KAIBURR TASK

Data Science Task:

DOMAIN: Telecommunications

OBJECTIVE: To use supervised learning techniques such as logistic regression in order to identify customers along with some intuitive findings from the data

The Tool is used is R

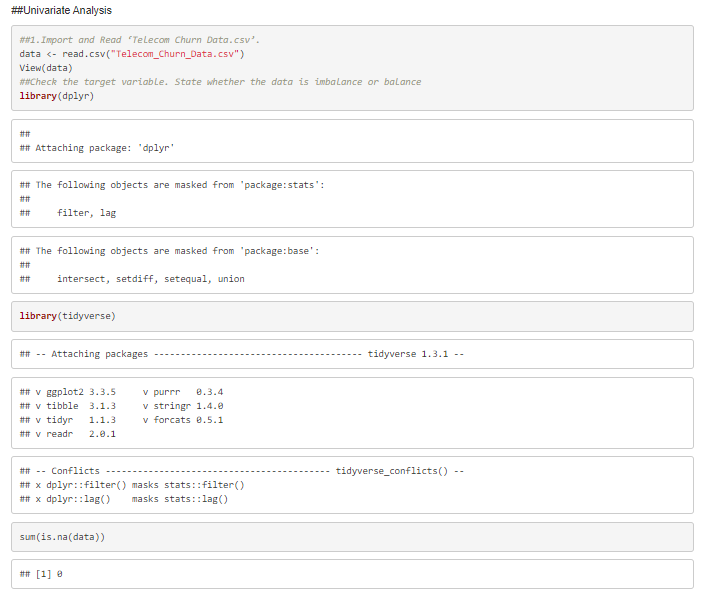
TASK GIVEN:

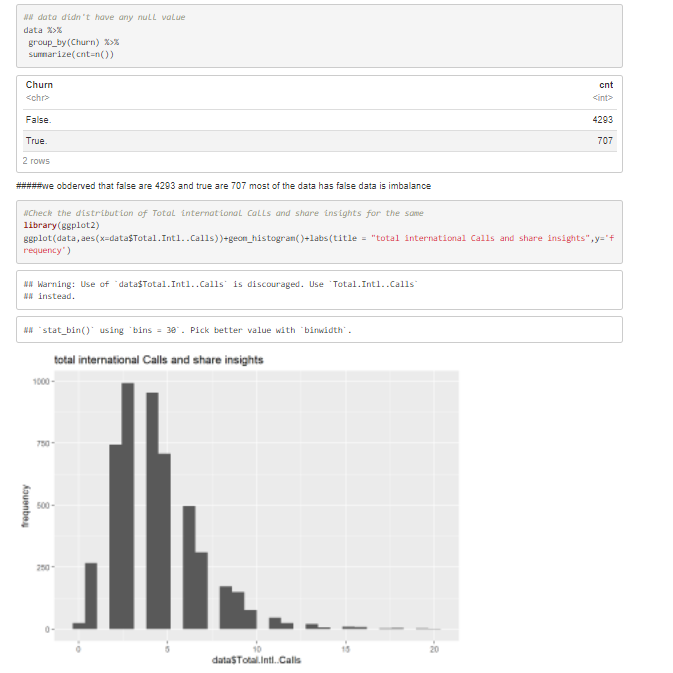
**Univariate Analysis:**

A. Import and Read ‘Telecom Churn Data.csv’.

B. Check the target variable. State whether the data is imbalance or balance.

C. Check the distribution of Total international Calls and share insights for the same.





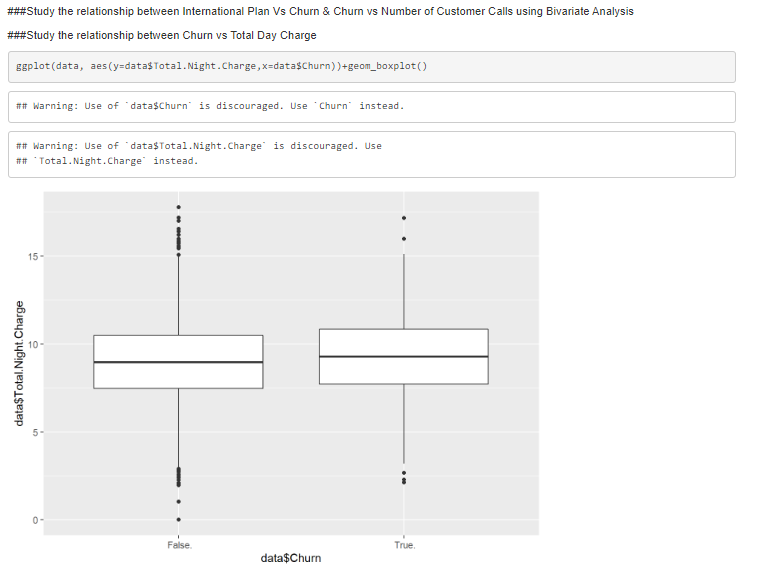
From the above Univariate Analysis, the following false data is in-balance and the above we can observe total international calls and share insights histogram plot

Bivariate Analysis:

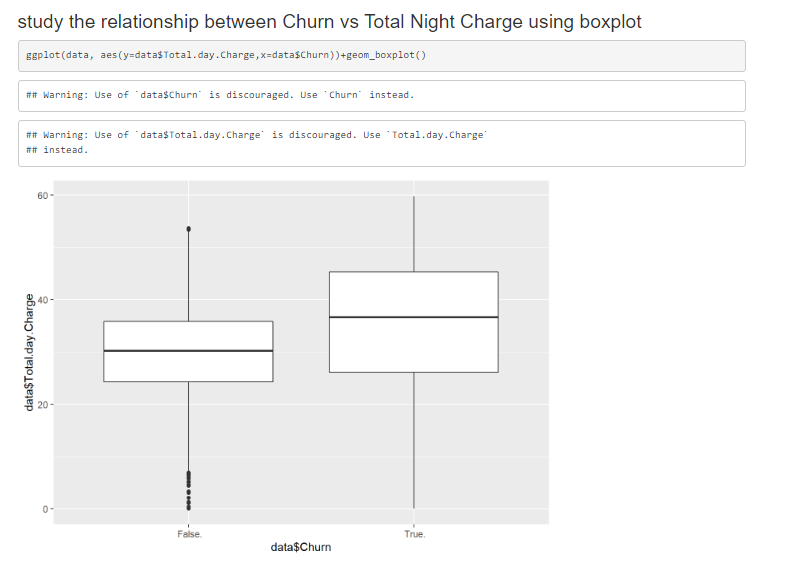
A. Study the relationship between International Plan Vs Churn & Churn vs Number of Customer Calls using Bivariate Analysis and state the inference clearly.

B. Study the relationship between Churn vs Total Day Charge & Churn vs Total Night Charge using boxplot. State the inference Clearly.

C. Split the dataset into 80:20 (i.e. 80% train and 20% test) .



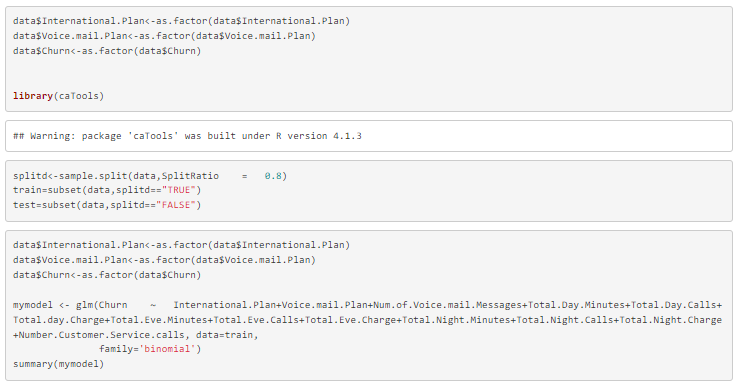
As the data is highly in balanced and both the given values are categorical values in Churn vs Number of Customer Calls thus creating any type of correlation or any other regression graph would be difficult or be not accurate.



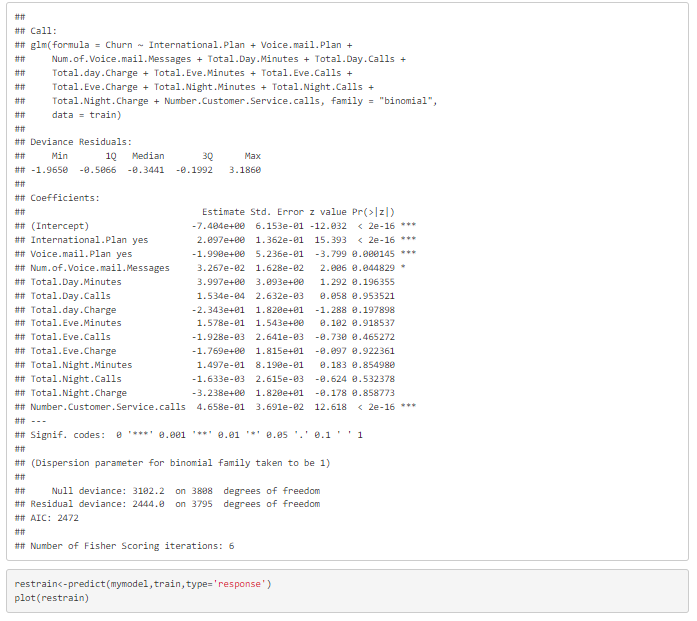
Got about the relationship between churn vs total night charges and we can observer the median, mean, maximum, minimum,3rd quartile,1st quartile lies in the above graph and thus it also has some outliers.

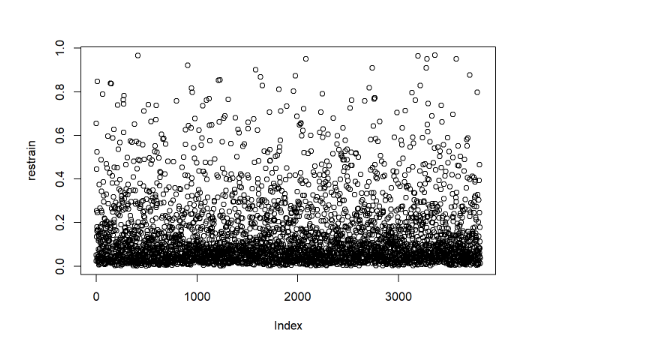
Model Training:

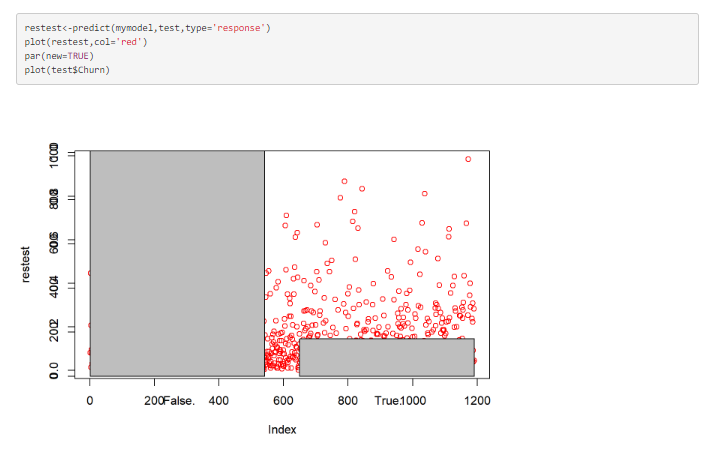
1. For further analysis train a logistic regression model and Naive Bayes model.
2. Evaluate the model using Accuracy, ROC-AUC curve(optional)

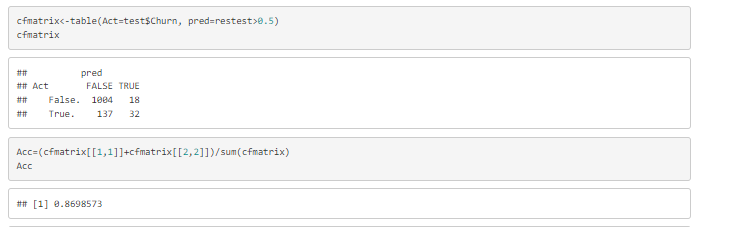


Splitting the training and testing to 80:20 ratio

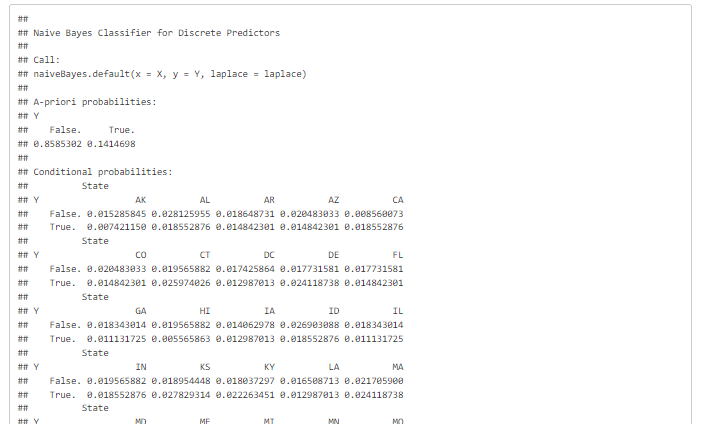




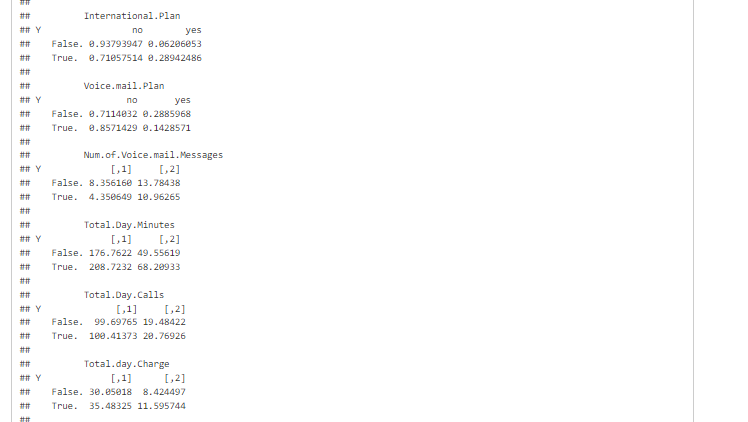


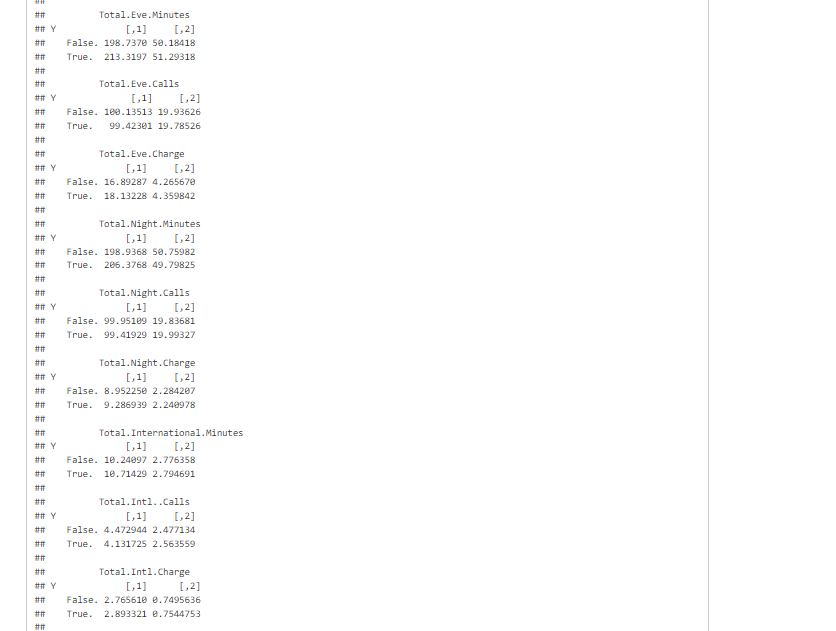


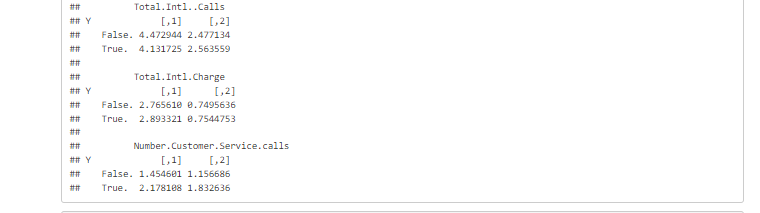




I have cropped the above result of naïve bayes as its very long output ..here I’m attaching the last end result of the output.







Hence, we can check the accuracy through the naïve bayes model proposed here for every attribute