Analysis of Sales Report of a Clothes Manufacturing Outlet

Background and Objective

A high-end fashion retail store is looking to expand its products. It wants to understand the market and find the current trends in the industry. It has a database of all products with attributes, such as style, material, season, and the sales of the products over a period of two months.

Analysis Tasks

To automate the process of recommendations, the store needs to analyze 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.

The **Attribute DataSet.xlsx** file contains all the attributes and the recommendations for each dress in binary. For binary output models, logistic regression can be used to build a suitable model to recommend products (1 denotes a positive recommendation and 0 denotes a negative recommendation). To perform logistic regression on a given dataset, we need to decide two major attributes of the model

Independent variables: All other variables, except the Dress_ID, since it is only an identifier.

Dependent variable: Recommendation

```
> library(xlsx)
> setwd("~/R Project_Clothes Manufacturing Outlet")
> attributeClothe = read.xlsx('Attribute DataSet.xlsx',header = TRUE,sheetName = 'Sheet1')
> head(attributeClothe)
    Dress_ID    Style   Price Rating Size Season NeckLine SleeveLength waiseline      Material FabricType Decoration
1 1006032852     Sexy     Low    4.6    M Summer   o-neck    sleevless    empire          null    chiffon    ruffles
2 1212192089   Casual     Low    0.0    L Summer   o-neck        Petal   natural    microfiber       null    ruffles
3 1190380701  vintage    High    0.0    L Automn   o-neck         full   natural      polyster       null       null
4  966005983    Brief Average    4.6    L Spring   o-neck         full   natural          silk    chiffon embroidary
5  876339541     cute     Low    4.5    M Summer   o-neck    butterfly   natural chiffonfabric    chiffon        bow
6 1068332458 bohemian     Low    0.0    M Summer   v-neck    sleevless    empire          null       null       null
  Pattern.Type Recommendation
1       animal              1
2       animal              0
3        print              0
4        print              1
5          dot              0
6        print              0
```

```
> model = glm(Recommendation ~ Style + Price + attributeClothe$Rating+Size+Season+NeckLine+SleeveLength+waiseline+Material+FabricType+
Decoration+Pattern.Type,data = attributeClothe)
> summary(model)

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

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9895  -0.3241   0.0000   0.3191   0.8808
```

```
Pricehigh              -0.140446   0.175119  -0.802  0.42310
PriceHigh              -0.129542   0.220973  -0.586  0.55810
Pricelow                0.086340   0.095125   0.908  0.36470
PriceLow                0.010431   0.063789   0.164  0.87021
PriceMedium             0.300730   0.111557   2.696  0.00737 **
Pricevery-high          0.223830   0.208403   1.074  0.28356
```

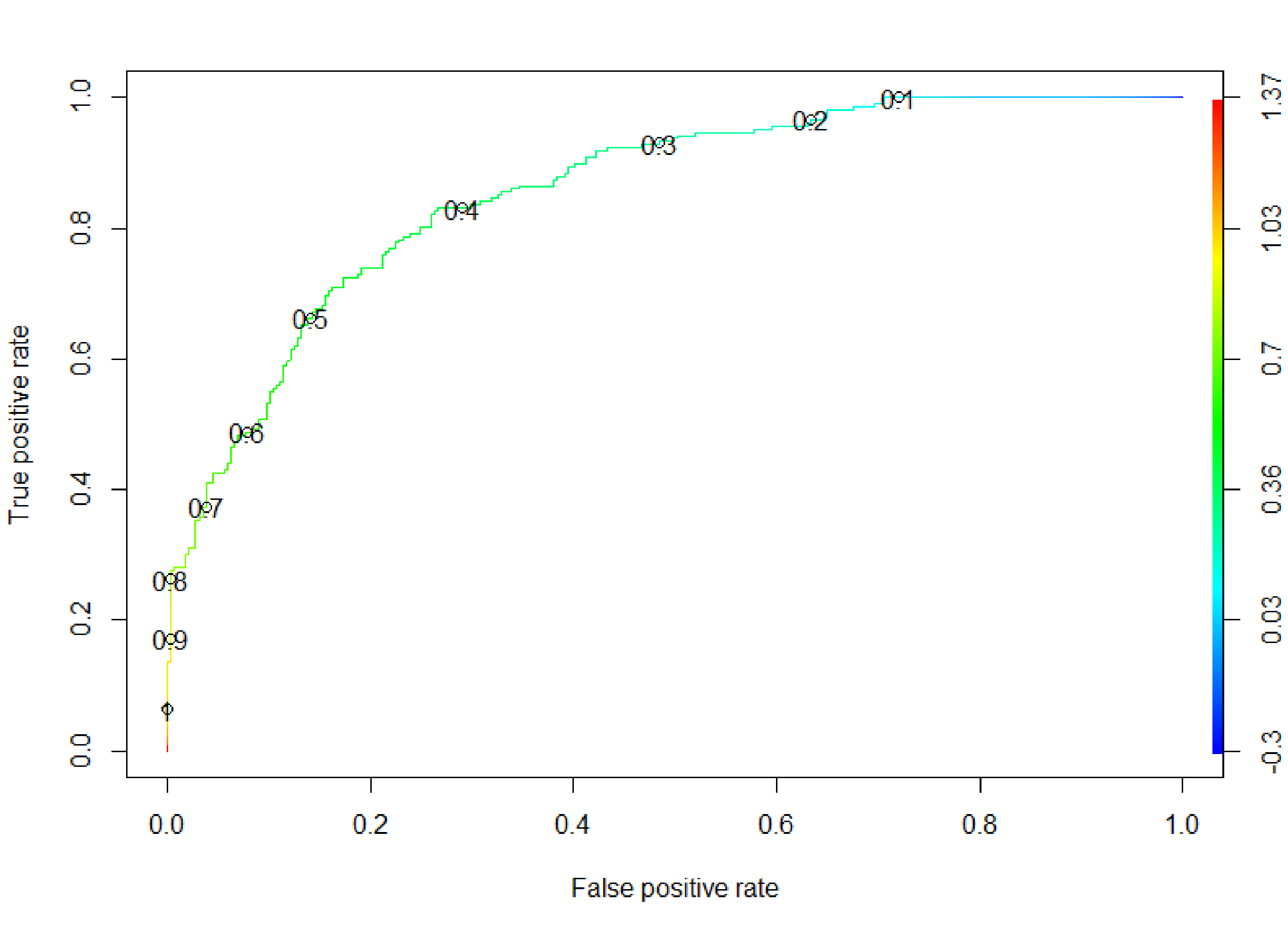
```
Materialcashmere        0.676362   0.406858   1.662  0.09734 .
Materialchiffonfabric   0.269077   0.316566   0.850  0.39592
Materialcotton          0.569308   0.297494   1.914  0.05649 .
Materialknitting       -0.234039   0.687674  -0.340  0.73381
Materiallace           -0.686514   0.967162  -0.710  0.47829
Materiallinen           0.093026   0.421718   0.221  0.82554
Materiallycra           0.214349   0.465173   0.461  0.64524
Materialmicrofiber     -0.054620   0.698670  -0.078  0.93773
Materialmilksilk        0.230922   0.372906   0.619  0.53616
Materialmix             0.471069   0.335452   1.404  0.16113
Materialmodal           0.080863   0.577976   0.140  0.88881
Materialmodel           1.222190   0.756022   1.617  0.10687
Materialnull            0.541787   0.295907   1.831  0.06797 .
Materialnylon           0.659877   0.348017   1.896  0.05878 .
Materialother           0.087055   0.596355   0.146  0.88402
Materialpolyster        0.385803   0.298936   1.291  0.19771
Materialrayon           0.761183   0.332962   2.286  0.02285 *
Materialshiffon         0.359398   0.508320   0.707  0.48002
Materialsilk            0.449012   0.311467   1.442  0.15032
Materialsill            0.359753   0.579957   0.620  0.53546
Materialspandex         0.213837   0.384351   0.556  0.57833
Materialviscos          0.728820   0.477683   1.526  0.12799
Materialwool           -0.075463   0.577508  -0.131  0.89611
```

```
SeasonAutumn           -0.183655   0.199303  -0.921  0.35744
Seasonspring            0.257196   0.355791   0.723  0.47024
SeasonSpring            0.182272   0.086329   2.111  0.03546 *
Seasonsummer           -0.601832   0.500358  -1.203  0.22988
SeasonSummer           -0.049393   0.083264  -0.593  0.55343
Seasonwinter            0.186837   0.108752   1.718  0.08669 .
SeasonWinter           -0.008118   0.090244  -0.090  0.92837
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.2230226)

    Null deviance: 120.611  on 495  degrees of freedom
Residual deviance:  77.166  on 346  degrees of freedom
  (4 observations deleted due to missingness)
AIC: 786.72

Number of Fisher Scoring iterations: 2
```

From the significance codes for each attribute, we can see that **Price, Material and Season** make an impact on the recommendation, both positively affecting the recommendation. Other than that, we can see that the increased number of factors and comparatively lesser number of entries make the predictions slightly difficult. However, the residual deviance is lower than the null deviance, which implies that using the independent variables makes it closer to predicting the actual values of recommendation. With the given model, the new data or attributes can be fed into the model to get recommendations.

```
> y = attributeClothe$Recommendation
> result <- predict(model,attributeClothe[,2:13],type = "response")
Warning message:
In predict.lm(object, newdata, se.fit, scale = 1, type = if (type ==  :
  prediction from a rank-deficient fit may be misleading
> library(ROCR)
> rocrPredic <- prediction(result,y)
> rocrPerf  <- performance(rocrPredic,"tpr","fpr")
> plot(rocrPerf,colorize=TRUE, print.cutoffs.at=seq(0.1,by=0.1))
```



From that plot we can see the best threshold point is 0.4

Now we are going to predict “Recommendation” with trained model. Through the confusion matrix we will know the accurate rate.

```
> train_PredSurvived <- ifelse(result > 0.4,1,0)
> table(predicted=train_PredSurvived,actualdata=y)
         actualdata
predicted   0   1
        0 206  35
        1  83 172
  .
```

Model Accuracy is 76.21%

Analysis Tasks

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 or more alternative days.

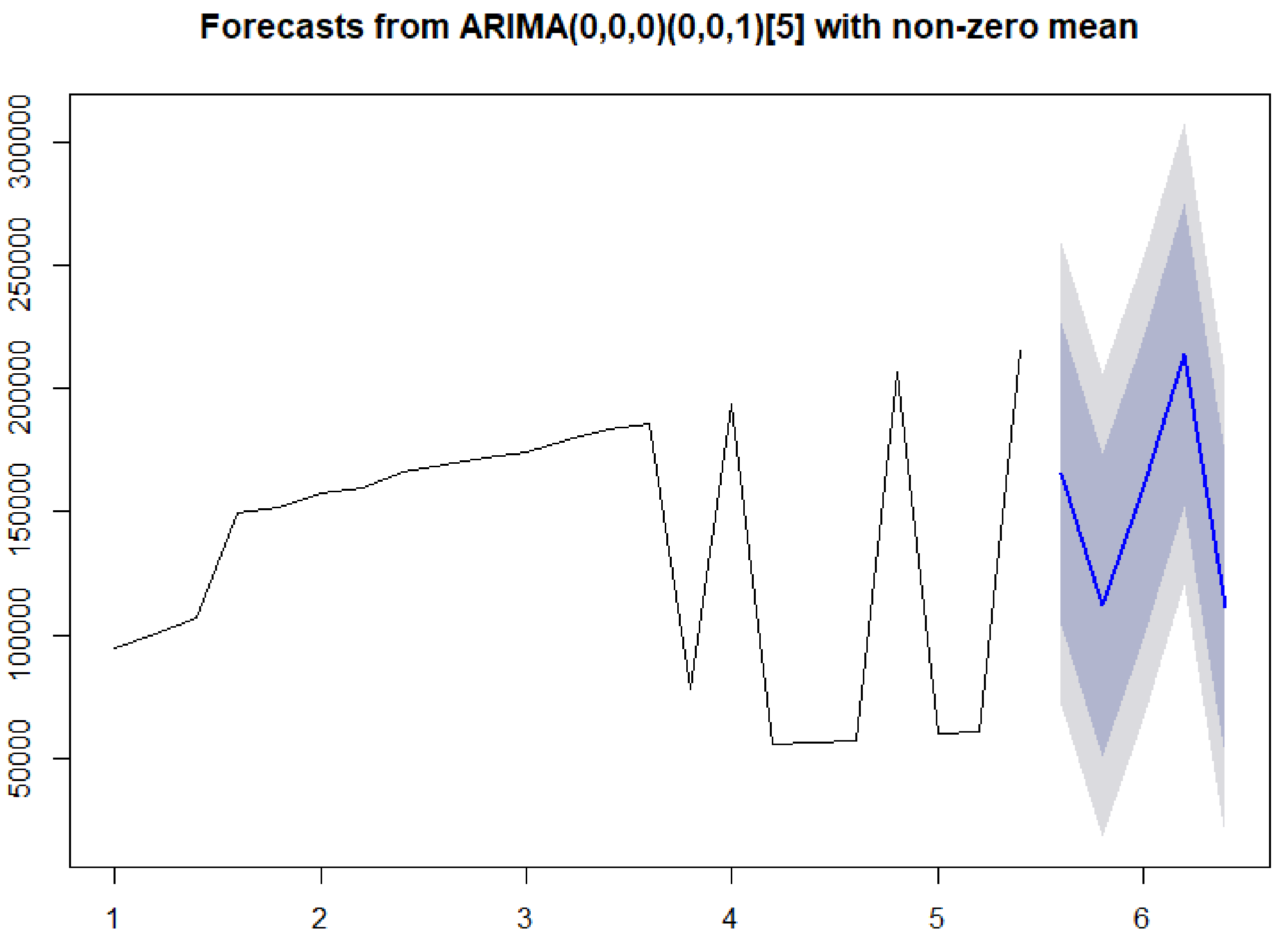
For this question, we will use the **Total Sales.xlsx** file. Since the time series should be a vector, we will work on the total sales per day (of all items). The total sales is calculated in Excel (using the SUM command) and saved in a file named totalSales3.csv. (The file is provided here for verification). The Auto.arima function is used to predict the trend for three more days. Note that any of the given time series functions can be used in the prediction of sales

```
> totalSales = read.csv('Total Sales.csv',header = TRUE)
> head(totalSales)
  Dress_ID X29.08.2013 X31.08.2013 X09.02.2013 X09.04.2013 X09.06.2013 X09.08.2013 X09.10.2013 X09.12.2013 X14.09.2013
1        1       94883      100483      107081      149336      151829      157647      159391      165962      169101
  X16.09.2013 X18.09.2013 X20.09.2013 X22.09.2013 X24.09.2013 X26.09.2013 X28.09.2013 X30.09.2013 X10.02.2013 X10.04.2013
1      171726      174360      179037      183261      185616       77934      193734       55412       56395       57405
  X10.06.2013 X10.08.2010 X10.10.2013 X10.12.2013
1      206334       59816       60757      215533
```

```
> matrixTotalSales <- as.matrix(totalSales[,-1])
> numericVector <-  as.numeric(as.vector(matrixTotalSales))
> timeseries <- ts(numericVector, start = 1,frequency = 5)
> fit <- auto.arima(timeseries)
> summary(fit)
Series: timeseries
ARIMA(0,0,0)(0,0,1)[5] with non-zero mean

Coefficients:
         sma1        mean
      -0.6502  139318.887
s.e.   0.3123    5523.818

sigma^2 estimated as 2.275e+09:  log likelihood=-280.71
AIC=567.42   AICc=568.68   BIC=570.82

Training set error measures:
                   ME    RMSE      MAE       MPE     MAPE      MASE       ACF1
Training set 668.9685 45571.28 38319.67 -17.89152 39.03761 0.5704893 0.04400276
```

With trained ARIMA model, we are going to forecast and plot total sales in next 5days.

```
> forecast(fit,5)
     Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
5.60       165379.4 104018.53 226740.2  71536.07 259222.6
5.80       112153.8  50792.95 173514.6  18310.49 205997.1
6.00       159545.1  98324.13 220766.1  65915.70 253174.6
6.20       214045.0 152823.96 275265.9 120415.52 307674.4
6.40       111295.8  50074.82 172516.8  17666.38 204925.2
> plot(forecast(fit,5))
```



The model built by the auto.arima function is ARIMA(0,0,0)(0,0,1). The coefficients, information criterion (used in comparing different models), and the error terms are given. A plot of the forecasted values show that there is a lot of fluctuation in the total sales, and hence we can see that the low and high values, which are 80% and 95%, have huge differences for the predicted values (depicted by the light grey and dark grey areas in the plot).

Similarly, the forecasting used for each dress to make individual predictions for the inventory. Before that, I have removed duplicate clothe ID and filled up with 0 for NAN value.

```
> indTotalSales <- read.csv("Dress Sales_Remove Duplicate ID.csv")
> indTotalSales[is.na(indTotalSales)] <- 0
> head(indTotalSales)
   Dress_ID X29.08.2013 X31.08.2013 X09.02.2013 X09.04.2013 X09.06.2013 X09.08.2013 X09.10.2013 X09.12.2013 X14.09.2013
1 444282011           0           0           0          76          76          78          80          82          81
2 510519284           0           0           0         190         192         195         197         204         205
3 511503677           0           0           0         623         630         636         640         645         645
4 520233308         524         550         594         603         608         615         616         635         641
5 522776523         432         466         491         510         513         527         534         559         574
6 531254082         203         309         341         348         351         355         355         362         365
  X16.09.2013 X18.09.2013 X20.09.2013 X22.09.2013 X24.09.2013 X26.09.2013 X28.09.2013 X30.09.2013 X10.02.2013 X10.04.2013
1          81          82          82          84          82          86          88          90          91          96
2         209         211         216         218         216           0         218           0           0           0
3         644         636         628         621         602         602         602         592         585         582
4         647         654         663         671         681           0         736           0           0           0
5         581         589         594         608         593           0         608           0           0           0
6         369         372         378         384         387           0         389           0           0           0
  X10.06.2013 X10.08.2010 X10.10.2013 X10.12.2013
1          96          98         101         102
2         213           0           0         204
3         570         564         555         547
4         785           0           0         818
5         619           0           0         616
6         403           0           0         416
```

Now we can forecast total sales for individual clothe.

```
> dataIndTotalSales = indTotalSales[,-1]
>
> i = 0
> n = nrow(dataIndTotalSales)
> tmp = 0
>
> while(i < n) {
+     matrixIndTotalSales <- as.matrix(dataIndTotalSales[i+1,])
+     numMatrixIndTotalSales <- as.numeric(as.vector(matrixIndTotalSales))
+     timeseries <- ts(numMatrixIndTotalSales, start = 1,frequency = 5)
+     fit <- auto.arima(timeseries)
+     predictedValue <-  as.data.frame(forecast(fit,3))
+     productwithpredictedvalue <- data.frame(dressid=indTotalSales[i+1,1],forecast1stday=predictedValue$`Point Forecast`[1],
+                                             forecast2ndday=predictedValue$`Point Forecast`[2],
+                                             forecast3rdday=predictedValue$`Point Forecast`[3])
+     tmp <- rbind(tmp,productwithpredictedvalue)
+     i <- i+1
+     print(tmp)
+ }
    dressid forecast1stday forecast2ndday forecast3rdday
1         0              0              0              0
2 444282011            102            102            102
    dressid forecast1stday forecast2ndday forecast3rdday
1         0         0.0000         0.0000         0.0000
2 444282011       102.0000       102.0000       102.0000
3 510519284       160.3375       112.6141       161.8902
    dressid forecast1stday forecast2ndday forecast3rdday
1         0         0.0000         0.0000         0.0000
2 444282011       102.0000       102.0000       102.0000
3 510519284       160.3375       112.6141       161.8902
4 511503677       526.2861       506.3566       487.1817
    dressid forecast1stday forecast2ndday forecast3rdday
1         0         0.0000         0.0000         0.0000
2 444282011       102.0000       102.0000       102.0000
3 510519284       160.3375       112.6141       161.8902
4 511503677       526.2861       506.3566       487.1817
5 520233308       336.4290       191.7102       388.8527
```

	dressid	forecast1stday	forecast2ndday	forecast3rdday
1	0	0.0000	0.00000	0.0000
2	444282011	102.0000	102.00000	102.0000
3	510519284	160.3375	112.61409	161.8902
4	511503677	526.2861	506.35656	487.1817
5	520233308	336.4290	191.71019	388.8527
6	522776523	194.0872	143.44370	378.2437
7	531254082	127.3409	99.18477	256.9353

Above data is just the sample only. By right, it will run 475 times and show the details in the table.

Analysis Tasks

To decide the pricing for various upcoming clothes, the store wishes 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

a. Firstly, calculate the total sales per dress ID and save it along with the Attribute DataSet file (with column name as Total.Sales). We need to find how style, season, and material affect the sale of a dress. Since they are categorical, let us first use the analysis of variances to see if the different types make an impact.

The required variables are:

Independent Variable: Total.Sales

Dependent Variables: Style, Season, Material

```
> dressSalesWTotalSales = read.csv('Dress Sales with Total Sales.csv')
> attributeClotheOrg = read.csv('Attribute DataSet.csv')
> indDressWTotalSales=cbind(attributeClotheOrg, Total.Sales=dressSalesWTotalSales$Total.sales)
> head(indDressWTotalSales)
    Dress_ID    Style   Price Rating Size Season NeckLine SleeveLength waiseline     Material FabricType Decoration
1 1006032852     Sexy     Low    4.6    M Summer   o-neck    sleevless    empire         null    chiffon    ruffles
2 1212192089   Casual     Low    0.0    L Summer   o-neck        Petal   natural  microfiber       null    ruffles
3 1190380701  vintage    High    0.0    L Automn   o-neck         full   natural    polyster       null       null
4  966005983    Brief Average    4.6    L Spring   o-neck         full   natural         silk    chiffon embroidary
5  876339541     cute     Low    4.5    M Summer   o-neck    butterfly   natural chiffonfabric    chiffon        bow
6 1068332458 bohemian     Low    0.0    M Summer   v-neck    sleevless    empire         null       null       null
  Pattern.Type Recommendation Total.Sales
1       animal              1       75979
2       animal              0       52256
3        print              0         223
4        print              1       39691
5          dot              0       44077
6        print              0         457
```

Let's find out how style, season, and material can affect Total Sales for individual cloth.

From the p-values we can see that out of the three, only season has a high p-value, thus showing that different seasons have a different impact on the sales.

```
> TestForStyle <- aov(Total.Sales ~ Style, data = indDressWTotalSales)
> summary(TestForStyle)
             Df    Sum Sq   Mean Sq F value Pr(>F)
Style        12 2.738e+09 228195052   1.527  0.111
Residuals   487 7.278e+10 149450545
>
> TestForSeason <- aov(Total.Sales ~ Season, data = indDressWTotalSales)
> summary(TestForSeason)
             Df    Sum Sq   Mean Sq F value  Pr(>F)
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
>
> TestForMaterial <- aov(Total.Sales ~ Material, data = indDressWTotalSales)
> summary(TestForMaterial)
             Df    Sum Sq   Mean Sq F value Pr(>F)
Material     24 3.221e+09 134204184   0.882  0.628
Residuals   475 7.230e+10 152210222
```

Next, we will try linear regression with Style, Season and Material to find out which of the factors affect the sales more.

```
> lmModel=lm(Total.Sales ~ Style+Season+Material, data = indDressWTotalSales)
> summary(lmModel)

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

Residuals:
   Min     1Q Median     3Q    Max 
-18248  -4652  -2017   1170 137639 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)   
(Intercept)             2458.6    15296.2   0.161   0.8724   
StyleBrief              6185.9     3886.8   1.591   0.1122   
StyleCasual             1770.3     2693.9   0.657   0.5114   
Stylecute               4255.2     3169.3   1.343   0.1801   
Stylefashion           -2462.1    12375.7  -0.199   0.8424   
StyleFlare             -1952.4     9111.6  -0.214   0.8304   
StyleNovelty             162.5     5045.3   0.032   0.9743   
StyleOL                -2299.6    12436.7  -0.185   0.8534   
Styleparty              -888.6     3125.8  -0.284   0.7763   
Stylesexy              13426.3     5615.3   2.391   0.0172 * 
StyleSexy               4815.1     2946.1   1.634   0.1029   
Stylevintage            4858.6     3520.3   1.380   0.1682   
Stylework               1534.8     3931.5   0.390   0.6964   
SeasonAutomn            3576.2     8751.6   0.409   0.6830   
SeasonAutumn            -903.7     9632.7  -0.094   0.9253   
Seasonspring           39044.1    12142.6   3.215   0.0014 **
SeasonSpring            4276.4     8650.9   0.494   0.6213   
Seasonsummer            3139.5    14828.4   0.212   0.8324   
SeasonSummer            2841.9     8645.0   0.329   0.7425   
Seasonwinter            -194.1     8806.0  -0.022   0.9824   
SeasonWinter            2682.4     8680.0   0.309   0.7574   
```

From below P-Value we can tell Style, Season and Material have big impact to Total Sales

```
Materialacrylic        -3574.8    14153.8  -0.253    0.8007
Materialcashmere       -4993.9    13731.5  -0.364    0.7163
Materialchiffonfabric   7298.1    12511.5   0.583    0.5600
Materialcotton         -2412.5    12279.1  -0.196    0.8443
Materialknitting       -3650.4    17300.2  -0.211    0.8330
Materiallace          -13165.1    17951.3  -0.733    0.4637
Materiallinen           -349.9    14155.0  -0.025    0.9803
Materiallycra          -1854.8    14087.4  -0.132    0.8953
Materialmicrofiber     11432.0    14073.0   0.812    0.4170
Materialmilksilk         954.1    13407.5   0.071    0.9433
Materialmix            -2706.2    12791.6  -0.212    0.8325
Materialmodal          -6349.9    17190.4  -0.369    0.7120
Materialmodel          -6681.9    17190.4  -0.389    0.6977
Materialnull           -1537.6    12263.1  -0.125    0.9003
Materialnylon          -3816.7    12876.8  -0.296    0.7671
Materialother          -3648.4    15047.8  -0.242    0.8085
Materialpolyster       -2001.9    12329.7  -0.162    0.8711
Materialrayon          -1855.1    12819.9  -0.145    0.8850
Materialshiffon        -4877.1    15100.6  -0.323    0.7469
Materialsilk           -3749.0    12493.5  -0.300    0.7643
Materialsill           -7396.7    17234.8  -0.429    0.6680
Materialspandex        -4822.7    13455.2  -0.358    0.7202
Materialviscos         -4803.1    14863.8  -0.323    0.7467
Materialwool           -2653.8    17325.2  -0.153    0.8783
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12070 on 455 degrees of freedom
Multiple R-squared:  0.1221,    Adjusted R-squared:  0.03719
F-statistic: 1.438 on 44 and 455 DF,  p-value: 0.03846
```

Secondly, to check if style is more influential than the price, let us construct a linear regression model as before, with only the attributes style and price.

Dependent Variable: Total.Sales

Independent Variables: Style, Price

```
> comStyPrModel <- lm(Total.Sales ~ Style + Price, data = indDressWTotalSales)
> summary(comStyPrModel)

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

Residuals:
   Min     1Q Median     3Q    Max 
-12610  -4953  -2411    602 143277 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)  
(Intercept)      561.03    9218.52   0.061   0.9515  
StyleBrief      6053.18    3809.42   1.589   0.1127  
StyleCasual     2916.15    2619.86   1.113   0.2662  
Stylecute       7551.39    3097.61   2.438   0.0151 *
Stylefashion   -1817.17   12434.48  -0.146   0.8839  
StyleFlare      -575.17    8979.89  -0.064   0.9490  
StyleNovelty     852.61    5010.34   0.170   0.8649  
StyleOL         1252.00   12620.71   0.099   0.9210  
Styleparty      3590.97    3317.13   1.083   0.2795  
Stylesexy      12377.13    5234.62   2.364   0.0185 *
StyleSexy       5176.98    2896.84   1.787   0.0746 .
Stylevintage    6121.80    3529.22   1.735   0.0835 .
Stylework       3949.46    3906.25   1.011   0.3125  
PriceAverage    1682.14    8889.79   0.189   0.8500  
Pricehigh      -2783.49    9219.08  -0.302   0.7628  
PriceHigh       1927.38   10150.10   0.190   0.8495  
Pricelow         -43.58    9083.79  -0.005   0.9962  
PriceLow        4537.90    8963.83   0.506   0.6129  
PriceMedium    -1384.03    9091.55  -0.152   0.8791  
Pricevery-high -2890.40    9012.44  -0.321   0.7486  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12160 on 480 degrees of freedom
Multiple R-squared:  0.05961,	Adjusted R-squared:  0.02239 
F-statistic: 1.602 on 19 and 480 DF,  p-value: 0.05153
```

Conclusion:

We can see that the **style** affects more compared to the price range. The price values hardly have any significance whereas the cute, sexy, and vintage style dresses positively affect the sales. The style ‘Sexy’ has a positive coefficient of 12377, from which we can safely conclude that sexy dresses are a safe bet when looking at the total sales. The p-value is almost 0.05 and hence we can conclude that these variables do affect the sales linearly.

However, the R-squared value is very low, specifying that these variables do not completely explain the significant changes in sales which means they cannot completely be used in predicting the total sales of dresses.

Analysis Tasks

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

We will again use linear regression, however, with all attribute variables to find which variables are most significant.

Dependent Variable: Total.Sales

Independent Variables: All the other variables, except Dress_ID

```
> modeluseAll <- lm(Total.Sales~.-Dress_ID,data = indDressWTotalSales)
> summary(modeluseAll)

Call:
lm(formula = Total.Sales ~ . - Dress_ID, data = indDressWTotalSales)

Residuals:
   Min     1Q Median     3Q    Max
-24359  -3794      0   1919  59098

Coefficients: (7 not defined because of singularities)
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          24704.90   17862.52   1.383  0.16754
StyleBrief            4874.30    3434.28   1.419  0.15671
StyleCasual           2096.88    2402.84   0.873  0.38345
Stylecute             1600.70    2854.88   0.561  0.57537
Stylefashion        -10126.75   18636.09  -0.543  0.58721
StyleFlare            2043.71    7842.89   0.261  0.79457
StyleNovelty          1737.32    4475.92   0.388  0.69815
StyleOL              15385.85   27300.01   0.564  0.57340
Styleparty            2172.36    3140.72   0.692  0.48961
Stylesexy             5348.73    5862.32   0.912  0.36220
StyleSexy             5628.23    2685.67   2.096  0.03684 *
Stylevintage          6121.19    3167.94   1.932  0.05415 .
Stylework             1804.85    3721.72   0.485  0.62802
PriceAverage          6754.38   11387.99   0.593  0.55349
Pricehigh             7296.15   11803.76   0.618  0.53690
PriceHigh             5407.89   12330.37   0.439  0.66124
Pricelow              4451.73   11567.35   0.385  0.70058
PriceLow              8196.48   11492.35   0.713  0.47620
PriceMedium           2799.93   11538.06   0.243  0.80841
Pricevery-high        6631.61   11793.75   0.562  0.57428
Rating                1236.47     259.59   4.763 2.81e-06 ***
SizeL                 4104.71    1520.05   2.700  0.00727 **
SizeM                 1186.31    1355.53   0.875  0.38210
Sizes               -11441.93   10707.61  -1.069  0.28601
SizeS                 1361.12    2112.62   0.644  0.51982
Sizesmall            -2138.72   11223.29  -0.191  0.84898
SizeXL                -178.18    3086.05  -0.058  0.95399
```

NeckLinepeterpan-collor	9690.80	25403.58	0.381	0.70309	
NeckLineruffled	150944.44	26865.54	5.619	3.97e-08	***
NeckLineScoop	8122.53	28961.38	0.280	0.77929	
NeckLineslash-neck	8527.08	24942.76	0.342	0.73266	
NeckLinesqare-collor	5682.05	24620.26	0.231	0.81762	
NeckLinesweetheart	13752.76	29058.81	0.473	0.63632	
NeckLineSweetheart	12628.88	24653.51	0.512	0.60880	
NeckLineturndowncollor	14496.57	25084.01	0.578	0.56369	
NeckLinev-neck	11237.33	24864.52	0.452	0.65159	
SleeveLengthcap-sleeves	-35827.44	13222.66	-2.710	0.00707	**
SleeveLengthcapsleeves	-33486.85	11988.62	-2.793	0.00551	**
SleeveLengthfull	-31076.84	10695.68	-2.906	0.00390	**
SleeveLengthhalf	-28542.05	18382.71	-1.553	0.12142	
SleeveLengthhalfsleeve	-29590.10	10755.18	-2.751	0.00625	**
SleeveLengthNULL	-23154.81	13742.64	-1.685	0.09291	.
SleeveLengthPetal	12969.91	18298.38	0.709	0.47893	
SleeveLengthshort	-32132.46	10530.82	-3.051	0.00246	**
SleeveLengthsleeevless	-41724.65	12768.75	-3.268	0.00119	**
SleeveLengthsleeveless	-34322.26	11836.31	-2.900	0.00397	**
SleeveLengthsleevless	-32975.66	10616.89	-3.106	0.00205	**
SleeveLengthsleveless	-38905.07	14646.61	-2.656	0.00827	**
SleeveLengththreequarter	-23959.51	11007.41	-2.177	0.03018	*
SleeveLengththreequater	-38754.71	17350.23	-2.234	0.02615	*
SleeveLengththressqatar	-33444.60	11122.03	-3.007	0.00283	**
SleeveLengthturndowncollor	-34870.83	14654.40	-2.380	0.01788	*
SleeveLengthurndowncollor	-39645.37	14621.42	-2.711	0.00703	**
Decorationapplique	-21074.89	11871.32	-1.775	0.07673	.
Decorationbeading	-25944.45	11862.27	-2.187	0.02940	*
Decorationbow	-19811.70	11810.51	-1.677	0.09436	.
Decorationbutton	-22991.52	12038.63	-1.910	0.05699	.
Decorationcascading	-16928.13	15219.10	-1.112	0.26679	
Decorationcrystal	-19369.51	15275.51	-1.268	0.20565	
Decorationdraped	-20105.50	15235.98	-1.320	0.18784	
Decorationembroidary	-17507.98	12694.60	-1.379	0.16874	
Decorationfeathers	-24561.89	13762.62	-1.785	0.07519	.
Decorationflowers	-25288.59	13450.45	-1.880	0.06093	.
Decorationhollowout	-22947.00	11736.28	-1.955	0.05136	.
Decorationlace	-21155.65	11541.54	-1.833	0.06766	.
Decorationnone	-21462.70	13654.65	-1.572	0.11691	
Decorationnull	-22178.38	11465.58	-1.934	0.05389	.
Decorationpearls	-21041.40	15726.74	-1.338	0.18180	
Decorationplain	-25071.10	15177.87	-1.652	0.09948	.
Decorationpleat	-33444.01	18586.91	-1.799	0.07284	.
Decorationpockets	-23620.37	12250.13	-1.928	0.05465	.
Decorationrivet	-21703.97	13017.56	-1.667	0.09637	.
Decorationruched	-26759.17	13279.98	-2.015	0.04468	*
Decorationruffles	-17410.62	11824.78	-1.472	0.14183	
Decorationsashes	-21381.24	11472.89	-1.864	0.06322	.
Decorationsequined	-21648.81	11811.62	-1.833	0.06769	.
Decorationtassel	-24952.86	15532.44	-1.607	0.10908	
DecorationTiered	NA	NA	NA	NA	

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9586 on 345 degrees of freedom
Multiple R-squared:  0.5802,	Adjusted R-squared:  0.3929
F-statistic: 3.097 on 154 and 345 DF,  p-value: < 2.2e-16
```

Since linear regression uses dummy variables for categorical variables, we have a lot of variables affecting the sales. Checking the output, we can make the following observations: –

a. Sexy style makes a positive impact

b. Rating is of a very high significance

c. Large size clothes are sold more

d. Spring season clothes make a positive impact on sales

e. Ruffled neckline clothes have a very significant positive impact

f. Sleeve length is significant; however, it affects the sale negatively (sold less)

g. Ruched and Beaded clothes make a negative impact on sale (sold less)

The multiple and adjusted R-squared values are 58% and 39% respectively. We can conclude that the model is quite robust and the p-value of almost 0 also suggests that there is definitely an impact of these variables on the total sales of the dresses.

Analysis Tasks

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

To find the relation between rating and total sales (both are numerical variables), perform a correlation of the two attributes

```
> cor.test(~Total.Sales+Rating,data= indDressWTotalSales)

        Pearson's product-moment correlation

data:  Total.Sales and Rating
t = 4.4531, df = 498, p-value = 1.046e-05
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1098837 0.2785985
sample estimates:
      cor
0.1956887
```

It can be clearly seen from the result that there is almost no correlation between the two variables. The correlation value is 0.2, which shows a very weak positive association, that is, a higher rating correlates with higher sales. Thus, the rating process has to be regularized.