FML-ASSIGN-4

Misba Faisal

2024-03-17

library(factoextra)

## Warning: package 'factoextra' was built under R version 4.3.3

## Loading required package: ggplot2

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

library(ggplot2)   
library(ISLR)   
library(NbClust)   
library(cluster)  
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.3.3

## Warning: package 'readr' was built under R version 4.3.3

## Warning: package 'forcats' was built under R version 4.3.3

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ lubridate 1.9.3 ✔ tibble 3.2.1  
## ✔ purrr 1.0.2 ✔ tidyr 1.3.1

## ── 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

## Uploding data from CSV file(pharmaceutical.csv)  
Pharma <- read.csv ("C:/Users/Syed's/Downloads/Pharmaceuticals.csv")

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

##Creating a new data frame ‘R’ by excluding data which is empty in ‘Pharmaceuticals’

R <- na.omit(Pharma)   
summary(R)

## Symbol Name Market\_Cap Beta   
## Length:21 Length:21 Min. : 0.41 Min. :0.1800   
## Class :character Class :character 1st Qu.: 6.30 1st Qu.:0.3500   
## Mode :character Mode :character Median : 48.19 Median :0.4600   
## Mean : 57.65 Mean :0.5257   
## 3rd Qu.: 73.84 3rd Qu.:0.6500   
## Max. :199.47 Max. :1.1100   
## PE\_Ratio ROE ROA Asset\_Turnover Leverage   
## Min. : 3.60 Min. : 3.9 Min. : 1.40 Min. :0.3 Min. :0.0000   
## 1st Qu.:18.90 1st Qu.:14.9 1st Qu.: 5.70 1st Qu.:0.6 1st Qu.:0.1600   
## Median :21.50 Median :22.6 Median :11.20 Median :0.6 Median :0.3400   
## Mean :25.46 Mean :25.8 Mean :10.51 Mean :0.7 Mean :0.5857   
## 3rd Qu.:27.90 3rd Qu.:31.0 3rd Qu.:15.00 3rd Qu.:0.9 3rd Qu.:0.6000   
## Max. :82.50 Max. :62.9 Max. :20.30 Max. :1.1 Max. :3.5100   
## Rev\_Growth Net\_Profit\_Margin Median\_Recommendation Location   
## Min. :-3.17 Min. : 2.6 Length:21 Length:21   
## 1st Qu.: 6.38 1st Qu.:11.2 Class :character Class :character   
## Median : 9.37 Median :16.1 Mode :character Mode :character   
## Mean :13.37 Mean :15.7   
## 3rd Qu.:21.87 3rd Qu.:21.1   
## Max. :34.21 Max. :25.5   
## Exchange   
## Length:21   
## Class :character   
## Mode :character   
##   
##   
##

# Adding new row name for the first column in the new dataset,creating a new dataset from coloumns 3 to 11 from ’A’and displaying the rows of the pharmacy data frame.

row.names(R) <- R[,1]  
Pharmacy <- R[,3:11]  
head(Pharmacy)

## 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

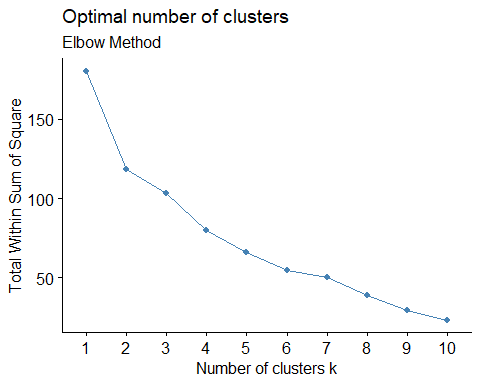
# scaling data in the standard value form and displaying in standard form.

Pharmacist <- scale (Pharmacy)  
head(Pharmacist)

## 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

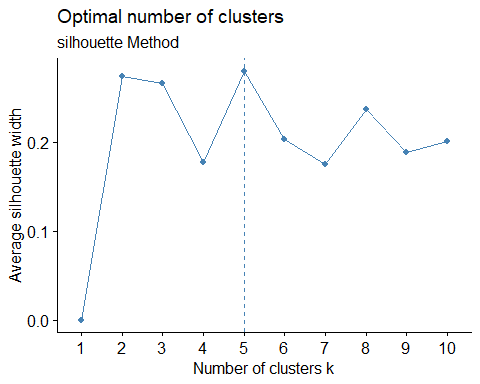
# display(1)

fviz\_nbclust(Pharmacist, kmeans, method = "wss") + labs(subtitle = "Elbow Method")



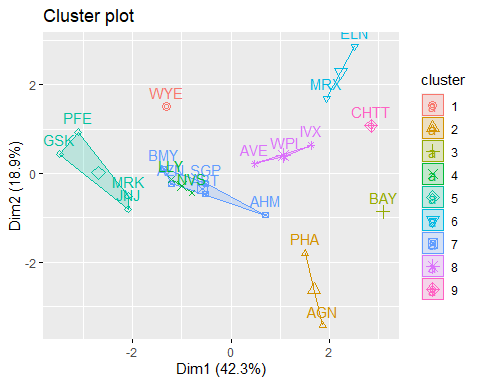
# display(2)

fviz\_nbclust(Pharmacist, kmeans, method = "silhouette") + labs(subtitle = "silhouette Method")



# Clustering algorithm by using k-means with presenting 9 as the centre.

set.seed(2731)  
k9 <- kmeans(Pharmacist, centers = 9, nstart = 81)  
fviz\_cluster(k9, data = Pharmacist)

 # calculating euclidean distance between matrix and pharmacist dataset.

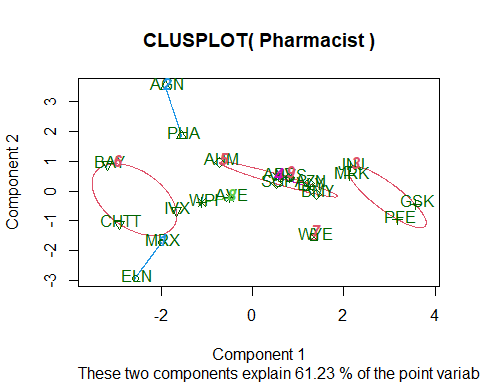
distance <- dist(Pharmacist, method = "euclidean")   
fviz\_dist(distance)

 # CRAN mirror

options(repos = c(CRAN = "https://cran.rstudio.com/"))  
fit <-kmeans(Pharmacist, 9)   
aggregate(Pharmacist, by = list(fit$cluster), FUN=mean)

## Group.1 Market\_Cap Beta PE\_Ratio ROE ROA  
## 1 1 -0.9668697 1.5162611 -0.57398880 -0.83826708 -0.9892673  
## 2 2 -0.4392513 -0.4701800 2.70002464 -0.83495252 -0.9234951  
## 3 3 1.6955811 -0.1780563 -0.19845823 1.23498791 1.3503431  
## 4 4 0.2301688 -1.0739023 0.05138455 0.19255994 0.4013445  
## 5 5 -0.3039339 -0.1098942 -0.41151435 -0.05768906 0.4483247  
## 6 6 -0.8705151 1.3409869 -0.05284434 -0.61840151 -1.1928478  
## 7 7 -0.1614497 0.4061910 -0.75792214 1.92938746 0.5422849  
## 8 8 -0.5535800 -0.9570529 -0.38085880 -0.64933737 -0.6322183  
## 9 9 0.6654710 -1.3076013 -0.23677768 -0.52338423 0.1288598  
## Asset\_Turnover Leverage Rev\_Growth Net\_Profit\_Margin  
## 1 -1.8450624 0.5302448 1.7123890 0.2445520  
## 2 0.2306328 -0.1417034 -0.1168459 -1.4165148  
## 3 1.1531640 -0.4680782 0.4671788 0.5912425  
## 4 -0.2306328 -0.1417034 -0.5879576 0.6179110  
## 5 0.8072148 -0.4808772 -0.4564898 0.1759758  
## 6 -0.4612656 1.3664470 -0.6912914 -1.3200002  
## 7 -0.4612656 0.6838330 -1.1776392 1.4941618  
## 8 -0.6918984 -0.4040831 1.3236427 -0.2583398  
## 9 -0.9225312 -0.6728624 -1.4536989 1.0217484

Pharmacist1 <- data.frame(Pharmacist, fit$cluster)   
clusplot(Pharmacist, fit$cluster, color = TRUE, shade = FALSE, labels = 2, lines = 0)

 ## (B). Interpret the clusters with respect to the numerical variables used in forming the clusters. \*\*The corporations that comprise Cluster 1 are ##JNJ, MRK, PFE, and GSK; these companies have the largest market capitalizations and depend on funding to operate efficiently. (less than 0.47 in leverage).

\*\*Due to their lower asset turnover and beta values, the stocks of Cluster 2 companies—AHM, WPI, and AVE—have the potential to perform better than the present market benchmark.

\*\*They are the least capitalized company on the market, with the fastest sales growth in Cluster 3, and can’t even raise financing to keep running their business. (MRX, CHTT, LVX, ELAN). These company stocks’ remarkable performance can be attributed in part to their high beta ratings.

\*\*Cluster 4: AGN, BAY, and RHA Because of their largest expense to earnings ratio, they have the lowest earnings. Additionally, the fact that these companies’ Return on Equity is less than 1 makes it unlikely that investing in them will yield the biggest profits.

\*\*Cluster-5 The group is composed of ABT, SGP, NVS, AZN, BMY, and WYE. They have the lowest rate of sales development, the highest asset turnover, and the highest net profit margin. The growth of these businesses has contributed to their success.

## (C). Is there a pattern in the clusters with respect to the numerical variables (10 to 12)? (those not used informing the clusters)

\*\*The equities in Cluster-1 are neither very strong, nor have they recently produced significant returns. Rather, their personality is average.

\*\*The companies in Cluster-2 are equally spread out around the world. Despite having a solid technological basis, the media has widely embraced these ideas.

\*\*Cluster 3: Despite having a high leverage ratio, they are only mildly advised due to the steadiness of their finances.

\*\*Cluster 4: The media claims that they should be preserved in case they turn out to be very valuable assets in the future.

\*\*Cluster No:5 are encouraged to stay for an extended period of time if they have a high net profit margin.

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

\*\*Cluster 1: Realistic strategy (as they are reputable stocks). A team of miners for gold.

\*\*Despite Cluster-2’s low beta, the market is very optimistic about them.

\*\*The first configuration, or Cluster-3, was made up of stocks with solid fundamentals, both financially and otherwise.

\*\*Cluster 4: The starting lineup (stocks with solid fundamentals and financials).

\*\*Cluster 5 is the recurring cluster. Including the equities in the portfolio is highly recommended because a substantial net profit margin indicates that the business is making a profit.