Telco Customer Churn

"Predicting Behavior to Retain Customers"

Group Members:

Akash Valathappan

Abdur Rahman Danish

Deepak Sri Ram murthy Bolem

Jonathan Gabriel

Project Objectives

The Main purpose of this project is to analyze and understand the factors that influence Customer Churn and develop strategies to retain customers.

To identify patterns and trends in customer behavior, and then using their information to design and implement targeted retention programs.

Aiming to help the telco company develop an effective retention strategy using outcomes of our model that will help the company to keep their customers satisfied and engaged.

Business Questions

What are the characteristics of customers who are more likely to churn, and can we identify them before they leave?

What are the most effective retention strategies for different types of users?

What kinds of services can be implemented or should be improved to decrease the rate of churn?

What is cost of acquiring new customers compared to the retaining existing ones?

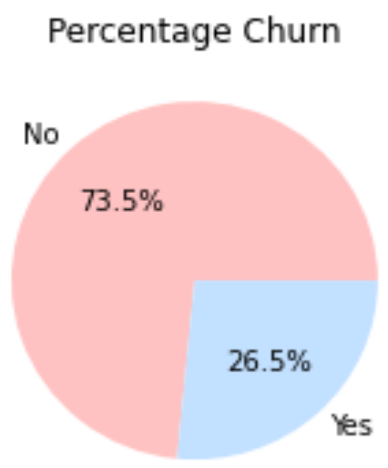
Dataset

Data Source: The Telco Customer Churn data was obtained from Kaggle and has originated from the IBM analytics community.

Data: The dataset contains 7043 rows and 21 features.

Variables: There are 17 categorical and 3 numeric variables. These include Customer demographics, Services signed up by customer and Account information

Churn: Our target variable 'Churn' has a classification ratio of 73%-27% (No – Yes)



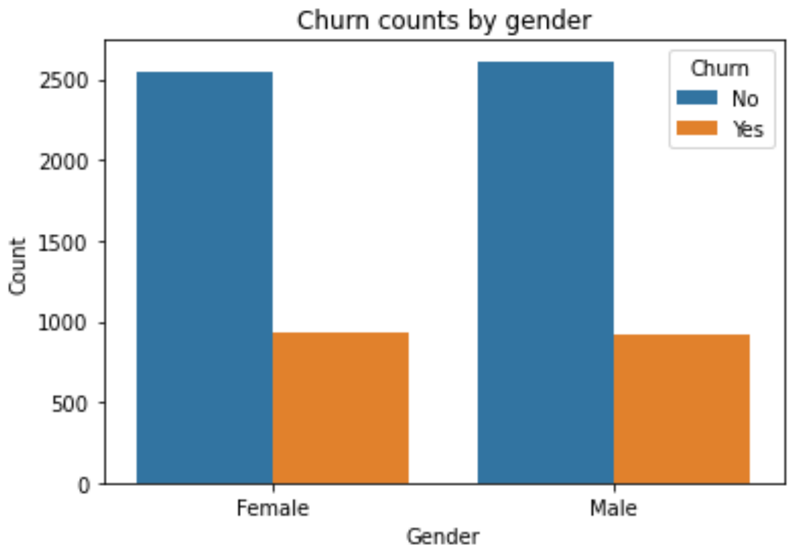
Data Wrangling

Data preprocessing is a crucial step in a data mining project as it helps to ensure data quality and consistency. We performed several data cleaning and preprocessing steps to prepare the data for further analysis:

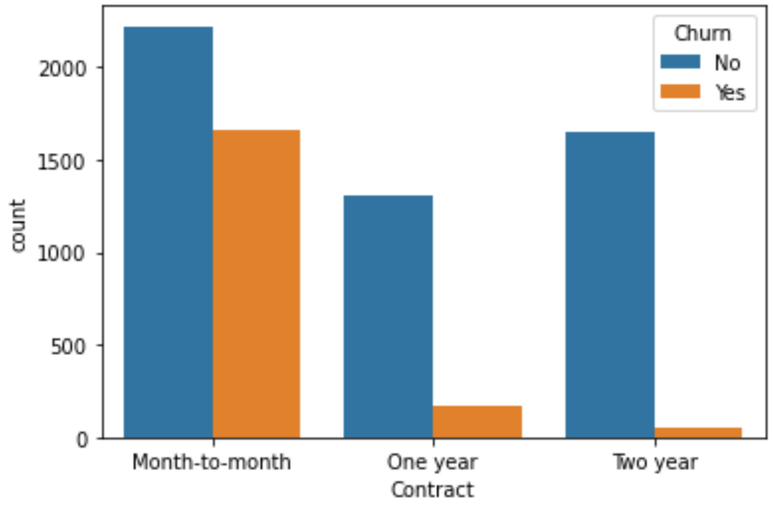
- We converted the 'TotalCharges' column from object to a numeric type.
- We identified and removed 22 duplicate rows in our dataset to ensure that each customer was only represented once.
- We checked and got rid of the 11 missing values.
- We performed outlier detection on our dataset and found no outliers.
- We created dummy variables for categorical variables to enable further analysis.

Exploratory Data Analysis

Exploratory Data Analysis

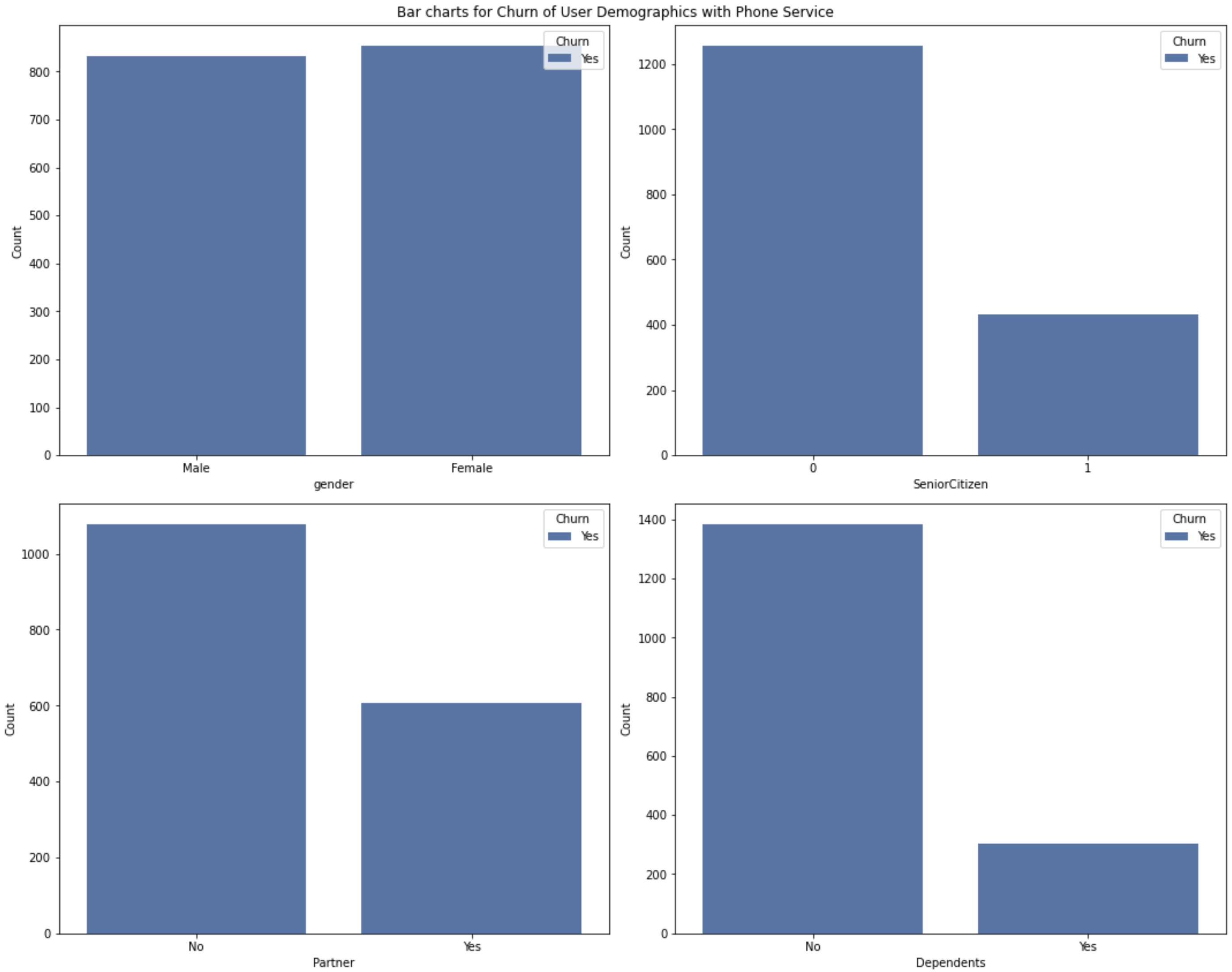


The churn within gender distribution was roughly equal with no significant difference between the number of males and females.



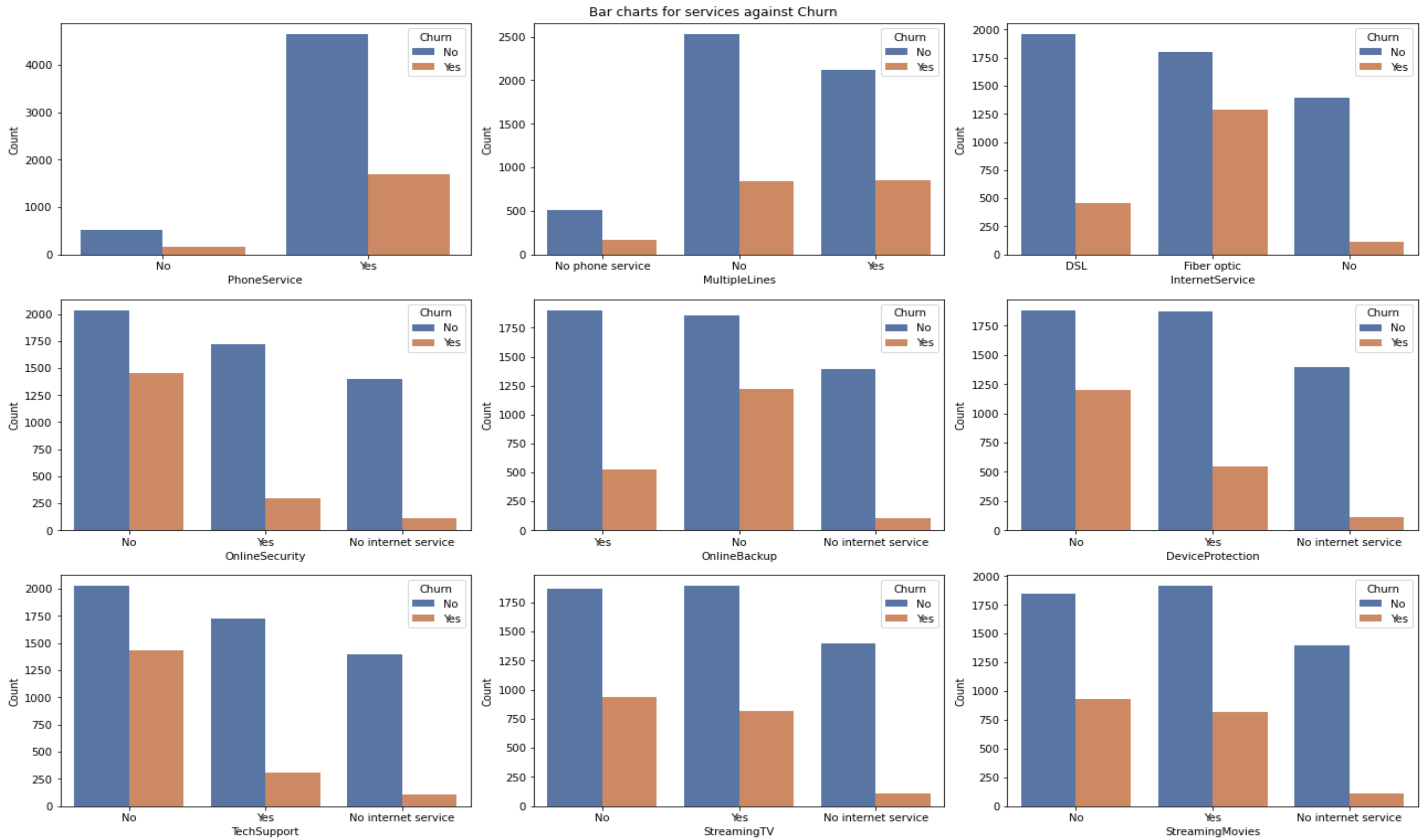
Customers who had month-to-month contract churned more than those who had one- or two-year contracts.

Exploratory Data Analysis

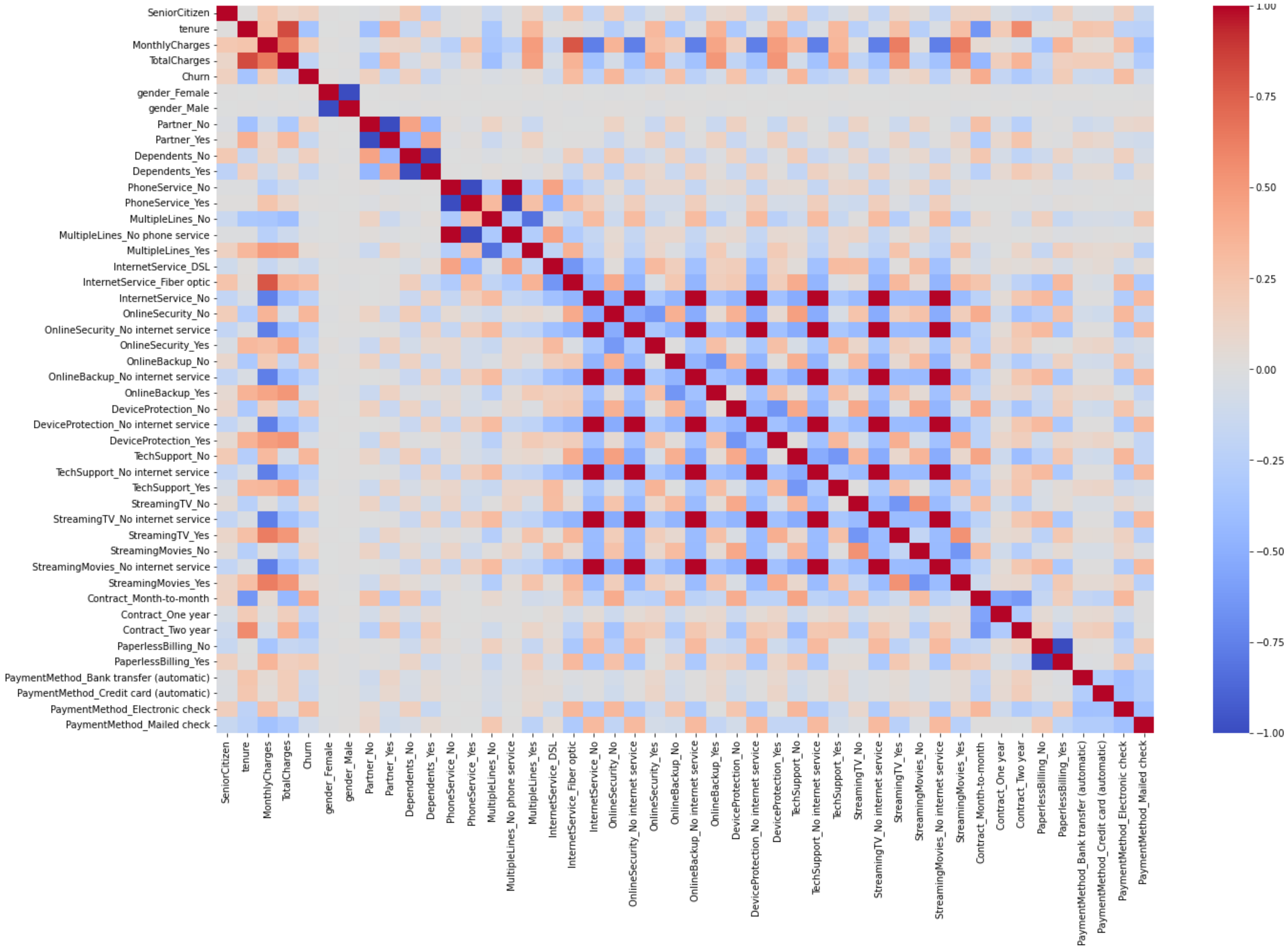


We can conclude that customers who use phone service
and churned are likely to be non-seniors, without partners and without dependents.

Exploratory Data Analysis



Correlation Matrix



The Correlation matrix showed that month-to-month contracts have the highest correlation with customer churn.

CORRELATED VARIABLES	Churn
Contract_Month-to-month	0.404346383
OnlineSecurity_No	0.34185304
TechSupport_No	0.336456244
InternetService_Fiber optic	0.307611936
PaymentMethod_Electronic check	0.301078799
OnlineBackup_No	0.266637331
DeviceProtection_No	0.251038065
MonthlyCharges	0.19400838
PaperlessBilling_Yes	0.190518382
Dependents_No	0.162365916

Data Mining Models Applied

Supervised Machine Learning Models

- Logistic Regression
- Decision Trees
- Random Forests

Model Improvement Techniques

- SelectKBest: to select top features with high impact on Churn variable.
- SMOTE Oversampling: to balance the data and resample.
- Hyperparameter Tuning: GridSearchCV to obtain best hyperparameters.

Results and Analysis

Model Comparison

Logistic Regression:

	Accuracy	F1 - Score
Logistic Regression	82%	59%
With K Best Feature Selection	82%	59%
With Smote	81%	60%
With Hyper Parameter Tuning	82%	60%

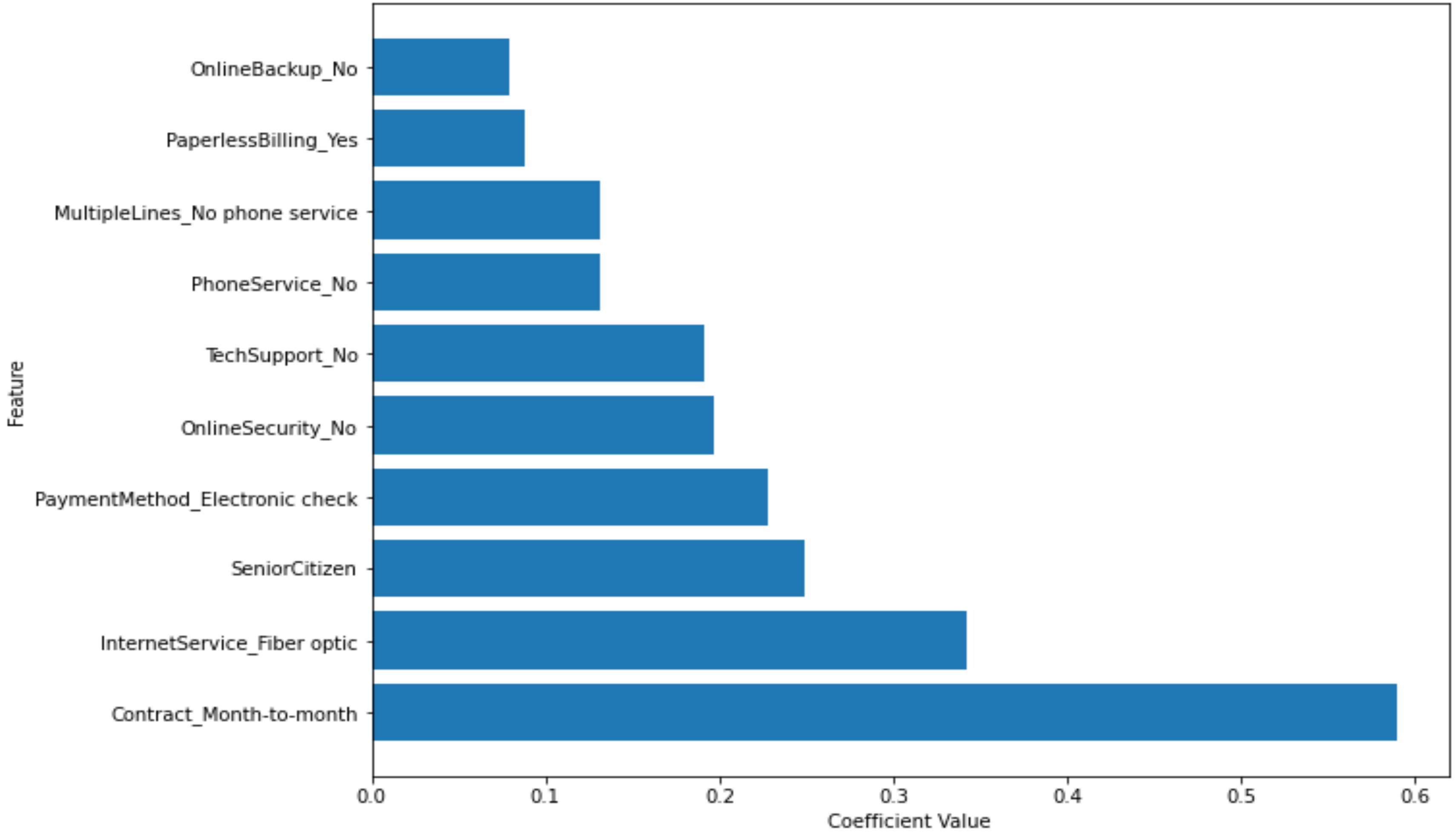
Random Forest:

	Accuracy	F1 - Score
Random Forest	78%	52%
With K Best Feature Selection	80%	54%
Smote	64%	23%

Decision Tree:

	Accuracy	F1 - Score
Decision Tree	73%	49%
With K Best Feature Selection	74%	50%
Smote	74%	50%

High Prediction variables for the Retention Rate



THANK YOU