

Project Report on the Project Retail

Business Scenario: The business is a cloth selling retail store in which we are required to find the current trend and the relation between the attributes of clothing with respect to the total sales for each product id. We are provided with the two files in excel format in which one contains the attributes like fashion, style, rating, etc, while other file is a sales data of alternate days.

Expectation/Goal: There are five objective which we need to achieve which are:

1. We need to find a model to predict the recommendation value in binary form (1 and 0) by analysis which factor affects the recommendation and influence.
2. We need to stock the inventory and hence need a forecasting model by which we get to know the predicted value for the Total sales for the next 3 alternative days.
3. We need to decide if the Pricing of clothes for future and hence we need a model which tells the relation of style, season and material on the sales of clothing and also find if the style is influencing the sales more than the price is doing.
4. We need to increase sales and for that we need a model which can tell about the influencing factors from all the attributes which are provided.
5. To regularize and find the efficiency of the rating procedure we need to find if the rating is influencing the total sales.

The Project can be done by following the CRISP DM methodology.

In this project the entries are very low and for a good model the entries should be more but performing the process in the limited data which is provided.

CODE:

```
#Project retail
```

```
library("caTools")
```

```
library("forecast")
```

```
library("tseries")
```

```
library("corrplot")
```

```
library("arules")
```

```
Data<-read.csv(file.choose(),header = T)
```

```
head(Data)
```

```
names(Data)
```

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

#Removed the First column as it is not required for the model for objective 1

```
Question1<-Data[,-1]
```

```
head(Question1)
```

```
class(Question1$Recommendation)
```

```
summary(Question1)
```

```
levels(Question1$Style)
```

#Model for this objective 1

```
model<-glm(Recommendation~Style + Price + Rating + Size + Season + NeckLine +
```

```
SleeveLength + waiseline + Material + FabricType + Decoration + Pattern.Type,Question1 ,family =  
"gaussian")
```

```
summary(model)
```

#For testing and validating if the model is good or not and is applicable to any random data or is just confined to this data set we do the following steps but since in the objective it is not asked to test and validate i will not be running the below codes

#Splitting data into 75%train and 25%test

```
set.seed(1234)
```

```
split1 <- sample.split(Question1,SplitRatio = 0.75)
```

```
train1 <- subset(Question1,split == TRUE)
```

```
test1 <- subset(Question1,split == FALSE)
```

```
levels(train1$Style)
```

```
levels(test1$Style)
```

#making a column for prediction of recommendation

```
test1$recomendpred<-0
```

```
#making logistic model to perform on train1 data
```

```
model1<-glm(Recommendation~Style + Price + Rating + Size + Season + NeckLine +  
            SleeveLength + waiseline + Material + FabricType + Decoration + Pattern.Type,train1,family =  
            "gaussian")
```

```
summary(model1)
```

```
1-pchisq(model1$null.deviance-model1$deviance,355-221)
```

```
table(test1$Style)
```

```
predict1<-predict(model1,test1,type="response")
```

```
class(test1$Style)
```

```
levels(test1$Style)
```

```
levels(train1$Style)
```

```
class(train1$Style)
```

```
head(test1$Style)
```

```
#taking cutoff as 0.7 for the probability
```

```
test1$recomendpred <- ifelse(predicted>=0.7,1,0)
```

```
test1$recomendpred
```

```
#testing by confusion matrix
```

```
table(test1$Recommendation,test1$recomendpred)
```

```
tab<-table(test1$Recommendation,test1$recomendpred)
```

```
#accuracy
```

```
sum(tab)
```

```
sum(diag(tab))/sum(tab)
```

```
#error
```

```
1-sum(diag(tab))/sum(tab)
```

#Objective 2: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.

```
Sales<-read.csv(file.choose(),header = T)
```

```
head(Sales)
```

```
timeseries <- ts(Sales, start = 1, frequency = 7)
```

```
plot(timeseries)
```

```
ArimaModel <- auto.arima(timeseries)
```

```
summary(ArimaModel)
```

```
forecast(ArimaModel,3)
```

```
plot(forecast(ArimaModel,3))
```

#Objective 3: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.

```
Data3<-read.csv(file.choose(),header = T)
```

```
head(Data3)
```

```
class(Data3$Total.Sales)
```

```
OutStyle<-aov(Data3$Total.Sales~Data3$Style)
```

```
Outseason<-aov(Data3$Total.Sales~Data3$Season)
```

```
Outmaterial<-aov(Data3$Total.Sales~Data3$Material)
```

```
summary(OutStyle)
```

```
summary(Outseason)
```

```
summary(Outmaterial)
```

```
model3a<-lm(Total.Sales~Style+Season+Material,data = Data3)
```

```
summary(model3a)
```

```
model3b<-lm(Total.Sales~Price+Style,data = Data3)
```

```
summary(model3b)
```

#Objective4:Also, to increase the sales, the management wants to analyze the attributes

#of dresses and find which are the leading factors affecting the sale of a

```
#dress.
```

```
model4<-lm(Total.Sales~.,data = Data3)
```

```
summary(model4)
```

```
#objective5:To regularize the rating procedure and find its efficiency, the store wants to
```

```
#find if the rating of the dress affects the total sales.
```

```
model5<-cor(Data3$Total.Sales,Data3$Rating)
```

```
model5
```

```
names(Data3)
```

```
Data3[,c(4,15)]
```

```
c<-cor(Data3[,c(4,15)])
```

```
corrplot(c)
```

Steps involved with the Screenshot of results

1.The Attribute data set is converted to CSV format.

2. Calling the packages from the library required.

3.reading the Attribute file which is in csv format and data manipulation.

```
Console Terminal x
~/
> #Project retail
> library("caTools")
> library("forecast")
> library("tseries")
> library("corrplot")
> library("arules")
> Data<-read.csv(file.choose(),header = T)
> head(Data)
  Dress_ID Style Price Rating Size Season NeckLine SleeveLength waiseline
1 1006032852 Sexy Low 4.6 M Summer o-neck sleeveless empire
2 1212192089 Casual Low 0.0 L Summer o-neck Petal natural
3 1190380701 vintage High 0.0 L Autumn o-neck full natural
4 966005983 Brief Average 4.6 L Spring o-neck full natural
5 876339541 cute Low 4.5 M Summer o-neck butterfly natural
6 1068332458 bohemian Low 0.0 M Summer v-neck sleeveless empire
  Material FabricType Decoration Pattern.Type Recommendation
1 null chiffon ruffles animal 1
2 microfiber null ruffles animal 0
3 polyester null null print 0
4 silk chiffon embroidery print 1
5 chiffonfabric chiffon bow dot 0
6 null null null print 0
> names(Data)
[1] "Dress_ID" "Style" "Price" "Rating" "Size"
[6] "Season" "NeckLine" "SleeveLength" "waiseline" "Material"
[11] "FabricType" "Decoration" "Pattern.Type" "Recommendation"
```

```

Console Terminal
~/Desktop/Week 4
> head(question1)
  Style Price Rating Size Season NeckLine SleeveLength waiseline Material
1 Sexy Low 4.6 M Summer o-neck sleeveless empire null
2 Casual Low 0.0 L Summer o-neck Petal natural microfiber
3 vintage High 0.0 L Autumn o-neck full natural polyester
4 Brief Average 4.6 L Spring o-neck full natural silk
5 cute Low 4.5 M Summer o-neck butterfly natural chiffonfabric
6 bohemian Low 0.0 M Summer v-neck sleeveless empire null
FabricType Decoration Pattern.Type Recommendation
1 chiffon ruffles animal 1
2 null ruffles animal 0
3 null null print 0
4 chiffon embroidery print 1
5 chiffon bow dot 0
6 null null print 0
> class(question1$Recommendation)
[1] "integer"
> summary(question1)
      Style      Price      Rating      Size      Season
Casual :232 Average :252 Min. :0.000 free :173 Summer :159
Sexy : 76 Low :129 1st Qu.:3.700 L : 96 Spring :122
party : 51 Low : 45 Median :4.600 M :177 winter : 99
cute : 45 Medium : 30 Mean :3.529 S : 1 Autumn : 61
vintage : 25 very-high: 21 3rd Qu.:4.800 S : 37 winter : 46
bohemian: 24 high : 15 Max. :5.000 small: 1 Autumn : 8
(other) : 47 (other) : 8 XL :15 (other): 5
      NeckLine      SleeveLength      waiseline      Material
o-neck :271 sleeveless :223 : 1 cotton :152
v-neck :124 full : 97 dropped : 4 null :127
slash-neck : 25 short : 96 empire :104 polyester : 99
boat-neck : 19 halfsleeve : 35 natural :304 silk : 26
Sweetheart : 14 threequarter: 17 null : 86 chiffonfabric: 25
turndowncollar: 13 thressqatar : 10 princess: 1 mix : 12
(other) : 34 (other) : 22 (other) : 59
      FabricType      Decoration      Pattern.Type      Recommendation
null :265 null :235 solid :203 Min. :0.00
Chiffon :135 lace : 70 null :108 1st Qu.:0.00
broadcloth: 31 sashes : 42 print : 71 Median :0.00
worsted : 19 beading : 22 patchwork: 48 Mean :0.42
jersey : 12 applique : 21 animal : 21 3rd Qu.:1.00

```

4. Running the model

```

Console Terminal
~/Desktop/Week 4
> model<-glm(Recommendation ~ Style + Price + Rating + Size + Season + NeckLine + SleeveLength + waiseline + Material + FabricType + Decoration + Pattern.Type
,question1,family = "gaussian")
> summary(model)

Call:
glm(formula = Recommendation ~ Style + Price + Rating + Size +
    Season + NeckLine + SleeveLength + waiseline + Material +
    FabricType + Decoration + Pattern.Type, family = "gaussian",
    data = question1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9907  -0.3228   0.0000   0.3124   0.8834

Coefficients: (7 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.208892   0.878935   0.238  0.8123
StyleBrief     -0.258159   0.168402  -1.533  0.1262
StyleCasual    -0.101504   0.118077  -0.860  0.3906
StyleCute      0.016297   0.140410   0.116  0.9077
StyleFashion  -1.411871   0.913873  -1.545  0.1233
StyleFlare    -0.021174   0.385666  -0.055  0.9562
StyleNovelty   0.078243   0.220182   0.355  0.7225
StyleLoL      -1.499962   1.340590  -1.119  0.2640
StyleParty    -0.097065   0.154204  -0.629  0.5295
StyleSexy     -0.088861   0.131551  -0.675  0.4998
StyleVintage  -0.119791   0.155760  -0.769  0.4424
StyleWork     -0.294834   0.182452  -1.616  0.1070
PriceAverage  -0.192621   0.360245  -0.344  0.7312
PriceHigh     -0.328613   0.580585  -0.566  0.5718
PriceHigh     -0.328478   0.606116  -0.542  0.5882
PriceLow      -0.107766   0.569079  -0.189  0.8499
PriceLow      -0.180664   0.565431  -0.320  0.7495
PriceMedium   0.112440   0.567749   0.198  0.8431
PriceVery-High 0.033481   0.580353   0.058  0.9540
Rating        0.020767   0.012714   1.633  0.1033
SizeL        -0.084013   0.074618  -1.126  0.2610
SizeM        -0.007366   0.066478  -0.111  0.9118
SizeS        -0.613489   0.525870  -1.167  0.2442
SizeS      -0.044645   0.103932  -0.430  0.6678

```

Console	Terminal				
SeasonAutumn	0.469318	0.492834	0.952	0.3416	
SeasonAutumn	0.284369	0.524252	0.542	0.5879	
Seasonspring	0.725422	0.603464	1.202	0.2301	
Seasonspring	0.652107	0.490957	1.328	0.1850	
Seasonsummer	-0.125857	0.684699	-0.184	0.8543	
Seasonsummer	0.419780	0.489556	0.857	0.3918	
Seasonwinter	0.654508	0.495484	1.321	0.1874	
Seasonwinter	0.462766	0.491295	0.942	0.3469	
Necklinebackless	-1.254530	1.419872	-0.884	0.3776	
Necklineboat-neck	-0.582017	1.230356	-0.473	0.6365	
Necklinebowneck	-1.312418	1.232928	-1.064	0.2879	
Necklinehalter	-0.025341	1.421140	-0.018	0.9858	
Necklinemandarin-collor		NA	NA	NA	
NecklineNULL	-0.582935	1.288168	-0.453	0.6512	
Necklineo-neck	-0.899317	1.220073	-0.737	0.4616	
Necklineopen	-0.049056	1.453411	-0.034	0.9731	
Necklinepeterpan-collor	-1.286560	1.247642	-1.031	0.3032	
Necklineruffled	-0.452511	1.321326	-0.342	0.7322	
Necklinescoop	-1.634963	1.422296	-1.150	0.2511	
Necklineslash-neck	-0.904593	1.226241	-0.738	0.4612	
Necklinesquare-collor	-0.971724	1.210128	-0.803	0.4225	
Necklinesweetheart	-1.365543	1.427795	-0.956	0.3395	
Necklinesweetheart	-0.741449	1.212362	-0.612	0.5412	
Necklineturndowncollor	-1.050643	1.232954	-0.852	0.3947	
Necklinev-neck	-0.834377	1.222459	-0.683	0.4954	
Sleevelengthcap-sleeves	0.023285	0.650632	0.036	0.9715	
Sleevelengthcapsleeves	0.731884	0.588428	1.244	0.2144	
Sleevelengthfull	0.458956	0.525569	0.873	0.3831	
Sleevelengthhalf	0.121217	0.904498	0.134	0.8935	
Sleevelengthhalfsleeve	0.386391	0.528625	0.731	0.4653	
SleevelengthNULL	-0.008000	0.676210	-0.012	0.9906	
SleevelengthPetal	0.167605	0.900130	0.186	0.8524	
Sleevelengthshort	0.250598	0.517820	0.484	0.6287	
Sleevelengthsleeveless	0.689700	0.627122	1.100	0.2722	
Sleevelengthsleeveless	0.230819	0.582148	0.396	0.6920	
Sleevelengthsleeveless	0.414537	0.521706	0.795	0.4274	
Sleevelengthsleeveless	1.247600	0.717578	1.739	0.0830	
Sleevelengththreequarter	0.416416	0.539294	0.772	0.4406	
Sleevelengththreesqatar	0.930522	0.852041	1.092	0.2755	
Sleevelengththreesqatar	0.349959	0.546789	0.640	0.5226	

Console	Terminal				
Sleevelengthturndowncollor	-0.127322	0.719238	-0.177	0.8596	
waisselinedropped	-0.343477	0.660342	-0.520	0.6033	
waisselineempire	-0.480361	0.596887	-0.805	0.4215	
waisselineanatural	-0.546476	0.598026	-0.914	0.3615	
waisselinenull	-0.596031	0.600455	-0.993	0.3216	
waisselineprincess	NA	NA	NA	NA	
Materialacrylic	0.076951	0.576831	0.133	0.8940	
Materialcashmere	0.754218	0.571489	1.320	0.1878	
Materialchiffonfabric	0.343663	0.519451	0.662	0.5087	
Materialcotton	0.646010	0.504324	1.281	0.2011	
Materialknitting	-0.159281	0.801115	-0.199	0.8425	
Materiallace	-0.502632	1.038433	-0.484	0.6287	
Materiallinen	0.172931	0.586291	0.295	0.7682	
Materiallycra	0.287412	0.598318	0.480	0.6313	
Materialmicrofiber	0.616761	0.737301	0.837	0.4034	
Materialmilkssilk	0.308028	0.546589	0.564	0.5734	
Materialmix	0.544851	0.525050	1.038	0.3001	
Materialmodal	0.163388	0.703590	0.232	0.8165	
Materialmodel	1.302791	0.855226	1.523	0.1286	
Materialnull	0.619651	0.504670	1.228	0.2203	
Materialnylon	0.734646	0.535278	1.372	0.1708	
Materialother	0.169479	0.721247	0.235	0.8144	
Materialpolyester	0.463705	0.501565	0.925	0.3559	
Materialrayon	0.839200	0.528499	1.588	0.1132	
Materialshiffon	0.428243	0.647806	0.661	0.5090	
Materialsilk	0.521678	0.508615	1.026	0.3058	
Materialsilk	0.431810	0.715948	0.603	0.5468	
Materialspandex	0.288049	0.565746	0.509	0.6110	
Materialviscos	0.801165	0.625779	1.280	0.2013	
Materialwool	NA	NA	NA	NA	
Fabrictypebatik	0.592984	0.373483	1.588	0.1133	
Fabrictypebroadcloth	-0.052443	0.158720	-0.330	0.7413	
Fabrictypechiffon	0.005769	0.138961	0.042	0.9669	
Fabrictypecorduroy	-0.424073	0.371707	-1.141	0.2547	
Fabrictypedobby	-0.144064	0.410118	-0.351	0.7256	
Fabrictypeflannel	-0.518545	0.503783	-1.029	0.3041	
Fabrictypeflannel	-0.483332	0.510752	-0.946	0.3446	
Fabrictypejersey	0.148630	0.209425	0.710	0.4784	
Fabricypeknitted	0.643860	0.603382	1.067	0.2867	
Fabricypeknitting	-0.482488	0.543402	-0.888	0.3752	
Fabricypelace	1.140769	0.880260	1.286	0.1959	

```

Console Terminal x
~/
Decoratnone 0.483074 0.671437 0.719 0.4723
Decoratnull 0.286159 0.563995 0.507 0.6122
Decoratpearls -0.168049 0.773849 -0.217 0.8282
Decoratplain 0.046837 0.746890 0.063 0.9500
Decoratpleat 1.069576 0.912849 1.172 0.2421
Decoratpockets 0.441867 0.602312 0.734 0.4637
Decoratnivet -0.047011 0.640574 -0.073 0.9415
Decoratnruched 0.366551 0.653202 0.561 0.5751
Decoratnuffles 0.312723 0.581649 0.538 0.5912
Decoratnashes 0.182605 0.564482 0.323 0.7465
Decoratnsequined 0.378175 0.580889 0.651 0.5155
Decoratntassel 0.707253 0.763386 0.926 0.3548
Decoratntiered NA NA NA NA
Pattern.Typeanimal 0.114176 0.190988 0.598 0.5504
Pattern.Typecharacter -0.115429 0.499387 -0.231 0.8173
Pattern.Typedot 0.316510 0.213568 1.482 0.1392
Pattern.Typefloral -0.494397 0.371798 -1.330 0.1845
Pattern.Typegeometric -0.193967 0.287792 -0.674 0.5008
Pattern.Typeleopard 0.932243 0.523410 1.781 0.0758
Pattern.Typeleopard -0.085287 0.317424 -0.269 0.7883
Pattern.Typeone 0.560240 0.515338 1.087 0.2777
Pattern.Typenull 0.211136 0.154874 1.363 0.1737
Pattern.Typepatchwork 0.155906 0.159827 0.975 0.3300
Pattern.Typeplaid -0.218180 0.324250 -0.673 0.5015
Pattern.Typeprint 0.039583 0.155269 0.255 0.7989
Pattern.Typesolid 0.082661 0.142600 0.580 0.5625
Pattern.Typesplice -0.009635 0.528989 -0.018 0.9855
Pattern.Typestriped NA NA NA NA
---
signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.2225075)

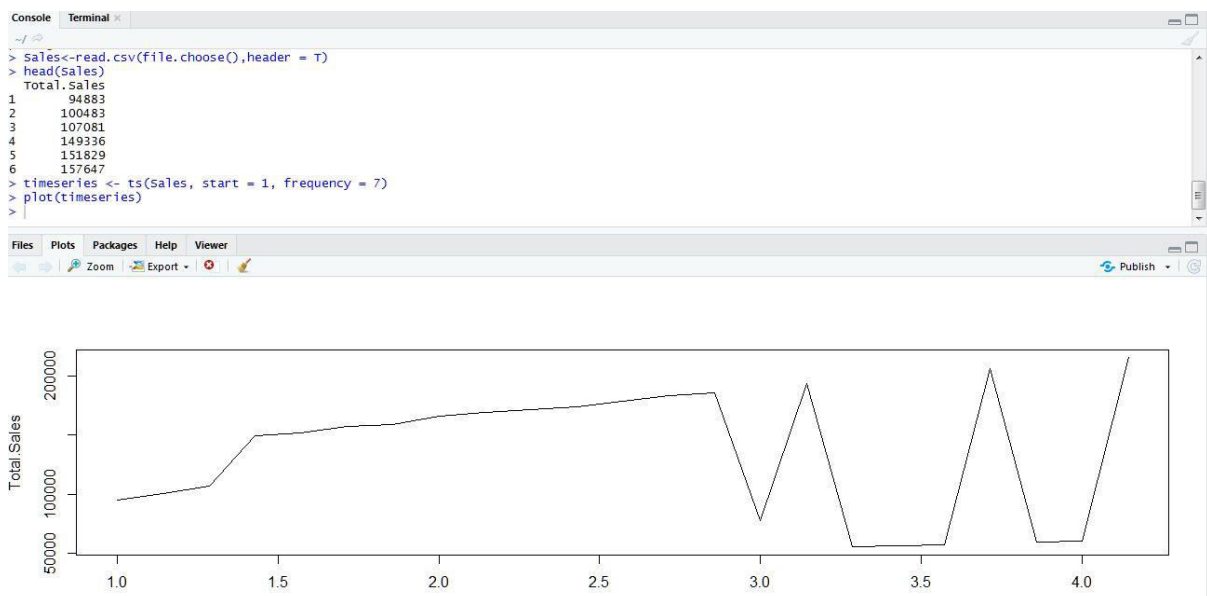
Null deviance: 121.80 on 499 degrees of freedom
Residual deviance: 77.21 on 347 degrees of freedom
AIC: 792.9

Number of Fisher Scoring iterations: 2

```

5. Making the Sales.csv data from the file dress_sales Extracting the total sum of sales for each alternative days provided.

6. reading the sales.csv file and applying the ts fuction for making timeseries plot.



7. Applying Auto Arima model on univariate and then forecasting the next 3 alternate days sales value.


```

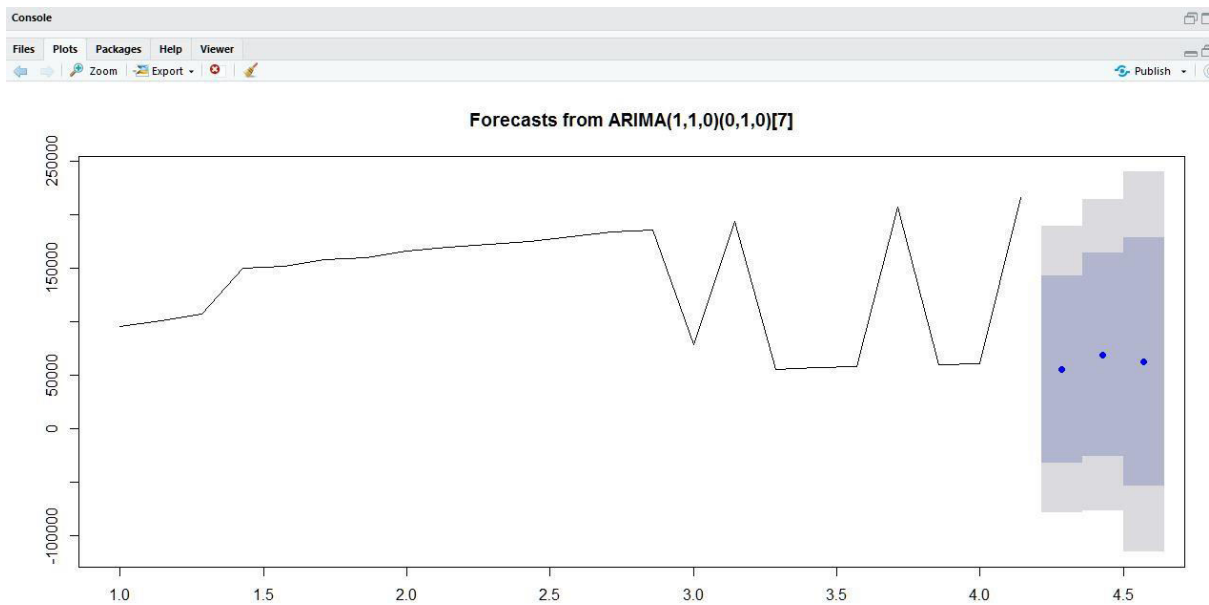
Console Terminal x
2 100483
3 107081
4 149336
5 151829
6 157647
> timeseries <- ts(sales, start = 1, frequency = 7)
> plot(timeseries)
> ArimaModel <- auto.arima(timeseries)
> summary(ArimaModel)
Series: timeseries
ARIMA(1,1,0)(0,1,0)[7]

Coefficients:
      ar1
    -0.5754
s.e.   0.1953

sigma^2 estimated as 4.68e+09: log likelihood=-187.97
AIC=379.93   AICc=380.93   BIC=381.35

Training set error measures:
      ME      RMSE      MAE      MPE
Training set -4326.728 53371.64 31649.38 -17.82185
      MAPE      MASE      ACF1
Training set 33.33584 0.5248468 -0.1716439
> forecast(ArimaModel,3)
      Point Forecast      Lo 80      Hi 80      Lo 95
4.285714      54784.79 -32884.31 142453.9 -79293.53
4.428571      68671.50 -26573.54 163916.5 -76993.22
4.571429      62256.90 -53763.30 178277.1 -115180.68
      Hi 95
4.285714 188863.1
4.428571 214336.2
4.571429 239694.5
> plot(forecast(ArimaModel,3))
>

```



8. Making another file of attribute in which adding a row of total sales per dress id.

9. Reading the new attribute file with total sales

```

Console Terminal x
~/
> Data3<-read.csv(file.choose(),header = T)
> head(Data3)
  Dress_ID Style Price Rating Size Season NeckLine SleeveLength waistline
1 1006032852 Sexy Low 4.6 M Summer o-neck sleeveless empire
2 1212192089 Casual Low 0.0 L Summer o-neck Petal natural
3 1190380701 vintage High 0.0 L Autumn o-neck full natural
4 966005983 Brief Average 4.6 L Spring o-neck full natural
5 876339541 cute Low 4.5 M Summer o-neck butterfly natural
6 1068332458 bohemian Low 0.0 M Summer v-neck sleeveless empire
  Material FabricType Decoration Pattern.Type Recommendation Total.Sales
1 null chiffon ruffles animal 1 75979
2 microfiber null ruffles animal 0 52256
3 polyester null null print 0 223
4 silk chiffon embroidery print 1 39691
5 chiffonfabric chiffon bow dot 0 44077
6 null null null print 0 457
> class(Data3$Total.Sales)
[1] "integer"
>

```

10.Applying ANOVA to find affect of style season and material separately on total sales

```

> outStyle<-aov(Data3$Total.Sales~Data3$Style)
> Outseason<-aov(Data3$Total.Sales~Data3$Season)
> Outmaterial<-aov(Data3$Total.Sales~Data3$Material)
> summary(outStyle)
              Df      Sum Sq   Mean Sq F value Pr(>F)
Data3$Style   11  2.340e+09  212693589   1.418  0.161
Residuals    488  7.318e+10 149961325
> summary(Outseason)
              Df      Sum Sq   Mean Sq F value   Pr(>F)
Data3$Season    8  3.700e+09  462453725   3.162 0.00167 **
Residuals     491  7.182e+10 146275206
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(Outmaterial)
              Df      Sum Sq   Mean Sq F value   Pr(>F)
Data3$Material 24  3.221e+09 134204184   0.882  0.628
Residuals     475  7.230e+10 152210222
>

```

11.Applying linear regression model to find the relation of all three attributes on the total sales which can also be helpful for predicting in future

```

Console Terminal x
~/
Residuals:      475 7.230e+10 152210222
> model3a<-lm(Total.Sales~Style+Season+Material,data = Data3)
> summary(model3a)

Call:
lm(formula = Total.Sales ~ Style + Season + Material, data = Data3)

Residuals:
    Min       1Q   Median       3Q      Max
-18312  -4768  -2023    1120 137575

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    2681.0    15324.4   0.175  0.86119
StyleBrief      6206.5     3894.1   1.594  0.11167
StyleCasual     1779.4     2698.9   0.659  0.51005
StyleCute       4249.7     3175.3   1.338  0.18144
StyleFashion   -2286.8     12398.5  -0.184  0.85375
StyleFlare     -1893.2      9128.7  -0.207  0.83580
StyleNovelty     82.7      5054.5   0.016  0.98695
StyleOL        -2339.6     12460.1  -0.188  0.85114
StyleParty     -972.9      3131.2  -0.311  0.75617
StyleSexy      5515.4      2920.8   1.888  0.05961
StyleVintage   4904.4      3526.8   1.391  0.16502
StyleWork      1492.3      3938.8   0.379  0.70496
SeasonAutumn    3426.3      8767.5   0.391  0.69613
SeasonAutumn   -1076.1      9650.2  -0.112  0.91126
SeasonSpring   39004.7     12165.4   3.206  0.00144 **
SeasonSpring    4291.5      8667.1   0.495  0.62074
SeasonSummer    3167.7     14856.2   0.213  0.83125
SeasonSummer    2703.8      8660.9   0.312  0.75504
SeasonWinter   -439.7      8821.4  -0.050  0.96027
SeasonWinter    2759.6      8696.2   0.317  0.75113
MaterialAcrylic -3682.5     14180.3  -0.260  0.79522
MaterialCashmere -5165.1     13756.9  -0.375  0.70749
MaterialChiffonFabric 7130.2     12534.6   0.569  0.56975
MaterialCotton  -2672.1     12301.1  -0.217  0.82813

```

```

Console Terminal x
~/
SeasonSpring    4291.5      8667.1   0.495  0.62074
SeasonSummer    3167.7     14856.2   0.213  0.83125
SeasonSummer    2703.8      8660.9   0.312  0.75504
SeasonWinter   -439.7      8821.4  -0.050  0.96027
SeasonWinter    2759.6      8696.2   0.317  0.75113
MaterialAcrylic -3682.5     14180.3  -0.260  0.79522
MaterialCashmere -5165.1     13756.9  -0.375  0.70749
MaterialChiffonFabric 7130.2     12534.6   0.569  0.56975
MaterialCotton  -2672.1     12301.1  -0.217  0.82813
MaterialKnitting -3959.0     17331.7  -0.228  0.81942
MaterialLace    -5326.8     17342.6  -0.307  0.75887
MaterialLinen   -418.2      14181.5  -0.029  0.97649
MaterialLycra   -2234.5     14112.0  -0.158  0.87426
MaterialMicrofiber 11264.0     14099.0   0.799  0.42475
MaterialMilkSilk 767.8      13432.2   0.057  0.95444
MaterialMix     -3001.0     12814.5  -0.234  0.81495
MaterialModal   -6443.3     17222.7  -0.374  0.70849
MaterialModel   -6775.3     17222.7  -0.393  0.69421
MaterialNull    -1647.5     12286.0  -0.134  0.89339
MaterialNylon   -4150.6     12899.5  -0.322  0.74777
MaterialOther   -3957.0     15075.0  -0.262  0.79306
MaterialPolyester -1936.4     12352.9  -0.157  0.87550
MaterialRayon   -2059.1     12843.4  -0.160  0.87270
MaterialShiffon -5111.9     15128.4  -0.338  0.73559
MaterialSilk    -4005.5     12516.0  -0.320  0.74909
MaterialSill    -8181.3     17260.7  -0.474  0.63574
MaterialSpandex -5174.0     13478.8  -0.384  0.70126
MaterialViscos  -5303.1     14888.7  -0.356  0.72187
MaterialWool    -2639.8     17357.8  -0.152  0.87919
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12090 on 456 degrees of freedom
Multiple R-squared:  0.1168,    Adjusted R-squared:  0.03356
F-statistic: 1.403 on 43 and 456 DF,  p-value: 0.0513

```

12. Applying Linear regression to find the affect of price and style on total sales.

```

Console Terminal x
~/
> model3b<-lm(Total.Sales~Price+Style,data = Data3)
> summary(model3b)

Call:
lm(formula = Total.Sales ~ Price + Style, data = Data3)

Residuals:
    Min       1Q   Median       3Q      Max
-12688  -5006  -2406    664 143199

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)      521.6      9230.0   0.057  0.9550
PriceAverage     1670.6      8900.9   0.188  0.8512
Pricehigh      -2820.5      9230.6  -0.306  0.7601
PriceHigh       1940.9     10162.8   0.191  0.8486
Pricelow       -206.6      9094.5  -0.023  0.9819
PriceLow        4634.3      8974.8   0.516  0.6058
PriceMedium    -1192.0      9102.0  -0.131  0.8959
Pricevery-high -2918.1      9023.7  -0.323  0.7465
StyleBrief       6074.1      3814.2   1.593  0.1119
StyleCasual      2940.3      2623.1   1.121  0.2629
Stylecute        7571.8      3101.5   2.441  0.0150 *
Stylefashion    -1766.2     12450.0  -0.142  0.8872
StyleFlare      -524.2      8991.1  -0.058  0.9535
StyleNovelty      878.1      5016.6   0.175  0.8611
StyleOL         1099.4     12636.1   0.087  0.9307
Styleparty       3630.4      3321.2   1.093  0.2749
StyleSexy        5867.8      2862.8   2.050  0.0409 *
Stylevintage     6147.6      3533.6   1.740  0.0825 .
Stylework        4006.2      3910.9   1.024  0.3062
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12180 on 481 degrees of freedom
Multiple R-squared:  0.0553,    Adjusted R-squared:  0.01994
F-statistic: 1.564 on 18 and 481 DF,  p-value: 0.06496

```

13. Applying linear regression on all the variables to know the affect on total sales.

```

Console Terminal x
~/
> #Objective4:Also, to increase the sales, the management wants to analyze the attributes
> #of dresses and find which are the leading factors affecting the sale of a
> #dress.
> model14<-lm(Total.Sales~.,data = Data3)
> summary(model14)

Call:
lm(formula = Total.Sales ~ ., data = Data3)

Residuals:
    Min       1Q   Median       3Q      Max
-24801  -3789   -200    1981  57866

Coefficients: (7 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.689e+04  1.819e+04   2.578  0.010340 *
Dress_ID     -1.606e-05  3.791e-06  -4.237  2.91e-05 ***
StyleBrief    4.689e+03  3.348e+03   1.400  0.162290
StyleCasual    1.898e+03  2.342e+03   0.810  0.418436
Stylecute     1.626e+03  2.782e+03   0.584  0.559364
Stylefashion   -3.020e+03  1.825e+04  -0.165  0.868653
StyleFlare     6.780e+02  7.648e+03   0.089  0.929410
StyleNovelty   3.452e+03  4.382e+03   0.788  0.431340
StyleOL        1.005e+04  2.664e+04   0.377  0.706249
Styleparty     1.854e+03  3.058e+03   0.606  0.544761
StyleSexy      5.834e+03  2.609e+03   2.237  0.025951 *
Stylevintage   6.836e+03  3.093e+03   2.210  0.027764 *
Stylework      2.221e+03  3.630e+03   0.612  0.541078
PriceAverage   1.017e+04  1.113e+04   0.914  0.361290
Pricehigh      9.122e+03  1.152e+04   0.792  0.428836
PriceHigh      9.613e+03  1.205e+04   0.798  0.425701
Pricelow       7.396e+03  1.130e+04   0.655  0.513071
PriceLow       1.158e+04  1.123e+04   1.031  0.303217
PriceMedium    5.046e+03  1.126e+04   0.448  0.654361
Pricevery-high 7.786e+03  1.150e+04   0.677  0.498895
Rating         4.841e+02  3.089e+02   1.567  0.118033
SizeL          4.757e+03  1.489e+03   3.194  0.001531 **
SizeM          1.788e+03  1.325e+03   1.350  0.177993
SizeS         -9.972e+03  1.044e+04  -0.955  0.340369
SizeS         2.027e+03  2.066e+03   0.981  0.327129
SizeS         4.421e+03  4.001e+03   1.105  0.267620

```


Console

Terminal

~/

PriceMedium	5.046e+03	1.126e+04	0.448	0.654361
Pricevery-high	7.786e+03	1.150e+04	0.677	0.498895
Rating	4.841e+02	3.089e+02	1.567	0.118033
SizeL	4.757e+03	1.489e+03	3.194	0.001531 **
SizeM	1.788e+03	1.325e+03	1.350	0.177993
Sizes	-9.972e+03	1.044e+04	-0.955	0.340369
Sizes	2.027e+03	2.066e+03	0.981	0.327129
Sizesmall	-1.119e+03	1.094e+04	-0.102	0.918608
SizeXL	6.414e+02	3.015e+03	0.213	0.831653
SeasonAutumn	-8.273e+02	9.778e+03	-0.085	0.932617
SeasonAutumn	-7.500e+03	1.039e+04	-0.722	0.470996
Seasonspring	2.795e+04	1.200e+04	2.329	0.020444 *
SeasonSpring	-6.369e+02	9.753e+03	-0.065	0.947968
Seasonsummer	9.403e+02	1.357e+04	0.069	0.944782
SeasonSummer	-1.715e+03	9.711e+03	-0.177	0.859884
Seasonwinter	-7.083e+03	9.842e+03	-0.720	0.472197
Seasonwinter	-1.266e+03	9.748e+03	-0.130	0.896711
NeckLinebackless	-6.107e+03	2.825e+04	-0.216	0.828982
NeckLineboat-neck	1.991e+03	2.446e+04	0.081	0.935185
NeckLinebowneck	3.543e+03	2.453e+04	0.144	0.885237
NeckLinehalter	6.249e+03	2.820e+04	0.222	0.824736
NeckLinemandarin-collor	NA	NA	NA	NA
NeckLineNULL	-3.445e+03	2.560e+04	-0.135	0.893048
NeckLineo-neck	3.616e+03	2.427e+04	0.149	0.881671
NeckLineopen	3.264e+03	2.904e+04	0.112	0.910578
NeckLinepeterpan-collor	1.103e+03	2.484e+04	0.044	0.964610
NeckLineeruffled	1.405e+05	2.630e+04	5.340	1.69e-07 ***
NeckLinescoop	4.483e+03	2.825e+04	0.159	0.873988
NeckLineslash-neck	-1.804e+02	2.440e+04	-0.007	0.994107
NeckLinesqare-collor	-2.726e+03	2.408e+04	-0.113	0.909924
NeckLinesweetheart	1.371e+03	2.848e+04	0.048	0.961635
NeckLineSweetheart	3.525e+03	2.413e+04	0.146	0.883938
NeckLineturndowncollor	5.745e+03	2.454e+04	0.234	0.815038
NeckLinev-neck	2.807e+03	2.432e+04	0.115	0.908184
Sleevelengthcap-sleeves	-3.788e+04	1.290e+04	-2.936	0.003544 **
Sleevelengthcapsleeves	-3.706e+04	1.171e+04	-3.164	0.001696 **
Sleevelengthfull	-3.356e+04	1.044e+04	-3.214	0.001431 **
Sleevelengthhalf	-2.917e+04	1.792e+04	-1.627	0.104550
Sleevelengthhalfsleeve	-3.193e+04	1.050e+04	-3.042	0.002527 **
SleevelengthNULL	-2.552e+04	1.341e+04	-1.904	0.057805 .

Console	Terminal x			
~/				
Materialacrylic	7.843e+02	1.143e+04	0.069	0.945352
Materialcashmere	-3.770e+03	1.142e+04	-0.330	0.741469
Materialchiffonfabric	1.104e+03	1.035e+04	0.107	0.915109
Materialcotton	-1.396e+03	1.004e+04	-0.139	0.889490
Materialknitting	1.709e+03	1.600e+04	0.107	0.914976
Materiallace	-1.152e+04	2.058e+04	-0.560	0.575951
Materiallinen	2.045e+03	1.163e+04	0.176	0.860512
Materiallycra	3.462e+03	1.188e+04	0.291	0.770963
Materialmicrofiber	-5.281e+03	1.463e+04	-0.361	0.718269
Materialmilkssilk	9.143e+02	1.085e+04	0.084	0.932877
Materialmix	-2.710e+03	1.043e+04	-0.260	0.795280
Materialmodal	7.247e+02	1.394e+04	0.052	0.958579
Materialmodel	-6.949e+03	1.700e+04	-0.409	0.683064
Materialnull	1.651e+02	1.005e+04	0.016	0.986895
Materialnylon	-2.009e+03	1.066e+04	-0.189	0.850578
Materialother	-6.281e+02	1.431e+04	-0.044	0.965018
Materialpolyester	-2.139e+02	9.965e+03	-0.021	0.982888
Materialrayon	-3.500e+02	1.055e+04	-0.033	0.973562
Materialshiffon	-3.608e+03	1.291e+04	-0.279	0.780051
Materialsilk	-3.277e+03	1.012e+04	-0.324	0.746155
Materialsill	-4.373e+03	1.420e+04	-0.308	0.758369
Materialspandex	-3.069e+03	1.122e+04	-0.274	0.784557
Materialviscos	-5.002e+03	1.248e+04	-0.401	0.688915
Materialwool	NA	NA	NA	NA
Fabrictypebatik	1.201e+03	7.439e+03	0.161	0.871841
Fabrictypebroadcloth	4.120e+03	3.184e+03	1.294	0.196454
Fabrictypechiffon	2.448e+03	2.771e+03	0.883	0.377672
Fabrictypecorduroy	6.004e+03	7.397e+03	0.812	0.417481
Fabrictypedobby	-8.597e+03	8.197e+03	-1.049	0.294979
Fabrictypeflannael	1.325e+03	1.000e+04	0.132	0.894726
Fabrictypeflannel	1.140e+04	1.013e+04	1.125	0.261524
Fabrictypejersey	7.247e+03	4.153e+03	1.745	0.081881
Fabrictypeknitted	2.148e+03	1.198e+04	0.179	0.857766
Fabrictypeknitting	2.889e+03	1.082e+04	0.267	0.789531
Fabrictypelace	-3.392e+03	1.752e+04	-0.194	0.846539
Fabrictypenull	9.141e+02	2.614e+03	0.350	0.726772
Fabrictypeorganza	NA	NA	NA	NA
Fabrictypeother	5.772e+03	1.444e+04	0.400	0.689598
Fabrictypepoplin	-5.569e+03	1.083e+04	-0.514	0.607387
Fabrictypesatin	1.313e+03	1.049e+04	0.125	0.900421
Fabrictypesattin	3.079e+03	6.044e+03	0.510	0.610700

```

Console Terminal x
~/
Decorat iont lowers      -2.107e+04  1.315e+04  -1.603  0.109959
Decorat ionhollowout    -1.720e+04  1.152e+04  -1.493  0.136331
Decorat ionlace         -1.631e+04  1.131e+04  -1.443  0.150046
Decorat ionnone         -1.710e+04  1.335e+04  -1.280  0.201278
Decorat ionnull         -1.769e+04  1.123e+04  -1.576  0.116009
Decorat ionpearls       -1.355e+04  1.543e+04  -0.878  0.380710
Decorat ionplain        -2.237e+04  1.481e+04  -1.510  0.131943
Decorat ionpleat        -2.980e+04  1.814e+04  -1.643  0.101384
Decorat ionpockets      -1.861e+04  1.200e+04  -1.551  0.121894
Decorat ionrivet        -1.852e+04  1.271e+04  -1.457  0.146039
Decorat ionruched       -2.427e+04  1.296e+04  -1.872  0.061999
Decorat ionruffles      -1.206e+04  1.160e+04  -1.040  0.299278
Decorat ionsashes      -1.623e+04  1.125e+04  -1.443  0.149995
Decorat ionsequined     -1.617e+04  1.159e+04  -1.396  0.163757
Decorat iontassel      -2.138e+04  1.517e+04  -1.409  0.159597
Decorat ionTiered       NA          NA          NA          NA
Pattern.Typeanimal      5.671e+03  3.788e+03  1.497  0.135319
Pattern.Typecharacter    3.124e+02  9.897e+03  0.032  0.974838
Pattern.Typedot         -4.149e+02  4.253e+03  -0.098  0.922335
Pattern.Typefloral      2.776e+03  7.412e+03  0.374  0.708278
Pattern.Typegeometric    8.367e+03  5.728e+03  1.461  0.145015
Pattern.Typeleopard     -9.167e+03  1.042e+04  -0.880  0.379629
Pattern.Typeleopard     -1.453e+02  6.293e+03  -0.023  0.981586
Pattern.Typenone        -4.851e+03  1.025e+04  -0.473  0.636490
Pattern.Typenull        -9.880e+02  3.078e+03  -0.321  0.748464
Pattern.Typepatchwork   -1.934e+03  3.183e+03  -0.608  0.543812
Pattern.Typeplaid       -2.664e+03  6.429e+03  -0.414  0.678838
Pattern.Typeprint       7.503e+02  3.081e+03  0.244  0.807727
Pattern.Typesolid       9.360e+02  2.841e+03  0.329  0.741981
Pattern.Typesplice      -4.217e+03  1.050e+04  -0.401  0.688320
Pattern.Typestriped     NA          NA          NA          NA
Recommendation         1.532e+03  1.064e+03  1.439  0.150973
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9346 on 345 degrees of freedom
Multiple R-squared:  0.601,    Adjusted R-squared:  0.4229
F-statistic: 3.374 on 154 and 345 DF,  p-value: < 2.2e-16

```

14. Applying correlation function to find if there is correlation between the rating and total sales.

```

Console Terminal x
~/
> #objective5:To regularize the rating procedure and find its efficiency, the store wants to
> #find if the rating of the dress affects the total sales.
> model5<-cor(Data3$Total.Sales,Data3$Rating)
> model5
[1] 0.1956887
> names(Data3)
[1] "Dress_ID"      "Style"          "Price"          "Rating"         "Size"
[6] "Season"        "NeckLine"       "sleeveLength"   "waixeline"      "Material"
[11] "FabricType"    "Decoration"     "Pattern.Type"   "Recommendation" "Total.Sales"
> c<-cor(Data3[,c(4,15)])
> corplot(c)
>

```



Analysis for the individual objectives

1. The significance code by checking the p value of each can be said and In my opinion the p value is not very low for specific factors. Though the AIC is low which implies a good model but the number of entries are less and there are a lot of factors which makes the prediction through this model very hard. The Residual deviance is lower than the null deviance which tells that the prediction through this model will be very close to the actual value. This model can be improved by more number of entries in the data set and reducing some levels from the individual factors on the basis of p value or providing different attributes to get the recommendations.
2. The Arimamodel is taking the ARIMA(1,1,0)(0,1,0) automatically. In the result the errors can be found and other coefficient and other information were used to select the model by comparing. The forecasted value are 54784.79,68671.50,62256.90. In the plot the light grey and grey area specifies that there is a lot of fluctuations in total sales and hence there can be error.
3. By seeing the ANOVA result we can conclude that individually the season is having the lowest p value which tells that it is impacting the total

sales, while the style is moderately effecting as it is more than 11% we may or may not accept its affect and the material factor is having very high p value hence it is not affecting and can be dropped. While doing the cumulative affects of style season and material by linear regression we can see the model3a is feasible as the p value is around 5% but the r square is very low and hence the model is not good, while in that the spring season is affecting as the p value is 0.1% while the style sexy is moderately affecting as its p value is around 6%. The other model3b tells us that the style is dominant in affecting the total sales over price and in particular the style cute, sexy, vintage is specially affecting the total sales. In this model the p value is around 6% which can be taken as feasible but the r square value is very low and hence model is not good.

4. The model5 is feasible as the p value very low but the model is not very good as the r square not very high. Style sexy and style vintage is making a positive impact, Mostly the size L are being sold, In the spring season the clothes makes a good impact on sales, Neckline ruffled is making a high impact on sales, Sleeve length is making impact but in a negative manner to the sales and hence are being sold less, Fabric jersey is making very moderate impact but in a positive way, the Beaded and ruched decoration is causing a mild negative impact on sales
5. It can be seen that the correlation value between Total sales and rating is 0.1956 which is suggesting a very weak but positive correlation which means the the higher rating comes when the higher sales are done which suggest that the result is not accurate and the rating procedure need to be regularized.