FML Assignment 4

Anjali

2023-10-30

```
getwd()
```

```
## [1] "C:/Users/Anjali/Desktop"
```

```
library(readr)
Pharma_data <- read_csv("C:/Users/Anjali/Desktop/Anjali FML Assignment 4/Pharmaceuticals.csv")
```

```
## Rows: 21 Columns: 14
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): Symbol, Name, Median_Recommendation, Location, Exchange
## dbl (9): Market_Cap, Beta, PE_Ratio, ROE, ROA, Asset_Turnover, Leverage, Rev...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
View(Pharma_data)
head(Pharma_data)
```

```
## # A tibble: 6 x 14
##   Symbol Name      Market_Cap  Beta PE_Ratio   ROE   ROA Asset_Turnover Leverage
##   <chr>  <chr>          <dbl> <dbl>    <dbl> <dbl> <dbl>          <dbl>    <dbl>
## 1 ABT    Abbott L~      68.4   0.32     24.7  26.4  11.8            0.7     0.42
## 2 AGN    Allergan~       7.58  0.41     82.5  12.9   5.5            0.9     0.6
## 3 AHM    Amersham~       6.3   0.46     20.7  14.9   7.8            0.9     0.27
## 4 AZN    AstraZen~      67.6   0.52     21.5  27.4  15.4            0.9     0
## 5 AVE    Aventis        47.2   0.32     20.1  21.8   7.5            0.6     0.34
## 6 BAY    Bayer AG       16.9   1.11     27.9   3.9   1.4            0.6     0
## # i 5 more variables: Rev_Growth <dbl>, Net_Profit_Margin <dbl>,
## #   Median_Recommendation <chr>, Location <chr>, Exchange <chr>
```

```
# a. Use only the numerical variables (1 to 9) to cluster the 21 firms. Justify the various choices mad

Numeric_vals <- Pharma_data [,3:11]
head(Numeric_vals)
```

```
## # A tibble: 6 x 9
##   Market_Cap  Beta PE_Ratio   ROE   ROA Asset_Turnover Leverage Rev_Growth
##        <dbl> <dbl>    <dbl> <dbl> <dbl>          <dbl>    <dbl>      <dbl>
```

```
## 1      68.4   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     0.6        9.16
## 3       6.3   0.46     20.7  14.9   7.8           0.9     0.27       7.05
## 4      67.6   0.52     21.5  27.4  15.4           0.9     0
## 5      47.2   0.32     20.1  21.8   7.5           0.6     0.34      26.8
## 6      16.9   1.11     27.9   3.9   1.4           0.6     0         -3.17
## # i 1 more variable: Net_Profit_Margin <dbl>
```

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

#We have scaled all numerical variables in the data frame

scaling_Numeric_vals <- scale (Numeric_vals)
head(scaling_Numeric_vals)
```

```
##      Market_Cap        Beta    PE_Ratio         ROE        ROA Asset_Turnover
## [1,]  0.1840960 -0.80125356 -0.04671323  0.04009035  0.2416121      0.0000000
## [2,] -0.8544181 -0.45070513  3.49706911 -0.85483986 -0.9422871      0.9225312
## [3,] -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700      0.9225312
## [4,]  0.1702742 -0.02225704 -0.24290879  0.10638147  0.9181259      0.9225312
## [5,] -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.5664461     -0.4612656
## [6,] -0.6953818  2.27578267  0.14948233 -1.45146000 -1.7127612     -0.4612656
##        Leverage Rev_Growth Net_Profit_Margin
## [1,] -0.2120979 -0.5277675        0.06168225
## [2,]  0.0182843 -0.3811391       -1.55366706
## [3,] -0.4040831 -0.5721181       -0.68503583
## [4,] -0.7496565  0.1474473        0.35122600
## [5,] -0.3144900  1.2163867       -0.42597037
## [6,] -0.7496565 -1.4971443       -1.99560225
```

```
#To determine the number of clusters we did the cluster analysis
Clusters <- (nrow(scaling_Numeric_vals)-1)*sum(apply(scaling_Numeric_vals,2,var))
Clusters
```

```
## [1] 180
```

```
for (i in 2:15) Clusters[i] <- sum(kmeans(scaling_Numeric_vals,centers=i)$withinss)
Clusters
```

```
##  [1] 180.000000 118.569343  99.455039  78.246004  70.043249  54.843448
##  [7]  43.019268  43.953025  37.663556  23.310806  26.005335  15.663185
## [13]  12.668848  11.289987   8.340019
```

```
#K-Means Cluster Analysis - Fit the data with 5 clusters

fit <- kmeans(scaling_Numeric_vals, 5)
aggregate(scaling_Numeric_vals,by=list(fit$cluster),FUN=mean)
```

```
##   Group.1 Market_Cap        Beta    PE_Ratio         ROE         ROA
```

```
## 1       1  0.9547543 -0.06120687 -0.35764816  1.0818081  1.10336187
## 2       2 -0.8705151  1.34098686 -0.05284434 -0.6184015 -1.19284783
## 3       3 -0.9668697  1.51626107 -0.57398880 -0.8382671 -0.98926727
## 4       4 -0.4392513 -0.47018004  2.70002464 -0.8349525 -0.92349509
## 5       5 -0.1799275 -0.81238208 -0.22714308 -0.3387161 -0.04563784
##   Asset_Turnover   Leverage  Rev_Growth Net_Profit_Margin
## 1      0.8566361 -0.2797499 -0.01818848         0.7082574
## 2     -0.4612656  1.3664470 -0.69129140        -1.3200002
## 3     -1.8450624  0.5302448  1.71238901         0.2445520
## 4      0.2306328 -0.1417034 -0.11684587        -1.4165148
## 5     -0.1976853 -0.4168821 -0.14141325         0.1923035
```

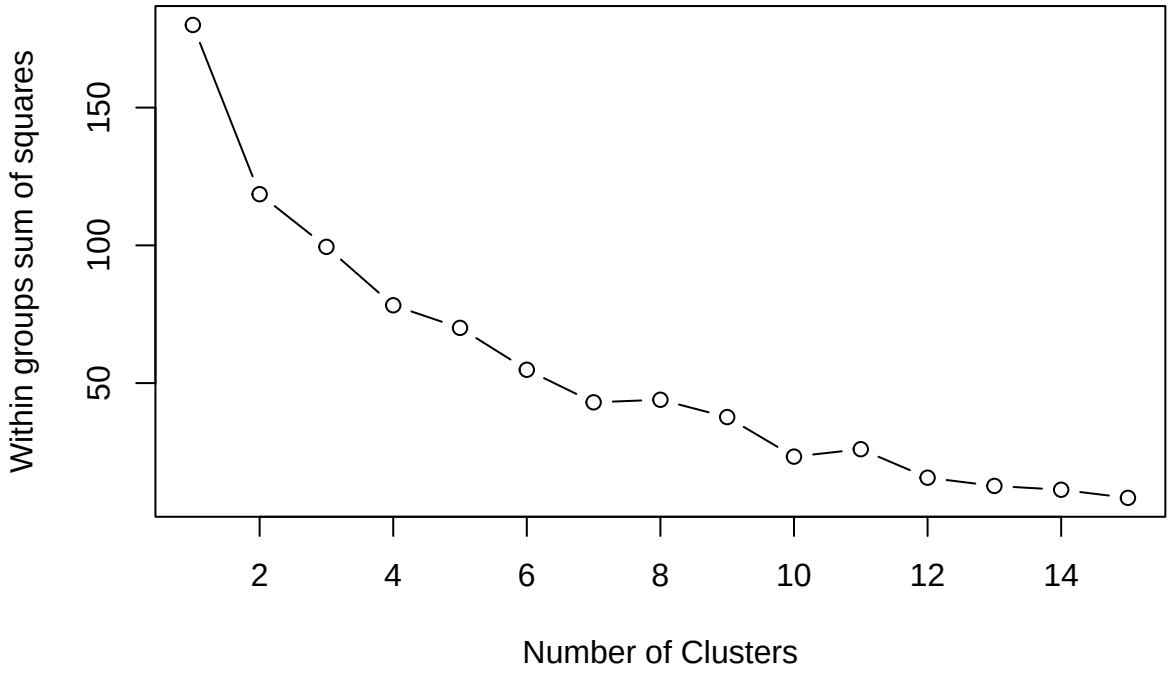
```
Cluster_Number<- data.frame(scaling_Numeric_vals, fit$cluster)

Cluster_Number
```

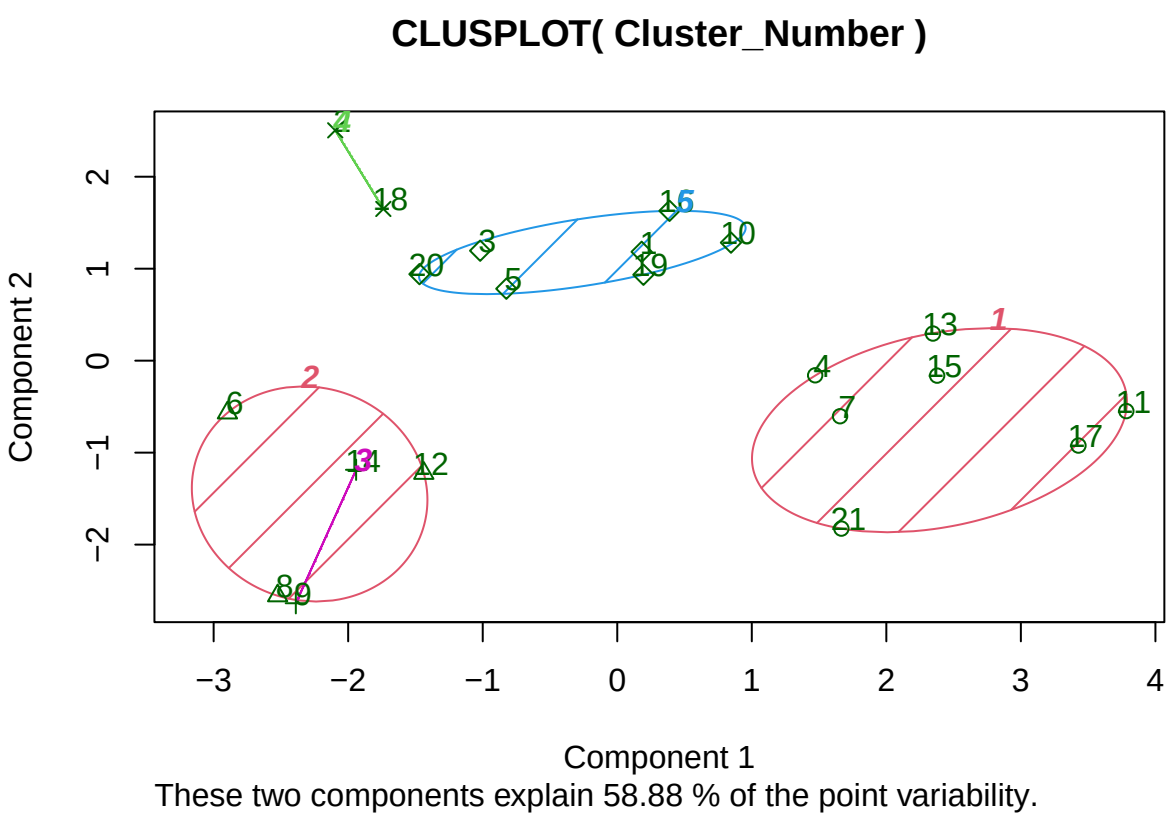
```
##    Market_Cap        Beta    PE_Ratio         ROE        ROA Asset_Turnover
## 1   0.1840960 -0.80125356 -0.04671323  0.04009035  0.2416121      0.0000000
## 2  -0.8544181 -0.45070513  3.49706911 -0.85483986 -0.9422871      0.9225312
## 3  -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700      0.9225312
## 4   0.1702742 -0.02225704 -0.24290879  0.10638147  0.9181259      0.9225312
## 5  -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.5664461     -0.4612656
## 6  -0.6953818  2.27578267  0.14948233 -1.45146000 -1.7127612     -0.4612656
## 7  -0.1078688 -0.10015669 -0.70887325  0.59693581  0.8617498      0.9225312
## 8  -0.9767669  1.26308721  0.03299122 -0.11237924 -1.1677918     -0.4612656
## 9  -0.9704532  2.15893320 -1.34037772 -0.70899938 -1.0174553     -1.8450624
## 10  0.2762415 -1.34655112  0.14948233  0.34502953  0.5610770     -0.4612656
## 11  1.0999201 -0.68440408 -0.45749769  2.45971647  1.8389364      1.3837968
## 12 -0.9393967  0.48409069 -0.34100657 -0.29136529 -0.6979905     -0.4612656
## 13  1.9841758 -0.25595600  0.18013789  0.18593083  1.0872544      0.9225312
## 14 -0.9632863  0.87358895  0.19240011 -0.96753478 -0.9610792     -1.8450624
## 15  1.2782387 -0.25595600 -0.40231769  0.98142435  0.8429577      1.8450624
## 16  0.6654710 -1.30760129 -0.23677768 -0.52338423  0.1288598     -0.9225312
## 17  2.4199899  0.48409069 -0.11415545  1.31287998  1.6322239      0.4612656
## 18 -0.0240846 -0.48965495  1.90298017 -0.81506519 -0.9047030     -0.4612656
## 19 -0.4018812 -0.06120687 -0.40231769 -0.21181593  0.5234929      0.4612656
## 20 -0.9281345 -1.11285216 -0.43297324 -1.03382590 -0.6979905     -0.9225312
## 21 -0.1614497  0.40619104 -0.75792214  1.92938746  0.5422849     -0.4612656
##       Leverage  Rev_Growth Net_Profit_Margin fit.cluster
## 1  -0.21209793 -0.52776752        0.06168225           5
## 2   0.01828430 -0.38113909       -1.55366706           4
## 3  -0.40408312 -0.57211809       -0.68503583           5
## 4  -0.74965647  0.14744734        0.35122600           1
## 5  -0.31449003  1.21638667       -0.42597037           5
## 6  -0.74965647 -1.49714434       -1.99560225           2
## 7  -0.02011273 -0.96584257        0.74744375           1
## 8   3.74279705 -0.63276071       -1.24888417           2
## 9   0.61983791  1.88617085       -0.36501379           3
## 10 -0.07130879 -0.64814764        1.17413980           5
## 11 -0.31449003  0.76926048        0.82363947           1
## 12  1.10620040  0.05603085       -0.71551412           2
## 13 -0.62166634 -0.36213170        0.33598685           1
## 14  0.44065173  1.53860717        0.85411776           3
## 15 -0.39128411  0.36014907       -0.24310064           1
```

```
## 16 -0.67286239 -1.45369888        1.02174835           5
## 17 -0.54487226  1.10143723        1.44844440           1
## 18 -0.30169102  0.14744734       -1.27936246           4
## 19 -0.74965647 -0.43544591        0.29026942           5
## 20 -0.49367621  1.43089863       -0.09070919           5
## 21  0.68383297 -1.17763919        1.49416183           1
```

```
plot(1:15, Clusters, type="b", xlab="Number of Clusters",ylab="Within groups sum of squares")
```




```
library(cluster)
clusplot(Cluster_Number, fit$cluster, color=TRUE, shade=TRUE, labels=2, lines=0)
```



#b. Interpret the clusters with respect to the numerical variables used in forming the clusters. Cluster 1 – 17,11,15,13

Cluster 2 – 2,18,20,5,3

Cluster 3- 10,16,19,7,21,1,4

Cluster 4 -14,9

Cluster 5 – 12,8,6

Cluster 1-We can interpret the cluster with help of Aggregate function aggregate(scaling_Numeric_vals, by=list(fit$cluster),FUN=mean)

Cluster 1- 17,11,15,13 It has the highest Market capitalization, ROE, ROA, Asset Turnover and lowest leverage When we compare with Csv file 2 are on hold and 2 are on buy median recommendation.

Cluster 2- 2,18,20,5,3 It has the Lowest negative beta, Highest P/E Ratio with strong asset turnover and second-lowest Profit Margin When we compare with CSV file we have 2 Moderate Buy, One Hold, One Moderate sell, , One Strong buy as per median of recommendation.

Cluster 3 - 10,16,19,7,21,1,4 In this Cluster, we have the highest Asset turnover, Net profit margin and second-lowest Leverage. When we compare with the CSV file, we have 5 Holds one moderate buy and one sell as per the median of recommendation.

Cluster 4- 14,9 It has the lowest Market capitalization, Price earning (P/E) Ratio, Return on equity (ROE), Second lowest ROE and Leverage, highest revenue Growth. When we compare with CSV file one moderate buy and one moderate sell as per median of recommendation.

Cluster 5 – 12,8,6 In this cluster, we have the lowest Asset Turnover, Profit margin, Second lowest market cap, Highest Net profit margin and second-highest beta. When we compare with CSV file we have 2 Holds and one buy as per the median of recommendation.

#c. pattern in the clusters with respect to the numerical variables Cluster 1- It shows the buy recommendations Pattern. As it is made up of the highest Market capitalization, ROE, ROA, Asset Turnover, and lowest leverage so undoubtedly investors can go for Buy or Hold recommendation. which is also similar to the median of recommendation.

Cluster 2- Highest Price earning ratio, strong asset turnover ratio with lowest or negative beta represents hold or buy, which is also like the median of recommendation.

Cluster 3- strong Assets turnover ratio represents the company is efficiently able to use their assets to generate the revenue having highest net profit margin and lowest leverage make the investors to go for hold recommendation which is mostly like Median of recommendation.

Cluster 4 - Cluster 4 has no pattern.

Cluster 5 – It has the highest beta with that it also has the lowest ROE, ROA, and Asset Turnover. Higher beta stocks are risky in nature but provide a high return so we can name those stocks volatile stocks for risk-averse person recommendation is going to sell it did not represent any pattern.

#d Provide an appropriate name for each cluster using any or all of the variables in the dataset Cluster 1 – Higher Market cap, ROE, ROA, and asset turnover Cluster.

Cluster2- High Price-earnings ratio, asset turnover ratio and negative beta cluster.

Cluster 3- High net profit margin and asset turnover Cluster.

Cluster 4- low market capitalization, price-earnings ratio and return on equity cluster.

Cluster 5- Highest Beta and lower ROE , ROA cluster.