

STOCK MARKET ANALYSIS – NIFTY-50 INDEX NSE INDIA-

Group 7

Group Members: 1. Aman Maheshwari 2. Saloni Bhutada

Introduction

Our Project focuses on Analysis of Stocks of Various Indian Companies where we have performed, probability, clustering, text mining & time series analysis.

We have considered two Datasets for our analysis. **Dataset 1** -> NIFTY-50 Stock Market Data (2000-2021) (Probability, Clustering & Time Series Analysis) This dataset consists of data (15 variables) of fifty companies listed in Nifty-50 Index from NSE India. Link -> <https://www.kaggle.com/rohanrao/nifty50-stock-market-data>

Dataset 2 -> Daily News for Stock Market Prediction (Text Analysis) Consists of 8 Years of daily news headlines to predict the Stock Market Movement. Link -> <https://www.kaggle.com/aaron7sun/stocknews>

Loading all the required libraries:

```
#Loading Libraries
library(stringr)
library(dplyr)
library(readxl)
library(tidyr)
library(magrittr)
library(lubridate)
library(lemon)
library(knitr)
library(ggplot2)
library(gridExtra)
library(libr)
library(lessR)
library(sf)
library(tidyverse)
library(corrplot)
library(forecast)
library(tseries)
library(TSA)
library(tibble)
library(TTR)
library(dygraphs)
library(assertthat)
library(xts)
library(urca)
library(TSstudio)
library(philentropy)
library(fpc)
library(clValid)
library(cluster)
library(factoextra)
library(ggfortify)
library(ggpubr)
library(zoo)
library(xts)
library(quantmod)
library(xml2)
library(wordcloud)
library(tm)
library(sentimentr)
library(readr)
library(zoo)
library()
```

STEP - 1 (DATA CLEANING & WRANGLING)

Loading the Dataset

```
# Loaded two datasets
nifty50_all <- read.csv("NIFTY50_all.csv")
head(nifty50_all)

stockmetadata <- read.csv("stock_metadata.csv")
head(stockmetadata)

News_data <- read.csv("Combined_News_DJIA.csv")
```

Data Cleaning and Wrangling:

```
#getting the summary of data
summary(stockmetadata)

#uniqueness of companies names
unique(nifty50_all$Symbol)
unique(stockmetadata$Symbol)
```

Removing the Duplicates from the tables:

```
#Change names of few companies that renamed their firm
nifty50_all$Symbol[nifty50_all$Symbol=="MUNDRAPORT"] <- "ADANIPOINTS"
nifty50_all$Symbol[nifty50_all$Symbol=="UTIBANK"] <- "AXISBANK"
nifty50_all$Symbol[nifty50_all$Symbol=="BHARTI"] <- "BHARTIARTL"
nifty50_all$Symbol[nifty50_all$Symbol=="BAJAAUTOFV"] <- "BAJFINANCE"
nifty50_all$Symbol[nifty50_all$Symbol=="BAJAAUTOFIN"] <- "BAJFINANCE"
nifty50_all$Symbol[nifty50_all$Symbol=="HEROHONDA"] <- "HEROMOTOCO"
nifty50_all$Symbol[nifty50_all$Symbol=="HINDALCO"] <- "HINDALCO"
nifty50_all$Symbol[nifty50_all$Symbol=="HINDLEVER"] <- "HINDUNILVR"
nifty50_all$Symbol[nifty50_all$Symbol=="INFOSYSTCH"] <- "INFY"
nifty50_all$Symbol[nifty50_all$Symbol=="JSWSTL"] <- "JSWSTEEL"
nifty50_all$Symbol[nifty50_all$Symbol=="KOTAKMAH"] <- "KOTAKBANK"
nifty50_all$Symbol[nifty50_all$Symbol=="ZEETELE"] <- "ZEEL"
nifty50_all$Symbol[nifty50_all$Symbol=="SESAGOA"] <- "VEDL"
nifty50_all$Symbol[nifty50_all$Symbol=="SSLT"] <- "VEDL"
nifty50_all$Symbol[nifty50_all$Symbol=="TISCO"] <- "TATASTEEL"
nifty50_all$Symbol[nifty50_all$Symbol=="TELCO"] <- "TATAMOTORS"
nifty50_all$Symbol[nifty50_all$Symbol=="UNIPHOS"] <- "UPL"
```

Joining the two main table – NIFTY50_ALL & stock_metadata

```
#join the table
complete_stock_data <- left_join(nifty50_all, stockmetadata, by="Symbol")
head(complete_stock_data)
```

Converting the Date column from String to Date format

```
#Convert to date from string
complete_stock_data$Date<- as.Date(complete_stock_data$Date)
typeof(complete_stock_data$Date)

complete_stock_data['Year'] <- as.numeric(format(complete_stock_data$Date, "%Y"))
complete_stock_data['Year-Month'] <- format(as.Date(complete_stock_data$Date), "%Y-%m")
```

Null Value Count Check

```
#Check the NAs
null_count <- sapply(complete_stock_data, function(x) sum(is.na(x)))
null_count
```

Insights: We obtained from the Dataset Trades, Deliverable.Volume and Delivery % has null values before 2011. Since we have considered data for 6 years for our analysis i.e. 2016-2021 we can drop the null values columns.

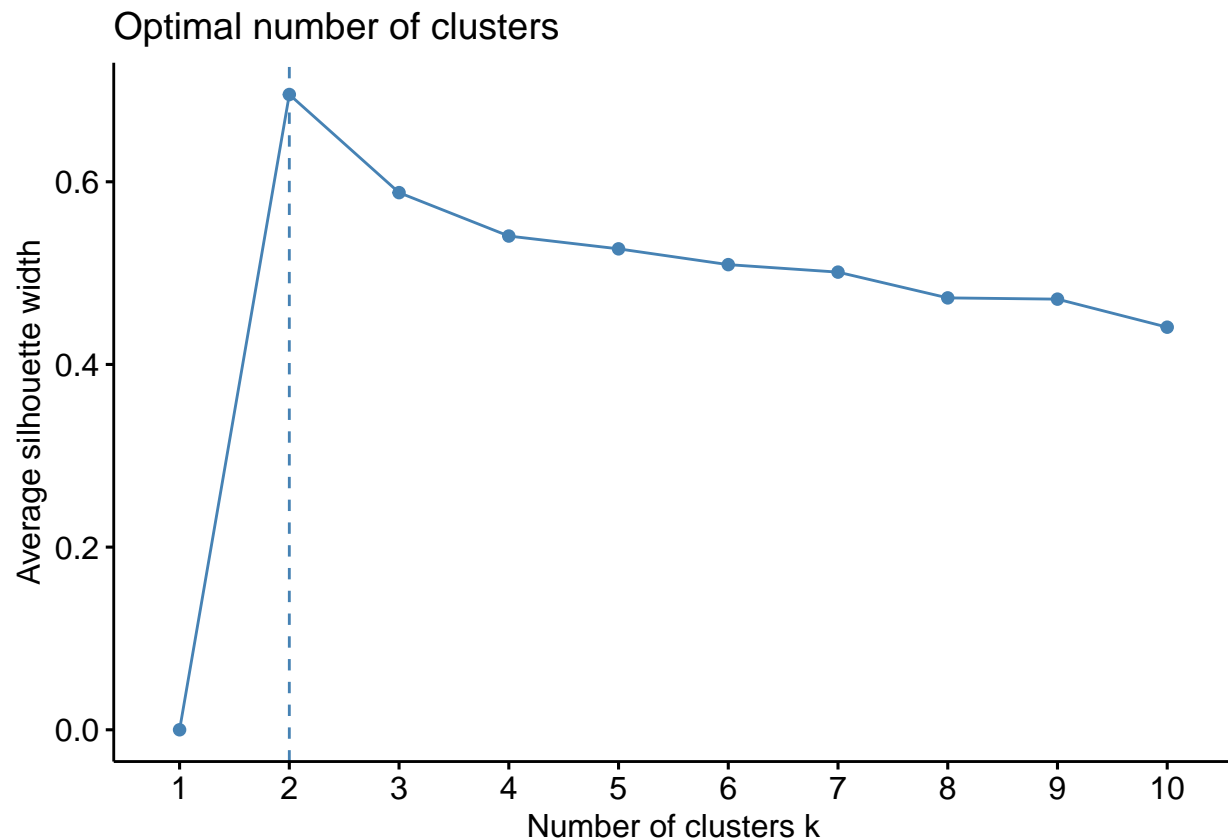
Clustering And Probability:

Q1. What's the overall situation of Nifty 50 stocks' volumes and turnovers before and after the pandemic?

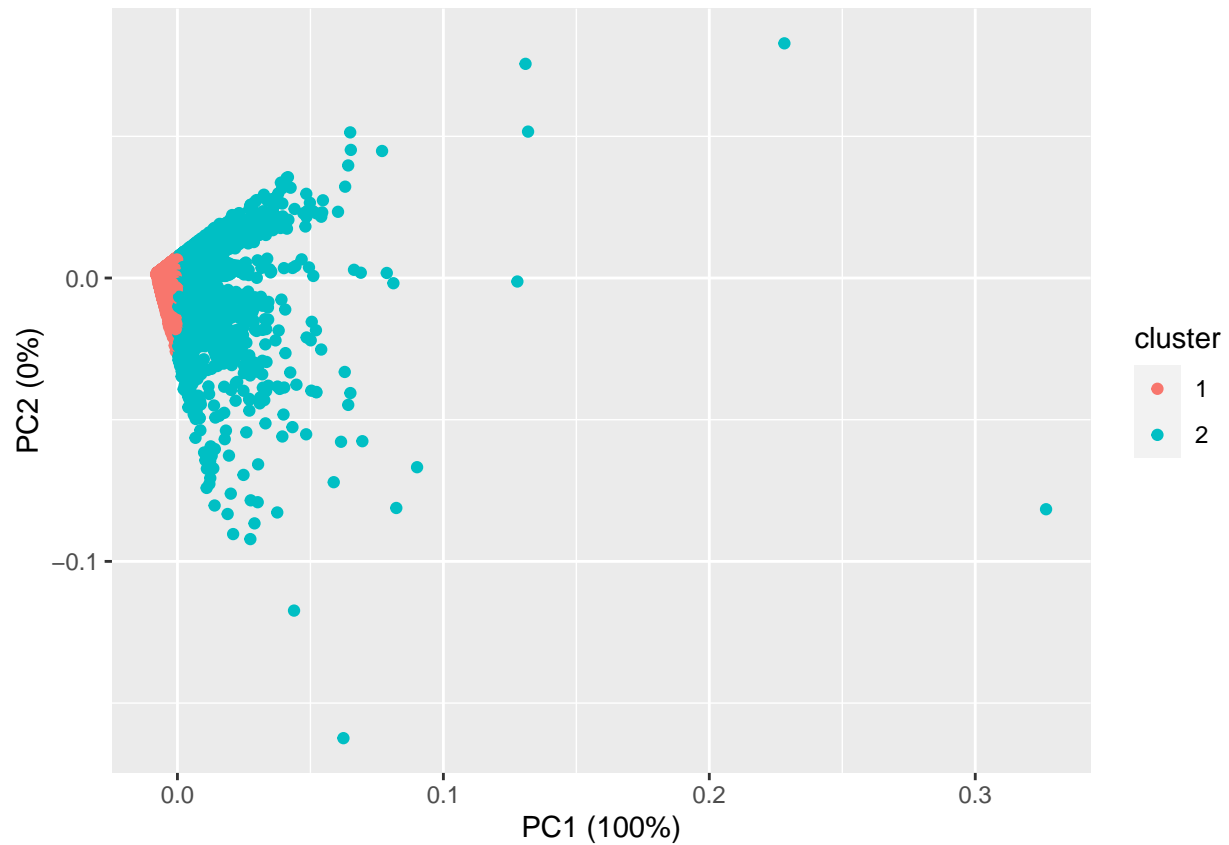
```
#Clustering and Probability
```

```
#All companies 2019
```

```
stockdata_2019<- complete_stock_data %>%  
  filter(Date > "2019-01-01" & Date <"2019-12-31") %>%  
  group_by(Date)  
stockdata_2019_num <- subset(stockdata_2019, select = c(Volume, Turnover))  
stockdata_2019_scale <- scale(stockdata_2019_num )  
fviz_nbclust(stockdata_2019_scale, kmeans, method = "silhouette") #plot shows 2 clusters
```

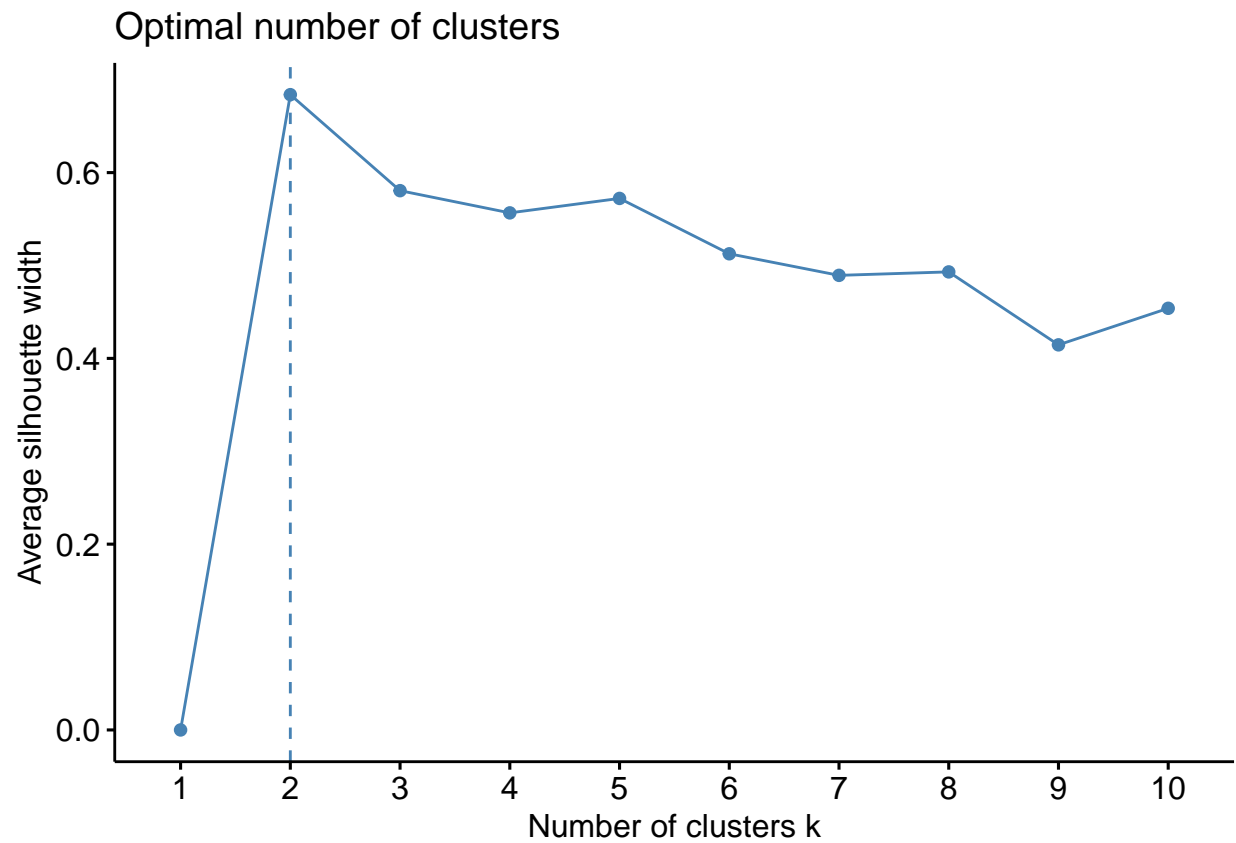


```
clara.res_19 <- clara(stockdata_2019_scale, 2, samples = 50, pamLike = TRUE)  
stockdata_2019_clara <- cbind(stockdata_2019, cluster = clara.res_19$cluster)  
autoplot(cluster::clara(stockdata_2019_num, 2), label=FALSE) #clustering plot
```

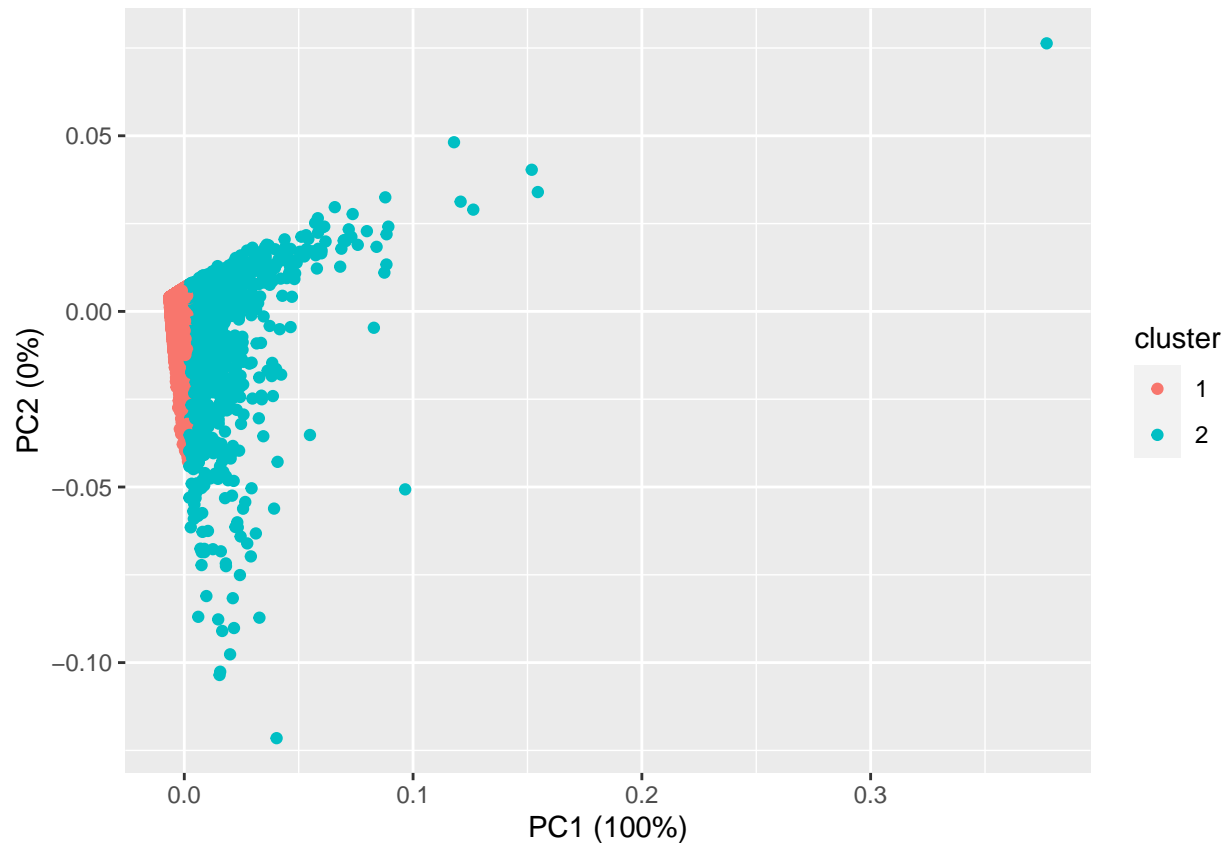


```
#PC1<0 days to cluster 1, PC1>0 days to cluster 2
```

```
#All companies 2020
stockdata_2020<- complete_stock_data %>%
  filter(Date > "2020-01-01" & Date <"2020-12-31") %>%
  group_by(Date)
stockdata_2020_num <- subset(stockdata_2020, select = c(Volume, Turnover))
stockdata_2020_scale <- scale(stockdata_2020_num )
fviz_nbclust(stockdata_2020_scale, kmeans, method = "silhouette") #Plot also shows 2 clusters
```



```
clara.res_20 <- clara(stockdata_2020_scale, 2, samples = 50, pamLike = TRUE)
stockdata_2019_clara <- cbind(stockdata_2020, cluster = clara.res_20$cluster)
autoplot(cluster::clara(stockdata_2020_num, 2), label=FALSE)
```

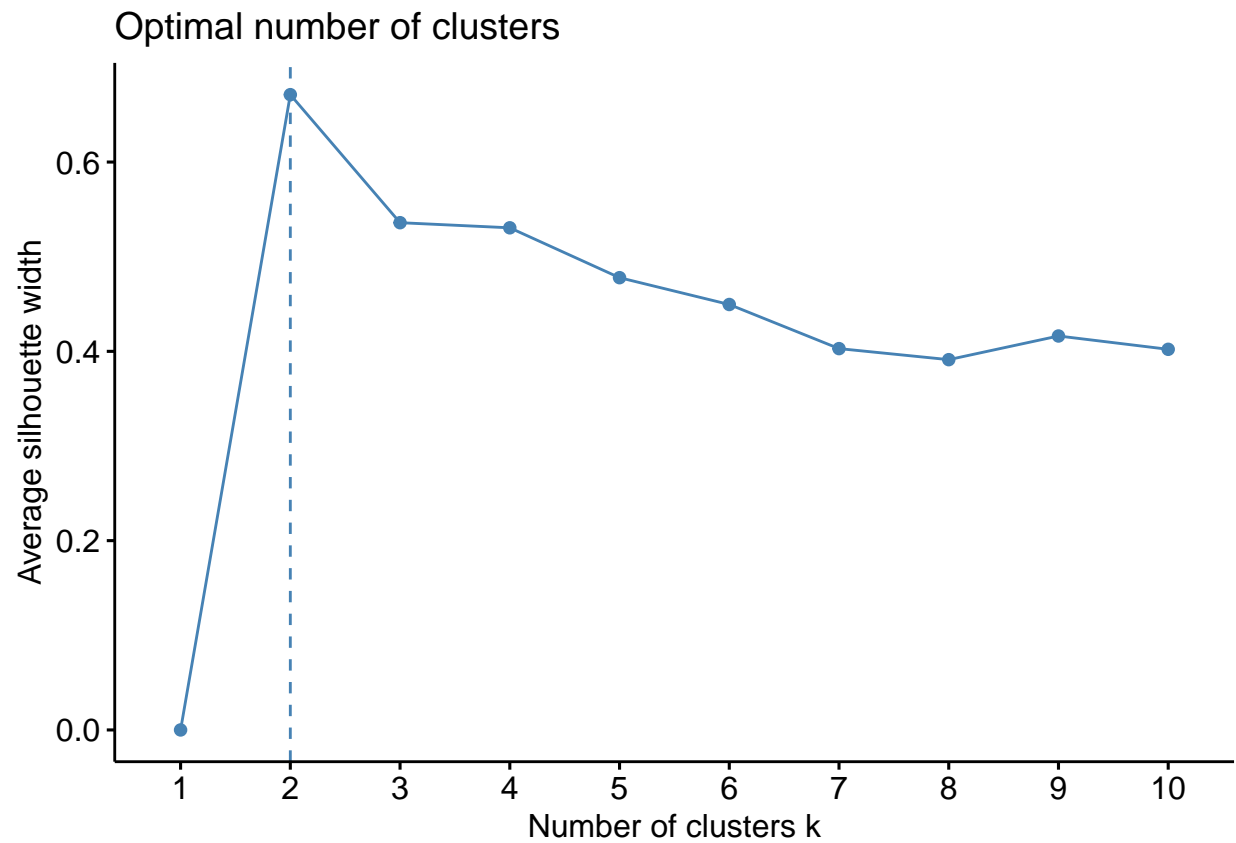


#Some more PC1>0 days now belong to cluster 1, the days are more dense in the interval of $0 < PC1 < 0.05$

Insights: From the clustering plots, we can see that the overall pattern of 2019 is pretty similar to 2020, however, by the measurement of PC1, the separation standard of the clusters shifts to the right. Because of this we can see that the pattern of volumes and turnovers experienced greater changes during 2020, and that is how the pandemic influenced the Nifty50 stock market.

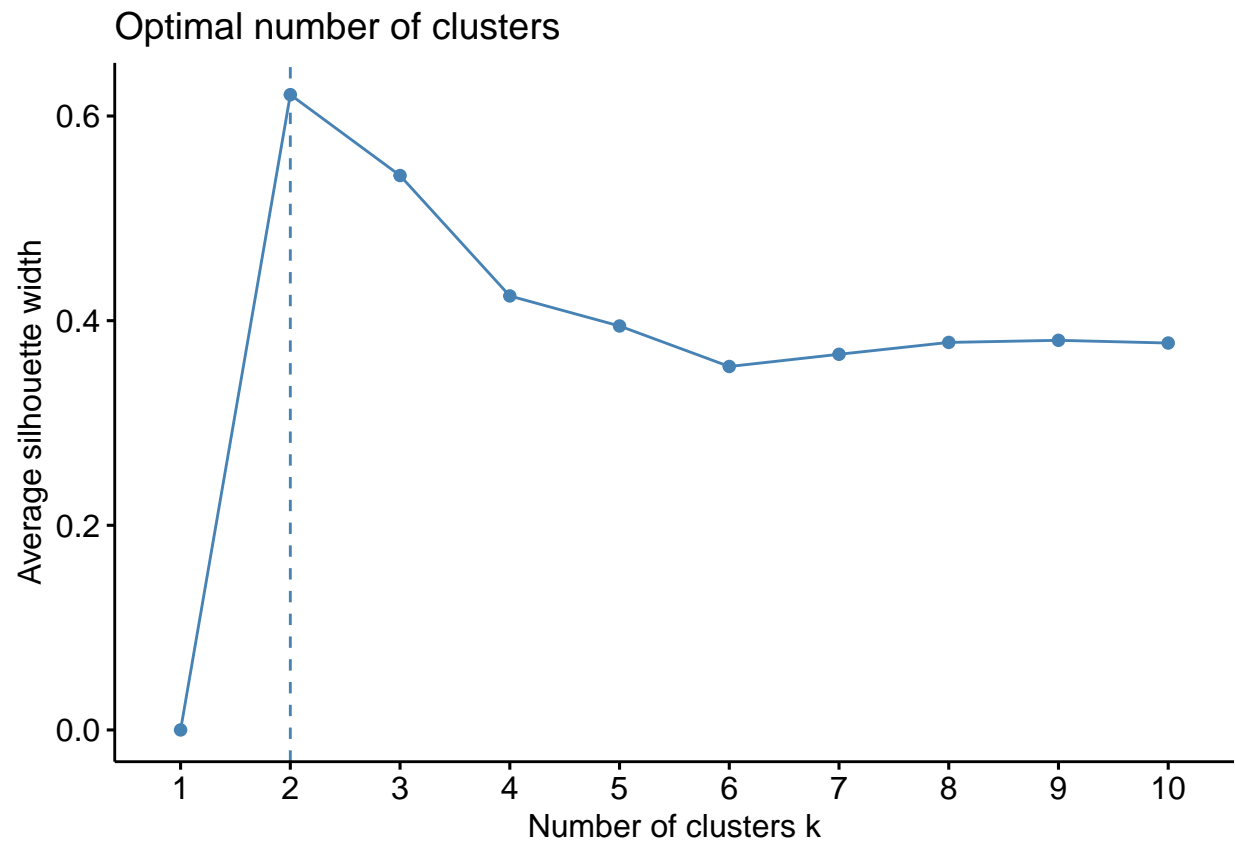
Q2. Business question: Any specific company examples of such changes?

```
#Tatamotors 2019
Tata_2019 <- subset(stockdata_2019, Symbol== "TATAMOTORS")
Tata_2019_num <- subset(Tata_2019, select = c(Volume, Turnover))
Tata_2019_scale <- scale(Tata_2019_num )
fviz_nbclust(Tata_2019_scale, kmeans, method = "silhouette")
```

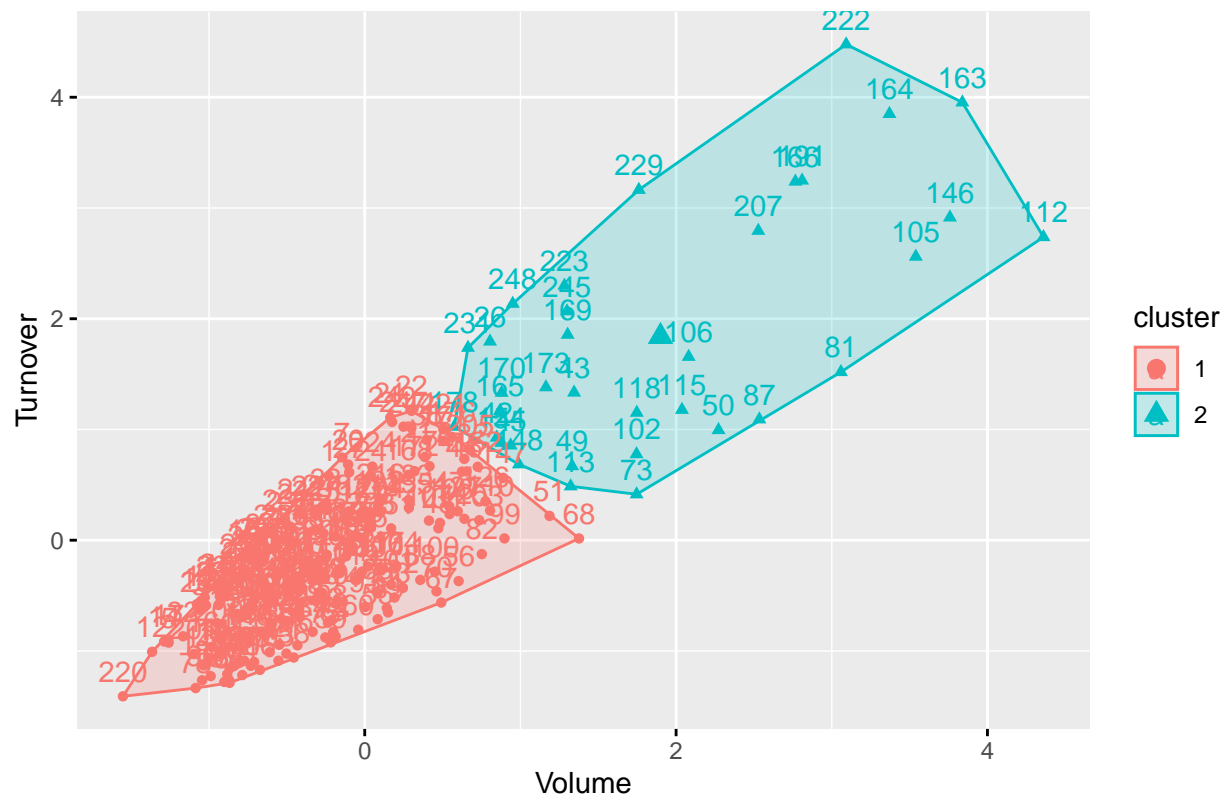
```
Tata_2019_kmeansed <- kmeans(Tata_2019_scale, 2)
fviz_cluster(Tata_2019_kmeansed, data = Tata_2019_scale, title="Tatamotors 2019")
```

```
#Tatamotors 2020
Tata_2020 <- subset(stockdata_2020, Symbol== "TATAMOTORS")
Tata_2020_num <- subset(Tata_2020, select = c(Volume, Turnover))
Tata_2020_scale <- scale(Tata_2020_num )
fviz_nbclust(Tata_2020_scale, kmeans, method = "silhouette")
```



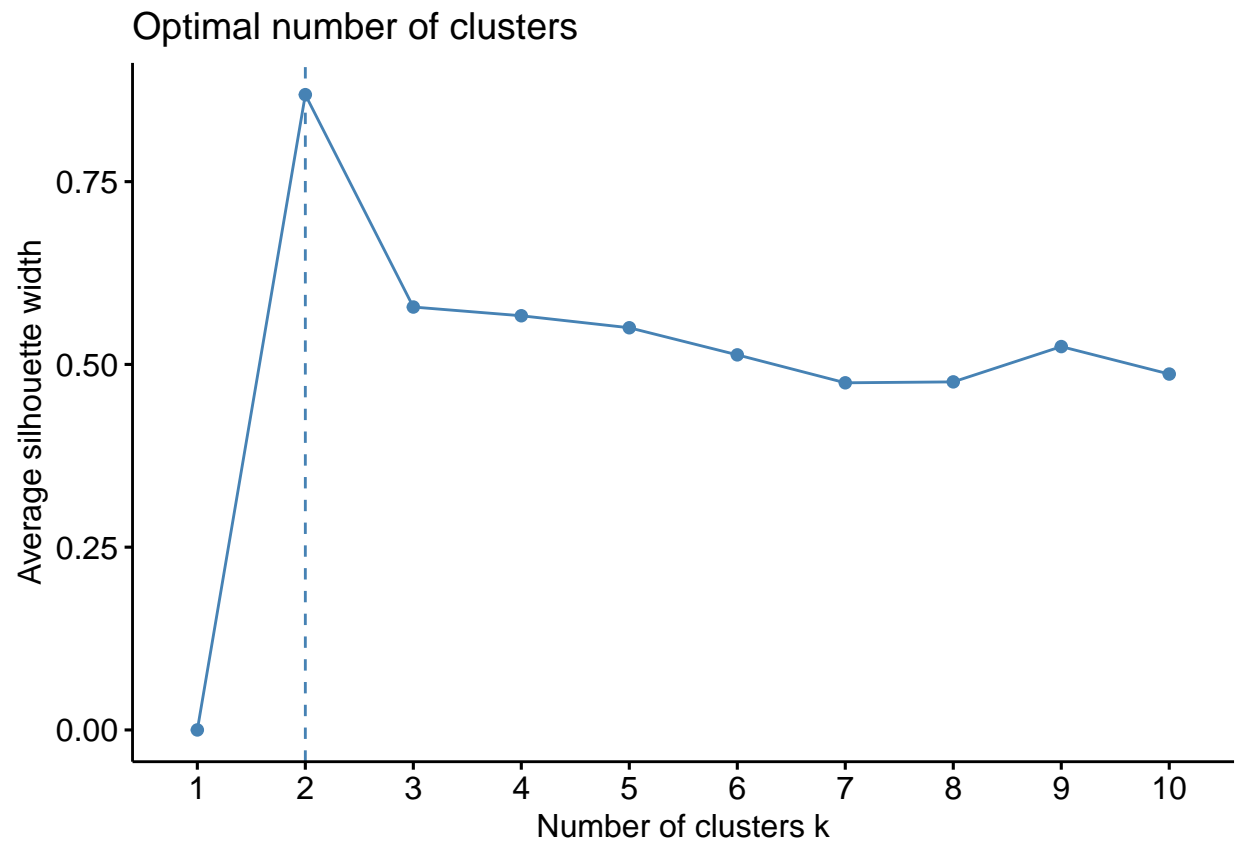
```
Tata_2020_kmeansed <- kmeans(Tata_2020_scale, 2)
fviz_cluster(Tata_2020_kmeansed, data = Tata_2020_scale, title="Tatamotors 2020")
```

Tatamotors 2020



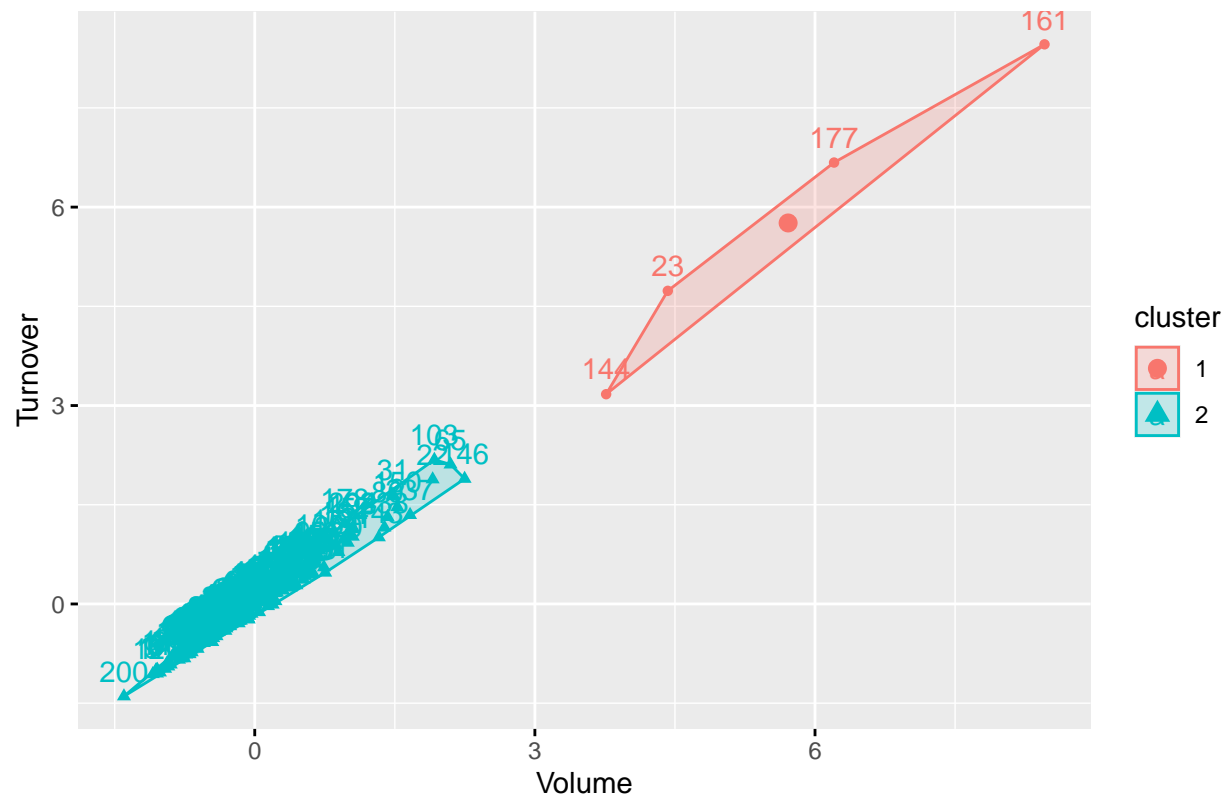
*#The line that separates the clusters shifts to the right
#same pattern as 2019*

```
#Hero 2019
Hero_2019 <- subset(stockdata_2019, Symbol== "HEROMOTOCO")
Hero_2019_num <- subset(Hero_2019, select = c(Volume, Turnover))
Hero_2019_scale <- scale(Hero_2019_num )
fviz_nbclust(Hero_2019_scale, kmeans, method = "silhouette")
```

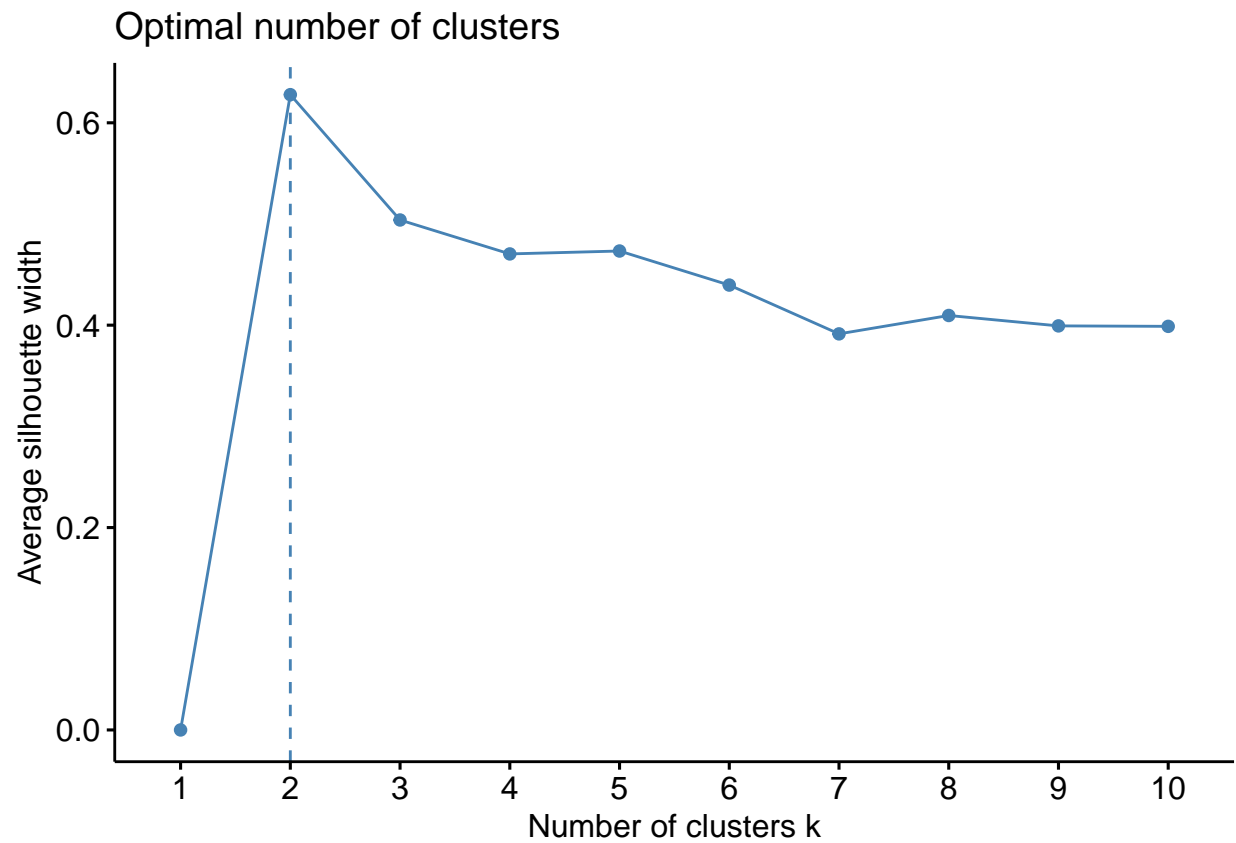


```
Hero_2019_kmeansed <- kmeans(Hero_2019_scale, 2)  
fviz_cluster(Hero_2019_kmeansed, data = Hero_2019_scale, title="Heromotoco 2019")
```

Heromotoco 2019



```
#Hero 2020
Hero_2020 <- subset(stockdata_2020, Symbol== "HEROMOTOCO")
Hero_2020_num <- subset(Hero_2020, select = c(Volume, Turnover))
Hero_2020_scale <- scale(Hero_2020_num )
fviz_nbclust(Hero_2020_scale, kmeans, method = "silhouette")
```

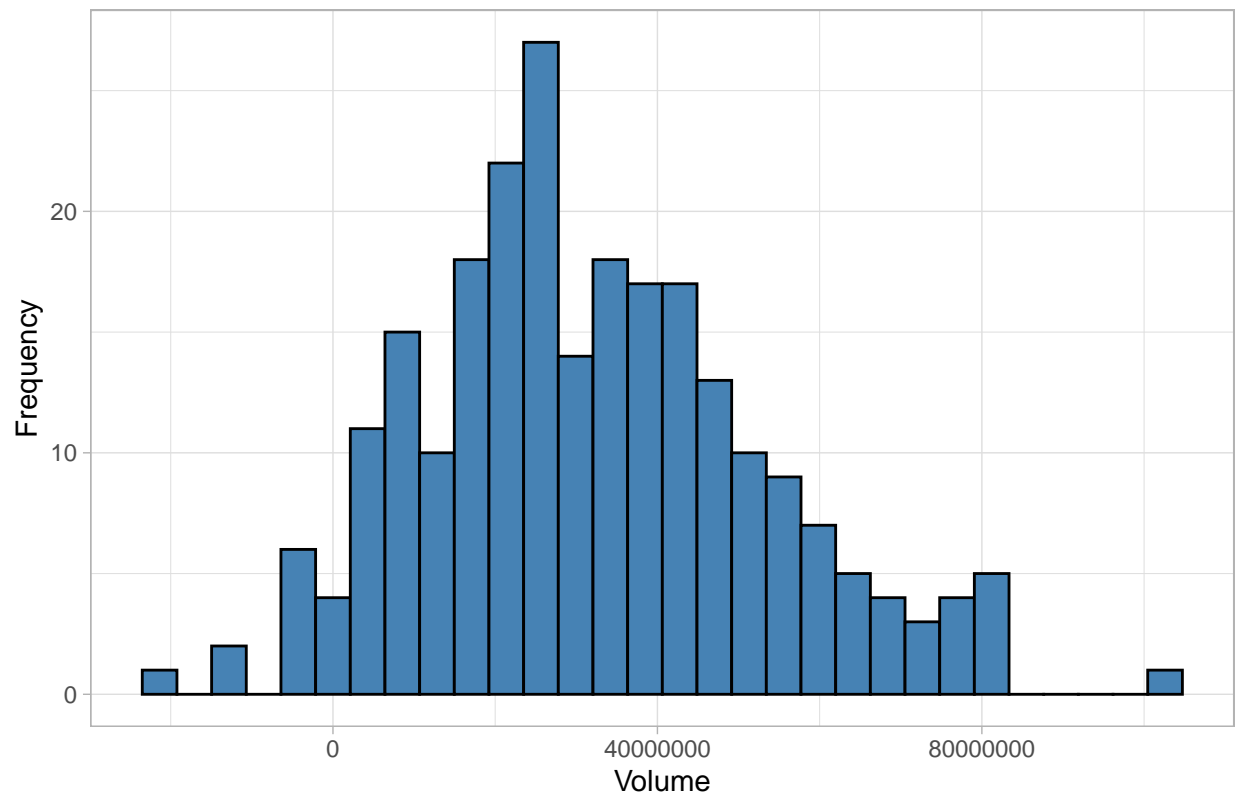


```
Hero_2020_kmeansed <- kmeans(Hero_2020_scale, 2)
fviz_cluster(Hero_2020_kmeansed, data = Hero_2020_scale, title="Heromotoco 2020")
```

Distribution of highly performed company (TATA MOTORS)

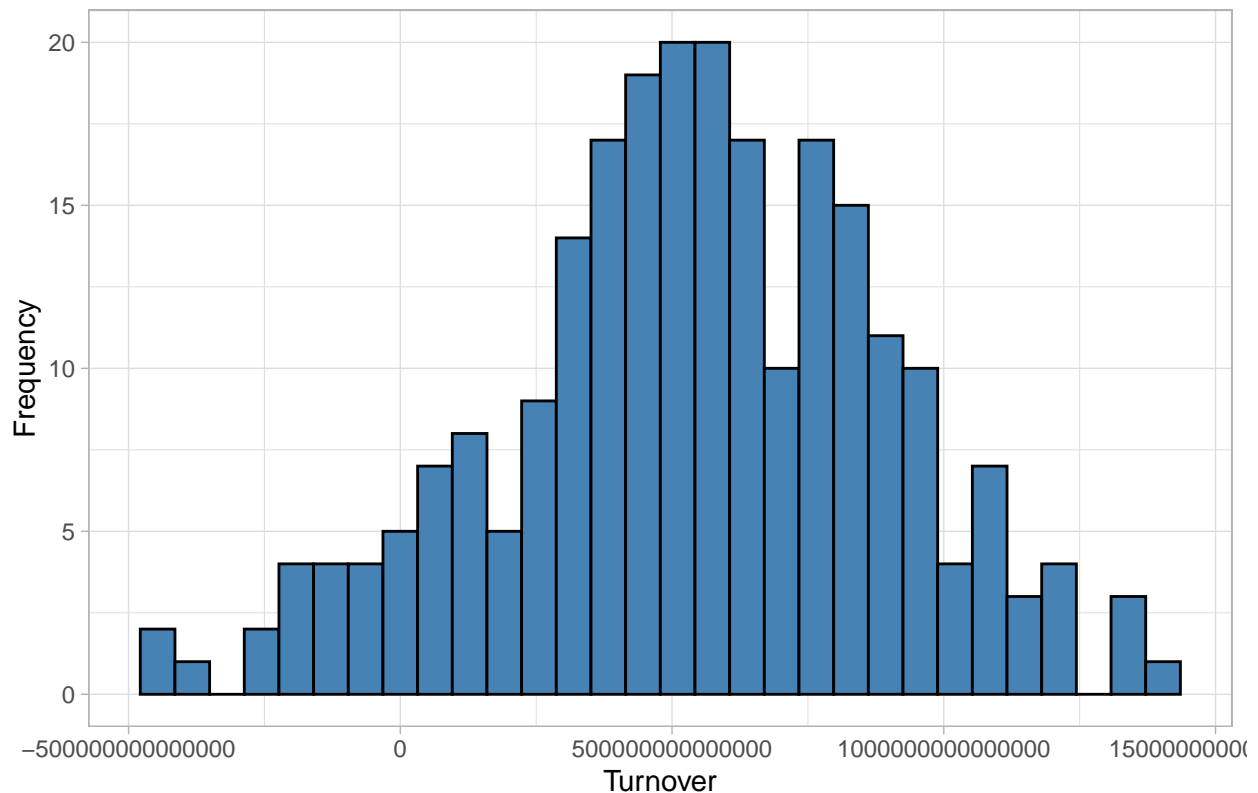
```
#Distribution of Tata in 2019
Volume_norm <- rnorm(n=nrow(Tata_2019_num),mean=mean(Tata_2019_num$Volume),sd=sd(Tata_2019_num$Volume))
ggplot(data.frame(Volume_norm), aes(Volume_norm)) +
  geom_histogram(fill="steelblue", color = 'black') +
  theme_light() +
  xlab('Volume') +
  ylab('Frequency') +
  ggtitle('Distribution Function Plot of Tata 2019 Volume')
```


Distribution Function Plot of Tata 2019 Volume



```
Turnover_norm <- rnorm(n=nrow(Tata_2019_num),mean=mean(Tata_2019_num$Turnover),sd=sd(Tata_2019_num$Turnover))
ggplot(data.frame(Turnover_norm), aes(Turnover_norm)) +
  geom_histogram(fill="steelblue", color = 'black') +
  theme_light() +
  xlab('Turnover') +
  ylab('Frequency') +
  ggtitle('Distribution Function Plot of Tata 2019 Turnover')
```

Distribution Function Plot of Tata 2019 Turnover



Insights: From the above graph we can analyze that the graph follows Normal distribution. We can see that the volume and turnover distribution of tata can be normalize. Despite the outliers, the most possible interval of volumes seems to be 0 to 400000000, and the most possible interval of turnovers is 0 to 7500000000000000. If you would like to invest Tata and purchase the stock shares of it. This is probably the amount of money you will have to prepare.

Correlation Plot:

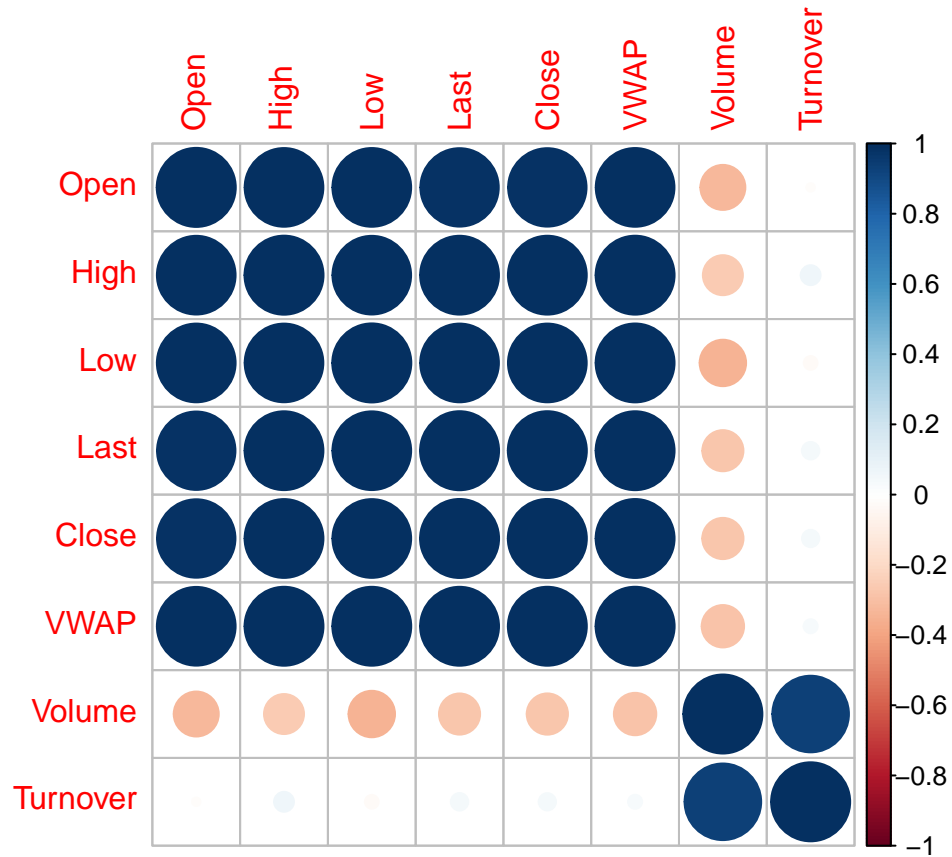
Q4. What are the most correlated elements that we can use to judge the turnovers?

```
#Correlation Plot
```

```
Tata_2019_num_lot <- subset(Tata_2019, select = c(Open, High, Low, Last, Close, VWAP, Volume, Turnover))
```

```
A <- cor(Tata_2019_num_lot)
```

```
corrplot(A)
```



Insights: From the correlation plot, we can see that the most correlated element to turnover is the volume.

Text Mining:

Data Preprocessing:

```
#Merge all text into one column

variables <- c(names(News_data[,c(3:27)]))
News_data$all_news <- apply( News_data[ , variables ] , 1 , paste , collapse = "-" )
colnames(News_data)[28] <- "all_news"
Combined_data <- Corpus(VectorSource(News_data$all_news))

# creating a function to erase special characters
toSpace <- content_transformer(function(x, pattern) gsub(pattern, " ", x))
# put space instead of special characters
Combined_data <- tm_map(Combined_data, toSpace, "/ | @ | \\b |-")
Combined_data <- tm_map(Combined_data, tolower)
# remove punctuation
Combined_data <- tm_map(Combined_data, removePunctuation)
# remove numbers
Combined_data <- tm_map(Combined_data, removeNumbers)
# remove stopwords
update_Stopwords <- c(stopwords('english'))
# Strip Whitespace
Combined_data <- tm_map(Combined_data, stripWhitespace)
```

Building the Document Term Matrix:

```
Dtm_of_text <- TermDocumentMatrix(Combined_data, control = list(stopwords=TRUE,minWordLength = 3))
inspect(Dtm_of_text[266:270,31:40])

# calculate the dimensions of TDM
dim(Dtm_of_text)

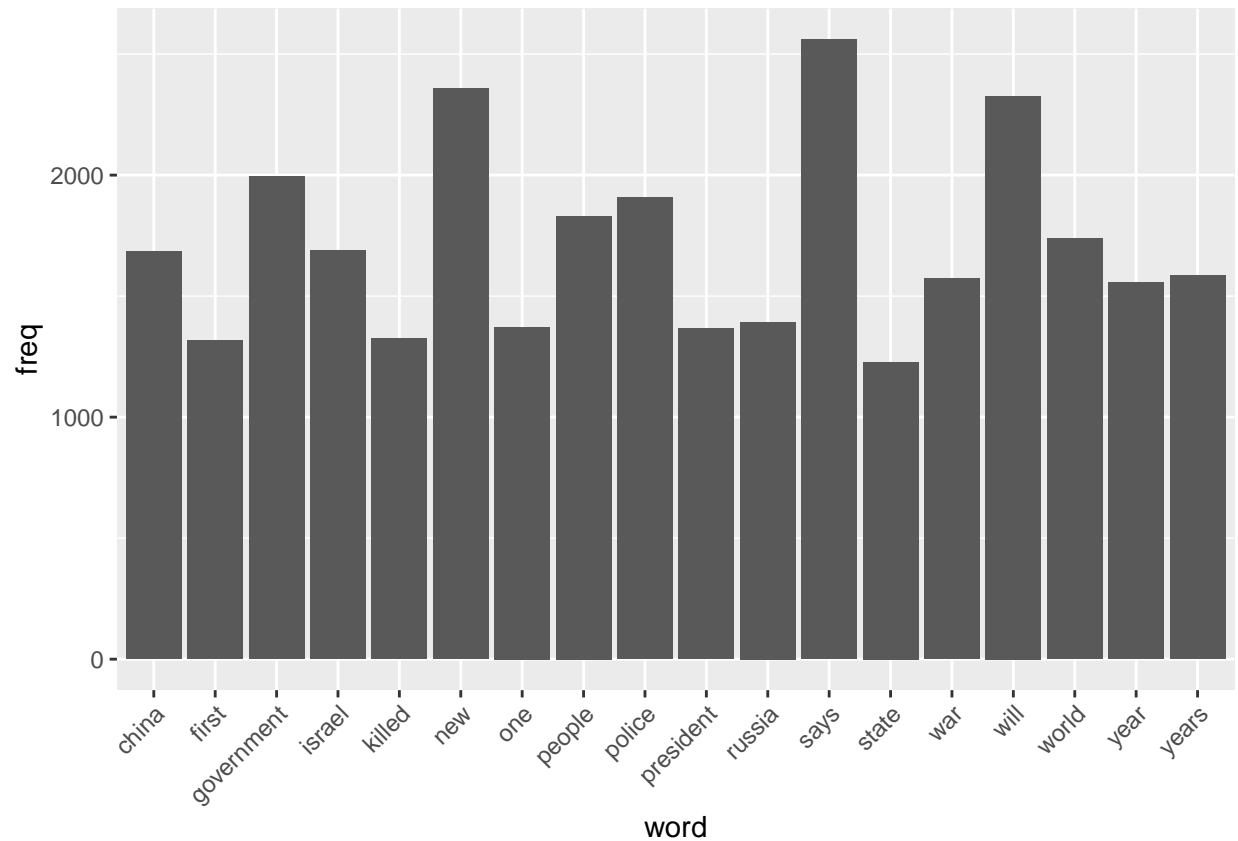
# TDM frequency count
matrix_of_DTM<-as.matrix(Dtm_of_text)
freq <- sort(rowSums(matrix_of_DTM), decreasing=TRUE)
word_freq <- data.frame(word=names(freq), freq=freq)
```

Frequency Plot:

Q5. What are the major words that are used in the news which impacted the stock market movement?

```
# create new dataframe of tdm
freq_words <- data.frame(word=word_freq[,1], freq= word_freq[,2])
# now sorting it and cutting it
freq_words <- subset(freq_words[with(freq_words, order(-freq)),],freq>300)

#plotting frequency of words
plot1 <- ggplot(subset(freq_words, freq>1200), aes(word, freq))
plot1 <- plot1 + geom_bar(stat="identity")
plot1 <- plot1 + theme(axis.text.x=element_text(angle=45, hjust=1))
plot1
```



Q6. What are the words used in the daily news affect the stock market movement based on its sentiment?

Wordcloud:

```
set.seed(2000)
wordcloud(words = freq_words$word, freq = freq_words$freq, min.freq = 1,
          max.words=500, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
```



Insights: Word cloud shows the same frequency plot in a visual manner. We can observe from the above frequency plot and wordcloud that words like government, news, people, says, etc have occurred highest number of times which might have given a significant impact over stock market movement.

Sensitivity Analysis:

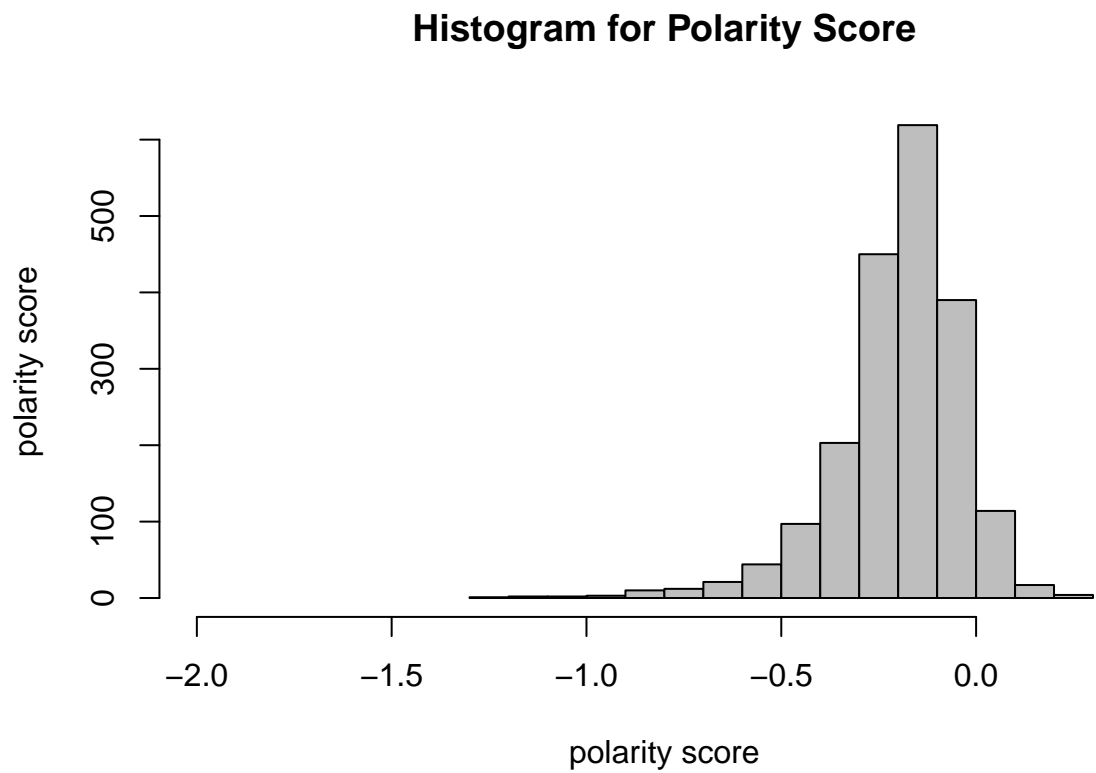
```
#creating a dataframe with the sentiment of the news
sentiment_data = as.data.frame(sentiment(News_data$all_news))
# aggregate the mean sentiment polarity for each element
sentiment_master = data.frame(News_data$Date, aggregate(sentiment_data$sentiment,
                                                         list(sentiment_data$element_id), mean))
colnames(sentiment_master)[1] <- "Date"
colnames(sentiment_master)[3] <- "polarity_score"
sentiment_master <- sentiment_master[,-c(2)]

# converting master sentiment into data frame
#sentiment_master <- as.data.frame(sentiment_master)
# merging it with the data
news_updated_data <- merge(News_data,sentiment_master,by.x = "Date",by.y = "Date")
```

Sensitivity Analysis Plot:

```
# plot histogram for polarity score
hist(news_updated_data$polarity_score,
      main="Histogram for Polarity Score",
      xlab="polarity score",
      ylab = "polarity score",
```

```
border="black",  
col="grey",  
xlim=c(-2.0,0.4)  
)
```



Insights: From the above graph we can conclude that the graph is skewed towards right that means the sentiment is Positive.

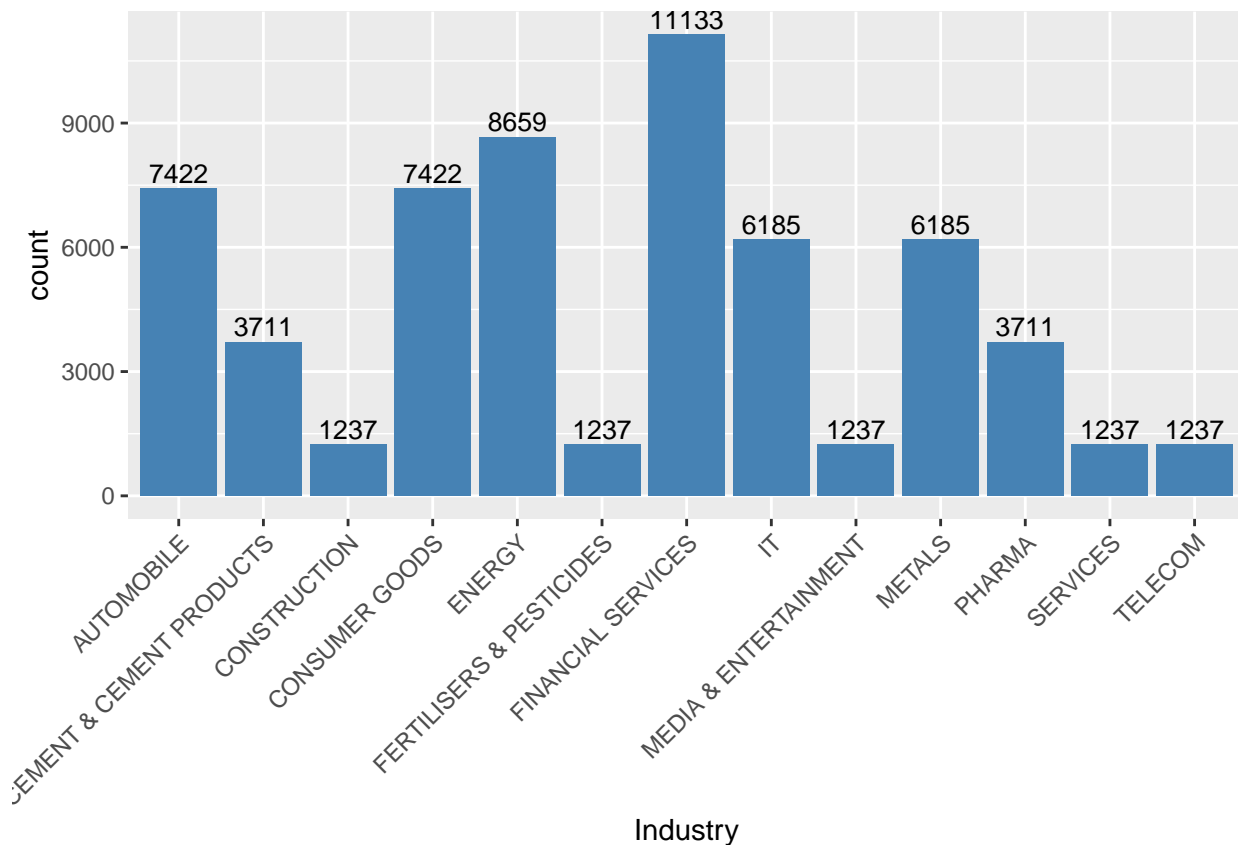
Time Series Analysis:

Q7. Barplot to check number of trades in all the sector for the last 5 years?

```
#Segregating data to 5 years
Fiveyears_stockdata<- complete_stock_data %>%
  filter(Date > "2016-01-01" & Date <"2021-01-01") %>%
  group_by(Industry) %>%
  summarize(count =n())

#Barplot to check number of trades in all the sector

Fiveyears_stockdata_Visual <-
  ggplot(Fiveyears_stockdata, aes(x=Industry, y = count)) +
  geom_bar(stat = "identity", fill="steelblue") +
  geom_text(aes(label=count), vjust=-0.3, size=3.5) +
  theme(axis.text.x=element_text(angle=45,hjust=1,vjust=1))
  theme_minimal() +
  labs(title="Count of Best Restuarants to visit in various countries",
        subtitle="Countrywise distributions")
Fiveyears_stockdata_Visual
```



Insight: From the above barchart we can depict that Financial Services has maximum amount of trades which happen every year. Then there is energy which has quite difference with the high amount of Financial Services, followed by Automobile and Consumer goods have almost similar trades.

Q8. What was the change in close price of the stock over time?


```

#Preparing the data to verify the distributions of each sectors over time.
Fiveyears_stockdata_cycle<- complete_stock_data %>%
  filter(Date > "2016-01-01" & Date <"2021-01-01",
         Industry %in%
           c('FINANCIAL SERVICES','AUTOMOBILE','CONSUMER GOODS', 'ENERGY','IT')) %>%
  group_by(Industry)
Fiveyears_stockdata_cycle

sample_num = 4
assert_that(sample_num < length(unique(Fiveyears_stockdata_cycle$Industry)))

sample_ticker <- as.character(sample(Fiveyears_stockdata_cycle$Industry, sample_num))
sample_ticker <- c(sample_ticker, 'FINANCIAL SERVICES')
candidate_ticker <- unique(sample_ticker)
candidate_ticker <- c('FINANCIAL SERVICES','AUTOMOBILE','CONSUMER GOODS', 'ENERGY','IT')
candidate_num <- length(candidate_ticker)
stock_list <- vector(mode="list", length=candidate_num)
names(stock_list) <- candidate_ticker
i = 1
for (ticker in candidate_ticker){
  stock_list[[i]] <- filter(Fiveyears_stockdata_cycle, Industry == ticker)
  # print(stock_list[[i]])
  i <- i+1
  # print(ticker)
}
str(stock_list)

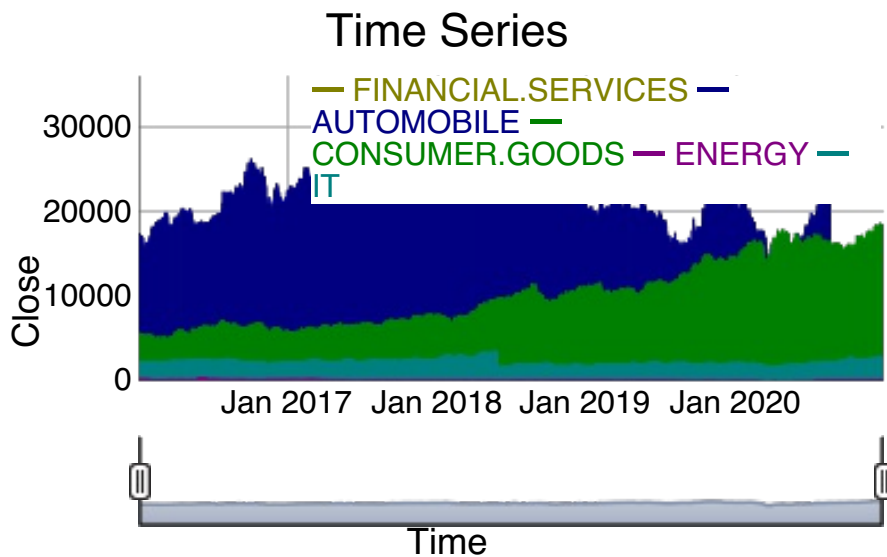
```

```

#Area Plot:
xts_list <- vector(mode="list", length=candidate_num)
ts_list <- vector(mode="list", length=candidate_num)
names(xts_list) = candidate_ticker
names(ts_list) = candidate_ticker

for (ticker in candidate_ticker){
  stock = stock_list[[ticker]]
  xts = xts(stock$Close, order.by=stock$Date)
  attr(xts, 'frequency') <- length(xts)/12
  ts = as.ts(xts, start = c(2016))
  xts_list[[ticker]] <- xts
  ts_list[[ticker]] <- ts
}
xts_table= do.call(cbind, xts_list)
dygraph(xts_table, xlab = "Time", ylab = "Close", main = "Time Series") %>%
  # dySeries(labels.default()) %>%
  # dyOptions(colors = c("red")) %>%
  dyRangeSelector()

```



Insight: Area Distribution Plot was being used to check the area wise distribution of each sector over the past five year Now that we have seen the distribution of each sector, we can further proceed with looking performance over years of each company in their respective sector.

Q9.How are the companies performing in the stock market of some highly recommended sectors?

```
#Dataframe for segrugation to check volume flow of companies in Automobile
stockvolume_AUTOMOBILE <- complete_stock_data %>%
  group_by(Date) %>%
  filter(Date > "2016-01-01" & Date <"2021-01-01", Industry == "AUTOMOBILE") %>%
  select(Year, Date, X.Deliverble, Volume, Company.Name)
stockvolume_AUTOMOBILE
```

```
#Dataframe for segrugation to check volume flow of companies in Financial Services
stockvolume_FS <- complete_stock_data %>%
  group_by(Date) %>%
  filter(Date > "2016-01-01" & Date <"2021-01-01", Industry == "FINANCIAL SERVICES") %>%
  select(Year,Date, X.Deliverble, Volume, Company.Name)
stockvolume_FS
```

```
#Dataframe for segrugation to check volume flow of companies in IT
stockvolume_IT <- complete_stock_data %>%
  group_by(Date) %>%
  filter(Date > "2016-01-01" & Date <"2021-01-01", Industry == "IT") %>%
  select(Year,Date, X.Deliverble, Volume, Company.Name)
```

```
stockvolume_IT
```

```
#Dataframe for segragation to check volume flow of companies in Energy
stockvolume_Energy <- complete_stock_data %>%
  group_by(Date) %>%
  filter(Date > "2016-01-01" & Date <"2021-01-01", Industry == "ENERGY") %>%
  select(Year,Date, X.Deliverble, Volume, Company.Name)
stockvolume_Energy
```

Stacked plot to rectify the performances of the companies in stock market for the last five years:

i. Automobile:

```
#Stacked plot to compare companies stock volume in Automobile sector
Auto<- ggplot(stockvolume_AUTOMOBILE, aes(fill=Company.Name, y= Volume/100000, x=Year)) +
  geom_bar(position="stack", stat="identity") +
  scale_fill_manual(values =
    c("cyan1","cadetblue1","deepskyblue2","deepskyblue3","deepskyblue4", "aquamarine"))
```

Insight: As per the graph results Tata Motors holds the highest share amongst all other companies in Automobile sector.

ii. Financial Services

```
#Stacked plot to compare companies stock volume in Financial Services sector
Fs<- ggplot(stockvolume_FS, aes(fill=Company.Name, y= Volume/100000, x=Year)) +
  geom_bar(position="stack", stat="identity") +
  scale_fill_manual(values =
    c("darkorange","darkorange1","darkorange2",
      "darkorange3","coral1","chocolate1","darkgoldenrod1",
      " darkgoldenrod2","darkgoldenrod3"))
```

Insight: As per the graph results State Bank of India and ICICI Bank are most preferred companies for investment.

iii. IT

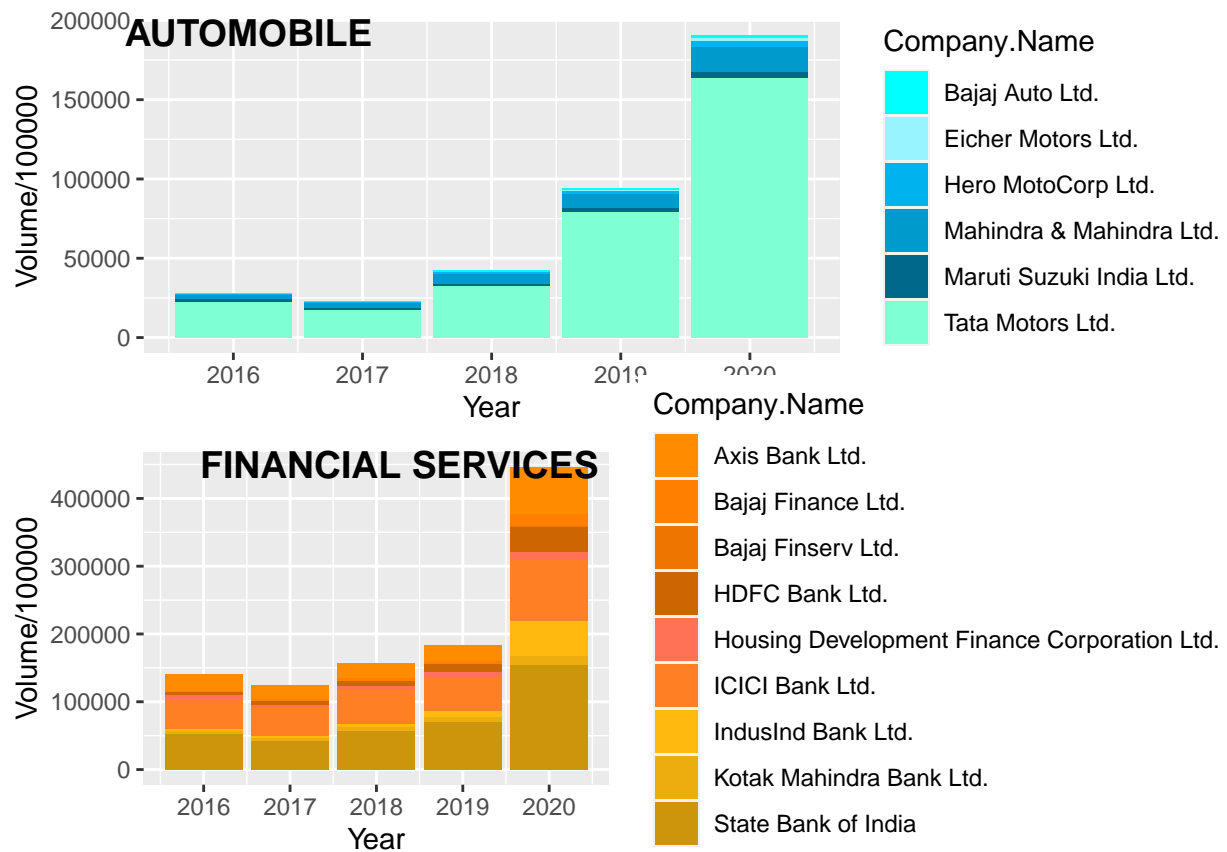
```
#Stacked plot to compare companies stock volume in IT sector
IT<- ggplot(stockvolume_IT, aes(fill=Company.Name, y= Volume/100000, x=Year)) +
  geom_bar(position="stack", stat="identity") +
  scale_fill_manual(values =
    c("darkolivegreen1","darkolivegreen2",
      "darkolivegreen3","aquamarine1","aquamarine2"))
```

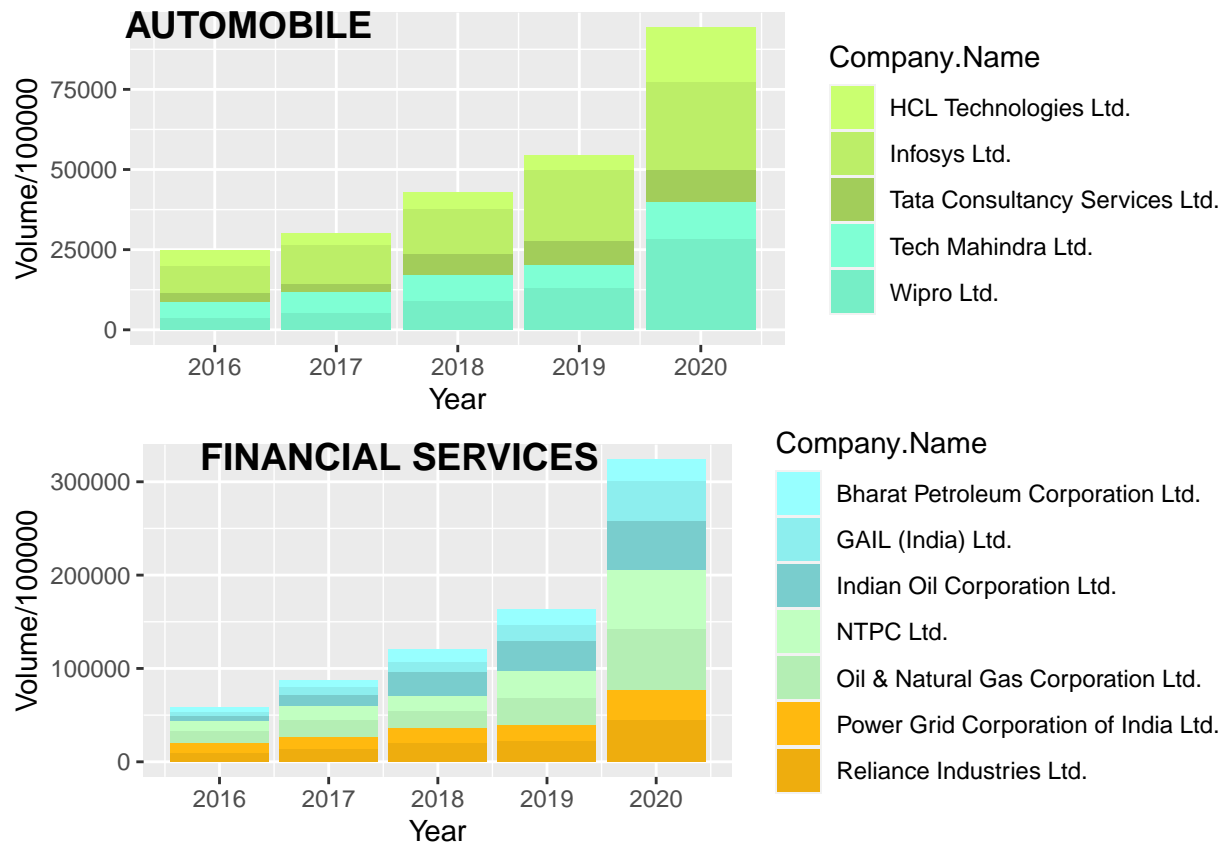
Insight: As per the analysis Infosys holds the highest share amongst all other companies in IT sector.

iv. Energy

```
Energy <- ggplot(stockvolume_Energy, aes(fill=Company.Name, y= Volume/100000, x=Year)) +
  geom_bar(position="stack", stat="identity") +
  scale_fill_manual(values =
    c("darkslategray1","darkslategray2",
      "darkslategray3","darkseagreen1","darkseagreen2","darkgoldenrod1"," darkgoldenrod2"))
```

```
ggarrange(Auto, Fs, IT, Energy,
  labels = c("AUTOMOBILE", "FINANCIAL SERVICES", "IT", "ENERGY"),
  ncol = 1, nrow = 2)
```





Insight: As per the analysis ONGC, NTPC and Indian Oil nearly equals in volume over the years.

Q10. How have the above-mentioned firms done in the last six years in comparison to one another?

```
#Tata Motors
stockdata_TataMotors<- complete_stock_data %>%
  group_by(Date) %>%
  filter(Date > "2016-01-01" & Date <"2021-01-01", Company.Name == "Tata Motors Ltd.") %>%
  select(Year, Date, Close, Industry, Company.Name)
stockdata_TataMotors
```

```
#ICICI Bank
stockdata_ICICIBank<- complete_stock_data %>%
  group_by(Date) %>%
  filter(Date > "2016-01-01" & Date <"2021-01-01", Company.Name == "ICICI Bank Ltd.") %>%
  select(Year,Date, Close, Industry, Company.Name)
stockdata_ICICIBank
```

```
#Infosys
stockdata_Infosys<- complete_stock_data %>%
  group_by(Date) %>%
  filter(Date > "2016-01-01" & Date <"2021-01-01", Company.Name == "Infosys Ltd.") %>%
  select(Year,Date, Close, Industry, Company.Name)
stockdata_Infosys
```

```

#Reliance Industries Ltd.
stockdata_Reliance<- complete_stock_data %>%
  group_by(Date) %>%
  filter(Date > "2016-01-01" & Date <"2021-01-01", Company.Name == "Reliance Industries Ltd.") %>%
  select(Year,Date, Close, Industry, Company.Name)
stockdata_Reliance

```

Time Series Plot - Close Price vs Years of Top Performing companies

```

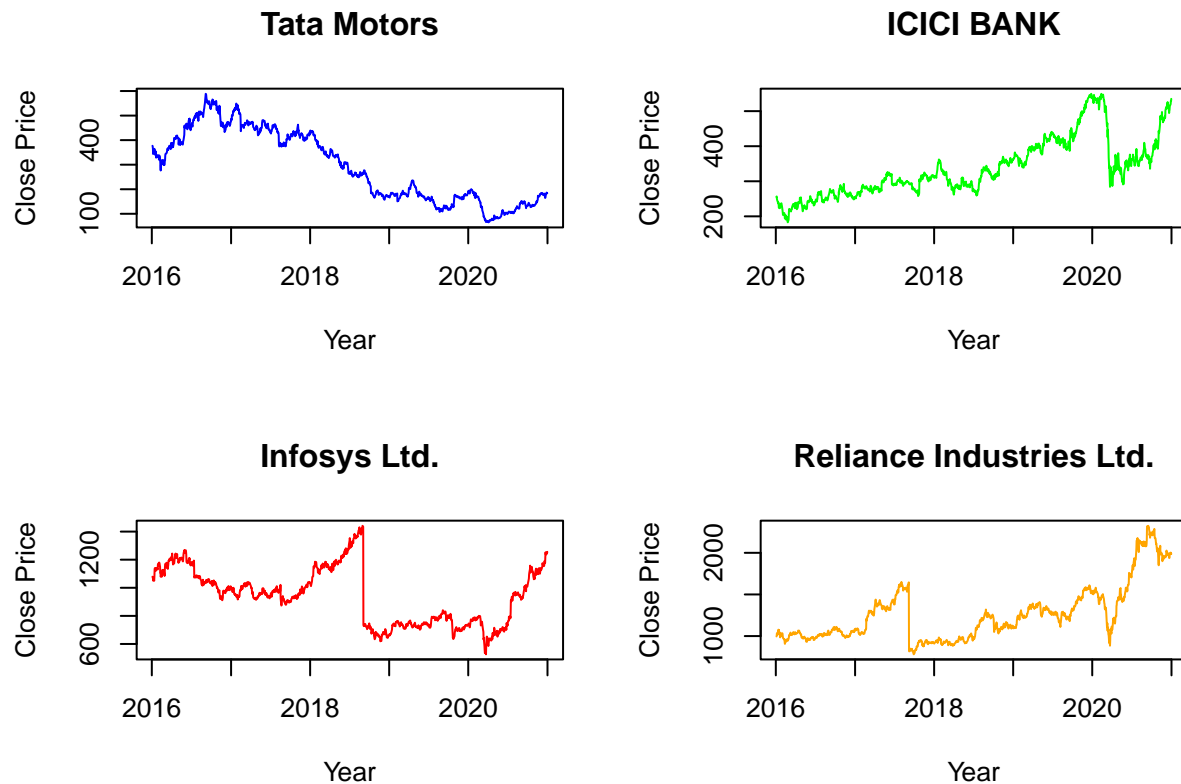
#Distribution vs time under Financial Sector
par(mfrow=c(2,2))
stockdata_TataMotors_plot <- plot(stockdata_TataMotors$Date,stockdata_TataMotors$Close,
                                lty = 1, lwd = 1, xlab = "Year", ylab = "Close Price", type = 'l',
                                main = 'Tata Motors', col = 'blue')

stockdata_ICICIBank_plot <- plot(stockdata_ICICIBank$Date,stockdata_ICICIBank$Close,
                                lty = 1, lwd = 1, xlab = "Year", ylab = "Close Price", type = 'l',
                                main = 'ICICI BANK', col = 'green')

stockdata_Infosys_plot <- plot(stockdata_Infosys$Date,stockdata_Infosys$Close,
                                lty = 1, lwd = 1, xlab = "Year", ylab = "Close Price", type = 'l',
                                main = 'Infosys Ltd.', col = 'red')

stockdata_Reliance_plot <- plot(stockdata_Reliance$Date,stockdata_Reliance$Close,
                                lty = 1, lwd = 1, xlab = "Year", ylab = "Close Price", type = 'l',
                                main = 'Reliance Industries Ltd.', col = 'orange')

```



Insight: 1. Based on the above Time series analysis Tata Motors shows a steady decline from the year 2017/18. 2. In Finance Sector - ICICI Bank has performed really well 2016-2020. After COVID-19 Pandemic the bank saw a huge loss in stocks. 3. In IT Sector - Infosys Ltd. stocks have remain constant throughout 2016 to mid 2018, after which it has increased steadily. 4. In Energy Sector - Reliance Industries Ltd. has show a excellent growth over the years. Only in mid-2017 & early-2020 the company saw a downfall in its stockprice. Reliance insdustries stocks have seen an overall excellent growth over the years.

Conclusion From this project we have learnt the basics of R which consists of Data Cleaning, Data wrangling, Clustring, Probability Distribution, Text Mining & Sentiment Analysis and Time Series Analysis.

Using various new libraries we were able to convert the data into excellent visualizations which gives us insights to all the business questions which are helpful to select the stocks for future investments and profit growth.

THANK YOU