1 -0.03142211 -0.4360989 -0.31724852 0.1950459 0.4083915 0.1729746  
## 2 -0.43925134 -0.4701800 2.70002464 -0.8349525 -0.9234951 0.2306328  
## 3 -0.87051511 1.3409869 -0.05284434 -0.6184015 -1.1928478 -0.4612656  
## 4 -0.76022489 0.2796041 -0.47742380 -0.7438022 -0.8107428 -1.2684804  
## 5 1.69558112 -0.1780563 -0.19845823 1.2349879 1.3503431 1.1531640  
## Leverage Rev\_Growth Net\_Profit\_Margin  
## 1 -0.27449312 -0.7041516 0.556954446  
## 2 -0.14170336 -0.1168459 -1.416514761  
## 3 1.36644699 -0.6912914 -1.320000179  
## 4 0.06308085 1.5180158 -0.006893899  
## 5 -0.46807818 0.4671788 0.591242521

clusterx$size

## [1] 8 2 3 4 4

fviz\_cluster(clusterx, data = Pharma2)

 Manhattan Distance when using Kmeans Clustering.

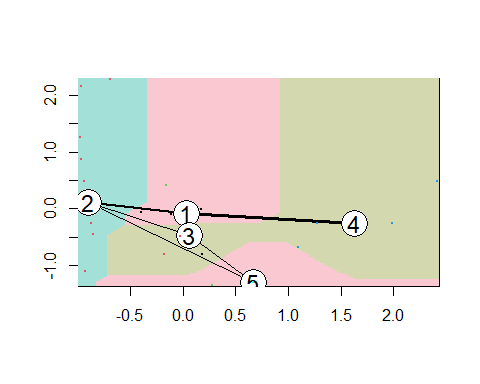
set.seed(21)  
# Use kcca function to create 5 clusters with Manhattan distance and k-medians algorithm  
clusterY <- kcca(Pharma2, k = 5, kccaFamily("kmedians"))  
# Print the results and visualize the clusters  
clusterY

## kcca object of family 'kmedians'   
##   
## call:  
## kcca(x = Pharma2, k = 5, family = kccaFamily("kmedians"))  
##   
## cluster sizes:  
##   
## 1 2 3 4 5   
## 4 10 2 4 1

clusters\_index <- predict(clusterY)  
dist(clusterY@centers)

## 1 2 3 4  
## 2 2.663586   
## 3 2.113529 3.531320   
## 4 2.359668 4.474483 3.022624   
## 5 2.582322 3.396689 2.360814 3.868401

image(clusterY)  
points(Pharma2, col = clusters\_index, pch = 19, cex = 0.3)



library(tidyverse)

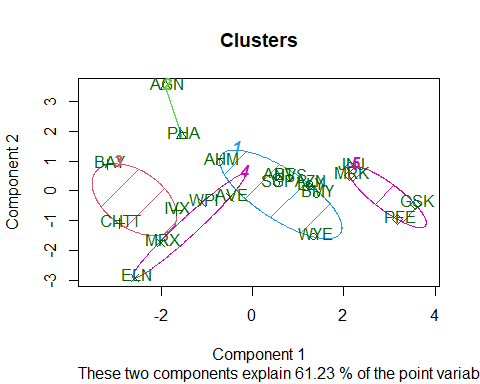
## Warning: package 'tidyverse' was built under R version 4.3.2

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.3 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ lubridate 1.9.2 ✔ tibble 3.2.1  
## ✔ purrr 1.0.2 ✔ tidyr 1.3.0  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

Pharma1 %>% mutate(Cluster = clusterx$cluster) %>% group\_by(Cluster) %>% summarise\_all("mean")

## # A tibble: 5 × 10  
## Cluster Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover Leverage  
## <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 55.8 0.414 20.3 28.7 12.7 0.738 0.371  
## 2 2 31.9 0.405 69.5 13.2 5.6 0.75 0.475  
## 3 3 6.64 0.87 24.6 16.5 4.17 0.6 1.65   
## 4 4 13.1 0.598 17.7 14.6 6.2 0.425 0.635  
## 5 5 157. 0.48 22.2 44.4 17.7 0.95 0.22   
## # ℹ 2 more variables: Rev\_Growth <dbl>, Net\_Profit\_Margin <dbl>

clusplot(Pharma2,clusterx$cluster, main="Clusters",color = TRUE,shade = TRUE, labels = 2,lines = 0)

 Companies are classified into different clusters as follows

Cluster1 : AHM,WYE,BMY,AZN,LLY,ABT,NVS,ABT and SGP

Cluster2 : AGN,PHA

Cluster3 : BAY,CHTT,IVX

Cluster4 : ELN,MRX,WPI,AVE

Cluster5 : JNJ,MRK,PFE,GSK

From the means of the cluster variables, it can be derived as follow:

Cluster1 has a medium risk

Cluster2 has very high PE Ratio

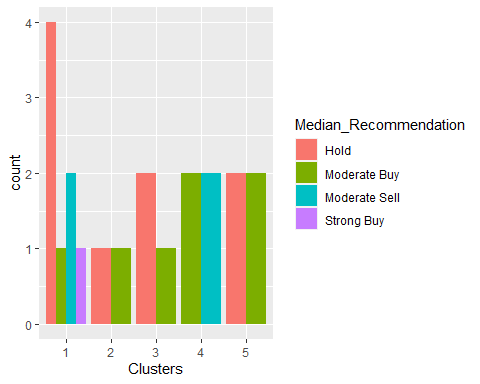
Cluster3 Despite having an excellent PE ratio, it is incredibly risky to own due to its extremely high risk, extremely high leverage, and poor Net Profit margin. Also very low in revenue growth.

Cluster4 has the best Net Profit Margin, the lowest PE ratio, and the fastest sales growth. It can be bought or kept on hand as a reserve.

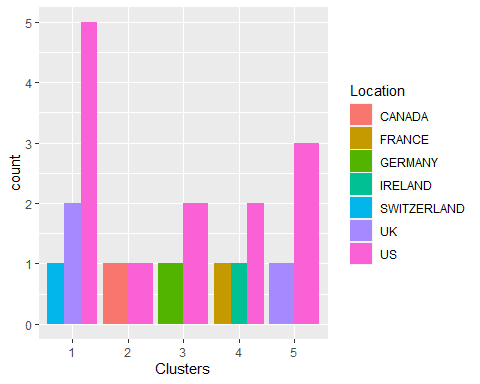
Cluster5 has strong market capitalization, ROI, ROA, ROA on assets, ROA on turnover of assets, and ROA on net profit margin. A low PE ratio indicates that the stock price is moderately valued and may thus be bought and kept. Revenue growth of 18.5% is also favorable.

Examining patterns by visualizing clusters against the variables

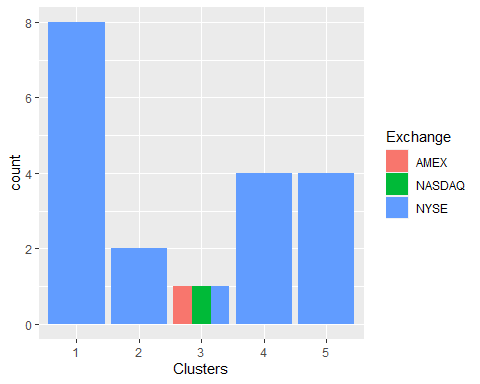
Pharma3 <- Pharma[12:14] %>% mutate(Clusters=clusterx$cluster)  
ggplot(Pharma3, mapping = aes(factor(Clusters), fill =Median\_Recommendation)) +  
 geom\_bar(position='dodge') + labs(x ='Clusters')



ggplot(Pharma3, mapping = aes(factor(Clusters),fill = Location)) +  
 geom\_bar(position = 'dodge') + labs(x ='Clusters')



ggplot(Pharma3, mapping = aes(factor(Clusters),fill = Exchange)) +  
 geom\_bar(position = 'dodge') + labs(x ='Clusters')



Provide an appropriate name for each cluster using any or all of the variables in the data set.

Cluster1: Attempt it Cluster

Cluster2: Significant Risk Cluster

Cluster3: Very Dangerous Cluster

Cluster4: Top Buying Cluster

Cluster5: A Perfect Asset