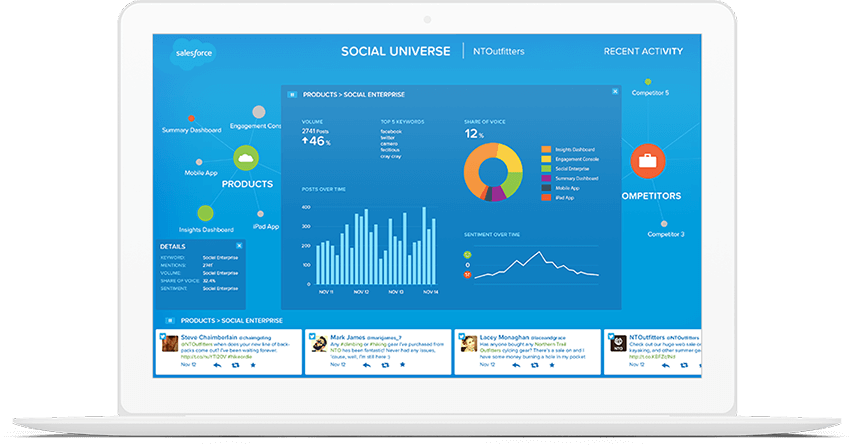
**Customer Churn Analysis**

****

**What is Customer Churn?**

Customer churn (also known as [customer attrition](https://www.optimove.com/resources/learning-center/customer-attrition)) refers to when a customer ceases his or her relationship with a company. Online businesses typically treat a customer as churned once a particular amount of time has elapsed since the customer’s last interaction with the site or service. The full cost of churn includes both lost revenue and the marketing costs involved with replacing those customers with new ones. Reducing churn is a key business goal of every online business.

### The Importance of Predicting Customer Churn

Customer retention can be achieved with good customer service and products. But the most effective way for a company to prevent attrition of customers is to truly know them. The vast volumes of data collected about customers can be used to build churn prediction models. Knowing who is most likely to defect means that a company can prioritize focused marketing efforts on that subset of their customer base.

Preventing customer churn is critically important to the telecommunications sector, as the barriers to entry for switching services are so low.

We have to examine customer data from IBM Sample Data Sets with the aim of building and comparing several customer churn prediction models.

**Project Goals :-**

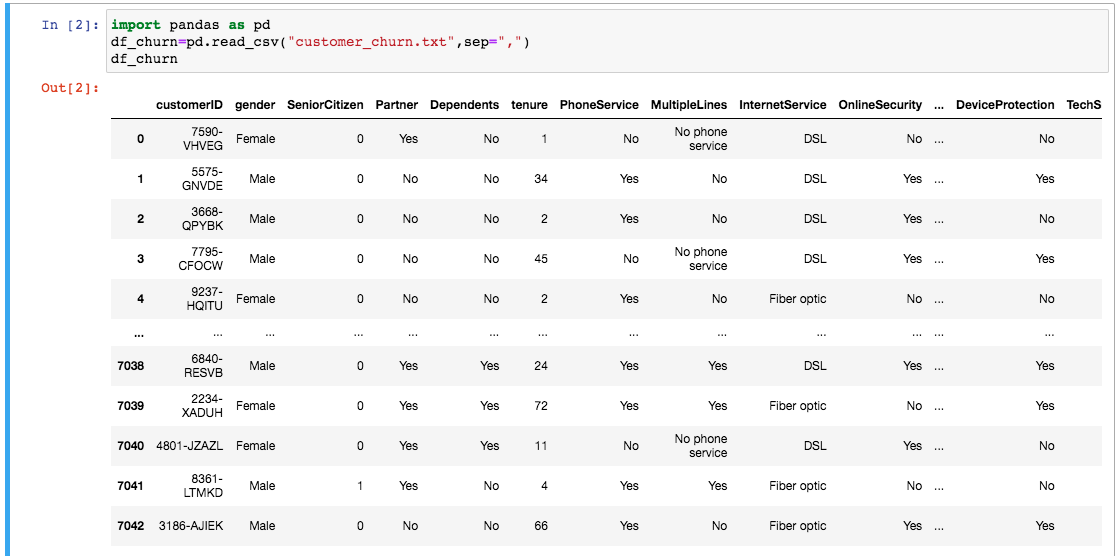
* **Data cleansing and preprocessing.**
* **Data visualization**
* **Statistical analysis of the data.**
* **Model generation for prediction of customer churn behavior.**
* **Application of Logistic Regression, SVM-Linear, Decision Tree Classifier and Random Forest algorithms on data and performance comparison.**

**Steps-**

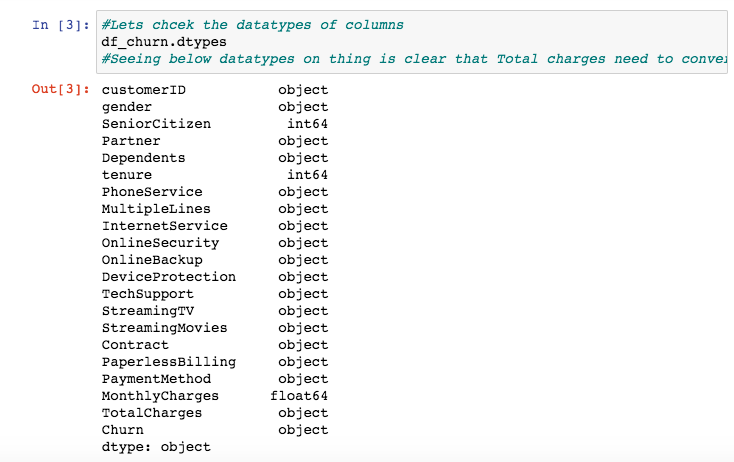
**Data Understanding**

Firstly we have to upload the data into the dataset and import into the Jupyter notebook.

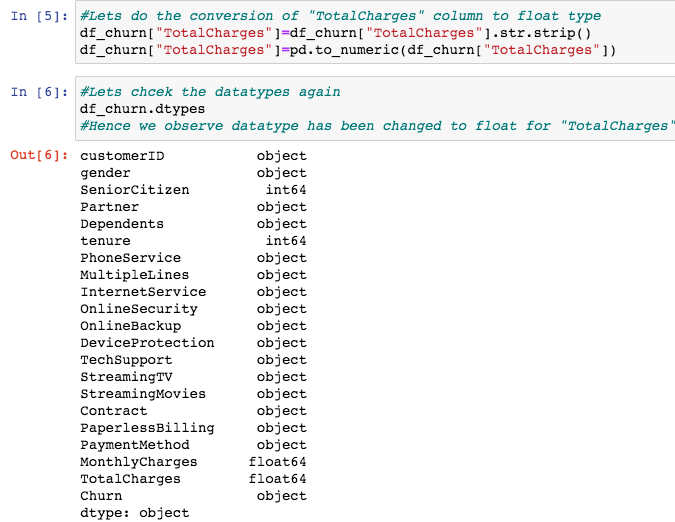
After uploading the data we will take the overview of our data and then we will check the data type of the data.

****

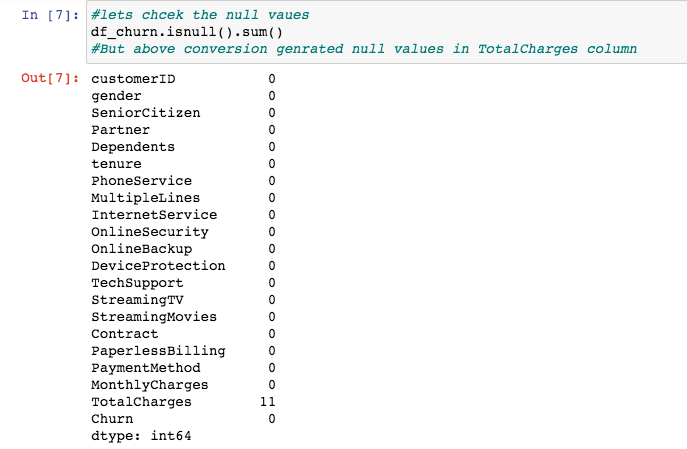
**Checking data types**

****

Converting the data type of “Total Charges” to float type after that checking the data types again.

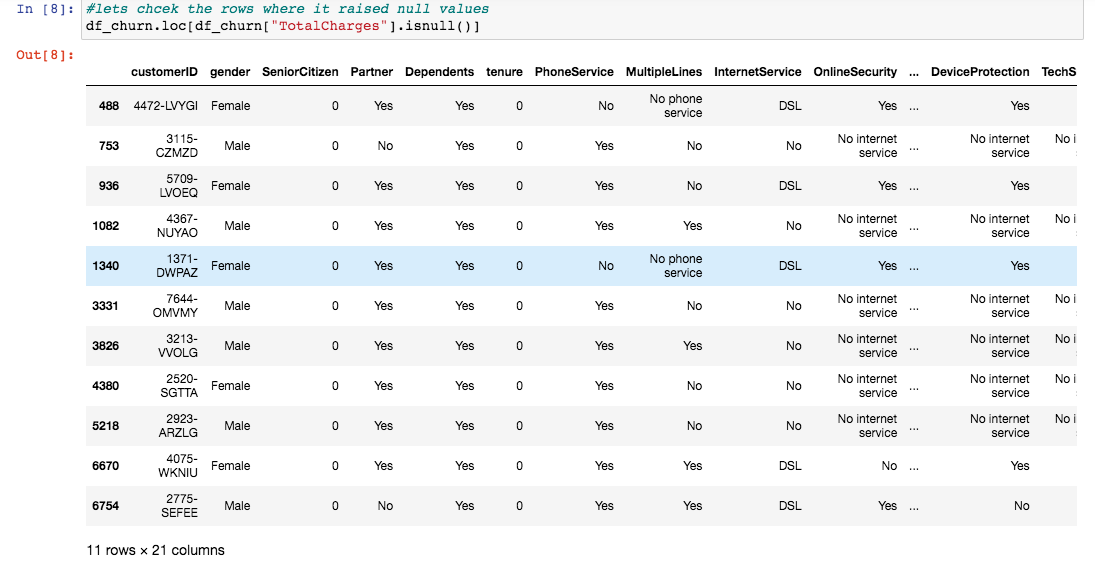
****

Checking the null values in the data and in this we have found that there are 11 null values in Total charges.

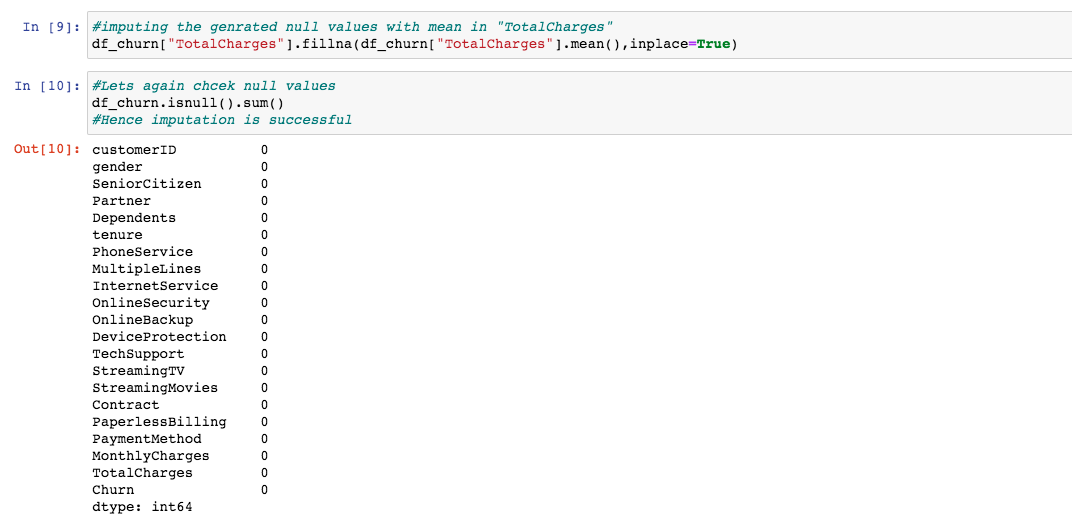
****

**Treating the Null Values**

Checking the rows where null values are present .



Removing the Null Values from rows .



Dropping the “Customer ID “ column as it is not necessary for the model.

Checking the Value Counts of each column of object datatype.

### /Users/ankitraj1/Desktop/customer churn/Screen Shot 2021-06-11 at 2.40.20 PM.png

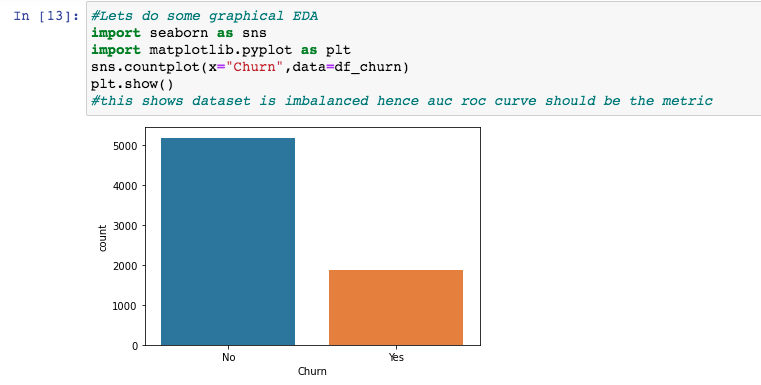
### Data Visualization and EDA :-

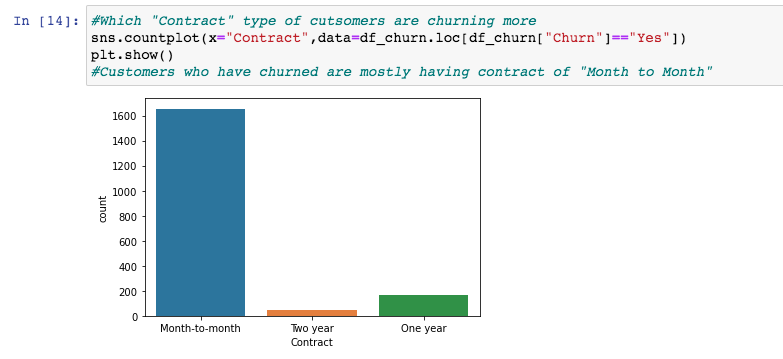
EDA usually involves a combination of the following methods:

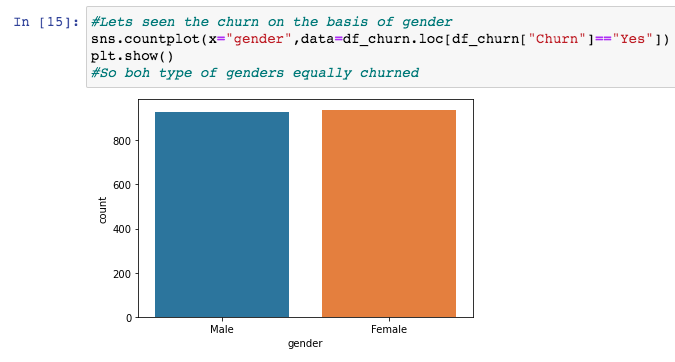
* Univariate visualization of and summary statistics for each field in the raw dataset.
* Bivariate visualization and summary statistics for assessing the relationship between each variable in the dataset and the target variable of interest (e.g. time until churn, spend)
* Multivariate visualizations to understand interactions between different fields in the data
* Dimensionality reduction to understand the fields in the data that account for the most variance between observations and allow for the processing of a reduced volume of data
* Clustering of similar observations in the dataset into differentiated groupings, which by collapsing the data into a few small data points, patterns of behavior can be more easily identified

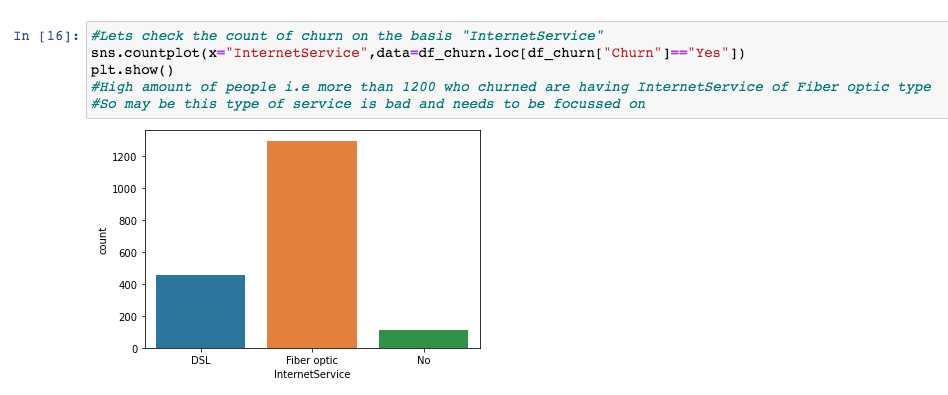
Through these methods, the data scientist validates assumptions and identifies patterns that will inform the understanding of the problem and model selection, builds an intuition for the data to ensure high quality analysis, and validates that the data has been generated in the way it was expected to.

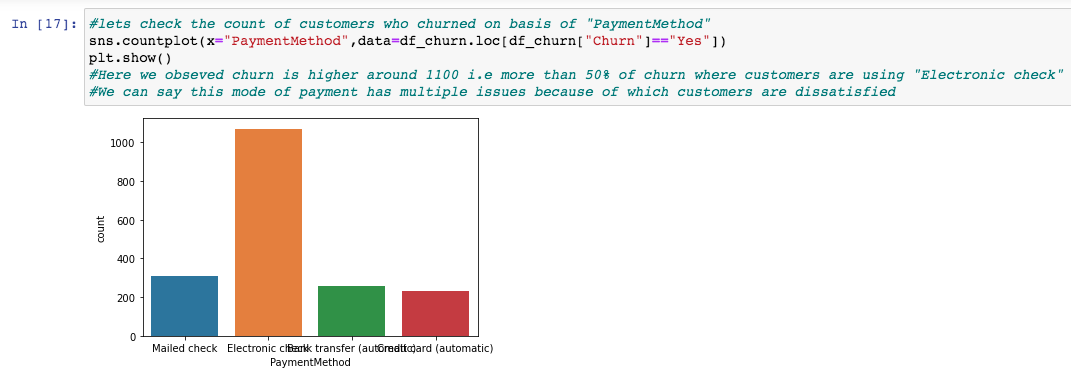
**Visualising the different comparison between objects of the model for better understanding of the dataset.**

****

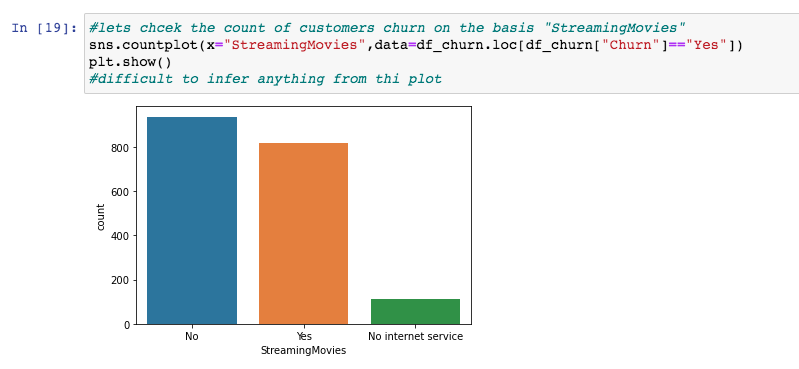


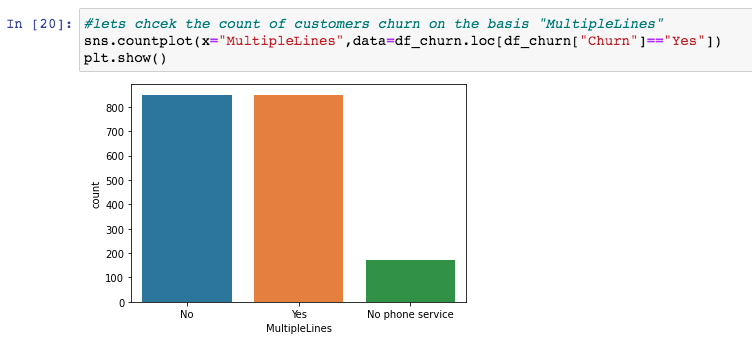


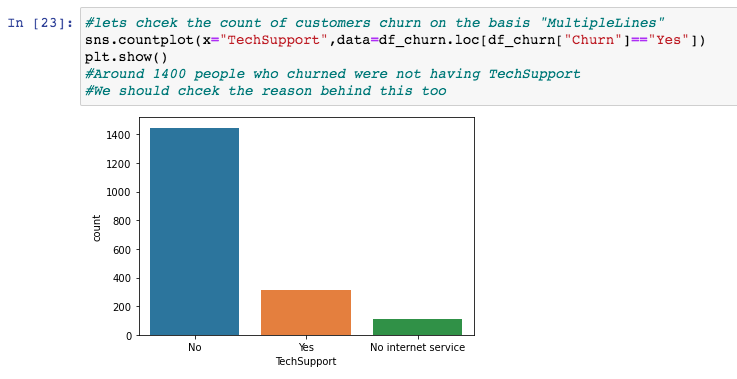


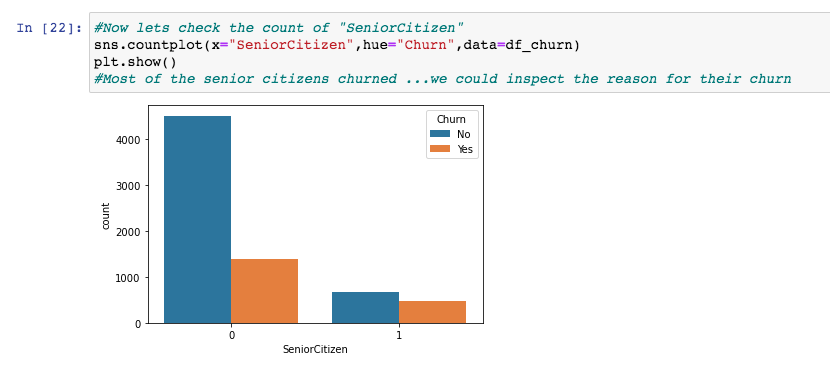


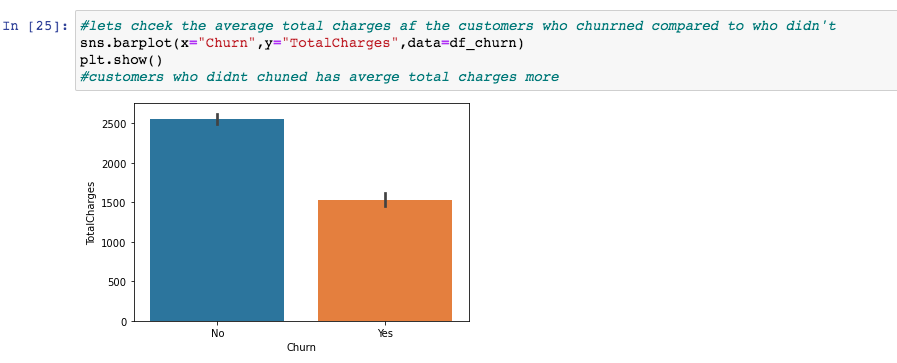
### /Users/ankitraj1/Desktop/customer churn/Screen Shot 2021-06-11 at 2.40.42 PM.png

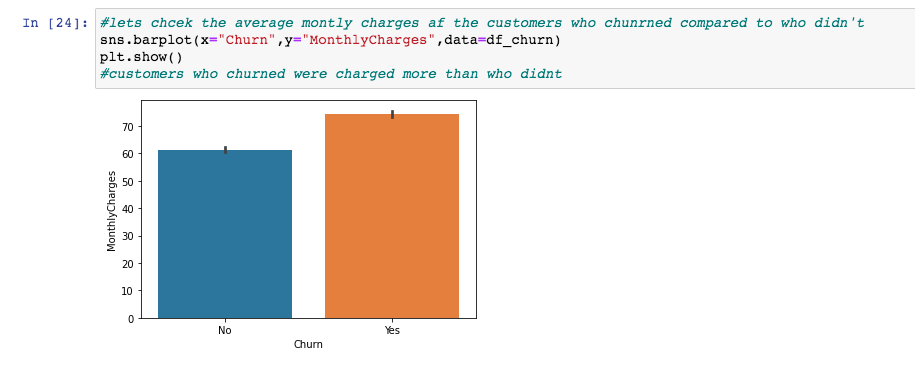












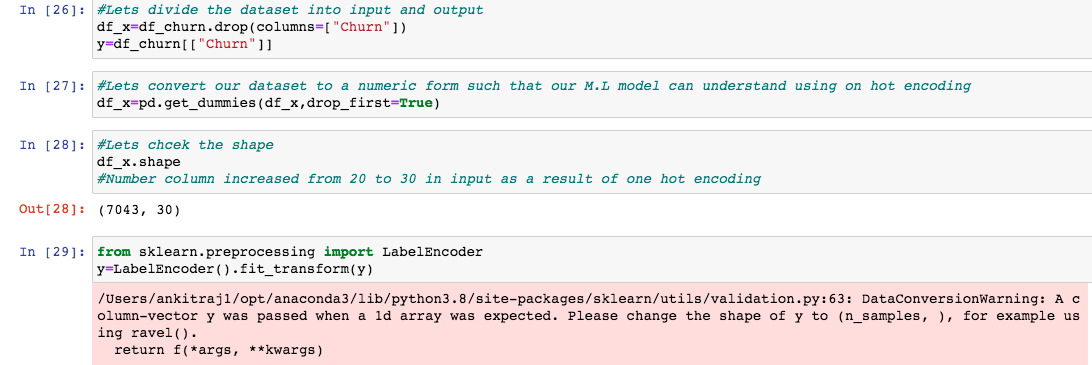
**Observations-**

* 1. The plot shows that there are many people who stay for a short time, and many people who stay for more than 60 months.
  2. Most customers have a Month-to-month contract.
  3. A lot of customers have a paperless billing.
  4. The different method by which customers do the payment are Electronic cheque, Mailed cheque ,Bank transfer, Credit card.

**FEATURE ENGINEERING**

Label Encoding :-

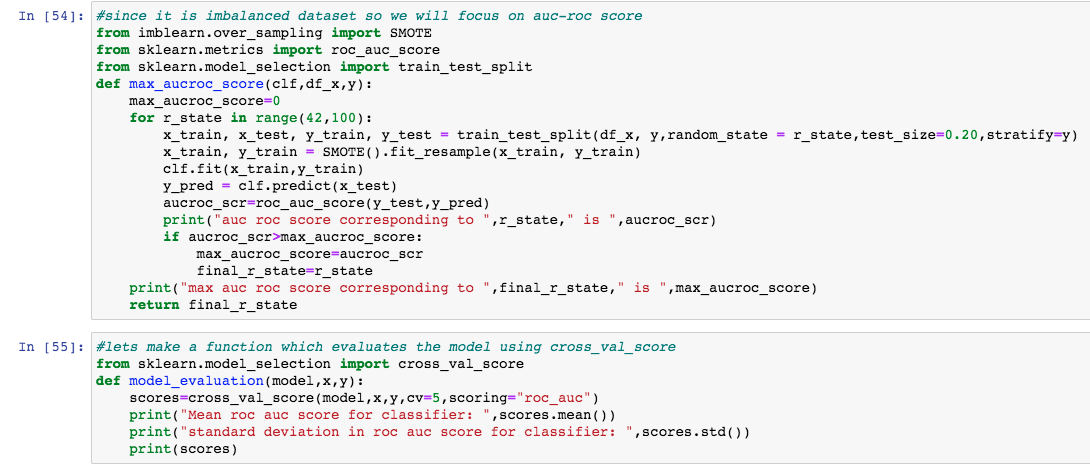
Converting our dataset to a numeric form such that our M.L model can understand using label encoding.



AUC – ROC Score :-

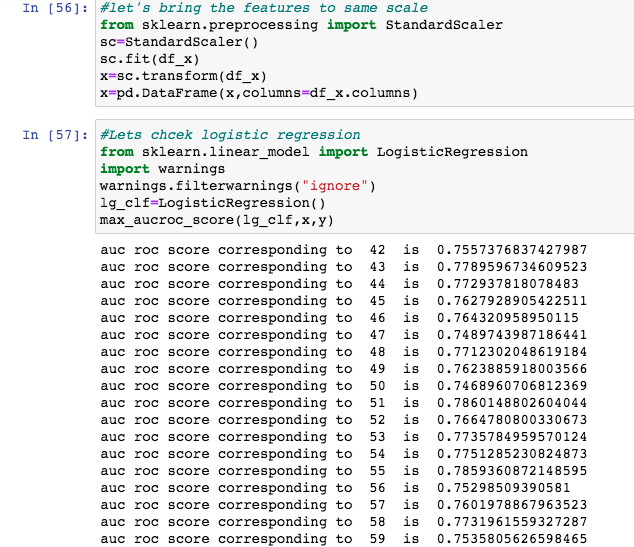
Since our dataset is imbalanced in nature so we will focus on auc - roc score.

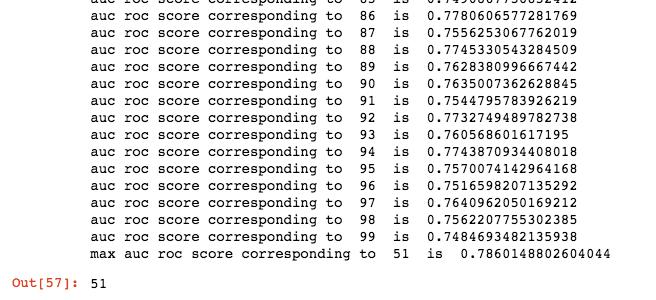
We will make a function to do Cross validation also.

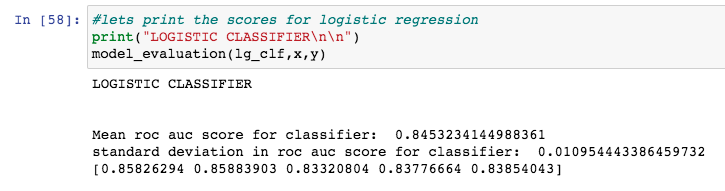


**Checking For Logistic Regression**

Printing the Scores for Logistic Regression

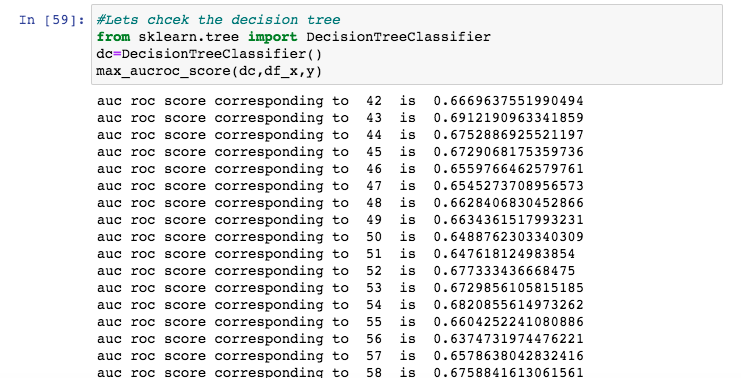
****

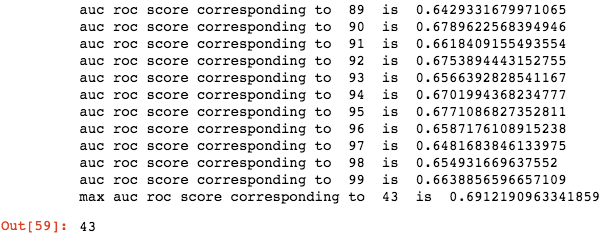
****

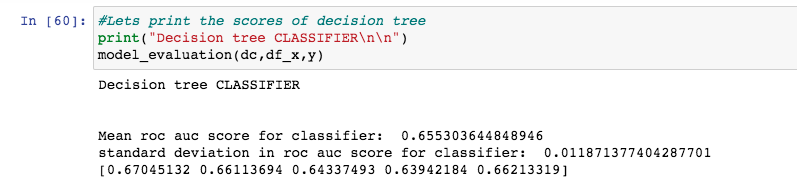
****

**Checking Decision Tree**

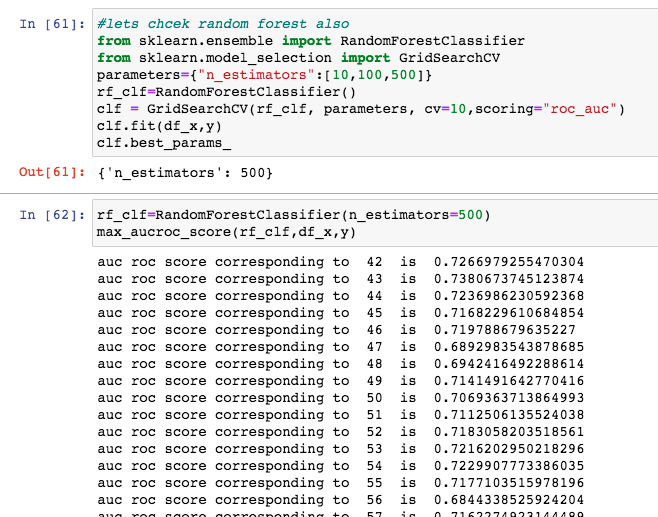
And printing scores for decision tree.

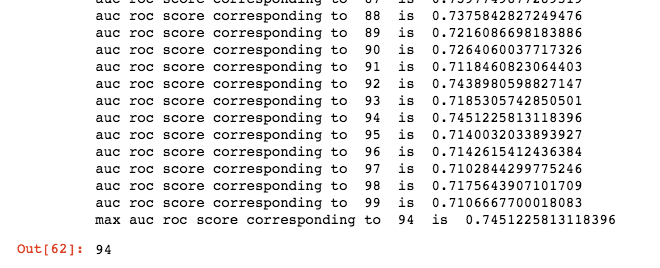




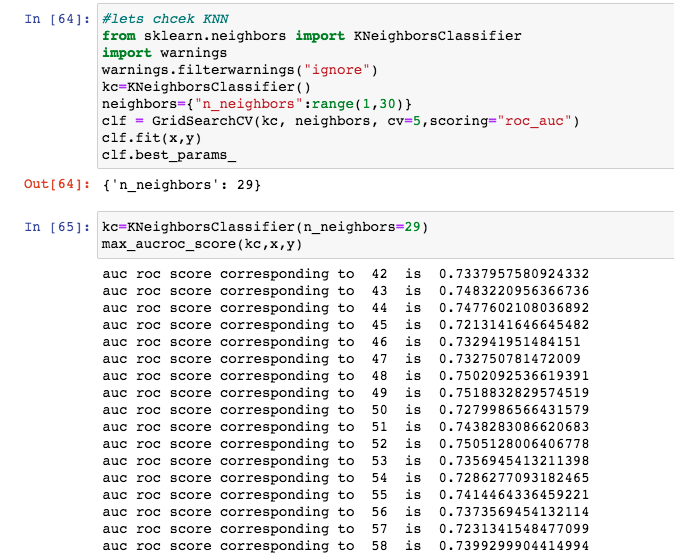


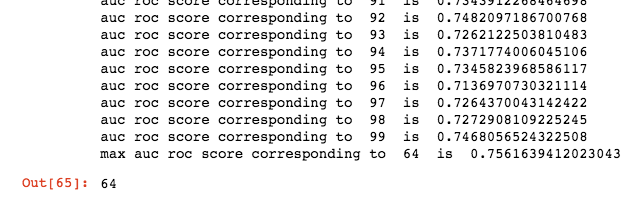
**Checking Random Forest Classifier**

****

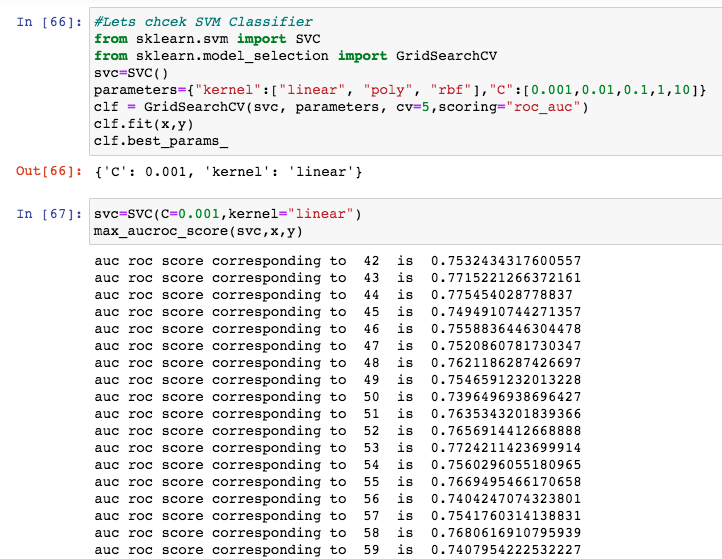
****

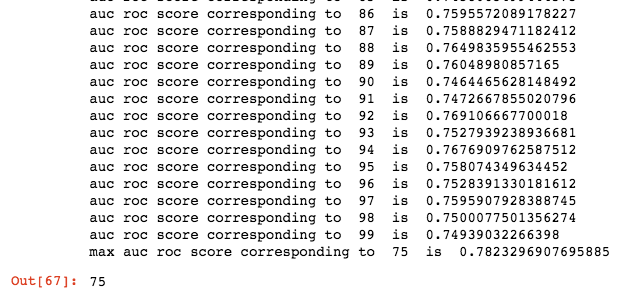
**Checking KNN and Auc – Roc Score**

****

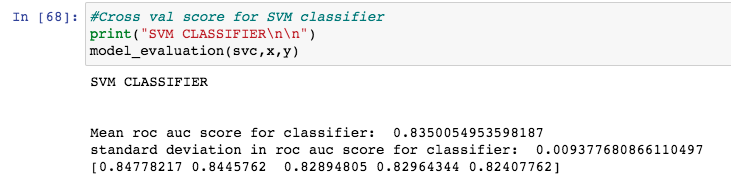
****

**Checking SVM model and Auc – Roc Score**

****

****

**We are Choosing SVM as our Final Model and we will Calculate the Cross validation Score for SVM Model .**

****

**Findings :-**

* Women are more likely to churn than men.
* Older (by age) customers are more likely to churn.
* Larger proportion of churned customers have poor credit scores.

**Predicting customers who are more likely to churn :-**

* Out of the prediction models, random forest predicted the churned customers with the highest accuracy of 82%.
* Age, Estimated Salary, Credit Scores, Number of products turned out to be significant variables.

**FEEDBACK**

**For any queries or question related to the Model or any suggestions please feel free to contact me on -**

**ankitverma87416@gmail.com**