CLUSTERS

Bhargavi kundrapu

2023-11-13

## Loading the Libraries

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(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  
## ✔ ggplot2 3.4.3 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.2

## ── 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

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

library(FactoMineR)

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

library(ggcorrplot)

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

## Setting the working directory

getwd()

## [1] "C:/Users/user/OneDrive/Desktop/fml assignment"

## Reading the data set

pharma <- read.csv("C:/Users/user/Downloads/Pharmaceuticals.csv")  
pharmacy <- pharma[3:11]  
head(pharmacy)

## 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

## Summary

summary(pharmacy)

## 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

## Normalizing the data

pharmacy1 <- scale(pharmacy)  
row.names(pharmacy1) <- pharma[,1]  
dis <- get\_dist(pharmacy1)  
corr <- cor(pharmacy1)  
fviz\_nbclust(pharmacy1, kmeans, method = "silhouette")



*Based on the above plot k value is 5*

### Question\_1 : Use only the numerical variables (1 to 9) to cluster the 21 firms. Justify the various choices made in conducting the cluster analysis, such as weights for different variables, the specific clustering algorithm(s) used, the number of clusters formed, and so on.

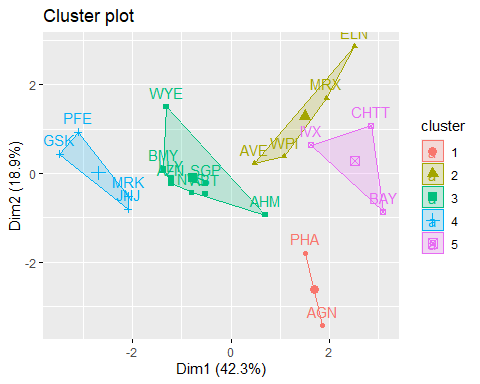
## Selecting the clusters

set.seed(150)  
k5 <- kmeans(pharmacy1, centers = 5, nstart = 20)  
k5$centers

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

## k-means

fviz\_cluster(k5, data = pharmacy1)



## Elbow

fviz\_nbclust(pharmacy1, kmeans, method = "wss")



## Manhattan

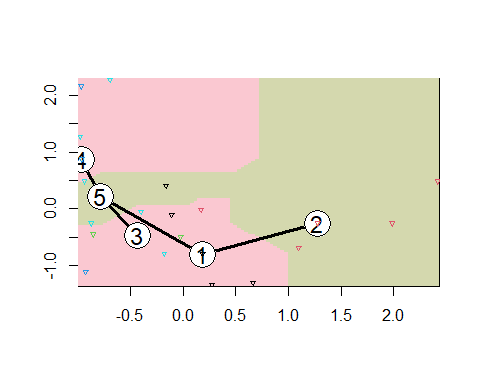
set.seed(50)  
k51 = kcca(pharmacy1, k = 5, kccaFamily("kmedians"))  
k51

## kcca object of family 'kmedians'   
##   
## call:  
## kcca(x = pharmacy1, k = 5, family = kccaFamily("kmedians"))  
##   
## cluster sizes:  
##   
## 1 2 3 4 5   
## 5 5 2 3 6

clusters\_index <- predict(k51)  
dist(k51@centers)

## 1 2 3 4  
## 2 2.558034   
## 3 4.451230 4.795056   
## 4 4.222539 4.954336 4.589219   
## 5 2.645989 3.581581 3.351236 2.857647

image(k51)  
points(pharmacy1, col=clusters\_index, pch= 25, cex=0.5)

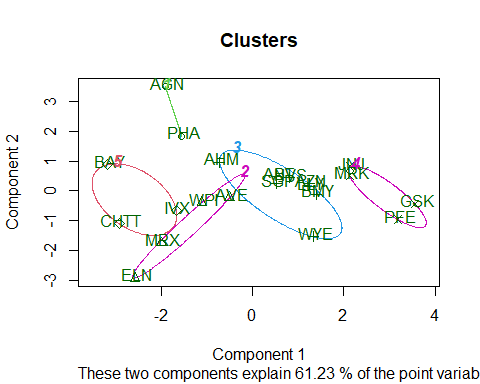


### Question\_2 : Interpret the clusters with respect to the numerical variables used in forming the clusters. Is there a pattern in the clusters with respect to the numerical variables (10 to 12)? (those not used in forming the clusters)

pharmacy%>% mutate(Cluster = k5$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 31.9 0.405 69.5 13.2 5.6 0.75 0.475  
## 2 2 13.1 0.598 17.7 14.6 6.2 0.425 0.635  
## 3 3 55.8 0.414 20.3 28.7 12.7 0.738 0.371  
## 4 4 157. 0.48 22.2 44.4 17.7 0.95 0.22   
## 5 5 6.64 0.87 24.6 16.5 4.17 0.6 1.65   
## # ℹ 2 more variables: Rev\_Growth <dbl>, Net\_Profit\_Margin <dbl>

clusplot(pharmacy1,k5$cluster, main="Clusters",color = TRUE, labels = 2,lines = 0)



## Companies are categorized into different clusters as follows:

**Cluster 1: ELN, MRX, WPI and AVE**

**Cluster 2: AGN and PHA**

**Cluster 3: AHM,WYE,BMY,AZN, LLY, ABT, NVS and SGP**

**Cluster 4: BAY, CHTT and IVX**

**Cluster 5: JNJ, MRK, PFE and GSK**

## Interpretation:

**Cluster 1** : Optimal Investment Opportunity Standing out with the finest Net Profit Margin, the lowest PE ratio, and rapid sales growth, Cluster 1 is a prime choice for investment or holding as a strategic reserve.

**Cluster 2** : Elevated PE Ratio Warning Marked by a notably high PE ratio, Cluster 2 raises a cautionary flag for potential overvaluation. Investors should approach this cluster with care, recognizing the elevated valuation.

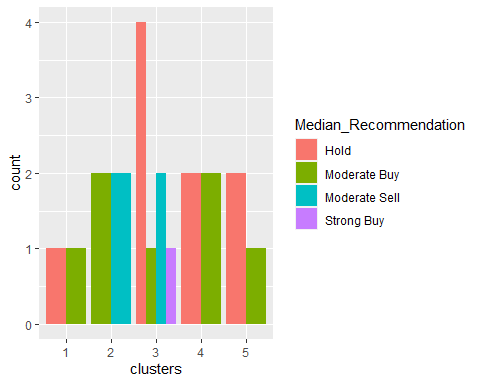
**Cluster 3**: Moderate Risk Class Cluster 3 Represents a moderate-risk category. While not as extreme as other clusters, entities in this group require careful consideration, balancing risk and potential return.

**Cluster 4** : High Risk, Despite Strong PE Ratio Despite boasting an excellent PE ratio, Cluster 4 carries exceptionally high risk due to elevated leverage, poor Net Profit Margin, and very low revenue growth. Ownership of entities in this cluster is considered highly precarious.

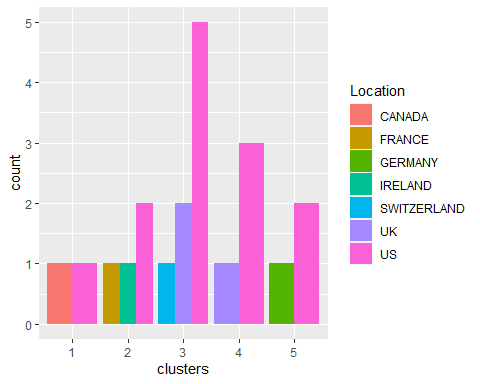
**Cluster 5** : Exceptional Metrics Showcase Cluster 5 demonstrates formidable market capitalization, ROI, ROA, asset turnover, and Net Profit Margin. Entities in this cluster, with a moderately valued PE ratio, are highly favorable for both purchase and retention. The substantial revenue growth of 18.5% further enhances the allure of this cluster.

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

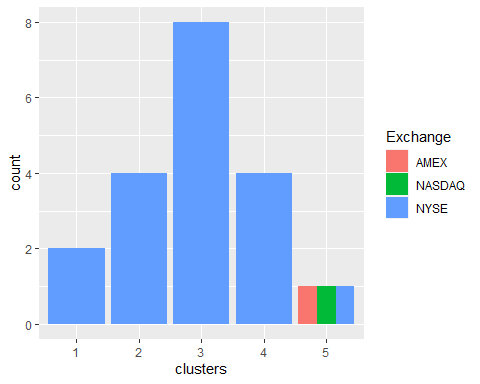
pharmacy2 <- pharma[12:14] %>% mutate(clusters=k5$cluster)  
ggplot(pharmacy2, mapping = aes(factor(clusters), fill = Median\_Recommendation)) + geom\_bar(position = 'dodge') + labs (x = 'clusters')



ggplot(pharmacy2, mapping = aes(factor(clusters), fill = Location)) + geom\_bar(position = 'dodge') + labs (x = 'clusters')



ggplot(pharmacy2, mapping = aes(factor(clusters), fill = Exchange)) + geom\_bar(position = 'dodge') + labs (x = 'clusters')



## Interpretation based of varibles 10 to 12 :

**cluster 1** : For Median Recommendation, Cluster 1 is characterized by a balanced mix of moderate buy and moderate sell recommendations. For Location, Within Cluster 1, there are three locations, with the United States being the predominant one. For Exchange, Exclusively listed on the NYSE, Cluster 1 operates on a single exchange.

**cluster 2** : The median recommendation for Cluster 2 suggests a combination of low hold and low buy. Cluster 2 is present in two locations, the United States and Canada, with an even distribution between them. Exclusive to the NYSE, Cluster 2 is listed on a single exchange.

**cluster 3** : The median recommendation for Cluster 3 leans strongly towards a robust hold, while for Cluster 4, it suggests a high level of moderate selling. In Cluster 3, there are three locations, with the United States being more prominent than the United Kingdom and Switzerland. Cluster 3 is exclusively listed on the NYSE, a widely participated exchange.

**cluster 4** : In Cluster 4 for Median Recommendation, there is a strong recommendation to hold, accompanied by a low recommendation to buy. Cluster 4 is present in two locations, with the United States having a higher ranking than Germany. Cluster 4 includes three exchanges (AMEX, NASDAQ, and NYSE), and these exchanges are evenly distributed within the cluster.

**cluster 5** : For Median Recommendation in cluster 5 has a high hold and a high buy, according to the median recommendation. has a high hold and a high buy, according to the median recommendation. Cluster 5 only has one exchange, which is NYSE.

There doesn't seem to be any discernable pattern among the clusters, locations, or exchanges other than the fact that the majority of the clusters/companies are listed on the NYSE and situated in the United States.\*

### Question\_3: Provide an appropriate name for each cluster using any or all of the variables in the dataset.

I have considered Net\_profit\_Margin, Leverage and Rev\_Growth to define the clusters.

## Interpretation

**Cluster 1** : Profitable Momentum Leaders This cluster excels with high Net Profit Margin, low Leverage, and significant Revenue Growth. Entities in this group showcase a favorable balance of profitability and growth momentum, making them standout investment opportunities.

**Cluster 2** : Overleveraged, Low Growth Caution Marked by high Leverage and limited Revenue Growth, Cluster 2 raises a cautionary flag for potential overleveraging. Investors should approach this cluster with care, recognizing the risks associated with high leverage and slow revenue growth.

**Cluster 3** : Balanced Performance Class Cluster 3 represents a balanced performance class with moderate Leverage, Revenue Growth, and Net Profit Margin. Entities in this group offer a stable mix of risk and return, making them suitable for investors seeking a balanced portfolio.

**Cluster 4** : High Risk, Limited Growth Despite a strong Net Profit Margin, Cluster 4 carries exceptionally high Leverage and low Revenue Growth, indicating a precarious financial position. Ownership of entities in this cluster is considered high risk due to the combination of high leverage and limited growth prospects.

**Cluster 5** : Exceptional Profitability and Growth Showcase Cluster 5 stands out with high Net Profit Margin, low Leverage, and substantial Revenue Growth. Entities in this cluster showcase exceptional profitability and growth metrics, making them an attractive choice for both purchase and retention.