12/20/2022

BY

JEENESH R JAIN

NANDAN S

HITESH S

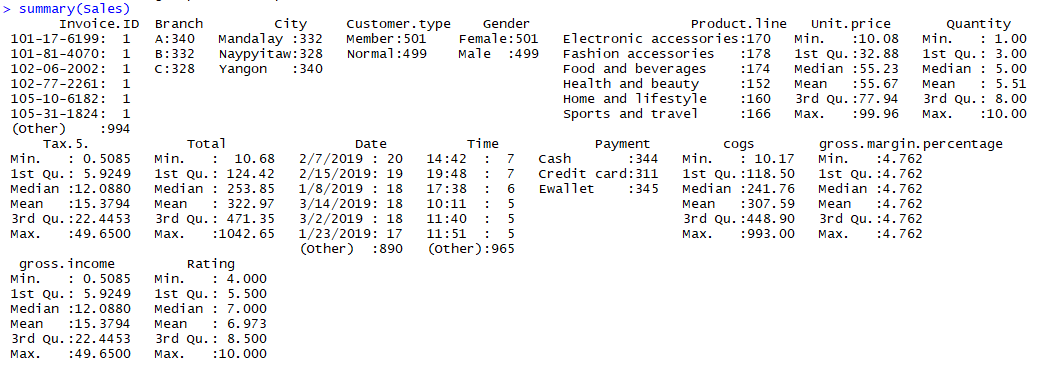
NEELESH

M-Smart Store Analysis with R

Final Project Report

# Dataset:

The Dataset is of a retail store, this data set contains 17 columns and 1000 rows of sales information which was recorded for a certain period of time. The dataset has the mix of columns. Glimpse of the dataset is the below screenshot for the reference.



# Libraries**:**

For calling any functions, Libraries and Packages play a very important role in R. Hence all the required libraries are imported in this Analysis. Tidyverse, Cowplot, dplyr and corrplot..etc are some important libraries amongst all. These are just quoted for the reference.

# Pre-Processing and Data Transformation –

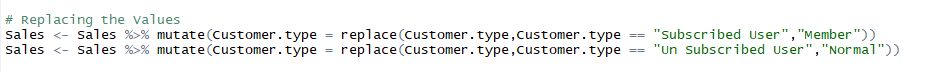
Once the Libraries are imported, it is now the Data Import and Data Cleaning operation that needs to be done. Below is the Source of the file that is being called from.



For us to initially look at the data in R, Below function is used as this defines and explains clearly about each attributes and Classes.



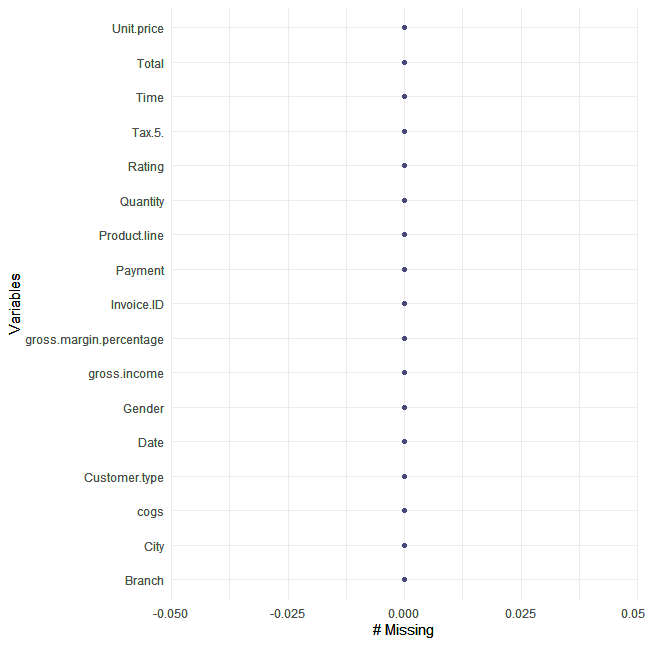
**Data Cleaning:** In this Module, Missing values are replaced and some data changes in the column has been performed.



**Cleaning Data –**

Data Cleaning is one of the crucial steps in the analysis. Here the primary objective of Data cleaning is to make the model to get the most out of the dataset. So All the assumptions are listed in this section before filling any values in the dataset. Below is the dataset values after cleaning the dataset. It is clearly visible from the below image that there are no missing/Blank values.

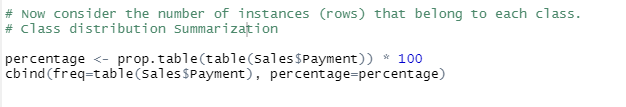
**Result After Cleaning Missing Values:**



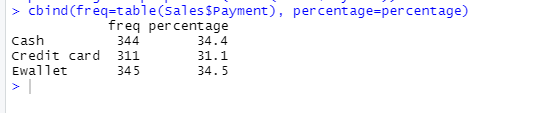
# Data Analysis:

Here we perform some of the operations to see the current visualizations and understand the metrics. As R is known for Statistical Analysis, Some analysis were performed on this existing data to check and draw some insights from this business.

**1. Payments Frequency:**

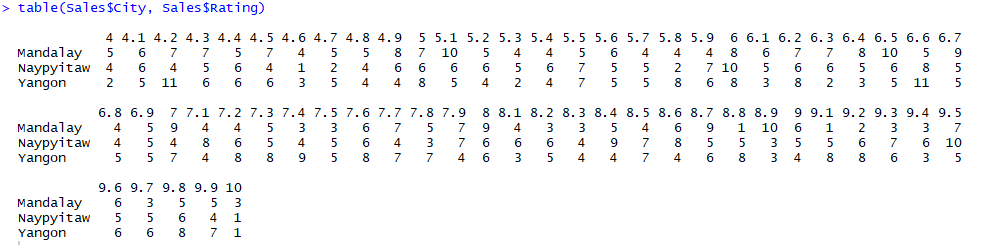


**Result:**



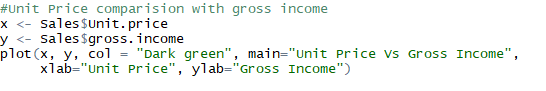
**2. Summarizing Data:**When the Business tries to Analyze the city rating, this could be the best approach. As the below insight gives a clear image on the ratings given. Even though this is not that great visual to capture all metrics in one view, Rounding off the values would help business to understand these numbers better.



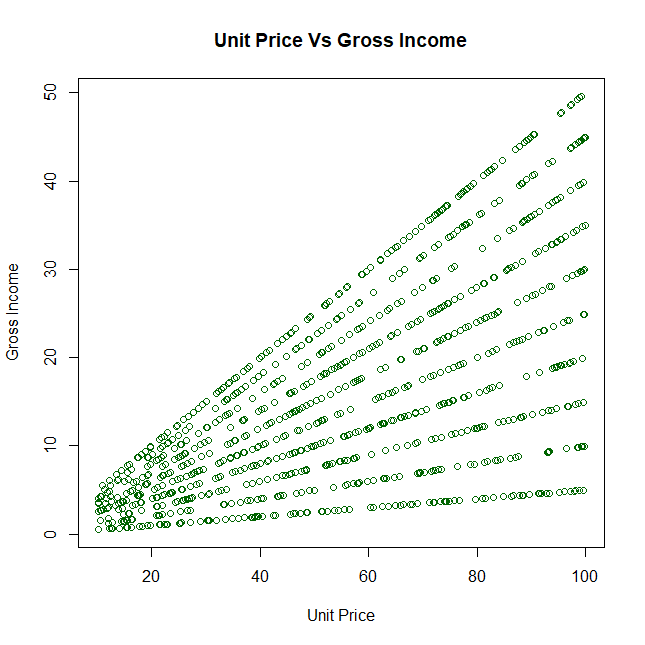
R**esult:**

**Unit Price Comparison with Gross Income:**

Here we analyze the unit price and its contribution in the gross income for the organization. Below are the screen shots for your reference.



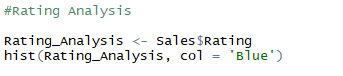
**Result: Unit Price vs Gross Income:**



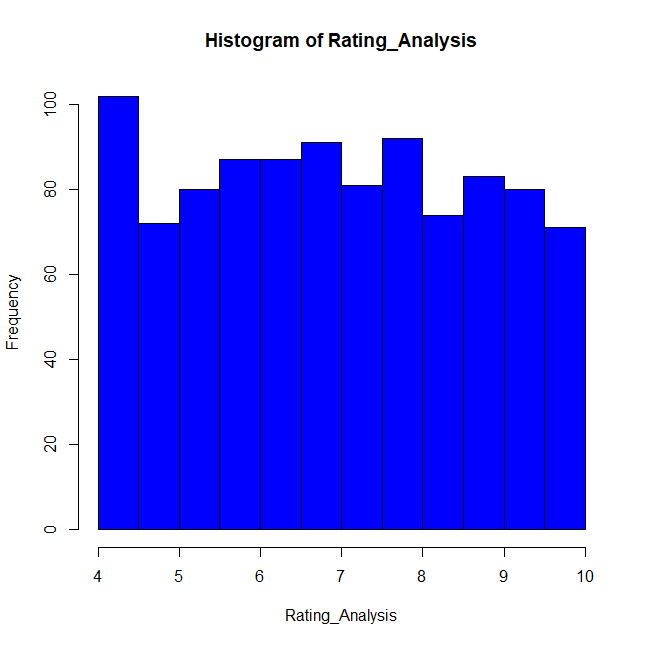
**Visualization:**

It is time to explore some visuals and see the results to understand the metrics better. Below are some analysis, which shows Histograms, Graphs..Etc.

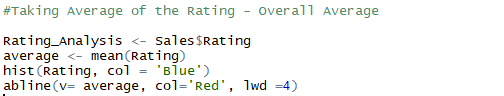
**Rating Analysis:**

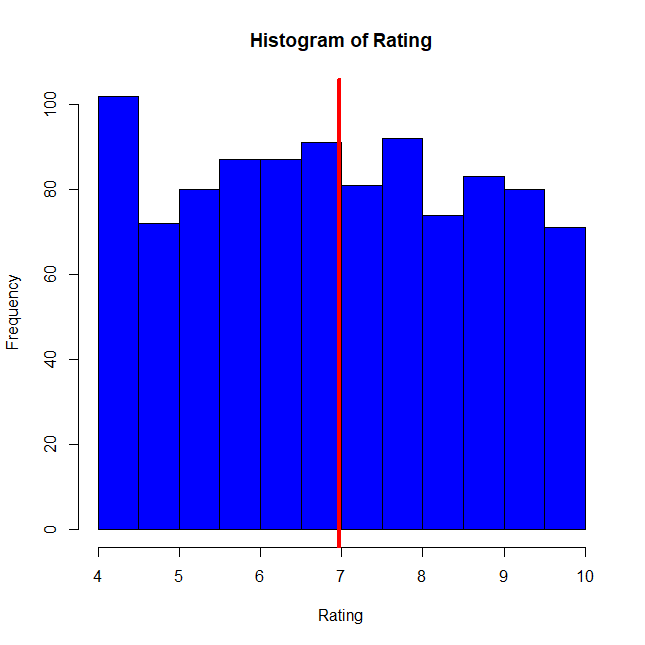
****

**Result:**

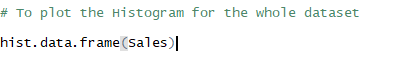


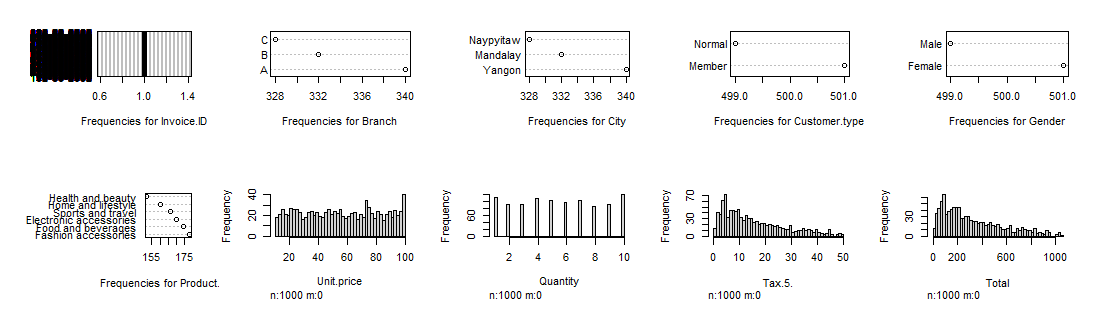
**Mean of Average:** Overall if we need an average rating for the data that we have analyzed so far, then this is definitely one of the methods. Overall Average rating has come up to 6.9 and close to 7.



**Result:**

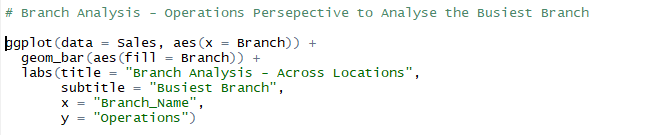
**Summarized ViOverall Dataset:**

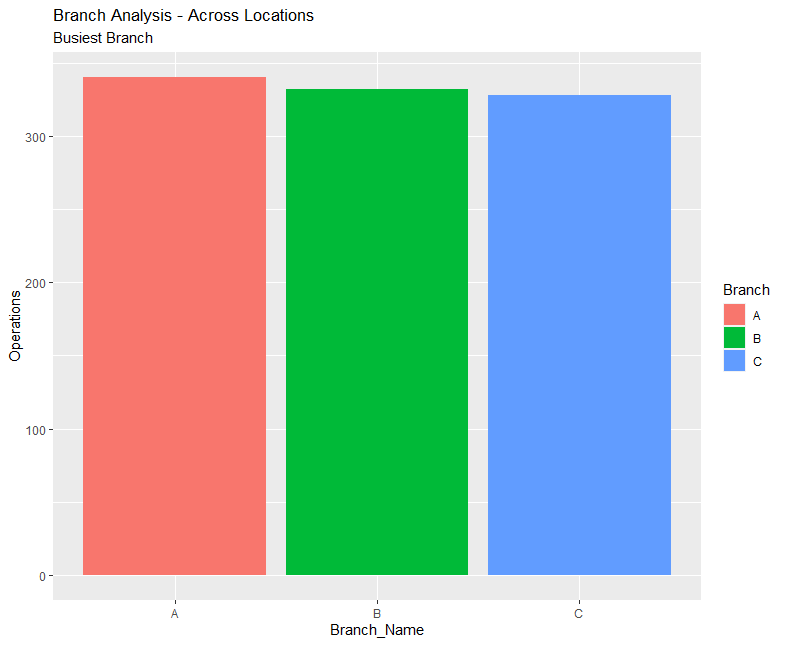


**Result:**

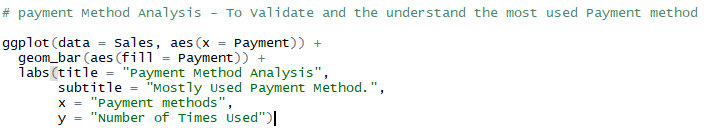
**Branch Analysis :**

There is competition everywhere, for business to review the performance, it is very important to have a insight of all the individual performances, which shall later be also considered for the growth calculations and revenue contribution for organization. To find out the busiest branch according to the operations.

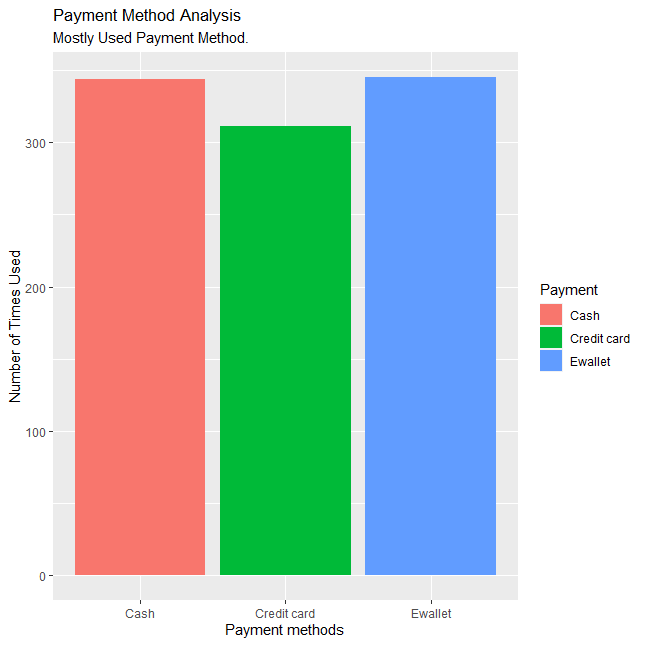




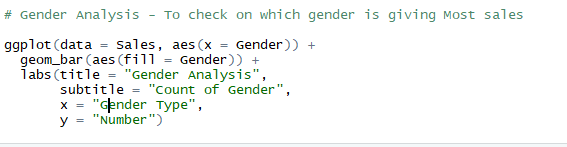
**Payment Method Analysis:** Due to Digitalization, it is observed that E wallet payments has increased.

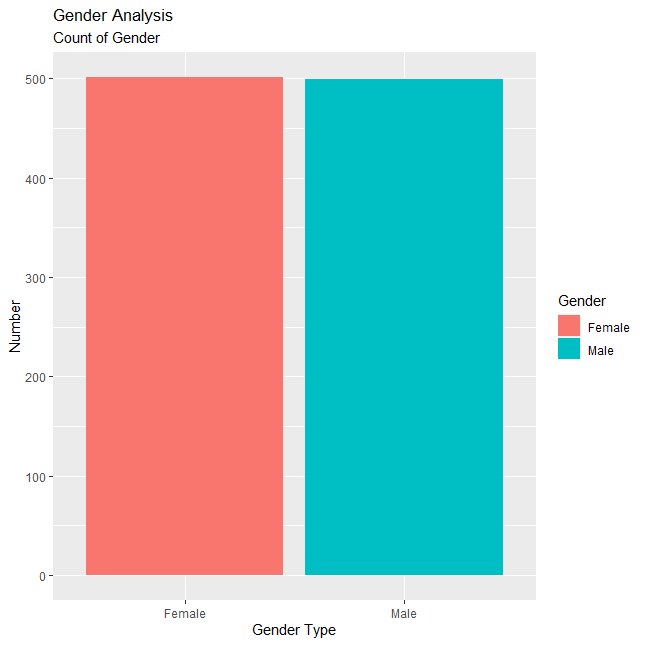


**Result:**

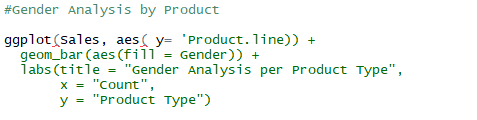


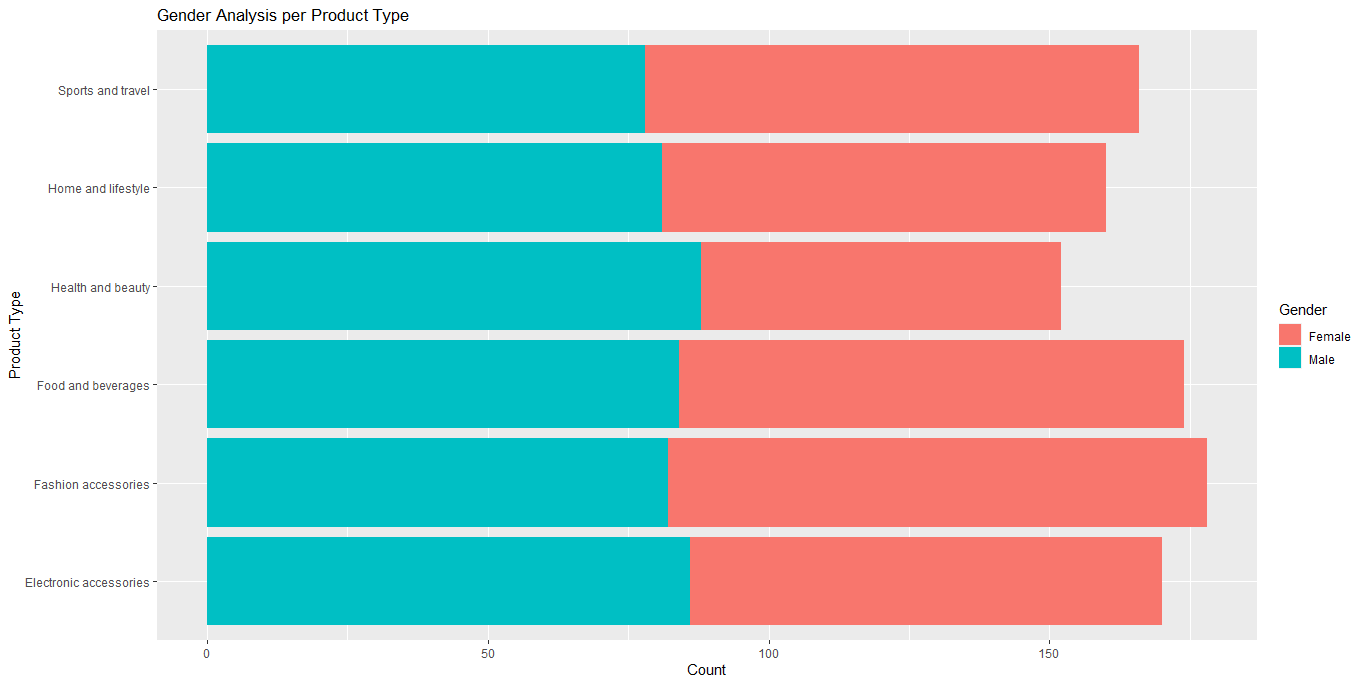
**Gender Analysis:**





**Product By Gender:**

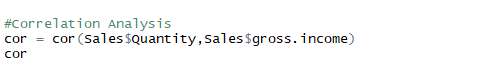


**Result:**

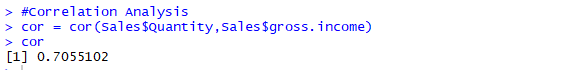
**Correlation Matrix:**

A correlation matrix is essentially a table that displays the correlation coefficients for various variables. The matrix displays the relationship between all possible pairings of values in a table. It is an effective tool for summarizing a huge dataset as well as identifying and visualizing trends in the data.

**Correlation Formulae:**



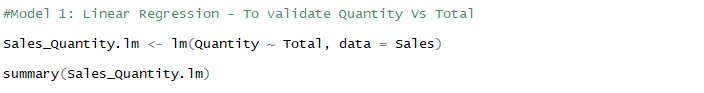
**Result of Correlation Matrix between these two parameters:**



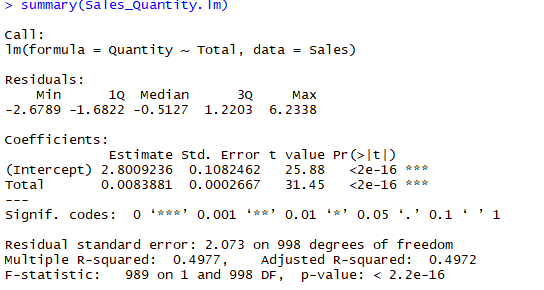
# Regression Models:

1. Simple Regression Model:

Regression Models are applied on the continuous solutions for better solutions. This analysis benefits such analysis and gives good results. Please find the algorithm applied below

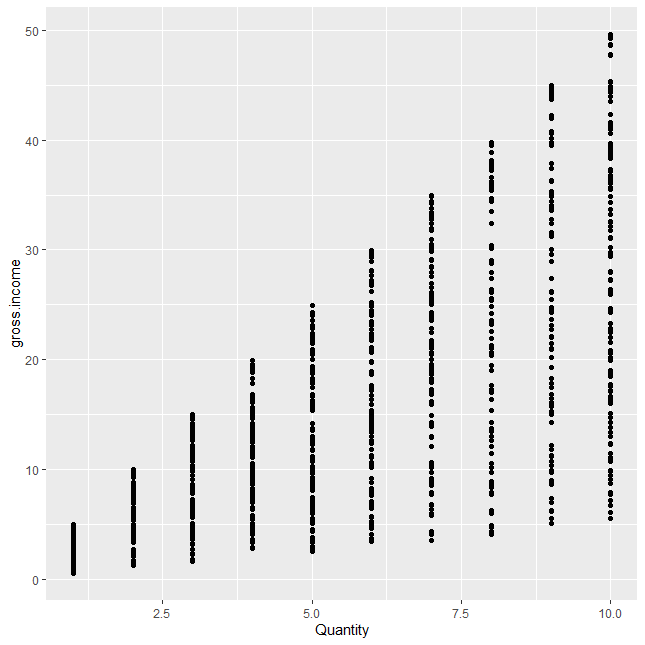


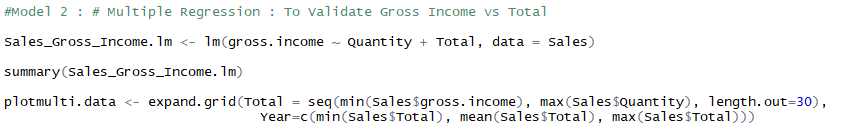
**Results:**



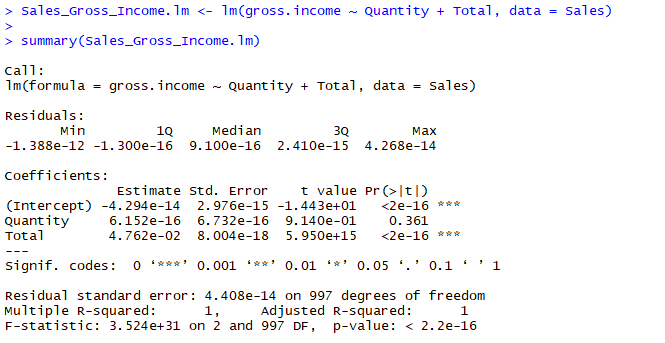
The considerable thing to be observed here is the p value is <2.2e-16) which says whether the model fits the data well or not

The value of y-intercept is 2.8009236 and the residual standard error is 2.073. T value is 31.45



**2. Multiple Regression Model:**

**Result:**



The considerable thing to be observed here is the p value is <2.2e-16) which says whether the model fits the data well.

The value of y-intercept is and the residual standard error is 8.004e-18. T value is 5.950

**Conclusion and Future Work:**

This Regression Models gave better results and would definitely need some more data for the model to predict the data. Retail Analysis is needed for all the businesses to understand the insights and take decisions. Technology is helping businesses to make the decisions by proving such a beautiful analysis.

This project can be extended by continuing this analysis on training the datasets. We have started working on Arima and Naïve Forecasting Models.