library(dplyr)

library(ggplot2)

library(tidyverse)

library(lubridate)

library(readxl)

data <- read\_excel("shaurya r project/data.xlsx")

View(data)

data <- read\_excel("C:/Users/shaurya khanna/Desktop/shaurya r project/data.xlsx")

View(data)

str(data)

data$Date<- as.numeric(data$Date)

#summary of Columns of HUL

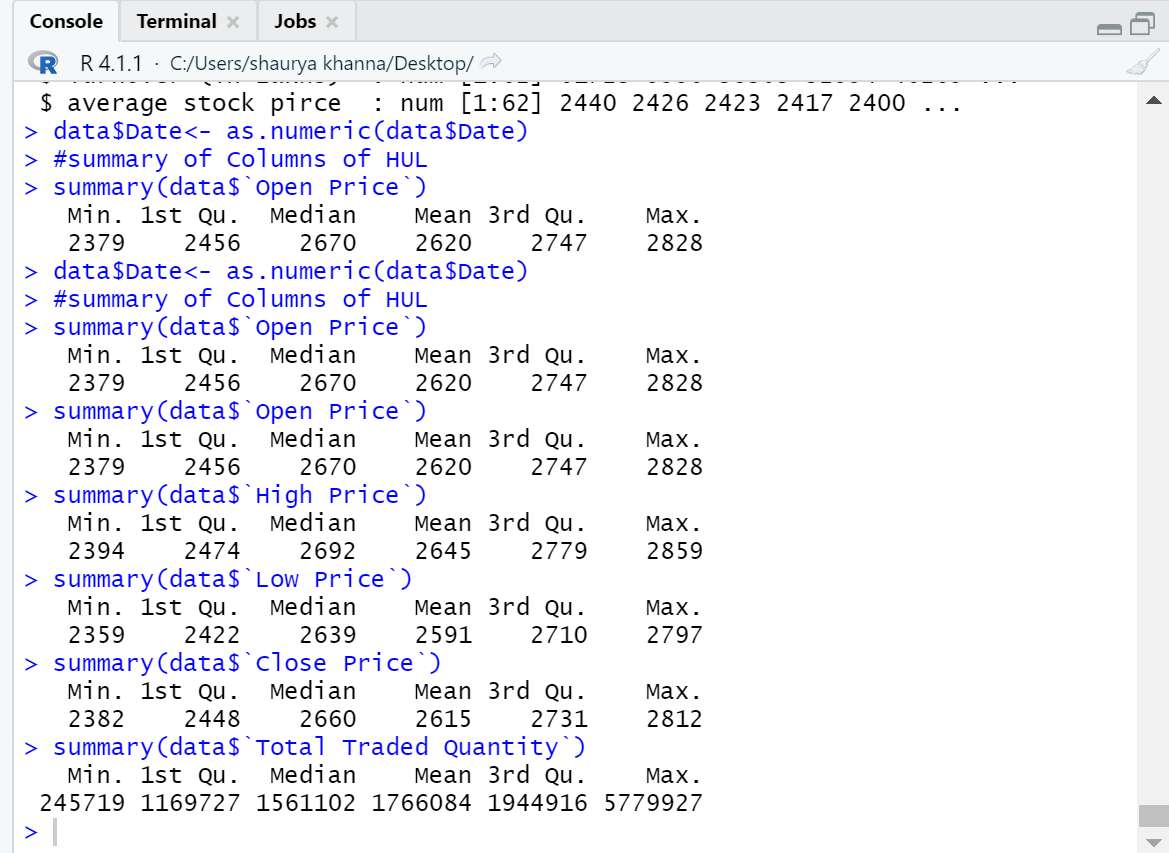
summary(data$`Open Price`)

summary(data$`High Price`)

summary(data$`Low Price`)

summary(data$`Close Price`)

summary(data$`Total Traded Quantity`)



Interpretation:

Mean is an essential concept in mathematics and statistics. The mean is **the average or the most common value in a collection of numbers**.

Median is a statistical measure that determines the middle value of a dataset listed in ascending order (i.e., from smallest to largest value).

The median value for the variables is higher than the mean value which is a good thing which means that mean is not supported by outliers but by a collective set of consistent quantitative values.

#Correlation Between Date and other Variables

cor.test(data$'Date',data$'Low Price', method = "pearson")

cor.test(data$'Date',data$'High Price',method = "pearson")

cor(data$Date,data$'Open Price')

cor(data$Date,data$'High Price')

cor(data$Date,data$'Low Price')

cor(data$Date,data$`Close Price`)

cor(data$Date,data$`Total Traded Quantity`)



The interpretation of this output is as follows

Positive correlation: A positive correlation would be 1. This means the two variables moved either up or down in the same direction together.

Negative correlation: A negative correlation is -1. This means the two variables moved in opposite directions.

Zero or no correlation: A correlation of zero means there is no relationship between the two variables. In other words, as one variable moves one way, the other moved in another unrelated direction.

Interpretation: Correlation of date with opening price, closing,high and low price is negative which means that which means that date and other variables move in opposite direction

Correlation between date and total traded quantity is positive which means moves in the same direction together.as the date invreases total traded quantity also increases.

#Correlation With different Variables

cor(data$'Open Price', data$`Close Price`)

cor(data$`Open Price` ,data$`High Price`)

cor(data$`Open Price`,data$`Low Price`)

cor(data$`Close Price`, data$`Low Price`)

cor(data$`Close Price`, data$`High Price`)

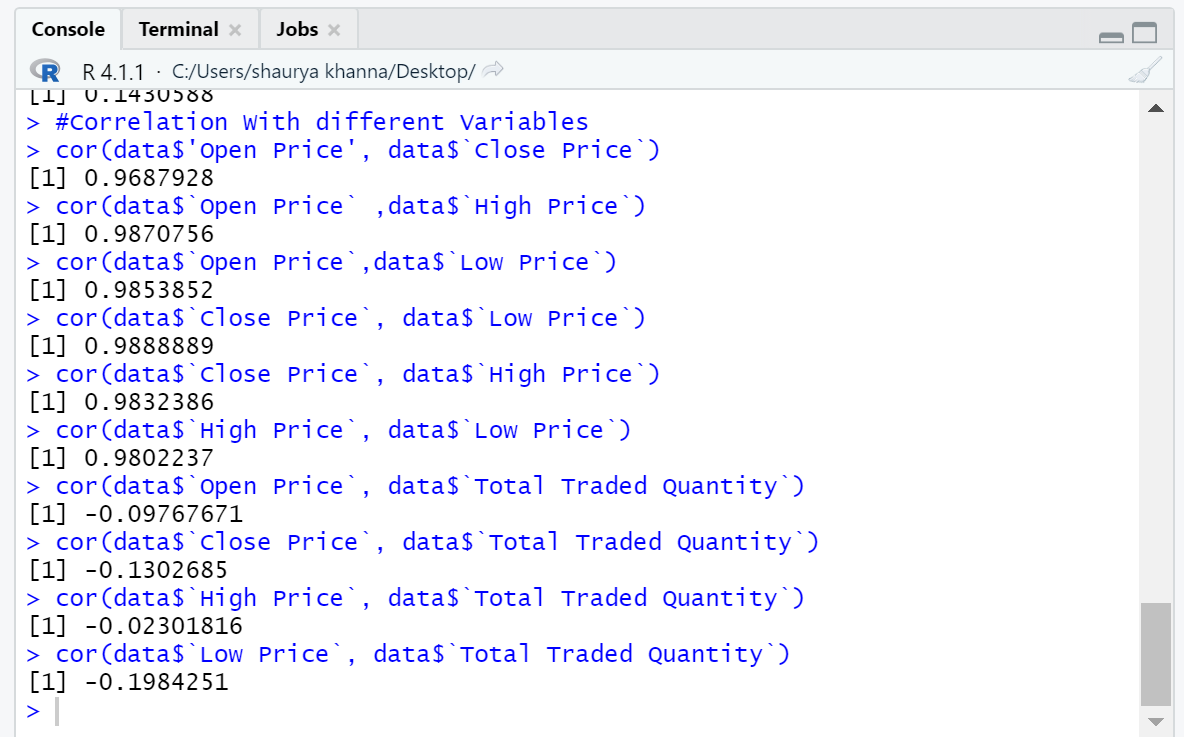
cor(data$`High Price`, data$`Low Price`)

cor(data$`Open Price`, data$`Total Traded Quantity`)

cor(data$`Close Price`, data$`Total Traded Quantity`)

cor(data$`High Price`, data$`Total Traded Quantity`)

cor(data$`Low Price`, data$`Total Traded Quantity`)



The interpretation of this output is as follows

Interpretation: Correlation of closing price with opening price is positive which means that which means that open price and close price move in same direction together.

Same is the interpretation opening and high price etc.

Correlation between opening price and total traded quantity is negative which means that they move in the different direction together as the date increases total traded quantity also increases.

Same is the interpretation for closing and total traded quantity although negative correlation is very high for high price and total traded quantity they would move in opposite less often for the variables where negative correlation is high.

#Correlation Test Between Date and other Variables

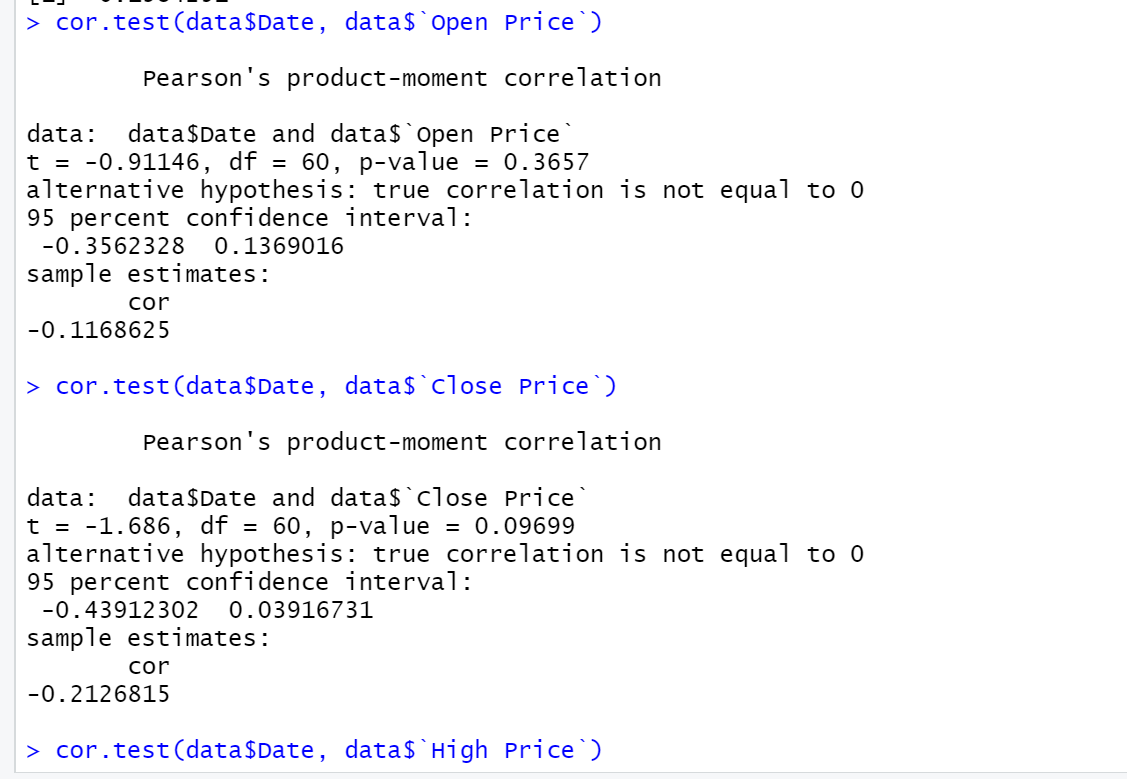
cor.test(data$Date, data$`Open Price`)

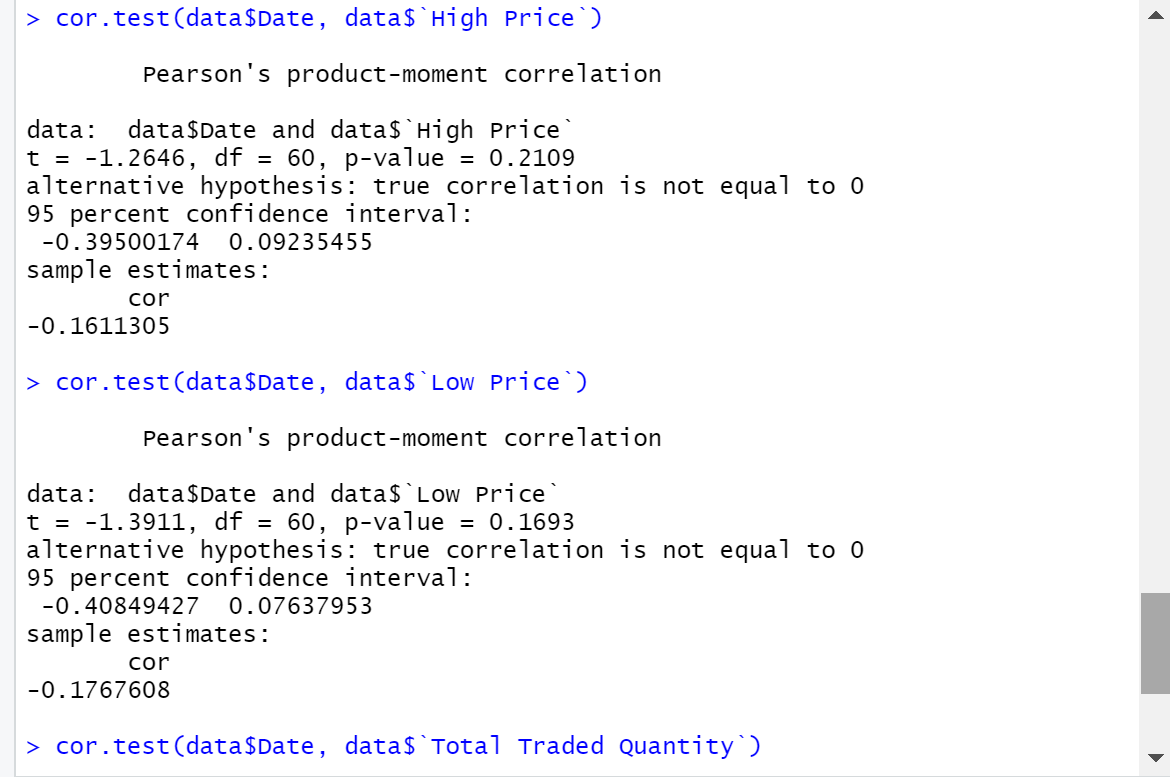
cor.test(data$Date, data$`Close Price`)

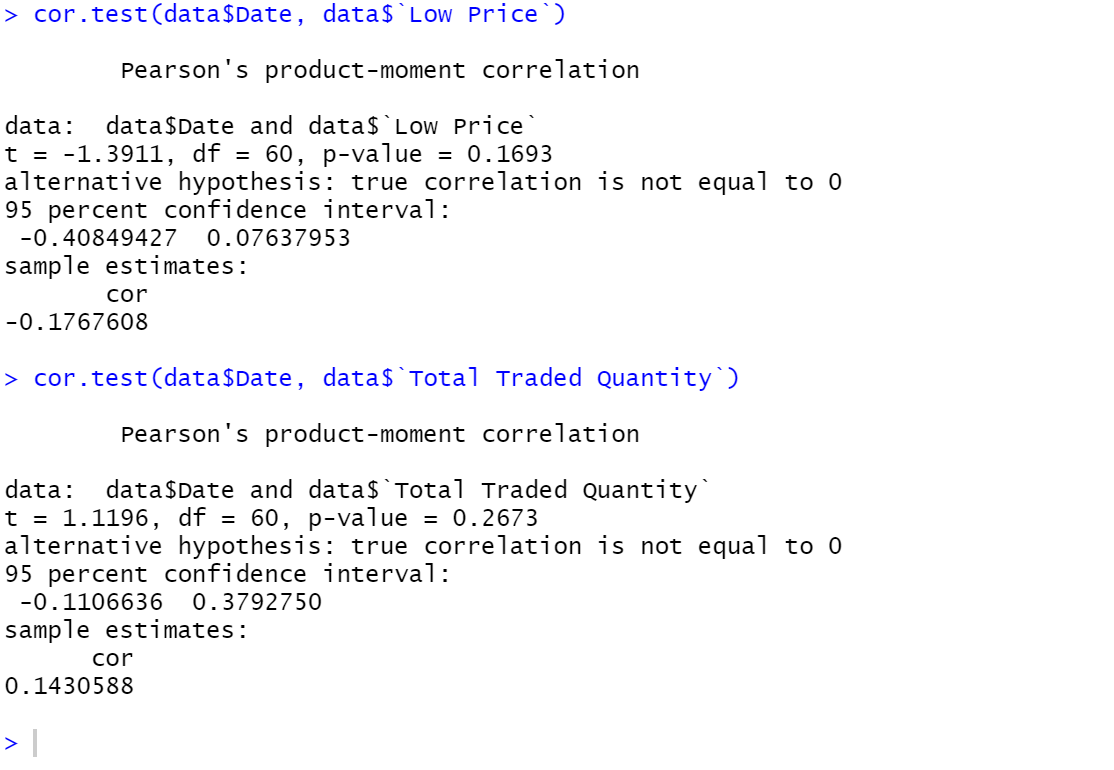
cor.test(data$Date, data$`High Price`)

cor.test(data$Date, data$`Low Price`)

cor.test(data$Date, data$`Total Traded Quantity`)







#Correlation Test With different Variables

cor.test(data$`Open Price`, data$`Close Price`)

cor.test(data$`Open Price` ,data$`High Price`)

cor.test(data$`Open Price`,data$`Low Price`)

cor.test(data$`Close Price`,data$`Low Price`)

cor.test(data$`Close Price`, data$`High Price`)

cor.test(data$`High Price`, data$`Low Price`)

cor.test(data$`Open Price`, data$`Total Traded Quantity`)

cor.test(data$`Close Price`, data$`Total Traded Quantity`)

cor.test(data$`High Price`, data$`Total Traded Quantity`)

cor.test(data$`Low Price`, data$`Total Traded Quantity`)

> cor.test(data$`Open Price`, data$`Close Price`)

**Pearson's product-moment correlation**

**data: data$`Open Price` and data$`Close Price`**

**t = 30.275, df = 60, p-value < 2.2e-16**

**alternative hypothesis: true correlation is not equal to 0**

**95 percent confidence interval:**

**0.9485483 0.9811489**

**sample estimates:**

**cor**

**0.9687928**

**> cor.test(data$`Open Price` ,data$`High Price`)**

**Pearson's product-moment correlation**

**data: data$`Open Price` and data$`High Price`**

**t = 47.71, df = 60, p-value < 2.2e-16**

**alternative hypothesis: true correlation is not equal to 0**

**95 percent confidence interval:**

**0.9785622 0.9922214**

**sample estimates:**

**cor**

**0.9870756**

**> cor.test(data$`Open Price`,data$`Low Price`)**

**Pearson's product-moment correlation**

**data: data$`Open Price` and data$`Low Price`**

**t = 44.809, df = 60, p-value < 2.2e-16**

**alternative hypothesis: true correlation is not equal to 0**

**95 percent confidence interval:**

**0.9757719 0.9912011**

**sample estimates:**

**cor**

**0.9853852**

**> cor.test(data$`Close Price`,data$`Low Price`)**

**Pearson's product-moment correlation**

**data: data$`Close Price` and data$`Low Price`**

**t = 51.528, df = 60, p-value < 2.2e-16**

**alternative hypothesis: true correlation is not equal to 0**

**95 percent confidence interval:**

**0.9815589 0.9933152**

**sample estimates:**

**cor**

**0.9888889**

**> cor.test(data$`Close Price`, data$`High Price`)**

**Pearson's product-moment correlation**

**data: data$`Close Price` and data$`High Price`**

**t = 41.773, df = 60, p-value < 2.2e-16**

**alternative hypothesis: true correlation is not equal to 0**

**95 percent confidence interval:**

**0.9722330 0.9899043**

**sample estimates:**

**cor**

**0.9832386**

**> cor.test(data$`High Price`, data$`Low Price`)**

**Pearson's product-moment correlation**

**data: data$`High Price` and data$`Low Price`**

**t = 38.368, df = 60, p-value < 2.2e-16**

**alternative hypothesis: true correlation is not equal to 0**

**95 percent confidence interval:**

**0.9672713 0.9880812**

**sample estimates:**

**cor**

**0.9802237**

**> cor.test(data$`Open Price`, data$`Total Traded Quantity`)**

**Pearson's product-moment correlation**

**data: data$`Open Price` and data$`Total Traded Quantity`**

**t = -0.76024, df = 60, p-value = 0.4501**

**alternative hypothesis: true correlation is not equal to 0**

**95 percent confidence interval:**

**-0.3391703 0.1558948**

**sample estimates:**

**cor**

**-0.09767671**

**> cor.test(data$`Close Price`, data$`Total Traded Quantity`)**

**Pearson's product-moment correlation**

**data: data$`Close Price` and data$`Total Traded Quantity`**

**t = -1.0177, df = 60, p-value = 0.3129**

**alternative hypothesis: true correlation is not equal to 0**

**95 percent confidence interval:**

**-0.3680611 0.1235186**

**sample estimates:**

**cor**

**-0.1302685**

**> cor.test(data$`High Price`, data$`Total Traded Quantity`)**

**Pearson's product-moment correlation**

**data: data$`High Price` and data$`Total Traded Quantity`**

**t = -0.17835, df = 60, p-value = 0.8591**

**alternative hypothesis: true correlation is not equal to 0**

**95 percent confidence interval:**

**-0.2712270 0.2280612**

**sample estimates:**

**cor**

**-0.02301816**

**> cor.test(data$`Low Price`, data$`Total Traded Quantity`)**

**Pearson's product-moment correlation**

**data: data$`Low Price` and data$`Total Traded Quantity`**

**t = -1.5682, df = 60, p-value = 0.1221**

**alternative hypothesis: true correlation is not equal to 0**

**95 percent confidence interval:**

**-0.42702961 0.05402036**

**sample estimates:**

**cor**

**-0.1984251**

model=lm(data$`Open Price`~data$`Low Price`)

model

summary(model)

model1=lm(data$`Open Price`~data$`High Price`)

model1

summary(model1)

model2=lm(data$`Close Price`~data$`Low Price`)

model2

summary(model2)

model3=lm(data$`Open Price`~data$`High Price`)

model3

summary(model3)

model4=lm(data$`High Price`~data$`Total Traded Quantity`)

model4

summary(model4)

model5=lm(data$`High Price`~data$`Low Price`)

model5

summary(model5)

model6=lm(data$`Open Price`~data$`Total Traded Quantity`)

model6

summary(model6)

model7=lm(data$`Close Price`~data$`Total Traded Quantity`)

model7

summary(model7)

model8=lm(data$`Low Price`~data$`Total Traded Quantity`)

model8

summary(model8)

model9=lm(data$'Date'~data$`Open Price`)

model9

summary(model9)

> model1=lm(data$`Open Price`~data$`High Price`)

> model1

Call:

lm(formula = data$`Open Price` ~ data$`High Price`)

Coefficients:

(Intercept) data$`High Price`

104.3360 0.9512

> summary(model1)

Call:

lm(formula = data$`Open Price` ~ data$`High Price`)

Residuals:

Min 1Q Median 3Q Max

-115.928 -7.342 4.662 15.060 30.663

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 104.33598 52.81140 1.976 0.0528 .

data$`High Price` 0.95118 0.01994 47.710 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 23.94 on 60 degrees of freedom

**Multiple R-squared: 0.9743, Adjusted R-squared: 0.9739**

F-statistic: 2276 on 1 and 60 DF, p-value: < 2.2e-16

>

> model2=lm(data$`Close Price`~data$`Low Price`)

> model2

Call:

lm(formula = data$`Close Price` ~ data$`Low Price`)

Coefficients:

(Intercept) data$`Low Price`

71.9964 0.9816

> summary(model2)

Call:

lm(formula = data$`Close Price` ~ data$`Low Price`)

Residuals:

Min 1Q Median 3Q Max

-24.696 -11.629 -5.581 4.646 116.883

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 71.99641 49.43199 1.456 0.15

data$`Low Price` 0.98163 0.01905 51.528 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 22.32 on 60 degrees of freedom

**Multiple R-squared: 0.9779, Adjusted R-squared: 0.9775**

F-statistic: 2655 on 1 and 60 DF, p-value: < 2.2e-16

>

> model3=lm(data$`Open Price`~data$`High Price`)

> model3

Call:

lm(formula = data$`Open Price` ~ data$`High Price`)

Coefficients:

(Intercept) data$`High Price`

104.3360 0.9512

> summary(model3)

Call:

lm(formula = data$`Open Price` ~ data$`High Price`)

Residuals:

Min 1Q Median 3Q Max

-115.928 -7.342 4.662 15.060 30.663

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 104.33598 52.81140 1.976 0.0528 .

data$`High Price` 0.95118 0.01994 47.710 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 23.94 on 60 degrees of freedom

**Multiple R-squared: 0.9743, Adjusted R-squared: 0.9739**

F-statistic: 2276 on 1 and 60 DF, p-value: < 2.2e-16

>

> model4=lm(data$`High Price`~data$`Total Traded Quantity`)

> model4

Call:

lm(formula = data$`High Price` ~ data$`Total Traded Quantity`)

Coefficients:

(Intercept) data$`Total Traded Quantity`

2.651e+03 -3.487e-06

> summary(model4)

Call:

lm(formula = data$`High Price` ~ data$`Total Traded Quantity`)

Residuals:

Min 1Q Median 3Q Max

-253.06 -171.37 45.99 133.82 216.49

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.651e+03 3.975e+01 66.689 <2e-16 \*\*\*

data$`Total Traded Quantity` -3.487e-06 1.956e-05 -0.178 0.859

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 155 on 60 degrees of freedom

**Multiple R-squared: 0.0005298, Adjusted R-squared: -0.01613**

F-statistic: 0.03181 on 1 and 60 DF, p-value: 0.8591

>

> model5=lm(data$`High Price`~data$`Low Price`)

> model5

Call:

lm(formula = data$`High Price` ~ data$`Low Price`)

Coefficients:

(Intercept) data$`Low Price`

42.301 1.005

> summary(model5)

Call:

lm(formula = data$`High Price` ~ data$`Low Price`)

Residuals:

Min 1Q Median 3Q Max

-43.316 -16.203 -5.748 6.963 157.727

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 42.30050 67.93614 0.623 0.536

data$`Low Price` 1.00455 0.02618 38.368 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 30.67 on 60 degrees of freedom

**Multiple R-squared: 0.9608, Adjusted R-squared: 0.9602**

F-statistic: 1472 on 1 and 60 DF, p-value: < 2.2e-16

>

> model6=lm(data$`Open Price`~data$`Total Traded Quantity`)

> model6

Call:

lm(formula = data$`Open Price` ~ data$`Total Traded Quantity`)

Coefficients:

(Intercept) data$`Total Traded Quantity`

2.645e+03 -1.426e-05

> summary(model6)

Call:

lm(formula = data$`Open Price` ~ data$`Total Traded Quantity`)

Residuals:

Min 1Q Median 3Q Max

-252.41 -158.00 48.63 125.23 205.44

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.645e+03 3.813e+01 69.37 <2e-16 \*\*\*

data$`Total Traded Quantity` -1.426e-05 1.876e-05 -0.76 0.45

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 148.6 on 60 degrees of freedom

**Multiple R-squared: 0.009541, Adjusted R-squared: -0.006967**

F-statistic: 0.578 on 1 and 60 DF, p-value: 0.4501

>

> model7=lm(data$`Close Price`~data$`Total Traded Quantity`)

> model7

Call:

lm(formula = data$`Close Price` ~ data$`Total Traded Quantity`)

Coefficients:

(Intercept) data$`Total Traded Quantity`

2.649e+03 -1.912e-05

> summary(model7)

Call:

lm(formula = data$`Close Price` ~ data$`Total Traded Quantity`)

Residuals:

Min 1Q Median 3Q Max

-246.59 -162.08 46.03 127.90 207.31

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.649e+03 3.818e+01 69.368 <2e-16 \*\*\*

data$`Total Traded Quantity` -1.912e-05 1.878e-05 -1.018 0.313

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 148.9 on 60 degrees of freedom

**Multiple R-squared: 0.01697, Adjusted R-squared: 0.000586**

F-statistic: 1.036 on 1 and 60 DF, p-value: 0.3129

>

> model8=lm(data$`Low Price`~data$`Total Traded Quantity`)

> model8

Call:

lm(formula = data$`Low Price` ~ data$`Total Traded Quantity`)

Coefficients:

(Intercept) data$`Total Traded Quantity`

2.642e+03 -2.934e-05

> summary(model8)

Call:

lm(formula = data$`Low Price` ~ data$`Total Traded Quantity`)

Residuals:

Min 1Q Median 3Q Max

-252.08 -160.07 46.14 133.99 221.15

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.642e+03 3.802e+01 69.490 <2e-16 \*\*\*

data$`Total Traded Quantity` -2.934e-05 1.871e-05 -1.568 0.122

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 148.2 on 60 degrees of freedom

**Multiple R-squared: 0.03937, Adjusted R-squared: 0.02336**

F-statistic: 2.459 on 1 and 60 DF, p-value: 0.1221

>

> model9=lm(data$'Date'~data$`Open Price`)

> model9

Call:

lm(formula = data$Date ~ data$`Open Price`)

Coefficients:

(Intercept) data$`Open Price`

1637054732 -1813

> summary(model9)

Call:

lm(formula = data$Date ~ data$`Open Price`)

Residuals:

Min 1Q Median 3Q Max

-4273608 -1766110 358030 1910546 3714885

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1637054732 5219287 313.655 <2e-16 \*\*\*

data$`Open Price` -1813 1989 -0.911 0.366

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2301000 on 60 degrees of freedom

**Multiple R-squared: 0.01366, Adjusted R-squared: -0.002782**

F-statistic: 0.8308 on 1 and 60 DF, p-value: 0.3657

Interpretation

For the variables where R squared value is the highest it indicates that the one variable impact the other variable the most those will be the highest impacting variables,

For the variables where R squared value is the lowest those variable will be low impacting variables.

For this output we can interpret that high and low price impact other the most since they have the highest r square value.

################Graphical Representation#####################

library(readxl)

data2 <- read\_excel("data2.xlsx")

View(data2)

#linechart Between date and open price

library(ggplot2)

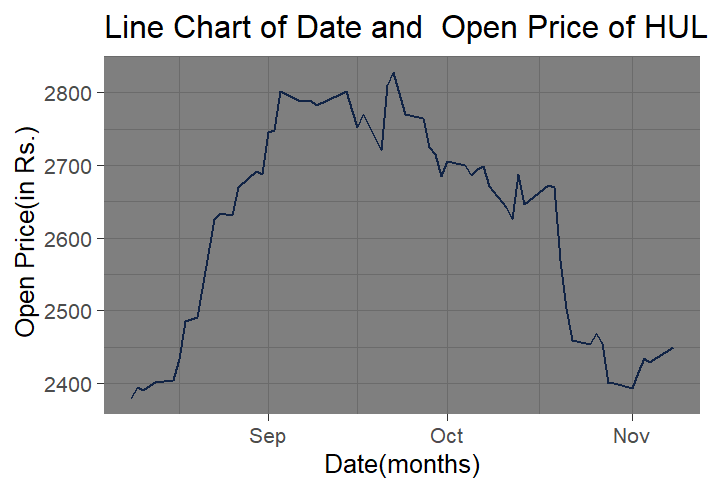
ggplot(data2) +

aes(x = Date, y = `Open Price`) +

geom\_line(size = 0.5, colour = "#112446") +

labs(x = "Date(months)", y = " Open Price(in Rs.)", title = "Line Chart of Date and Open Price of HUL") +

theme\_dark()



Interpretation

We can see that HUL achieved its highest opening price around September 20th we can also see the negative correlation being displayed by states.

#linechart Between date and close Price

library(ggplot2)

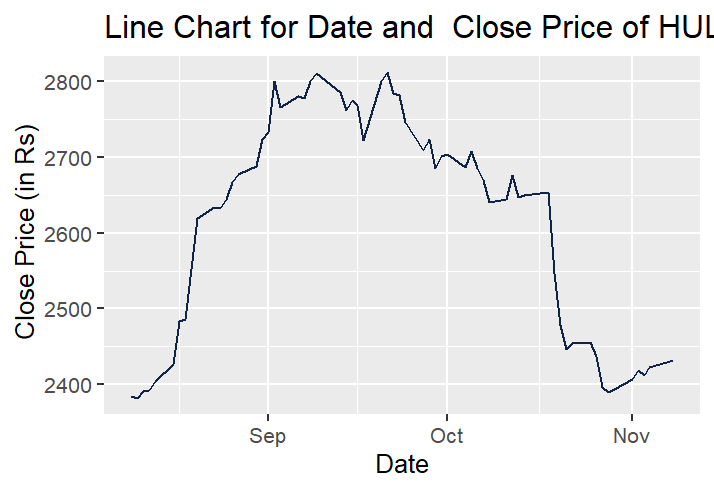
ggplot(data2) +

aes(x = Date, y = `Close Price`) +

geom\_line(size = 0.5, colour = "#112446") +

labs(x = "Date", y = " Close Price (in Rs)", title = "Line Chart for Date and Close Price of HUL") +

theme\_gray()



We can also see the negative correlation between date and closing price being displayed by the graph.

#linechart Between Date and High Price

library(ggplot2)

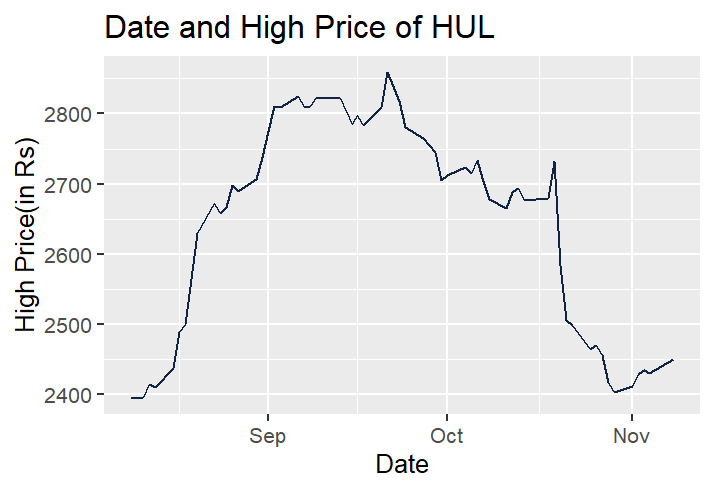
ggplot(data2) +

aes(x = Date, y = `High Price`) +

geom\_line(size = 0.5, colour = "#112446") +

labs(x = "Date", y = "High Price(in Rs)", title = "Date and High Price of HUL") +

theme\_gray()

HUL achieved it’s highest high price around September 20th and lowest high price around October 25th.

#linechart Between Date and Low Price

library(ggplot2)

ggplot(data2) +

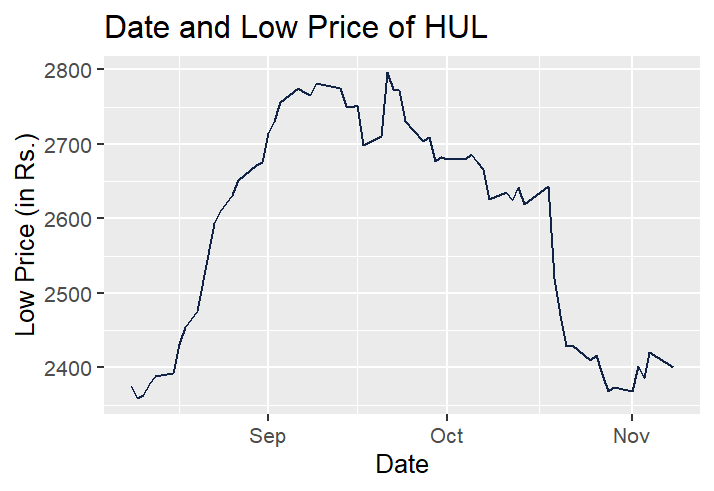
aes(x =Date, y = `Low Price`) +

geom\_line(size = 0.5, colour = "#112446") +

labs(x = "Date",

y = "Low Price (in Rs.)", title = "Date and Low Price of HUL") +

theme\_gray()



HUL achieved it’s highest low price around September 20th and lowest low price around October 25th.

#linechart Between Date and Total Traded Quantity

ggplot(data2) +

aes(x = Date, y = `Total Traded Quantity`) +

geom\_line(size = 0.5, colour = "#112446") +

labs(

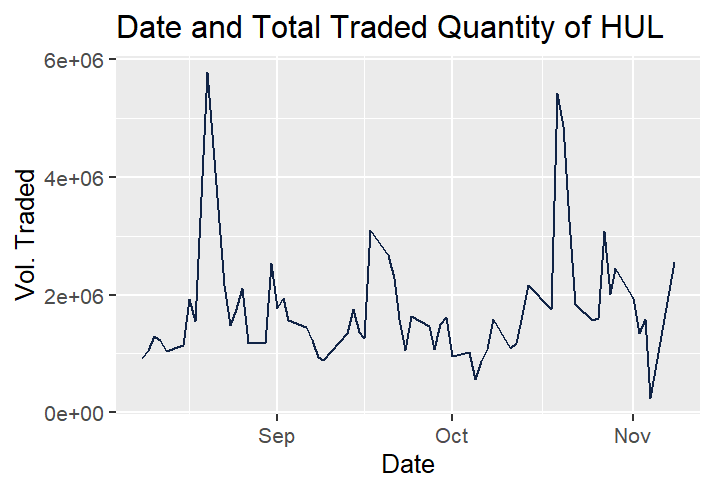
x = "Date",

y = "Vol. Traded",

title = "Date and Total Traded Quantity of HUL"

) +

theme\_gray()



Highest quantity that was traded as around august 18th and lowest around November 5th.

library(rvest)

library(dplyr)

url="https://www.moneycontrol.com/stocks/company\_info/stock\_news.php?sc\_id=HL&scat=&pageno=3&next=0&durationType=Y&Year=2021&duration=1&news\_type="

page1=read\_html(url)

page1

date1=page1%>%html\_nodes(".a\_10dgry")%>%html\_text()

date1

news1=page1%>%html\_nodes(".g\_14bl strong")%>%html\_text()

news1

desc1=page1%>%html\_nodes(".company-news-listing-txt p+ p")%>%html\_text()

newsmodel2=data.frame(date1,news1)

View(newsmodel2)

write.csv(newsmodel2,"newsmodel2.csv")

library(readxl)

newsmodel2 <- read\_excel("C:/Users/shaurya khanna/Desktop/newsmodel2.xlsx")

View(newsmodel2)

Country <- c("Positive", "Negative", "Positive", "Negative", "Positive", "Negative",

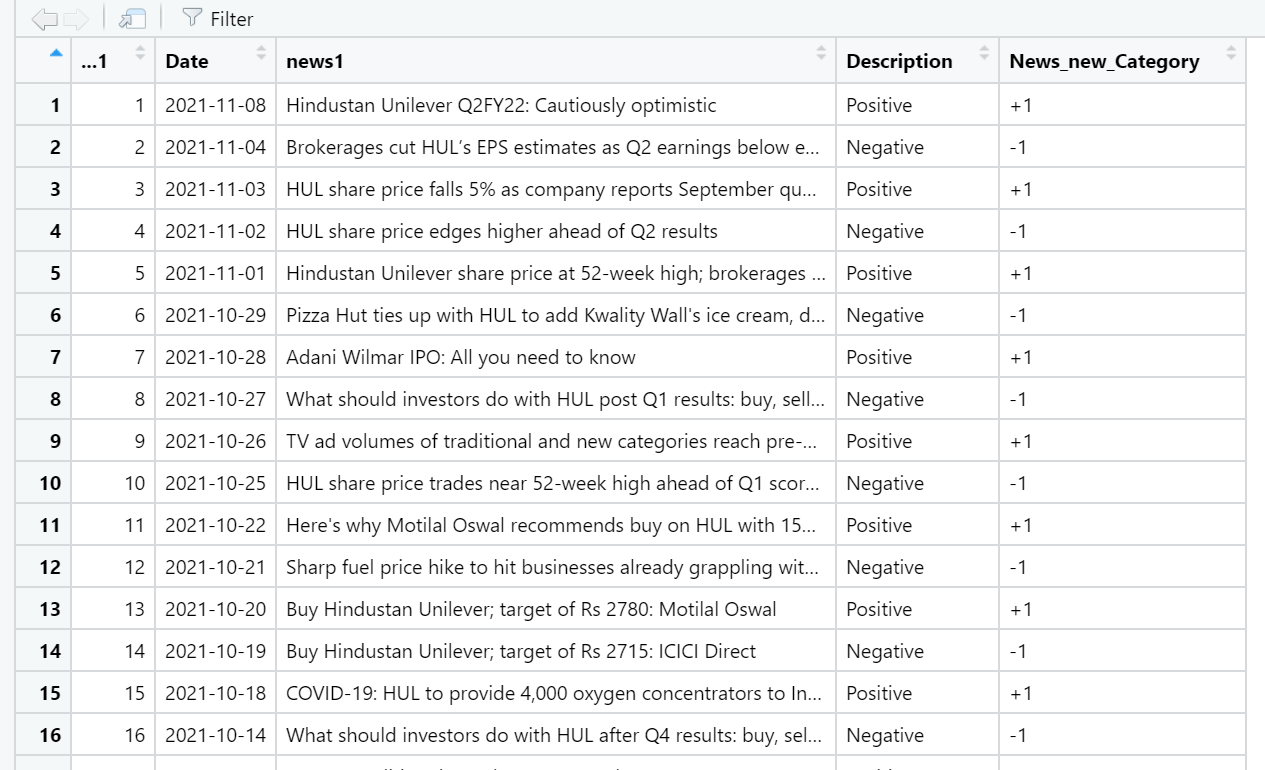
"Positive", "Negative", "Positive", "Negative","Positive", "Negative","Positive", "Negative","Positive", "Negative","Positive", "Negative","Positive", "Negative")

newsmodel2['Description'] <- Country

newsmodel2$Description

newsmodel2$News\_new\_Category=ifelse(newsmodel2$Description=="Positive", "+1", ifelse(newsmodel2$Description=="Negative","-1",ifelse(newsmodel2$Description=="Nuetral", "0"," ")))

View(newsmodel2)

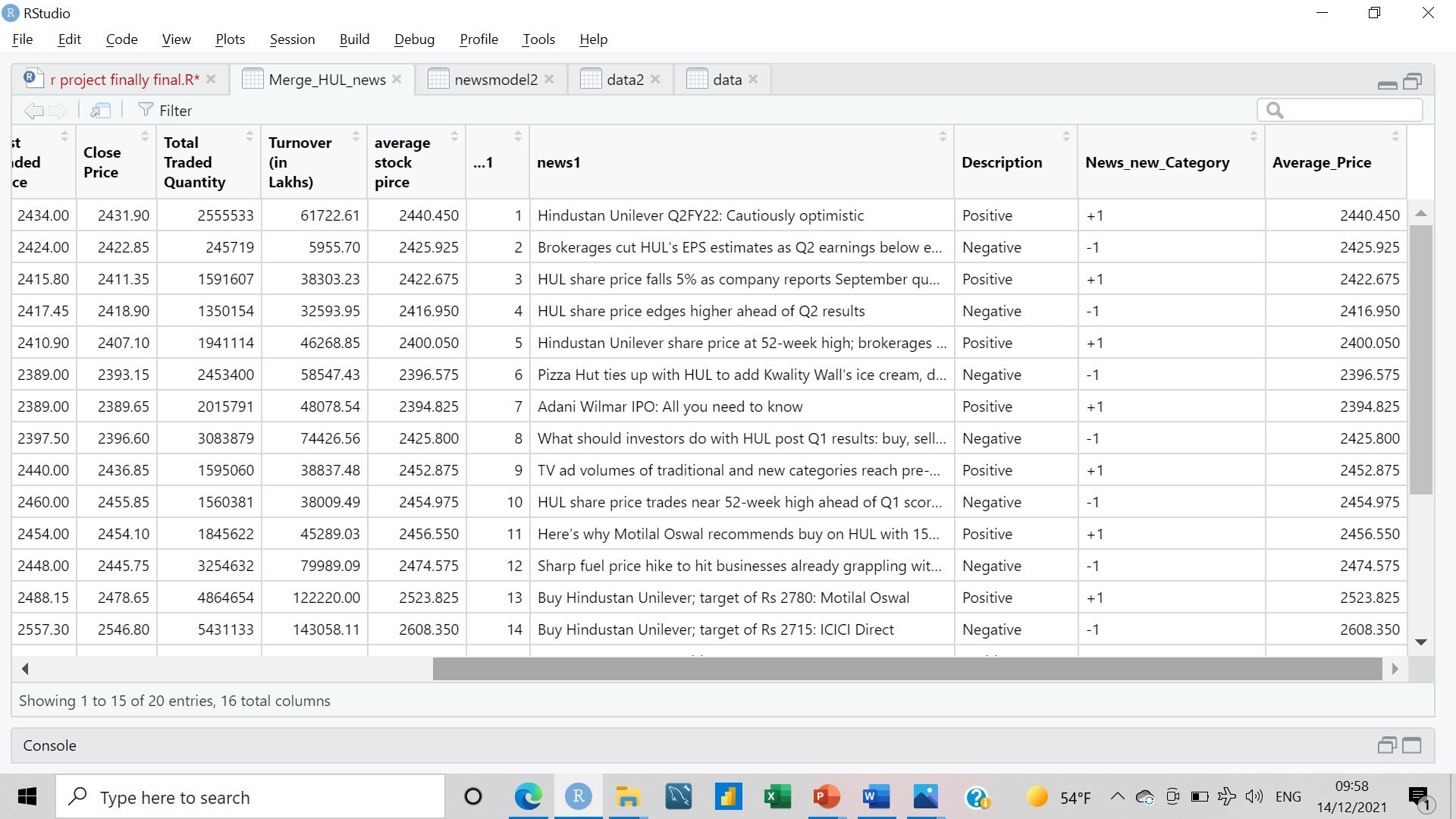


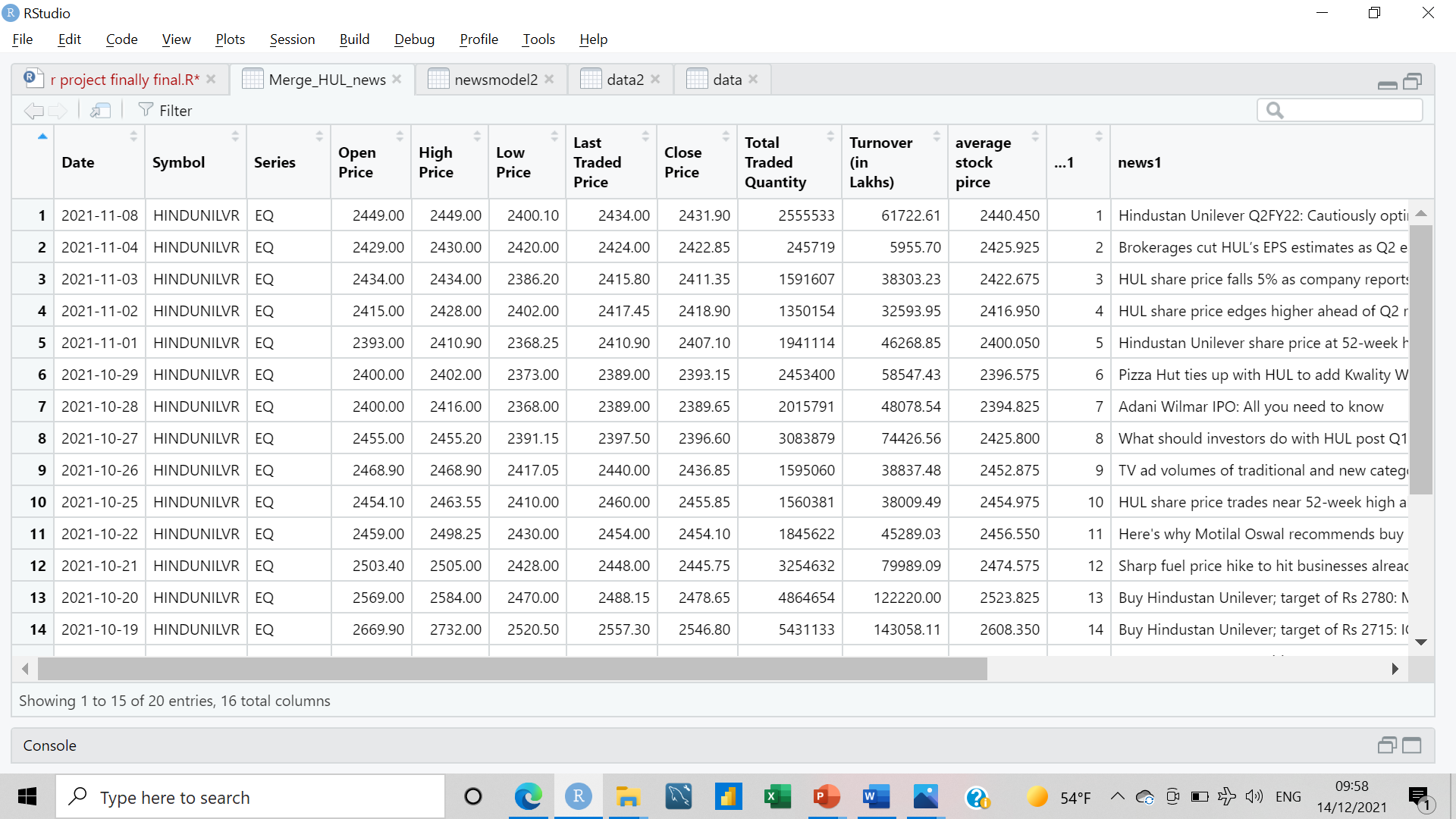
library(dplyr)

newsmodel2$Date<-as\_date(newsmodel2$Date)

Merge\_HUL\_news<- inner\_join(data2,newsmodel2,by="Date")

View(Merge\_HUL\_news)





names(Merge\_HUL\_news)

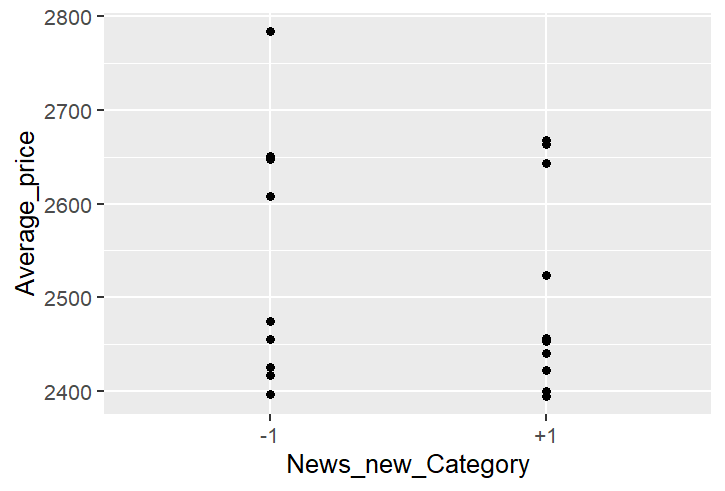
Merge\_HUL\_news$Average\_Price<-((Merge\_HUL\_news$`Open Price`+Merge\_HUL\_news$`Close Price`)/2)

model=lm(Merge\_HUL\_news$News\_new\_Category~Merge\_HUL\_news$Average\_Price)

summary(model)

names(Merge\_HUL\_news)

ggplot(data = Merge\_HUL\_news,aes(y=Merge\_HUL\_news$Average\_Price,x=Merge\_HUL\_news$News\_new\_Category)) + geom\_point() + labs(y='Average\_price',x="News\_new\_Category")



The interpretation that we can draw from this graph is as follows:-

* Average price for the days where the news is negative is not very much lower than the days where the news is positive this means that negative news about the company had minimal impact on it’s price.
* Average price for the days where the news is positive is not very much higher than the days where the news is negative this means that positive news about the company had minimal impact on its valuation.
* One of the interpretation we can also draw is that HUL is not a volatile stock whose valuation will swing according to nature of news in the market.
* HUL is a fundamentally stable whose valuation is well defined.