Assignment_4

SIVAREDDY

2022-03-20

```
getwd()
```

```
## [1] "C:/Users/sivap/OneDrive/Desktop/SIVA"
```

```
library(readr)
Pharmaceuticals <-
read_csv("C:/Users/sivap/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>
```

a. 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.

We have selected the variables of 3 to 11 which are 1 to 9

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

```
## # 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
```

#We have scaled all numerical variables in the data frame

```
scaling_Numericalvariables <- scale (Numericalvariables)
head(scaling_Numericalvariables)
```

```
##      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_Numericalvariables)-
1)*sum(apply(scaling_Numericalvariables,2,var))
Clusters
```

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

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

```
##  [1] 180.00000 118.56934  95.99420  79.21748  67.16336  54.55263  43.01927
##  [8]  41.12605  36.26082  23.31081  21.76304  18.67084  12.61696  12.73515
## [15]  10.27798
```

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

fit <- kmeans(scaling_Numericalvariables, 5)
aggregate(scaling_Numericalvariables,by=list(fit$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 -0.57238455 -0.6220844  0.86927480 -0.7381675 -0.7242993
## 3       3  0.08926902 -0.4618336 -0.32086149  0.3260892  0.5396003
## 4       4 -0.96686975  1.5162611 -0.57398880 -0.8382671 -0.9892673
## 5       5 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478
##   Asset_Turnover   Leverage Rev_Growth Net_Profit_Margin
## 1   1.153164e+00 -0.4680782  0.4671788         0.5912425
## 2   1.776140e-16 -0.2991312  0.3682951        -0.8069490
## 3   6.589509e-02 -0.2559803 -0.7230135         0.7343816
## 4  -1.845062e+00  0.5302448  1.7123890         0.2445520
## 5  -4.612656e-01  1.3664470 -0.6912914        -1.3200002
```

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

Cluster_Number
```

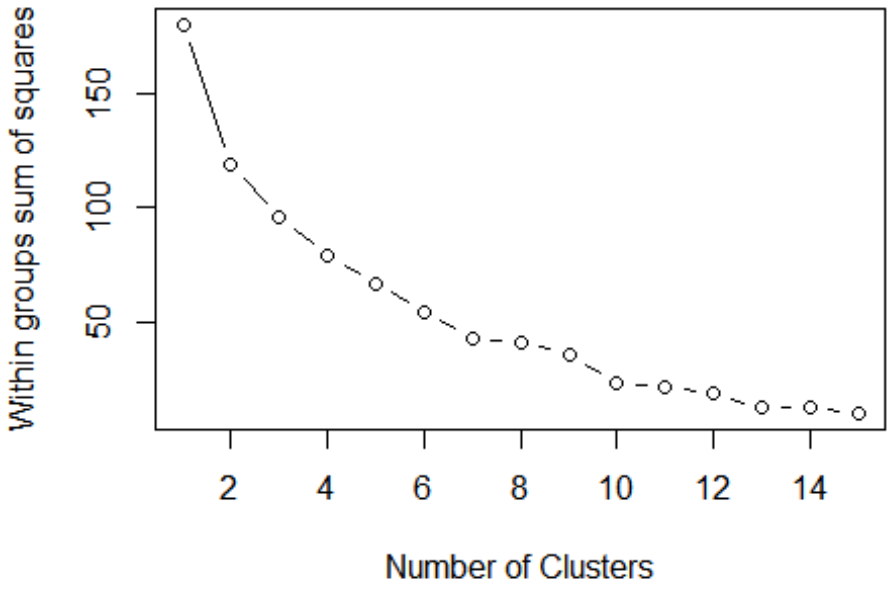
```
##    Market_Cap        Beta    PE_Ratio         ROE        ROA
Asset_Turnover
## 1   0.1840960 -0.80125356 -0.04671323  0.04009035  0.2416121
0.0000000
## 2  -0.8544181 -0.45070513  3.49706911 -0.85483986 -0.9422871
0.9225312
## 3  -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700
0.9225312
## 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           3
## 2   0.01828430 -0.38113909       -1.55366706           2
## 3  -0.40408312 -0.57211809       -0.68503583           2
## 4  -0.74965647  0.14744734        0.35122600           3
## 5  -0.31449003  1.21638667       -0.42597037           2
## 6  -0.74965647 -1.49714434       -1.99560225           5
## 7  -0.02011273 -0.96584257        0.74744375           3
## 8   3.74279705 -0.63276071       -1.24888417           5
## 9   0.61983791  1.88617085       -0.36501379           4
## 10 -0.07130879 -0.64814764        1.17413980           3
## 11 -0.31449003  0.76926048        0.82363947           1
## 12  1.10620040  0.05603085       -0.71551412           5
## 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           3
## 17 -0.54487226  1.10143723        1.44844440           1
## 18 -0.30169102  0.14744734       -1.27936246           2
```

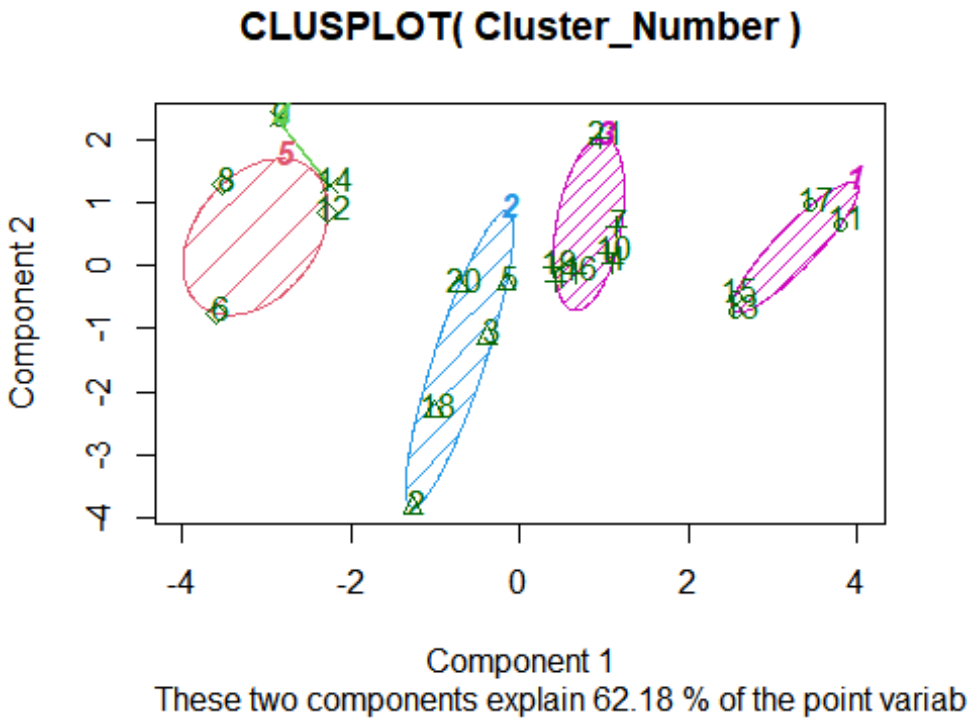
```
## 19 -0.74965647 -0.43544591        0.29026942           3
## 20 -0.49367621  1.43089863       -0.09070919           2
## 21  0.68383297 -1.17763919        1.49416183           3
```

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

The plot shows 5 clusters are sufficient to capture most variations in the data.




```
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_Numericalvariables, 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