Analysis of Sales Report of a Clothes Manufacturing Outlet.

Q1) To automate the process of recommendations, the store needs to analyse the given attributes of the product, like the style, season, etc., and come up with a model to predict the recommendation of products (in binary output – 0 or 1) accordingly.

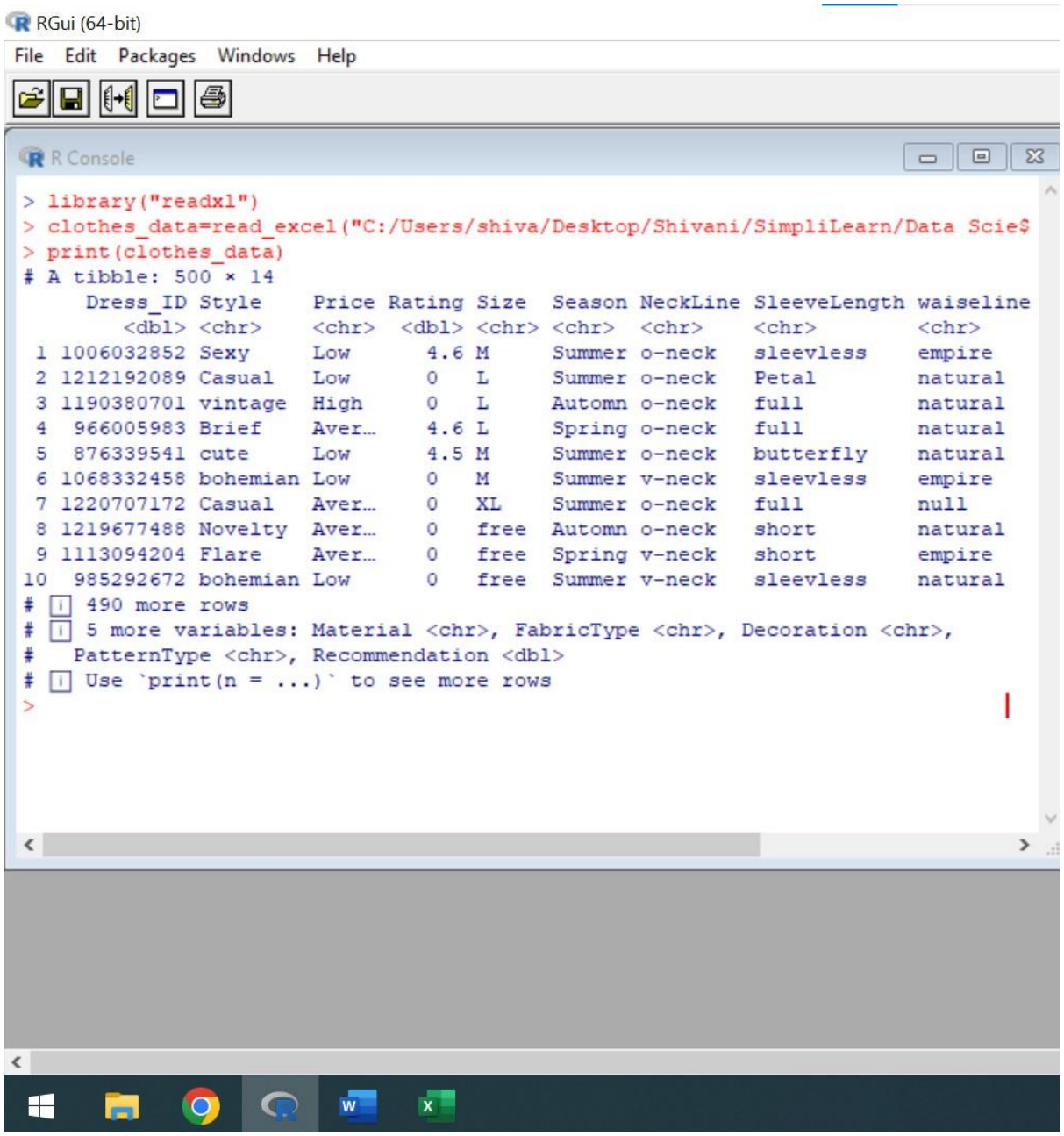
Ans) The various attributes of the product are Style,Season,Material,Fabric Type,Decoartion,Pattern Type etc.Here, I am using the concept of Regression Analysis with Multiple Variables. The dependent variable is the Recommendation while Season,Style,Fabric Type,Decoration are the independent variables. Below mention is the code of Regression analysis with Multiple variables

```
clothes_data=read_excel("C:/Users/shiva/Desktop/Shivani/SimpliLearn/Data Science with R/Attributes.xlsx")

print(clothes_data)

View(clothes_data)

str(clothes_data)
```

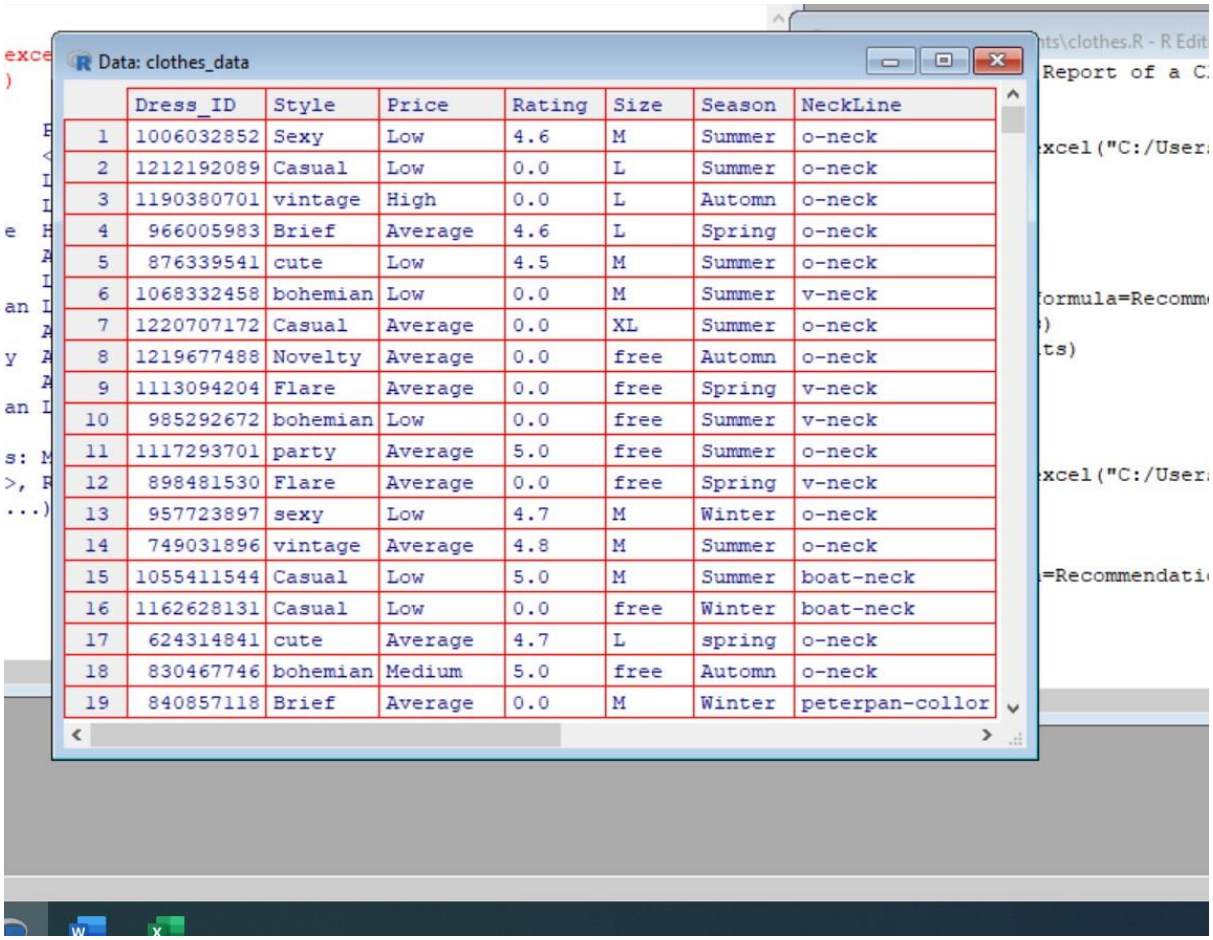
```
dress_results<-
lm(formula=Recommendation~Season+Style+Material+FabricType+Decoration+PatternType+Size+NeckLine+SleeveLength+waiseline,data=clothes_data)

print(dress_results)

summary(dress_results)
```

The output of the above mentioned Regression Analysis with Multiple variables is mention below:

The first screenshot represents the style,price,rating and other attributes of various dresses along with their dress ID’s

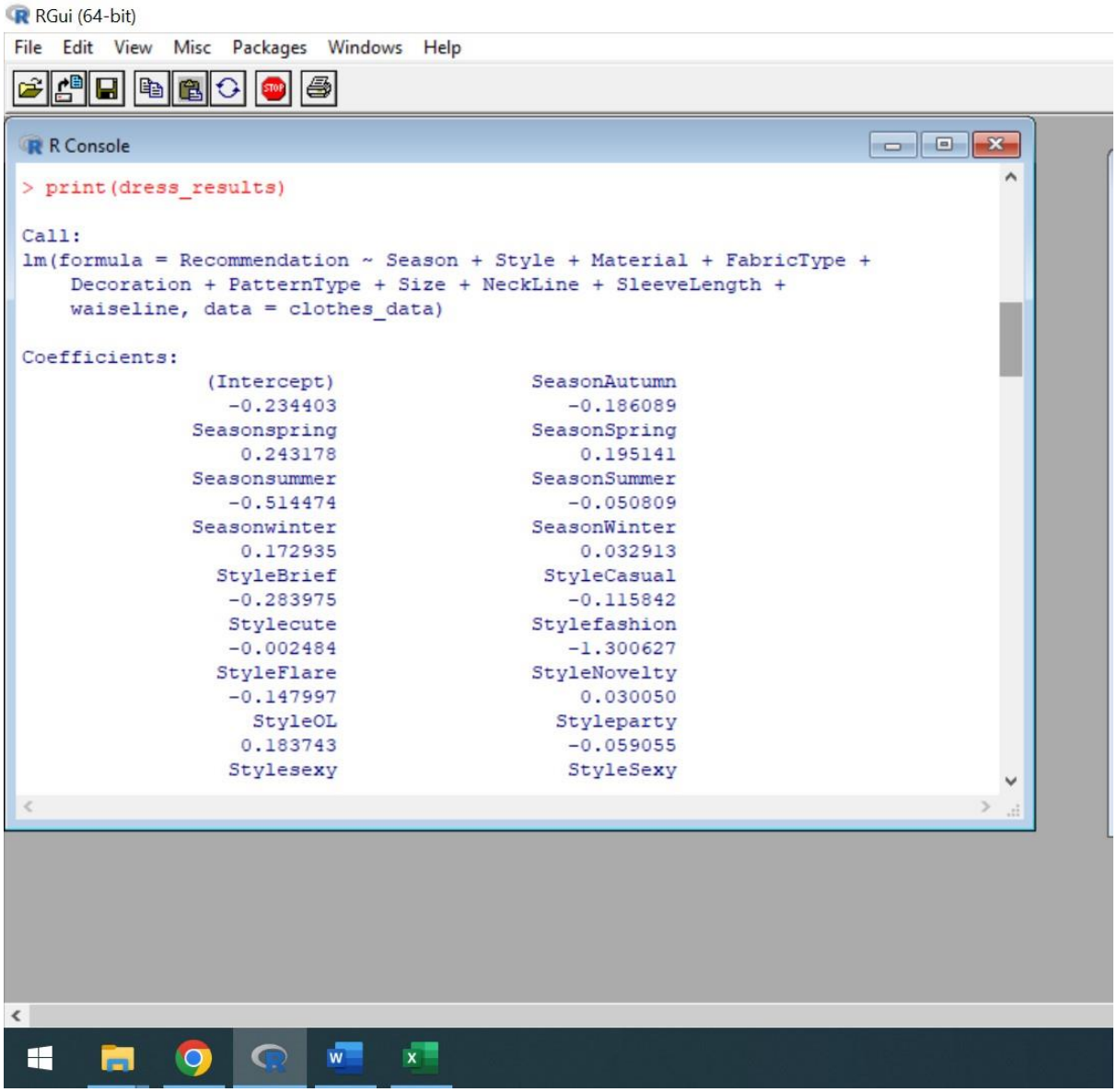


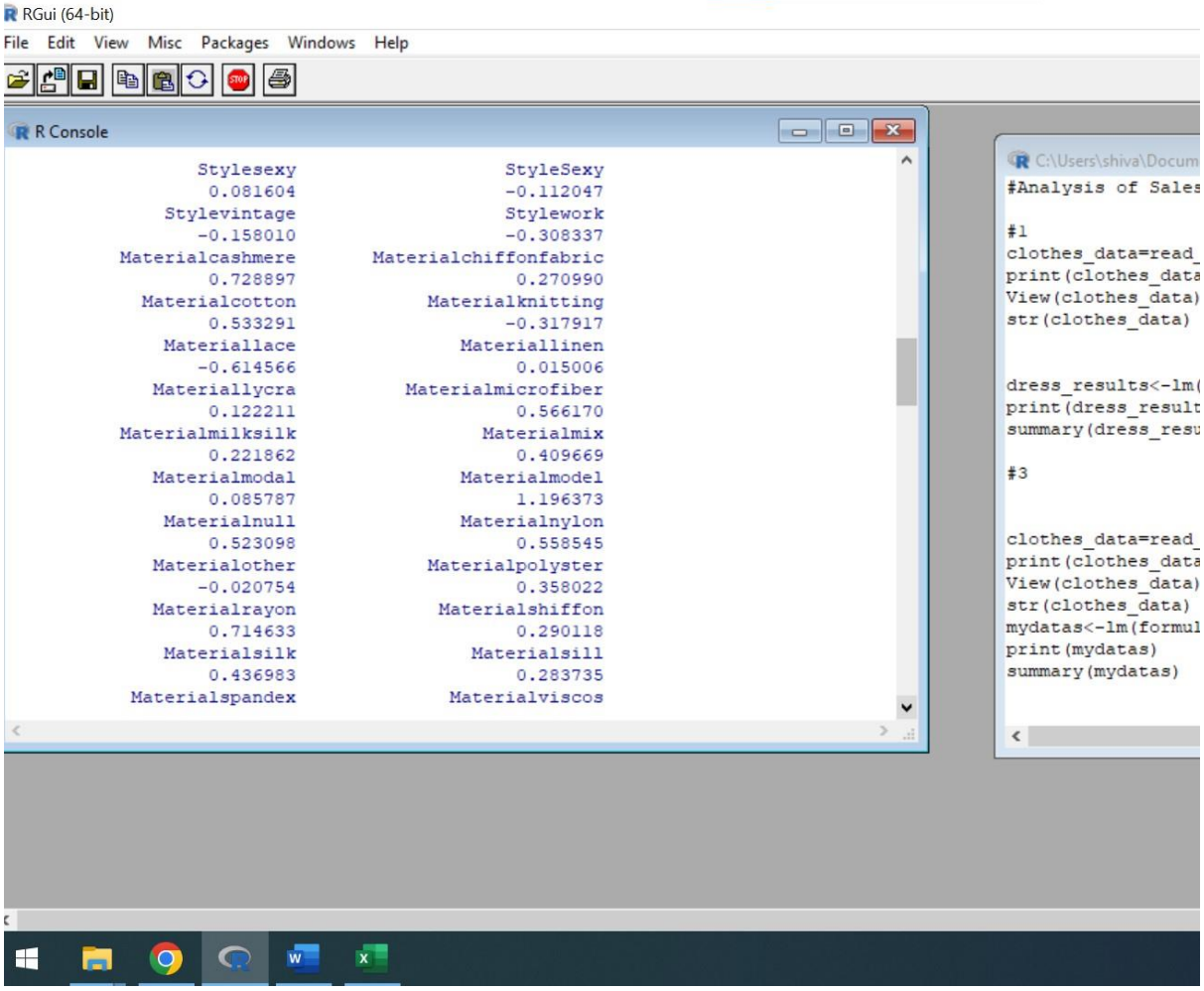
The second screenshot represents the tabular view of dress attributes

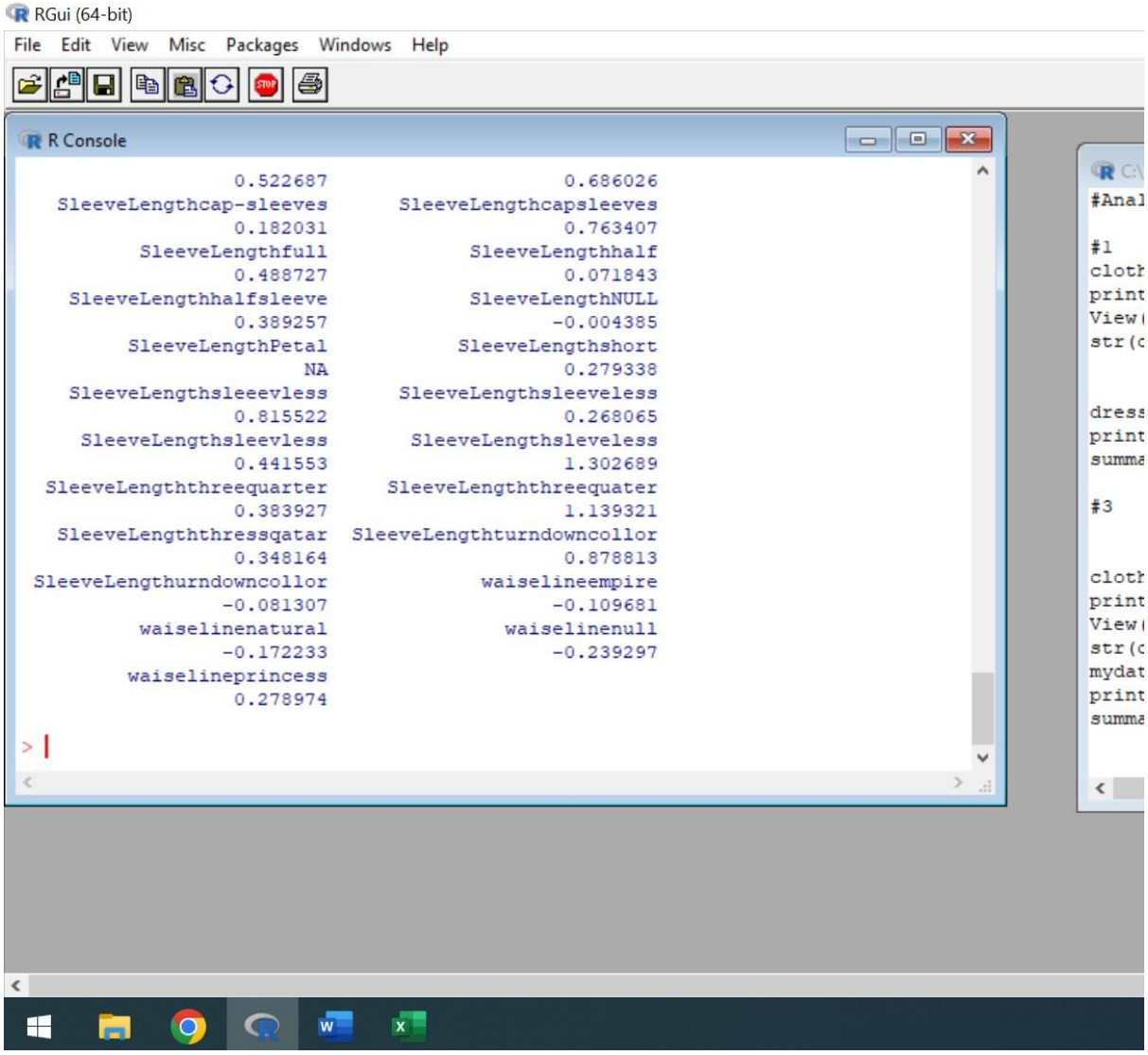


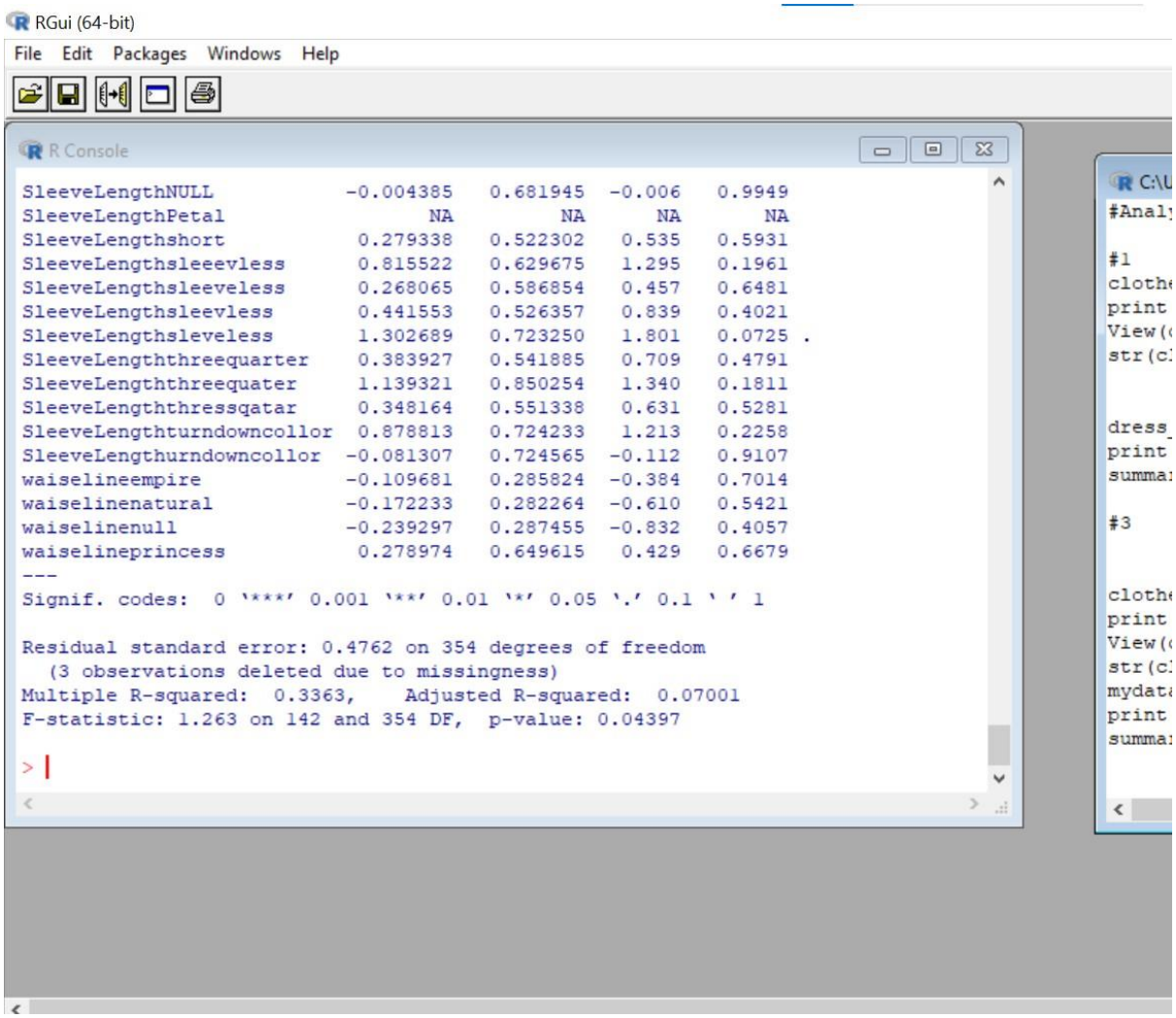
	Dress_ID	Style	Price	Rating	Size	Season	NeckLine
1	1006032852	Sexy	Low	4.6	M	Summer	o-neck
2	1212192089	Casual	Low	0.0	L	Summer	o-neck
3	1190380701	vintage	High	0.0	L	Automn	o-neck
4	966005983	Brief	Average	4.6	L	Spring	o-neck
5	876339541	cute	Low	4.5	M	Summer	o-neck
6	1068332458	bohemian	Low	0.0	M	Summer	v-neck
7	1220707172	Casual	Average	0.0	XL	Summer	o-neck
8	1219677488	Novelty	Average	0.0	free	Automn	o-neck
9	1113094204	Flare	Average	0.0	free	Spring	v-neck
10	985292672	bohemian	Low	0.0	free	Summer	v-neck
11	1117293701	party	Average	5.0	free	Summer	o-neck
12	898481530	Flare	Average	0.0	free	Spring	v-neck
13	957723897	sexy	Low	4.7	M	Winter	o-neck
14	749031896	vintage	Average	4.8	M	Summer	o-neck
15	1055411544	Casual	Low	5.0	M	Summer	boat-neck
16	1162628131	Casual	Low	0.0	free	Winter	boat-neck
17	624314841	cute	Average	4.7	L	spring	o-neck
18	830467746	bohemian	Medium	5.0	free	Automn	o-neck
19	840857118	Brief	Average	0.0	M	Winter	peterpan-collor

For expanding the sales of the product,a high end fashion retail store has to follow styling,fabric type.material,decoration and other attributes wrt to season. Here, the dependent variable is Recommendation while season,style,material,Fabric Type and other summation fields are independent variables


RGui (64-bit)
File Edit View Misc Packages Windows Help
R Console
> print(dress_results)
Call:
lm(formula = Recommendation ~ Season + Style + Material + FabricType +
Decoration + PatternType + Size + NeckLine + SleeveLength +
waiseline, data = clothes_data)
Coefficients:
(Intercept) SeasonAutumn
-0.234403 -0.186089
Seasonspring SeasonSpring
0.243178 0.195141
Seasonsummer SeasonSummer
-0.514474 -0.050809
Seasonwinter SeasonWinter
0.172935 0.032913
StyleBrief StyleCasual
-0.283975 -0.115842
Stylecute Stylefashion
-0.002484 -1.300627
StyleFlare StyleNovelty
-0.147997 0.030050
StyleOL Styleparty
0.183743 -0.059055
Stylesexy StyleSexy


RGui (64-bit)
File Edit View Misc Packages Windows Help
R Console
Stylesexy StyleSexy
0.081604 -0.112047
Stylevintage Stylework
-0.158010 -0.308337
Materialcashmere Materialchiffonfabric
0.728897 0.270990
Materialcotton Materialknitting
0.533291 -0.317917
Materiallace Materiallinen
-0.614566 0.015006
Materiallycra Materialmicrofiber
0.122211 0.566170
Materialmilksilk Materialmix
0.221862 0.409669
Materialmodal Materialmodel
0.085787 1.196373
Materialnull Materialnylon
0.523098 0.558545
Materialother Materialpolyster
-0.020754 0.358022
Materialrayon Materialshiffon
0.714633 0.290118
Materialsilk Materialsill
0.436983 0.283735
Materialspandex Materialviscos
#Analysis of Sales
#1
clothes_data=read_
print(clothes_data
View(clothes_data)
str(clothes_data)
dress_results<-lm(
print(dress_result
summary(dress_resu
#3
clothes_data=read_
print(clothes_data
View(clothes_data)
str(clothes_data)
mydatas<-lm(formul
print(mydatas)
summary(mydatas)


RGui (64-bit)
File Edit View Misc Packages Windows Help
R Console
0.522687 0.686026
SleeveLengthcap-sleeves SleeveLengthcapsleeves
0.182031 0.763407
SleeveLengthfull SleeveLengthhalf
0.488727 0.071843
SleeveLengthhalfsleeve SleeveLengthNULL
0.389257 -0.004385
SleeveLengthPetal SleeveLengthshort
NA 0.279338
SleeveLengthsleeevless SleeveLengthsleeveless
0.815522 0.268065
SleeveLengthsleevless SleeveLengthsleveless
0.441553 1.302689
SleeveLengththreequarter SleeveLengththreequater
0.383927 1.139321
SleeveLengththressqatar SleeveLengthturndowncollor
0.348164 0.878813
SleeveLengthurndowncollor waiselineempire
-0.081307 -0.109681
waiselinenatural waiselinenull
-0.172233 -0.239297
waiselineprincess
0.278974
>
#Anal
#1
cloth
print
View
str(c
dress
print
summa
#3
cloth
print
View
str(c
mydat
print
summa



Residual is the difference between the dependent variable and independent variable. By applying above mentioned formula one can analyse the 70% of values fit to the model because the adjusted R-squared is 0.07001 and residual standard error is also 0.47

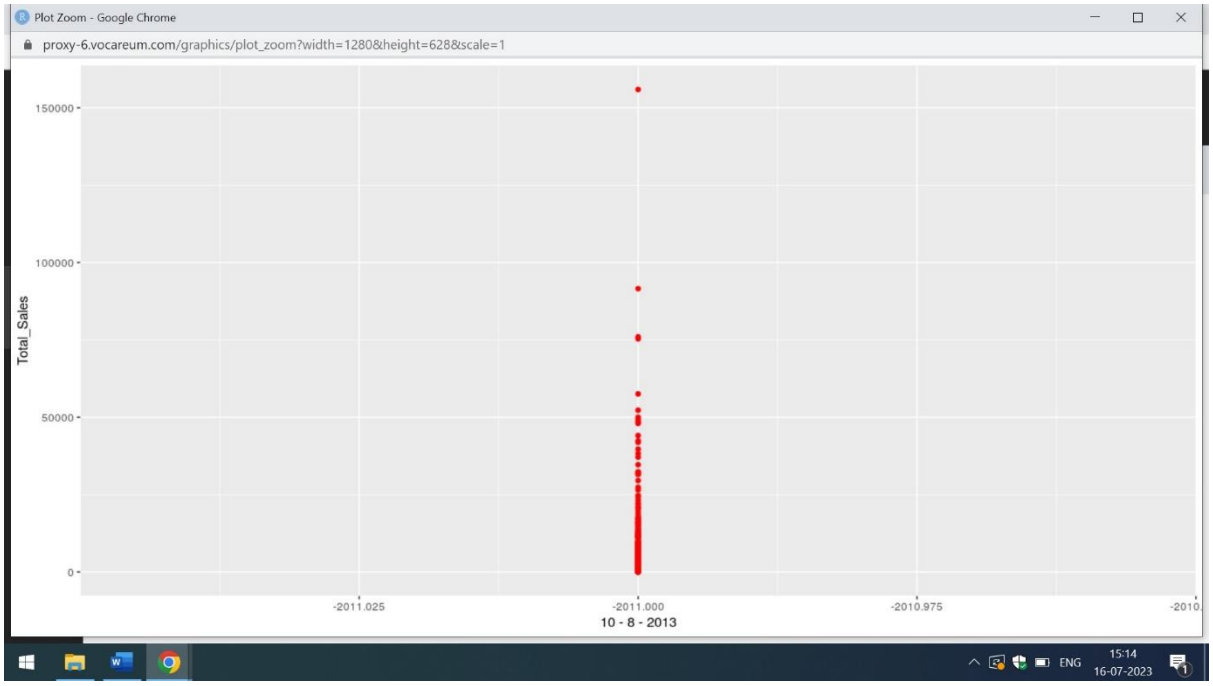
Q2) In order to stock the inventory, the store wants to analyze the sales data and predict the trend of total sales for each dress for an extended period of three more alternative days.

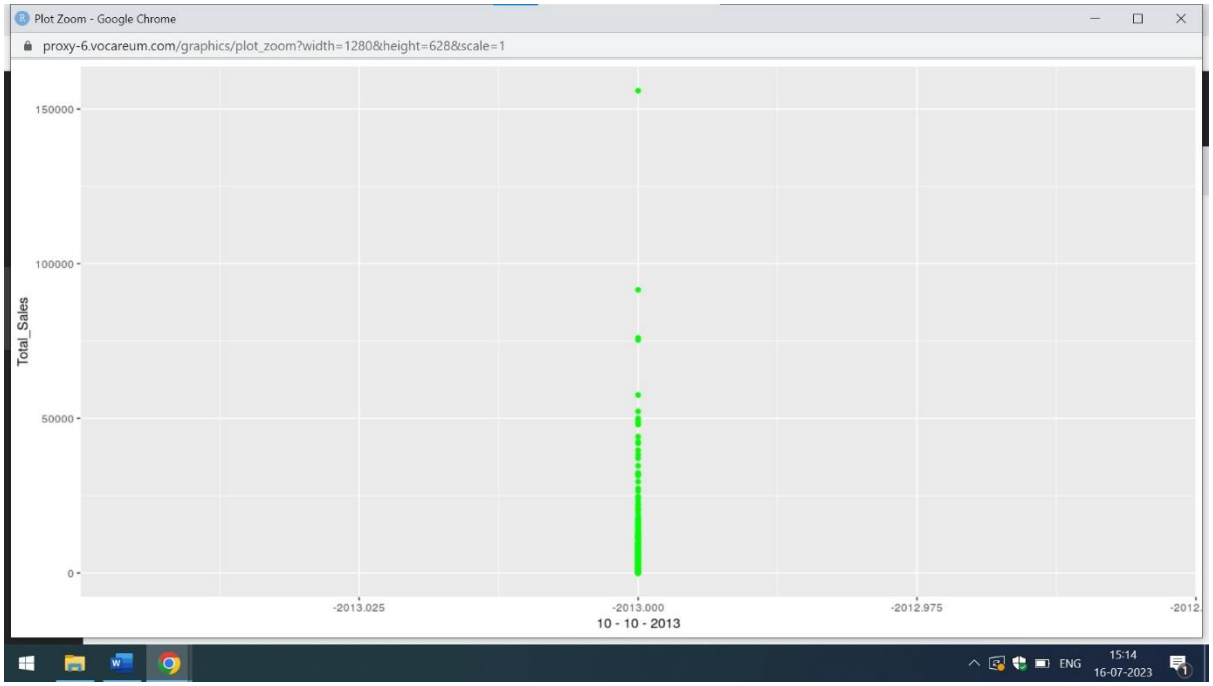
Ans) For analysing the sales data and predicting the trend of total sales for each dress for an extended period, I am using the concept of ggplot. Ggplot in R is helpful in data visualization. The below attached are the ggplot of Total Sales for October Month. The dates 8,10 and 12 are taken in consideration. Below mentioned are the code in R

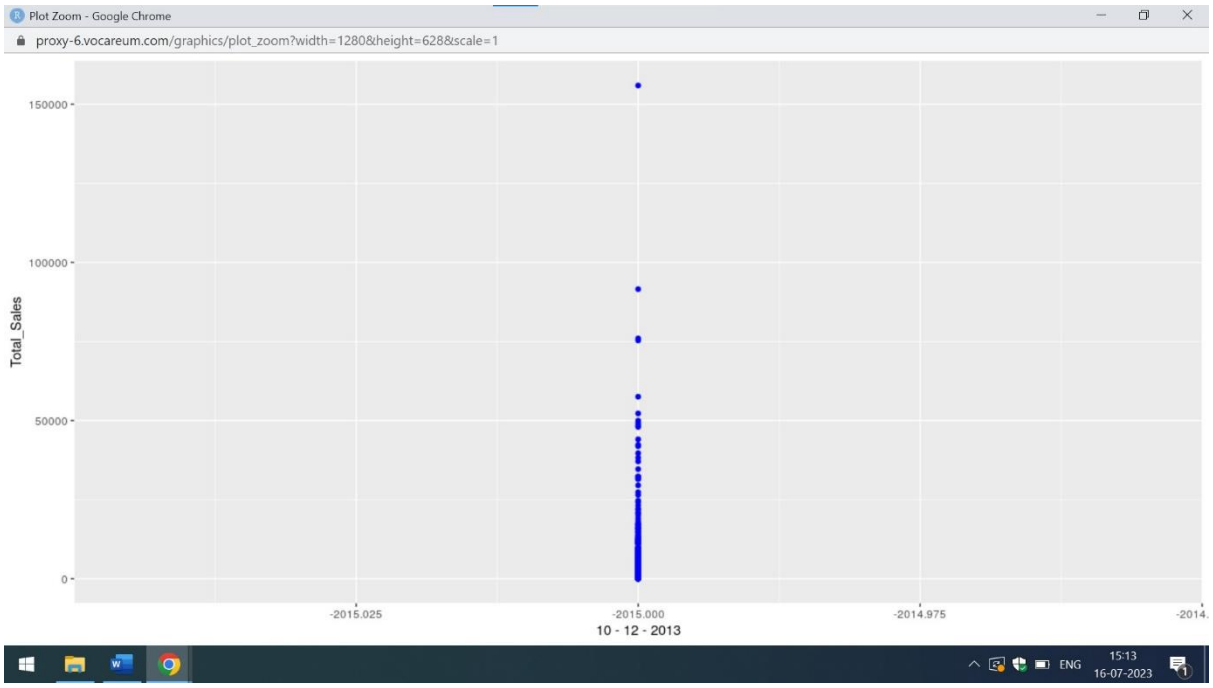
```
dress_data<-
read.csv("https://raw.githubusercontent.com/shivanipriya89/mydress/main/Dress%20Sales.csv")

print(dress_data)

View(dress_data)
```

```
myplot<-ggplot(data=dress_data,mapping=aes(x=10-12-2013,y=Total_Sales))
myplot+geom_point(color="blue")

myplot<-ggplot(data=dress_data,mapping=aes(x=10-10-2013,y=Total_Sales))
myplot+geom_point(color="green")

myplot<-ggplot(data=dress_data,mapping=aes(x=10-08-2013,y=Total_Sales))
myplot+geom_point(color="red")
```

Below attached are the ggplot of 8th,10th and 12th October 2013







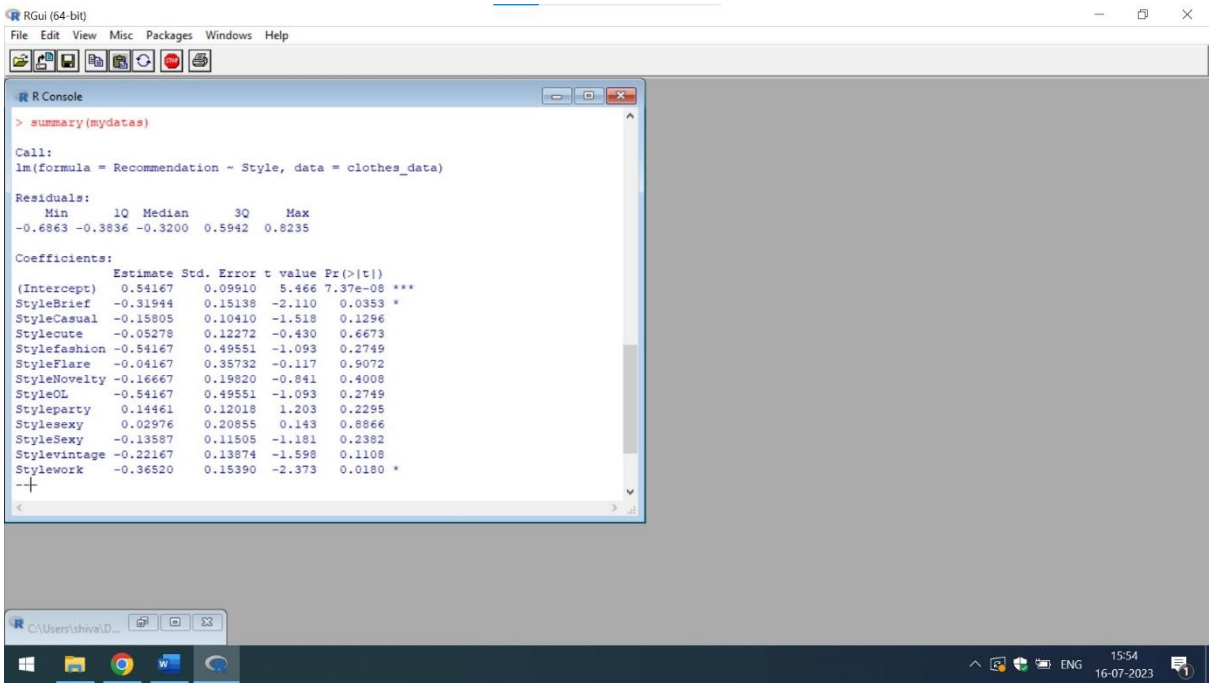
From,the above mentioned plots it is clear that total sales is increasing for the consecutive days

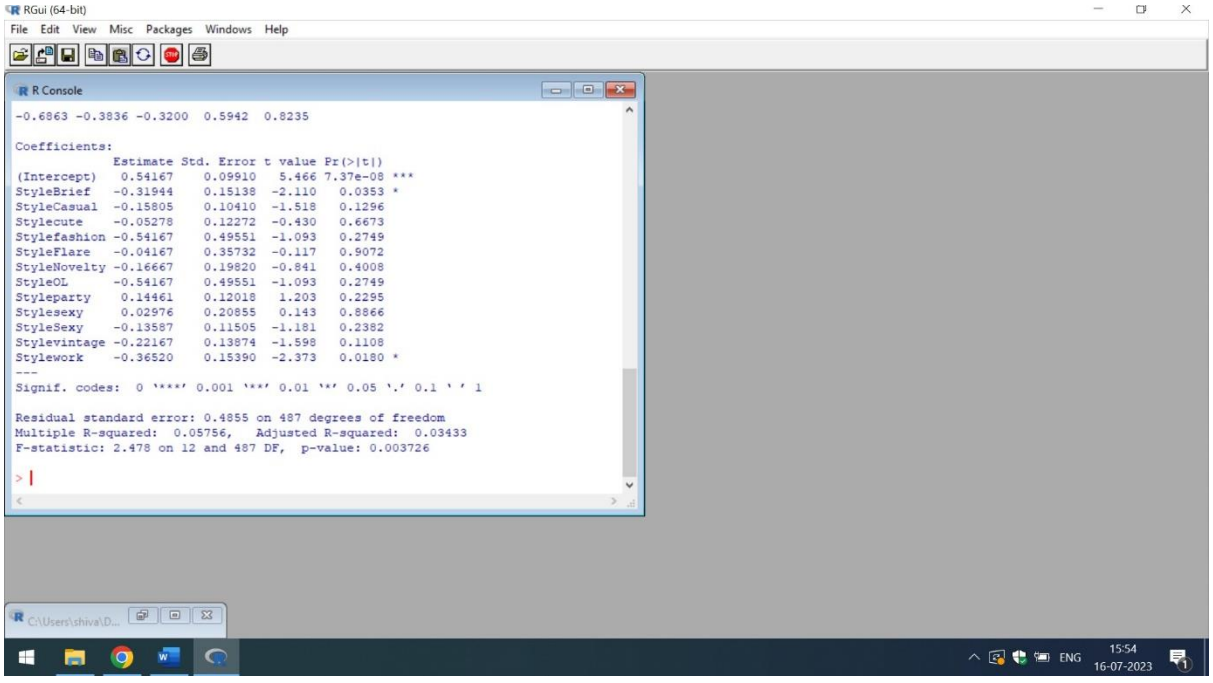
Q3) To decide the pricing for various upcoming clothes, they wish to find how the style, season, and material affect the sales of a dress and if the style of the dress is more influential than its price.

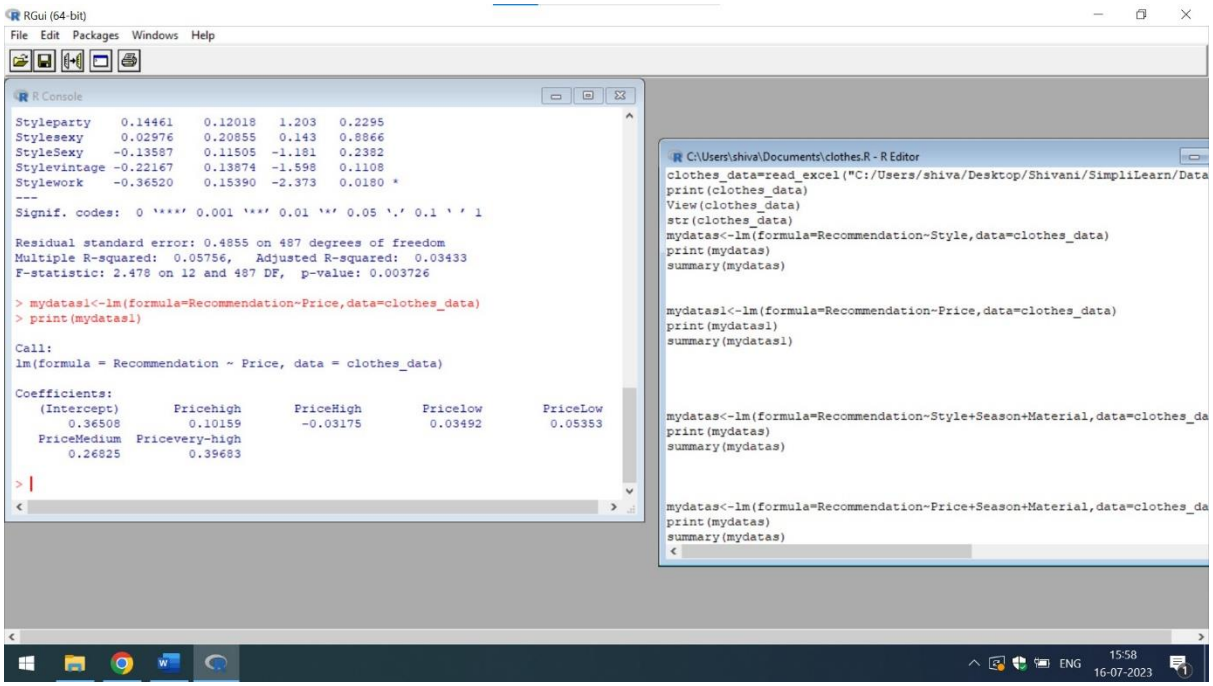
Ans) For deciding the pricing for various upcoming clothes,style,season and material attributes plays an important role.Here, I am using the concept of Simple Linear Regression for the analysis of Style and Price attributes on dependent variable ie Recommendation

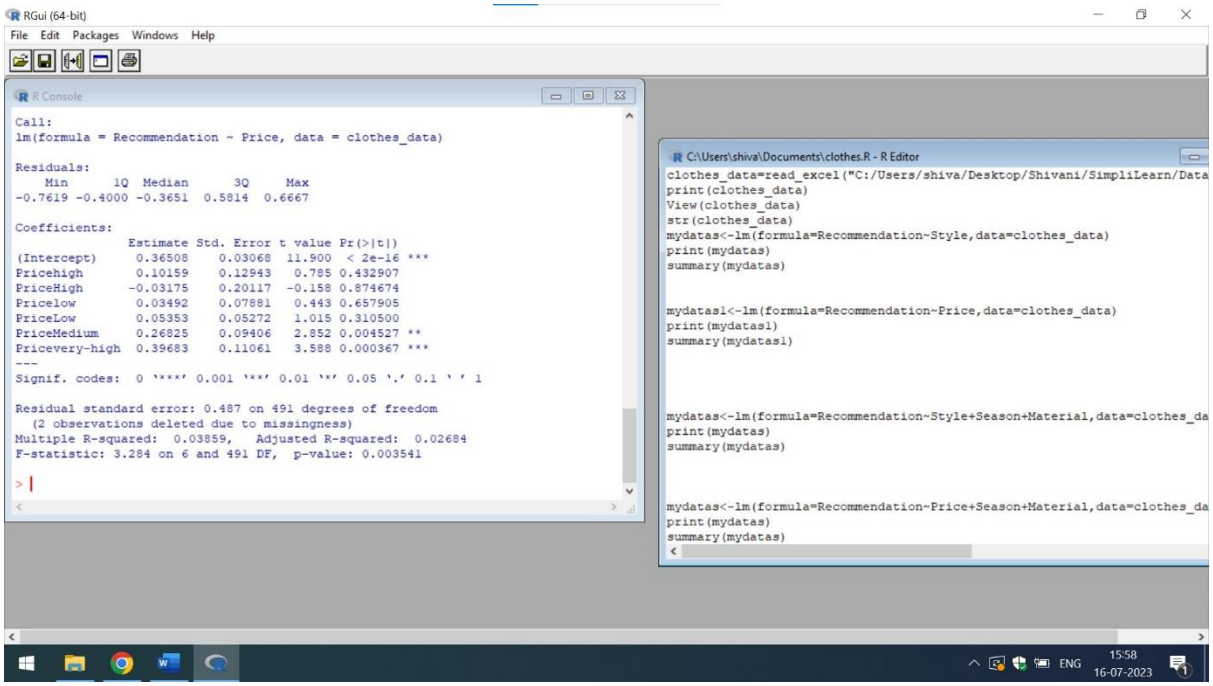
```
clothes_data=read_excel("C:/Users/shiva/Desktop/Shivani/SimpliLearn/Data Science with R/Attributes.xlsx")

print(clothes_data)

View(clothes_data)

str(clothes_data)

mydatas<-lm(formula=Recommendation~Style,data=clothes_data)

print(mydatas)

summary(mydatas)



mydatas1<-lm(formula=Recommendation~Price,data=clothes_data)

print(mydatas1)

summary(mydatas1)



mydatas<-lm(formula=Recommendation~Style+Season+Material,data=clothes_data)

print(mydatas)

summary(mydatas)



mydatas<-lm(formula=Recommendation~Price+Season+Material,data=clothes_data)

print(mydatas)

summary(mydatas)
```

The below attached is the output of the Linear Regression ie Recommendation on Style vs Recommendation on Price





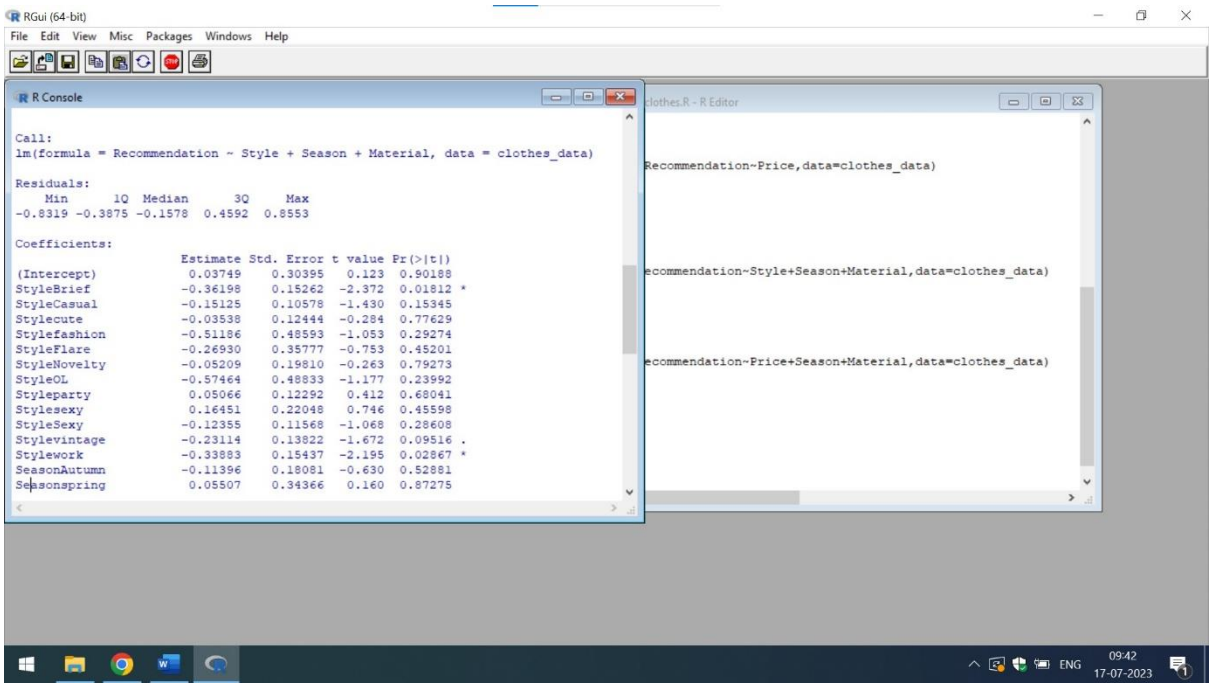


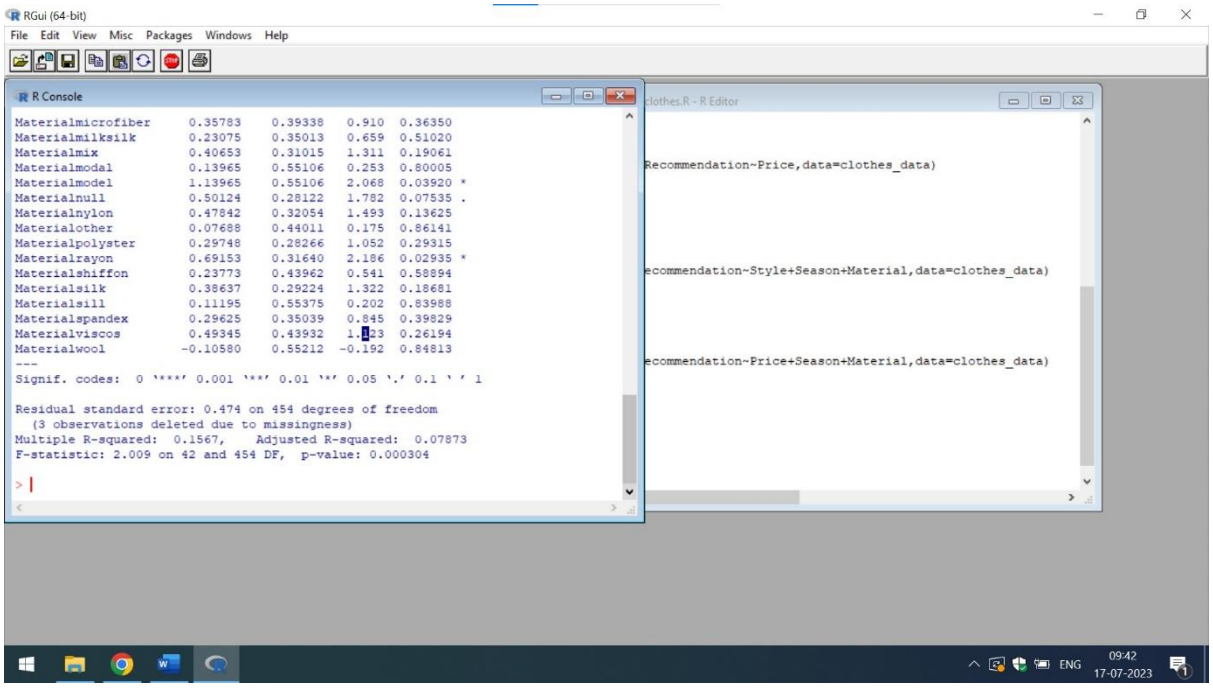


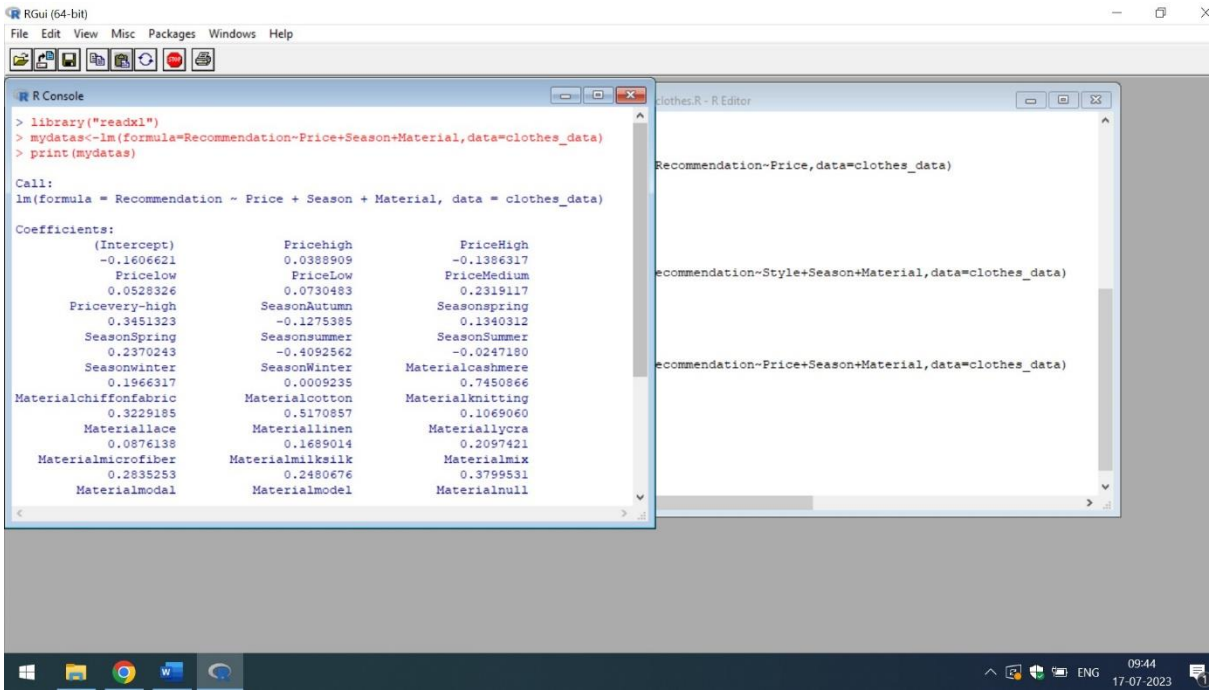
From,the above mentioned screenshots,it is clear that residual standard error of Recommendation vs Style is 0.4855 on 487 degrees of freedom. While the residual standard error of Recommendation vs Price is 0.487 on 491 degrees of freedom. Smaller residual standard means predictions are better. Hence, the style of the dress is more influential than its price. Moreover, the adjusted R square of Recommendation on Style is 0.3433 which shows

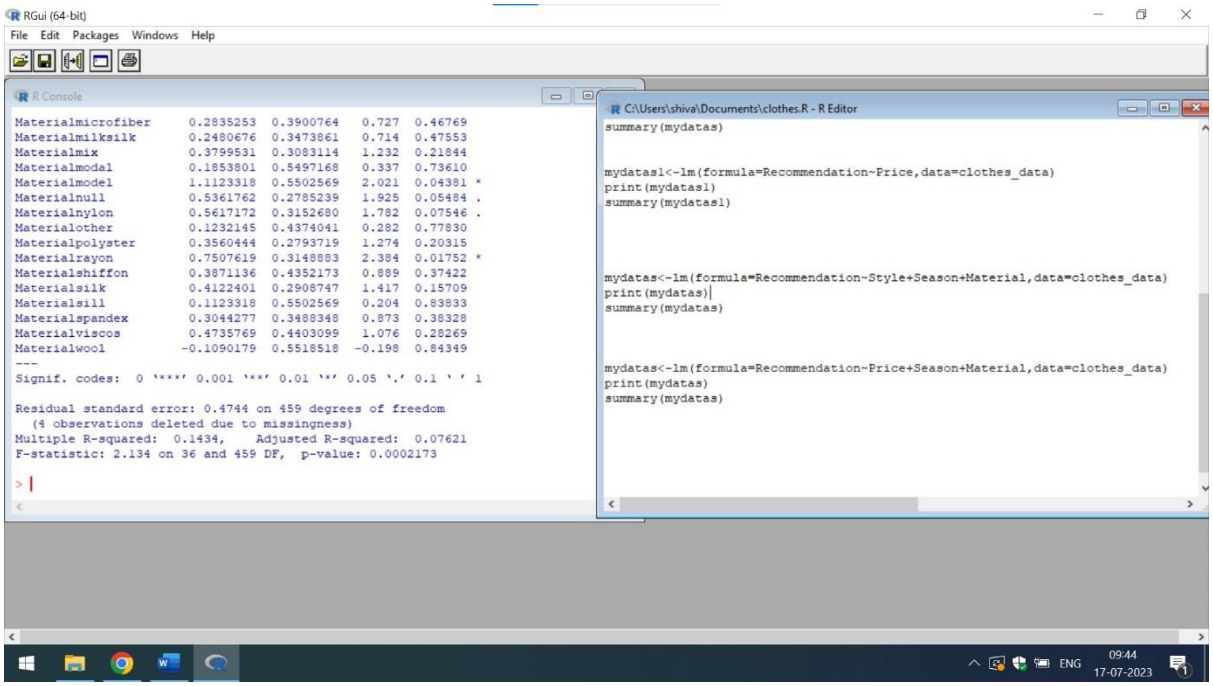
34% of values fit to the model and the adjusted R square of Recommendation on Price is 0.268 ie almost 27% of values fit to the model. Hence, from both analysis it is clear that style of the dress is more influential than its price

The below attached is the output of the Linear Regression ie Recommendation on Style vs Recommendation on Price with multiple attributes









It is also very clear that from the simple Linear Regression with multiple attributes that Adjusted R square of Recommendation vs Style with multiple attributes is 0.0787 which shows that almost 78% of values fit to the model but the adjusted R square of Recommendation vs Price with multiple attributes is 0.07621 which indicates that 76% of values fit to the model. Hence, from both analysis it is clear that style of the dress is more influential than its price

Q4) Also, to increase sales, the management wants to analyze the attributes of dresses and find which are the leading factors affecting the sale of a dress.

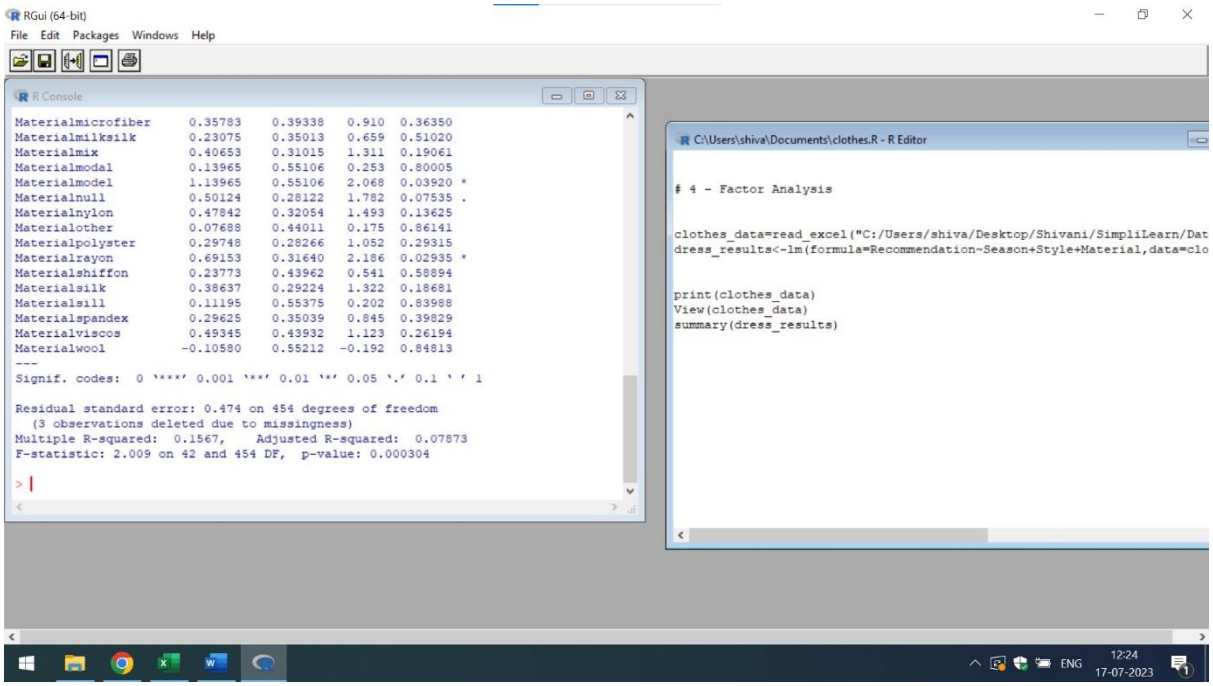
Ans) The Decision Tree Algorithm helps in analysing the attributes of dresses. It even gives an idea in discovering the factors affecting the sales of a dress. Simple Linear Regression with multiple attributes gives an idea of the impact of independent variables on dependent.

The below attached is the code in R of impact of Simple Linear Regression with Multiple variables including the decision tree algorithm

```
clothes_data=read.csv("https://raw.githubusercontent.com/shivanipriya89/Dress/main/Attry.csv")

dress_results<-lm(formula=Recommendation~Season+Style+Material,data=clothes_data)



print(clothes_data)

View(clothes_data)

str(clothes_data)


png(file = "decision_tree.png")


# Create the tree.

output.tree <- ctree(Recommendation~Season+Style+Material,data = clothes_data)


# Plot the tree.

plot(output.tree)


# Save the file.

dev.off()
```

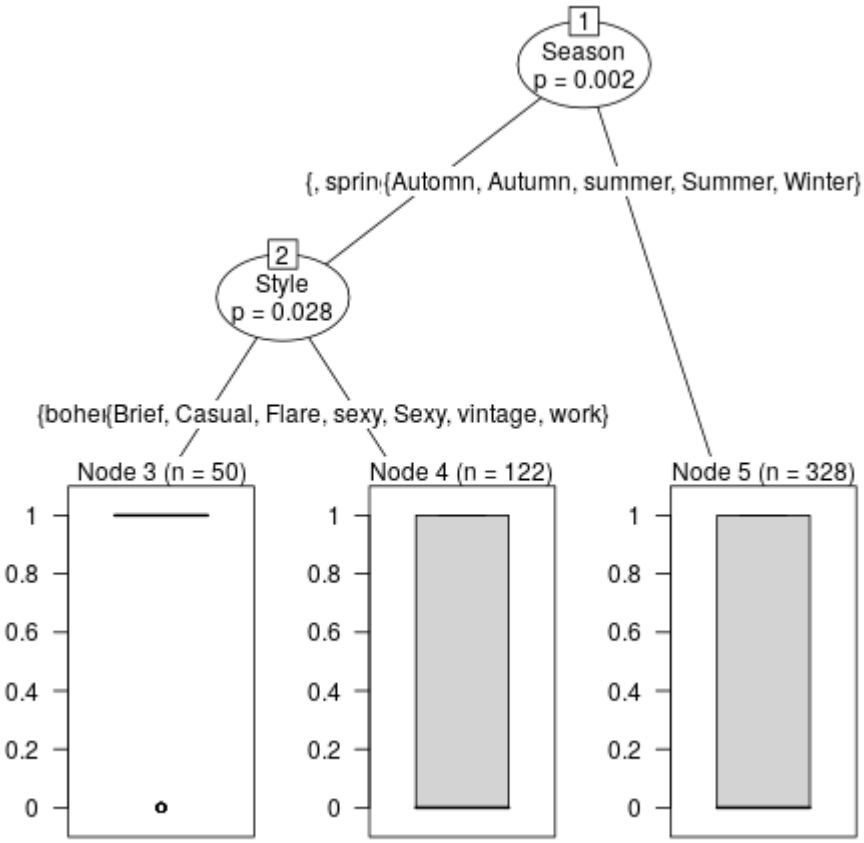
The below attached is the output of

```
dress_results<-lm(formula=Recommendation~Season+Style+Material,data=clothes_data)
```



From the above mentioned output of the Simple Linear Regression with Multiple Attributes

It is clear that season,style and material are the most important factors in increasing sales. The adjusted R-square is 0.0787 which shows that 79% of values fit to model. The decision tree of Recommendation on Season,Style and Material is mention below



From the above mention graph,it is clear that casual,brief and bohemian are not the good parameters of style of Autumn and spring season. Hence, we can ignore these values

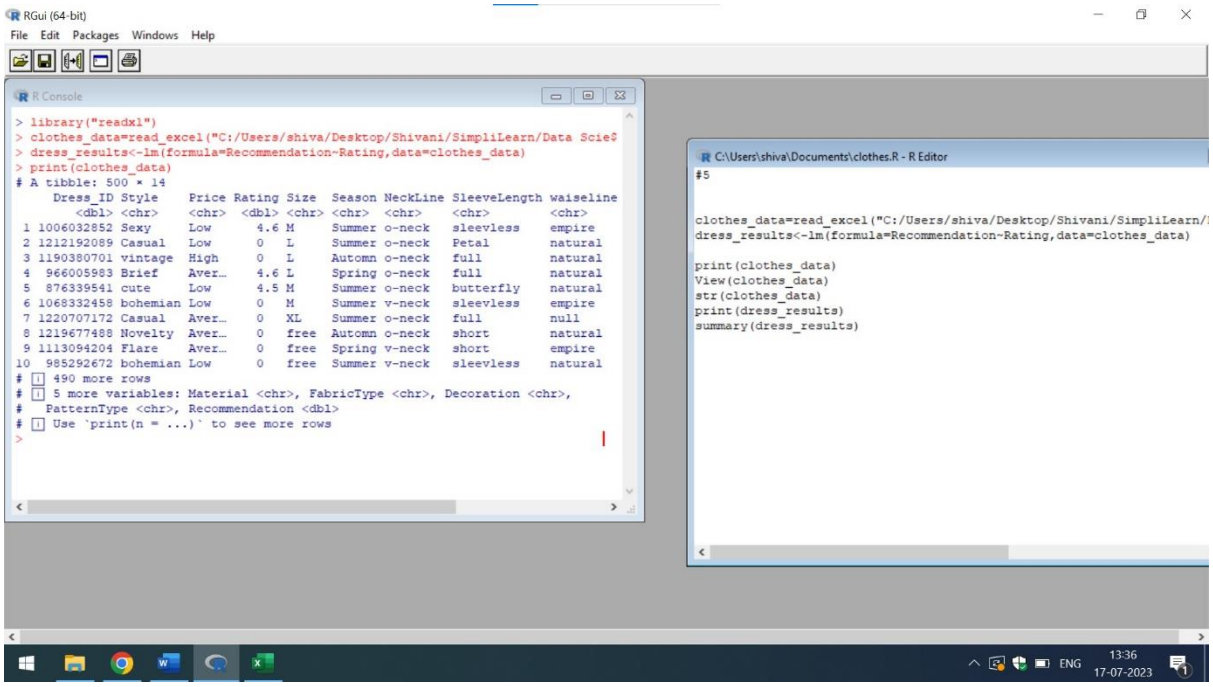
Q5) To regularize the rating procedure and find its efficiency, the store wants to find if the rating of the dress affects the total sales

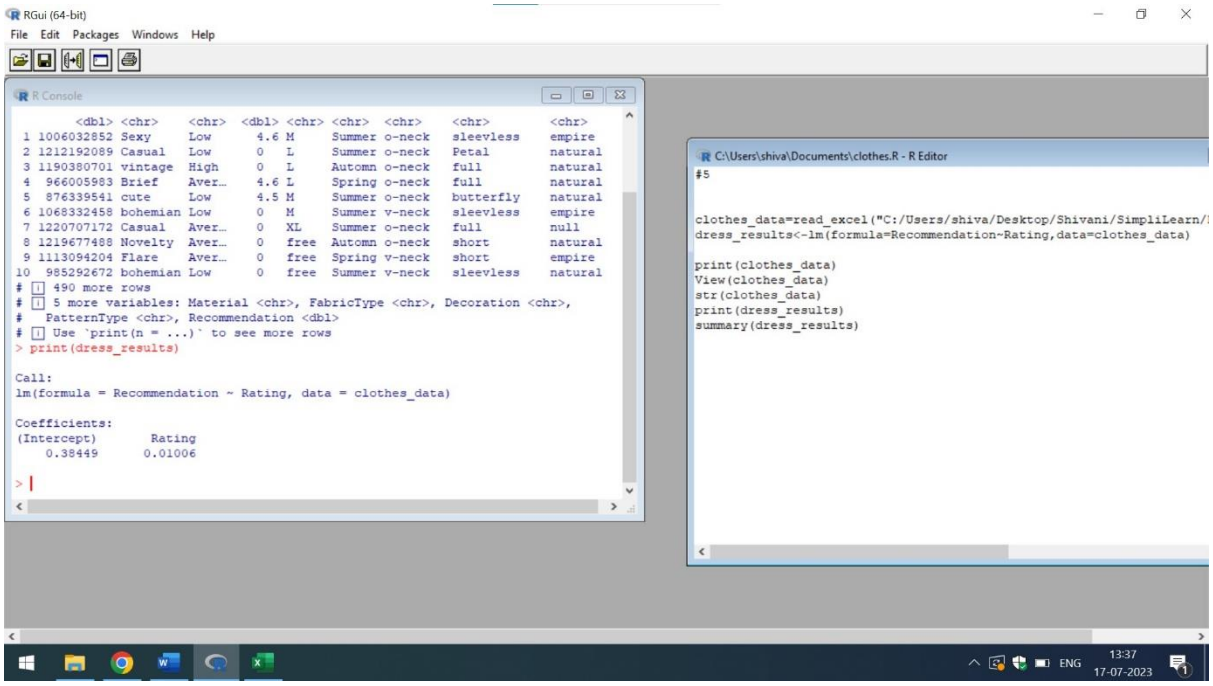
Ans) For analysing the impact of rating,we can use the concept of Simple Linear Regression ie impact of Ratings on Recommendation. The code attached below written in R

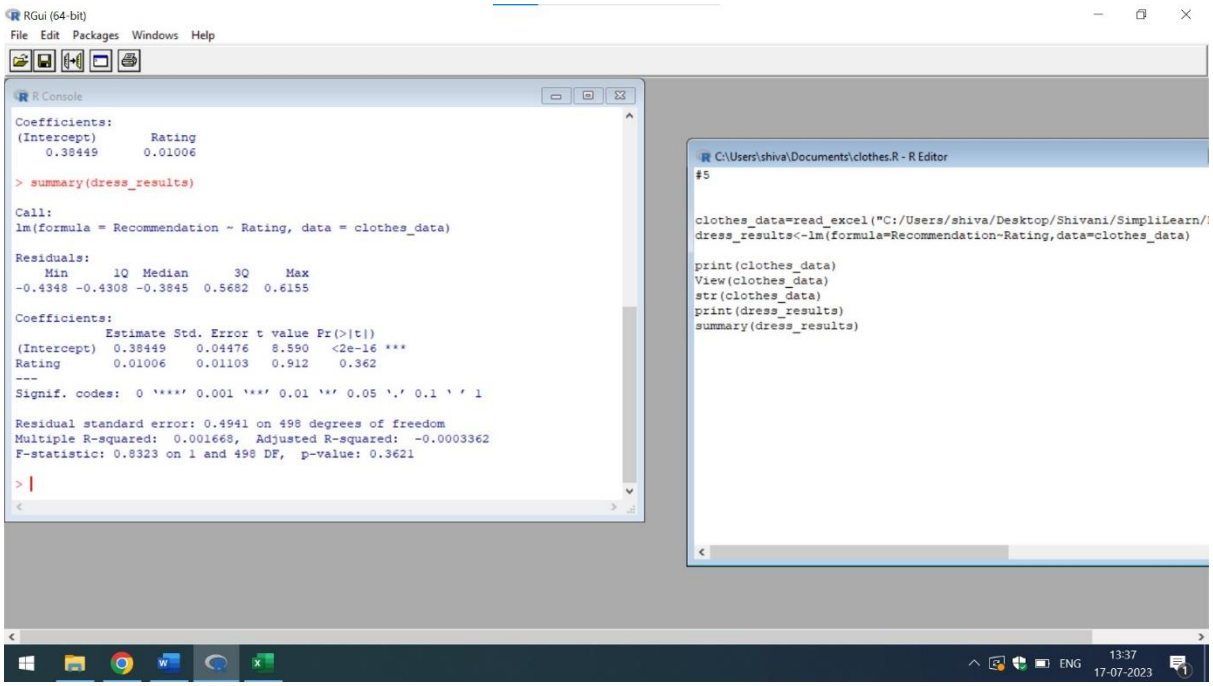
```
clothes_data=read_excel("C:/Users/shiva/Desktop/Shivani/SimpliLearn/Data Science with R/Attributes.xlsx")

dress_results<-lm(formula=Recommendation~Rating,data=clothes_data)


print(clothes_data)

View(clothes_data)

str(clothes_data)

print(dress_results)
```

summary(dress_results)

The below attached is the output of the dress results







It is very clear from the simple Linear Regression that Rating has no impact on Recommendation because the value of adjusted R square is in negative digits that indicates none of the values fit to the model even the values of the residuals are in negative digits.

Hence we can assume rating is an independent variable and has no impact on Recommendation, hence rating has no impact on sales also