FML ASSIGNMENT-4

Rachana

2022-11-06

library("tidyverse")

## Warning: package 'tidyverse' was built under R version 4.2.2

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4   
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.4.1   
## ✔ readr 2.1.2 ✔ forcats 0.5.2

## Warning: package 'forcats' was built under R version 4.2.2

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library("factoextra")

## Warning: package 'factoextra' was built under R version 4.2.2

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

library("ggplot2")  
library("dplyr")

#read data

data <- read.csv("C:/Users/kurra/Downloads/Pharmaceuticals.csv")

#Normalization

data.1 <- scale(data[,-c(1:2,12:14)])

#A.using only numerical variables (1-9) to cluster the 21 firms.

#finding the optimal k

wss <- fviz\_nbclust(data.1,kmeans,method="wss")  
wss



silhouette <- fviz\_nbclust(data.1,kmeans,method="silhouette")  
silhouette

 #The optimal k obtained through wss method is k = 2 , and the optimal value obtained through silhouette method is k = 5.

#Formulation of clusters using K-Means with k = 2 (WSS)

wss\_kmeans <- kmeans(data.1,centers = 2,nstart=10)  
wss\_kmeans

## K-means clustering with 2 clusters of sizes 11, 10  
##   
## Cluster means:  
## Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover  
## 1 0.6733825 -0.3586419 -0.2763512 0.6565978 0.8344159 0.4612656  
## 2 -0.7407208 0.3945061 0.3039863 -0.7222576 -0.9178575 -0.5073922  
## Leverage Rev\_Growth Net\_Profit\_Margin  
## 1 -0.3331068 -0.2902163 0.6823310  
## 2 0.3664175 0.3192379 -0.7505641  
##   
## Clustering vector:  
## [1] 1 2 2 1 2 2 1 2 2 1 1 2 1 2 1 1 1 2 1 2 1  
##   
## Within cluster sum of squares by cluster:  
## [1] 43.30886 75.26049  
## (between\_SS / total\_SS = 34.1 %)  
##   
## Available components:  
##   
## [1] "cluster" "centers" "totss" "withinss" "tot.withinss"  
## [6] "betweenss" "size" "iter" "ifault"

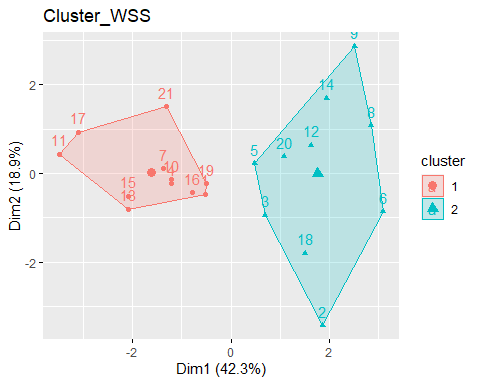
*Formulation of clusters using K-Means with k = 5 (Silhouette)*

silhouette\_kmeans <- kmeans(data.1,centers=5,nstart=10)  
silhouette\_kmeans

## K-means clustering with 5 clusters of sizes 8, 4, 3, 4, 2  
##   
## Cluster means:  
## Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover  
## 1 -0.03142211 -0.4360989 -0.31724852 0.1950459 0.4083915 0.1729746  
## 2 1.69558112 -0.1780563 -0.19845823 1.2349879 1.3503431 1.1531640  
## 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 -0.43925134 -0.4701800 2.70002464 -0.8349525 -0.9234951 0.2306328  
## Leverage Rev\_Growth Net\_Profit\_Margin  
## 1 -0.27449312 -0.7041516 0.556954446  
## 2 -0.46807818 0.4671788 0.591242521  
## 3 1.36644699 -0.6912914 -1.320000179  
## 4 0.06308085 1.5180158 -0.006893899  
## 5 -0.14170336 -0.1168459 -1.416514761  
##   
## Clustering vector:  
## [1] 1 5 1 1 4 3 1 3 4 1 2 3 2 4 2 1 2 5 1 4 1  
##   
## Within cluster sum of squares by cluster:  
## [1] 21.879320 9.284424 15.595925 12.791257 2.803505  
## (between\_SS / total\_SS = 65.4 %)  
##   
## Available components:  
##   
## [1] "cluster" "centers" "totss" "withinss" "tot.withinss"  
## [6] "betweenss" "size" "iter" "ifault"

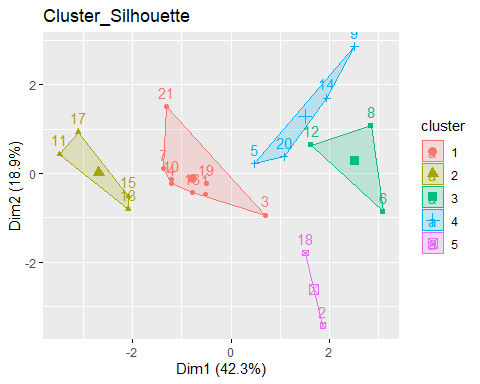
#plotting cluster for WSS

fviz\_cluster(wss\_kmeans,data.1,main="Cluster\_WSS")

 #we got 2 clusters using WSS method of size 11,10

#plotting cluster for Silhouette

fviz\_cluster(silhouette\_kmeans,data.1,main="Cluster\_Silhouette")

 #we got 5 clusters of size 4, 8, 5, 3 and 1`using silhouette

#grouping the clusters to the original data frame for analysis

clusters\_wss <- wss\_kmeans$cluster  
clusters\_silhouette <- silhouette\_kmeans$cluster  
group1 <- cbind(data,clusters\_wss)  
group2 <- cbind(data,clusters\_silhouette)

#Aggregating the clusters to interpret wss

interpret\_wss<- aggregate(data.1,by=list(group1$clusters\_wss),FUN="median")  
print(interpret\_wss[,-1])

## Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover  
## 1 0.2762415 -0.2559560 -0.2429088 0.3450295 0.8429577 0.4612656  
## 2 -0.9021972 0.1140673 -0.1294832 -0.7686614 -0.9234951 -0.4612656  
## Leverage Rev\_Growth Net\_Profit\_Margin  
## 1 -0.3912841 -0.4354459 0.7474438  
## 2 -0.1417034 0.1017391 -0.7002750

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

#from first cluster we can tell that there is high cahnce of success rate by reffering to attributes : “Market Capital”, ROE - Return on Expenditure, ROA - Return on Assets, Asset Turnover and Net Profit Margin.

#from the second cluster, we can interpret that it is having poor performance metrics compared to the first cluster. Return on Expenditure (ROE), Return on Assets (ROA), Asset Turnover, Net Profit Margin are low.the risk level in here is high which can be concluded by beta and leverage values, as they are high in this frim.

#Aggregating the clusters to interpret the attribute Silhouette

interpret\_s <- aggregate(group2[,-c(1:2,12:14)],by=list(group2$clusters\_silhouette),FUN="median")  
print(interpret\_s[,-1])

## Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover Leverage Rev\_Growth  
## 1 59.480 0.480 21.10 26.90 13.35 0.75 0.345 6.630  
## 2 153.245 0.460 21.25 43.10 17.75 0.95 0.220 19.610  
## 3 2.600 0.850 26.00 21.40 4.30 0.60 1.450 6.380  
## 4 2.230 0.535 19.25 13.15 6.10 0.40 0.635 29.775  
## 5 31.910 0.405 69.50 13.20 5.60 0.75 0.475 12.080  
## Net\_Profit\_Margin clusters\_silhouette  
## 1 19.3 1  
## 2 19.5 2  
## 3 7.5 3  
## 4 14.2 4  
## 5 6.4 5

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

#The first cluster indicates that there is high risk in the firms since there is high beta and leverage and also market cap and net profit margin are low

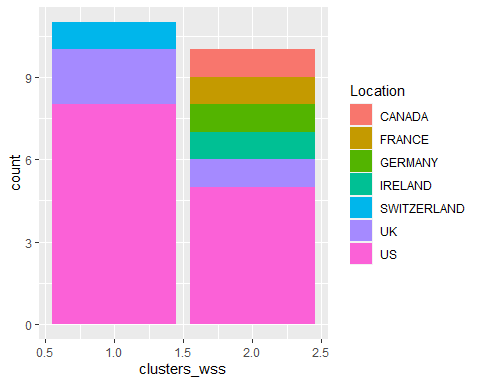
#The Second Cluster also indicate the same as the first one.

#Third Cluster indicates that it can fit properly for firm industry with less risks as pE ratio and market capital is good # fourth indicates less risk and high valued

# fifth cluster indicates low returning vales.

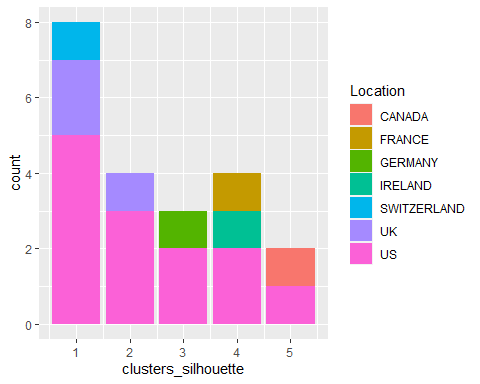
#c.Pattern in the clustrers with respect to numerical values (10-12)

ggplot(group1,aes(x=clusters\_wss,fill=Location)) + geom\_bar()

 #From the bars we can observe that for 2 clusters based on location we can interpret that pharmaceutical exchange are mostly US based. i.e. Cluster 1 have greater exchange ratio for US based companies.

##Pattern in the clustrers with respect to numerical values (10-12)

ggplot(group2,aes(x=clusters\_silhouette,fill=Location)) + geom\_bar()

 #here in sihouette method we can say that location based is similar to wss cluster. more clusters are US based.

#majority of pharma locations are cluster1 and cluster2 and for pharma exchange is NYSE.

#d.Naming for the clusters:

#WSS: #hold cluster- moderate risk #buy cluster- high risk

#sihouette

#high risk #low market cap #moderate fit #high valued #low returns high investment