

Telco - Customer Churn

10.02.2018

Annapoorani

Springboard

Capstone project proposal



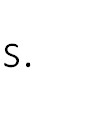
# Problem Statement

Predict behavior to retain customers. To can analyze all relevant customer data and develop focused customer retention programs.

# Outcomes

Telco - a Telecommunication company will be beneficial by this project. This will help them in the following ways

* Improve user experience with their products
* Identifying important features that are key for customer retention

● To come up with new Marketing Strategies.

# Data

Telco data collected from Kaggle.

https://www.kaggle.com/blastchar/telco-customer-churn

# Approach

The Steps to solve the identified problems are as below

* Data Wrangling using Python Pandas, so that data is cleaned and readily available for analysis
* EDA to discover the underlying facts about the data and to study the different features
* Applying machine learning algorithms like Logistic regression, Decision Trees and Deep Learning.

# Data Wrangling

Success of analysis depends upon how the data is cleaned up as usable for analysis with columns as separate features and each row as single observation. As it is highlighted always as Data Wrangling is time consuming, wrangling for this project also was very challenging.

As this data is downloaded from Kaggle it is already cleaned. Two main steps performed is

1. Converting the Total and Monthly charges to numeric
2. Checked for any NA values in the columns, identified very few rows with Total Charges as NA and those rows are dropped

# Data Pre-Processing

Most features of the Telco – Customer Churn Dataset are categorical variables. One hot encoding is performed on those categories and redundant columns are removed.

One major column of interest is the Internet Service, if the column is ‘No’ then obviously the concern customer won’t have Streaming TV or Streaming Movies.

Henceforth “Internet\_Service\_No” column is retained by dropping Streaming\_TV\_No and Streaming\_Movies\_No columns.

# Exploratory Data Analysis

EDA helps in Visualizing underlying insights from the data. It helps in Feature Engineering also paves way for identifying new scope of data. As churn is the key field of study, first part of the notebook deals with the picturization of churn data based on different features.

## Churn Based on Gender

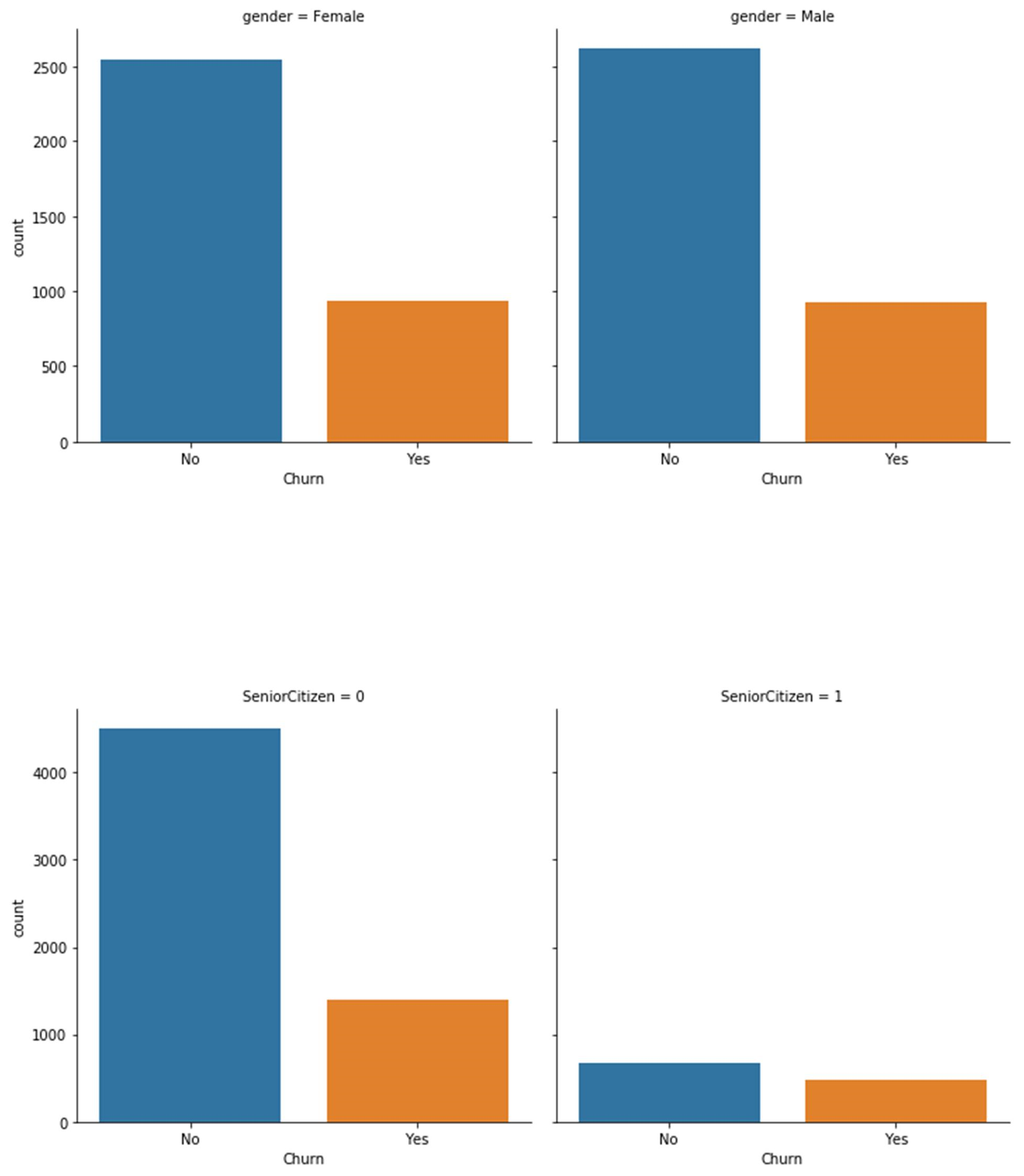
As from above chart churning

rate looks similar for both male and female so there is

no

significant difference in churning rate based on gender.

Churn Behavior with Senior Citizen



It seems Senior citizens churned and not churned numbers have very less difference, also the number of Total Senior Citizens is less compared to remaining population.

## Partners Churn

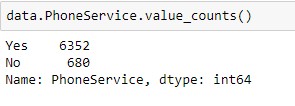
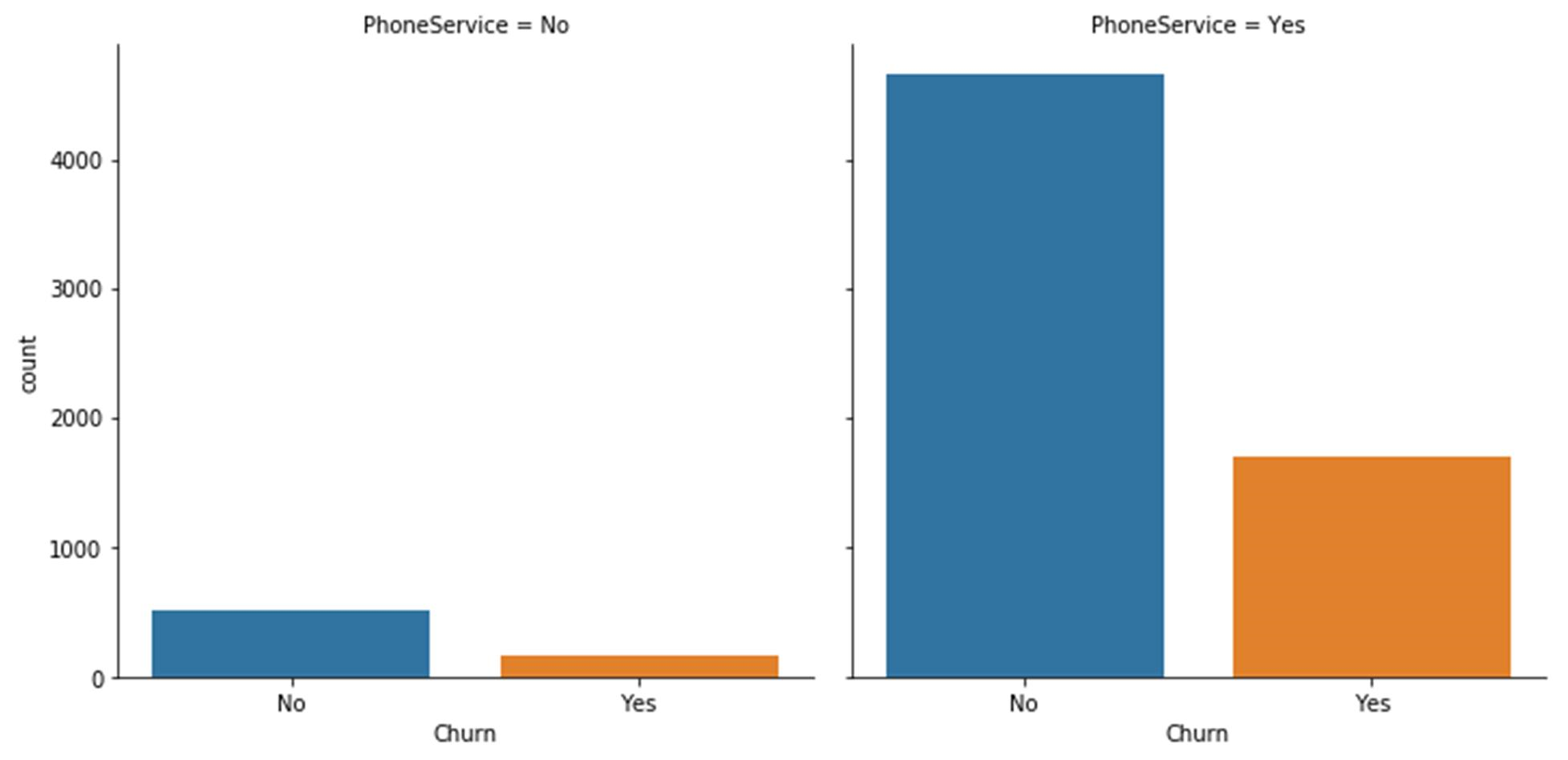
If there is no partner then churn rate is little higher.

Churn Numbers for Dependents



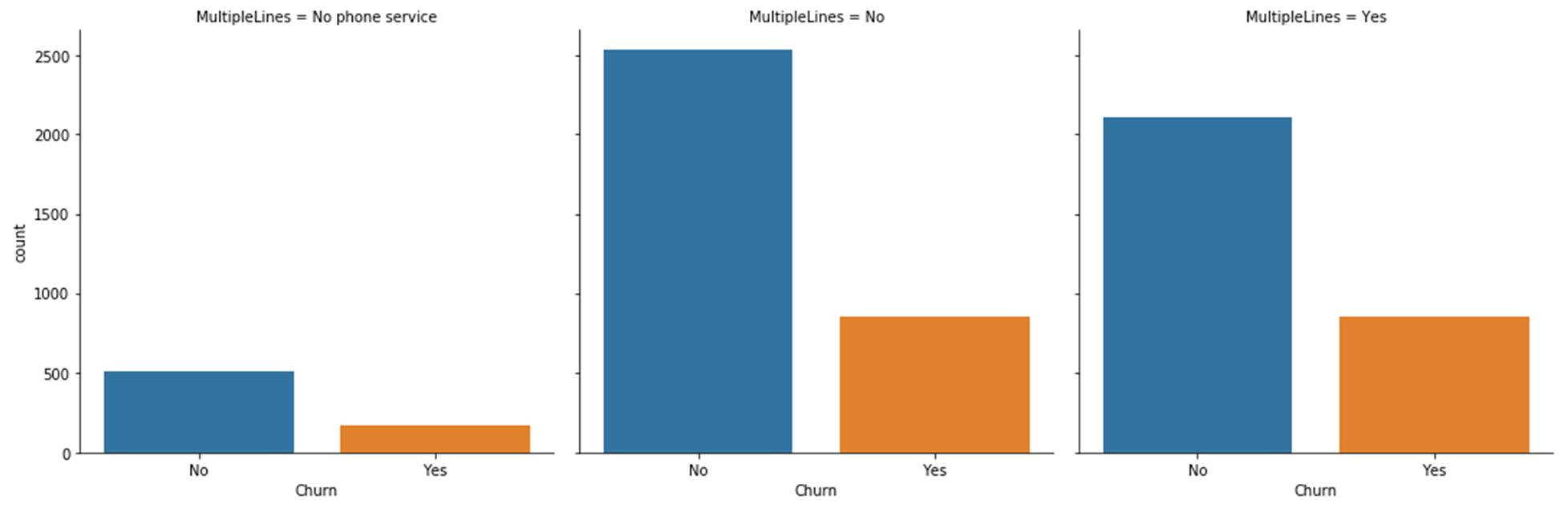
If there are no Dependents then Churning numbers are higher

## Customer Churn with Phone Service



Most of the Telcom customers have Phone service too, nearly 10% of them Phone Service customers are churned

## Churn based on Multiple Lines



Irrespective of customer have multiple lines or not the Churning numbers are same.

## Churn based on Device Protection

Customer without device protection churned higher

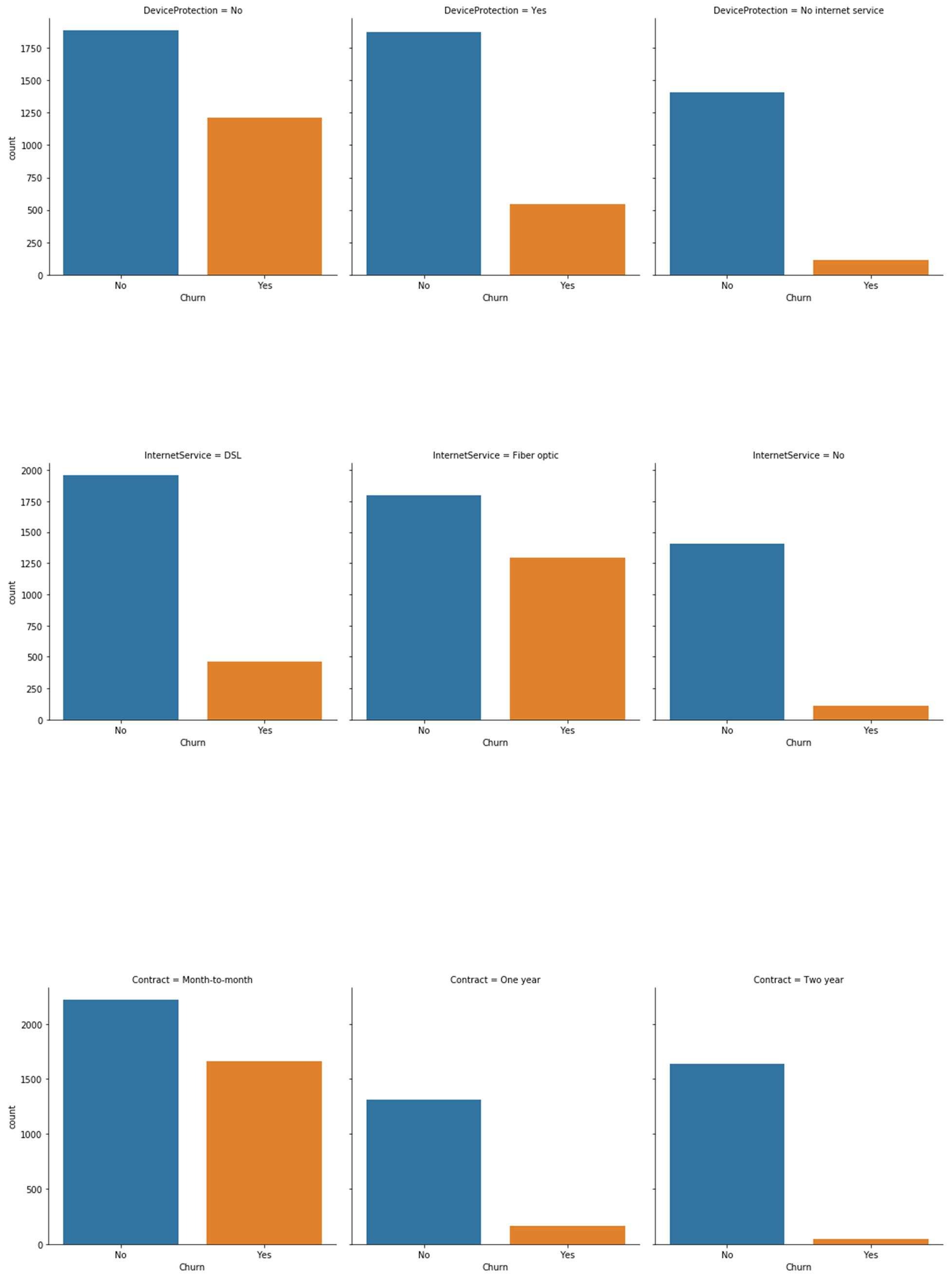
Churn based on Customer option on Internet service

Customers getting Internet Service by Fiber optics Churned more, so definetly that may be a

important feature.

Churn based on Contr

act Pattern

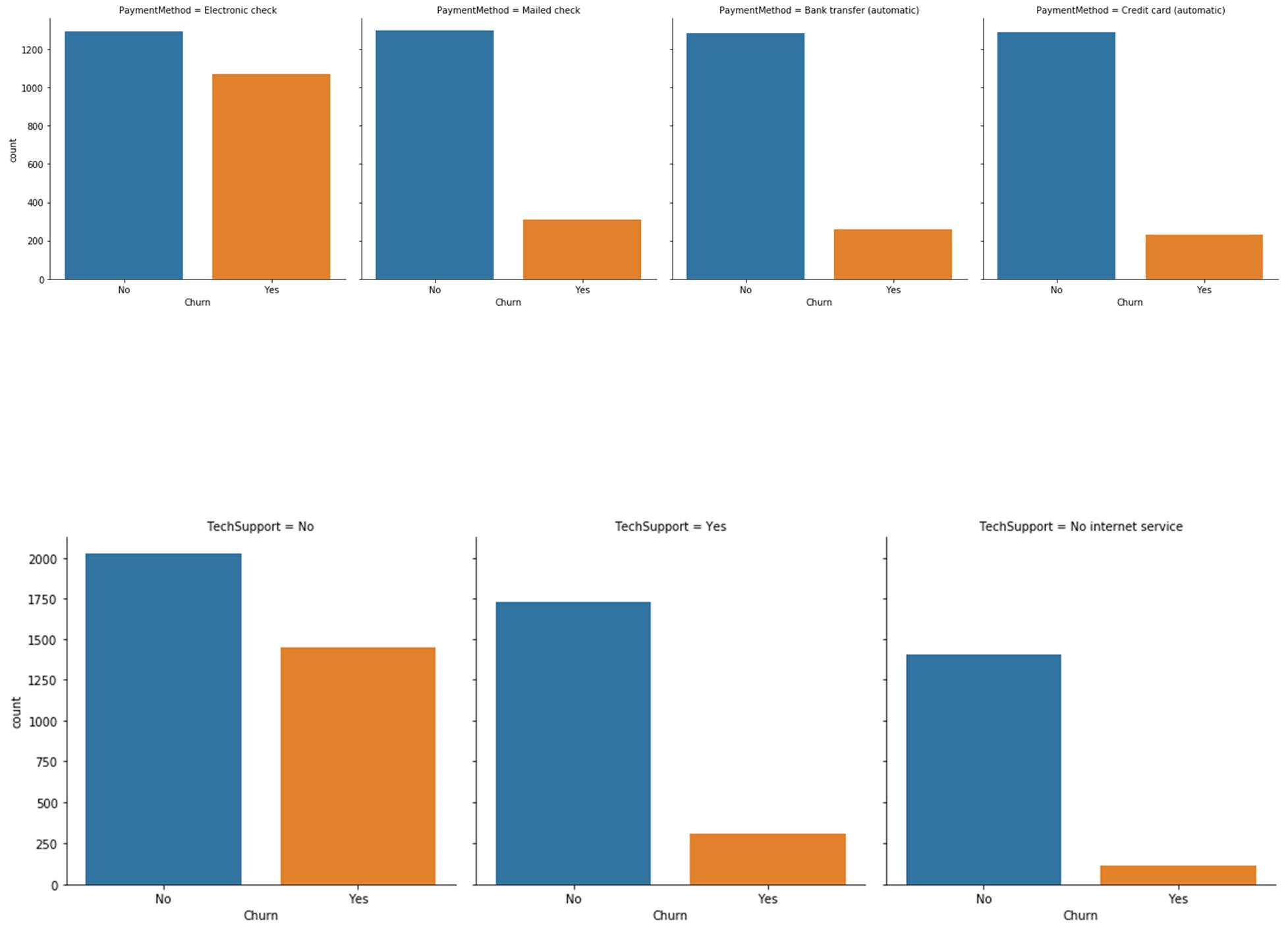


Customers chosen Month-to-month contract churned higher, this also may be important feature

## Payment Method – Churn Behavior

Customer paid by electronic check churned more

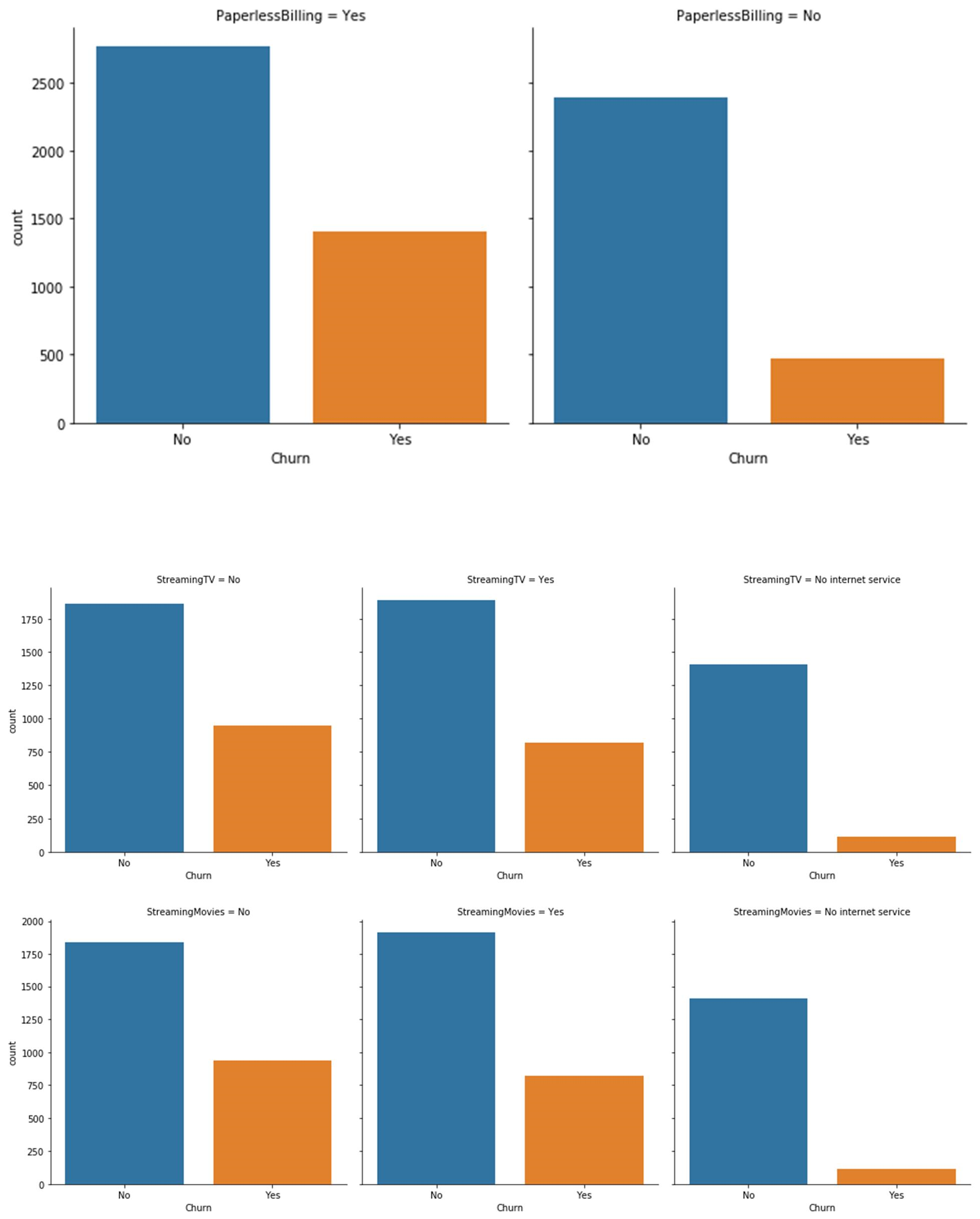
Customer Churn based on support



Customer without tech support churned higher than customer with tech support

# Churn based on Billing Method

Churn by Streaming Services

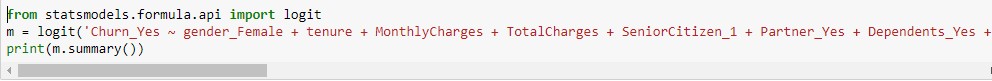


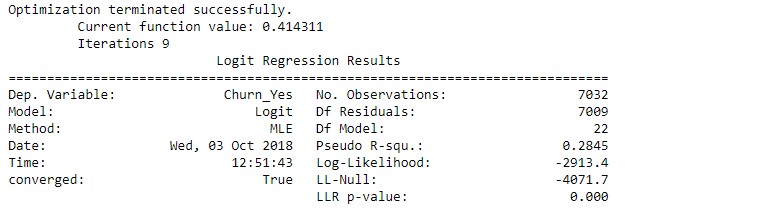
Both Streaming Tv and Streaming Movies have greater churn rate

## Feature Engineering

From EDA Internet Service, Contract, Payment Method, Streaming Tv and Streaming Movies, Tech Support are the Primary features and Paperless billing and Senior Citizen can be of secondary importance.

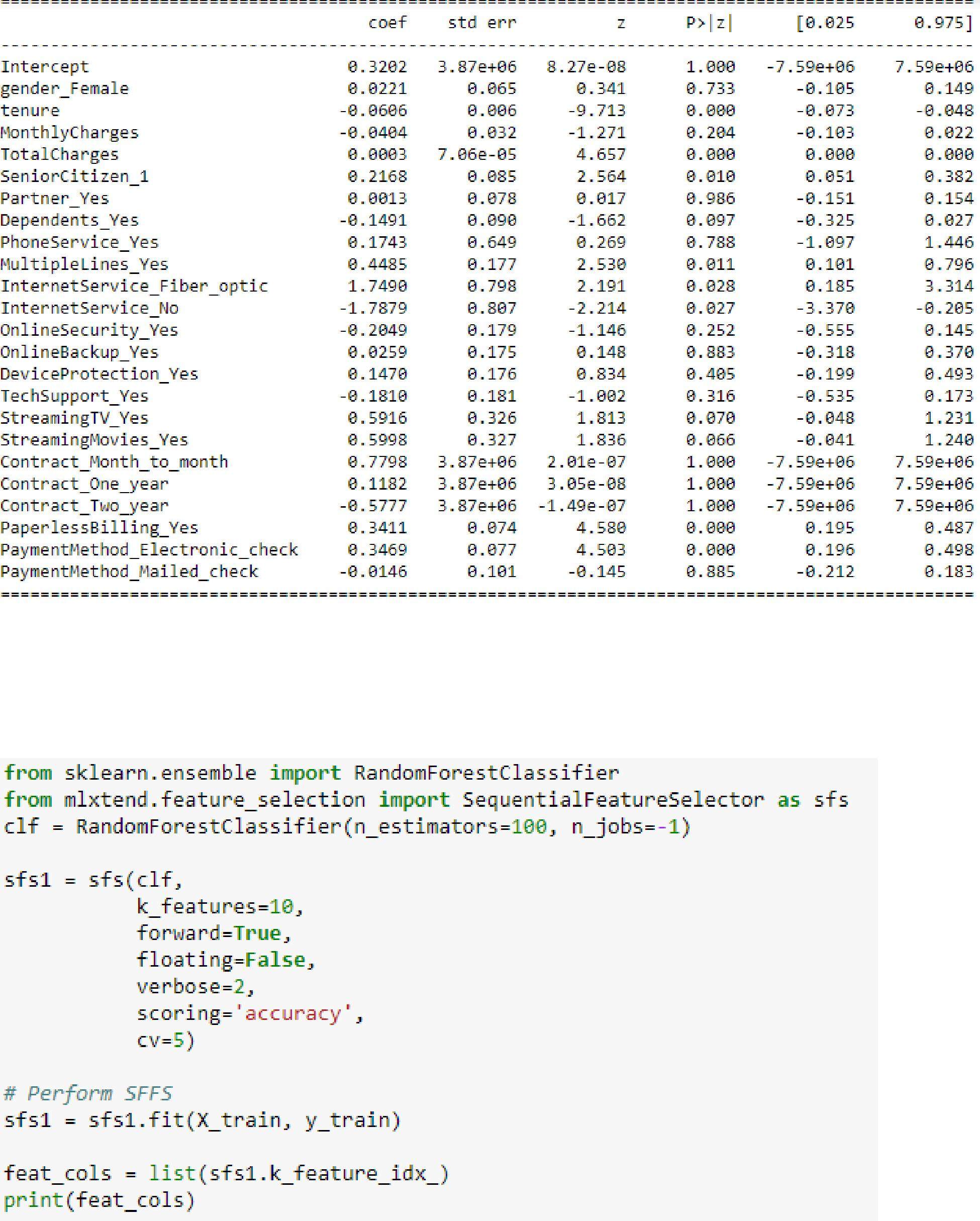
Second way is by applying stats model Logit function

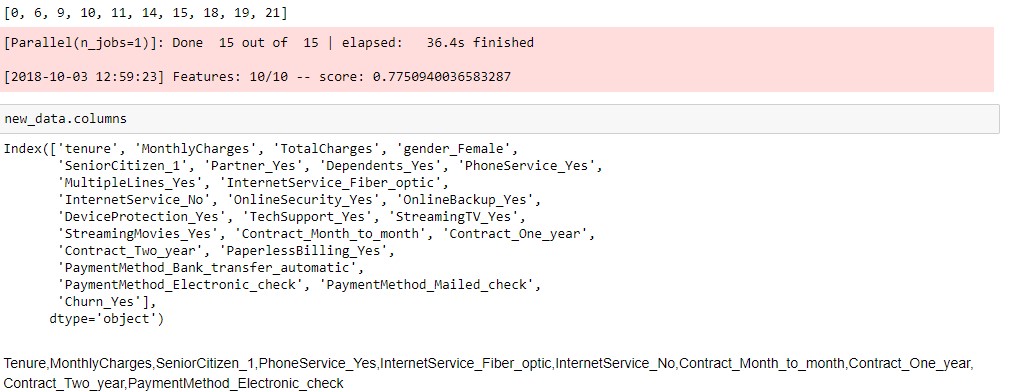




The third approach is to apply Random Forest Clas

ssifier to identify the important approach



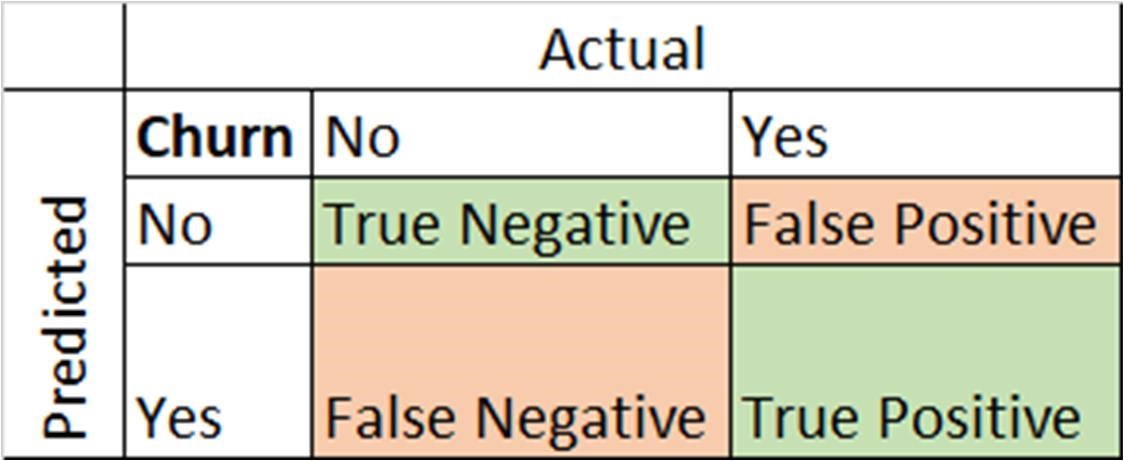


From all the above three methods the concluded features are Tenure, Monthly Charges, SeniorCitizen\_1, PhoneService\_Yes, InternetService\_Fiber\_optic, InternetService\_No, Contract\_Month\_to\_month, Contract\_One\_year, Contract\_Two\_year, PaymentMethod\_Electronic\_check

## CLASSIFICATION

Approaches for Predicting Qualitative Responses are called Classification. There are many classification techniques, widely used are Logistic Regression, Decision Trees, SVM, Random Forest and XG Boost. Deep Learning also can be applied for classification problems.

Confusion Matrix



The above table is called the Confusion Matrix, it clearly helps in studying our model performance.

* True Positive - Actual Data point is positive and Predicted as positive
* True Negative - Actual Data point is negative and Predicted as negative  False Positive - Actual Data point is negative and Predicted as positive  False Negative - Actual Data point is positive and Predicted as negative.

From above four values two metrics are calculated

* Precision = True Positives/ (True Positives + False Positives)
* Recall = True Positives/ (True Positives + False Negatives)

Precision talks about out of Total predicted positive values how many are actually Positive. Recall talks about the Total positive values predicted right.

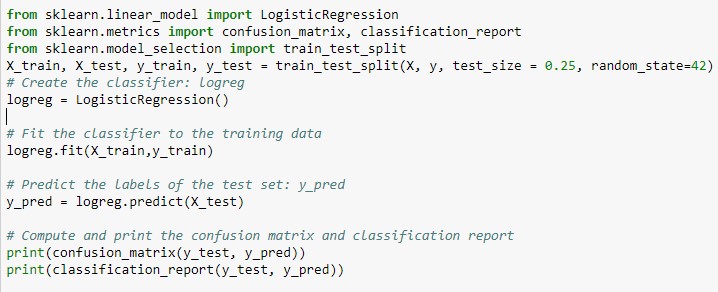
* Cost of False Positive is high, then Precision is a good measure  Cost of False Negative is high, then recall is a good measure.

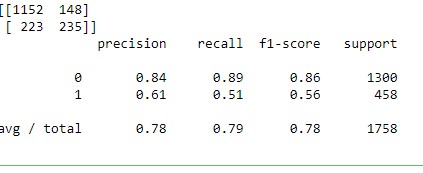
For Customer churn classification, False Negative cost is high that is if a customer churned, is identified as not churned customer. So, Recall is an important measure for this dataset.

### Logistic Regression

Logistic Regression models the probability that the output variable belongs to a particular category. Logistic Regression uses Logistic function as follows

p(X) = e^(b0 + b1\*X) / (1 + e^(b0 + b1\*X))

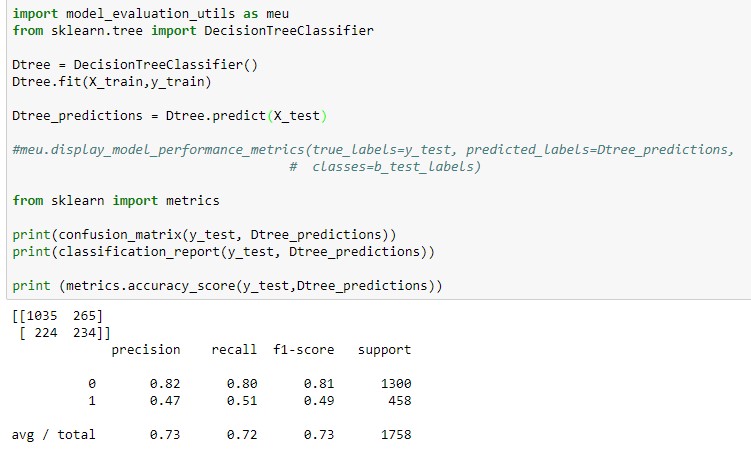




### Decision Trees

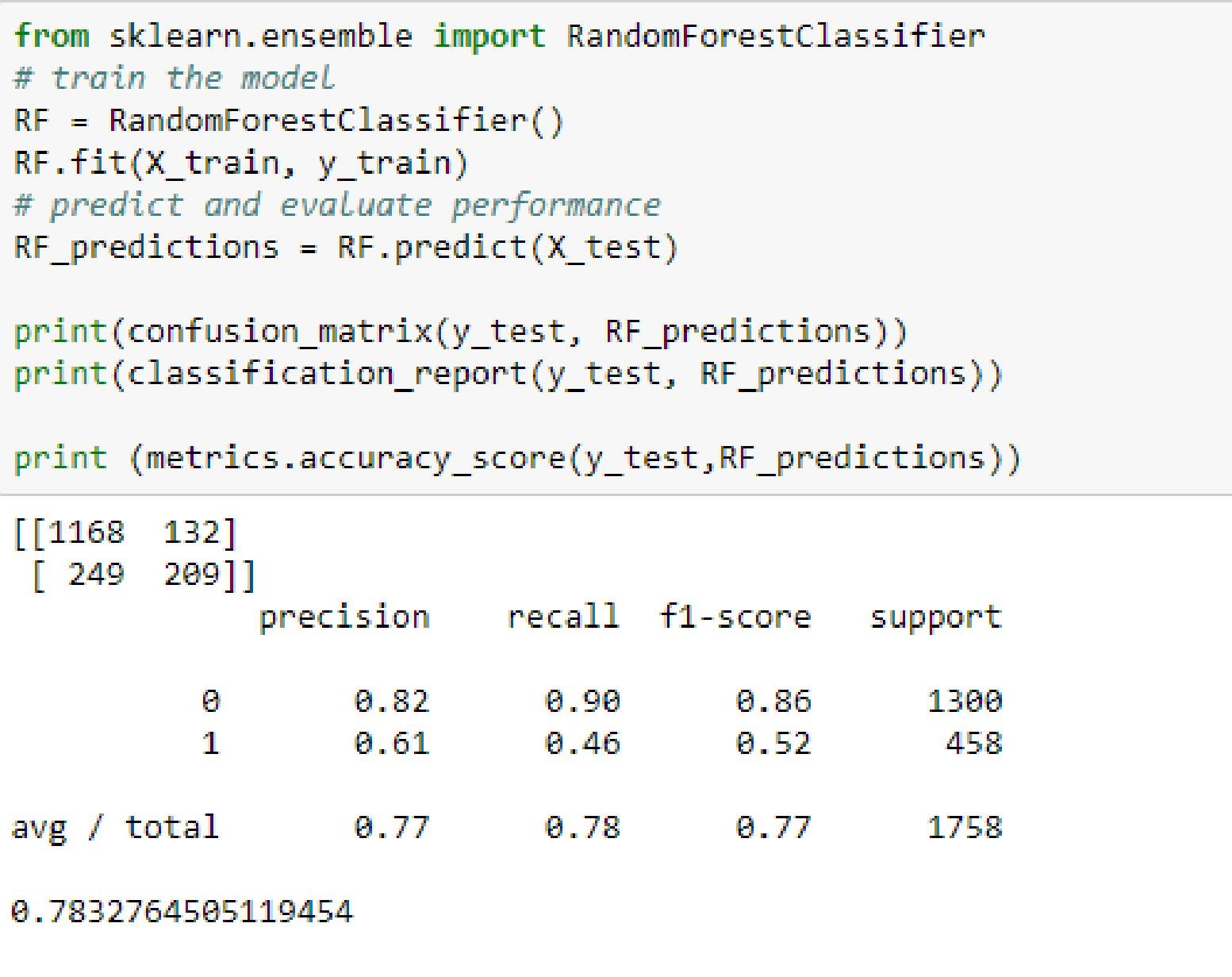
Decision Trees are non- non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.

* Decision Trees can be able to handle both continuous and categorical data.
* The main disadvantage of Decision trees is it will create overcomplex trees that lead to Overfitting. It also creates biased tree if one class dominates.



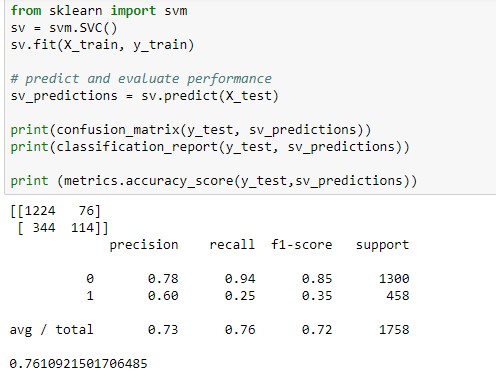
### Random Forest

In Random Forest each tree is built from samples wit replacement from training set. Split is not chosen as the best split among features instead built from random subset of features.



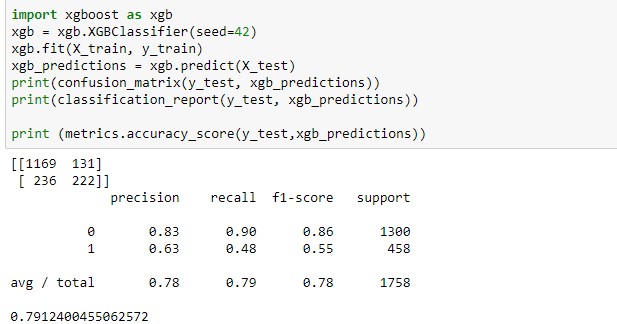
### SVM

* SVM - Support Vector Machines are very effective in high dimensional spaces.
* Rather than simply drawing a zero-width line between the classes, margin of some width can be drawn around each line, up to the nearest point.



### XG Boost

XGBoost is an implementation of gradient boosted decision trees designed for speed and performance that is dominative competitive machine learning.



### Deep Learning

* the type of network that work well on this kind of problems is a simple stack of fully connected Dense layers with relu activations.
* The argument being passed to each Dense layer is the number of hidden units of layer. Here it is 64.



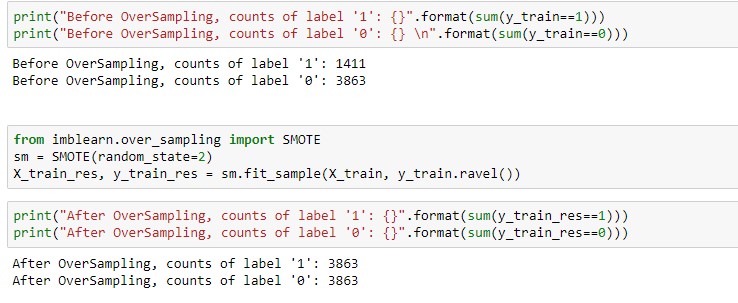
As the point of interest is recall, from above all the models have very less recall value for Customer Churn (Value 1) - 0.48 to 0.51, but recall for not churned if from 0.79 to 0.94. Considering Precision, the values are higher for classifier label 0 than 1. This leads to the doubt about the distribution of rows for both classifiers.

### SMOTE

Imbalanced data sets affect the performance and predictions of a model. From the code below the number of customers churn (1) is in very less number compared to not churned. There can be three types of solutions for these problems, they are

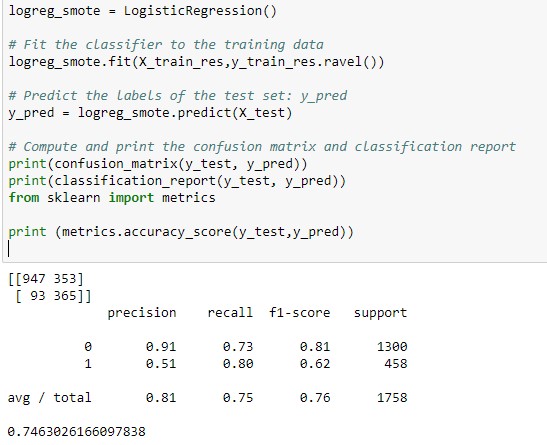
* Over-sample the minority class.
* Under-sample the majority class. 
* Synthesize new minority classes.

SMOTE (Synthetic Minority Over-sampling Technique) is the process of creating a new minority classes from the datasets.



After applying SMOTE the distribution of both are classifiers are synthesized equally. Now we can apply all the algorithms to see any considerable differences in model performance and predictions.

### Logistic regression after SMOTE



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Before SMOTE |  | After  SMOTE |  |
|  |  | Precision | Recall | Precision | Recall |
| Logistic | 0 | 0.84 | 0.89 | 0.91 | 0.73 |
|  | 1 | 0.61 | 0.51 | 0.51 | 0.8 |
| Decision | 0 | 0.83 | 0.79 | 0.83 | 0.78 |
|  | 1 | 0.47 | 0.53 | 0.47 | 0.54 |
| Random | 0 | 0.82 | 0.89 | 0.82 | 0.88 |
|  | 1 | 0.59 | 0.44 | 0.57 | 0.44 |
| svm | 0 | 0.78 | 0.94 | 0.81 | 0.84 |
|  | 1 | 0.68 | 0.25 | 0.49 | 0.44 |
| XG Boost | 0 | 0.83 | 0.9 | 0.85 | 0.85 |
|  | 1 | 0.63 | 0.48 | 0.58 | 0.58 |

* Decision Trees and Random Forest didn't show any considerable difference in recall and precision values before and after applying SMOTE
* SVM and XGBoost though recall value for classifier 1 increased but the precision for the same met with a decrease in value.
* Comparing other models Logistic Regression gives best value for recall and precision value with little drop in precision of Classifier 1

### Classification with Important Features

Considering only the important features from Feature Engineering analysis, the features are Tenure, MonthlyCharges, SeniorCitizen\_1, PhoneService\_Yes, InternetService\_Fiber\_optic, InternetService\_No, Contract\_Month\_to\_month, Contract\_One\_year, Contract\_Two\_year, PaymentMethod\_Electronic\_check



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Before SMOTE |  | After  SMOTE |  | SMOTE  less  Features |  |
|  |  | Precision | Recall | Precision | Recall | Precision | Recall |
| Logistic | 0 | 0.84 | 0.89 | 0.91 | 0.73 | 0.91 | 0.66 |
|  | 1 | 0.61 | 0.51 | 0.51 | 0.8 | 0.46 | 0.81 |
| Decision | 0 | 0.83 | 0.79 | 0.83 | 0.78 | NA | NA |
|  | 1 | 0.47 | 0.53 | 0.47 | 0.54 | NA | NA |
| Random | 0 | 0.82 | 0.89 | 0.82 | 0.88 | 0.9 | 0.68 |
|  | 1 | 0.59 | 0.44 | 0.57 | 0.44 | 0.47 | 0.79 |
| svm | 0 | 0.78 | 0.94 | 0.81 | 0.84 | 0.91 | 0.64 |
|  | 1 | 0.68 | 0.25 | 0.49 | 0.44 | 0.45 | 0.83 |
| XG  Boost | 0 | 0.83 | 0.9 | 0.85 | 0.85 | 0.9 | 69 |
|  | 1 | 0.63 | 0.48 | 0.58 | 0.58 | 0.47 | 0.78 |

As only a smaller number of features selected precision and recall values suffered a considerable amount. Recall for classifier 0 and Precision for 1 suffers the most

From the above analysis Logistic Regression after SMOTE application performs best for this problem implying that not always complex algorithms are essential for better performance.

# Conclusion

Always it is not the accuracy is only important. Based on business question and problem we try to solve other measures are equally important. As for the problems like Customer Churn, Credit defaulters and Spam emails mainly based on recall value. Same way there is not any golden rule that complex algorithms gives best performance and accuracy. Each problem is different so gauging against all the algorithms will be a better choice.

For future analysis, still improvement on Feature Engineering and fine tuning the parameters will give better results.

