FML assignment 4

SAKSHI

2023-11-10

Question

An equities analyst is studying the pharmaceutical industry and would like your help in exploring and understanding the financial data collected by her firm. Her main objective is to understand the structure of the pharmaceutical industry using some basic financial measures. Financial data gathered on 21 firms in the pharmaceutical industry are available in the file Pharmaceuticals.csv Download Pharmaceuticals.csv. For each firm, the following variables are recorded:

-Market capitalization (in billions of dollars) -Beta -Price/earnings ratio -Return on equity -Return on assets -Asset turnover -Leverage -Estimated revenue growth -Net profit margin -Median recommendation (across major brokerages) -Location of firm's headquarters -Stock exchange on which the firm is listed -Use cluster analysis to explore and analyze the given dataset as follows:

1.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.

2.Interpret the clusters with respect to the numerical variables used in forming the clusters. Is there a pattern in the clusters with respect to the numerical variables (10 to 12)? (those not used in forming the clusters).

3.Provide an appropriate name for each cluster using any or all of the variables in the dataset.

Summary

1.k-means is the most appropriate clustering algorithm with k=5.

2.we got 5 clusters-

one with highest Market_Cap, highest ROE, highest ROA, highest Asset_Turnover has equal Hold and Moderate Buy Recommendation.

one with lowest PE_Ratio and lowest Asset_Turnover has Hold Recommendation.

one with highest Beta, highest Leverage has mostly Moderate Buy Recommendation.

one with highest PE_Ratio has Hold Recommendation.

one with highest Net_Profit_Margin has mostly Hold Recommendation.

3.

Loading all the required packages

```r
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.3     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.4.3     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(cluster)
library(factoextra)
```

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

```r
library(ISLR)
library(caret)
```

```
Loading required package: lattice

Attaching package: 'caret'

The following object is masked from 'package:purrr':

    lift
```

```r
library(tidyr)
library(readr)
library(knitr)
library(dplyr)
library(stats)
library(dbscan)
```

```

Attaching package: 'dbscan'

The following object is masked from 'package:stats':

    as.dendrogram
```

Loading the dataset into the R

```r
pharma.df.2 <- read.csv("/Users/sakshibansal/Downloads/Pharmaceuticals.csv")
```

Understanding the data structure

```r
dim(pharma.df.2)
```

```
[1] 21 14
```

```r
head(pharma.df.2, n=10)
```

```
   Symbol                         Name Market_Cap Beta PE_Ratio  ROE  ROA
1     ABT          Abbott Laboratories      68.44 0.32     24.7 26.4 11.8
2     AGN               Allergan, Inc.       7.58 0.41     82.5 12.9  5.5
3     AHM                 Amersham plc       6.30 0.46     20.7 14.9  7.8
4     AZN              AstraZeneca PLC      67.63 0.52     21.5 27.4 15.4
5     AVE                      Aventis      47.16 0.32     20.1 21.8  7.5
6     BAY                     Bayer AG      16.90 1.11     27.9  3.9  1.4
7     BMY Bristol-Myers Squibb Company      51.33 0.50     13.9 34.8 15.1
8    CHTT                Chattem, Inc       0.41 0.85     26.0 24.1  4.3
9     ELN        Elan Corporation, plc       0.78 1.08      3.6 15.1  5.1
10    LLY         Eli Lilly and Company      73.84 0.18     27.9 31.0 13.5
   Asset_Turnover Leverage Rev_Growth Net_Profit_Margin Median_Recommendation
1             0.7     0.42       7.54              16.1          Moderate Buy
2             0.9     0.60       9.16               5.5          Moderate Buy
3             0.9     0.27       7.05              11.2            Strong Buy
4             0.9     0.00      15.00              18.0         Moderate Sell
5             0.6     0.34      26.81              12.9          Moderate Buy
6             0.6     0.00      -3.17               2.6                  Hold
7             0.9     0.57       2.70              20.6         Moderate Sell
8             0.6     3.51       6.38               7.5          Moderate Buy
9             0.3     1.07      34.21              13.3         Moderate Sell
10            0.6     0.53       6.21              23.4                  Hold
   Location Exchange
1        US     NYSE
2    CANADA     NYSE
3        UK     NYSE
4        UK     NYSE
5    FRANCE     NYSE
6   GERMANY     NYSE
7        US     NYSE
8        US   NASDAQ
9   IRELAND     NYSE
10       US     NYSE
```

```r
str(pharma.df.2)
```

```
'data.frame':   21 obs. of  14 variables:
 $ Symbol                : chr  "ABT" "AGN" "AHM" "AZN" ...
 $ Name                  : chr  "Abbott Laboratories" "Allergan, Inc." "Amersham plc" "AstraZeneca PLC"
```

```
 $ Market_Cap           : num  68.44 7.58 6.3 67.63 47.16 ...
 $ Beta                 : num  0.32 0.41 0.46 0.52 0.32 1.11 0.5 0.85 1.08 0.18 ...
 $ PE_Ratio             : num  24.7 82.5 20.7 21.5 20.1 27.9 13.9 26 3.6 27.9 ...
 $ ROE                  : num  26.4 12.9 14.9 27.4 21.8 3.9 34.8 24.1 15.1 31 ...
 $ ROA                  : num  11.8 5.5 7.8 15.4 7.5 1.4 15.1 4.3 5.1 13.5 ...
 $ Asset_Turnover       : num  0.7 0.9 0.9 0.9 0.6 0.6 0.9 0.6 0.3 0.6 ...
 $ Leverage             : num  0.42 0.6 0.27 0 0.34 0 0.57 3.51 1.07 0.53 ...
 $ Rev_Growth           : num  7.54 9.16 7.05 15 26.81 ...
 $ Net_Profit_Margin    : num  16.1 5.5 11.2 18 12.9 2.6 20.6 7.5 13.3 23.4 ...
 $ Median_Recommendation: chr  "Moderate Buy" "Moderate Buy" "Strong Buy" "Moderate Sell" ...
 $ Location             : chr  "US" "CANADA" "UK" "UK" ...
 $ Exchange             : chr  "NYSE" "NYSE" "NYSE" "NYSE" ...
```

```r
summary(pharma.df.2)
```

```
    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  
```

Cleaning the data to remove every n/a values

```r
pharma.df.2 <- na.omit(pharma.df.2)
```

1a.Use only the numerical variables (1 to 9) to cluster the 21 firms

Dropping all the non-numeric varaibles

```
pharma.df = pharma.df.2[ , c(3:11)]
pharma.df
```

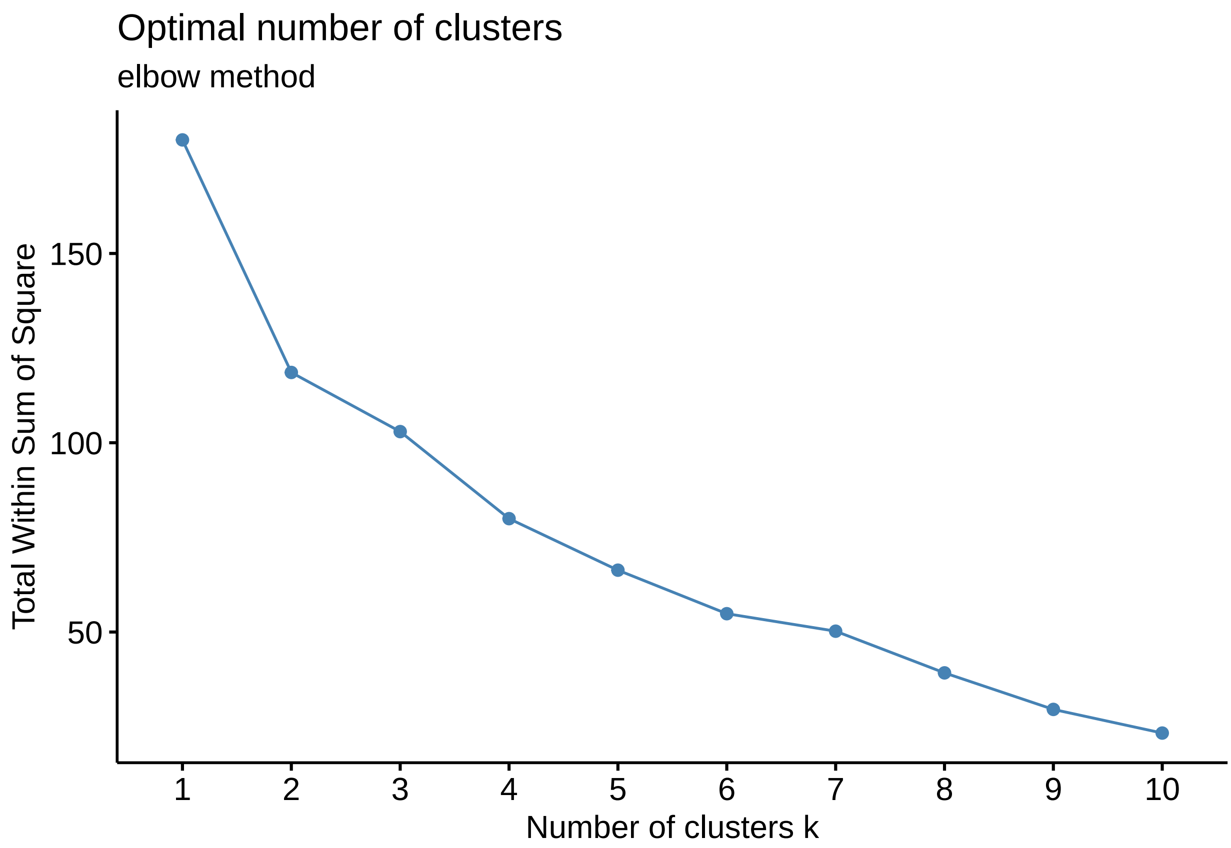
```
   Market_Cap Beta PE_Ratio  ROE  ROA Asset_Turnover Leverage Rev_Growth
1       68.44 0.32     24.7 26.4 11.8            0.7     0.42       7.54
2        7.58 0.41     82.5 12.9  5.5            0.9     0.60       9.16
3        6.30 0.46     20.7 14.9  7.8            0.9     0.27       7.05
4       67.63 0.52     21.5 27.4 15.4            0.9     0.00      15.00
5       47.16 0.32     20.1 21.8  7.5            0.6     0.34      26.81
6       16.90 1.11     27.9  3.9  1.4            0.6     0.00      -3.17
7       51.33 0.50     13.9 34.8 15.1            0.9     0.57       2.70
8        0.41 0.85     26.0 24.1  4.3            0.6     3.51       6.38
9        0.78 1.08      3.6 15.1  5.1            0.3     1.07      34.21
10      73.84 0.18     27.9 31.0 13.5            0.6     0.53       6.21
11     122.11 0.35     18.0 62.9 20.3            1.0     0.34      21.87
12       2.60 0.65     19.9 21.4  6.8            0.6     1.45      13.99
13     173.93 0.46     28.4 28.6 16.3            0.9     0.10       9.37
14       1.20 0.75     28.6 11.2  5.4            0.3     0.93      30.37
15     132.56 0.46     18.9 40.6 15.0            1.1     0.28      17.35
16      96.65 0.19     21.6 17.9 11.2            0.5     0.06      -2.69
17     199.47 0.65     23.6 45.6 19.2            0.8     0.16      25.54
18      56.24 0.40     56.5 13.5  5.7            0.6     0.35      15.00
19      34.10 0.51     18.9 22.6 13.3            0.8     0.00       8.56
20       3.26 0.24     18.4 10.2  6.8            0.5     0.20      29.18
21      48.19 0.63     13.1 54.9 13.4            0.6     1.12       0.36
   Net_Profit_Margin
1               16.1
2                5.5
3               11.2
4               18.0
5               12.9
6                2.6
7               20.6
8                7.5
9               13.3
10              23.4
11              21.1
12              11.0
13              17.9
14              21.3
15              14.1
16              22.4
17              25.2
18               7.3
19              17.6
20              15.1
21              25.5
```

Normalizing the dataset

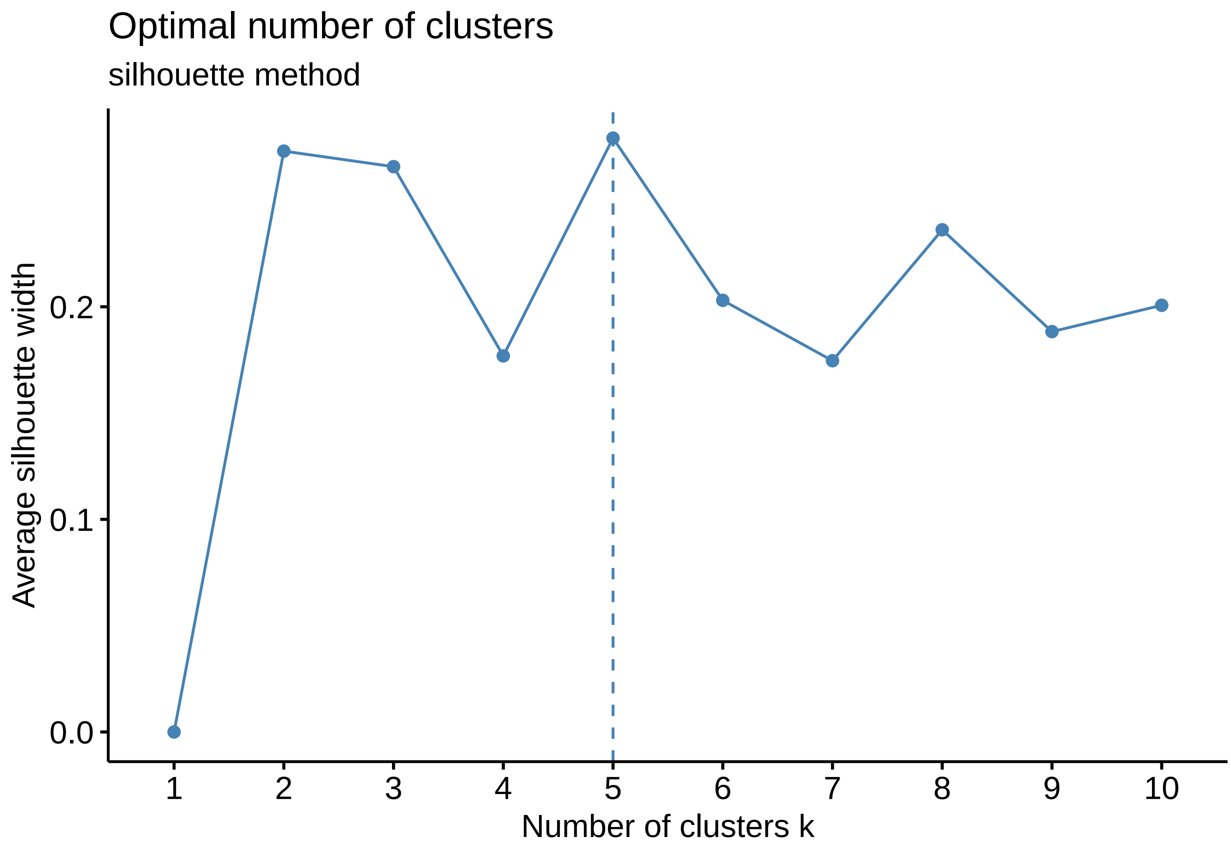
```r
df.norm = preProcess(pharma.df , method = c("center", "scale"))
pharma.norm = predict(df.norm , pharma.df)
```

Finding the value of k using elbow and silhouette method

```r
fviz_nbclust(pharma.norm , kmeans , method = "wss") + labs(subtitle = "elbow method")
```




```r
fviz_nbclust(pharma.norm , kmeans , method = "silhouette") + labs(subtitle = "silhouette method")
```



The chart shows that the elbow point 5 provides the best value for k. While WSS will continue to drop for larger values of k, we have to make the trade off between over fitting, i.e., a model fitting both noise and signal, to a model having bias. Here, the elbow point provides that compromise where WSS, while still decreasing beyond k = 5, decreases at a much smaller rate. In other words, k = 5 provides the best value between bias and over fitting. Again, we see that 5 is the ideal number of clusters by looking at the large values for the Silhouette Width (Y Axis). This proves that K = 5 is the appropriate number.

Now making clusters and visualizing it using k-means

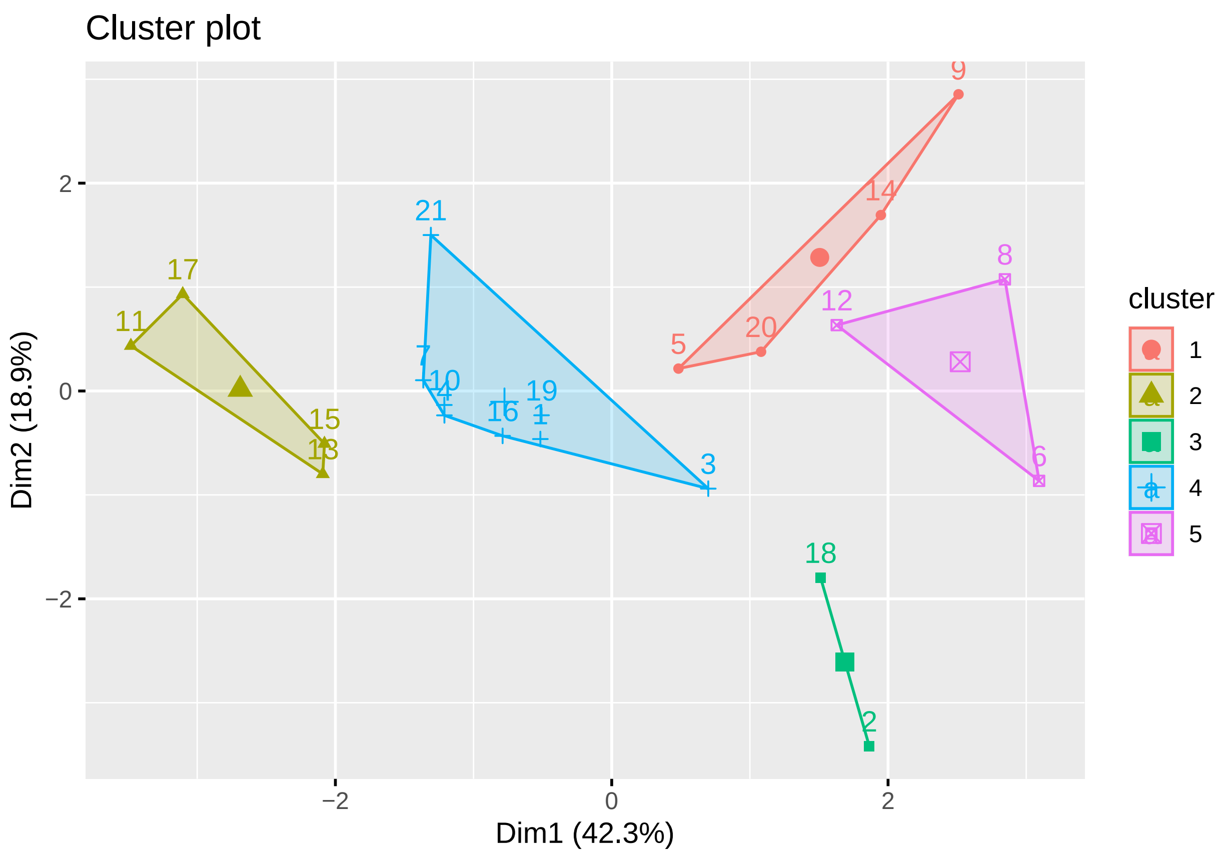
```
set.seed(250)
kcluster <- kmeans(pharma.norm , centers = 5 , nstart = 25)
kcluster$centers
```

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

```r
kcluster$size
```

```
[1] 4 4 2 8 3
```

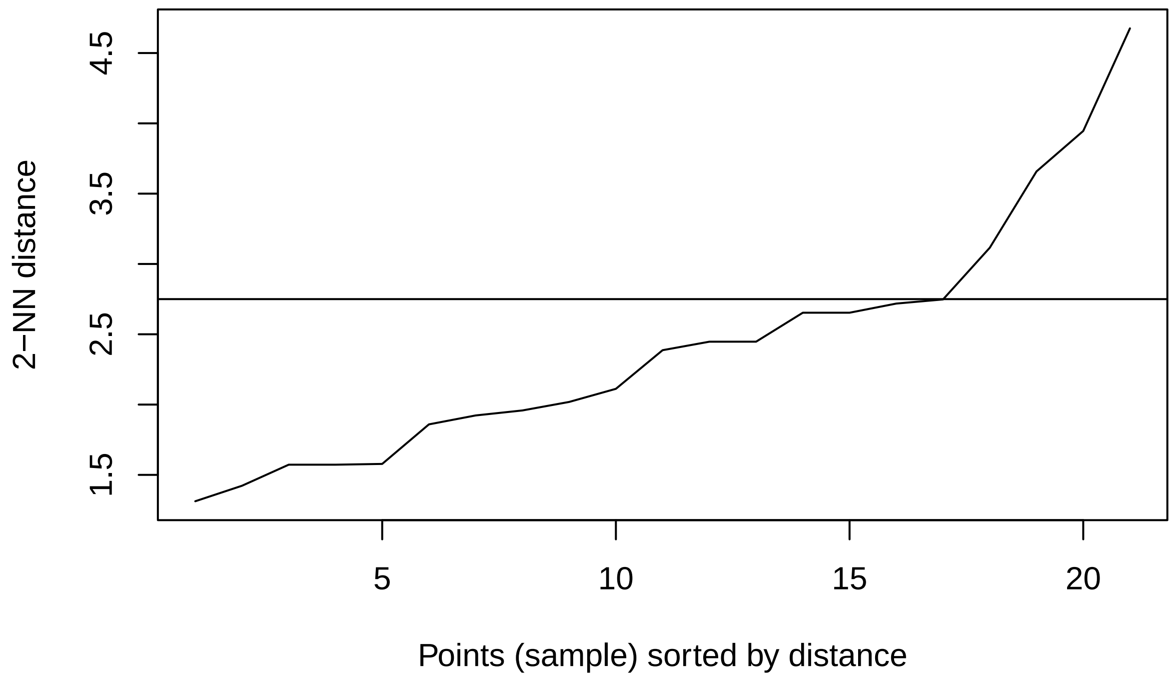
```r
fviz_cluster(kcluster , data = pharma.norm)
```



We can clearly see the 5 clusters being made.

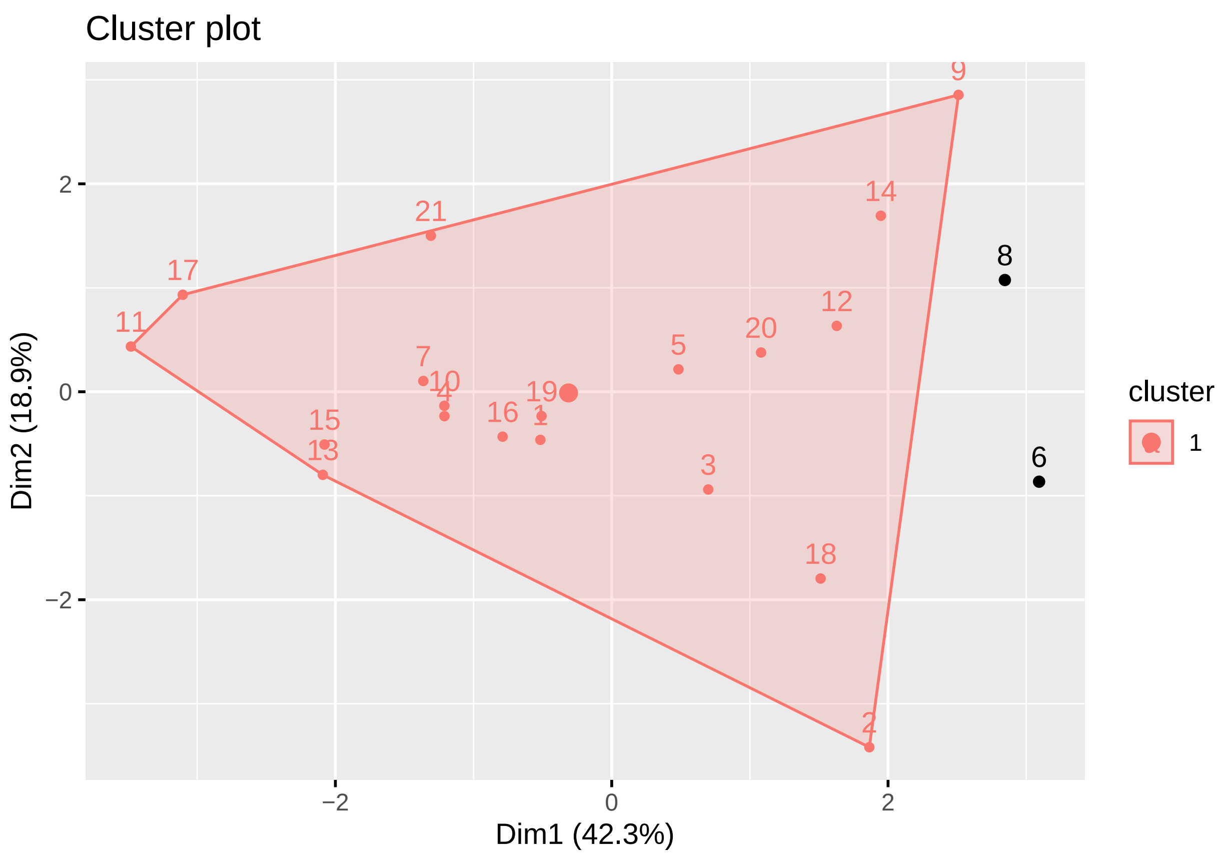
Making cluster using the dbscan method

```r
# Finding the optimal value of eps
dbscan::kNNdistplot(pharma.norm , k=2)
abline(h = 2.75)
```




```
pharma.db <- dbscan::dbscan(pharma.norm , eps = 2.75 , minPts = 2)

fviz_cluster(pharma.db , pharma.norm)
```



We can clearly see the 1 cluster being made with two outliers.

Making cluster using the hierarchical method

```
# Using agnes and finding the best linkage method

pharma.single <- agnes(pharma.norm , method = "single")
pharma.complete <- agnes(pharma.norm , method = "complete")
pharma.average <- agnes(pharma.norm , method = "average")

# Comparing their agglomerative function

print(pharma.single$ac)
```

```
[1] 0.4600348
```

```
print(pharma.complete$ac)
```

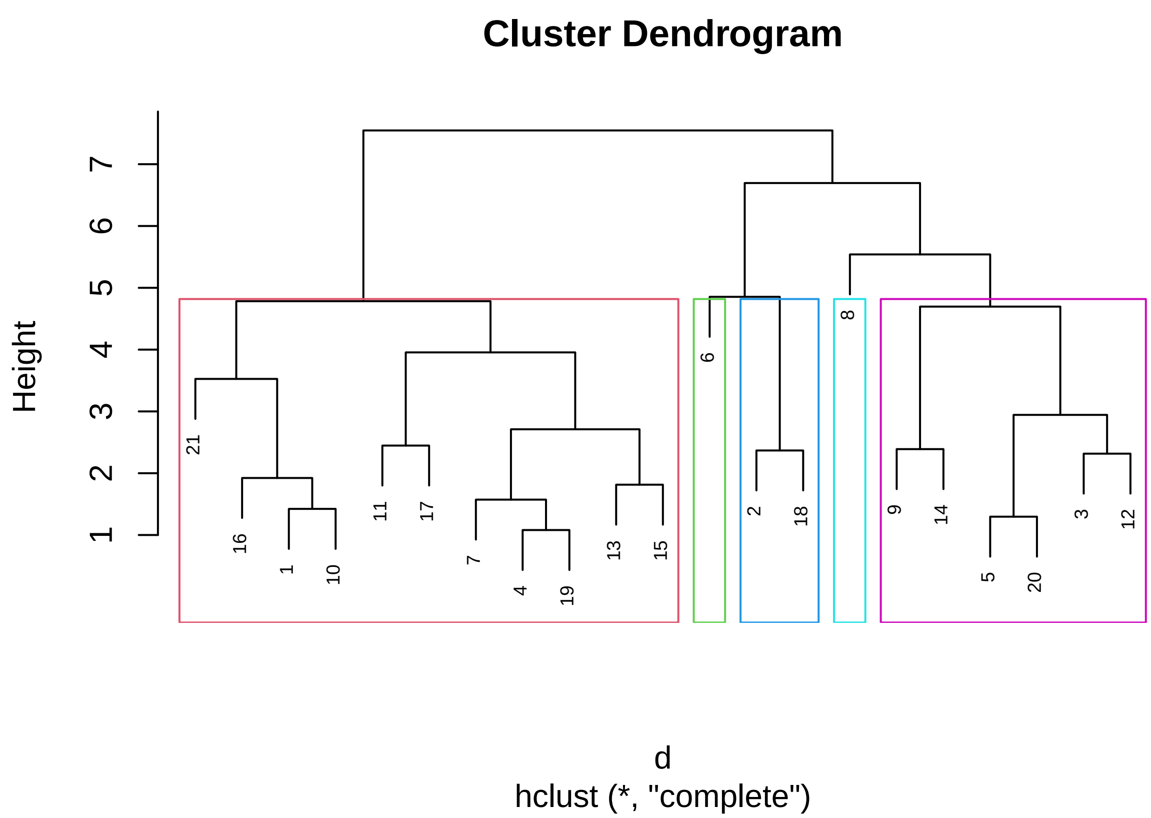
```
[1] 0.6990833
```

```
print(pharma.average$ac)
```

```
[1] 0.5600652
```

```r
# We are going with the complete one since it has highest agglomerative function

# Using hclust to make the clusters
d <- dist(pharma.norm , method = "euclidean")
pharma.complete.2 <- hclust(d, method = "complete")
plot(pharma.complete.2 , cex = 0.6)
rect.hclust(pharma.complete.2 , k=5 , border=2:6)
```



We can clearly see the 5 clusters being made.

1b.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.

Interpretation

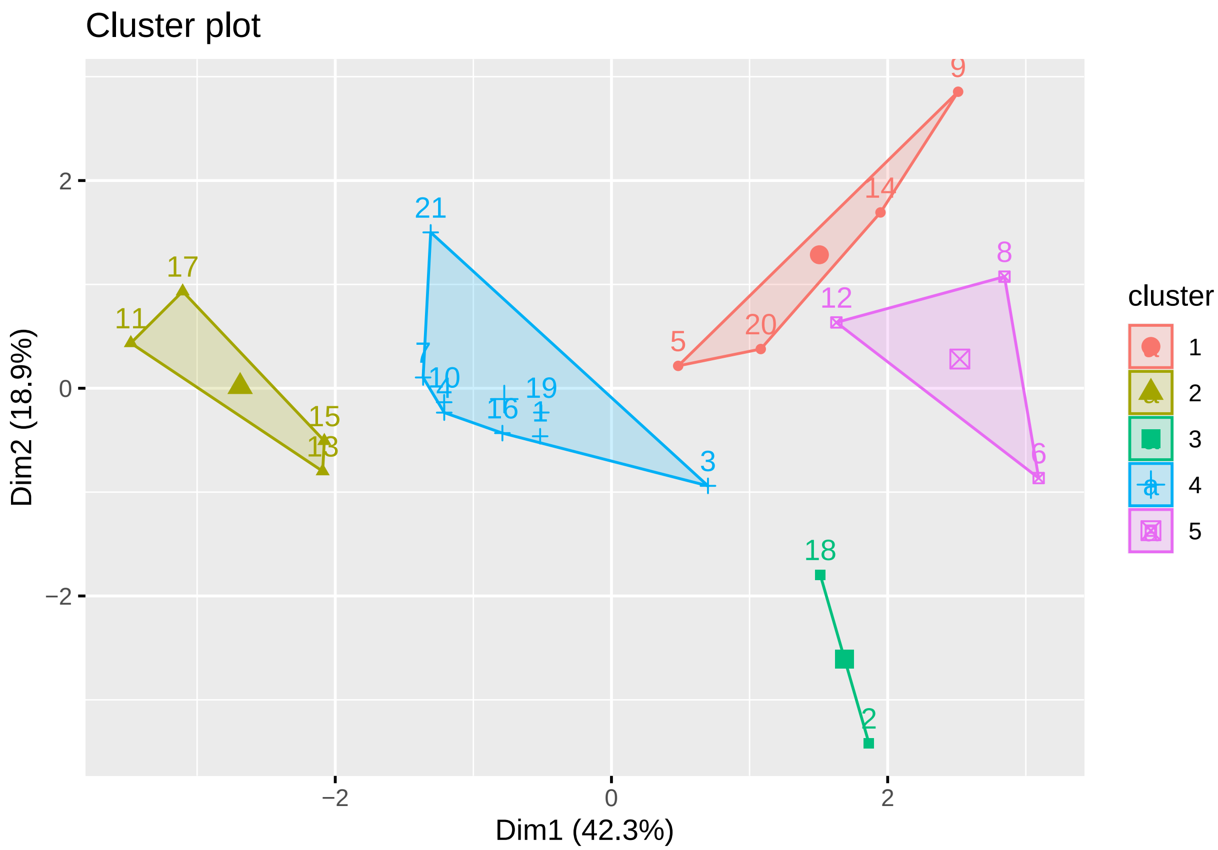
As we don't want the clustering algorithm to depend to an arbitrary variable unit, we started by normalizing the data and setting equal weights to all the variables so we can get a more clear picture of the clusters.

I used all the 3 methods for clustering that it- k-means , dbscan , hierarchical to analyze how differently data can be distributed across the platform. k-means gave me 5 cluster which I was able to find out by using the both elbow and silhouette method. Furthermore, I provided the same k for hierarchical clustering to get a comparison. Lastly, dbscan gave me a single cluster with 2 outliers which indicated that it won't be an appropriate technique to apply on the given dataset. Furthermore, I was able to observe that dbscan and hierarchical method gave me same outliers.

Lastly, between k-means , dbscan and hierarchical method, I believe k-means is a more appropriate fit for clustering this dataset because both the other methods gave us outliers which should have been included in the dataset since they are very close to the other points as shown in k-means clustering. Furthermore, this is a financial dataset and there is no hierarchy in it which means we can't do clustering on the basis of the same.

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

```r
fviz_cluster(kcluster , data = pharma.norm)
```



Now we are going to view the data points in order to intepret the 4 clusters

```r
cluster1 <- kable(pharma.norm[c(5,9,14,20),], align = "c")
cluster1
```

	Market_Cap	Beta	PE_Ratio	ROE	ROA	Asset_Turnover	Leverage	Rev_Growth	Net_Profit_Margin
5	-0.1790256	-0.8012536	-0.3287443	-0.2648488	-0.5664461	-0.4612656	-0.3144900	1.216387	-0.4259704
9	-0.9704532	2.1589332	-1.3403777	-0.7089994	-1.0174553	-1.8450624	0.6198379	1.886171	-0.3650138
14	-0.9632863	0.8735889	0.1924001	-0.9675348	-0.9610792	-1.8450624	0.4406517	1.538607	0.8541178

	Market_Cap	Beta	PE_Ratio	ROE	ROA	Asset_Turnover	Leverage	Rev_Growth	Net_Profit_Margin
20	-0.9281345	-1.1128522	-0.4329732	-1.0338259	-0.6979905	-0.9225312	-0.4936762	1.430899	-0.0907092

This cluster has low values for Marketcap, PE-ratio, ROE, ROA and Net_Profit , high Beta, high Leverage but all other values are average.

```
cluster2 <- kable(pharma.norm[c(11,15,17,13),], align = "c")
cluster2
```

	Market_Cap	Beta	PE_Ratio	ROE	ROA	Asset_Turnover	Leverage	Rev_Growth	Net_Profit_Margin
11	1.099920	-0.6844041	-0.4574977	2.4597165	1.8389364	1.3837968	-0.3144900	0.7692605	0.8236395
15	1.278239	-0.2559560	-0.4023177	0.9814243	0.8429577	1.8450624	-0.3912841	0.3601491	-0.2431006
17	2.419990	0.4840907	-0.1141555	1.3128800	1.6322239	0.4612656	-0.5448723	1.1014372	1.4484444
13	1.984176	-0.2559560	0.1801379	0.1859308	1.0872544	0.9225312	-0.6216663	-0.3621317	0.3359869

All companies in this cluster have high values besides Beta, PE ratio and Leverage.

```
cluster3 <- kable(pharma.norm[c(2,18),], align = "c")
cluster3
```

	Market_Cap	Beta	PE_Ratio	ROE	ROA	Asset_Turnover	Leverage	Rev_Growth	Net_Profit_Margin
2	-0.8544181	-0.4507051	3.497069	-0.8548399	-0.9422871	0.9225312	0.0182843	-0.3811391	-1.553667
18	-0.0240846	-0.4896550	1.902980	-0.8150652	-0.9047030	-0.4612656	-0.3016910	0.1474473	-1.279362

This cluster has a high PE ratio but very low values for other variables.

```
cluster4 <- kable(pharma.norm[c(1,3,4,7,10,16,19,21),], align = "c")
cluster4
```

	Market_Cap	Beta	PE_Ratio	ROE	ROA	Asset_Turnover	Leverage	Rev_Growth	Net_Profit_Margin
1	0.1840960	-0.8012536	-0.0467132	0.0400903	0.2416121	0.0000000	-0.2120979	-0.5277675	0.0616823
3	-0.8762600	-0.2559560	-0.2919577	-0.7222576	-0.5100700	0.9225312	-0.4040831	-0.5721181	-0.6850358
4	0.1702742	-0.0222570	-0.2429088	0.1063815	0.9181259	0.9225312	-0.7496565	0.1474473	0.3512260
7	-0.1078688	-0.1001567	-0.7088733	0.5969358	0.8617498	0.9225312	-0.0201127	-0.9658426	0.7474438

	Market_Cap	Beta	PE_Ratio	ROE	ROA	Asset_Turnover	Leverage	Rev_Growth	Net_Profit_Margin
10	0.2762415	-1.3465511	0.1494823	0.3450295	0.5610770	-0.4612656	-0.0713088	-0.6481476	1.1741398
16	0.6654710	-1.3076013	-0.2367777	-0.5233842	0.1288598	-0.9225312	-0.6728624	-1.4536989	1.0217484
19	-0.4018812	-0.0612069	-0.4023177	-0.2118159	0.5234929	0.4612656	-0.7496565	-0.4354459	0.2902694
21	-0.1614497	0.4061910	-0.7579221	1.9293875	0.5422849	-0.4612656	0.6838330	-1.1776392	1.4941618

This cluster has highest net profit margin.

```r
cluster5 <- kable(pharma.norm[c(6,8,12),], align = "c")
cluster5
```

	Market_Cap	Beta	PE_Ratio	ROE	ROA	Asset_Turnover	Leverage	Rev_Growth	Net_Profit_Margin
6	-0.6953818	2.2757827	0.1494823	-1.4514600	-1.7127612	-0.4612656	-0.7496565	-1.4971443	-1.9956023
8	-0.9767669	1.2630872	0.0329912	-0.1123792	-1.1677918	-0.4612656	3.7427970	-0.6327607	-1.2488842
12	-0.9393967	0.4840907	-0.3410066	-0.2913653	-0.6979905	-0.4612656	1.1062004	0.0560309	-0.7155141

This cluster has high beta value , average per ration and all other variables have mixed values.

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

Identifying the patter of remaining variables

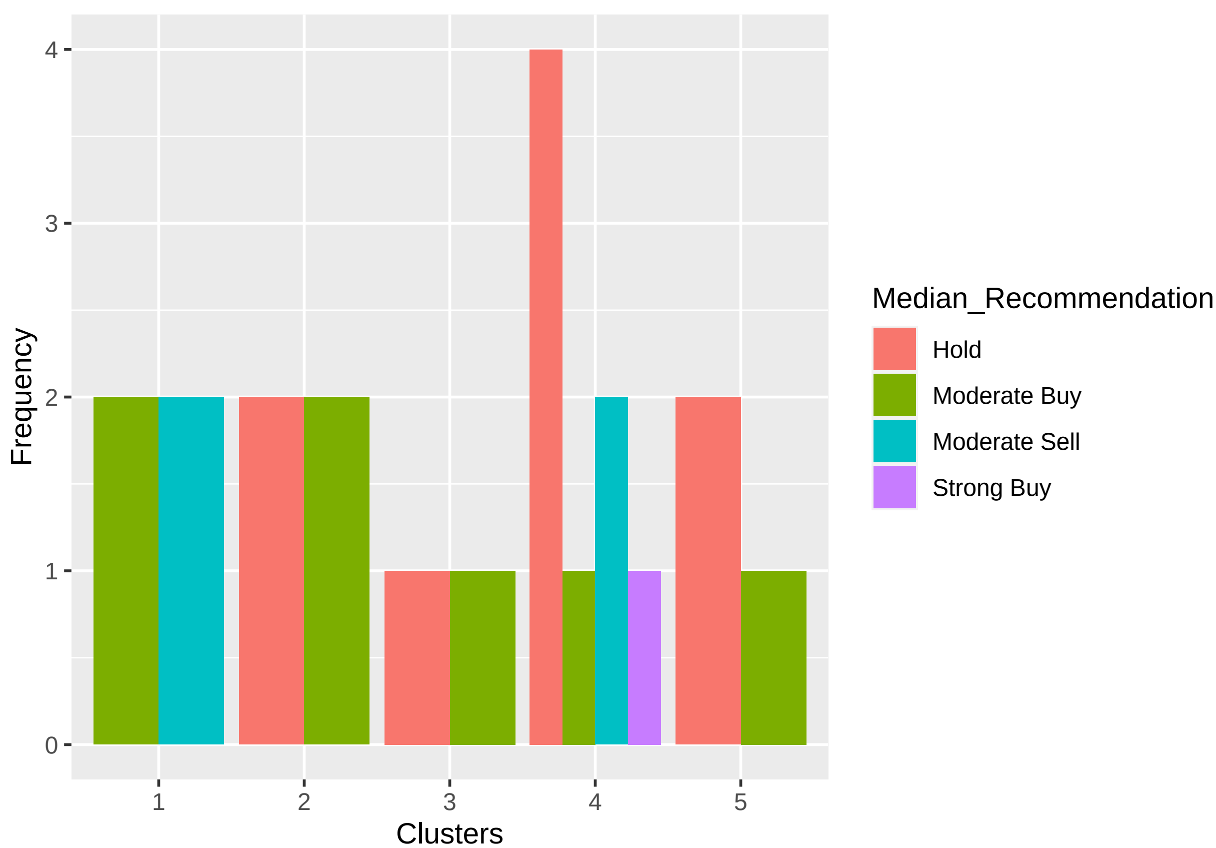
```r
cluster_Pattern <- pharma.df.2 %>%
  select(c(1,12,13,14)) %>%
  mutate(Cluster = kcluster$clust)
print(cluster_Pattern)
```

```
   Symbol Median_Recommendation    Location Exchange Cluster
1     ABT           Moderate Buy         US     NYSE       4
2     AGN           Moderate Buy     CANADA     NYSE       3
3     AHM             Strong Buy         UK     NYSE       4
4     AZN          Moderate Sell         UK     NYSE       4
5     AVE           Moderate Buy     FRANCE     NYSE       1
6     BAY                   Hold    GERMANY     NYSE       5
7     BMY          Moderate Sell         US     NYSE       4
8    CHTT           Moderate Buy         US   NASDAQ       5
9     ELN          Moderate Sell    IRELAND     NYSE       1
10    LLY                   Hold         US     NYSE       4
11    GSK                   Hold         UK     NYSE       2
12    IVX                   Hold         US     AMEX       5
```

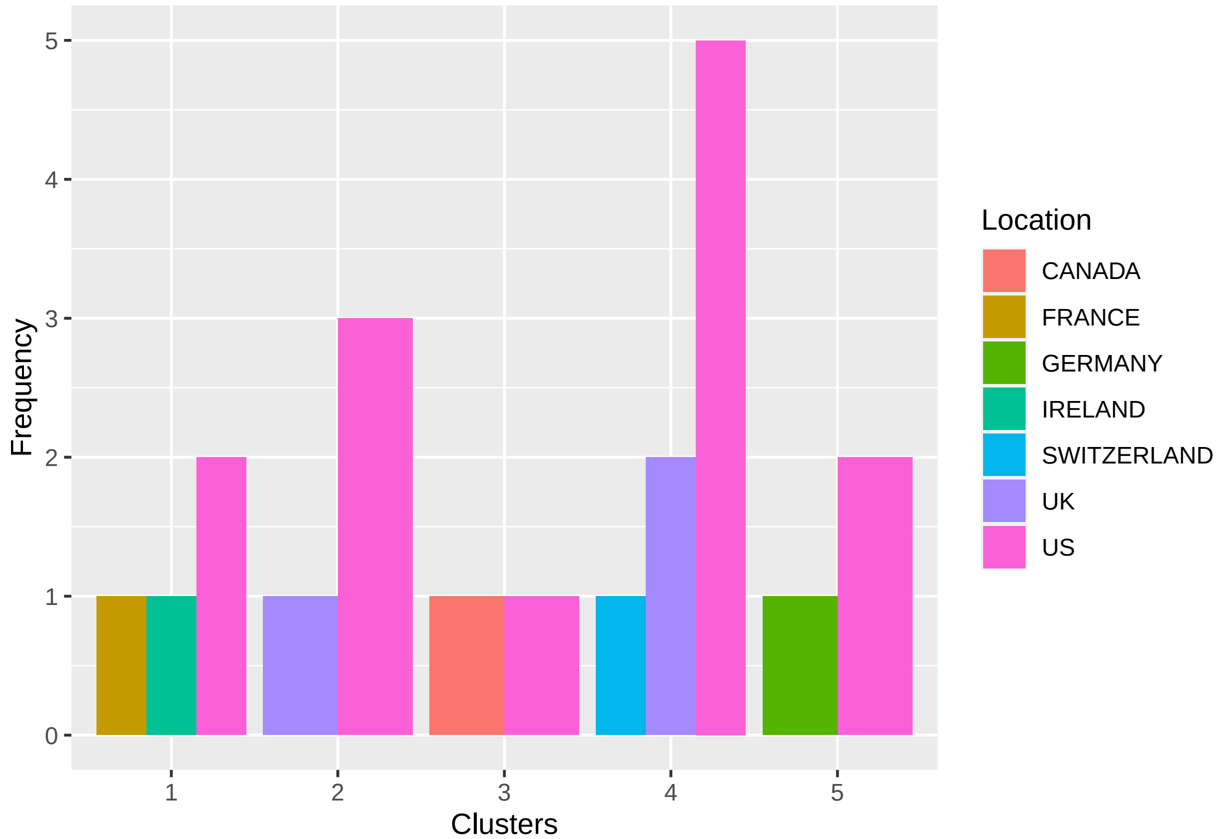
```
13     JNJ          Moderate Buy          US      NYSE       2
14     MRX          Moderate Buy          US      NYSE       1
15     MRK                  Hold          US      NYSE       2
16     NVS                  Hold SWITZERLAND      NYSE       4
17     PFE          Moderate Buy          US      NYSE       2
18     PHA                  Hold          US      NYSE       3
19     SGP                  Hold          US      NYSE       4
20     WPI         Moderate Sell          US      NYSE       1
21     WYE                  Hold          US      NYSE       4
```

Now visualizing the data points to find the pattern

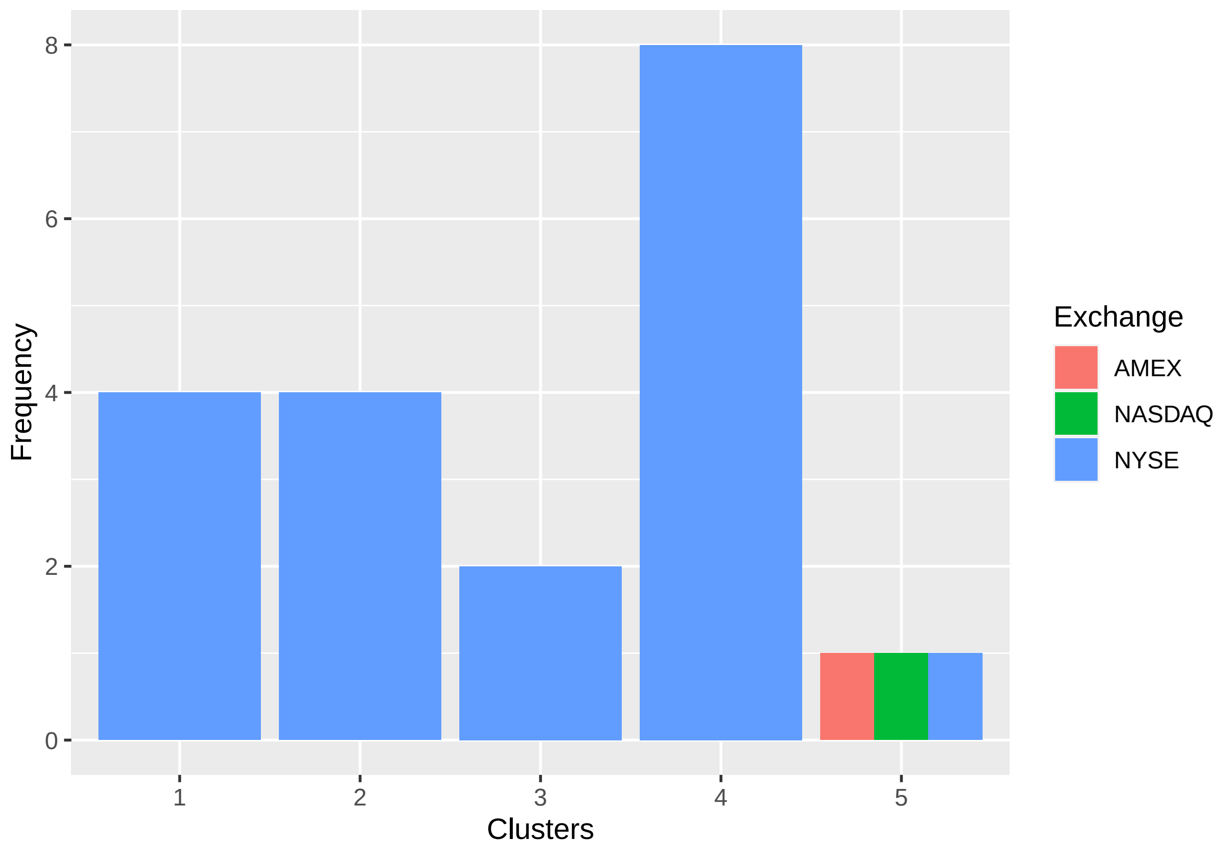
```
ggplot(cluster_Pattern, mapping = aes(factor(Cluster), fill=Median_Recommendation)) + geom_bar(position
```




```
ggplot(cluster_Pattern, mapping = aes(factor(Cluster), fill=Location)) + geom_bar(position = 'dodge') +
```




```
ggplot(cluster_Pattern, mapping = aes(factor(Cluster), fill=Exchange)) + geom_bar(position = 'dodge') +
```



Interpretation

From all the above graphs we can say-

cluster1 - All the companies are listed in NYSE and trades from various locations like France, Germany and US. Also, they are indulged in partially equal moderate buy and sell.

cluster2 - All the companies are listed in NYSE and trades from two locations which are UK AND US. Also, they are indulged in partially hold and moderate buy.

cluster3 - All the companies are listed in NYSE and trades from two locations which are Canada AND US. Also, they are indulged in very low hold and moderate buy.

cluster4 - All the companies are listed in NYSE and trades from various locations like Switzerland, UK and US being the dominant one. Also, they are indulged in highest hold with very low moderate/strong buy.

cluster5 - All the companies are listed in all 3 exchanges and trades from Germany and US. Also, they are indulged in average hold and low moderate buy.

In terms of variables, a pattern can be seen among the clusters (10 to 12).

Clusters 1,2 has mostly Moderate Buy Recommendation

Clusters 2,3,4,5 has Hold Recommendation

3.Provide an appropriate name for each cluster using any or all of the variables in the dataset.

Based on the entire analysis and looking at the characterstics of the clusters, 21 pharmaceutical industries can be categorized into 5 different groups:

Cluster 1 - “Startups”: A company with low asset turnover and high revenue growth may indicate that the company has significant growth potential but does not have sufficient investment to meet the needs. We can say that this cluster consists of companies who observes low risk and are growth oriented.

Cluster 2 - “MNC”: Companies with high market capitalization are typically large and well-established companies that have a significant market presence and a strong financial position. We can say that this cluster consists of settled and profitable companies.

Cluster 3 - “Gamble”: since it has high price-to-earnings (PE) ratio and a low net profit margin means that such companies can be risky, as they may not be able to meet the market’s expectations and may experience a decline in stock price in the future.We can say that this cluster consists of overpriced and risky companies

Cluster 4 - “Unicorns”: company with normal levels across financial metrics can be considered that the company is operating efficiently and effectively.We can say that this cluster consists of stable and efficient companies

Cluster 5 - “Leveraged”: Companies with high leverage and low net profit margin & ROA may indicate that the company is taking on a significant amount of debt to finance its operations, while not generating a sufficient level of profitability or returns on assets.We can say that this cluster consists of debt-ridden and very risky companies.