CUSTOMER CHURN PREDICTION IN TELECOM INDUSTRY

HOW CUSTOMER CHURN AND QUALITY OF SERVICE ARE RELATED?

Let us consider the situation of India. In India there are nearly ten telecom providers. If a customer is dissatisfied, he has other options to explore.

Different customers tend to have different requirements. Consider a population of 15 teenage boys, 15 working persons and 15 senior citizens. The teenage boys tend to spend more time on internet, the working class spends less time over both internet and voice call, and senior citizens tend to spend more time on calls as compared to teenage boys. There is a need to understand the requirements of each of the class.

The Churn Prediction Models takes into account various demographic aspects along with financial burden and usage pattern of various groups. Attaching weightage to the features will enable the telecom industry to direct specific services towards group of customers.

Providing sufficient Quality of Service (QoS) across IP networks is becoming an increasingly important aspect of today's enterprise IT infrastructure. Not only is QoS necessary for voice and video streaming over the network, it's also an important factor in supporting the growing Internet of Things (IoT).

The Churn Prediction Model will aid the telecom service providers to concentrate to improve the QoS of services provided by them.

INTRODUCTION

Customer attrition, also known as customer churn, customer turnover, or customer defection, is the loss of clients or customers.

Telephone service companies, Internet service providers, pay TV companies, insurance firms, and alarm monitoring services, often use customer attrition analysis and customer attrition rates as one of their key business metrics because the cost of retaining an existing customer is far less than acquiring a new one. Companies from these sectors often have customer service branches which attempt to win back defecting clients, because recovered long-term customers can be worth much more to a company than newly recruited clients.

Companies usually make a distinction between voluntary churn and involuntary churn. Voluntary churn occurs due to a decision by the customer to switch to another company or service provider, involuntary churn occurs due to circumstances such as a customer's relocation to a long-term care facility, death, or the relocation to a distant location. In most applications, involuntary reasons for churn are excluded from the analytical models. Analysts tend to concentrate on voluntary churn, because it typically occurs due to factors of the company-customer relationship which companies control, such as how billing interactions are handled or how after-sales help is provided.

Predictive analytics use churn prediction models that predict customer churn by assessing their propensity of risk to churn. Since these models generate a small prioritized list of potential defectors, they are effective at focusing customer retention marketing programs on the subset of the customer base who are most vulnerable to churn.

DATA OVERVIEW

➢ CONTENT

Each row represents a customer, each column contains customer's attributes described on the column Metadata.
The data set includes information about:

- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents.

➢ DATA DESCRIPTION

customerID – Unique Identification number assigned to a customer
gender – Customer's Gender **(MALE, FEMALE)**
SeniorCitizen – Whether the customer is a senior citizen. **(0, 1)**
Partner – Whether the customer has a partner. **(YES, NO)**
Dependents – Whether the customer has any dependent. **(YES, NO)**
Tenure – Number of month customer has subscribed.
PhoneService – Whether the customer has Phone service. **(YES, NO)**
MultipleLines – Whether the customer has multiple lines under same subscription. **(Yes, No, No Phone Service)**
InternetService – Whether the customer has Internet service, and what type. **(Fibre optic, DSL No)**
OnlineSecurity – Whether the customer has online security. **(Yes, No)**
OnlineBackup – Whether the customer has online backup facility. **(Yes, No)**
DeviceProtection – Whether the customer has device protection plan. **(Yes, No)**
TechSupport – Whether the customer is entitled to tech support. **(Yes, No)**
StreamingTV – Whether the customer has streaming of TV enabled. **(Yes, No)**
StreamingMovies – Whether the customer has streaming of movies enabled
Contract – The basis of contract of the customer. **(Month-to-Month, One Year, Two Year)**
PaperlessBilling – Whether the customer has opted for paperless billing. **(Yes, No)**
PaymentMethod – Payment method of the customer **(Electronic Check, Mailed Check, Bank Transfer (automatic), Credit Card (automatic))**
MonthlyCharges – Monthly charges of the customer.
TotalCharges – Total charges billed to the customer during their tenure.
Churn – Whether the customer churned or not

➢ LOADING DATASET INTO DATAFRAME

```
telcom = pd.read_csv('C:/Users\Raktim\Desktop\FINAL YEAR PROJECT\WA_Fn-UseC_-Telco-Customer-Churn.csv')
#first few rows
telcom.head()
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSup
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes	
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No	
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes	
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No	

5 rows × 21 columns

➢ NO. OF UNIQUE VALUES IN EACH COLUMN

```
Unique values :
 customerID          7043
gender                 2
SeniorCitizen          2
Partner                2
Dependents             2
tenure                73
PhoneService           2
MultipleLines          3
InternetService        3
OnlineSecurity         3
OnlineBackup           3
DeviceProtection       3
TechSupport            3
StreamingTV            3
StreamingMovies        3
Contract               3
PaperlessBilling       2
PaymentMethod          4
MonthlyCharges      1585
TotalCharges        6531
Churn                  2
dtype: int64
```

EXPLORATORY DATA ANALYSIS

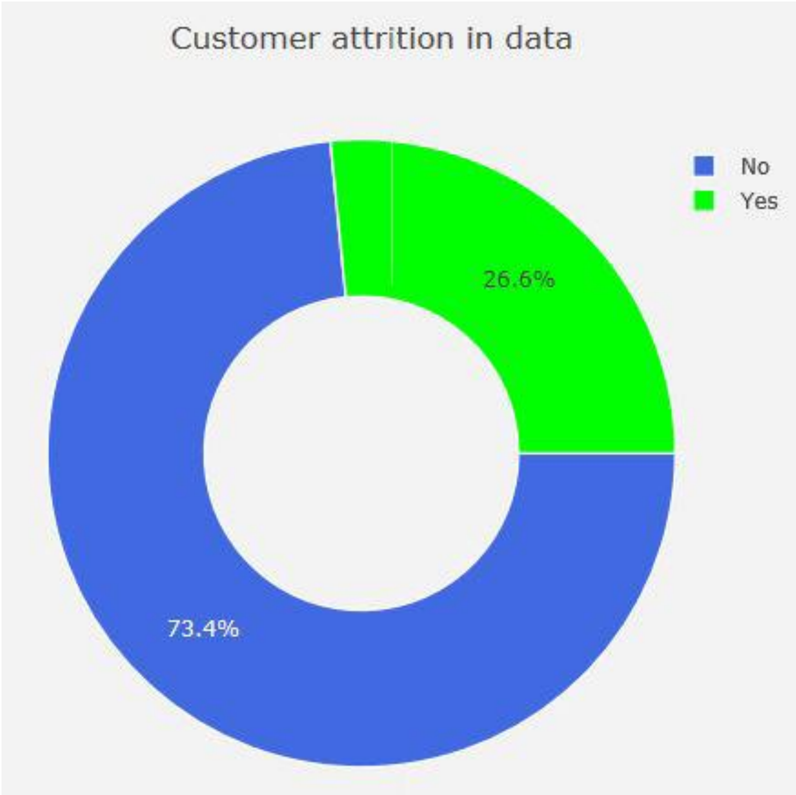


FIG: CUSTOMER ATTRITION IN DATA

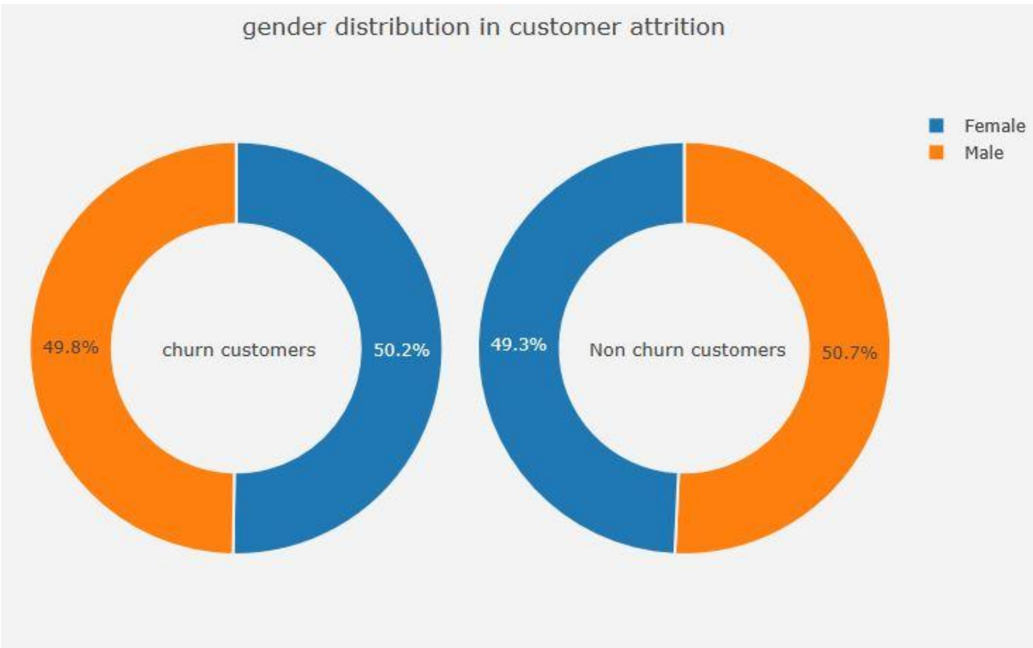


FIG: GENDER DISTRIBUTION IN CUSTOMER ATTRITION

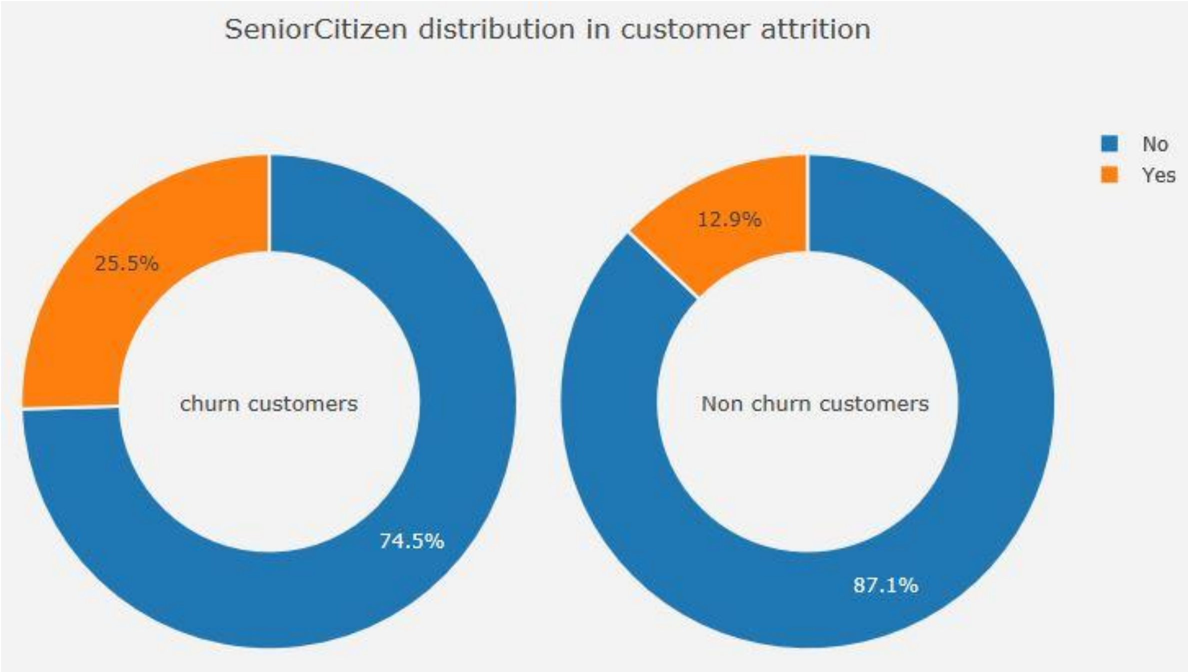


FIG: SENIOR CITIZEN DISTRIBUTION IN CUSTOMER ATTRITION

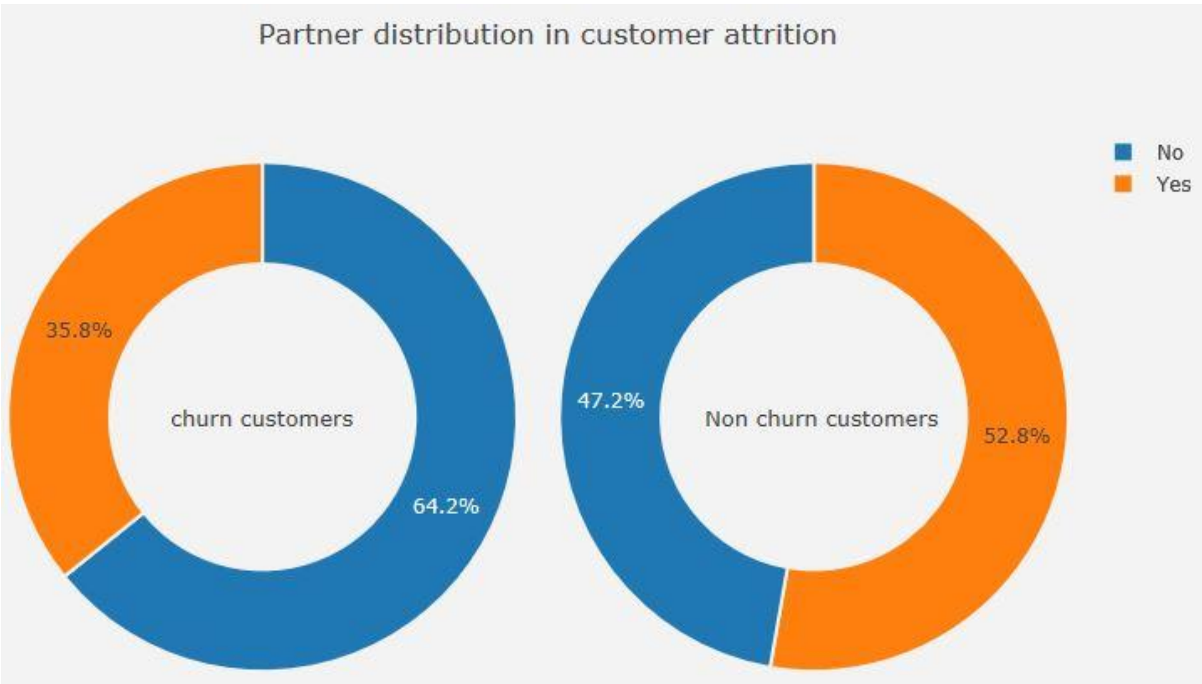


FIG: PARTNER DISTRIBUTION IN CUSTOMER ATTRITION

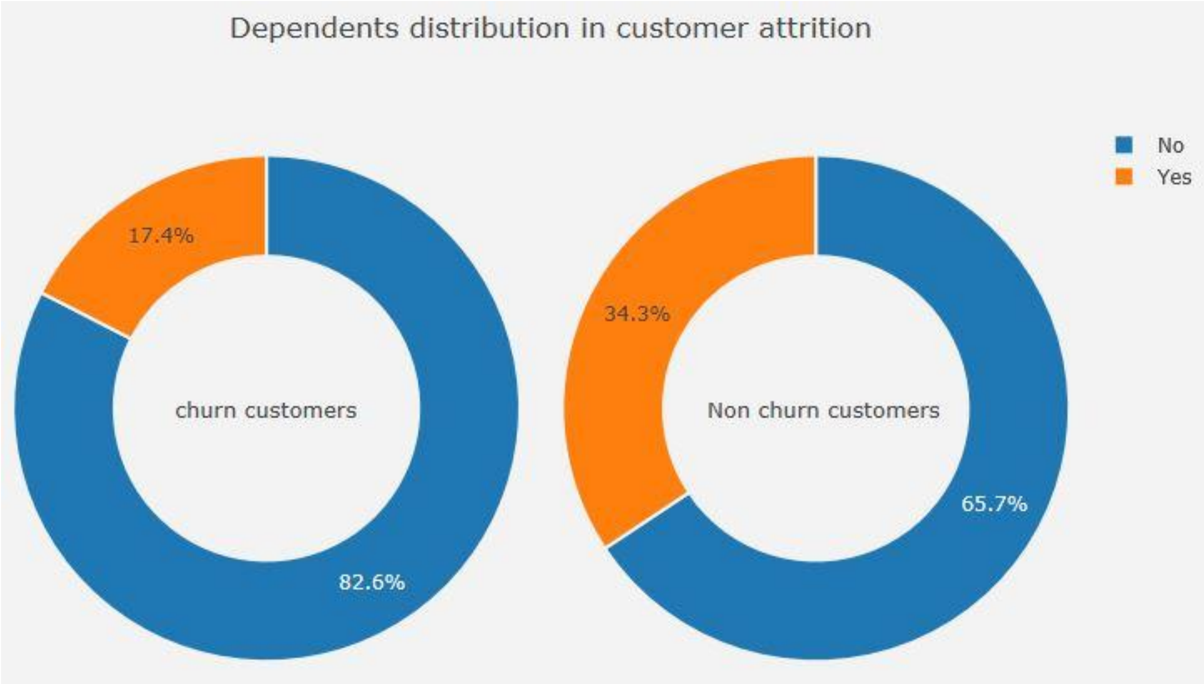


FIG: DEPENDENTS DISTRIBUTION IN CUSTOMER ATTRITION

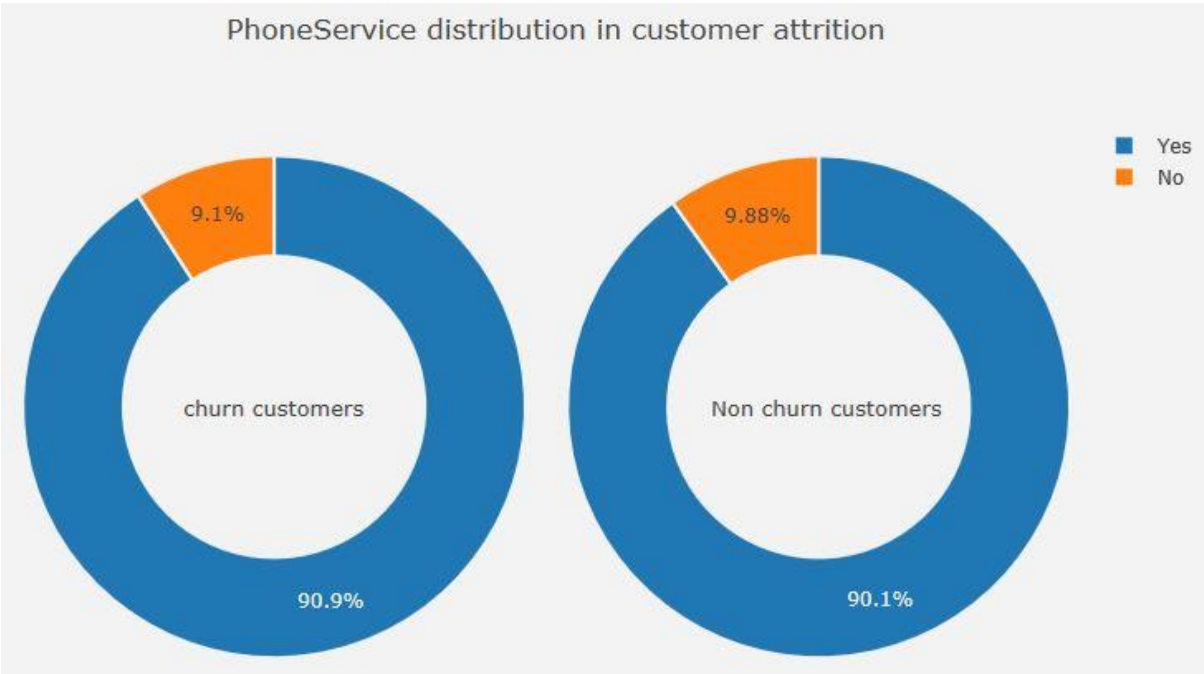


FIG: PHONESERVICE DISTRIBUTION IN CUSTOMER ATTRITION

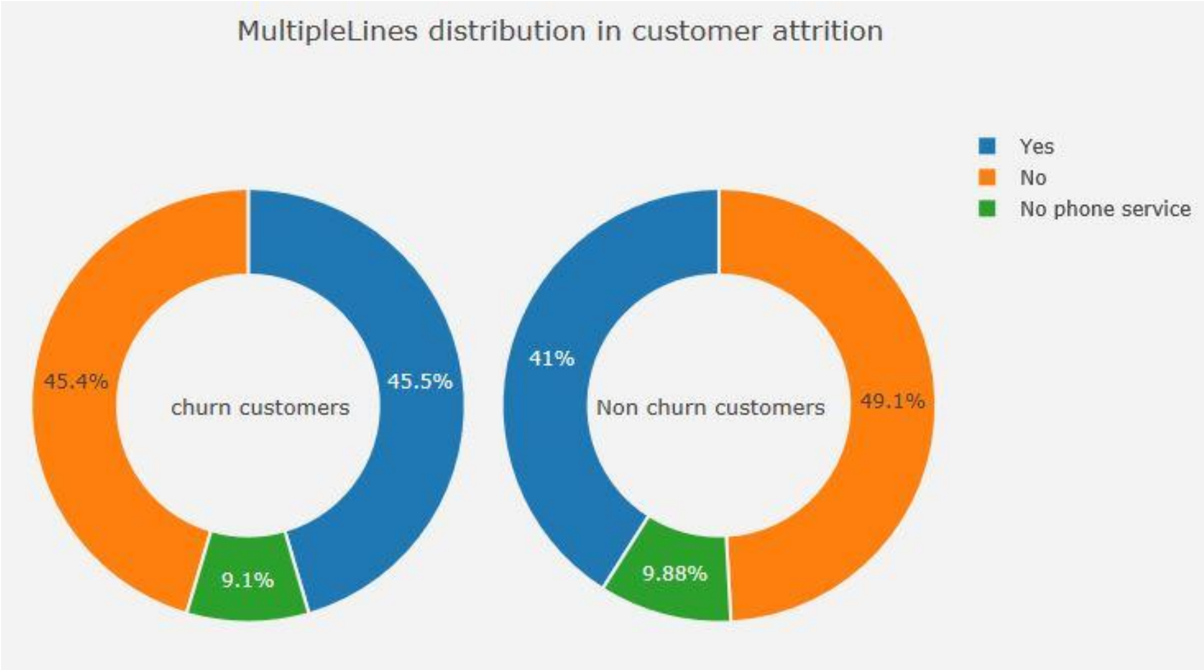


FIG: MULTIPLELINES DISTRIBUTION IN CUSTOMER ATTRITION

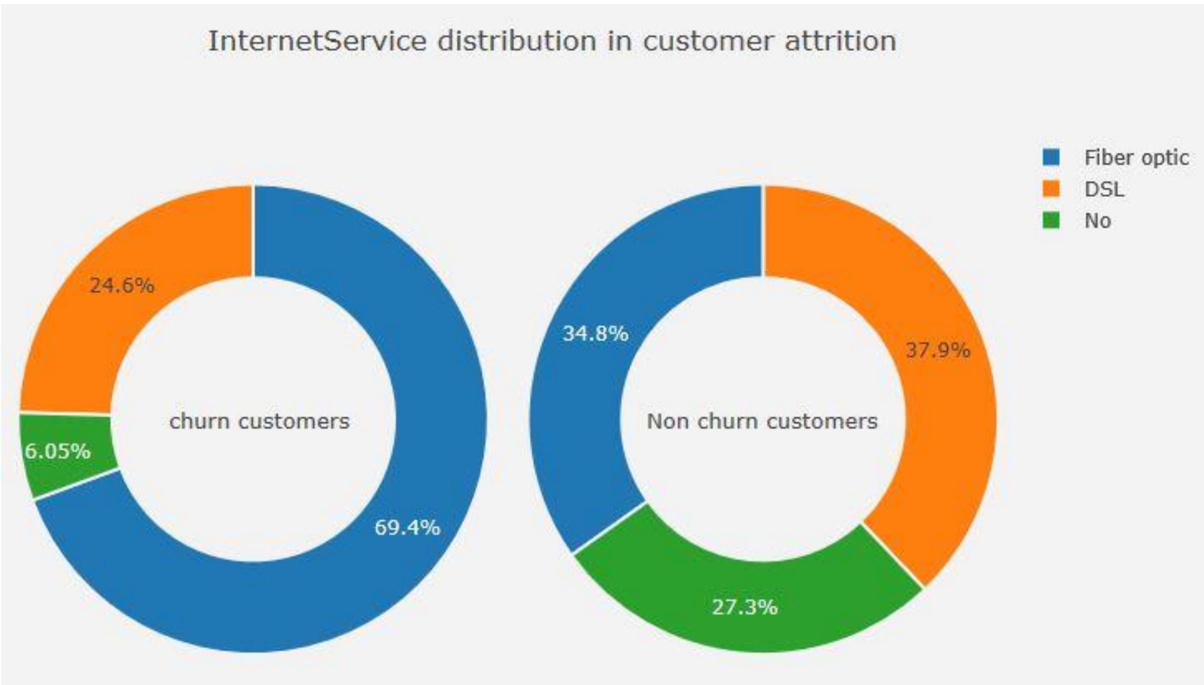


FIG: INTERNET SERVICE DISTRIBUTION IN CUSTOMER ATTRITION

FIG: ONLINESECURITY DISTRIBUTION IN CUSTOMER ATTRITIONS

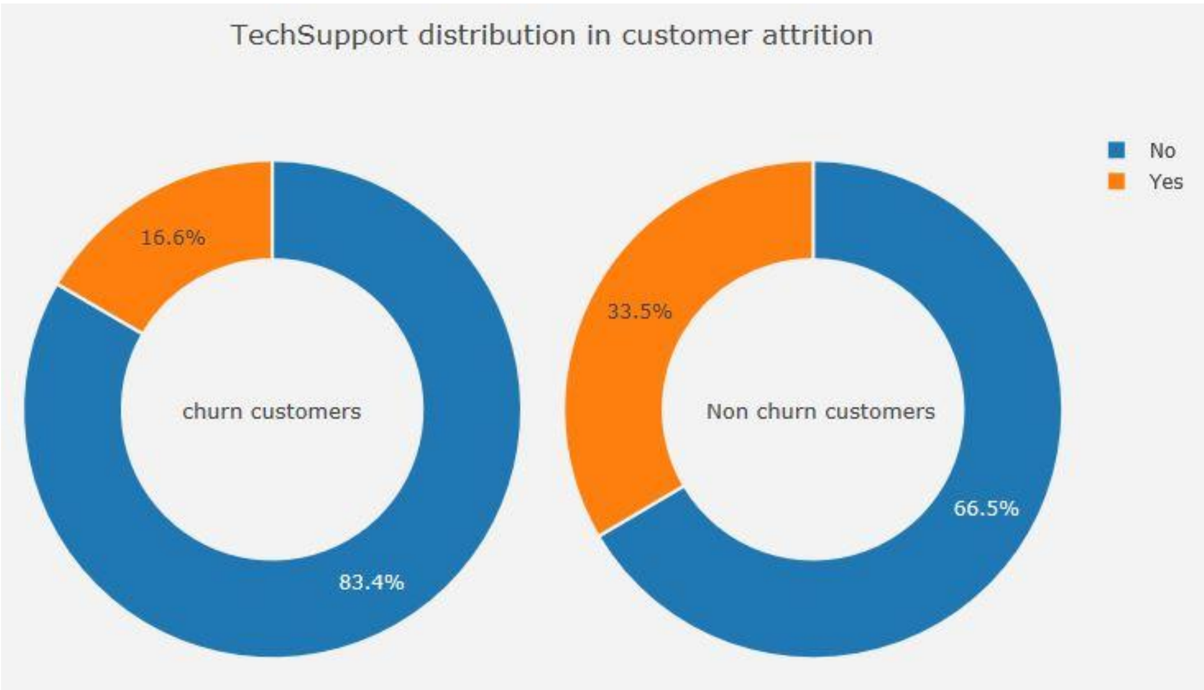


FIG: TECHSUPPORT DISTRIBUTION IN CUSTOMER ATTRITION

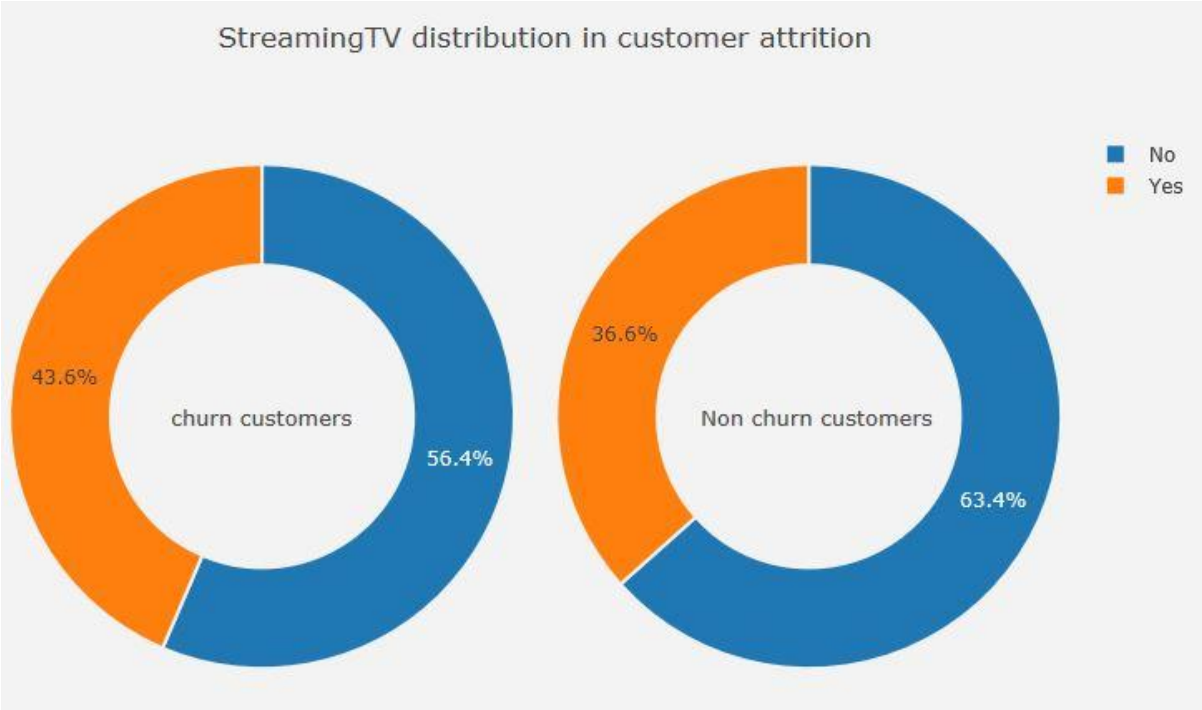


FIG: STREAMINGTV DISTRIBUTION IN CUSTOMER ATTRITION

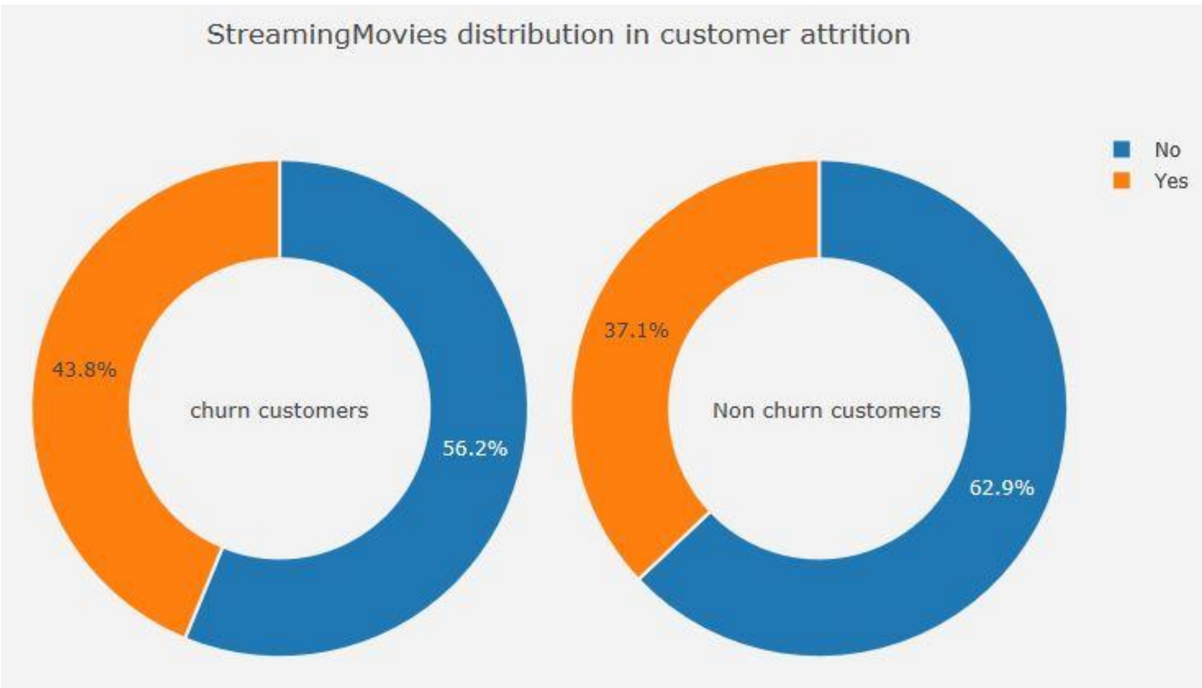


FIG: STREAMINGMOVIES DISTRIBUTION IN CUSTOMER ATTRITION

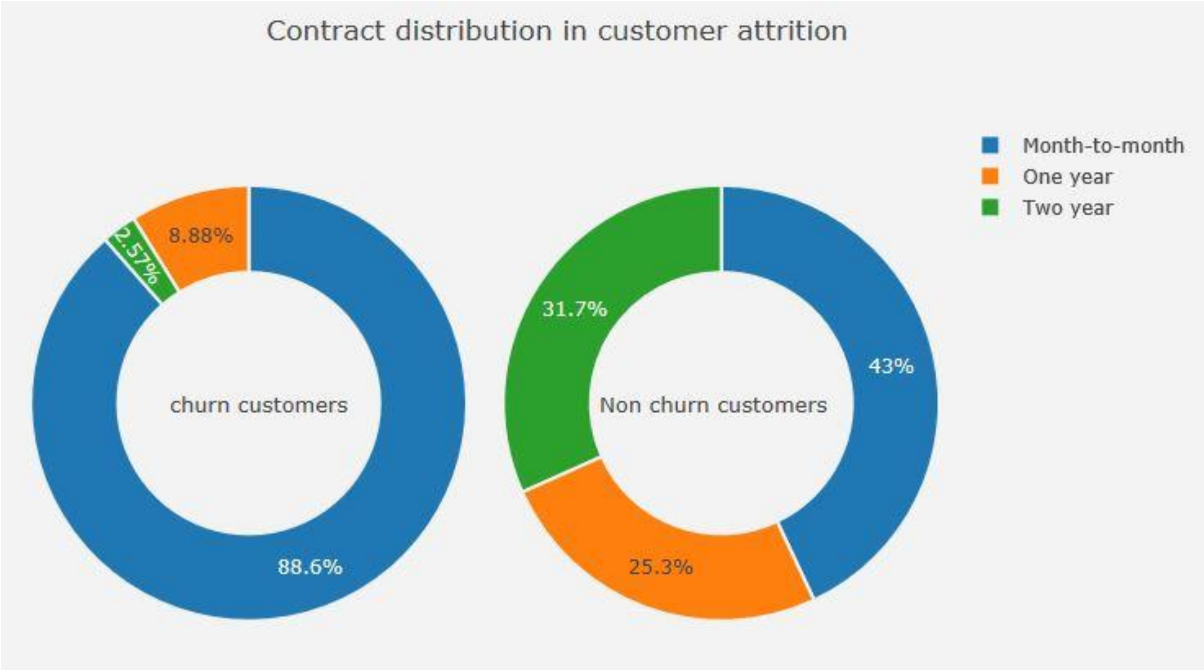


FIG: CONTRACT DISTRIBUTION IN CUSTOMER ATTRITION

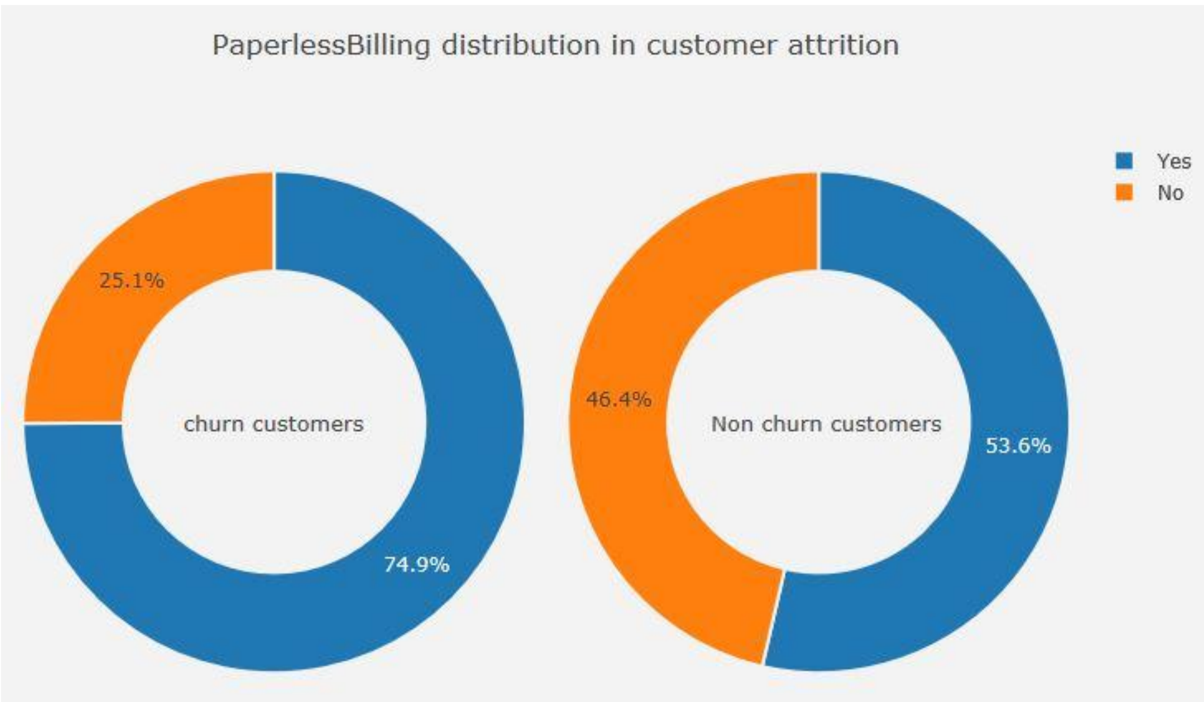


FIG: PAPERLESSLESSBILLING DISTRIBUTION IN CUSTOMER ATTRITION

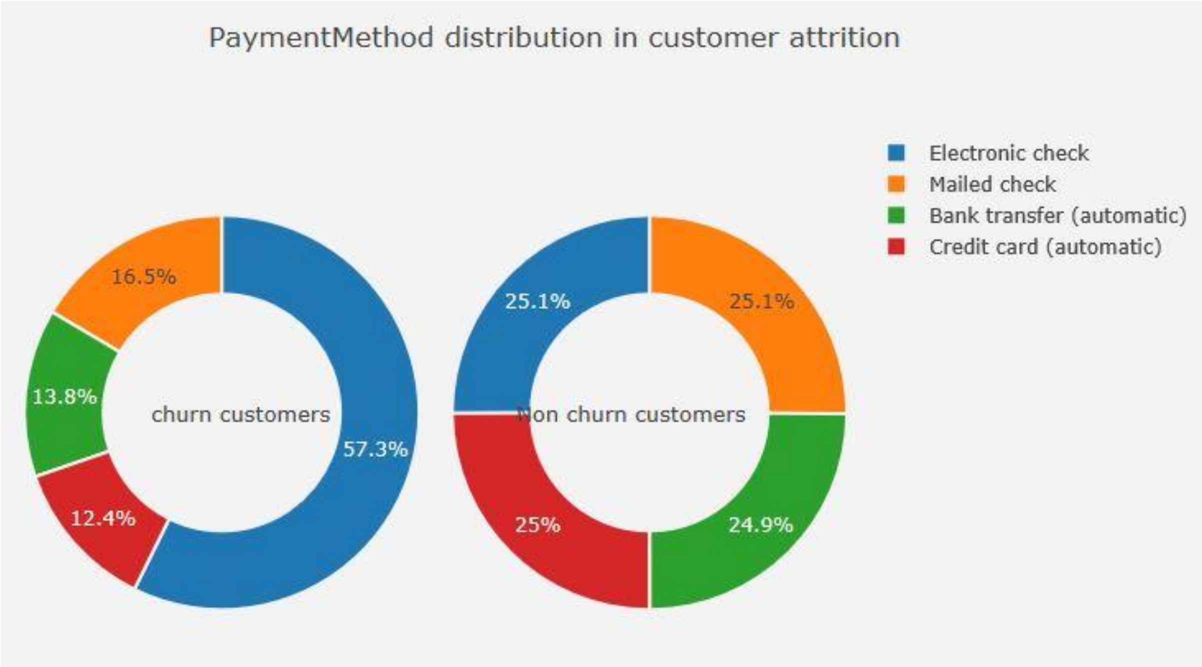


FIG: PAYMENTMETHOD DISTRIBUTION IN CUSTOMER ATTRITION

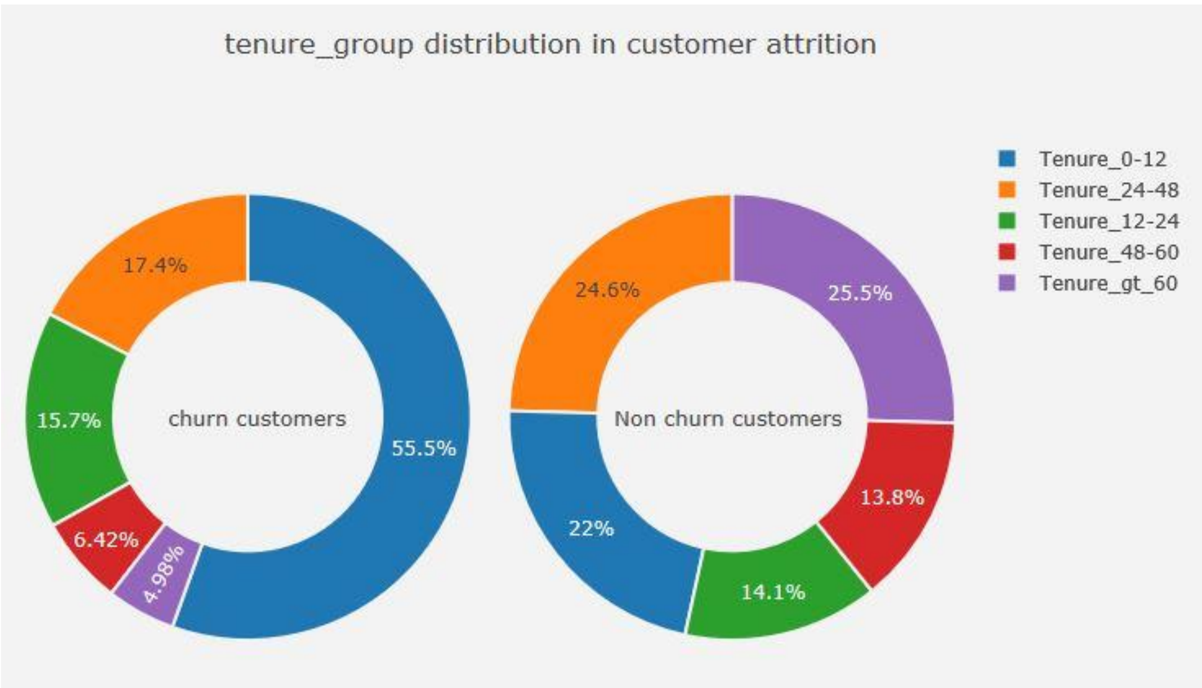


FIG: TENURE GROUP DISTRIBUTION IN CUSTOMER ATTRITION

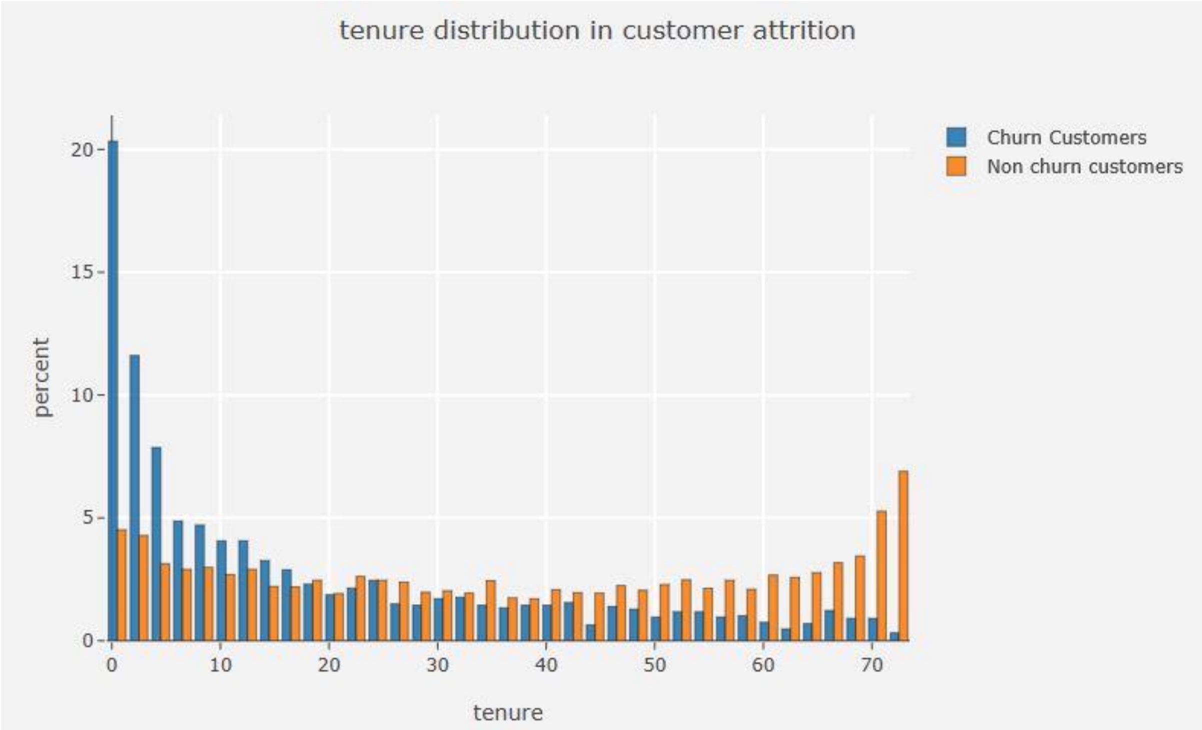


FIG: TENURE DISTRIBUTION IN CUSTOMER ATTRITION

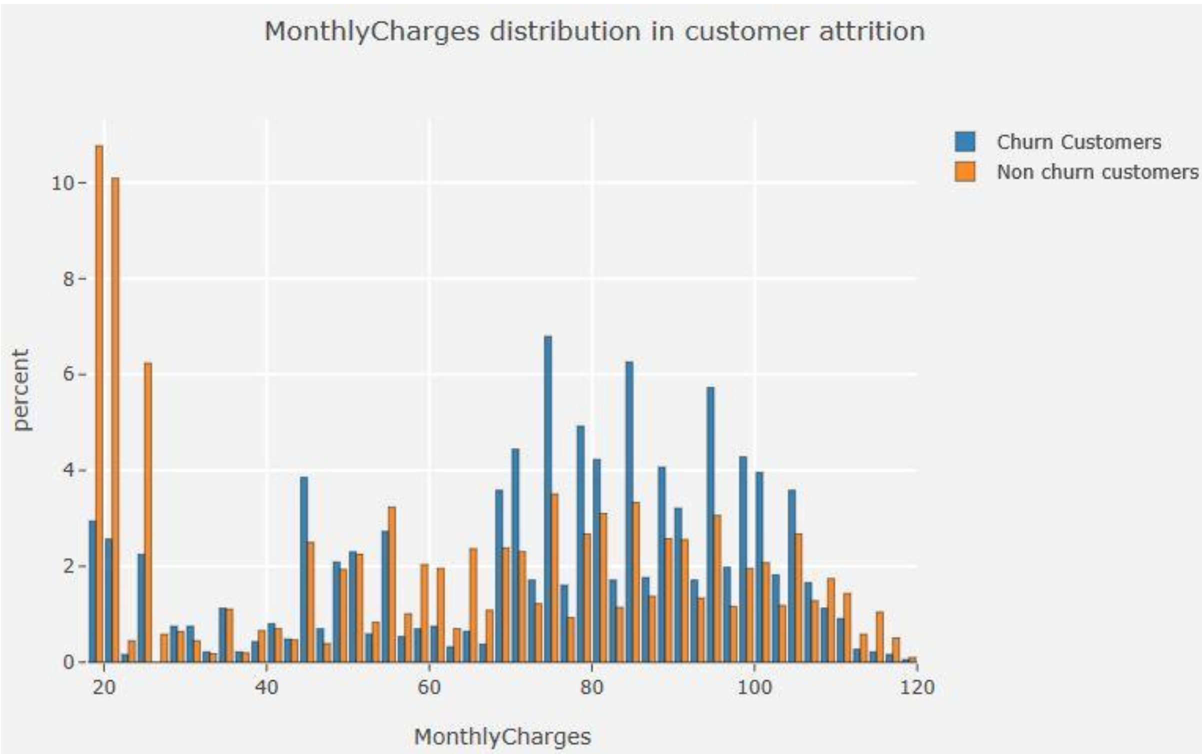


FIG: MONTHLY CHARGES DISTRIBUTION IN CUSTOMER ATTRITION

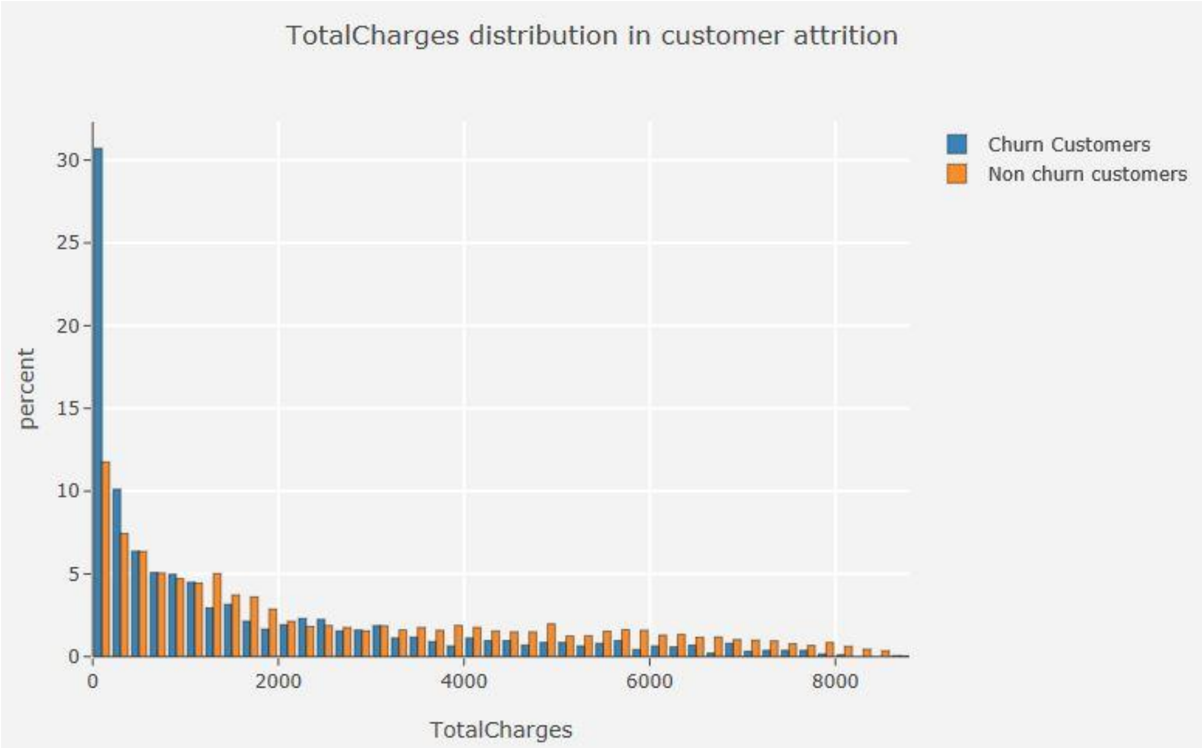


FIG: TOTAL CHARGES DISTRIBUTION IN CUSTOMER ATTRITION

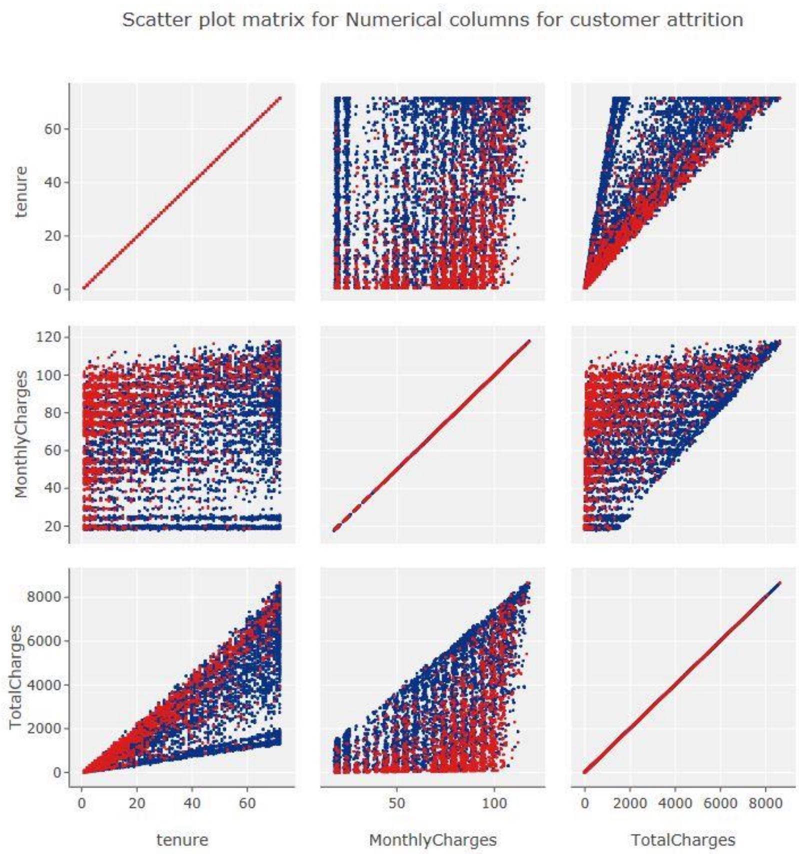


FIG: SCATTERPLOT MATRIX FOR NUMERICAL COLUMNS FOR CUSTOMER ATTRITION

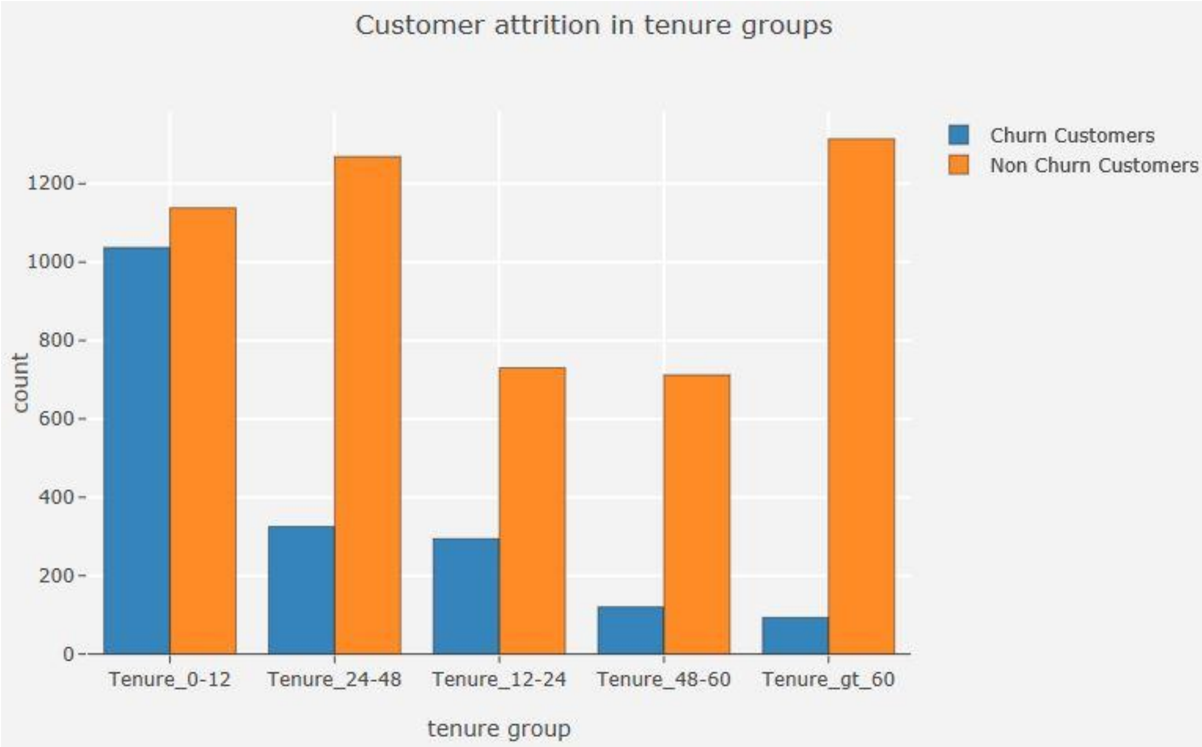


FIG: CUSTOMER ATTRITION IN TENURE GROUPS

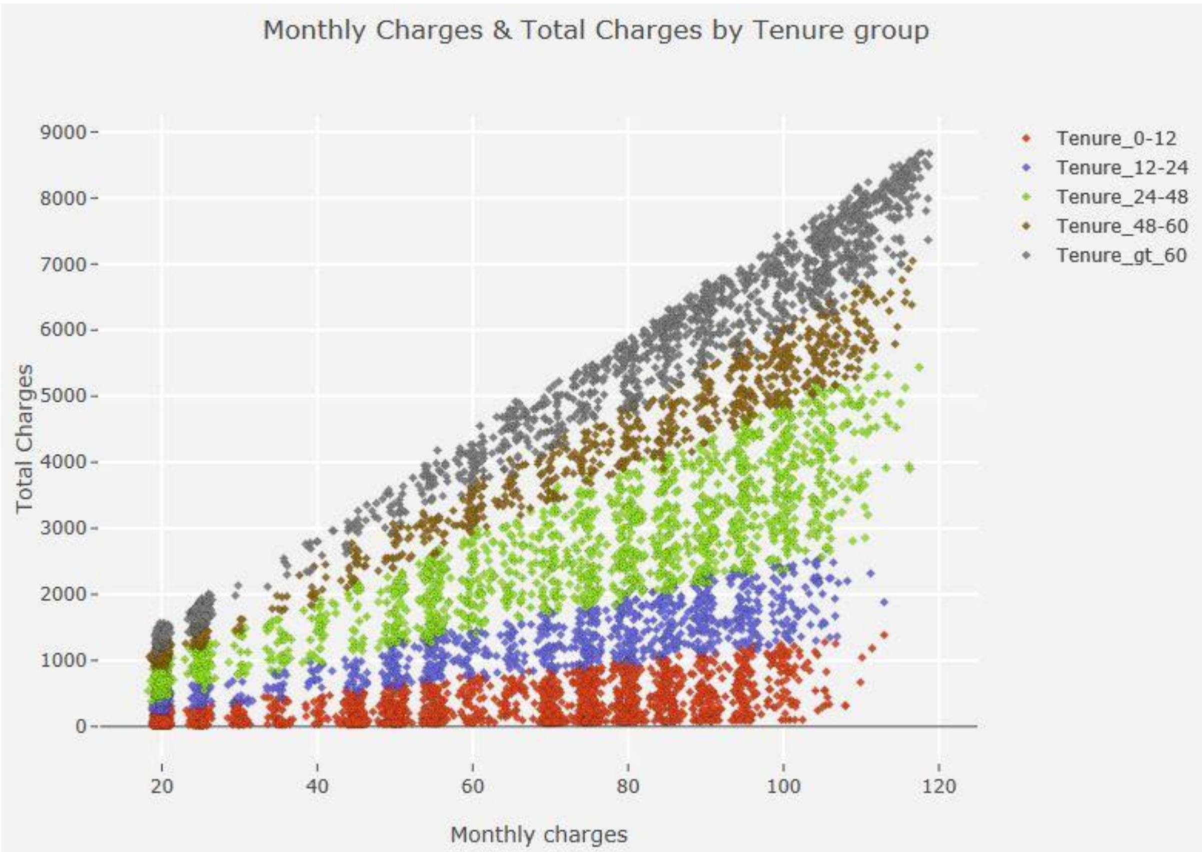


FIG: MONTHLY CHARGES AND TOTAL CHARGES BY TENURE GROUP

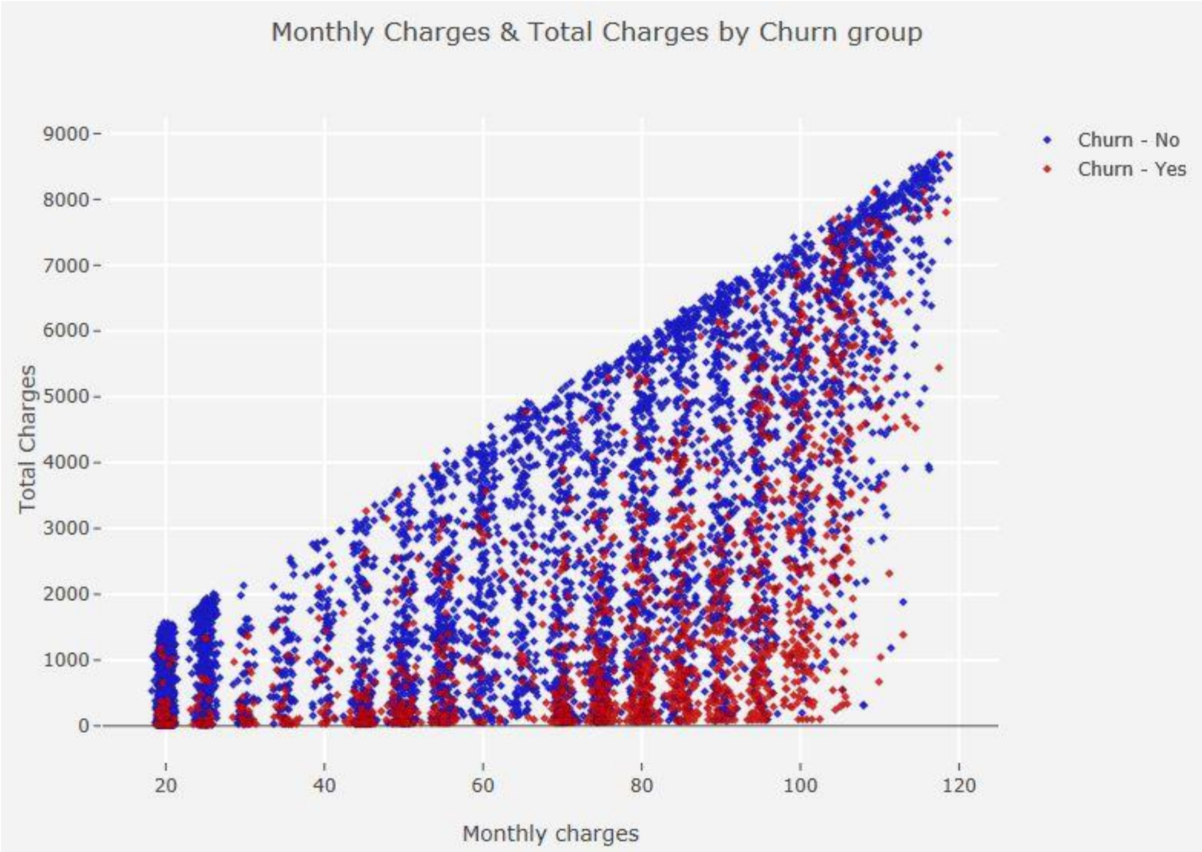


FIG: MONTHLY CHARGES AND TOTAL CHARGES BY CHURN GROUP

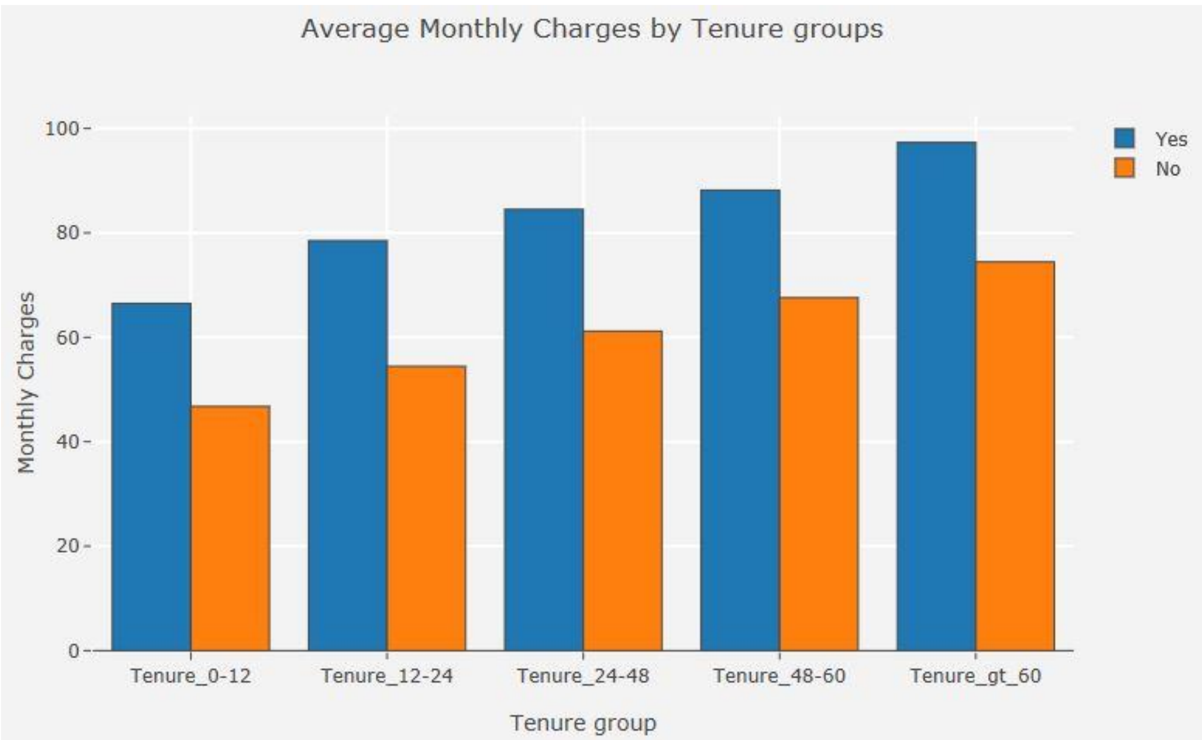


FIG: AVERAGE MONTHLY CHARGES BY TEURE GROUP

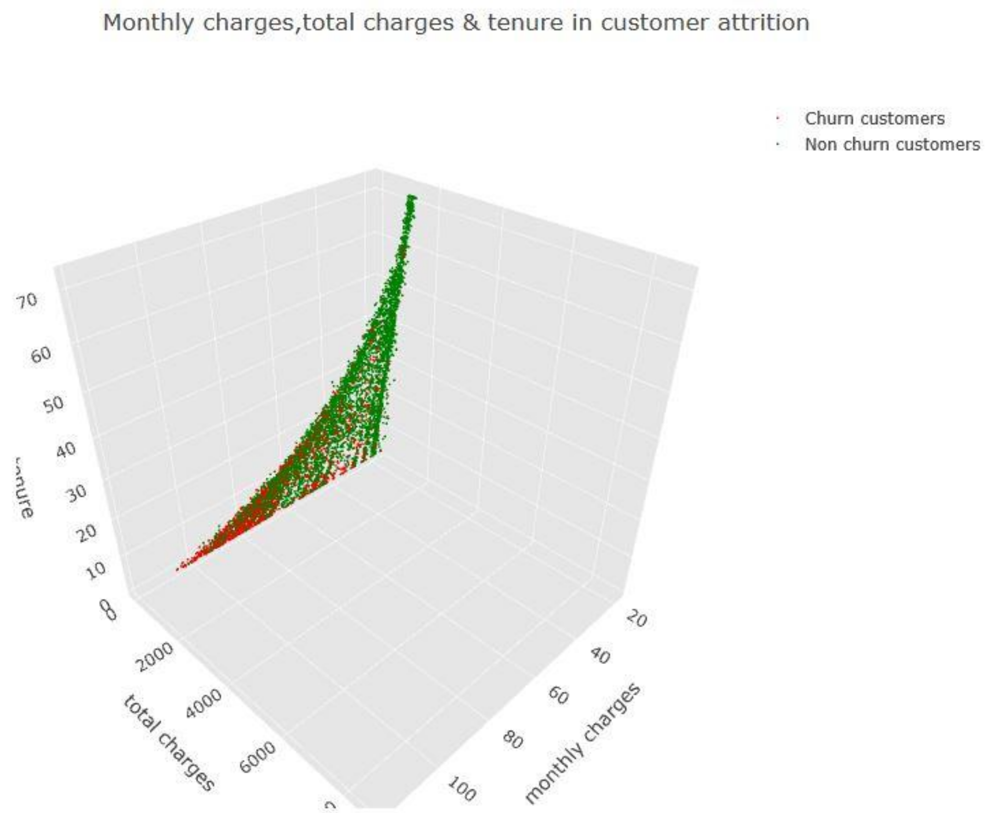


FIG: MONTHLY CHARGES, TOTAL CHARGES AND TENURE IN CUSTOMER ATTRITION

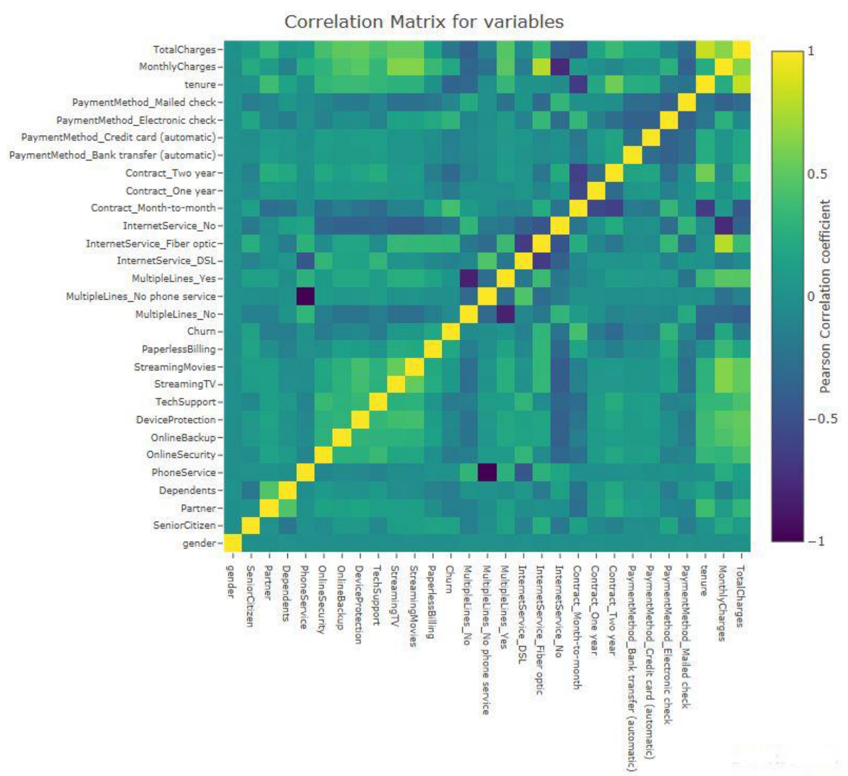


FIG: CORRELATION MATRIX

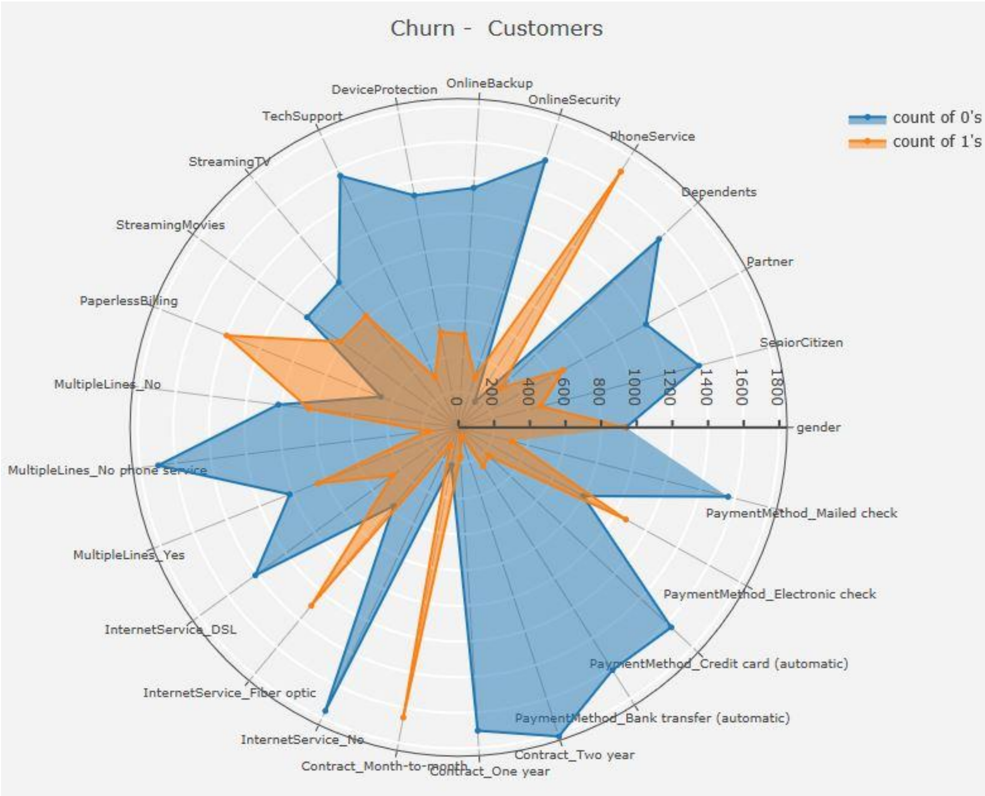


FIG: BINARY VARIABLE DISTRIBUTION IN CHURN CUSTOMERS

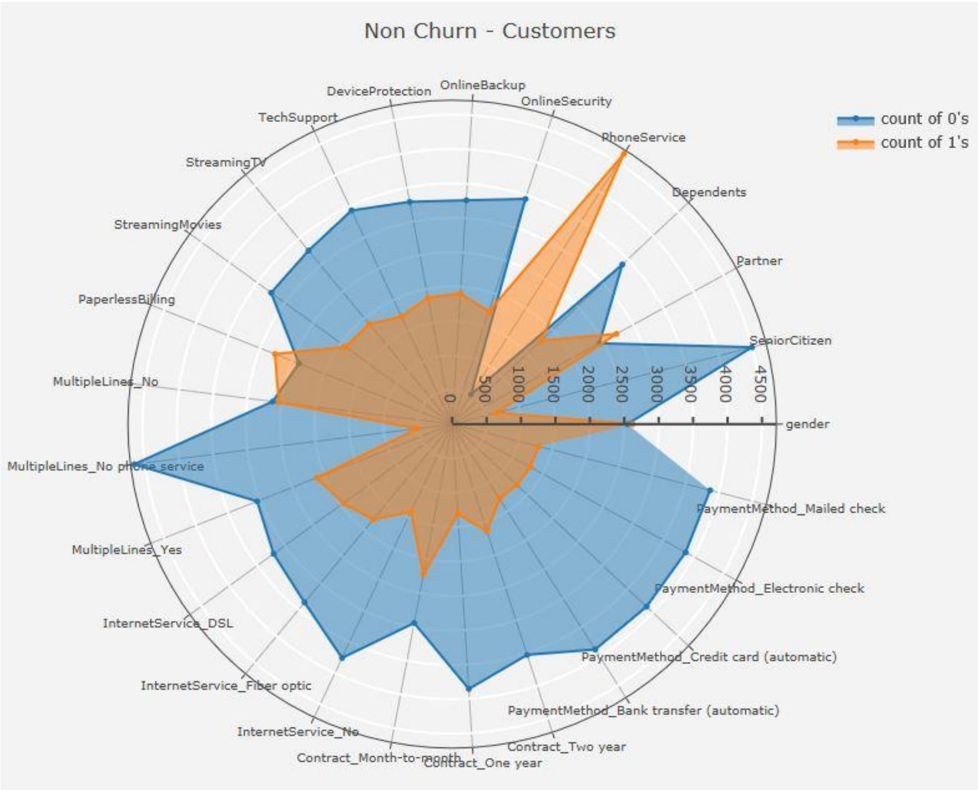
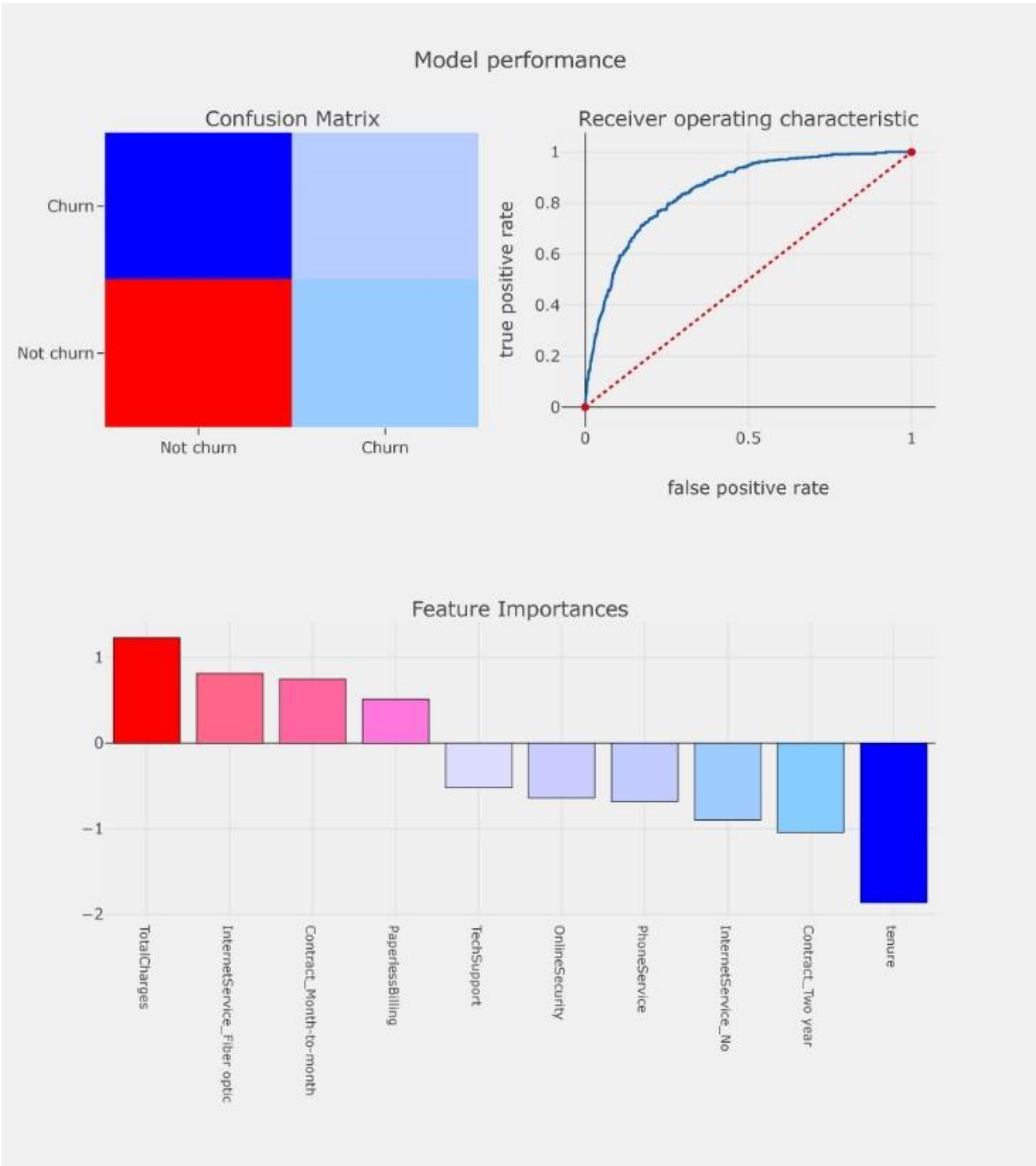
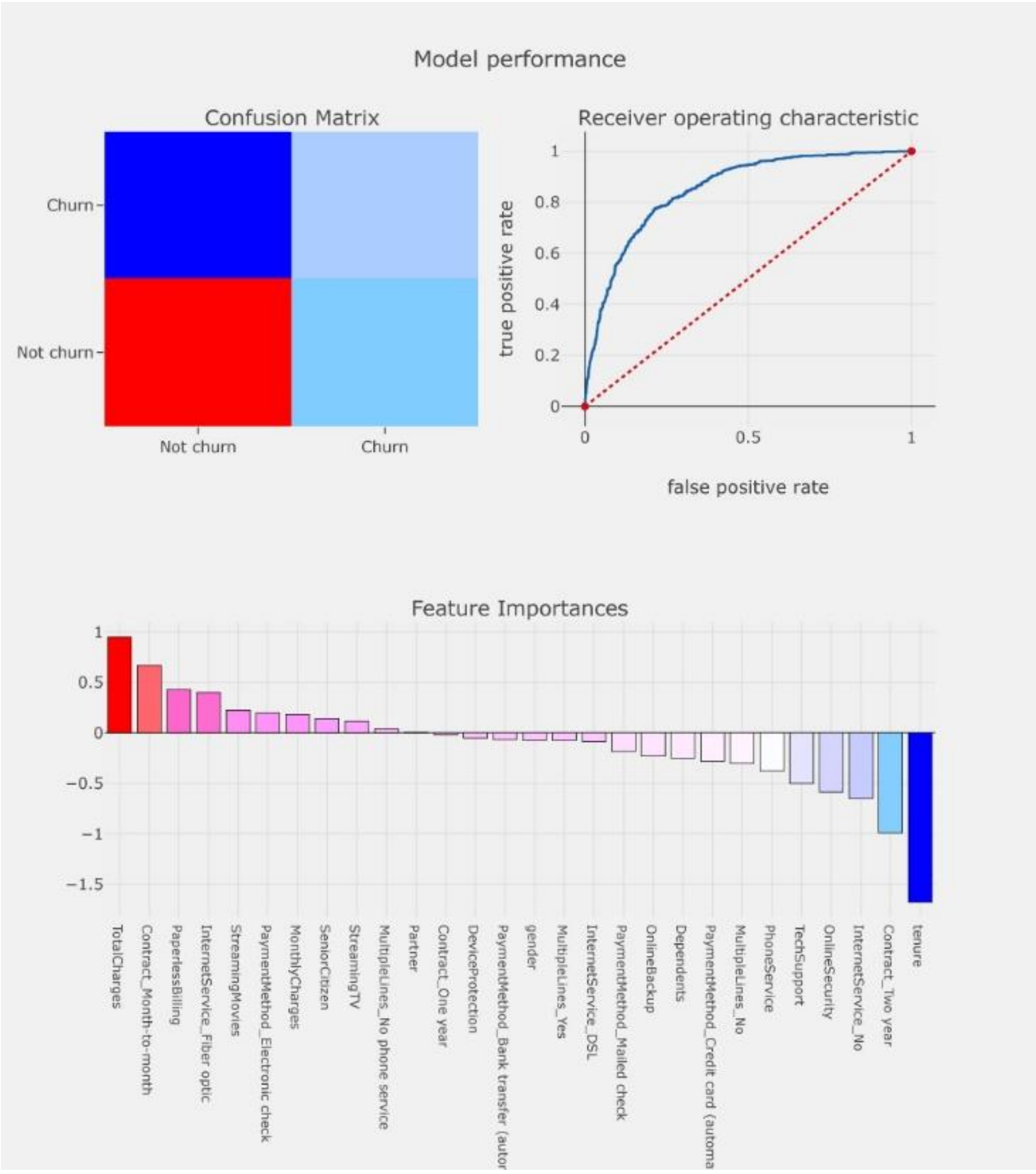


FIG: BINARY VARIABLE DISTRIBUTION IN NON-CHURN CUSTOMERS

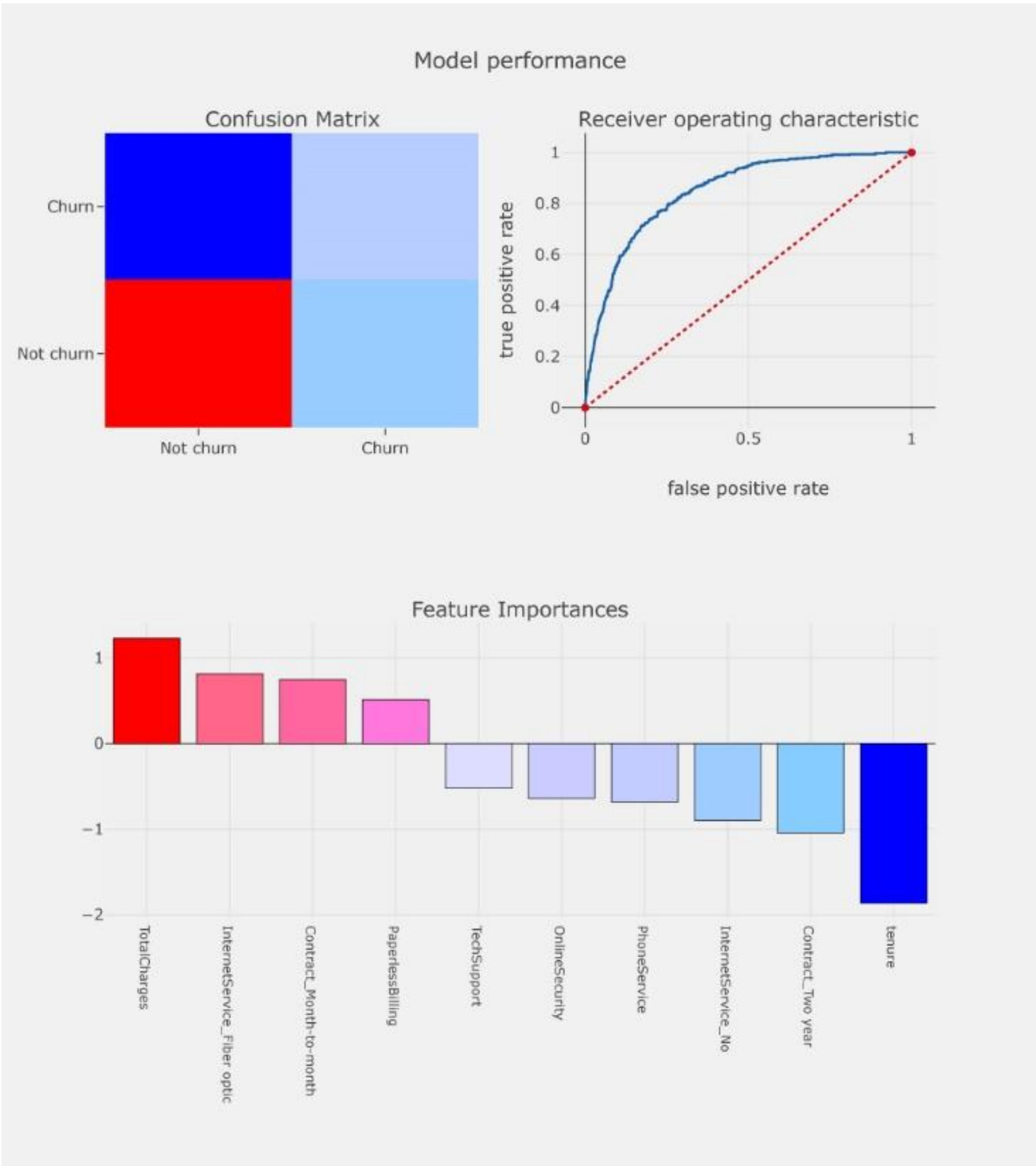
LOGISTIC REGRESSION BASELINE MODEL



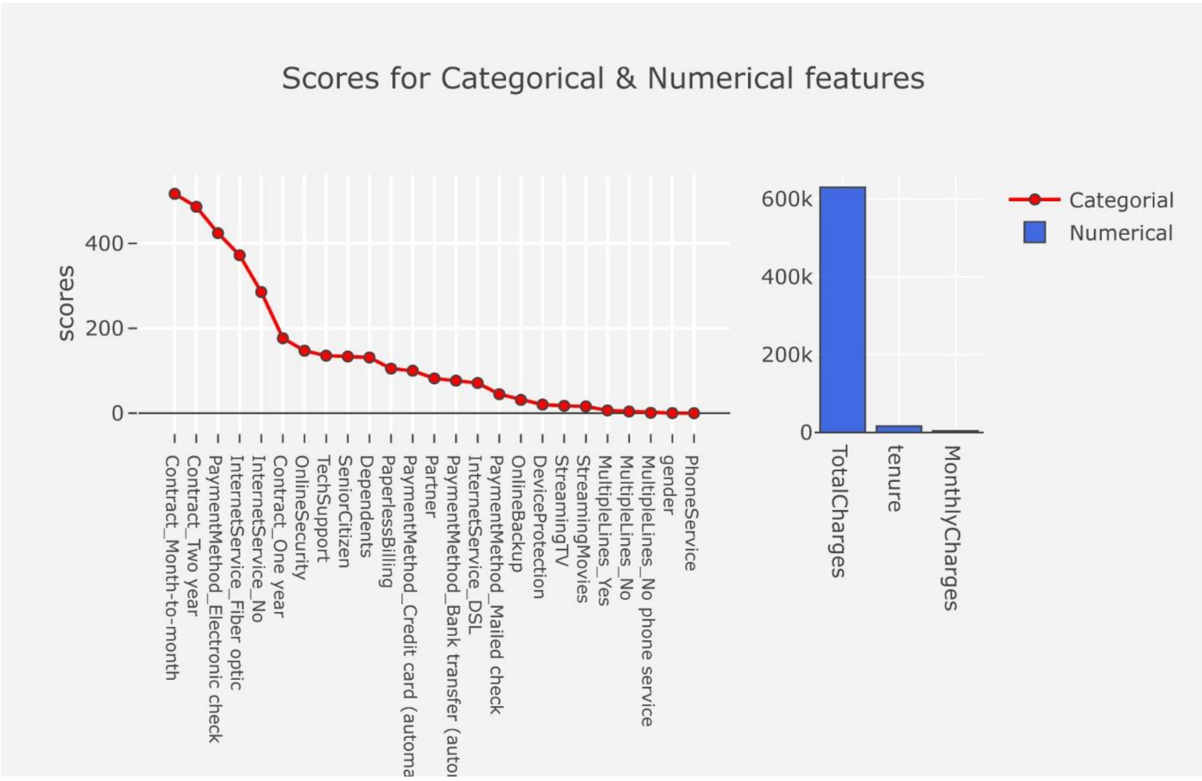
LOGISTIC REGRESSION USING SMOTE



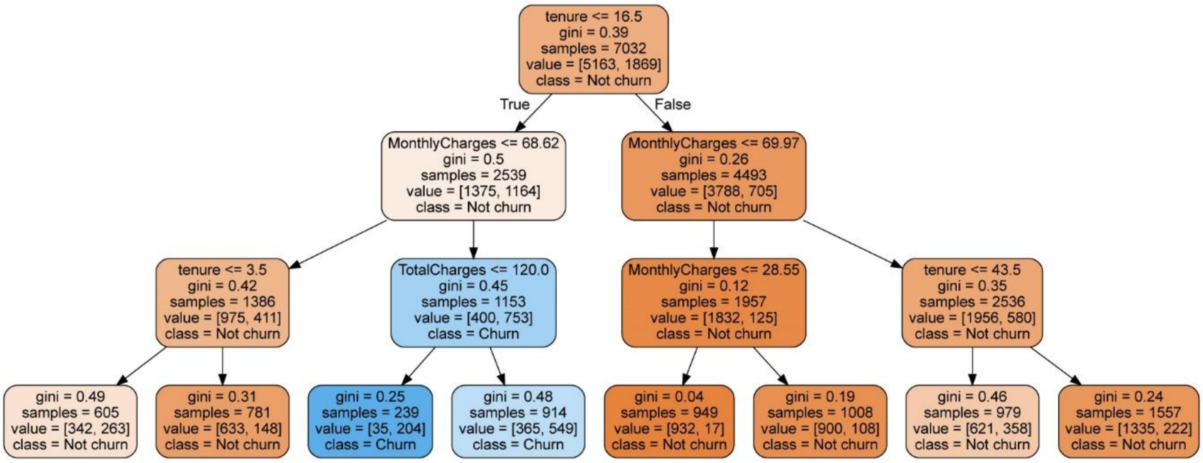
RECURSIVE FEATURE ELIMINATION MODEL

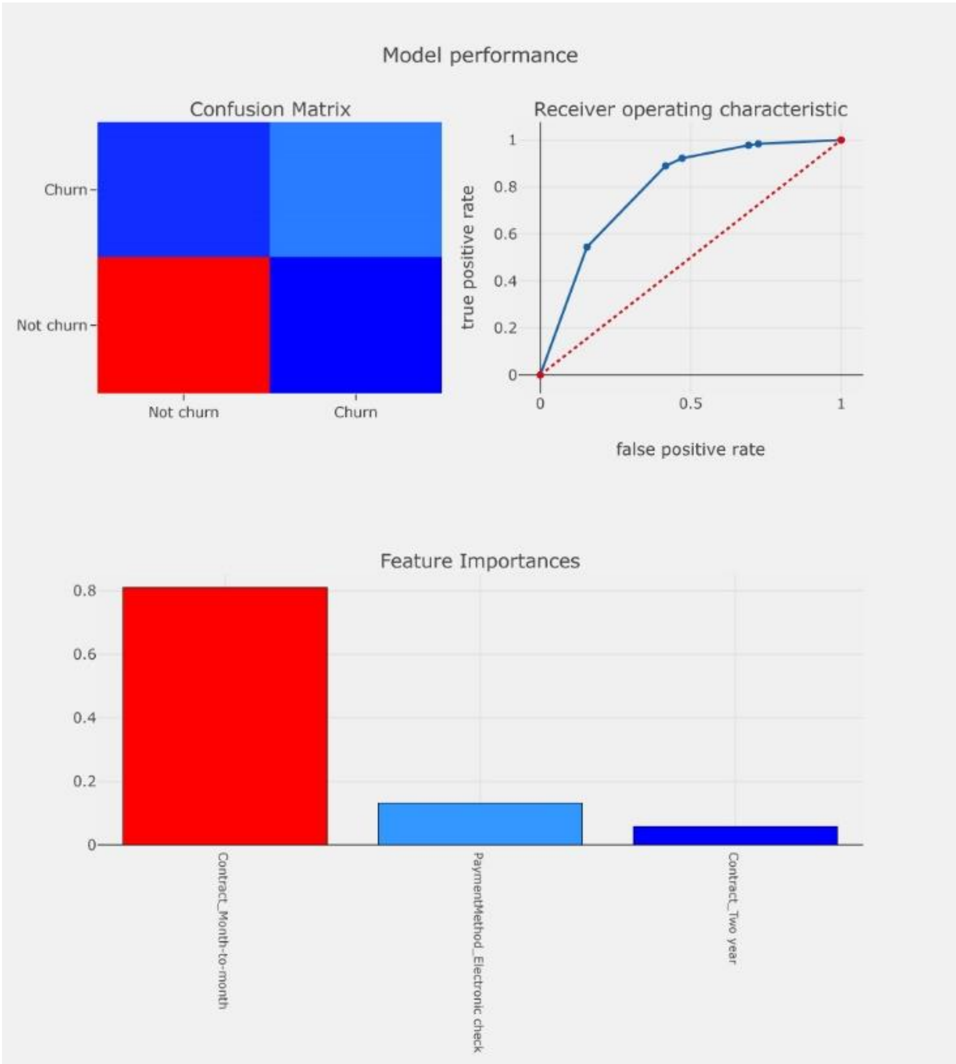


UNIVARIATE SELECTION MODEL



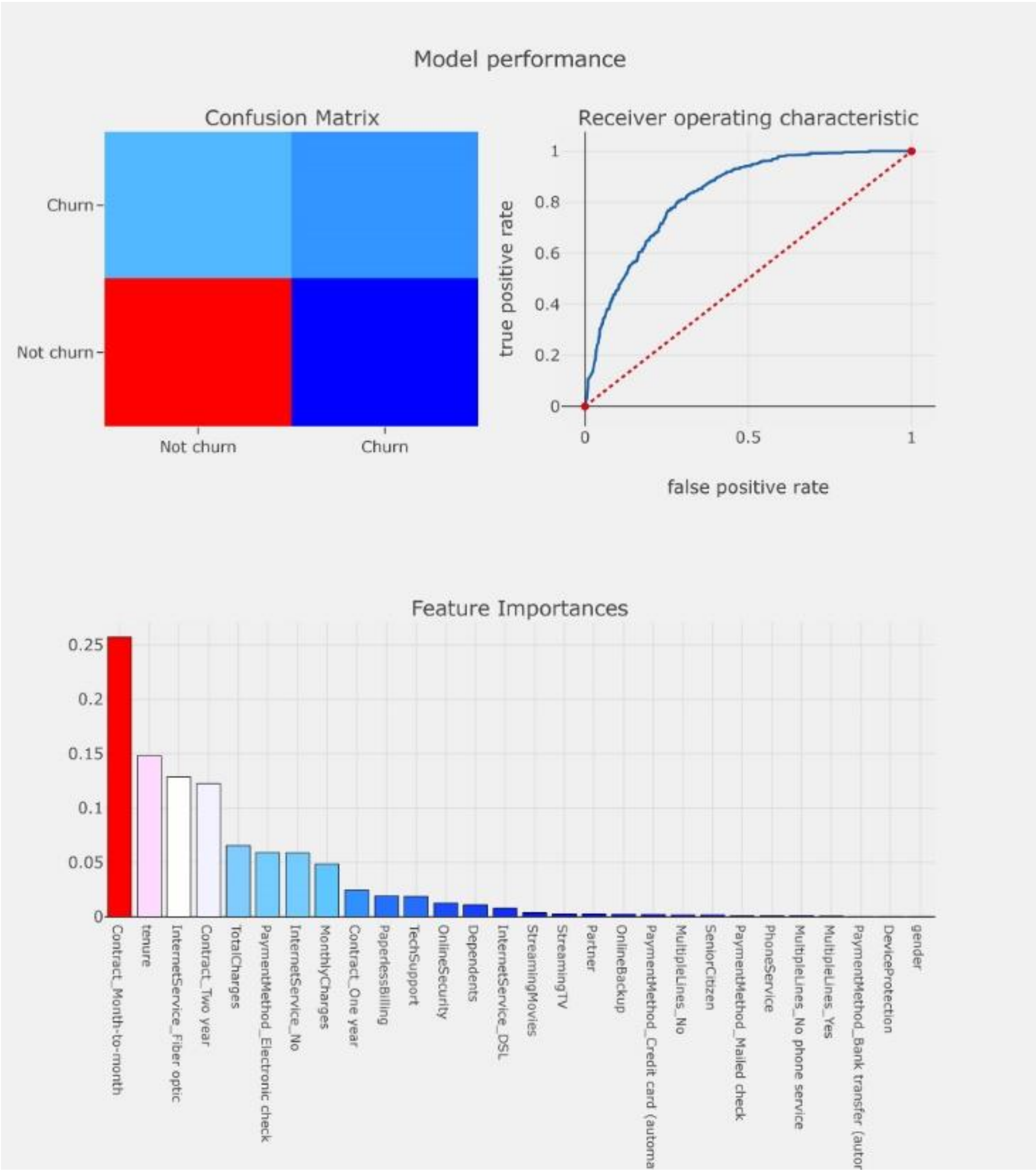
DECISION TREE



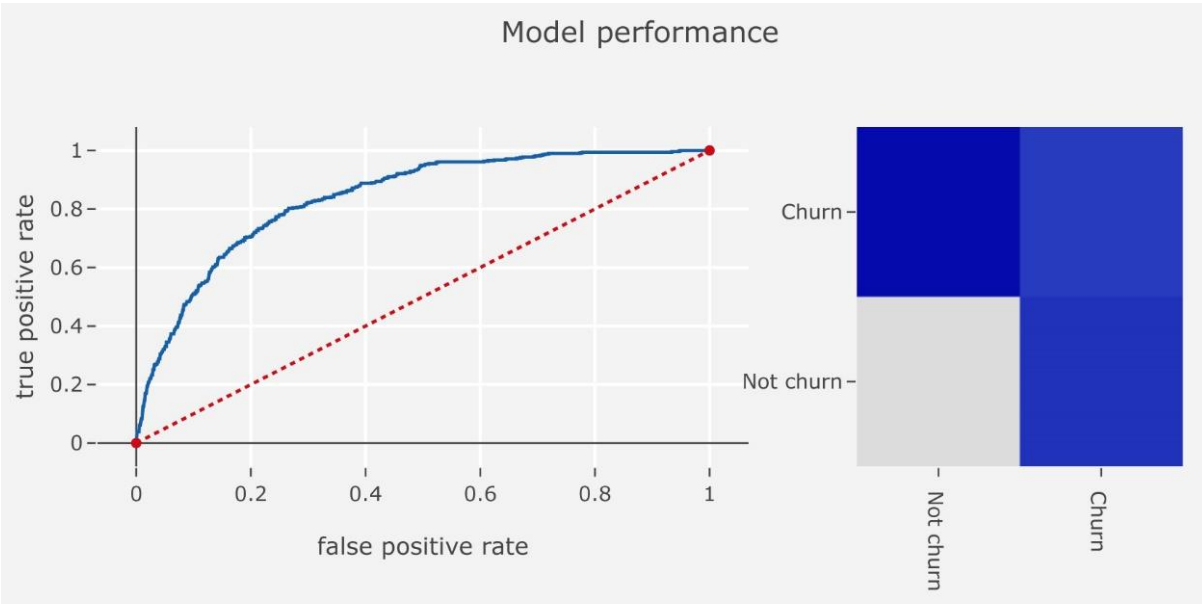


K-NEAREST NEIGHBOUR

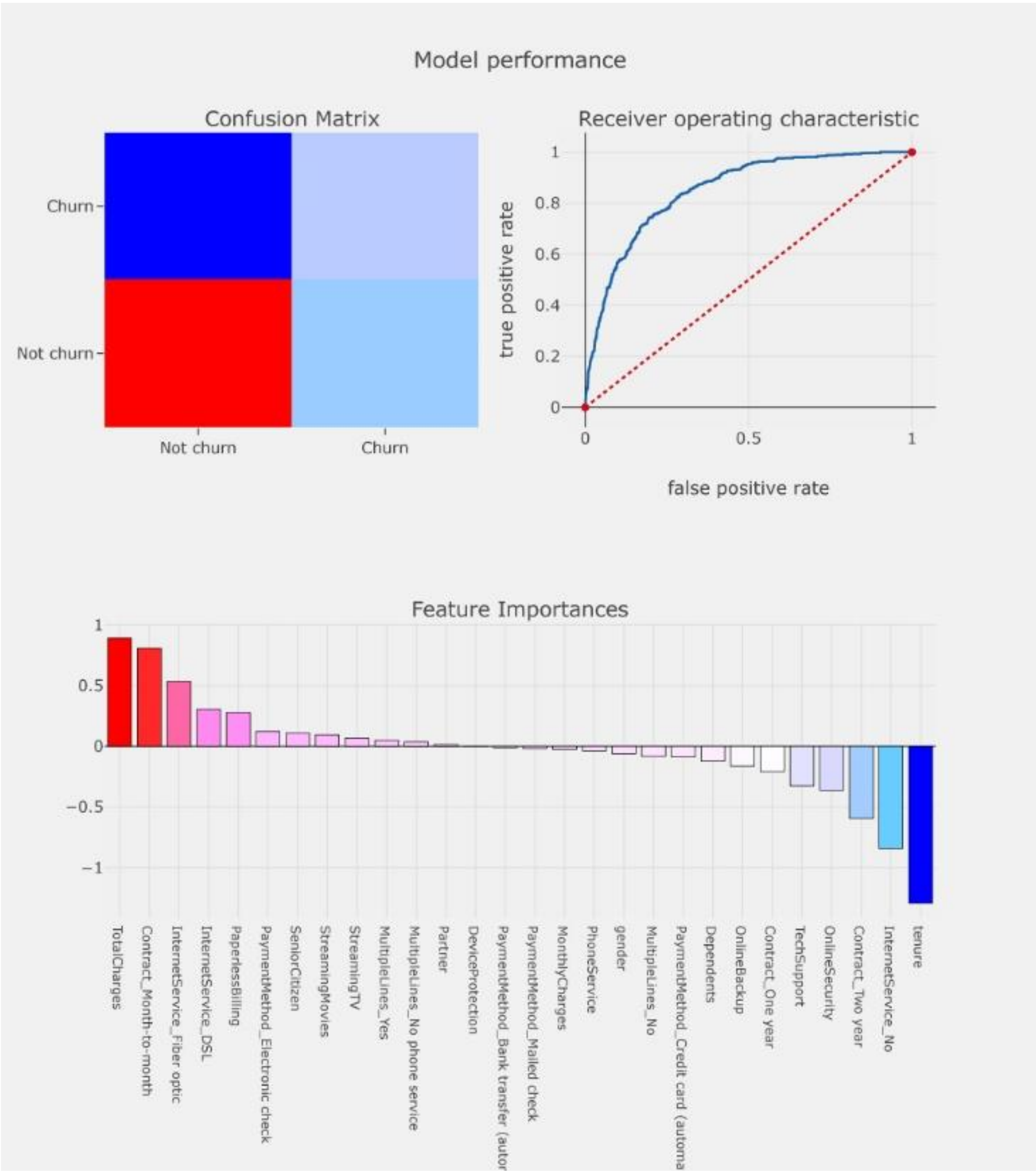
RANDOM FOREST CLASSIFIER



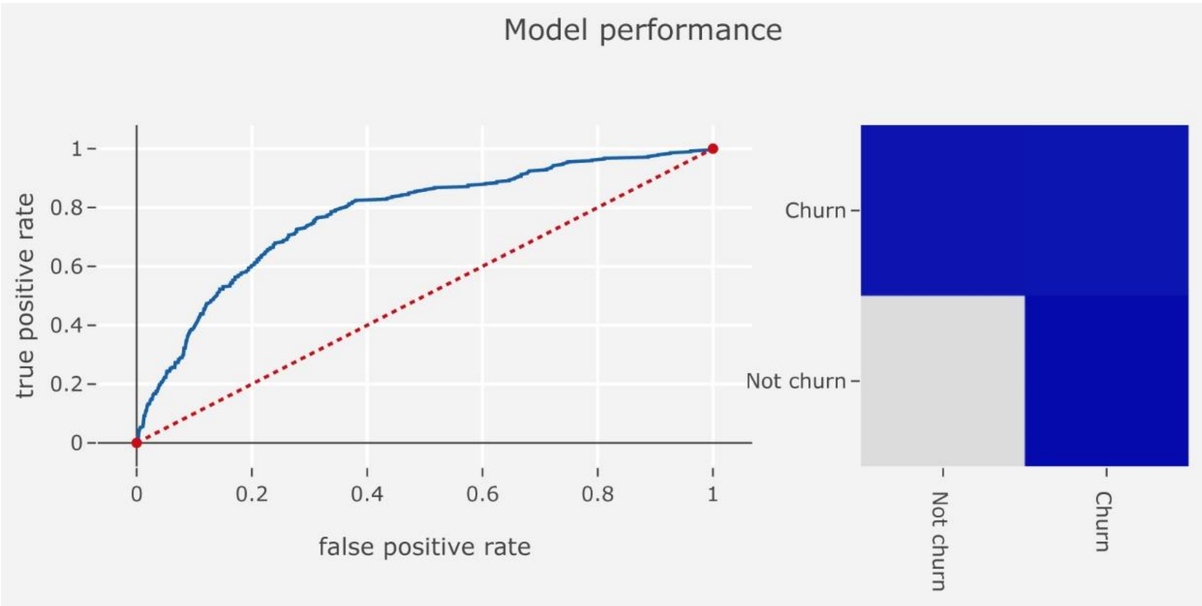
GAUSSIAN NAÏVE-BAYES MODEL



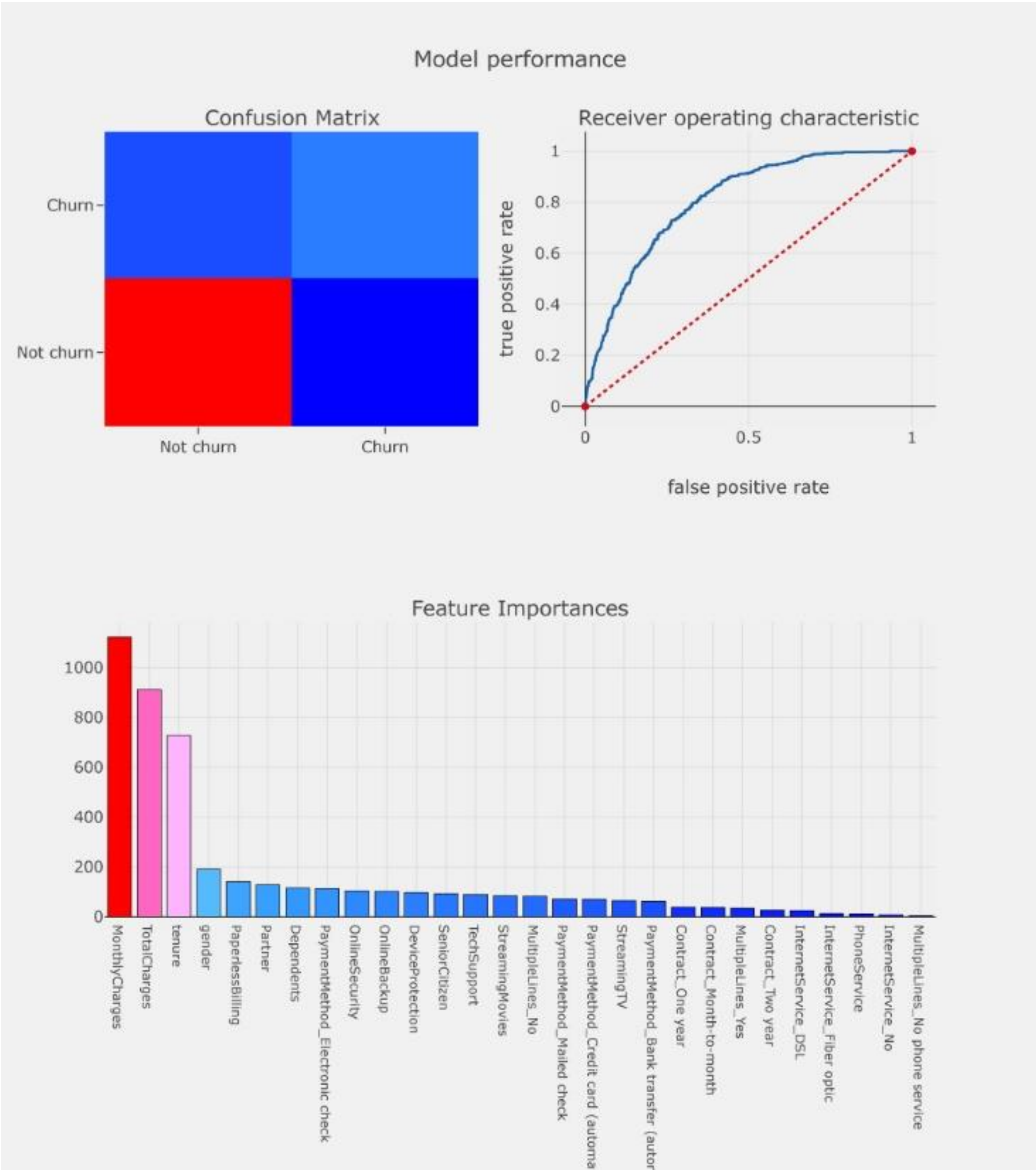
SUPPORT VECTOR MACHINE MODEL



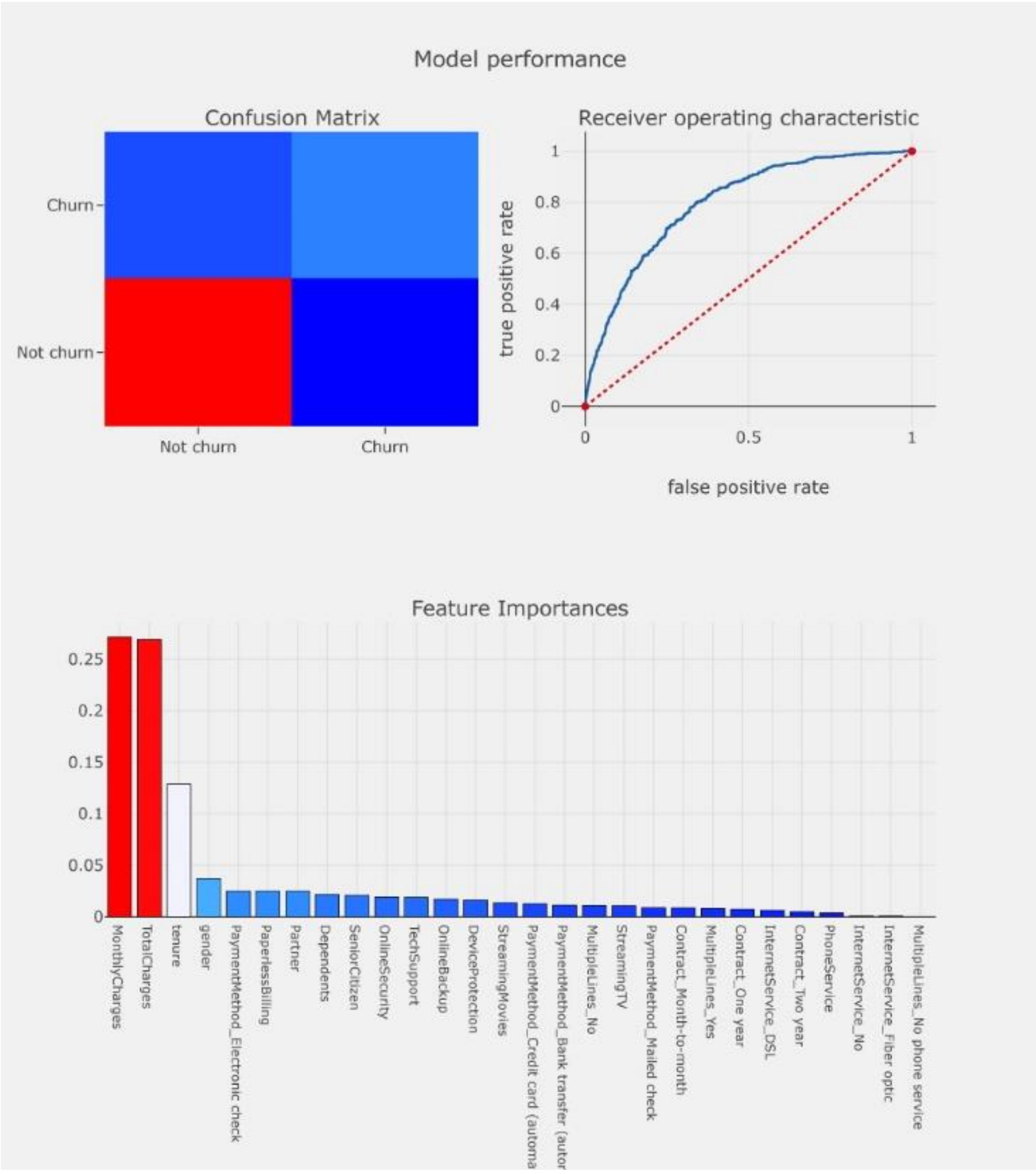
SUPPORT VECTOR MACHINE MODEL AFTER TUNING PARAMETERS



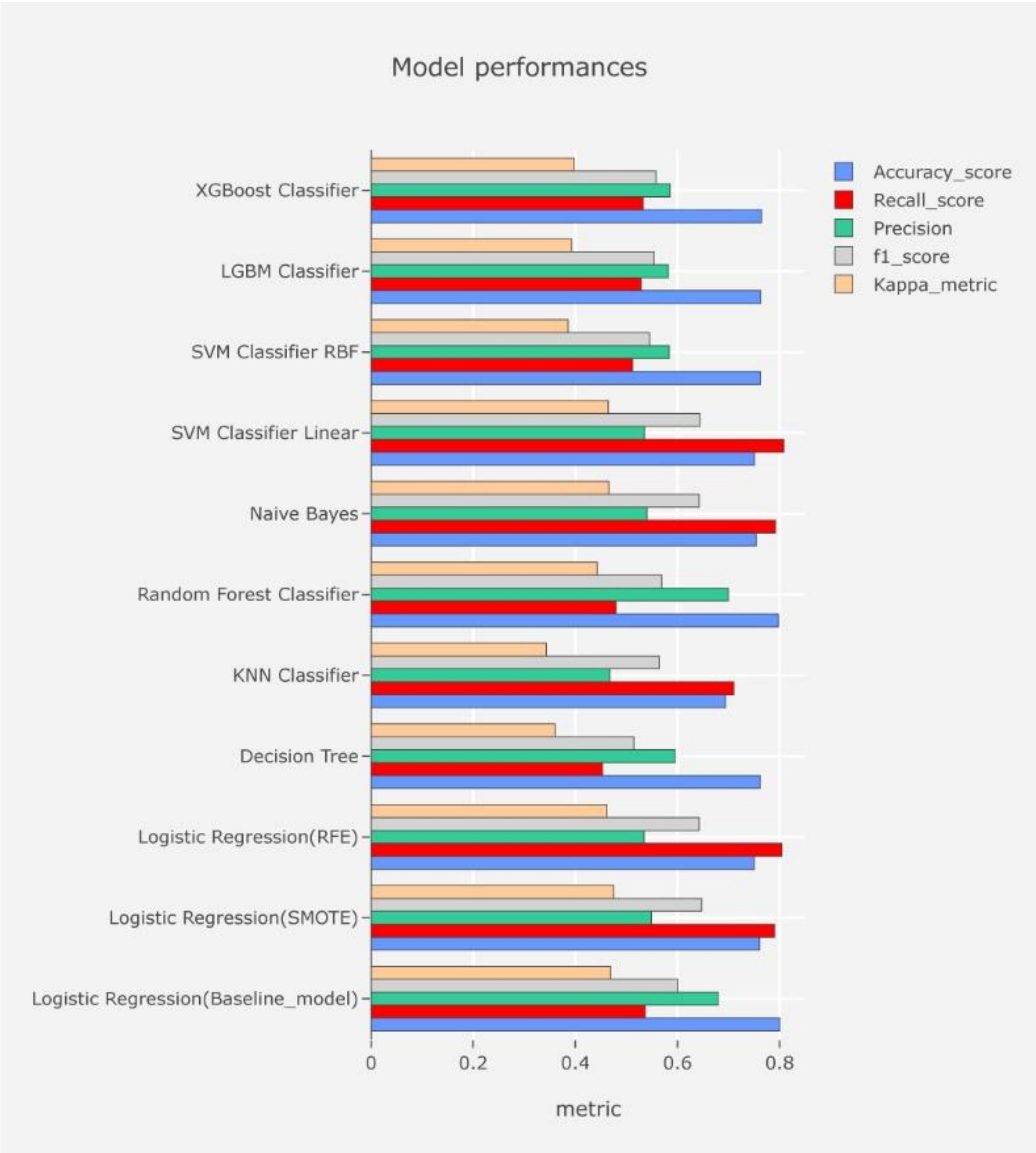
LIGHTGBM CLASSIFIER MODEL



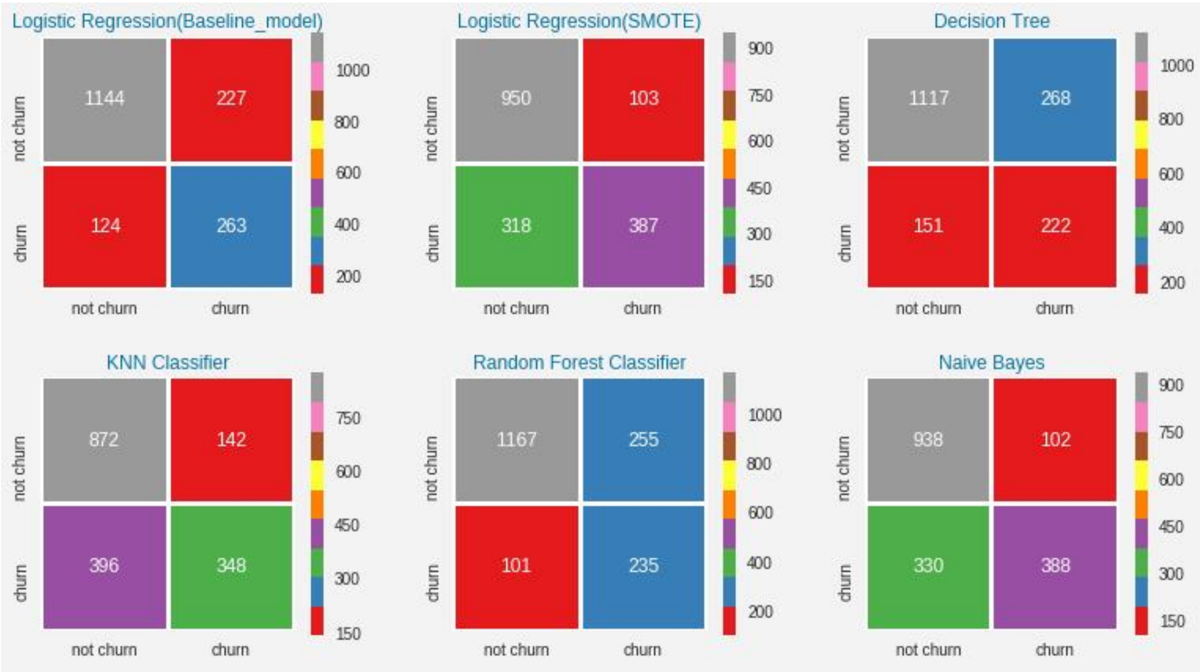
XGBOOST CLASSIFIER MODEL

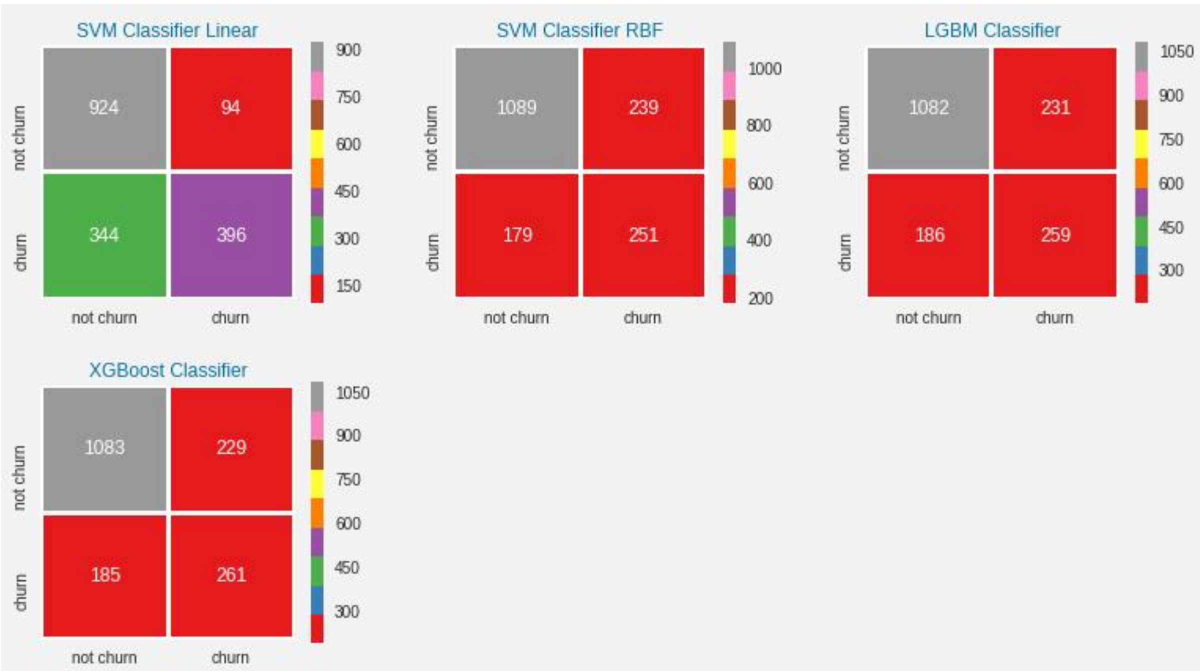


MODEL PERFORMANCES

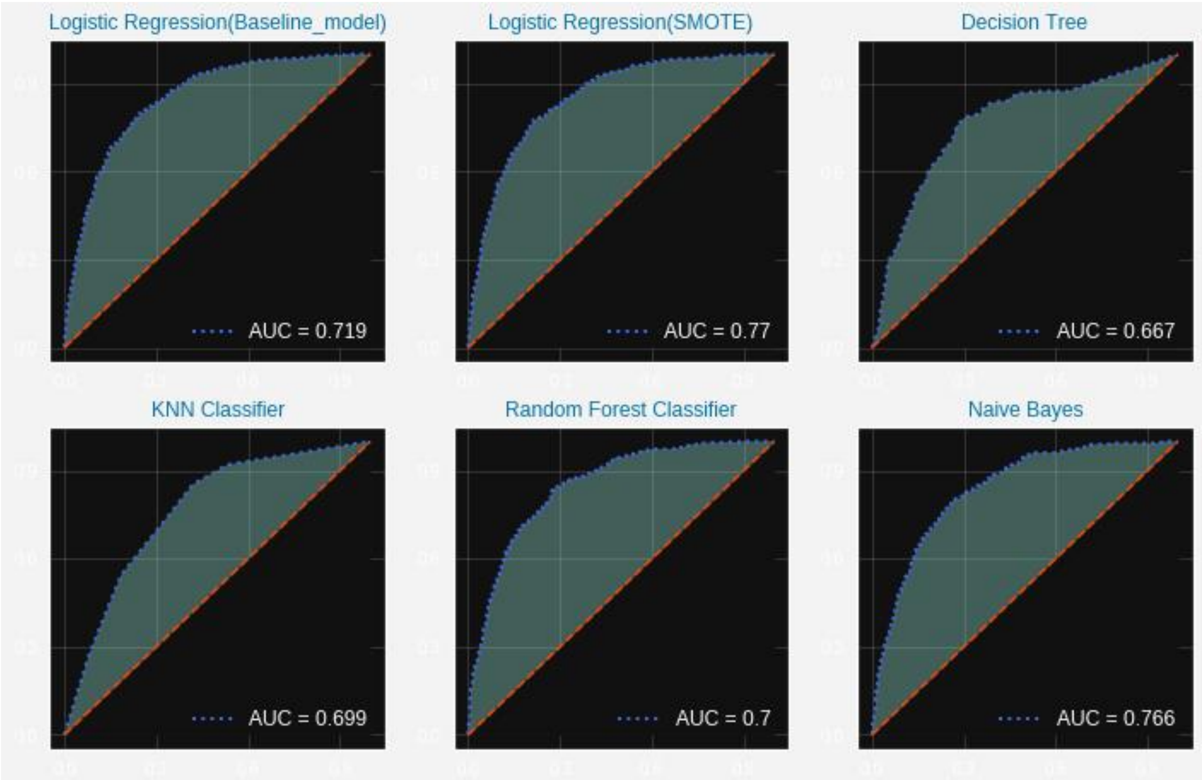


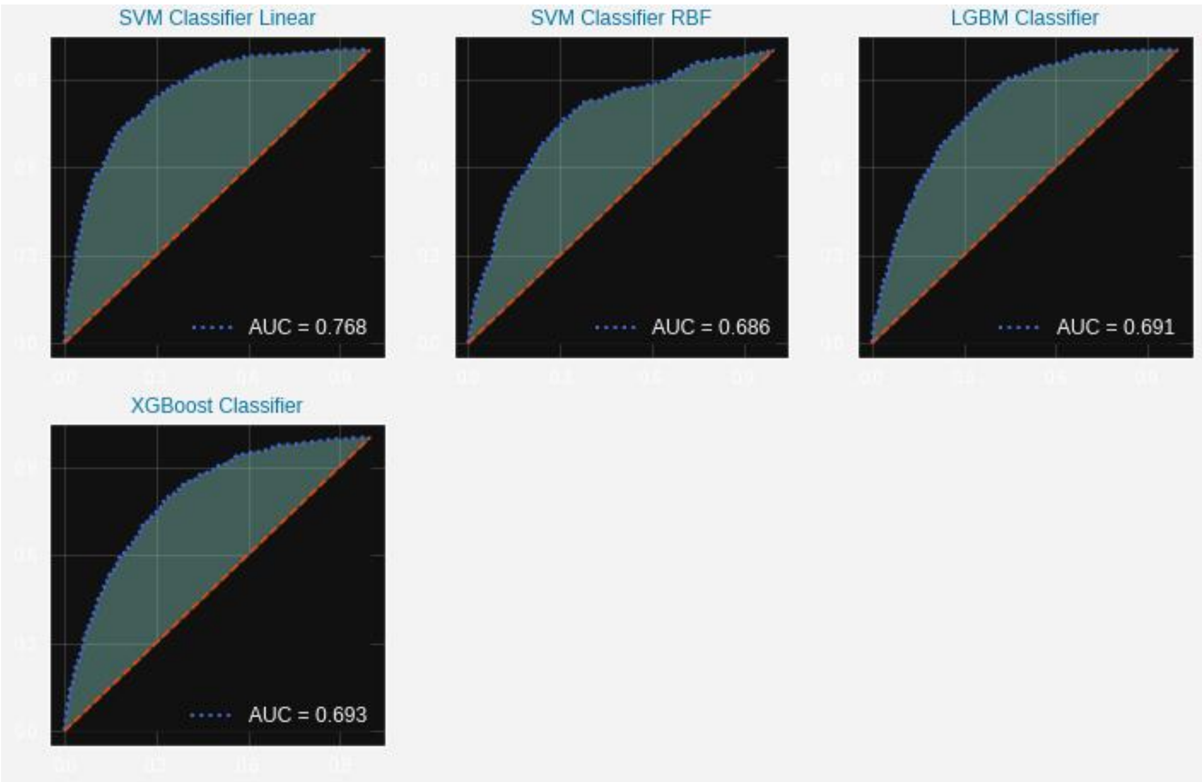
CONFUSION MATRICES





RECEIVER OPERATING CHARACTERISTIC CURVES





OBSERVATIONS

We can see that some variables have a negative relation to our predicted variable (Churn), while some have positive relation. Negative relation means that likeliness of churn decreases with that variable. Let us summarize some of the interesting features below:

- Having DSL internet service also reduces the probability of Churn

Lastly, total charges, monthly contracts, fibre optic internet services and seniority can lead to higher churn rates. This is interesting because although fibre optic services are faster, customers are likely to churn because of it.

FUTURE SCOPE OF WORK

- Use of unsupervised learning techniques such as cluster analysis.
- We consider two prominent approaches for our study, namely, Neural Networks (NN) and Bayesian Networks (BN). We will formulate the CAC process as a classification problem.

- The Neural Network model will consist of 15 input neurons (3 input variables of 5 states each), 3 neurons in the hidden layer and 5 neurons in the output layer.