Project Name : Analysis of Sales Report of a Clothes Manufacturing Outlet.

Course: Data Science with R

Created by: Shahroo Akhtar

**Overview**: A high end Fashion outlet has data for two months. They want to study the current market trends and understand the fashion industry.

**Objectives**: The objectives of the analysis are as follows:

* 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.
* 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.
* 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.
* 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.
* 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 Tasks:** The following are performed:

* Since the output for the recommendations is in binary form, I have used logistic regression Model. Dataset for the product attributes have been provided. After data exploration and feature selection, model was applied.

**Result:** Accuracy test was performed using confusion matrix. Predicted value came out higher than the actual value proving the efficiency of the model.

* I deduced average per day sale from total sale and predicted sales for next 3 alternative days.

**Result**: Sale was predicted.

* For this task I had a choice of choosing between two models, Multiple Linear Regression or Applying Apriori Algorithim. I used the later after creating the subset of features from the original dataset. After data exploration, model was applied.
* **Result:** Outcome showed that Style of the dress is definitely more influential than the price.
* Summary of the dataset from step 3 was analyzed and inferences were made for the factors affecting the product.

**Result**: Casual style, summer season, average pricing and cotton material have great selling rate.

* In order to regularize the rating process I decided to choose linear regression model to find the relationship between the dependent variable that is, Total sales and independent variables, that is Rating variable.

**Result:** The outcome showed positive relationship between independent and dependent variable. The higher the rating, the higher the chance of the product getting sold soon.

It was a great learning experience especially since I don’t have a programming background. I was able to learn a lot.