BA64060_Assignment4

Durga Prasad Gandi

2023-11-10

Load the required packages

```r
library(flexclust)
```

```
## Warning: package 'flexclust' was built under R version 4.3.2
```

```
## Loading required package: grid
```

```
## Loading required package: lattice
```

```
## Loading required package: modeltools
```

```
## Loading required package: stats4
```

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

```
## Warning: package 'factoextra' was built under R version 4.3.2
```

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

This library is used to extract and visualize information from the results of multivariate data analyses

```r
library(FactoMineR)
```

```
## Warning: package ’FactoMineR’ was built under R version 4.3.2
```

```r
library(ggcorrplot)
```

```
## Warning: package ’ggcorrplot’ was built under R version 4.3.2
```

```r
library(dbscan)
```

```
## Warning: package ’dbscan’ was built under R version 4.3.2

##
## Attaching package: ’dbscan’

## The following object is masked from ’package:stats’:
##
##     as.dendrogram
```

Load the dataset

```r
pharmaceutical = read.csv("C:/Users/gdurg/Downloads/Pharmaceuticals.csv")
pharmaceutical1 = pharmaceutical[3:11]
head(pharmaceutical1)
```

```
##   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
##   Net_Profit_Margin
## 1              16.1
## 2               5.5
## 3              11.2
## 4              18.0
## 5              12.9
## 6               2.6
```

First entity has the highest market cap at 68.44, while 3rd has the lowest at 6.30.With a beta of 1.11, entity 6 has the highest volatility when compared to the market.** With a PE ratio of 82.5, Entity 2 has the highest, indicating a rather expensive valuation.The PE ratio, which shows how much investors are ready to pay for each dollar of earnings, is calculated by dividing the company’s stock price by its earnings per share.With the lowest asset turnover (0.6), Entity 5 may be using its assets less effectively to drive sales. With a leverage ratio of 0.60, Entity 2 has the highest level of debt in its capital structure. Leverage quantifies how much debt a business uses to fund its operations. 5th Entity has experienced a significant rise in sales, as evidenced by its highest revenue growth of 26.81%. The percentage rise in revenues over a given period for a corporation is known as revenue growth. With a net profit margin of 16.1%, Entity 1 has the largest, suggesting a comparatively greater percentage of revenue kept as profit.

```r
summary(pharmaceutical1)
```

```
##    Market_Cap          Beta           PE_Ratio          ROE      
##  Min.   :  0.41   Min.   :0.1800   Min.   : 3.60   Min.   : 3.9  
##  1st Qu.:  6.30   1st Qu.:0.3500   1st Qu.:18.90   1st Qu.:14.9  
##  Median : 48.19   Median :0.4600   Median :21.50   Median :22.6  
##  Mean   : 57.65   Mean   :0.5257   Mean   :25.46   Mean   :25.8  
##  3rd Qu.: 73.84   3rd Qu.:0.6500   3rd Qu.:27.90   3rd Qu.:31.0  
##  Max.   :199.47   Max.   :1.1100   Max.   :82.50   Max.   :62.9  
##       ROA        Asset_Turnover    Leverage        Rev_Growth   
##  Min.   : 1.40   Min.   :0.3    Min.   :0.0000   Min.   :-3.17  
##  1st Qu.: 5.70   1st Qu.:0.6    1st Qu.:0.1600   1st Qu.: 6.38  
##  Median :11.20   Median :0.6    Median :0.3400   Median : 9.37  
##  Mean   :10.51   Mean   :0.7    Mean   :0.5857   Mean   :13.37  
##  3rd Qu.:15.00   3rd Qu.:0.9    3rd Qu.:0.6000   3rd Qu.:21.87  
##  Max.   :20.30   Max.   :1.1    Max.   :3.5100   Max.   :34.21  
##  Net_Profit_Margin
##  Min.   : 2.6     
##  1st Qu.:11.2     
##  Median :16.1     
##  Mean   :15.7     
##  3rd Qu.:21.1     
##  Max.   :25.5     
```

These organizations have market capitalizations that range widely, from 0.41 to 199.47. With a median market cap of 48.19 and an average of 57.65, the distribution appears to be slightly positively biased. With a market capitalization of 68.44, Entity 1 has the largest, and Entity 6 has the lowest, at 16.90.

A stock's beta, or volatility relative to the market, can be anywhere between 0.18 and 1.11. With Entity 6 having the highest volatility at 1.11, the average beta is 0.5257. With a median beta of 0.46, most entities have a central tendency around this value.

The PE ratios show that there is a wide range between 3.60 and 82.50. With a median of 21.50, the mean PE ratio is 25.46. The significant discrepancy between the mean and median points to the possibility of higher-order outliers affecting the mean.

A company's profitability as a percentage of shareholder equity, or ROE, can range from 3.9% to 62.9%. With a mean ROE of 25.8%, the average return is large. The distribution looks dispersed, though, with Entity 4 having the highest ROE (27.3).

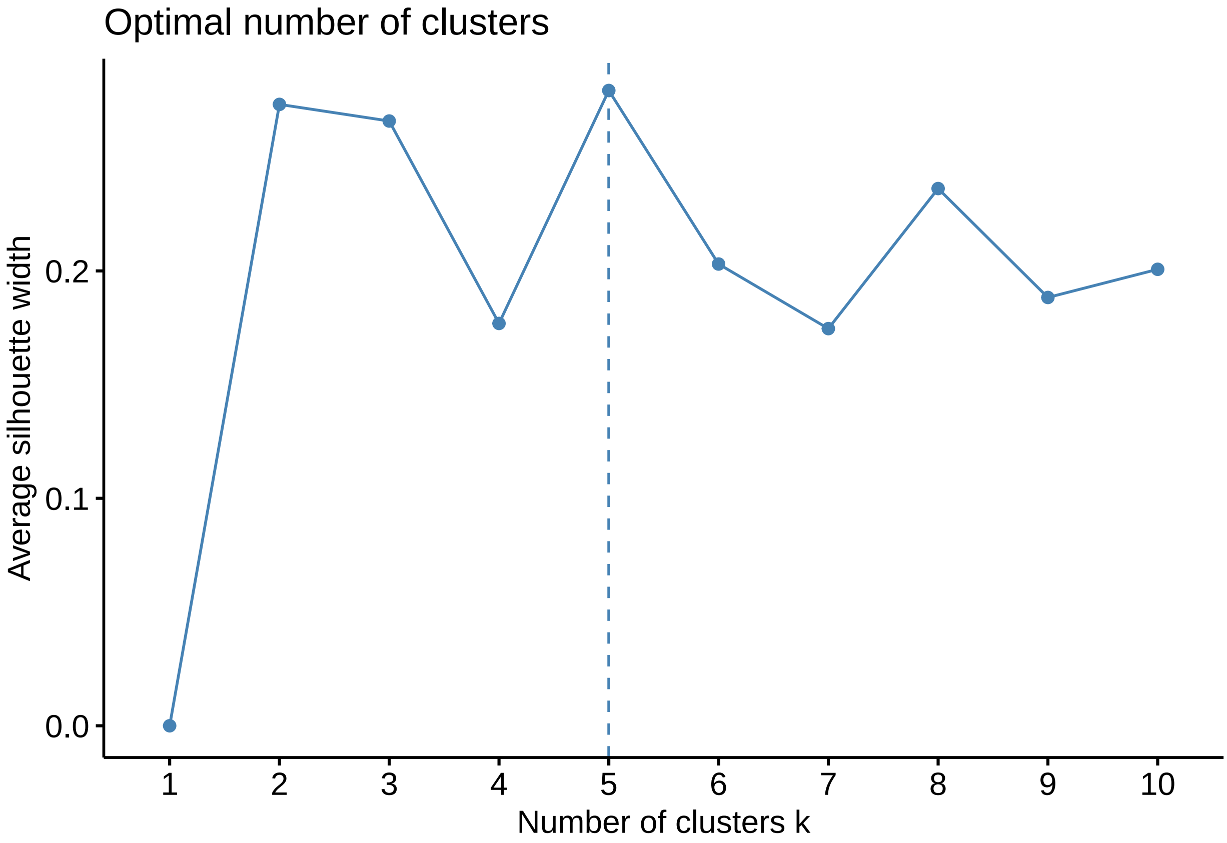
Growth rates range from -3.17% to 34.21%, indicating varied revenue trajectories. Entity 5 experiences the highest growth at 26.81%.

ROE varies from 3.9% to 62.9%, indicating diverse profitability. Entity 4 stands out with the highest ROE at 27.4, showcasing effective use of equity.

Net profit margin ranges from 2.6% to 25.5%, representing profitability. Entity 1 leads with the highest margin at 16.1%.

The information provides a thorough understanding of the financial dynamics of these companies, including market value, risk exposure, valuation, profitability, effective use of assets, and potential for growth. In order to understand the complex financial profiles of each organization, interpretation should take into consideration both central tendencies and the distribution of values.

```
pharmaceutical2 <- scale(pharmaceutical1)
row.names(pharmaceutical2) <- pharmaceutical[,1]
distance <- get_dist(pharmaceutical2)
corr <- cor(pharmaceutical2)
fviz_nbclust(pharmaceutical2, kmeans, method = "silhouette")
```



The cluster k=5 is the highest cluster with having average of silhouette width more than 0.2

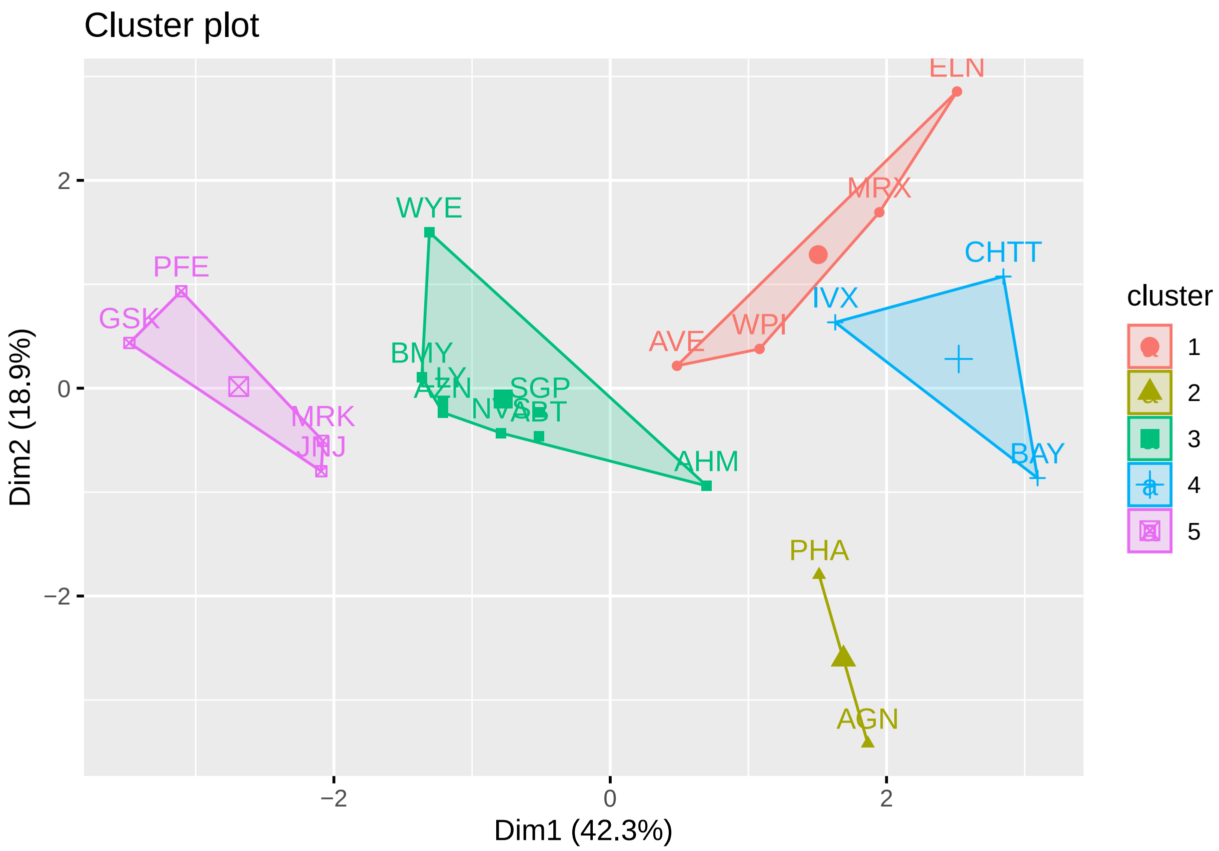
```
set.seed(1)
k5 <- kmeans(pharmaceutical2, centers = 5, nstart = 25) # k = 5 i.e number of restarts =25
k5$centers
```

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

```
k5$size
```

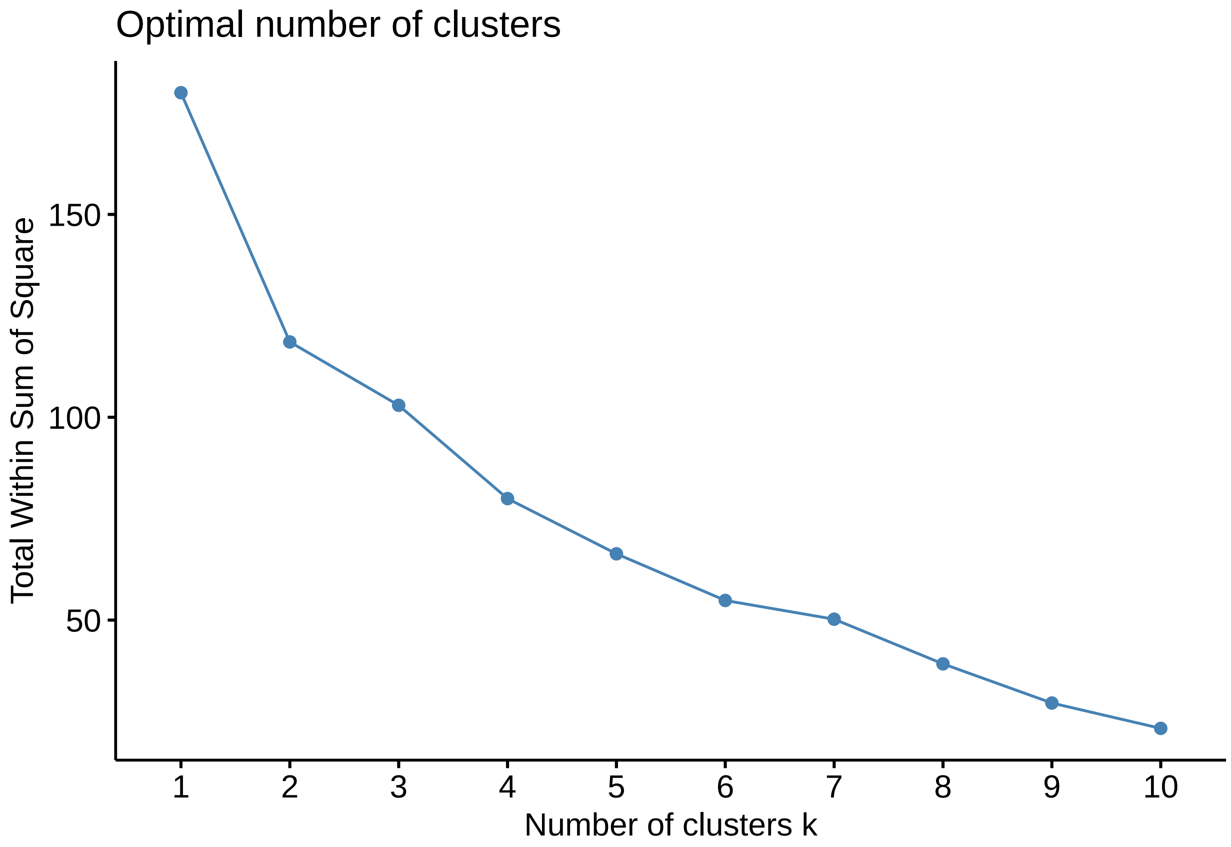
```
## [1] 4 2 8 3 4
```

```
fviz_cluster(k5,data = pharmaceutical2)
```



There are five different clusters where cluster 5 is below the average for both dim1 and dim2 .Cluster 3,4 is almost same as cluster5 but former(3,4) has more in dim2 than cluster5.Cluster 2 is more than average for dim1 but less than average for dim2,Interestingly Cluster 1 has both dim1 and dim2 values higher than average.

```
fviz_nbclust(pharmaceutical2, kmeans, method = "wss")
```



Observing the chart, we can say that the elbow point 4 offers the optimal value for k. WSS will decrease with increasing k values, but we must weigh the trade-off between over fitting—a model that fits both noise and signal—and bias in the model. In this particular instance,The elbow point offers that middle ground where WSS can be reduced without going below reduces significantly more slowly when k = 4. Stated otherwise, k=4 yields the optimal value for bias and over fitting.

```
set.seed(35)
k51 = kcca(pharmaceutical2, k=5, kccaFamily("kmedians"))
k51
```

```
## kcca object of family 'kmedians'
##
## call:
## kcca(x = pharmaceutical2, k = 5, family = kccaFamily("kmedians"))
##
## cluster sizes:
##
## 1 2 3 4 5
## 3 1 4 7 6
```

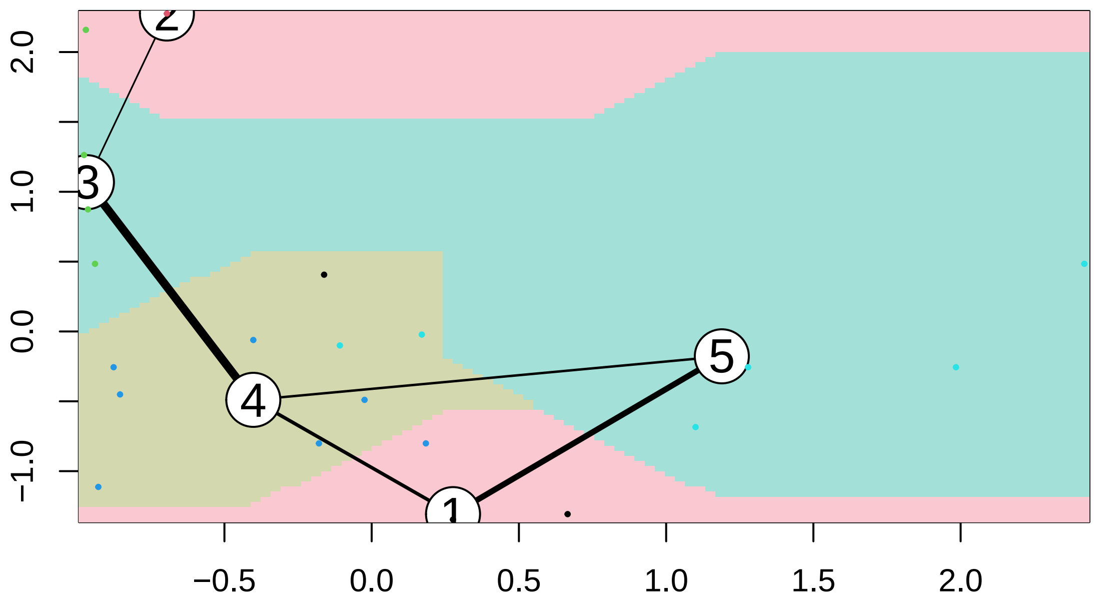
set.seed is used to get the same cluster for same code if we run and k medians is used to determine median rather than means ,here Cluster 4 is the largest, with 7 observations, followed by Cluster 5 with 6observations.Cluster 2 is the smallest, with only 1 observation.

```
clusters_index <- predict(k51)
dist(k51@centers)
```

```
##          1        2        3        4
## 2 5.732015
## 3 4.295187 3.676156
## 4 2.639505 3.726363 2.666800
## 5 2.652823 5.807428 4.407195 3.099168
```

Distance function is representing the pairwise distances between cluster centers.if the distance is smaller for cluster then it indicate they are closer,Here distance between cluster 2 and cluster4 is approx 3.12

```
image(k51)
points(pharmaceutical2, col=clusters_index, pch=19, cex=0.3)
```

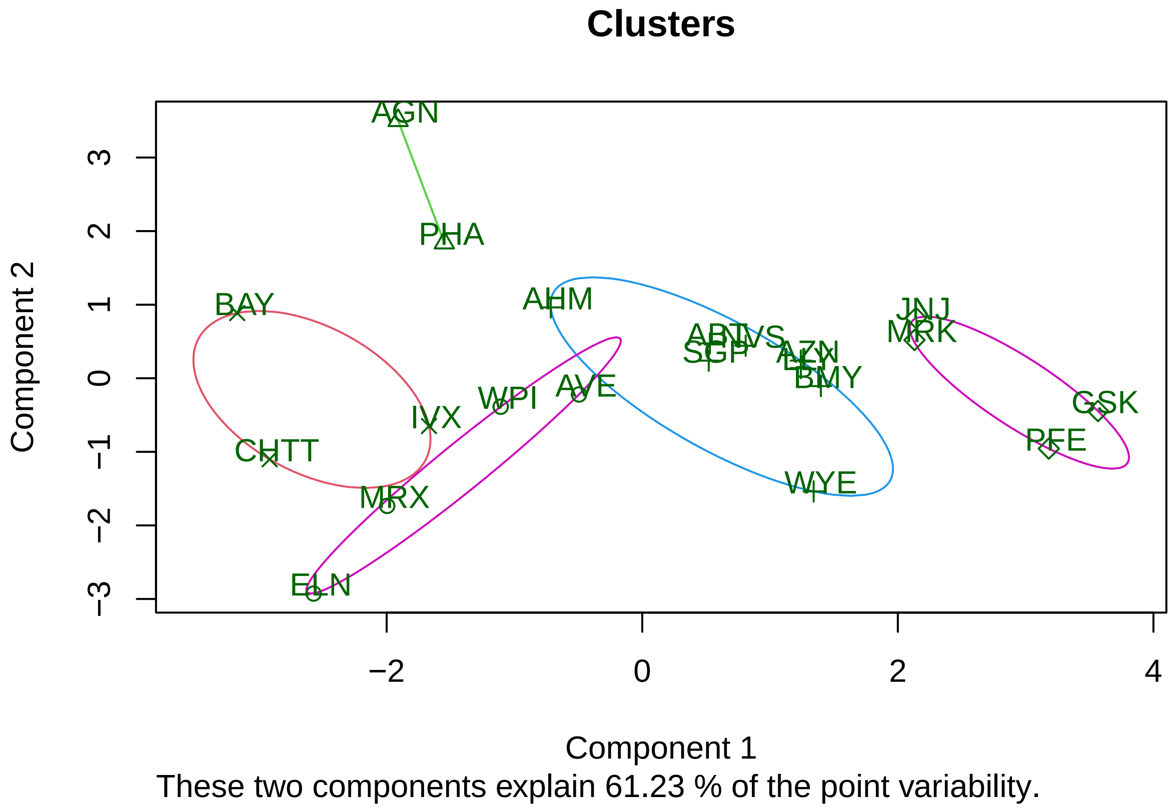


```
pharmaceutical1 %>% mutate(Cluster = k5$cluster) %>% group_by(Cluster) %>%
summarise_all("mean")
```

```
## # A tibble: 5 x 10
##   Cluster Market_Cap  Beta PE_Ratio   ROE   ROA Asset_Turnover Leverage
##     <int>      <dbl> <dbl>    <dbl> <dbl> <dbl>          <dbl>    <dbl>
## 1       1       13.1 0.598     17.7  14.6  6.2           0.425    0.635
## 2       2       31.9 0.405     69.5  13.2  5.6           0.75     0.475
## 3       3       55.8 0.414     20.3  28.7 12.7           0.738    0.371
## 4       4       6.64 0.87      24.6  16.5  4.17          0.6      1.65
## 5       5     157.   0.48      22.2  44.4 17.7           0.95     0.22
## # i 2 more variables: Rev_Growth <dbl>, Net_Profit_Margin <dbl>
```

The above results useful for Investors or analysts understanding the average financial characteristics of companies within each cluster. For example, Cluster 5 has the highest average Market Cap, ROE, ROA, and Asset Turnover, suggesting that companies in this cluster may be considered high-performing and have a strong market position. On the other hand, Cluster 4 has a relatively low Market Cap and high Leverage, indicating potential financial risk.

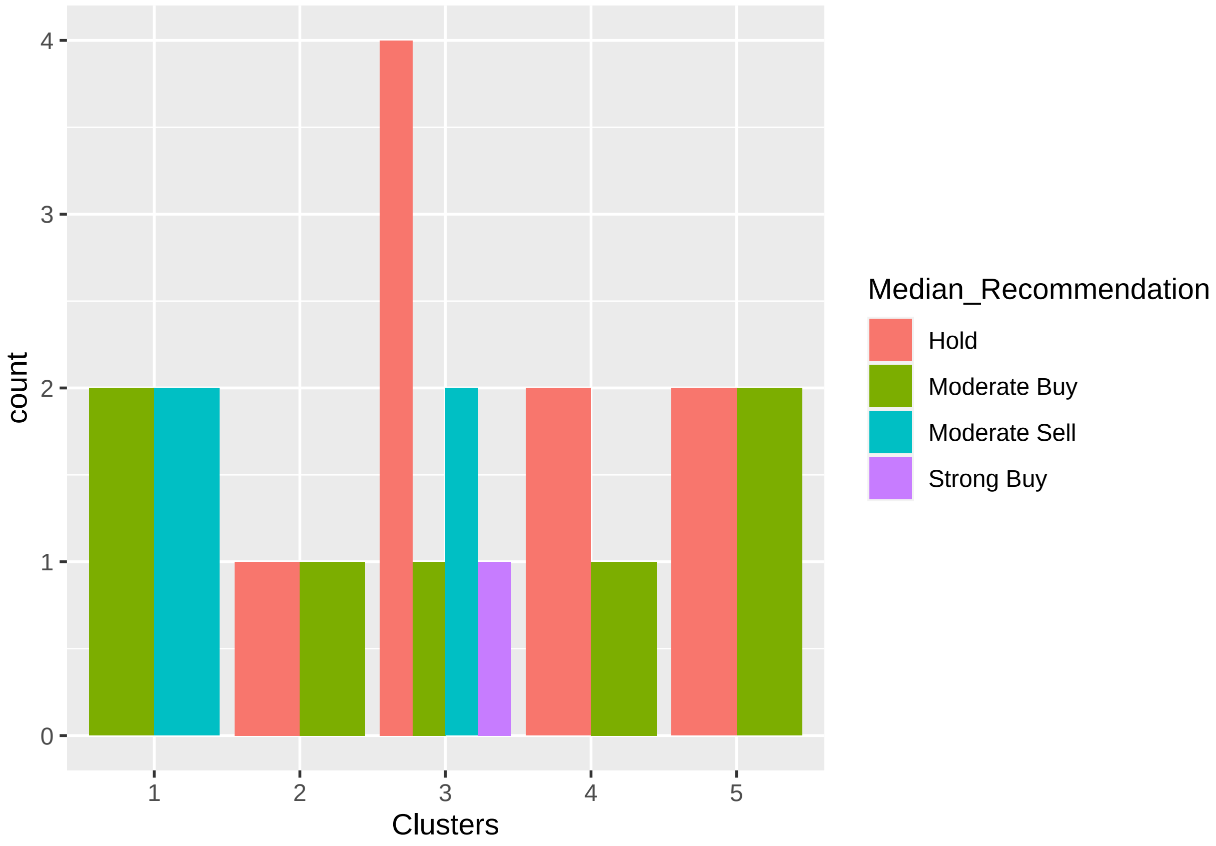
```
clusplot(pharmaceutical2,k5$cluster, main="Clusters",color = TRUE, labels = 3,lines =0)
```



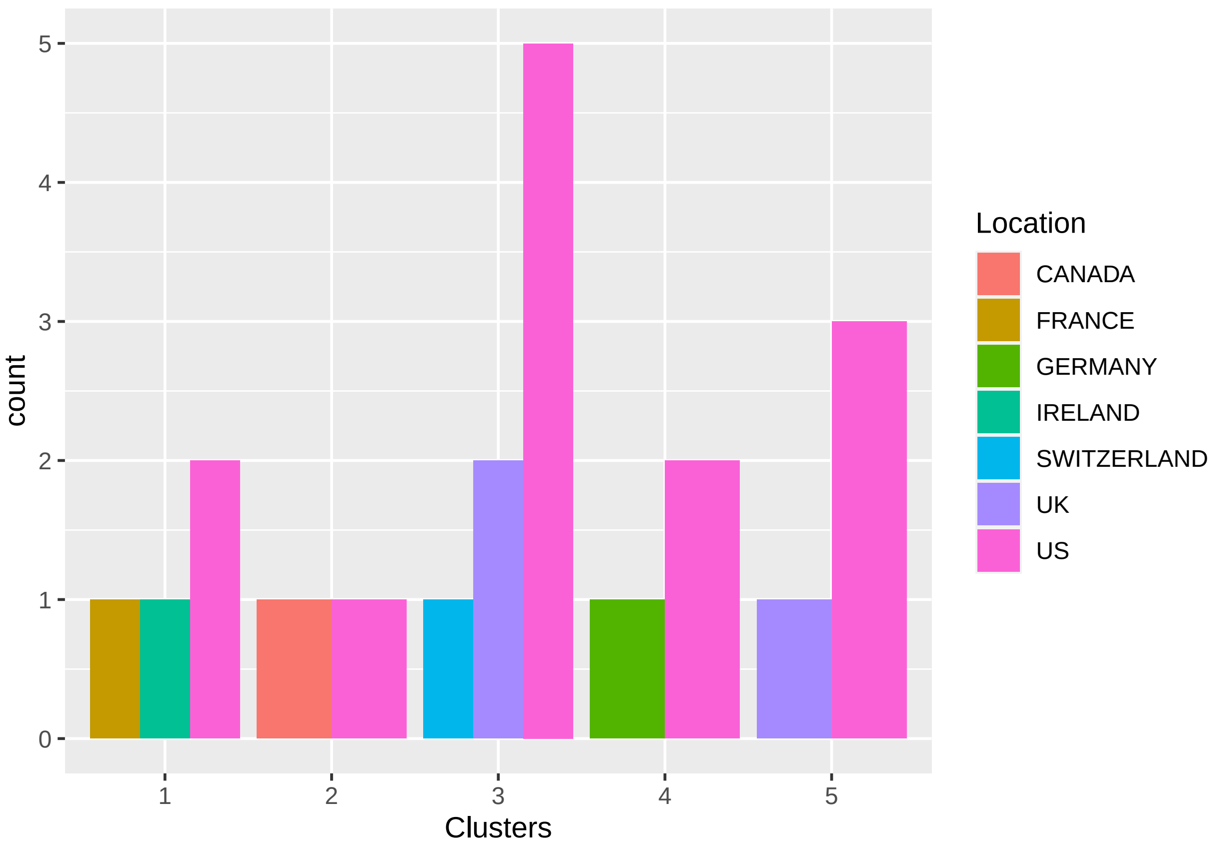
The plot describes the data points in a two-dimensional, with each point representing an observation and its position are determined by the variables in your data set .Points belonging to the same cluster are likely to have the same color, making it easy to visually identify clusters.

Question3

```
pharmaceutical3 <- pharmaceutical[12:14] %>% mutate(Clusters=k5$cluster)
ggplot(pharmaceutical3, mapping = aes(factor(Clusters), fill
=Median_Recommendation))+geom_bar(position='dodge')+labs(x ='Clusters')
```

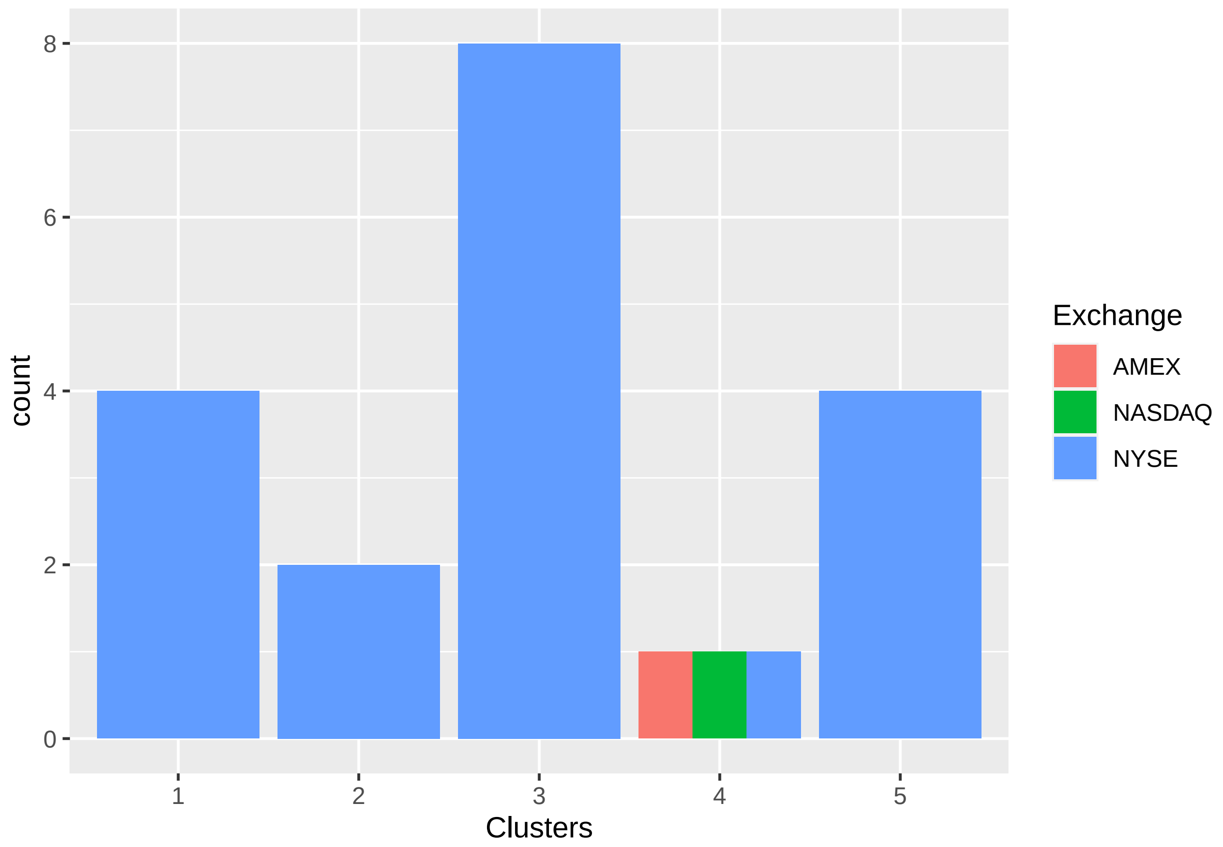


```r
ggplot(pharmaceutical3, mapping = aes(factor(Clusters),fill =
Location))+geom_bar(position = 'dodge')+labs(x ='Clusters')
```



Here the Cluster3 which has median (hold) of greater values compared to all clusters Cluster 1,2 both have moderate buy count,and for cluster1 moderate buy=moderate sell. For 5th cluster and cluster2 moderate buy=hold but cluster5 has more count than 5th cluster. We can observe that only cluster3 has most of the recommendations,Strong buy count whereas other cluster does not have,also same count as cluster2

```
ggplot(pharmaceutical3, mapping = aes(factor(Clusters),fill =
Exchange))+geom_bar(position = 'dodge')+labs(x ='Clusters')
```



The resulting plot is a grouped bar plot where each cluster (Clusters) is represented on the x-axis as a separate group, and the bars within each group are colored based on the Exchange variable. Overall comparing from chart Cluster 3 has highest exchange count,where as 1 and 5 has same cluster. Cluster4 has all types of exchange rates but of smaller count Cluster2 has more count than cluster4 but less than the remaining exchanges types. Each bar graph represents the frequency or count of observations comes under into a specific combination of cluster and exchange.It helps in visualizing the distribution of observations across different clusters and exchanges. Plot above we can say that they are useful when you want to compare the distribution of a categorical variable (Exchange) across different levels of another categorical variable (Clusters). plot gives us insights into how the distribution of exchanges changes with in clusters.

Summary

1)

Reason for Choosing Market_Cap, Beta, PE_Ratio, ROE, ROA, Asset_Turnover, Leverage, Rev_Growth, and Net_Profit_Margin The selected variables Market_Cap, Beta, PE_Ratio, ROE, ROA, Asset_Turnover, Leverage, Rev_Growth, and Net_Profit_Margin are common financial metrics used to evaluate and compare the performance of companies. These variables collectively provide a comprehensive overview of a firm's financial health, profitability, and efficiency.

K-means has been selected over DBSCAN because, K-means is often used in exploratory data analysis to identify patterns and groupings within the data, K-means clustering can provide insights into the financial profiles of pharmaceutical firms. It may reveal groups of firms with similar financial characteristics, aiding in strategic decision-making or investment analysis, easy to interpret, and DBSCAN is effective for datasets with dense regions.

Market_Cap: Range will be from 0.41 to 199.47. Indicates the overall size and valuation of the pharmaceutical firms.

2. Beta: Ranges from 0.18 to 1.11. Measures the sensitivity of a firm's returns to market fluctuations.

3. PE_Ratio: Ranges from 3.6 to 82.5. Represents the valuation of a firm's stock relative to its earnings.

4. ROE: Ranges from 3.9 to 62.9. Indicates how effectively a firm utilizes shareholder equity to generate profit.

5. ROA: Ranges from 0.3 to 1.1. Measures a firm's ability to generate profit from its assets.

6. Asset_Turnover: Ranges from 0.5 to 1.1. Represents how efficiently a firm utilizes its assets to generate revenue.

7. Leverage: Ranges from 0 to 3.51. Reflects the extent to which a firm uses debt to finance its operations.

8. Rev_Growth: Ranges from -3.17 to 34.21. Indicates the percentage change in revenue over a specific period.

9.Net_Profit_Margin: Ranges from 2.6 to 25.54. Represents the percentage of revenue that turns into profit.

2) Pattern in the clusters with respect to the numerical variables (10 to 12)

Cluster 1: It has moderate buy and moderate sell,also 1st cluster has three locations in which US is the highest and has only one exchange that is NYSE.

Cluster 2:

Median Recommendation for the Cluster 2 has low hold and low buy with locations(US and Canada) that were distributed evenly .Surprisingly it has the same as cluster1 i.e only one exchange that is NYSE.

Cluster 3:

Cluster 3 Median Recommendation is very strong hold and high moderate sell and also clear that it has three locations, US has more numbers, next UK and Switzerland NYSE which is very high in number in case of Exchange rate for the Cluster 3 Cluster 4:

Cluster 4 has Median Recommendation which is strong hold and low buy. and has two locations in which US is high when compared to Germany. Exchanges(AMEX, NASDAQ, NYSE) for the Cluster 4 and all of them are distributed equally .

Cluster 5:

Median Recommendation for Cluster 5 has high hold and high buy.It also has two locations in which US is in significantly high compared to UK which is too less and only one exchange that is NYSE.

3)appropriate name for each cluster:

For Cluster 1 it can be named as Profitable Ventures

For Cluster 2 it can be named as Risk-Reward Seekers

For Cluster 3 it can be named as Stable Giants

For Cluster 4 it can be named as Beta Boosted Enterprises

For Cluster 5 it can be named as Market Dominators

Indications or interpret for each of clusters

Cluster 1: AVE, WPI, MRX, ELN indicates moderate values across Market_Cap, Beta, PE_Ratio, ROE, ROA, Asset_Turnover, Leverage, Rev_Growth, and Net_Profit_Margin.

Cluster 2:PHA, AGN indicates lower Market_Cap, Beta, and PE_Ratio.

Cluster 3: WYE, BMY,AZN, SGP, AHM, LLY,NVS, ABT indicates higher Market_Cap, Beta, PE_Ratio, Rev_Growth, and Net_Profit_Margin compared to other clusters.

Cluster 4:IVX, CHTT, BAY indicates lower Market_Cap and PE_Ratio.

Cluster 5: GSK, PFE, MRK, JNJ indicates higher values across Market_Cap, Beta, PE_Ratio, ROE, ROA, Asset_Turnover, Rev_Growth, and Net_Profit_Margin.