# DAB 303: Marketing Analytics

# INDIVIDUAL Project

# Topic: Churn prediction analysis

# Name: Gurmanjit Singh Sahni

# Student ID: 0734775

# Dataset #: Project7.Csv

1-customers are coming to purchase the product every week means they are regular(not churn)

0-customers are not coming to purchase the product are not regular,either they are not coming or they are coming on alternate weeks.(churn)

Q.What is Churn Analysis?

Ans. It is the percentage of customers that stopped using the company's product or service during a certain time frame and process of how company can get back their customers by improving their services.

Q2.Import Data

**Ans.** Project7.csv dataset was imported in the Python File.

* All necessary Libraries was also imported.

Q3. Data Overview

Ans: An Overview of the dataset was completed in the Python File.

* An overview of the data shows 20,000 Rows and 37 Columns.

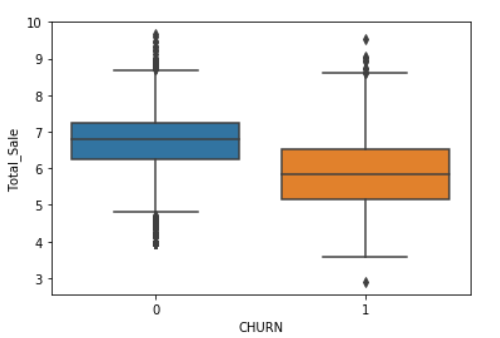
Q4. Data Cleaning

Answer: Initial Data Cleaning was completed in the Python File.

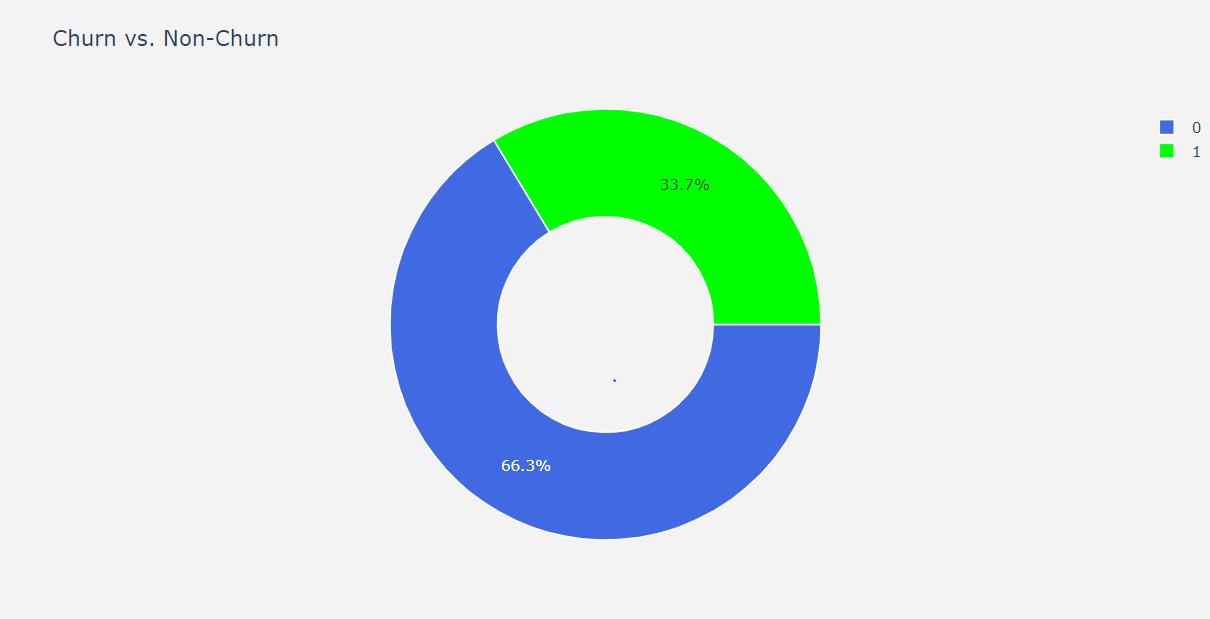
* The ‘Unnamed:0’ column was dropped as this wasn’t informative.

**Q5.Data exploration and analysis**

**Ans.BoxPlot**



Boxplot has two variables churn and Total Sale. As Total Sale increases it effects the churn rate (0), which means there are a greater number of people who are not regular visitors. Also median for the people who are not regular visitors is higher than the people who are regular visitors to the company. There are some outliers seen in boxplot for the people who are having churn rate ‘0’ which means that the total sales for the company is high irrespective of customers not visiting company regularly.

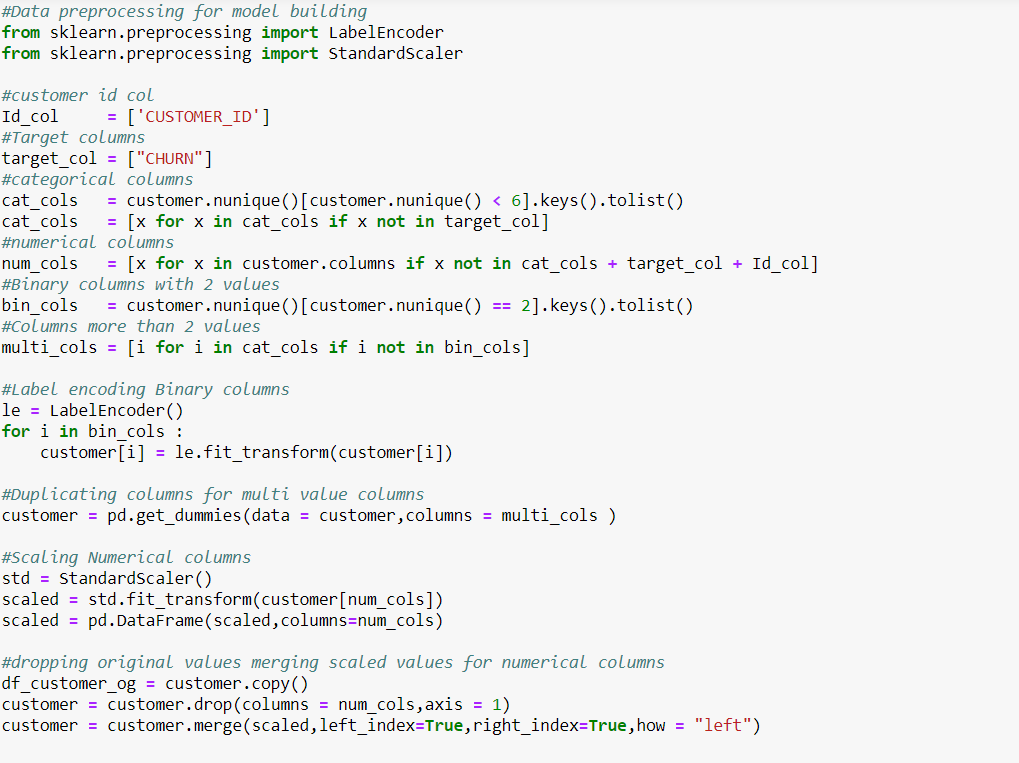


Doughnut Chart

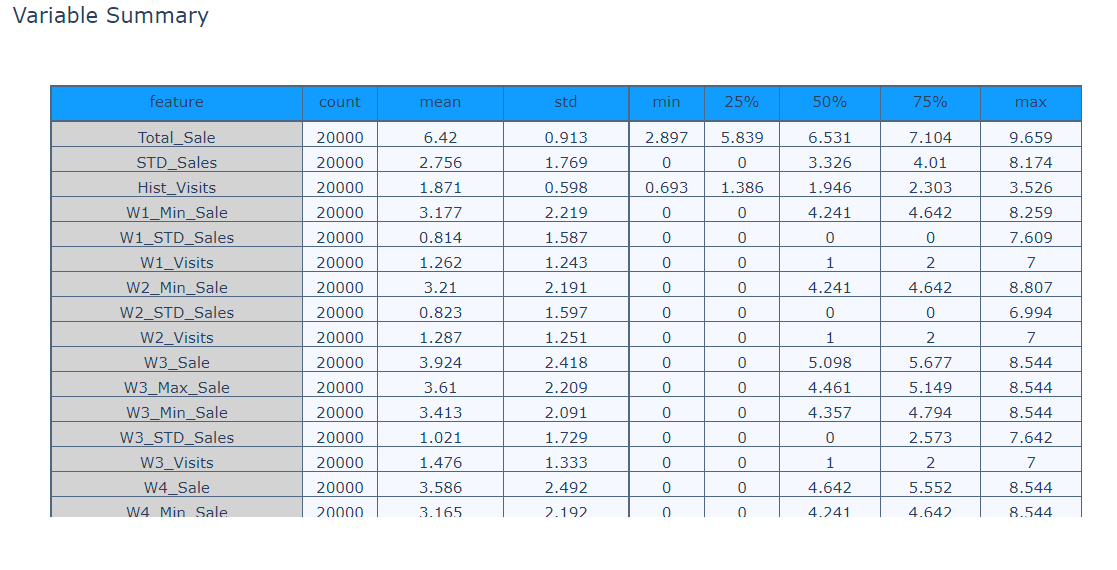
This chart explains about the churn customers and non churn customers.66.3% people were not regular in their visits to company that is 13,269 and 33.7 people were regular customers for the company that is 6731.

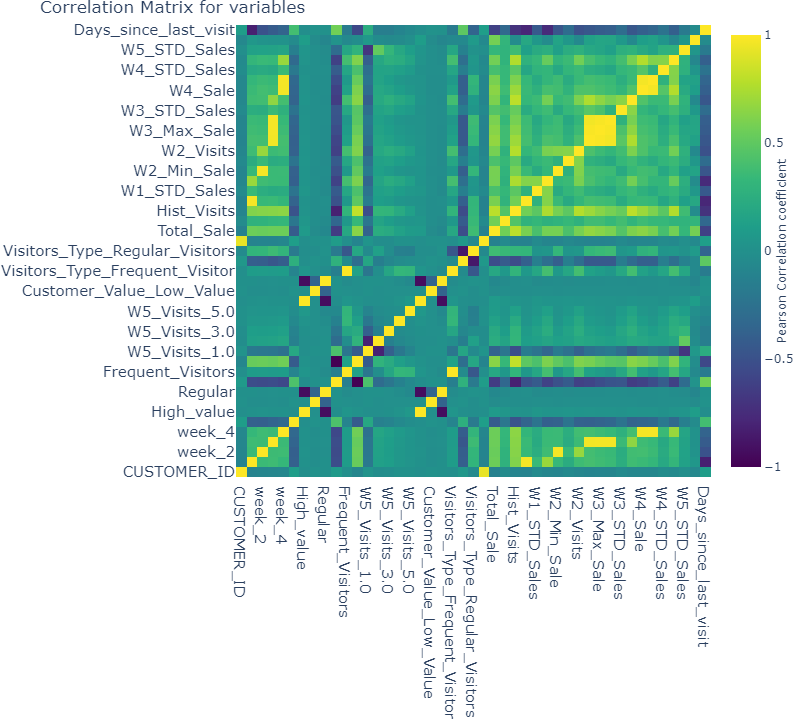
QUESTION 6: Data Preprocessing for Model Building

Ans. In data preprocessing we did scaling of data so that data becomes normalize and give better prediction.

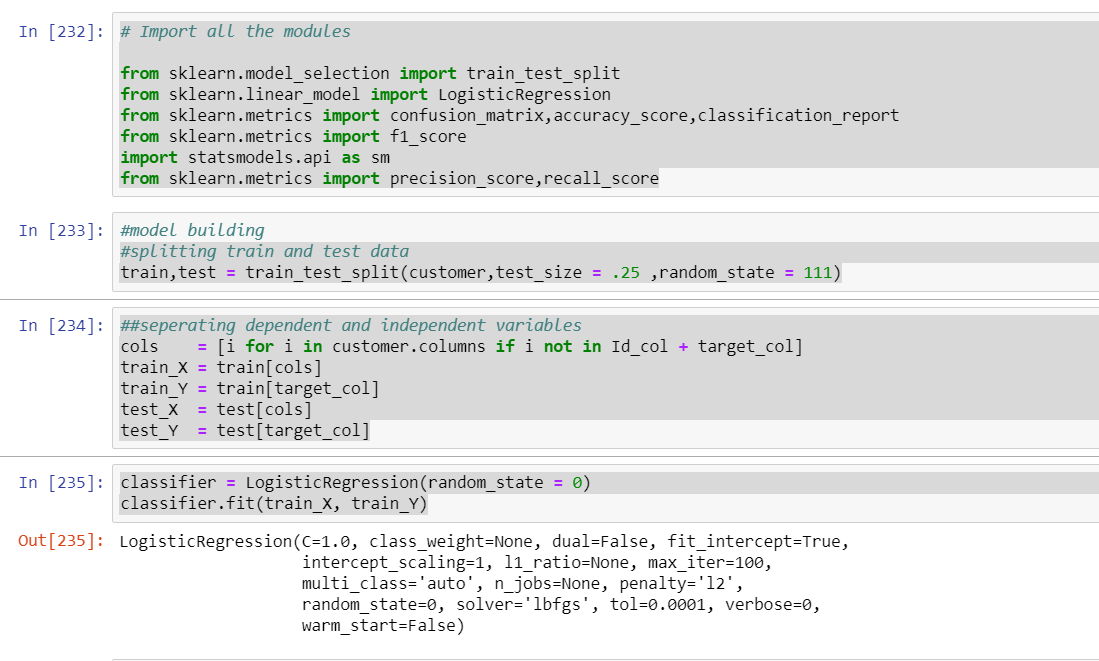


Que7: Variable Summary

**Answer:** 

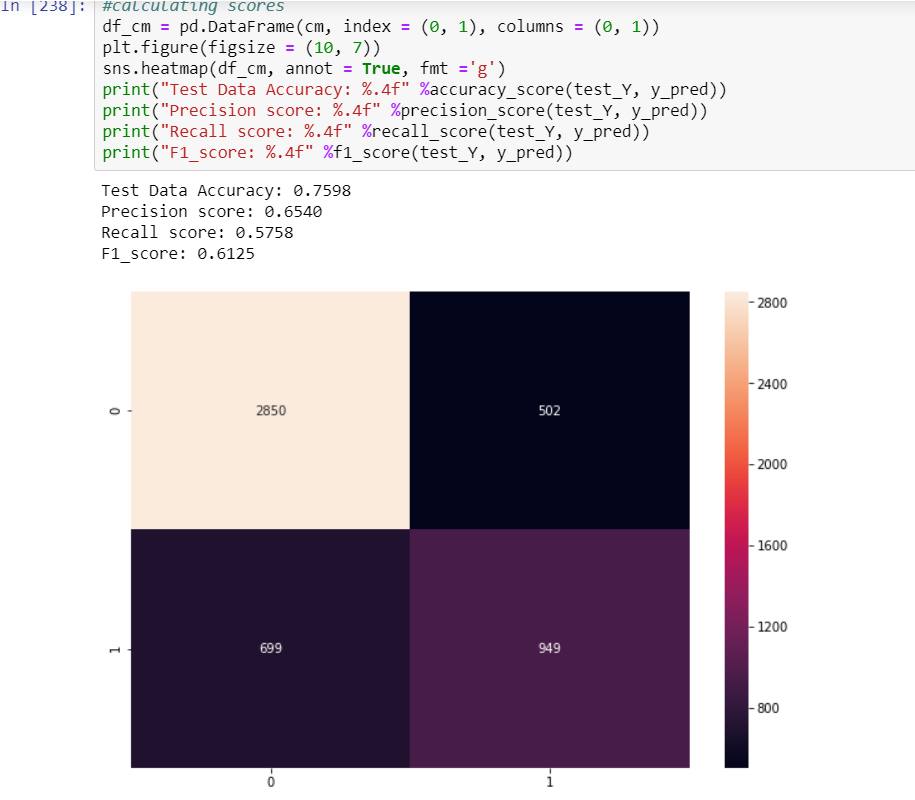


Que8: Model Building



Logistic regression has an accuracy of 75.98

Que9.Confusion Matrix



Test data accuracy means 76% of the data that we predicted is correct from the dataset

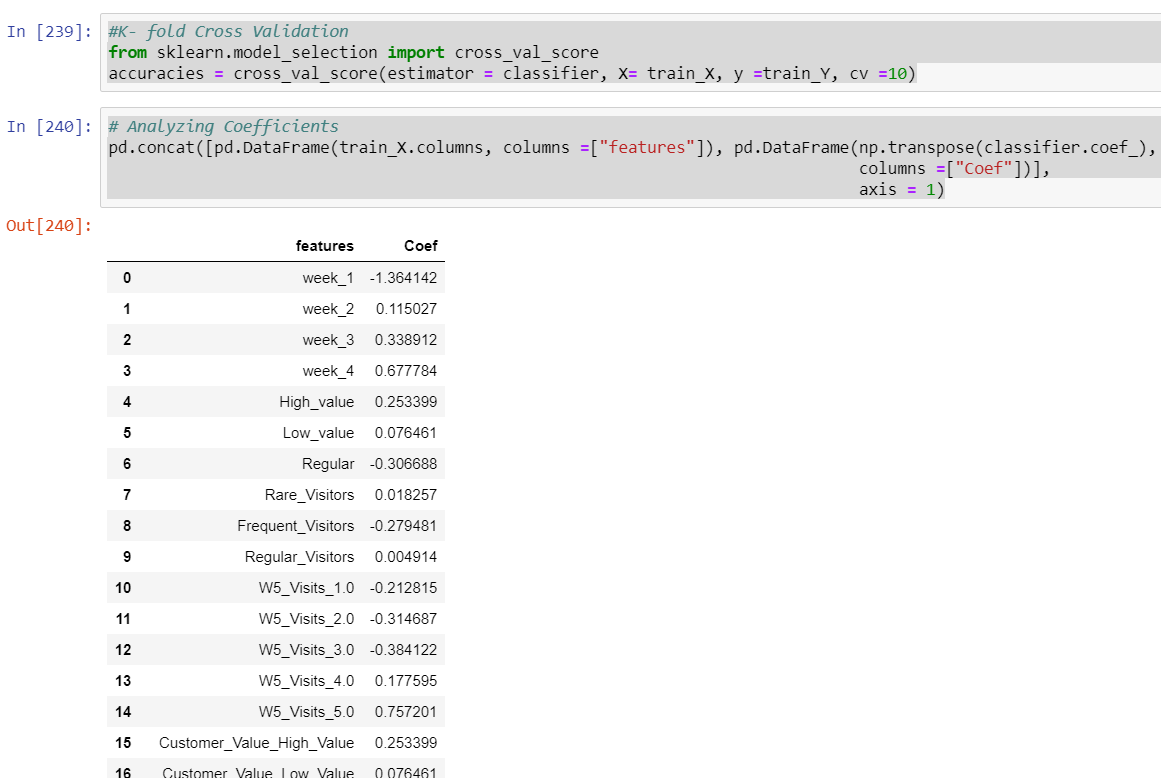
Precision score means that out of all positive predictions 65% were correct positive predictions.

Recall score means almost 58% of the positive predictions were true out of the total number of positives.

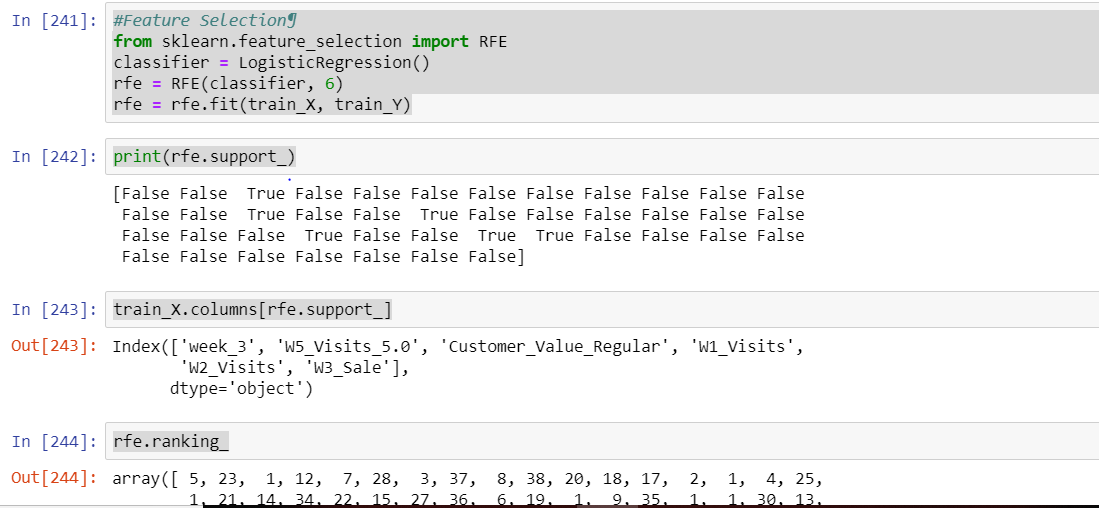
F1\_score means the weighted average of precision and recall that is 61% which is good from the dataset point of view..

Yes the model that we build is good fit depending on the scores of the model. As the model having accuracy above 50% is said to be a good fit for the model

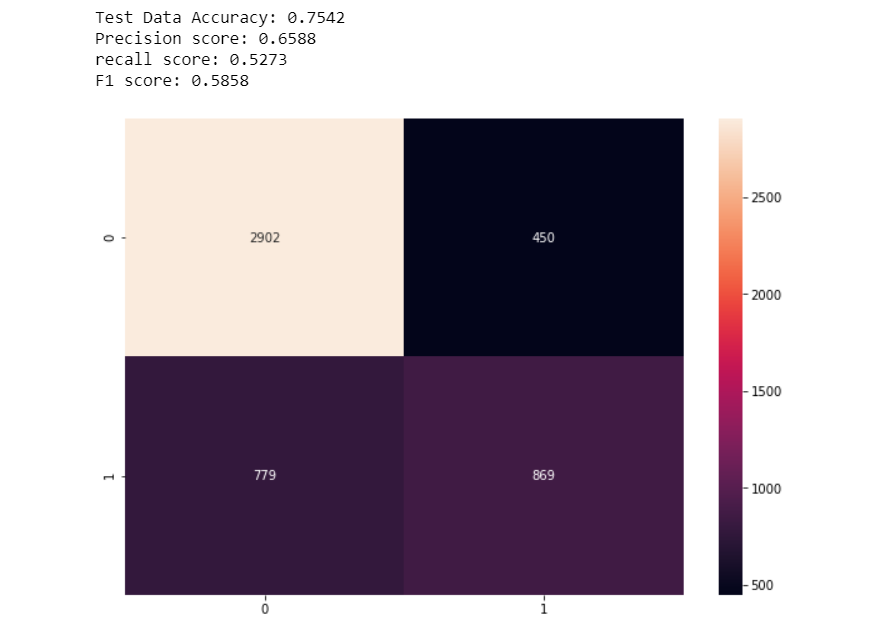
Que10 K fold cross validation



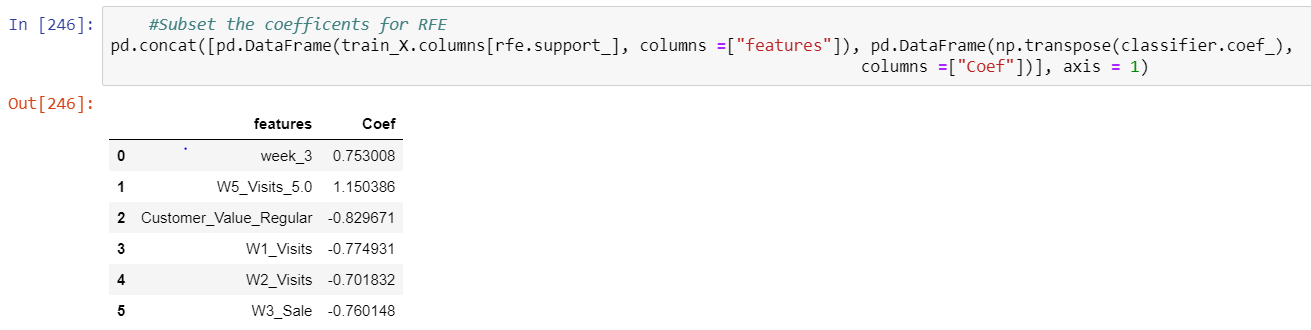
Que11.Feature Selection



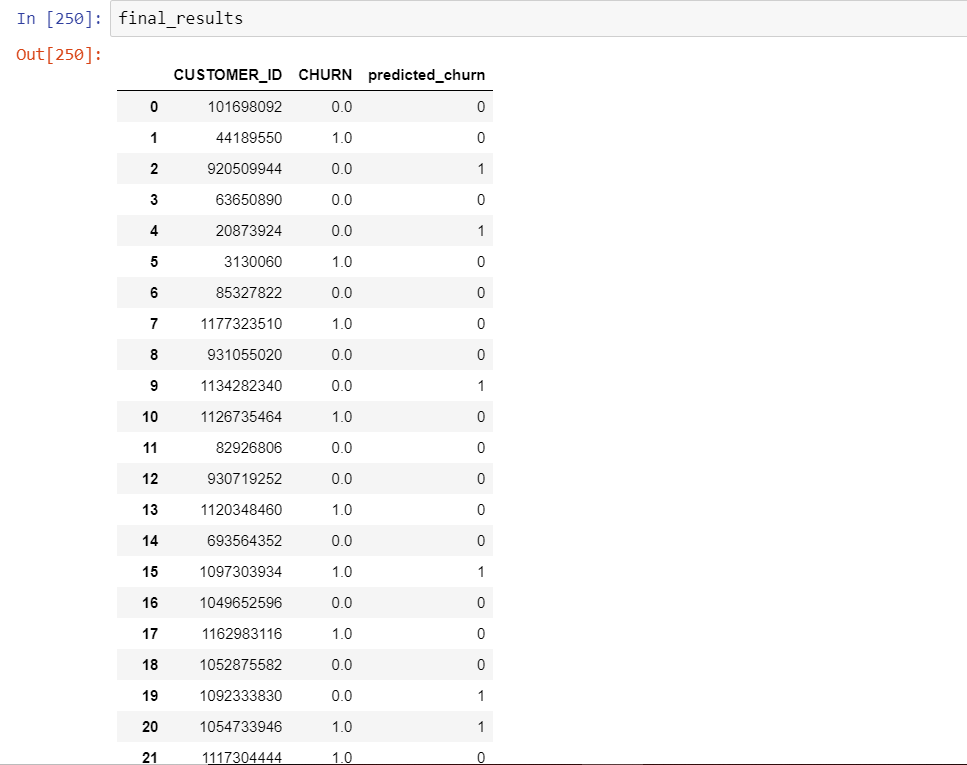
Que12.After doing feature selection what happens to the scores



After doing feature selection accuracy has decreased a little,recall has also decreased and f\_1 score is also less ,only precision score has increased



Que13.Final Results



Que14.Recommendation based on feature selection?

week 5 visits have more correlation coeff,so we can say that the customers are visiting more in that particular week,this can also mean that the customers are visting the company in alternate weeks.

Customer value Regular have least correlation coeff than other features because customers are not visiting the company regularly and are not getting the high value satisfaction with the company

Que15.What company should focus on ?

The company should focus on customer value by offering some promotional offers and festive offers to attract customers and one reason of customers not being regular can be due to poor after sales service of the company after selling the product to customer and the products of the company may be defective due to which there is less visitors.