Analyze the Sales Report of a Clothes Manufacturing Outlet

Problem Statement

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.

Analysis Approach :

A Logistic regression Model to Classify the recommendation Outcome based on the Attributes is created for Analysis .

Conclusions:

From the significance codes for each attribute, we can see that very few attributes have a major impact on the Predicting the Outcome .

However of the different Independent Variables used in the Analysis only a few appear to significant Impact like Price ,Fabric Type,
Material & SleeveLength.

(Intercept)	-0.077393	1.071343	-0.072	0.94245	
StyleBrief	-0.256643	0.168631	-1.522	0.12894	
StyleCasual	-0.100074	0.118257	-0.846	0.39801	
Stylecute	0.018336	0.140647	0.130	0.89635	
Stylefashion	-1.415854	0.914973	-1.547	0.12267	
StyleFlare	-0.027686	0.386389	-0.072	0.94292	
StyleNovelty	0.079978	0.220471	0.363	0.71701	
StyleOL	-0.236936	0.893565	-0.265	0.79105	
Styleparty	-0.092998	0.154652	-0.601	0.54801	
Stylesexy	0.025667	0.288813	0.089	0.92924	
StyleSexy	-0.094051	0.132217	-0.711	0.47736	
Stylevintage	-0.119594	0.155941	-0.767	0.44365	
Stylework	-0.293947	0.182674	-1.609	0.10850	
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	
Rating	0.020516	0.012741	1.610	0.10826	
SizeL	-0.085147	0.074748	-1.139	0.25544	
SizeM	-0.009807	0.066780	-0.147	0.88333	
Sizes	-0.615002	0.526489	-1.168	0.24356	
SizeS	-0.044350	0.104054	-0.426	0.67021	
Sizesmall	-0.835915	0.551102	-1.517	0.13023	
SizeXL	0.001798	0.152039	0.012	0.99057	
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	
NeckLineboat-neck	0.675626	0.725180	0.932	0.35216	
NeckLinebowneck	-0.052399	0.742296	-0.071	0.94376	
NeckLinehalter	1.235546	0.979549	1.261	0.20804	
NeckLinemandarin-collor	NA	NA	NA	NA	

```
SleeveLengthsleveless        1.242864   0.718486   1.730  0.08455 .
SleeveLengththreequarter     0.395897   0.541878   0.731  0.46552
SleeveLengththreequater      0.920010   0.853352   1.078  0.28173
SleeveLengththressqatar      0.343183   0.547632   0.627  0.53129
SleeveLengthturndowncollor   0.834707   0.720575   1.158  0.24750
SleeveLengthurndowncollor   -0.135579   0.720309  -0.188  0.85081
waiselineempire             -0.134535   0.285102  -0.472  0.63731
waiselinenatural            -0.201059   0.281484  -0.714  0.47554
waiselinenull               -0.251726   0.286949  -0.877  0.38096
waiselineprincess            0.349400   0.661240   0.528  0.59756
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
FabricTypebroadcloth        -0.645005   0.363767  -1.773  0.07709 .
FabricTypechiffon           -0.587900   0.358124  -1.642  0.10158
FabricTypeCorduroy          -1.015422   0.497612  -2.041  0.04205 *
FabricTypedobby             -0.735158   0.535390  -1.373  0.17060
FabricTypeflannael          -1.105589   0.599357  -1.845  0.06595 .
```

.

More Conclusions based on Deviance Values :

The residual deviance is lower than the null deviance, which implies that using the independent variables does help in making better Predictions .

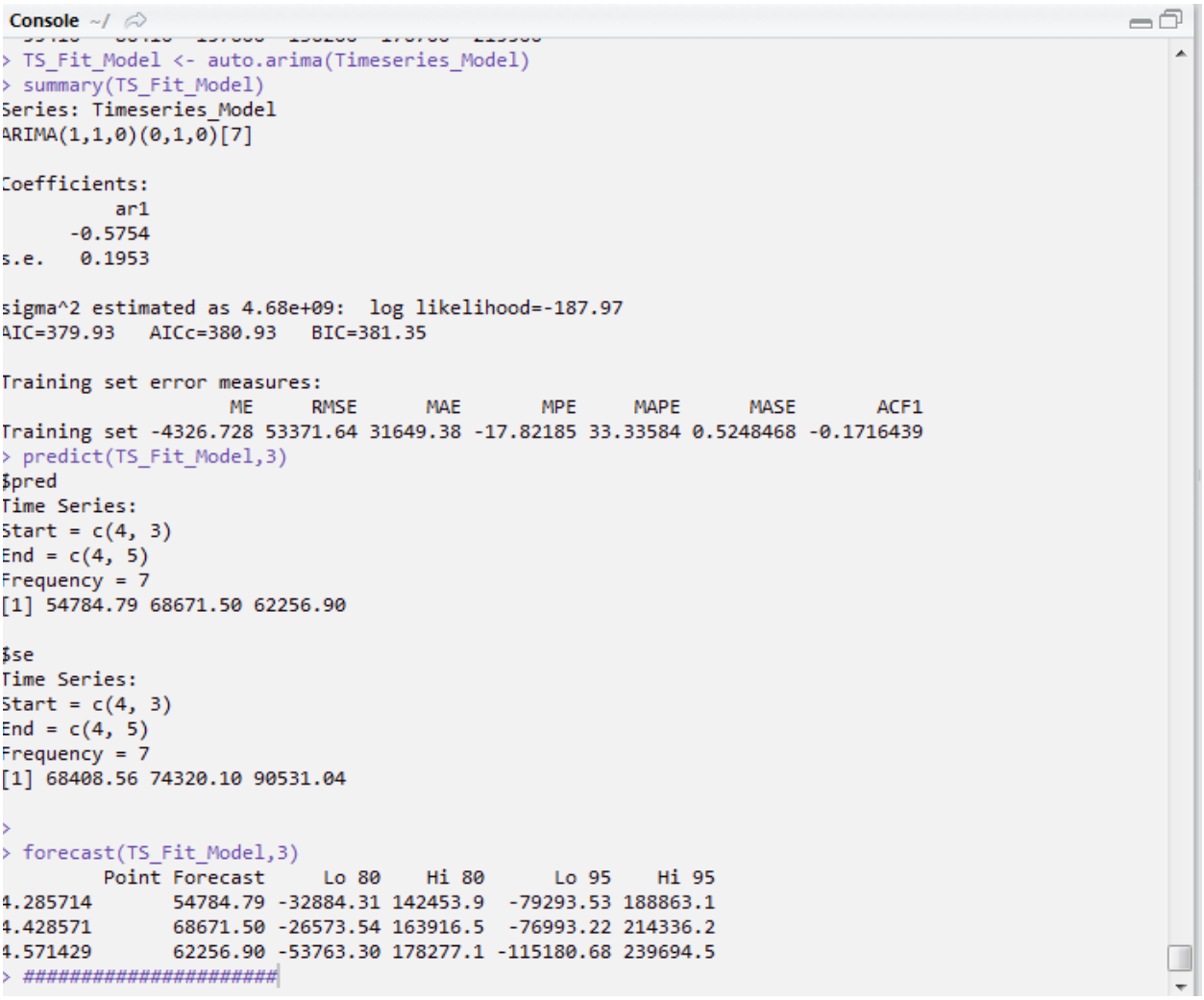
```
Pattern_Typecharacter      -0.229939   0.496882  -0.463  0.64383
Pattern_Typedot             0.202009   0.200745   1.006  0.31498
Pattern_Typefloral         -0.611082   0.370631  -1.649  0.10010
Pattern_Typegeometric      -0.320867   0.285274  -1.125  0.26147
Pattern_Typeleapord         0.813441   0.521981   1.558  0.12006
Pattern_Typeleopard        -0.196433   0.312895  -0.628  0.53055
Pattern_Typenone            0.445333   0.509353   0.874  0.38256
Pattern_Typenull            0.095522   0.141139   0.677  0.49899
Pattern_Typepatchwork       0.041482   0.151910   0.273  0.78496
Pattern_Typeplaid          -0.334197   0.319458  -1.046  0.29623
Pattern_Typeprint          -0.078888   0.141944  -0.556  0.57873
Pattern_Typesolid          -0.034288   0.136950  -0.250  0.80245
Pattern_Typesplice         -0.128351   0.528873  -0.243  0.80839
Pattern_Typestriped        -0.115098   0.191220  -0.602  0.54762
---
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
```

Problem Statement:

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.

Analysis Approach :

Auto.ARIMA Model has been used for Predicting the Trend .



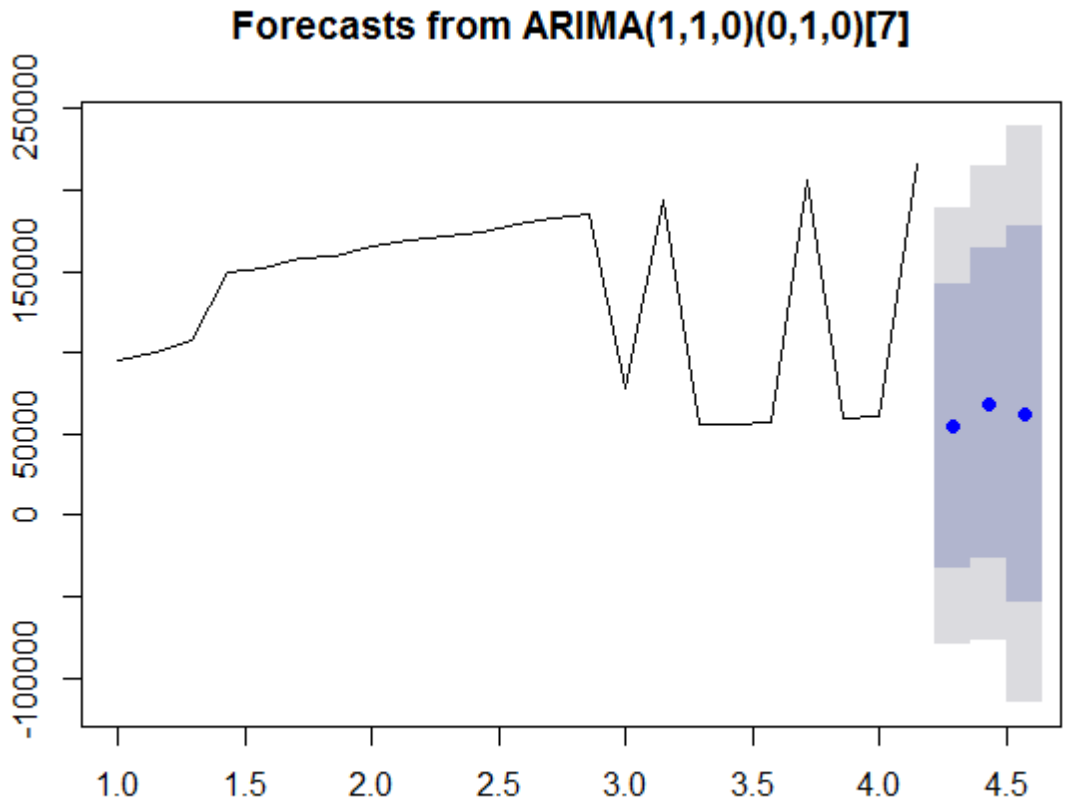
Conclusion:

Forecast for the Next Three Days :

The forecasted values are 54784, 68671, and 62256 respectively for the next three dates.

Plot Inference:

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, have huge differences for the predicted values (depicted by the light grey and dark grey areas in the plot).



Problem Statement:

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.

Analysis Approach:

An Analysis of Variance is performed to find the Impact of the various Attributes on the Sales.

To find the relative Impact , a LR Model is also created for further analysis.

Conclusions :

Season has an Impact on the Sales while Style and Material don't seem to have as much impact .

```
>
> summary(AOV_Style)
             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
> summary.aov(AOV_Season)
             Df    Sum Sq   Mean Sq F value   Pr(>F)
Season        7 3.642e+09 520333710    3.55 0.000983 ***
Residuals   490 7.182e+10 146571608
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
2 observations deleted due to missingness
> summary.aov(AOV_Material)
             Df    Sum Sq   Mean Sq F value Pr(>F)
Material     23 3.217e+09 139889390   0.919  0.573
Residuals   475 7.230e+10 152210222
1 observation deleted due to missingness
>
```

To Break down into further details and find a Coefficient for every attribute , Linear Regression Model has been used.

Conclusions from Attribute Level Analysis of Linear Regression Model:

From the result, we can see that spring season (coeff: 39044 and p-value:0.0014) and style "Sexy"(coeff:13426 and p-value:0.0172) directly impact the sales

```
>
> LR_Model_SSM <- lm(Total_Sales~Style + Season + Material, data = Dress_attributes)
> summary(LR_Model_SSM)

Call:
lm(formula = Total_Sales ~ Style + Season + Material, data = Dress_attributes)

Residuals:
   Min     1Q Median     3Q    Max
-18249  -4653  -2024   1169 137638

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          2457.77    7749.25   0.317   0.7513
StyleBrief           6182.91    3891.00   1.589   0.1127
StyleCasual          1776.23    2696.93   0.659   0.5105
Stylecute            4254.03    3172.68   1.341   0.1806
Stylefashion        -2468.93   12388.91  -0.199   0.8421
StyleFlare          -1963.95    9121.51  -0.215   0.8296
StyleNovelty          162.33    5050.64   0.032   0.9744
StyleOL             -2313.84   12450.13  -0.186   0.8526
Styleparty           -921.89    3133.98  -0.294   0.7688
Stylesexy           13426.57    5621.30   2.389   0.0173 *
StyleSexy            4813.39    2949.23   1.632   0.1034
Stylevintage         4857.87    3524.02   1.378   0.1687
Stylework            1533.81    3935.63   0.390   0.6969
SeasonAutumn        -4483.56    4609.69  -0.973   0.3313
Seasonspring        35465.54    8761.59   4.048 6.08e-05 ***
SeasonSpring          706.39    2013.99   0.351   0.7259
Seasonsummer         -452.43   12241.95  -0.037   0.9705
SeasonSummer         -737.27    1915.04  -0.385   0.7004
Seasonwinter        -3768.62    2455.42  -1.535   0.1255
SeasonWinter         -889.36    2079.89  -0.428   0.6691
Materialcashmere    -1425.42    9370.84  -0.152   0.8792
Materialchiffonfabric 10871.16  7476.60   1.454   0.1466
Materialcotton       1174.43    7155.94   0.164   0.8697
Materialknitting      -83.64   14103.74  -0.006   0.9953
Materiallace        -9588.35   14971.16  -0.640   0.5222
```

More Conclusions based on R-Squared & p-values :

However, a very low R-squared value suggests that they do not entirely affect the total sales, that is, there are other factors as well that affect the total sales.

The low p-value however indicates that there is a definite linear relationship between the variables season, style, material and the sales which have been used in this Model.

```
Residual standard error: 12080 on 454 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.1214,    Adjusted R-squared:  0.04016
F-statistic: 1.494 on 42 and 454 DF,  p-value: 0.02735

>
```

Problem Statement

To check if style is more influential than the price on Total Sales

Analysis Approach :

For this analysis, a new linear regression model with only the attributes style and price is created .

C**onclusions from Attribute Level Analysis:**

Except for the Low Price Category other Price Ranges don't have much impact on the Sales .However the cute, sexy, and vintage style dresses positively affect the sales.
The style '**Sexy'** has a positive coefficient of 12377, that has a huge impact on the Total Sales

```
> LR_Price_Style<- lm(Total_Sales ~ Style+Price,data=Dress_attributes)
> summary(LR_Price_Style)

Call:
lm(formula = Total_Sales ~ Style + Price, data = Dress_attributes)

Residuals:
   Min     1Q Median     3Q    Max
-12610  -4966  -2425    603 143277

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     2243.2     2583.5   0.868   0.3857
StyleBrief      6053.2     3813.4   1.587   0.1131
StyleCasual     2916.2     2622.6   1.112   0.2667
Stylecute       7551.4     3100.8   2.435   0.0152 *
Stylefashion   -1817.2    12447.4  -0.146   0.8840
StyleFlare      -575.2     8989.3  -0.064   0.9490
StyleNovelty     852.6     5015.6   0.170   0.8651
StyleOL         1252.0    12633.9   0.099   0.9211
Styleparty      3591.0     3320.6   1.081   0.2801
Stylesexy      12377.1     5240.1   2.362   0.0186 *
StyleSexy       5177.0     2899.9   1.785   0.0749 .
Stylevintage    6121.8     3532.9   1.733   0.0838 .
Stylework       3949.5     3910.3   1.010   0.3130
Pricehigh      -4465.6     3410.3  -1.309   0.1910
PriceHigh        245.2     5115.9   0.048   0.9618
Pricelow       -1725.7     2007.2  -0.860   0.3903
PriceLow        2855.8     1367.7   2.088   0.0373 *
PriceMedium    -3066.2     2411.9  -1.271   0.2043
Pricevery-high -4572.5     3241.2  -1.411   0.1590
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

More Conclusions based on R-Squared & p-values :

However, the R-squared value is very low, specifying that these variables do not completely explain the significant changes in sales.

Style and price cannot completely be used in predicting the total sales of dresses.

```
Residual standard error: 12180 on 479 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.0595,    Adjusted R-squared:  0.02416
F-statistic: 1.684 on 18 and 479 DF,  p-value: 0.03872
```

Problem Statement

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.

Analysis Approach:

For this we reconstruct a Linear Regression Model with all the Dress Attributes incorporated to identify the Attributes that have the most impact on the Sales.

C**onclusions from Attribute Level Analysis:**

- **Style :**Sexy style makes a positive impact
- **Rating :** Rating has a very high significance on the Sales.
- **Size**:Large size clothes are sold more
- **Season** :Spring season clothes make a positive impact on sales
- **Neck Line:**Ruffled neckline clothes have a significant positive impact
- **Sleeve Length :** Significant, however, it affects the sale negatively .
- **Pattern**: Patchwork pattern also have a negative Impact on the Sales.

```
> All_Feature_LM_Model <- lm(Total_Sales~Style + Price + Rating + Size + Recommendation+
+                              Season + NeckLine +SleeveLength + waiseline + Material +
+                              FabricType + Decoration + Pattern_Type, data = Dress_attributes)
> summary(All_Feature_LM_Model)

Call:
lm(formula = Total_Sales ~ Style + Price + Rating + Size + Recommendation +
    Season + NeckLine + SleeveLength + waiseline + Material +
    FabricType + Decoration + Pattern_Type, data = Dress_attributes)

Residuals:
   Min     1Q Median     3Q    Max
-24359  -3809      0   1955  59098

Coefficients: (2 not defined because of singularities)
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               22097.48   21746.05   1.016  0.31027
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                   12134.03   18139.22   0.669  0.50398
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
Pricehigh                   541.77    3557.84   0.152  0.87906
PriceHigh                 -1346.49    4487.50  -0.300  0.76432
Pricelow                  -2302.65    1933.13  -1.191  0.23441
PriceLow                   1442.10    1294.82   1.114  0.26616
PriceMedium               -3954.45    2288.01  -1.728  0.08482 .
Pricevery-high             -122.77    4237.16  -0.029  0.97690
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	
Recommendation	1371.84	1091.22	1.257	0.20954	
SeasonAutumn	-5520.31	4050.37	-1.363	0.17380	
Seasonspring	32233.96	7227.21	4.460	1.11e-05	***
SeasonSpring	74.98	1763.53	0.043	0.96611	
Seasonsummer	2650.20	10177.38	0.260	0.79471	
SeasonSummer	-1103.10	1690.94	-0.652	0.51460	
Seasonwinter	-5987.56	2216.82	-2.701	0.00725	**
SeasonWinter	-806.20	1831.78	-0.440	0.66013	
NeckLineboat-neck	7132.59	14738.00	0.484	0.62872	
NeckLinebowneck	7816.22	15067.08	0.519	0.60426	
NeckLinehalter	9466.67	19928.33	0.475	0.63506	
NeckLinemandarin-collor	NA	NA	NA	NA	
NeckLineNULL	1568.53	16727.22	0.094	0.92535	
NeckLineo-neck	8679.29	14473.68	0.600	0.54913	
NeckLineopen	15900.19	20992.72	0.757	0.44932	
NeckLinepeterpan-collor	6438.98	15071.35	0.427	0.66948	
NeckLineruffled	147692.62	17706.44	8.341	1.78e-15	***
NeckLineScoop	4870.71	14428.80	0.338	0.73589	
NeckLineslash-neck	5275.26	14728.57	0.358	0.72044	
NeckLinesqare-collor	2430.23	15455.55	0.157	0.87515	
NeckLinesweetheart	10500.94	20185.19	0.520	0.60324	
NeckLineSweetheart	9377.06	15253.54	0.615	0.53913	
NeckLineturndowncollor	11244.75	14840.99	0.758	0.44916	
NeckLinev-neck	7985.51	14542.77	0.549	0.58329	
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	-29151.43	19625.20	-1.485	0.13835	
SleeveLengthhalfsleeve	-29590.10	10755.18	-2.751	0.00625	**
SleeveLengthNULL	-23154.81	13742.64	-1.685	0.09291	.
SleeveLengthPetal	12360.53	19522.81	0.633	0.52707	
SleeveLengthshort	-32132.46	10530.82	-3.051	0.00246	**
SleeveLengthsleeevless	-41724.65	12768.75	-3.268	0.00119	**

```
Decorationflowers           -4213.70    7633.72  -0.552  0.58132
Decorationhollowout         -1872.11    3720.51  -0.503  0.61515
Decorationlace                -80.76    3057.83  -0.026  0.97895
Decorationnone               -387.81    8093.14  -0.048  0.96181
Decorationnull              -1103.49    2914.50  -0.379  0.70520
Decorationpearls               33.49   11220.25   0.003  0.99762
Decorationplain             -3996.21   10192.80  -0.392  0.69525
Decorationpockets           -2545.48    5495.97  -0.463  0.64355
Decorationrivet              -629.08    6819.20  -0.092  0.92655
Decorationruched            -5684.28    7296.23  -0.779  0.43647
Decorationruffles            3664.27    3868.97   0.947  0.34425
Decorationsashes             -306.35    3247.60  -0.094  0.92490
Decorationsequined           -573.92    4137.67  -0.139  0.88976
Decorationtassel            -3877.97   10616.91  -0.365  0.71514
DecorationTiered            21074.89   11871.32   1.775  0.07673 .
Pattern_Typecharacter       -5137.55   10088.72  -0.509  0.61091
Pattern_Typedot             -5559.76    4080.65  -1.362  0.17394
Pattern_Typefloral           -753.02    7552.47  -0.100  0.92064
Pattern_Typegeometric        4314.14    5801.00   0.744  0.45757
Pattern_Typeleapord        -16571.45   10632.18  -1.559  0.12000
Pattern_Typeleopard         -5547.15    6354.70  -0.873  0.38331
Pattern_Typenone            -7942.38   10350.15  -0.767  0.44339
Pattern_Typenull            -6770.48    2866.70  -2.362  0.01874 *
Pattern_Typepatchwork       -7022.03    3083.76  -2.277  0.02339 *
Pattern_Typeplaid           -8533.42    6494.53  -1.314  0.18974
Pattern_Typeprint           -4799.14    2882.44  -1.665  0.09683 .
Pattern_Typesolid           -4096.34    2780.04  -1.473  0.14153
Pattern_Typesplice          -7458.23   10735.87  -0.695  0.48771
Pattern_Typestriped         -6236.87    3883.37  -1.606  0.10918
---
```

More Conclusions based on R-Squared & p-values :

A good R-Squared Value of 58% and a low p-value ~ 0 indicates that it is a fairly good model in predicting the Total Sales .

```
Residual standard error: 9586 on 345 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.5799,    Adjusted R-squared:  0.3972
F-statistic: 3.174 on 150 and 345 DF,  p-value: < 2.2e-16

> ##############################################################
```

Problem Statement

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

Analysis :

As seen from the LR Model , Rating has an Impact .
To find the Correlation as a metric , we can use the Correlation Test

```
> cor.test(Dress_attributes$Total_Sales, Dress_attributes$Rating)

        Pearson's product-moment correlation

data:  Dress_attributes$Total_Sales and Dress_attributes$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
```

Conclusion :

The correlation value is 0.2, which shows a very weak positive association. A higher rating correlates with higher sales. Thus, the rating process has to be regularized.