Assignment_4

Akhila Sri

2022-03-20

```
library(readr)

Pharmaceuticals <-
read_csv("C:/Users/akhila/OneDrive/Desktop/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(Pharmaceuticals)
head(Pharmaceuticals)
```

```
## # 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
## # ... with 5 more variables: Rev_Growth <dbl>, Net_Profit_Margin <dbl>,
## #   Median_Recommendation <chr>, Location <chr>, Exchange <chr>
```

```
df <- Pharmaceuticals [,3:11]
head(df)
```

```
## # 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         15
## 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
## # ... with 1 more variable: Net_Profit_Margin <dbl>

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

## Warning: package 'gridExtra' was built under R version 4.1.3

df_Scaling <- scale (df)
head(df_Scaling)

##      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

clusters <- (nrow(df_Scaling)-1)*sum(apply(df_Scaling,2,var))
clusters

## [1] 180

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

```
##  [1] 180.000000 118.569343  97.318844  78.246004  62.354431  60.188620
##  [7]  43.096609  49.532873  29.185914  28.576262  18.466690  15.472353
## [13]  15.965202  12.136719   7.130262
```

Cluster Analysis

```
Cluster_5 <- kmeans(df_Scaling, 5)
aggregate(df_Scaling,by=list(Cluster_5$cluster),FUN=mean)
```

```
##   Group.1  Market_Cap       Beta    PE_Ratio        ROE        ROA
## 1       1  1.69558112 -0.1780563 -0.19845823  1.2349879  1.3503431
## 2       2 -0.03142211 -0.4360989 -0.31724852  0.1950459  0.4083915
## 3       3 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478
## 4       4 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428
## 5       5 -0.43925134 -0.4701800  2.70002464 -0.8349525 -0.9234951
##   Asset_Turnover    Leverage Rev_Growth Net_Profit_Margin
## 1      1.1531640 -0.46807818  0.4671788       0.591242521
## 2      0.1729746 -0.27449312 -0.7041516       0.556954446
## 3     -0.4612656  1.36644699 -0.6912914      -1.320000179
## 4     -1.2684804  0.06308085  1.5180158      -0.006893899
## 5      0.2306328 -0.14170336 -0.1168459      -1.416514761
```

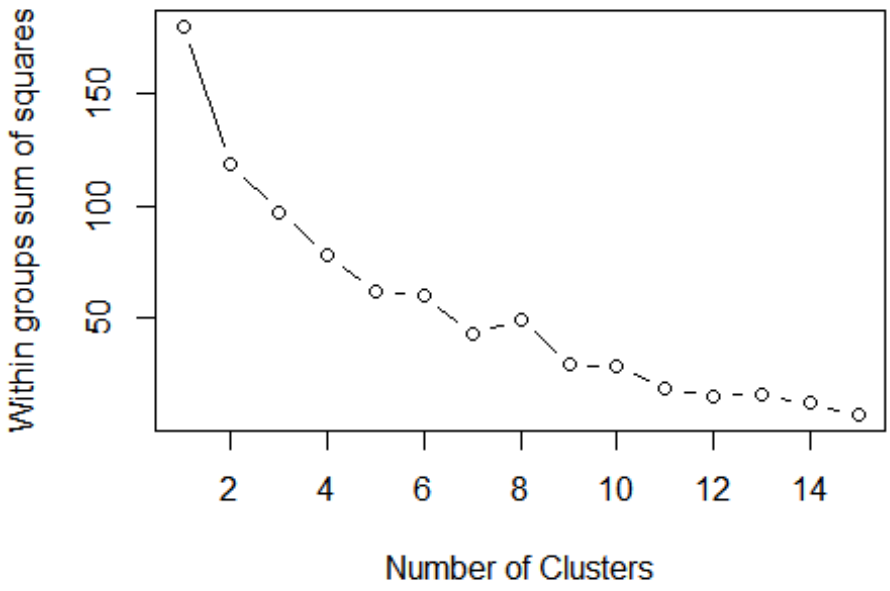
```
df_2 <- data.frame(df_Scaling, Cluster_5$cluster)
df_2
```

```
##    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
```

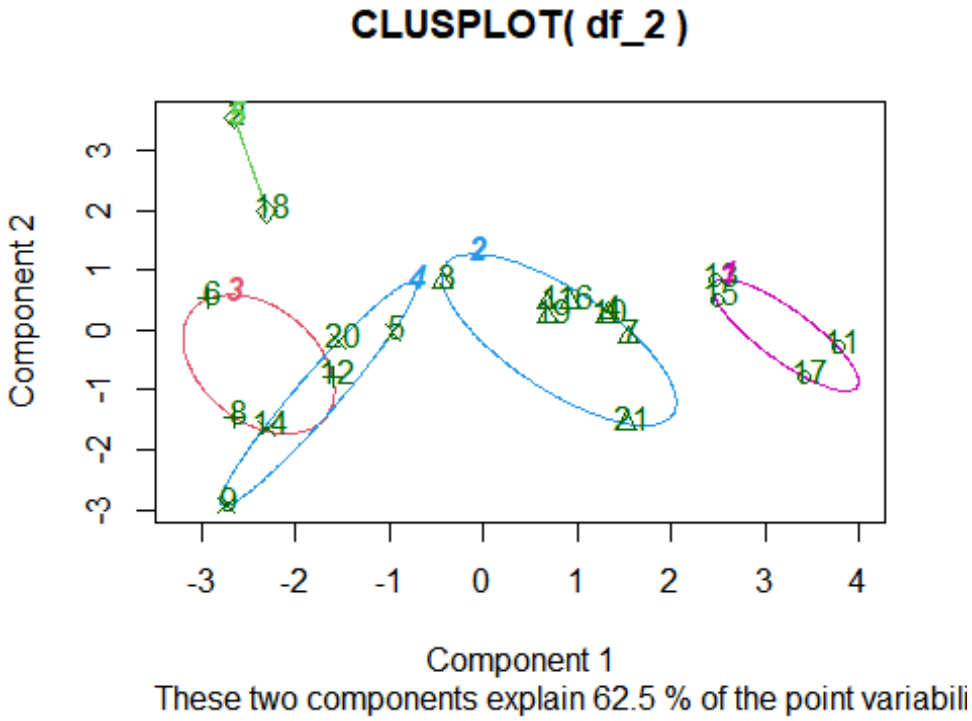
```
## 13  1.9841758 -0.25595600  0.18013789  0.18593083  1.0872544
0.9225312
## 14 -0.9632863  0.87358895  0.19240011 -0.96753478 -0.9610792     -
1.8450624
## 15  1.2782387 -0.25595600 -0.40231769  0.98142435  0.8429577
1.8450624
## 16  0.6654710 -1.30760129 -0.23677768 -0.52338423  0.1288598     -
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 Cluster_5.cluster
## 1  -0.21209793 -0.52776752        0.06168225                 2
## 2   0.01828430 -0.38113909       -1.55366706                 5
## 3  -0.40408312 -0.57211809       -0.68503583                 2
## 4  -0.74965647  0.14744734        0.35122600                 2
## 5  -0.31449003  1.21638667       -0.42597037                 4
## 6  -0.74965647 -1.49714434       -1.99560225                 3
## 7  -0.02011273 -0.96584257        0.74744375                 2
## 8   3.74279705 -0.63276071       -1.24888417                 3
## 9   0.61983791  1.88617085       -0.36501379                 4
## 10 -0.07130879 -0.64814764        1.17413980                 2
## 11 -0.31449003  0.76926048        0.82363947                 1
## 12  1.10620040  0.05603085       -0.71551412                 3
## 13 -0.62166634 -0.36213170        0.33598685                 1
## 14  0.44065173  1.53860717        0.85411776                 4
## 15 -0.39128411  0.36014907       -0.24310064                 1
## 16 -0.67286239 -1.45369888        1.02174835                 2
## 17 -0.54487226  1.10143723        1.44844440                 1
## 18 -0.30169102  0.14744734       -1.27936246                 5
## 19 -0.74965647 -0.43544591        0.29026942                 2
## 20 -0.49367621  1.43089863       -0.09070919                 4
## 21  0.68383297 -1.17763919        1.49416183                 2
```

Graph

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




```
library(cluster)
clusplot(df_2, Cluster_5$cluster, color=TRUE, shade=FALSE, labels=2, lines=0)
```



b. Interpreting the clusters:

Cluster 1- 17,11,13,15

-In this cluster Market capitalization, ROA, ROE, Asset Turnover and Net-Profit Margin is higher and leverage was the lowest.

-Most Of them are 2 are Hold and 2 are moderate-buy - median recommendation

Cluster2- 1,3,4,7,10,16,19,21

- second highest ROA and ROE and lowest revenue growth

-moderate buy, strong buy, 2 moderate sell, 4 hold- Median recommendation

Cluster3- 6,8,12

-lowest Market capital and ROA, high leverage, and Beta

-Hold, Moderate buy, Hold- Median recommendation

Cluster4- 20,9,5,14

-lowest P/E ratio, Asset turn over and highest revenue growth

Moderate buy, sell, moderate buy, moderate sell- Median recommendation

Cluster5-18,2

-low beta, ROE, NET-profit margin, and high P/E ratio

-Moderate buy, Hold- Median recommendation

c. pattern in the clusters with respect to the numerical variables

Cluster 1- There is a pattern in cluster 1 as it has high Market capitalization, ROE, ROA, Asset turn over and net profit margin it suggests HOLD/BUY which is similar to the median recommendation.

Cluster 2- As it has the lowest revenue growth but second-highest ROA, ROE, Asset-turn over ratio it is better to Hold/Sell as which most of the companies in the median recommendation suggests.

Cluster 3- No pattern

Cluster 4- No Pattern

Cluster 5- Here we have low beta which is low risk, low ROE, Net-profit margin with High P/E ratio it is better Hold/BUY which is similar to the Median recommendation

D. An appropriate name for each cluster:

Cluster 1- High market cap Roe Roa asset turnover net profit cluster

Cluster 2-low revenue high ROA ROE cluster

Cluster 3- low market cap ROA high leverage beta cluster

Cluster 4- low P/E asset t/o high revenue cluster

Cluster 5- High P/E low beta cluster