Analysis of Sales Report of a Clothes Manufacturing Outlet

Andrei Enescu

10/12/2020

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.

Domain:

Retail

Dataset Description:

There are two files provided: Atribute DataSet.xlsx and Dress Sales.xlsx

Analysis Tasks:

There are 5 goals / questions for this project
1. 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.
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.
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.
4. 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.
5. To regularize the rating procedure and find its efficiency, the store wants to find if the rating of the dress affects the total sales

---

These are the packages / libraries used

```
library(plyr)
library(dplyr)
library(rio)
library(raster)
library(data.table)
library(lattice)
library(ggplot2)
library(caret)
library(caTools)
library(forecast)
library(corrplot)
```

First, I imported the two datasets:
1. Attribute DataSet.xlsx contains all the attributes of the dresses and also a recomandation column (binary value suggesting a recommendation (1) or not (0))
2. Dress Sales.xlsx contains the sales for each dress on a particular date. Date ranges from 29/8/2013 to 12/10/2013, and the sales are registered for alternative days.

```
Attrset <- import("Attribute DataSet.xlsx")
Sales.dates <- import("Dress Sales.xlsx")

View(Attrset)
str(Attrset)
```

```
## 'data.frame':    500 obs. of  14 variables:
##  $ Dress_ID      : num  1.01e+09 1.21e+09 1.19e+09 9.66e+08 8.76e+08 ...
##  $ Style         : chr  "Sexy" "Casual" "vintage" "Brief" ...
##  $ Price         : chr  "Low" "Low" "High" "Average" ...
##  $ Rating        : num  4.6 0 0 4.6 4.5 0 0 0 0 0 ...
##  $ Size          : chr  "M" "L" "L" "L" ...
##  $ Season        : chr  "Summer" "Summer" "Automn" "Spring" ...
##  $ NeckLine      : chr  "o-neck" "o-neck" "o-neck" "o-neck" ...
##  $ SleeveLength  : chr  "sleevless" "Petal" "full" "full" ...
##  $ waiseline     : chr  "empire" "natural" "natural" "natural" ...
##  $ Material      : chr  "null" "microfiber" "polyster" "silk" ...
##  $ FabricType    : chr  "chiffon" "null" "null" "chiffon" ...
##  $ Decoration    : chr  "ruffles" "ruffles" "null" "embroidary" ...
##  $ Pattern Type  : chr  "animal" "animal" "print" "print" ...
##  $ Recommendation: num  1 0 0 1 0 0 0 0 1 1 ...
```

Data Preparation

Converting from Character to Factor

```
Attrset <- mutate_if(Attrset, is.character, as.factor)
```

Renaming the factor levels as there were misstypes in the original dataset

```
levels(Attrset$Style)
```

```
##  [1] "bohemian" "Brief"    "Casual"   "cute"     "fashion"  "Flare"
##  [7] "Novelty"  "OL"       "party"    "sexy"     "Sexy"     "vintage"
## [13] "work"
```

```
Attrset$Style <-  revalue(Attrset$Style, c("sexy" = "Sexy"))

levels(Attrset$Price)
```

```
## [1] "Average"   "high"      "High"      "low"       "Low"       "Medium"
## [7] "very-high"
```

```
Attrset$Price <- revalue(Attrset$Price, c("high" = "High"))
Attrset$Price <- revalue(Attrset$Price, c("low" = "Low"))

levels(Attrset$Size)
```

```
## [1] "free"  "L"     "M"     "s"     "S"     "small" "XL"
```

```
Attrset$Size <- revalue(Attrset$Size, c("s" = "S"))
Attrset$Size <- revalue(Attrset$Size, c("small" = "S"))

levels(Attrset$Season)
```

```
## [1] "Automn" "Autumn" "spring" "Spring" "summer" "Summer" "winter" "Winter"
```

```
Attrset$Season <- revalue(Attrset$Season, c("Automn" = "Autumn"))
Attrset$Season <- revalue(Attrset$Season, c("spring" = "Spring"))
Attrset$Season <- revalue(Attrset$Season, c("summer" = "Summer"))
Attrset$Season <- revalue(Attrset$Season, c("winter" = "Winter"))

levels(Attrset$NeckLine)
```

```
##  [1] "backless"        "boat-neck"       "bowneck"         "halter"
##  [5] "mandarin-collor" "NULL"            "o-neck"          "open"
##  [9] "peterpan-collor" "ruffled"         "Scoop"           "slash-neck"
## [13] "sqare-collor"    "sweetheart"      "Sweetheart"      "turndowncollor"
## [17] "v-neck"
```

```
Attrset$NeckLine <-
  revalue(Attrset$NeckLine, c("sweetheart" = "Sweetheart"))

levels(Attrset$SleeveLength)
```

```
##  [1] "butterfly"      "cap-sleeves"    "capsleeves"     "full"
##  [5] "half"           "halfsleeve"     "NULL"           "Petal"
##  [9] "short"          "sleeevless"     "sleeveless"     "sleevless"
## [13] "sleveless"      "threequarter"   "threequater"    "thressqatar"
## [17] "turndowncollor" "urndowncollor"
```

```
Attrset$SleeveLength <- revalue(Attrset$SleeveLength,
                                c("cap-sleeves" = "capsleeves"))
Attrset$SleeveLength <- revalue(Attrset$SleeveLength,
                                c("half" = "halfsleeve"))
Attrset$SleeveLength <- revalue(Attrset$SleeveLength,
                                c("sleeevless" = "sleeveless"))
Attrset$SleeveLength <- revalue(Attrset$SleeveLength,
                                c("sleevless" = "sleeveless"))
Attrset$SleeveLength <- revalue(Attrset$SleeveLength,
                                c("sleveless" = "sleeveless"))
Attrset$SleeveLength <- revalue(Attrset$SleeveLength,
                                c("threequater" = "threequarter"))
Attrset$SleeveLength <- revalue(Attrset$SleeveLength,
                                c("thressqatar" = "threequarter"))
Attrset$SleeveLength <- revalue(Attrset$SleeveLength,
                                c("urndowncollor" = "turndowncollor"))

levels(Attrset$Material)
```

```
##  [1] "acrylic"       "cashmere"      "chiffonfabric" "cotton"
##  [5] "knitting"      "lace"          "linen"         "lycra"
##  [9] "microfiber"    "milksilk"      "mix"           "modal"
## [13] "model"         "null"          "nylon"         "other"
## [17] "polyster"      "rayon"         "shiffon"       "silk"
## [21] "sill"          "spandex"       "viscos"        "wool"
```

```
Attrset$Material <- revalue(Attrset$Material, c("model" = "modal"))

levels(Attrset$FabricType)
```

```
##  [1] "batik"      "broadcloth" "chiffon"    "Corduroy"   "dobby"
##  [6] "flannael"   "flannel"    "jersey"     "knitted"    "knitting"
## [11] "lace"       "null"       "organza"    "other"      "poplin"
## [16] "satin"      "sattin"     "shiffon"    "terry"      "tulle"
## [21] "wollen"     "woolen"     "worsted"
```

```
Attrset$FabricType <-
  revalue(Attrset$FabricType, c("shiffon" = "chiffon"))
Attrset$FabricType <-
  revalue(Attrset$FabricType, c("flannael" = "flannel"))
Attrset$FabricType <-
  revalue(Attrset$FabricType, c("knitting" = "knitted"))
Attrset$FabricType <-
  revalue(Attrset$FabricType, c("sattin" = "satin"))
Attrset$FabricType <-
  revalue(Attrset$FabricType, c("wollen" = "woolen"))

levels(Attrset$Decoration)
```

```
##  [1] "applique"   "beading"    "bow"        "button"     "cascading"
##  [6] "crystal"    "draped"     "embroidary" "feathers"   "flowers"
## [11] "hollowout"  "lace"       "none"       "null"       "pearls"
```

```
## [16] "plain"      "pleat"      "pockets"    "rivet"      "ruched"
## [21] "ruffles"    "sashes"     "sequined"   "tassel"     "Tiered"
```

```
Attrset$Decoration <-
  revalue(Attrset$Decoration, c("none" = "null"))

levels(Attrset$`Pattern Type`)
```

```
##  [1] "animal"    "character" "dot"       "floral"    "geometric" "leapord"
##  [7] "leopard"   "none"      "null"      "patchwork" "plaid"     "print"
## [13] "solid"     "splice"    "striped"
```

```
Attrset$`Pattern Type` <- revalue(Attrset$`Pattern Type`,
                                  c("leapord" = "leopard"))
Attrset$`Pattern Type` <- revalue(Attrset$`Pattern Type`,
                                  c("none" = "null"))
```

Second dataset - data preparation

```
View(Sales.dates)

setnames(
  Sales.dates,
  old = c(
    "41314",
    "41373",
    "41434",
    "41495",
    "41556",
    "41617",
    "41315",
    "41374",
    "41435",
    "40400",
    "41557",
    "41618"
  ),
  new = c(
    "2/9/2013",
    "4/9/2013",
    "6/9/2013",
    "8/9/2013",
    "10/9/2013",
    "12/9/2013",
    "2/10/2013",
    "4/10/2013",
    "6/10/2013",
    "8/10/2013",
    "10/10/2013",
    "12/10/2013"
  )
)
```

Replacing the NA values with the mean of the row and also replacing some character values with NA. After Rowmean was calculated I formatted the data frame to have the values without decimals

```
sapply(Sales.dates, class)
```

```
##    Dress_ID   29/8/2013   31/8/2013    2/9/2013    4/9/2013    6/9/2013
##   "numeric"   "numeric"   "numeric"   "numeric"   "numeric"   "numeric"
##    8/9/2013   10/9/2013   12/9/2013   14/9/2013   16/9/2013   18/9/2013
##   "numeric"   "numeric" "character" "character" "character" "character"
##   20/9/2013   22/9/2013   24/9/2013   26/9/2013   28/9/2013   30/9/2013
## "character" "character"   "numeric"   "numeric"   "numeric"   "numeric"
##   2/10/2013   4/10/2013   6/10/2013   8/10/2013  10/10/2013  12/10/2013
##   "numeric"   "numeric"   "numeric"   "numeric"   "numeric"   "numeric"
```

```
Sales.dates[Sales.dates == "removed"] <- NA
Sales.dates[Sales.dates == "Removed"] <- NA
Sales.dates[Sales.dates == "Orders"] <- NA

Sales.dates[9:14] <-
  mutate_if(Sales.dates[9:14], is.character , as.numeric)

Sales.dates <-
  (is.na(Sales.dates)) * rowMeans(Sales.dates[2:24], na.rm = T)[row(Sales.dates)] +
  replace(Sales.dates, is.na(Sales.dates), 0)

Sales.dates <-
  Sales.dates %>% mutate_at(vars(-Dress_ID), funs(round(., 0)))
```

Calculating the Total sales from the Sales.dates data frame (both row and column wise); this will be used later in the analysis

```
Totalsales <- rowSums(Sales.dates[2:24])
TotalSalescol <- colSums(Sales.dates[2:24])
Sales.dates <- data.frame(Sales.dates, Totalsales)
```

Merging the two data frames; this will be used later in the analysis

```
TotalDF <- merge(Attrset, Sales.dates)
```

Question 1

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.

Bellow I use a logistic regression model. For this the variables are:
1. Dependent variable: Recommendation
2. Remaining variables, except Dress_ID, which is an identifier

For this model I am creating a new data frame without the Dress_ID and NA rows

```r
Attrib <- Attrset[, -1]
Attrib <- na.omit(Attrib)

fit1 <-
  glm(
    Recommendation ~ Style + Price + Rating + Size + Season +
      NeckLine + SleeveLength + waiseline + Material + FabricType +
      Decoration + `Pattern Type`,
    family = binomial(),
    Attrib
  )
summary(fit1)
```

```
##
## Call:
## glm(formula = Recommendation ~ Style + Price + Rating + Size +
##     Season + NeckLine + SleeveLength + waiseline + Material +
##     FabricType + Decoration + `Pattern Type`, family = binomial(),
##     data = Attrib)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.26507  -0.88021  -0.00025   0.87680   2.29953
##
## Coefficients: (1 not defined because of singularities)
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -5.216e+01  1.248e+04  -0.004  0.99667
## StyleBrief                 -1.607e+00  8.715e-01  -1.844  0.06525 .
## StyleCasual                -4.059e-01  5.592e-01  -0.726  0.46785
## Stylecute                   1.289e-01  6.793e-01   0.190  0.84950
## Stylefashion               -2.310e+01  1.130e+04  -0.002  0.99837
## StyleFlare                 -1.790e-01  1.695e+00  -0.106  0.91589
## StyleNovelty                6.284e-01  1.032e+00   0.609  0.54242
## StyleOL                     1.435e+01  1.130e+04   0.001  0.99899
## Styleparty                 -3.687e-01  7.340e-01  -0.502  0.61540
## StyleSexy                  -5.087e-01  6.195e-01  -0.821  0.41155
## Stylevintage               -7.849e-01  7.947e-01  -0.988  0.32332
## Stylework                  -1.668e+00  9.669e-01  -1.725  0.08445 .
## PriceHigh                  -1.105e+00  7.174e-01  -1.541  0.12343
## PriceLow                    2.159e-01  2.877e-01   0.750  0.45308
## PriceMedium                 1.940e+00  6.269e-01   3.095  0.00197 **
## Pricevery-high              4.133e-01  8.352e-01   0.495  0.62072
## Rating                      1.590e-01  6.750e-02   2.356  0.01847 *
## SizeL                      -4.403e-01  3.655e-01  -1.205  0.22839
## SizeM                      -2.656e-02  3.193e-01  -0.083  0.93371
## SizeS                      -6.380e-01  4.990e-01  -1.279  0.20106
## SizeXL                      8.005e-03  8.152e-01   0.010  0.99217
## SeasonSpring                1.087e+00  4.014e-01   2.709  0.00675 **
## SeasonSummer               -1.898e-01  3.903e-01  -0.486  0.62681
## SeasonWinter                4.204e-01  3.948e-01   1.065  0.28688
## NeckLineboat-neck           3.588e+01  9.224e+03   0.004  0.99690
## NeckLinebowneck             3.225e+01  9.224e+03   0.003  0.99721
## NeckLinehalter              7.118e+01  1.305e+04   0.005  0.99565
```

```
## NeckLinemandarin-collor            NA         NA      NA       NA
## NeckLineNULL                 3.521e+01  9.224e+03   0.004  0.99695
## NeckLineo-neck               3.403e+01  9.224e+03   0.004  0.99706
## NeckLineopen                 3.837e+01  1.305e+04   0.003  0.99765
## NeckLinepeterpan-collor      1.580e+01  9.540e+03   0.002  0.99868
## NeckLineruffled              5.184e+01  1.130e+04   0.005  0.99634
## NeckLineScoop               -1.727e+00  9.224e+03   0.000  0.99985
## NeckLineslash-neck           3.416e+01  9.224e+03   0.004  0.99704
## NeckLinesqare-collor         3.280e+01  9.224e+03   0.004  0.99716
## NeckLineSweetheart           3.471e+01  9.224e+03   0.004  0.99700
## NeckLineturndowncollor       3.310e+01  9.224e+03   0.004  0.99714
## NeckLinev-neck               3.441e+01  9.224e+03   0.004  0.99702
## SleeveLengthcapsleeves       1.875e+01  6.523e+03   0.003  0.99771
## SleeveLengthfull             1.863e+01  6.523e+03   0.003  0.99772
## SleeveLengthhalfsleeve       1.850e+01  6.523e+03   0.003  0.99774
## SleeveLengthNULL             1.699e+01  6.523e+03   0.003  0.99792
## SleeveLengthPetal            1.894e+01  1.130e+04   0.002  0.99866
## SleeveLengthshort            1.758e+01  6.523e+03   0.003  0.99785
## SleeveLengthsleeveless       1.851e+01  6.523e+03   0.003  0.99774
## SleeveLengththreequarter     1.858e+01  6.523e+03   0.003  0.99773
## SleeveLengthturndowncollor   1.790e+01  6.523e+03   0.003  0.99781
## waiselineempire             -5.595e-01  1.319e+00  -0.424  0.67140
## waiselinenatural            -9.873e-01  1.305e+00  -0.757  0.44923
## waiselinenull               -1.295e+00  1.327e+00  -0.976  0.32928
## waiselineprincess            1.661e+01  6.523e+03   0.003  0.99797
## Materialcashmere             2.065e+01  3.112e+03   0.007  0.99471
## Materialchiffonfabric        1.821e+01  3.112e+03   0.006  0.99533
## Materialcotton               1.925e+01  3.112e+03   0.006  0.99506
## Materialknitting             7.496e-01  7.227e+03   0.000  0.99992
## Materiallace                -2.399e+00  1.172e+04   0.000  0.99984
## Materiallinen                2.489e-01  4.342e+03   0.000  0.99995
## Materiallycra                1.830e+01  3.112e+03   0.006  0.99531
## Materialmicrofiber           1.425e+00  7.227e+03   0.000  0.99984
## Materialmilksilk             1.769e+01  3.112e+03   0.006  0.99547
## Materialmix                  1.870e+01  3.112e+03   0.006  0.99521
## Materialmodal                1.854e+01  3.112e+03   0.006  0.99525
## Materialnull                 1.919e+01  3.112e+03   0.006  0.99508
## Materialnylon                1.986e+01  3.112e+03   0.006  0.99491
## Materialother                1.420e+00  7.227e+03   0.000  0.99984
## Materialpolyster             1.834e+01  3.112e+03   0.006  0.99530
## Materialrayon                2.042e+01  3.112e+03   0.007  0.99477
## Materialshiffon              1.822e+01  3.112e+03   0.006  0.99533
## Materialsilk                 1.866e+01  3.112e+03   0.006  0.99522
## Materialsill                 2.301e+00  7.227e+03   0.000  0.99975
## Materialspandex              1.720e+01  3.112e+03   0.006  0.99559
## Materialviscos               2.006e+01  3.112e+03   0.006  0.99486
## Materialwool                 5.108e-01  7.227e+03   0.000  0.99994
## FabricTypebroadcloth        -1.974e+01  4.300e+03  -0.005  0.99634
## FabricTypechiffon           -1.944e+01  4.300e+03  -0.005  0.99639
## FabricTypeCorduroy          -3.764e+01  6.274e+03  -0.006  0.99521
## FabricTypedobby             -2.051e+01  4.300e+03  -0.005  0.99620
## FabricTypeflannel           -3.835e+01  6.302e+03  -0.006  0.99514
## FabricTypejersey            -1.842e+01  4.300e+03  -0.004  0.99658
## FabricTypeknitted           -1.954e+01  4.300e+03  -0.005  0.99637
```

```
## FabricTypelace              3.494e+01  1.209e+04   0.003  0.99769
## FabricTypenull             -1.954e+01  4.300e+03  -0.005  0.99637
## FabricTypeorganza           1.440e+01  1.018e+04   0.001  0.99887
## FabricTypeother            -2.241e+01  1.018e+04  -0.002  0.99824
## FabricTypepoplin           -2.449e+00  7.813e+03   0.000  0.99975
## FabricTypesatin            -1.830e+01  4.300e+03  -0.004  0.99660
## FabricTypeterry            -2.408e+00  7.813e+03   0.000  0.99975
## FabricTypetulle            -1.941e+01  4.300e+03  -0.005  0.99640
## FabricTypewoolen           -1.690e+01  4.300e+03  -0.004  0.99686
## FabricTypeworsted          -1.945e+01  4.300e+03  -0.005  0.99639
## Decorationbeading           2.213e-01  9.245e-01   0.239  0.81084
## Decorationbow              -4.198e-01  1.110e+00  -0.378  0.70519
## Decorationbutton           -5.367e-01  1.340e+00  -0.400  0.68884
## Decorationcascading        -1.588e+01  6.523e+03  -0.002  0.99806
## Decorationcrystal           1.600e+01  6.523e+03   0.002  0.99804
## Decorationdraped           -1.912e+01  6.523e+03  -0.003  0.99766
## Decorationembroidary        1.689e+00  1.271e+00   1.328  0.18410
## Decorationfeathers          1.836e+01  4.609e+03   0.004  0.99682
## Decorationflowers           7.492e-01  1.507e+00   0.497  0.61899
## Decorationhollowout        -1.022e+00  9.619e-01  -1.063  0.28784
## Decorationlace              5.653e-01  7.324e-01   0.772  0.44026
## Decorationnull              3.733e-01  6.999e-01   0.533  0.59380
## Decorationpearls           -1.777e+01  6.523e+03  -0.003  0.99783
## Decorationplain             8.889e-01  1.918e+00   0.463  0.64302
## Decorationpockets           1.050e+00  1.243e+00   0.845  0.39828
## Decorationrivet            -1.782e+01  3.610e+03  -0.005  0.99606
## Decorationruched            3.459e-01  1.829e+00   0.189  0.85000
## Decorationruffles           5.561e-01  9.480e-01   0.587  0.55747
## Decorationsashes           -2.449e-01  7.734e-01  -0.317  0.75148
## Decorationsequined          8.959e-01  9.609e-01   0.932  0.35112
## Decorationtassel            1.878e+01  6.523e+03   0.003  0.99770
## DecorationTiered           -1.686e+01  6.523e+03  -0.003  0.99794
## ‘Pattern Type‘character    -1.725e+01  6.523e+03  -0.003  0.99789
## ‘Pattern Type‘dot           1.355e+00  9.964e-01   1.360  0.17392
## ‘Pattern Type‘floral       -1.931e+01  3.986e+03  -0.005  0.99614
## ‘Pattern Type‘geometric    -1.809e+00  1.546e+00  -1.170  0.24209
## ‘Pattern Type‘leopard       1.368e-01  1.344e+00   0.102  0.91894
## ‘Pattern Type‘null          5.639e-01  6.843e-01   0.824  0.40993
## ‘Pattern Type‘patchwork     2.232e-01  7.307e-01   0.305  0.76004
## ‘Pattern Type‘plaid        -1.748e+01  3.638e+03  -0.005  0.99617
## ‘Pattern Type‘print        -3.667e-01  6.858e-01  -0.535  0.59290
## ‘Pattern Type‘solid        -6.757e-02  6.691e-01  -0.101  0.91956
## ‘Pattern Type‘splice       -1.704e+01  6.523e+03  -0.003  0.99792
## ‘Pattern Type‘striped      -5.916e-01  9.402e-01  -0.629  0.52919
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 673.98  on 495  degrees of freedom
## Residual deviance: 484.96  on 372  degrees of freedom
## AIC: 732.96
##
## Number of Fisher Scoring iterations: 17
```

Observations related to Question 1

In the above model we can see that **Price, Season** and **Rating** are the main attributes of the dress that are affecting the recommendation.
Furthermore the Residual deviance is lower than the Null deviance, this means that using this model for prediction will be close to the actual values of recommendation.

I am also testing and validating the model. Splitting the data into 80% Training and 20% Testing

```
set.seed(1)
inTrain <-
  createDataPartition(Attrib$Recommendation, p = 0.8, list = FALSE)
Training <- Attrib[inTrain, ]
Testing <- Attrib[-inTrain, ]

fit2 <-
  glm(
    Recommendation ~ Style + Price + Rating + Size + Season + NeckLine +
      SleeveLength + waiseline + Material + FabricType + Decoration +
      `Pattern Type`,
    family = binomial(),
    Training
  )
summary(fit2)
```

```
##
## Call:
## glm(formula = Recommendation ~ Style + Price + Rating + Size +
##     Season + NeckLine + SleeveLength + waiseline + Material +
##     FabricType + Decoration + `Pattern Type`, family = binomial(),
##     data = Training)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.01772  -0.73477  -0.00014   0.67936   2.36359
##
## Coefficients: (3 not defined because of singularities)
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                3.805e+01  1.407e+04   0.003  0.99784
## StyleBrief                -1.499e+00  1.044e+00  -1.436  0.15112
## StyleCasual               -2.839e-01  7.392e-01  -0.384  0.70094
## Stylecute                  7.475e-01  9.105e-01   0.821  0.41165
## Stylefashion              -2.040e+01  1.019e+04  -0.002  0.99840
## StyleNovelty               1.741e+00  1.361e+00   1.279  0.20091
## StyleOL                   -2.119e+01  6.523e+03  -0.003  0.99741
## Styleparty                 6.085e-01  1.015e+00   0.600  0.54884
## StyleSexy                 -1.554e-01  8.096e-01  -0.192  0.84781
## Stylevintage              -5.352e-01  1.008e+00  -0.531  0.59540
## Stylework                 -2.726e+00  1.243e+00  -2.193  0.02831 *
## PriceHigh                 -1.323e+00  8.833e-01  -1.498  0.13410
## PriceLow                   1.706e-01  3.552e-01   0.480  0.63094
## PriceMedium                2.259e+00  7.602e-01   2.971  0.00296 **
## Pricevery-high             2.500e+00  1.352e+00   1.850  0.06434 .
## Rating                     2.816e-01  8.583e-02   3.281  0.00103 **
```

```
## SizeL                      -5.689e-01  4.435e-01  -1.283  0.19956
## SizeM                      -2.295e-01  3.824e-01  -0.600  0.54833
## SizeS                      -1.834e+00  6.937e-01  -2.644  0.00820 **
## SizeXL                     -5.743e-01  1.100e+00  -0.522  0.60150
## SeasonSpring                1.978e+00  5.047e-01   3.920 8.87e-05 ***
## SeasonSummer                5.615e-01  4.962e-01   1.132  0.25777
## SeasonWinter                8.954e-01  4.960e-01   1.805  0.07106 .
## NeckLinebowneck            -4.879e+00  1.748e+00  -2.792  0.00524 **
## NeckLinehalter              3.610e+01  9.224e+03   0.004  0.99688
## NeckLinemandarin-collor            NA         NA      NA       NA
## NeckLineNULL                1.710e+01  6.523e+03   0.003  0.99791
## NeckLineo-neck             -2.336e+00  8.612e-01  -2.713  0.00668 **
## NeckLineopen               -4.801e-01  7.829e+03   0.000  0.99995
## NeckLinepeterpan-collor    -1.975e+01  3.128e+03  -0.006  0.99496
## NeckLineScoop              -3.654e+01  9.224e+03  -0.004  0.99684
## NeckLineslash-neck         -2.914e+00  1.183e+00  -2.463  0.01377 *
## NeckLinesqare-collor       -2.214e+00  1.954e+00  -1.133  0.25703
## NeckLineSweetheart         -2.205e+00  1.503e+00  -1.467  0.14247
## NeckLineturndowncollor     -3.996e+00  1.394e+00  -2.866  0.00416 **
## NeckLinev-neck             -2.155e+00  9.153e-01  -2.354  0.01856 *
## SleeveLengthcapsleeves      1.488e+01  6.523e+03   0.002  0.99818
## SleeveLengthfull            1.896e+01  6.523e+03   0.003  0.99768
## SleeveLengthhalfsleeve      1.898e+01  6.523e+03   0.003  0.99768
## SleeveLengthNULL           -2.007e+01  1.130e+04  -0.002  0.99858
## SleeveLengthPetal           1.941e+01  1.130e+04   0.002  0.99863
## SleeveLengthshort           1.773e+01  6.523e+03   0.003  0.99783
## SleeveLengthsleeveless      1.868e+01  6.523e+03   0.003  0.99771
## SleeveLengththreequarter    1.859e+01  6.523e+03   0.003  0.99773
## SleeveLengthturndowncollor -1.211e+00  9.224e+03   0.000  0.99990
## waiselineempire            -5.627e+01  1.130e+04  -0.005  0.99603
## waiselinenatural           -5.745e+01  1.130e+04  -0.005  0.99594
## waiselinenull              -5.775e+01  1.130e+04  -0.005  0.99592
## waiselineprincess          -4.123e+01  1.305e+04  -0.003  0.99748
## Materialcashmere            2.244e+01  2.989e+03   0.008  0.99401
## Materialchiffonfabric       1.916e+01  2.989e+03   0.006  0.99489
## Materialcotton              2.002e+01  2.989e+03   0.007  0.99466
## Materialknitting            2.843e+00  7.175e+03   0.000  0.99968
## Materiallace                1.581e+00  1.062e+04   0.000  0.99988
## Materiallinen               1.100e+00  4.045e+03   0.000  0.99978
## Materiallycra               1.785e+01  2.989e+03   0.006  0.99524
## Materialmicrofiber          3.218e+00  7.175e+03   0.000  0.99964
## Materialmilksilk            1.864e+01  2.989e+03   0.006  0.99502
## Materialmix                 2.055e+01  2.989e+03   0.007  0.99451
## Materialmodal               1.873e+01  2.989e+03   0.006  0.99500
## Materialnull                2.011e+01  2.989e+03   0.007  0.99463
## Materialnylon               2.106e+01  2.989e+03   0.007  0.99438
## Materialother              -2.024e+01  8.380e+03  -0.002  0.99807
## Materialpolyster            1.891e+01  2.989e+03   0.006  0.99495
## Materialrayon               2.184e+01  2.989e+03   0.007  0.99417
## Materialshiffon             1.123e+00  7.175e+03   0.000  0.99988
## Materialsilk                2.029e+01  2.989e+03   0.007  0.99458
## Materialsill                2.208e+00  7.175e+03   0.000  0.99975
## Materialspandex             1.739e+01  2.989e+03   0.006  0.99536
## Materialviscos              2.091e+01  2.989e+03   0.007  0.99442
```

```
## Materialwool                1.722e+00  7.175e+03   0.000  0.99981
## FabricTypebroadcloth       -1.970e+01  4.330e+03  -0.005  0.99637
## FabricTypechiffon          -1.960e+01  4.330e+03  -0.005  0.99639
## FabricTypeCorduroy         -3.690e+01  7.829e+03  -0.005  0.99624
## FabricTypedobby            -2.079e+01  4.330e+03  -0.005  0.99617
## FabricTypeflannel          -3.866e+01  6.316e+03  -0.006  0.99512
## FabricTypejersey           -1.918e+01  4.330e+03  -0.004  0.99646
## FabricTypeknitted          -2.056e+01  4.330e+03  -0.005  0.99621
## FabricTypelace              3.127e+01  1.210e+04   0.003  0.99794
## FabricTypenull             -1.982e+01  4.330e+03  -0.005  0.99635
## FabricTypeorganza           1.128e+01  1.019e+04   0.001  0.99912
## FabricTypeother                    NA         NA      NA       NA
## FabricTypepoplin                   NA         NA      NA       NA
## FabricTypesatin            -1.732e+01  4.330e+03  -0.004  0.99681
## FabricTypeterry            -3.479e+00  7.829e+03   0.000  0.99965
## FabricTypetulle            -3.798e+01  7.829e+03  -0.005  0.99613
## FabricTypewoolen           -1.785e+01  4.330e+03  -0.004  0.99671
## FabricTypeworsted          -2.015e+01  4.330e+03  -0.005  0.99629
## Decorationbeading          -6.126e-01  1.409e+00  -0.435  0.66375
## Decorationbow              -9.982e-01  1.700e+00  -0.587  0.55716
## Decorationbutton           -4.136e-01  1.702e+00  -0.243  0.80794
## Decorationcascading        -1.460e+01  6.523e+03  -0.002  0.99821
## Decorationcrystal           1.440e+01  6.523e+03   0.002  0.99824
## Decorationdraped           -1.967e+01  6.523e+03  -0.003  0.99759
## Decorationembroidary        1.032e+00  1.752e+00   0.589  0.55593
## Decorationfeathers          1.905e+01  3.976e+03   0.005  0.99618
## Decorationflowers           2.928e-02  1.903e+00   0.015  0.98772
## Decorationhollowout        -5.412e-01  1.456e+00  -0.372  0.71014
## Decorationlace              1.073e+00  1.163e+00   0.923  0.35623
## Decorationnull              9.547e-01  1.132e+00   0.843  0.39913
## Decorationpearls           -1.643e+01  6.523e+03  -0.003  0.99799
## Decorationplain             2.337e+00  2.167e+00   1.078  0.28085
## Decorationpockets           2.810e+00  1.714e+00   1.639  0.10117
## Decorationrivet            -1.714e+01  3.741e+03  -0.005  0.99634
## Decorationruched           -1.619e+01  6.523e+03  -0.002  0.99802
## Decorationruffles           4.798e-01  1.449e+00   0.331  0.74055
## Decorationsashes           -2.093e-01  1.181e+00  -0.177  0.85936
## Decorationsequined          1.042e+00  1.395e+00   0.747  0.45494
## DecorationTiered           -1.734e+01  6.523e+03  -0.003  0.99788
## ‘Pattern Type‘character    -1.692e+01  6.523e+03  -0.003  0.99793
## ‘Pattern Type‘dot           2.086e+00  1.242e+00   1.679  0.09311 .
## ‘Pattern Type‘floral       -2.012e+01  3.660e+03  -0.005  0.99561
## ‘Pattern Type‘geometric    -1.292e+00  1.793e+00  -0.721  0.47113
## ‘Pattern Type‘leopard      -2.183e-01  1.692e+00  -0.129  0.89736
## ‘Pattern Type‘null          5.542e-01  8.491e-01   0.653  0.51400
## ‘Pattern Type‘patchwork     8.089e-01  9.221e-01   0.877  0.38032
## ‘Pattern Type‘plaid        -1.652e+01  3.686e+03  -0.004  0.99642
## ‘Pattern Type‘print        -2.400e-01  8.429e-01  -0.285  0.77587
## ‘Pattern Type‘solid         1.242e-01  8.213e-01   0.151  0.87978
## ‘Pattern Type‘striped      -4.040e-02  1.131e+00  -0.036  0.97150
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 537.59  on 396  degrees of freedom
## Residual deviance: 337.88  on 280  degrees of freedom
## AIC: 571.88
##
## Number of Fisher Scoring iterations: 17
```

Creating a new column for prediction of recommendation

```
Testing$Predict.recommend <- 0
```

There were some errors after the split, because some levels of the attributes had values only in the Testing set. The values are small and the impact on predication is insignificant so I decided to have the same values for the bellow levels in both, Training and in Testing data sets

```
fit2$xlevels[["Style"]] <-
  union(fit2$xlevels[["Style"]], levels(Testing[["Style"]]))
fit2$xlevels[["NeckLine"]] <-
  union(fit2$xlevels[["NeckLine"]], levels(Testing[["NeckLine"]]))
fit2$xlevels[["Decoration"]] <-
  union(fit2$xlevels[["Decoration"]], levels(Testing[["Decoration"]]))
fit2$xlevels[["Pattern Type"]] <-
  union(fit2$xlevels[["Pattern Type"]], levels(Testing[["Pattern Type"]]))

Pred <- predict(fit2, Testing, type = "response")
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from a rank-deficient fit may be misleading
```

Setting threshold as 0.7 for the probability

```
Testing$Predict.recommend <- ifelse(Pred >= 0.7, 1, 0)
Testing$Predict.recommend
```

```
##  [1] 1 0 1 1 1 0 1 1 0 1 0 1 0 1 1 1 1 1 0 1 0 1 1 1 0 0 0 0 1 1 0 1 0 0 0 1 1 1
## [39] 0 1 0 0 1 1 0 0 1 1 1 1 0 1 0 0 1 0 1 0 0 1 0 1 1 0 0 1 1 1 1 1 0 1 0 1 0 1
## [77] 0 1 1 0 0 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0
```

Testing by confusion matrix

```
table(Testing$Recommendation,
      Testing$Predict.recommend,
      dnn = list("Actual", "Predicted"))
```

```
##       Predicted
## Actual  0  1
##      0 25 30
##      1 16 28
```

```
conf.matrix <-
  table(Testing$Recommendation, Testing$Predict.recommend)
```

Accuracy and error of the model:

```
sum(conf.matrix)
```

```
## [1] 99
```

```
sum(diag(conf.matrix)) / sum(conf.matrix)
```

```
## [1] 0.5353535
```

```
1 - sum(diag(conf.matrix)) / sum(conf.matrix)
```

```
## [1] 0.4646465
```

Overall observation

The accuracy of the test model is only 53.54%

Question 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.

In order to predict the trend of total sales for an extended period of three more days I used the *auto.arima* function.
In this part of the analysis I am using the *TotalSalescol* created in the Data Preparation part.

```
is.vector(TotalSalescol)
```

```
## [1] TRUE
```

```
TotalSalescol
```

```
##  29/8/2013  31/8/2013   2/9/2013   4/9/2013   6/9/2013   8/9/2013  10/9/2013
##      94883     100483     107081     149336     151829     157647     159391
##  12/9/2013  14/9/2013  16/9/2013  18/9/2013  20/9/2013  22/9/2013  24/9/2013
##     166330     169469     172094     174728     179405     184261     185616
##  26/9/2013  28/9/2013  30/9/2013  2/10/2013  4/10/2013  6/10/2013  8/10/2013
##     172853     193734     172008     173339     174193     206334     175684
## 10/10/2013 12/10/2013
##     176625     215533
```

```r
Trend.Pred <- ts(TotalSalescol, start = 1, frequency = 7)
summary(Trend.Pred)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   94883  158519  172853  165776  178015  215533
```

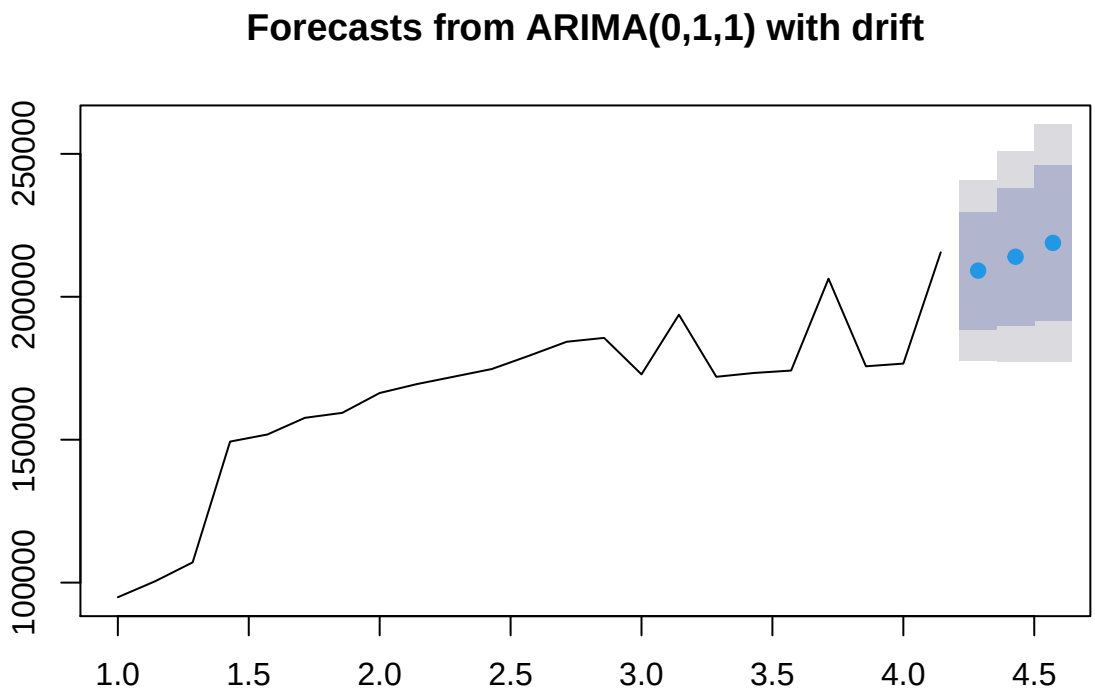
```r
Trend.Pred <- auto.arima(Trend.Pred)
summary(Trend.Pred)
```

```
## Series: Trend.Pred
## ARIMA(0,1,1) with drift
##
## Coefficients:
##           ma1     drift
##       -0.3944  4846.298
## s.e.   0.1938  2072.032
##
## sigma^2 estimated as 258352544:  log likelihood=-243.32
## AIC=492.64   AICc=493.98   BIC=495.92
##
## Training set error measures:
##                     ME     RMSE      MAE       MPE    MAPE      MASE        ACF1
## Training set 191.8268 14988.48 10419.99 0.1309597 5.99801 0.4103329 -0.01880812
```

```r
forecast(Trend.Pred, 3)
```

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 4.285714       209155.1 188556.3 229754.0 177652.0 240658.3
## 4.428571       214001.4 189920.1 238082.8 177172.3 250830.6
## 4.571429       218847.7 191727.5 245968.0 177370.9 260324.6
```

```r
plot(forecast(Trend.Pred, 3))
```



Observation

As we can see from the plot and the ARIMA model output, there is a high fluctuation in the confidence interval

---

Question 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.

In this part of the analysis I am using the *TotalSales* created in the Data Preparation part.
Also I am using the ANOVA model to find the effect of **Style, Season** and **Material** on Total Sales.
After ANOVA I am using the linear regression model in order to find the relation of the three attributes on Total Sales and then just for the **Price** and **Styele**

```
ANOVA.Style <- aov(TotalDF$Totalsales ~ TotalDF$Style)
ANOVA.Season <- aov(TotalDF$Totalsales ~ TotalDF$Season)
ANOVA.Material <- aov(TotalDF$Totalsales ~ TotalDF$Material)
summary(ANOVA.Style)
```

```
##                Df    Sum Sq   Mean Sq F value Pr(>F)
## TotalDF$Style  11 2.920e+09 265472353    1.48  0.135
## Residuals     538 9.647e+10 179318899
```

```r
summary(ANOVA.Season)
```

```
##                  Df    Sum Sq   Mean Sq F value Pr(>F)
## TotalDF$Season    3 9.663e+08 322084143   1.782   0.15
## Residuals       544 9.835e+10 180782604
## 2 observations deleted due to missingness
```

```r
summary(ANOVA.Material)
```

```
##                   Df    Sum Sq   Mean Sq F value Pr(>F)
## TotalDF$Material  22 3.643e+09 165604109    0.91  0.582
## Residuals        526 9.575e+10 182033429
## 1 observation deleted due to missingness
```

```r
fit3 <- lm(Totalsales ~ Style + Season + Material, TotalDF)
summary(fit3)
```

```
##
## Call:
## lm(formula = Totalsales ~ Style + Season + Material, data = TotalDF)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -20077  -5684  -2598   1069 135810
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1587.9     8504.2   0.187   0.8520
## StyleBrief                6654.8     4100.8   1.623   0.1052
## StyleCasual               2165.0     2854.8   0.758   0.4486
## Stylecute                 4618.9     3390.0   1.363   0.1736
## Stylefashion             -3169.1    13756.1  -0.230   0.8179
## StyleFlare               -3851.7    10115.5  -0.381   0.7035
## StyleNovelty             -1204.4     4837.4  -0.249   0.8035
## StyleOL                  -2547.3    13805.3  -0.185   0.8537
## Styleparty               -1182.1     3355.5  -0.352   0.7248
## StyleSexy                 5916.3     3084.5   1.918   0.0557 .
## Stylevintage              5246.8     3801.2   1.380   0.1681
## Stylework                 1917.8     4221.3   0.454   0.6498
## SeasonSpring              1705.6     2030.4   0.840   0.4013
## SeasonSummer              -917.0     1923.1  -0.477   0.6337
## SeasonWinter             -1655.8     1970.0  -0.841   0.4010
## Materialcashmere          -592.7    10397.0  -0.057   0.9546
## Materialchiffonfabric    12204.9     8239.3   1.481   0.1391
## Materialcotton            3044.2     7921.0   0.384   0.7009
## Materialknitting          2569.5    12391.0   0.207   0.8358
## Materiallace              1011.9    15724.1   0.064   0.9487
## Materiallinen             5770.6    11170.8   0.517   0.6057
## Materiallycra             2498.8    11090.6   0.225   0.8218
## Materialmicrofiber       15697.3    11119.9   1.412   0.1587
## Materialmilksilk          5905.0     9921.9   0.595   0.5520
## Materialmix               2448.9     8687.6   0.282   0.7782
```

```
## Materialmodal          -2280.8    12379.5  -0.184   0.8539
## Materialnull            3668.6     7933.5   0.462   0.6440
## Materialnylon           1408.3     9040.0   0.156   0.8763
## Materialother           1907.0    12391.0   0.154   0.8777
## Materialpolyster        3415.7     7981.7   0.428   0.6689
## Materialrayon           3676.2     8939.3   0.411   0.6811
## Materialshiffon        -1303.0    11131.5  -0.117   0.9069
## Materialsilk             184.4     8238.7   0.022   0.9822
## Materialsill           -2908.1    15673.0  -0.186   0.8529
## Materialspandex         -106.1     9906.2  -0.011   0.9915
## Materialviscos           213.5    12415.2   0.017   0.9863
## Materialwool            -716.0    15617.8  -0.046   0.9635
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13440 on 510 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.07173,    Adjusted R-squared:  0.006207
## F-statistic: 1.095 on 36 and 510 DF,  p-value: 0.3278
```

```
fit4 <- lm(Totalsales ~ Price + Style, TotalDF)
summary(fit4)
```

```
##
## Call:
## lm(formula = Totalsales ~ Price + Style, data = TotalDF)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -12168  -6090  -2943    348 143719
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3740.51    2705.97   1.382   0.1675
## PriceHigh      -3664.79    3165.86  -1.158   0.2475
## PriceLow        1402.63    1310.78   1.070   0.2851
## PriceMedium    -3226.27    2593.32  -1.244   0.2140
## Pricevery-high -5568.54    3513.34  -1.585   0.1136
## StyleBrief      6690.38    3996.23   1.674   0.0947 .
## StyleCasual     2867.40    2758.94   1.039   0.2991
## Stylecute       7065.36    3256.84   2.169   0.0305 *
## Stylefashion   -3194.51   13654.39  -0.234   0.8151
## StyleFlare     -2072.51    9842.88  -0.211   0.8333
## StyleNovelty    -418.07    4711.49  -0.089   0.9293
## StyleOL          -85.24   13856.34  -0.006   0.9951
## Styleparty      3103.80    3498.00   0.887   0.3753
## StyleSexy       5782.36    2997.04   1.929   0.0542 .
## Stylevintage    6506.45    3742.36   1.739   0.0827 .
## Stylework       3466.16    4131.97   0.839   0.4019
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13380 on 532 degrees of freedom
##   (2 observations deleted due to missingness)
```

```
## Multiple R-squared:  0.04121,    Adjusted R-squared:  0.01418
## F-statistic: 1.524 on 15 and 532 DF,  p-value: 0.09161
```

Observation

By using the ANOVA model individually on the three attributes we can conclude that Style has the lowest p value (0.135) among the three attributes so it has the most impact in the Total Sales. In the linear regression model *fit3* we can see that the Style Sexy is affecting (in a low manner) the Total Sales. Also for model *fit4*, the Style has a higher impact then the price in the Total Sales of the dress. However both models have a low R and R adjusted values, which means that the models using only these attributes can't be used in predicting the Total Sales.

Question 4

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.

I am creating a new linear regression models with TotalSales as a dependent variable and the rest (except Dress_ID) as independent variable.

```
TotalData <- TotalDF[,-c(1, 15:37)]
fit5 <- lm(Totalsales ~ ., TotalData)
summary(fit5)
```

```
##
## Call:
## lm(formula = Totalsales ~ ., data = TotalData)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -34442  -5036  -1087   2345  83797
##
## Coefficients: (1 not defined because of singularities)
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 15241.62   25233.85   0.604 0.546159
## StyleBrief                   6615.57    3752.51   1.763 0.078631 .
## StyleCasual                  2701.59    2603.62   1.038 0.300038
## Stylecute                    2648.16    3141.14   0.843 0.399676
## Stylefashion                -9333.60   21741.81  -0.429 0.667930
## StyleFlare                   -360.52    9128.12  -0.039 0.968514
## StyleNovelty                 1373.93    4496.94   0.306 0.760116
## StyleOL                      8699.48   21147.39   0.411 0.681008
## Styleparty                   3087.53    3442.69   0.897 0.370319
## StyleSexy                    6604.32    2898.17   2.279 0.023180 *
## Stylevintage                 6687.10    3533.75   1.892 0.059130 .
## Stylework                    1338.08    4005.21   0.334 0.738482
## PriceHigh                   -1177.91    3190.11  -0.369 0.712135
## PriceLow                      441.33    1328.18   0.332 0.739845
## PriceMedium                 -3166.26    2598.40  -1.219 0.223700
```

```
## Pricevery-high              -2148.82    3885.71  -0.553 0.580552
## Rating                       1289.62     292.37   4.411 1.31e-05 ***
## SizeL                        5962.77    1677.52   3.555 0.000422 ***
## SizeM                        1517.27    1487.78   1.020 0.308398
## SizeS                        1858.26    2285.92   0.813 0.416727
## SizeXL                       -159.85    3429.84  -0.047 0.962850
## SeasonSpring                 1320.34    1839.67   0.718 0.473338
## SeasonSummer                 -569.42    1779.04  -0.320 0.749071
## SeasonWinter                -1651.50    1811.18  -0.912 0.362377
## NeckLineboat-neck            2661.32   17222.16   0.155 0.877267
## NeckLinebowneck              4435.72   17537.69   0.253 0.800449
## NeckLinehalter               5544.20   23342.48   0.238 0.812372
## NeckLinemandarin-collor           NA         NA      NA       NA
## NeckLineNULL                 -69.36   19581.05  -0.004 0.997175
## NeckLineo-neck               6620.75   16911.91   0.391 0.695637
## NeckLineopen                15556.58   24535.34   0.634 0.526394
## NeckLinepeterpan-collor      5181.20   17485.12   0.296 0.767131
## NeckLineruffled            143396.32   20677.36   6.935 1.54e-11 ***
## NeckLineScoop                1647.37   16881.61   0.098 0.922309
## NeckLineslash-neck           2836.08   17179.69   0.165 0.868957
## NeckLinesqare-collor         -779.20   18035.81  -0.043 0.965560
## NeckLineSweetheart           7493.45   17690.03   0.424 0.672076
## NeckLineturndowncollor       7542.24   17278.35   0.437 0.662688
## NeckLinev-neck               6211.88   16987.56   0.366 0.714793
## SleeveLengthcapsleeves     -31244.75   13164.99  -2.373 0.018077 *
## SleeveLengthfull           -28709.98   12388.40  -2.317 0.020956 *
## SleeveLengthhalfsleeve     -27608.16   12433.01  -2.221 0.026913 *
## SleeveLengthNULL           -21048.85   16017.86  -1.314 0.189533
## SleeveLengthPetal           11482.57   21132.48   0.543 0.587169
## SleeveLengthshort          -30219.64   12213.89  -2.474 0.013747 *
## SleeveLengthsleeveless     -31296.91   12297.93  -2.545 0.011287 *
## SleeveLengththreequarter   -26096.38   12485.07  -2.090 0.037199 *
## SleeveLengthturndowncollor -33833.61   14842.29  -2.280 0.023134 *
## waiselineempire              7679.85    6738.10   1.140 0.255033
## waiselinenatural             5440.48    6661.84   0.817 0.414583
## waiselinenull                7368.57    6758.89   1.090 0.276247
## waiselineprincess            2633.90   15022.92   0.175 0.860908
## Materialcashmere              407.27    9449.60   0.043 0.965643
## Materialchiffonfabric        5074.80    7391.75   0.687 0.492745
## Materialcotton               1729.72    7040.30   0.246 0.806043
## Materialknitting               52.97   11028.09   0.005 0.996170
## Materiallace               -10351.76   22224.74  -0.466 0.641616
## Materiallinen                7127.23    9993.01   0.713 0.476104
## Materiallycra                4661.37   10099.16   0.462 0.644635
## Materialmicrofiber           -938.28   13845.49  -0.068 0.946003
## Materialmilksilk             4367.78    8848.43   0.494 0.621831
## Materialmix                  -184.73    7818.07  -0.024 0.981160
## Materialmodal               -3340.83   11951.88  -0.280 0.779981
## Materialnull                 3235.72    6994.19   0.463 0.643869
## Materialnylon                2317.44    8142.30   0.285 0.776078
## Materialother                4076.03   14075.21   0.290 0.772273
## Materialpolyster             1327.43    7081.39   0.187 0.851395
## Materialrayon                3962.12    7936.10   0.499 0.617862
## Materialshiffon             -4951.20   10090.73  -0.491 0.623917
```

```
## Materialsilk               -1554.79    7328.78  -0.212 0.832093
## Materialsill               -3541.12   13759.49  -0.257 0.797028
## Materialspandex            -3723.38    9090.70  -0.410 0.682321
## Materialviscos               222.06   11291.14   0.020 0.984319
## Materialwool               -1939.29   13648.74  -0.142 0.887080
## FabricTypebroadcloth        3405.25    8660.17   0.393 0.694365
## FabricTypechiffon             65.26    8529.52   0.008 0.993899
## FabricTypeCorduroy          1987.75   11886.05   0.167 0.867266
## FabricTypedobby           -16017.24   11178.14  -1.433 0.152627
## FabricTypeflannel           8468.71   11867.68   0.714 0.475873
## FabricTypejersey            5790.97    9137.64   0.634 0.526589
## FabricTypeknitted           2870.57   12504.73   0.230 0.818546
## FabricTypelace              7877.02   22223.99   0.354 0.723188
## FabricTypenull             -1062.72    8454.90  -0.126 0.900036
## FabricTypeorganza          -4572.08   19284.73  -0.237 0.812708
## FabricTypeother             3682.75   19083.74   0.193 0.847069
## FabricTypepoplin          -10592.80   14990.09  -0.707 0.480173
## FabricTypesatin              41.02   10043.63   0.004 0.996743
## FabricTypeterry            -6428.51   15238.05  -0.422 0.673334
## FabricTypetulle             5153.42   12761.73   0.404 0.686551
## FabricTypewoolen           -1017.27   10895.38  -0.093 0.925657
## FabricTypeworsted           -799.77    8817.09  -0.091 0.927769
## Decorationbeading          -1933.15    4212.32  -0.459 0.646523
## Decorationbow               3651.07    4597.87   0.794 0.427598
## Decorationbutton             542.68    5433.93   0.100 0.920496
## Decorationcascading         2940.40   11954.91   0.246 0.805835
## Decorationcrystal           4926.03   12445.38   0.396 0.692444
## Decorationdraped            3915.34   12054.11   0.325 0.745484
## Decorationembroidary        5370.20    6616.93   0.812 0.417488
## Decorationfeathers         -1958.20    9085.03  -0.216 0.829449
## Decorationflowers           -788.57    7115.96  -0.111 0.911814
## Decorationhollowout         -277.64    4004.39  -0.069 0.944756
## Decorationlace              2408.79    3239.42   0.744 0.457541
## Decorationnull               686.05    3060.43   0.224 0.822736
## Decorationpearls            5134.49   12742.00   0.403 0.687184
## Decorationplain            -3177.41    9909.30  -0.321 0.748635
## Decorationpockets          -1602.86    6177.41  -0.259 0.795398
## Decorationrivet             1475.73    7578.03   0.195 0.845692
## Decorationruched           -5572.74    8316.23  -0.670 0.503159
## Decorationruffles           6935.31    4220.56   1.643 0.101084
## Decorationsashes            3438.46    3453.17   0.996 0.319948
## Decorationsequined         -2683.87    4389.39  -0.611 0.541234
## Decorationtassel           -6543.68   12305.32  -0.532 0.595161
## DecorationTiered           31109.41   13776.87   2.258 0.024451 *
## `Pattern Type`character    -1609.37   11801.41  -0.136 0.891593
## `Pattern Type`dot          -1857.61    4564.51  -0.407 0.684238
## `Pattern Type`floral         949.63    8719.14   0.109 0.913323
## `Pattern Type`geometric    14286.73    6377.53   2.240 0.025601 *
## `Pattern Type`leopard      -2897.95    6077.86  -0.477 0.633749
## `Pattern Type`null         -5901.52    3103.46  -1.902 0.057906 .
## `Pattern Type`patchwork    -4335.86    3360.15  -1.290 0.197628
## `Pattern Type`plaid        -6456.94    7479.72  -0.863 0.388486
## `Pattern Type`print        -2605.65    3147.40  -0.828 0.408212
## `Pattern Type`solid        -2149.40    3028.17  -0.710 0.478220
```

```
## `Pattern Type`splice        -5723.22    9554.25  -0.599 0.549480
## `Pattern Type`striped       -3087.51    4117.66  -0.750 0.453782
## Recommendation                907.02    1168.51   0.776 0.438055
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11310 on 421 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.4578, Adjusted R-squared:  0.2981
## F-statistic: 2.867 on 124 and 421 DF,  p-value: 1.271e-15
```

Observation

We can see in the output of the model that we have a lot of variables affecting the Total Sales. We can split the variables in high and moderate impact and also positive and negative impact.

1. High - positive impact variables

a) Rating
b) Size L
c) NeckLine ruffled

2. Moderate - positive impact variables

a) Style Sexy
b) Decoration Tiered
c) Pattern Type geometric

3. Moderate - negative impact

a) SleevLength

The multiple and adjusted R-squared values are 46% and 30% respectively. We can conclude that the model is quite fit 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.

Question 5

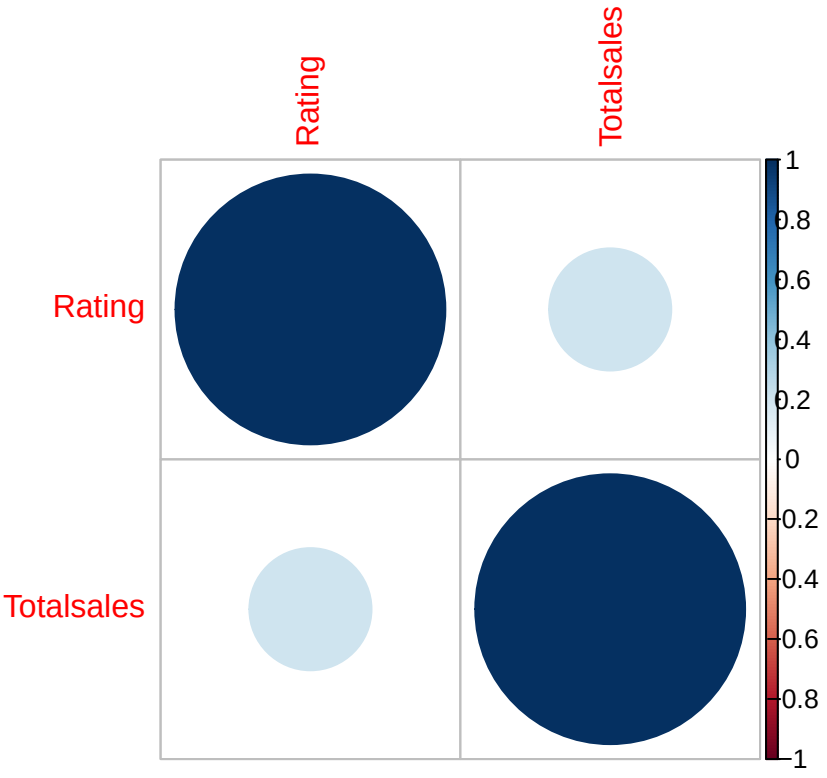
To regularize the rating procedure and find its efficiency, the store wants to find if the rating of the dress affects the total sales

In order to find if there is a relation between *Rating* and *Total Sales*, I use *cor.test* function.

```
corr <- cor.test(TotalDF$Totalsales, TotalDF$Rating)
corr
```

```
##
##  Pearson's product-moment correlation
##
## data:  TotalDF$Totalsales and TotalDF$Rating
## t = 4.8992, df = 548, p-value = 1.269e-06
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1233527 0.2835959
## sample estimates:
##       cor
## 0.2048465
```

```r
corr4plot <- cor(TotalDF[, c(4, 38)])
corrplot(corr4plot)
```



Observation

Based on the output of this function, we have a correlation of 0.2 between the Total Sales and the Rating. This means that there is a weak - positive correlation between the two attributes. In conclusion a higher rating will correlate with higher sales which suggests that the rating approach requires regularization.