**Customer Churn Analysis**



Customer churn is when a company’s customers stop doing business with that company. Businesses are very keen on measuring churn because keeping an existing customer is far less expensive than acquiring a new customer. New business involves working leads through a sales funnel, using marketing and sales budgets to gain additional customers. Existing customers will often have a higher volume of service consumption and can generate additional customer referrals.

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 -**

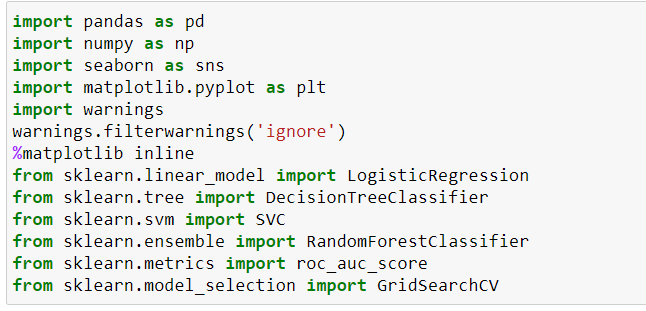
* Data cleansing and preprocessing.
* Data visualization and Exploratory Data Analysis.
* Statistical analysis of the data.
* Model generation for prediction of customer churn behavior.
* Application of Logistic Regression, SVM-Linear, SVM-RBF and Random Forest algorithms on data and performance comparison.

**Steps-**

* Data Understanding-

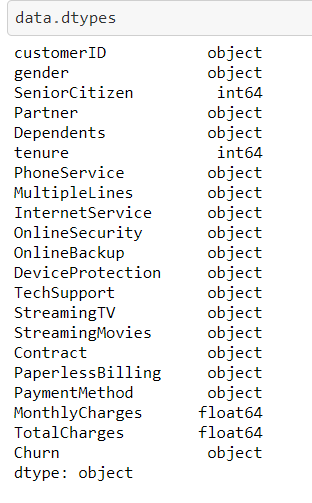
Firstly importing all the necessary libraries in the Jupyter notebook .

These, are the necessary libraries which will be used in our dataset.

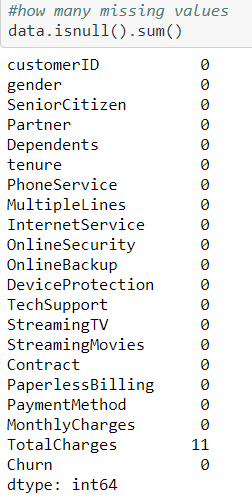


Then we have to upload the data into the dataset and import into the Jupyter notebook.

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



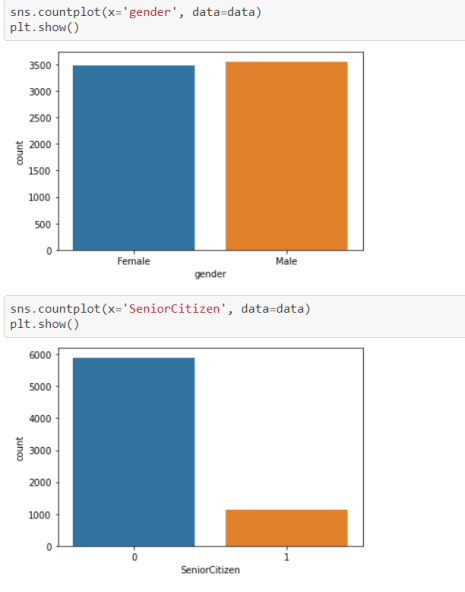
Checking the null values in the data and in this we have found that there are 11 null values in Total charges.

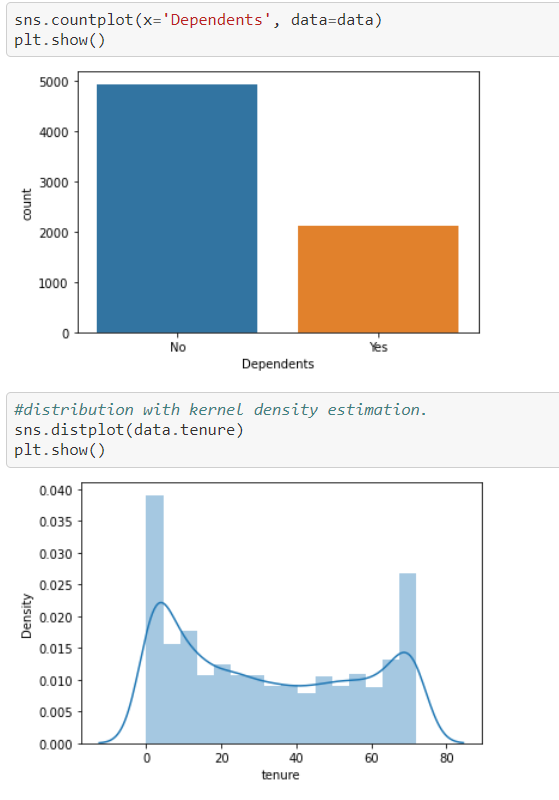


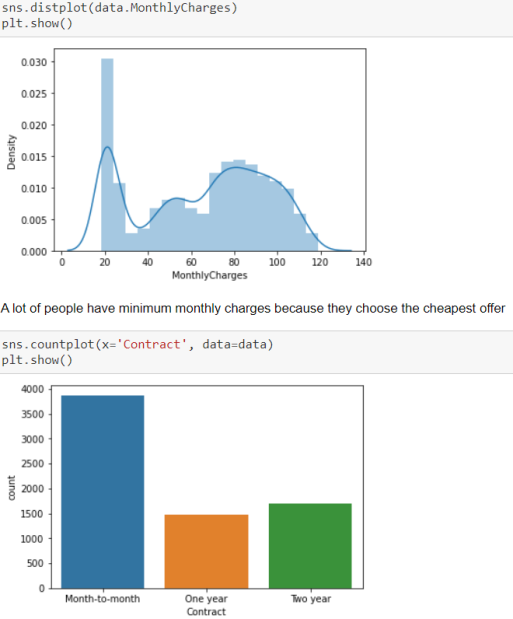
* Exploratory Data Analysis –



* Univariate Analysis-

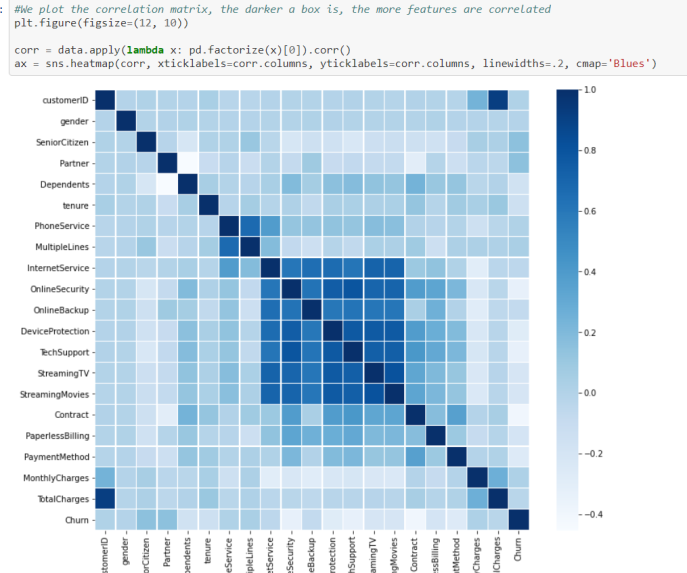






Observations-

* 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.
* Correlation Matrix-



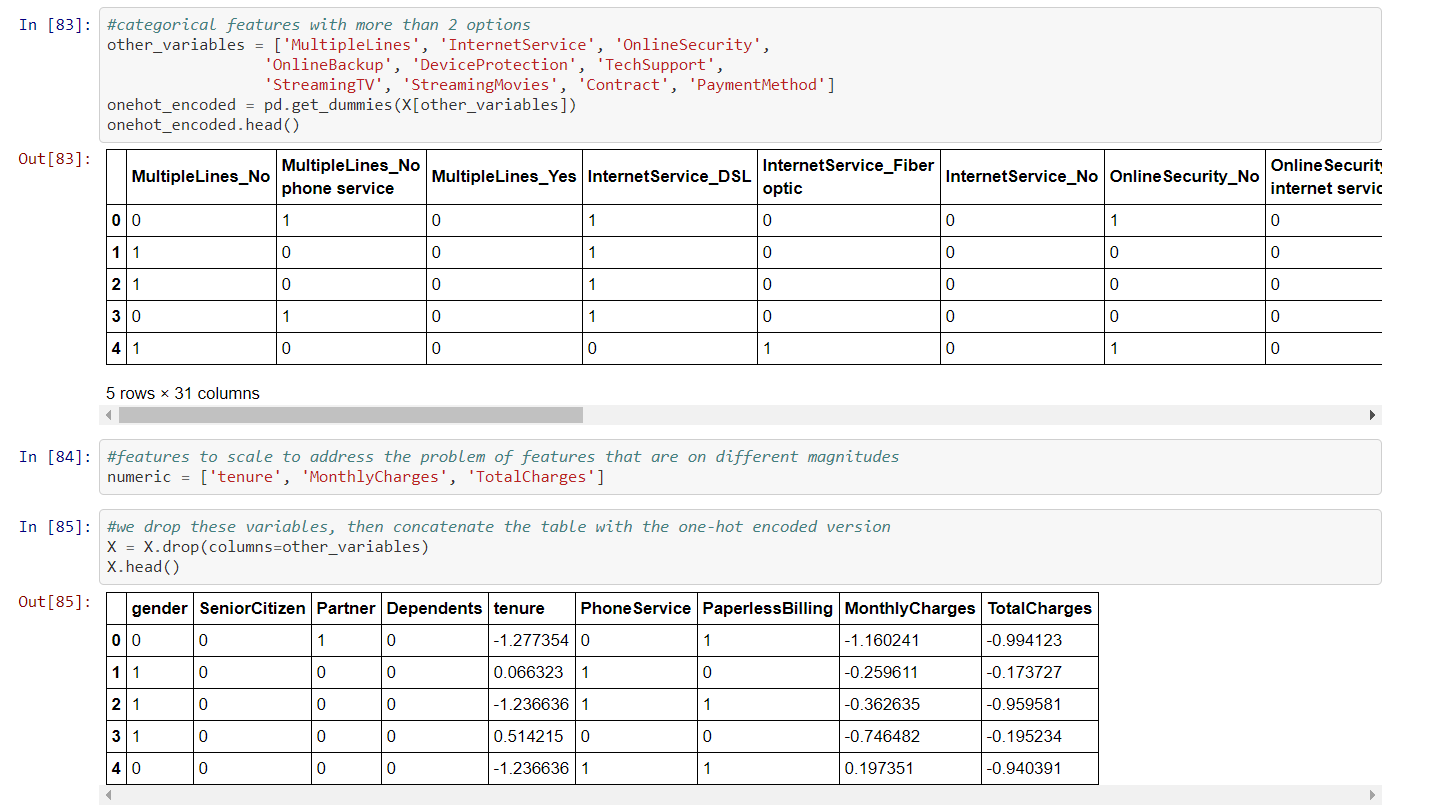
\*Internet service, Online security, Online Backup, DeviceProtection, Tech Support and streaming are highly correlated features.

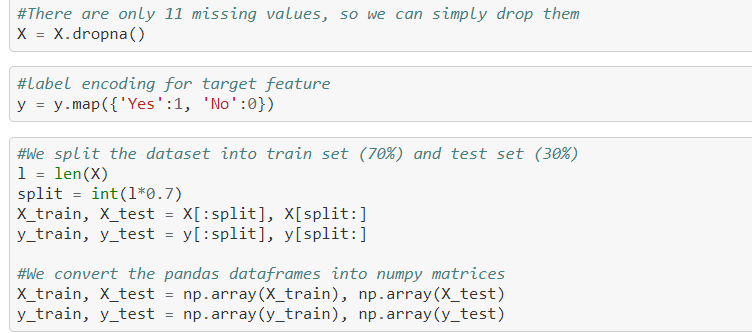
\*Total charges and customer ID are also very correlated, maybe the ID is chosen according to high-potential customers.

\*The most correlated to churn : Senior, Partner, Multiple lines, online backup, Monthly charges.

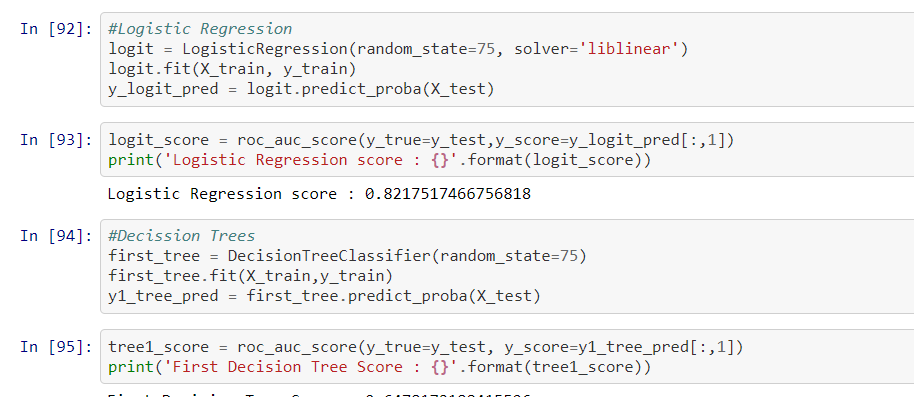
* Model Selection-







* Training Models-



The best Score is of Logistic regression which is 82%.

**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.

**Recommendations:**

* Increase awareness amongst customers on importance of maintaining a healthy credit score.
* Customized financial advisory and investment offerings to different segments of customers.
* Launch Digital products and diverse schemes to woo customers from different age brackets.