Telecom Churn Prediction Analysis

Sailaja Karra

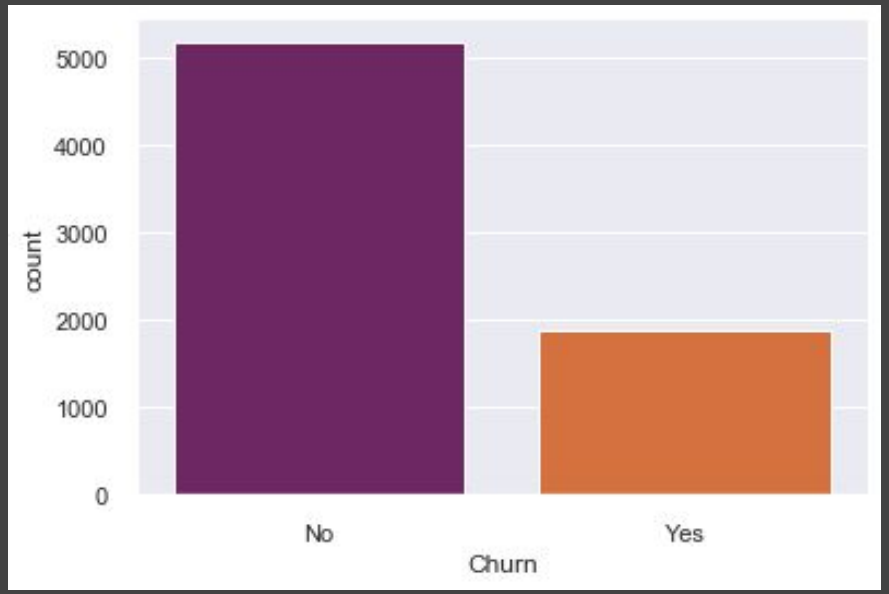
Introduction

What is Churn?

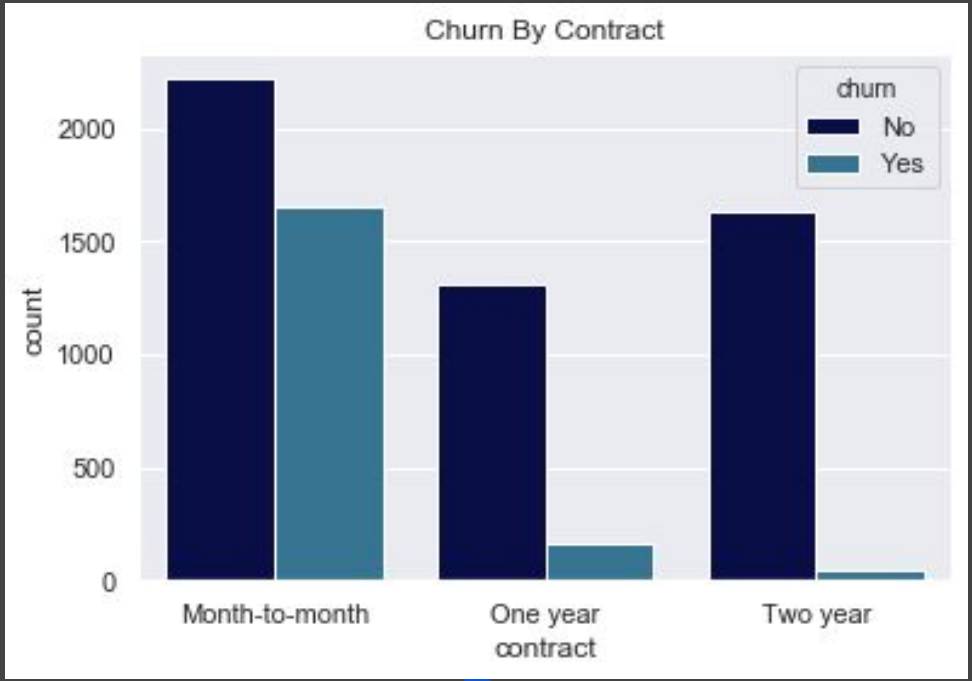
Customer churn is the loss of clients or customers.

Data Collected from:
IBM Sample Data Sets
(via Kaggle)

Total Churns per month in the data set is 26.53%

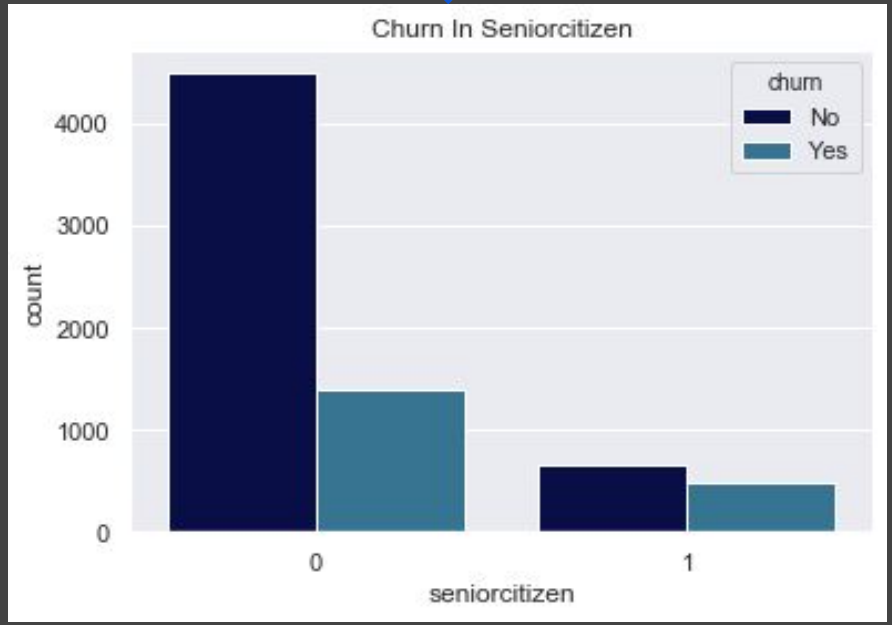


Exploratory Data Analysis

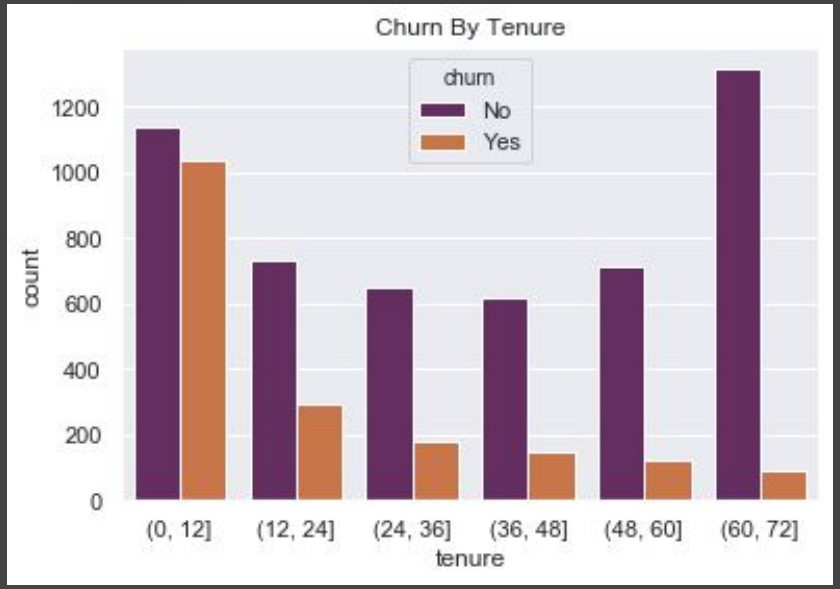


This shows Churn by contract type, as expected the churn is very low for customers with one and two year contracts

This graph shows Churn in senior citizens, 0 is Not Senior Citizen

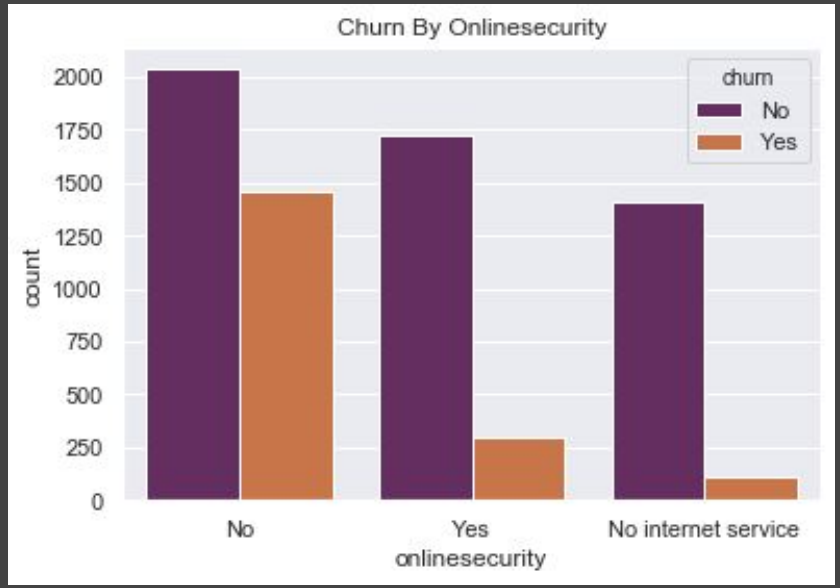


Further Exploratory Data Analysis



This shows Churning after 'binning' by tenure

This shows Churn based on customer buying 'online security' with the company.



FEATURE ENGINEERING

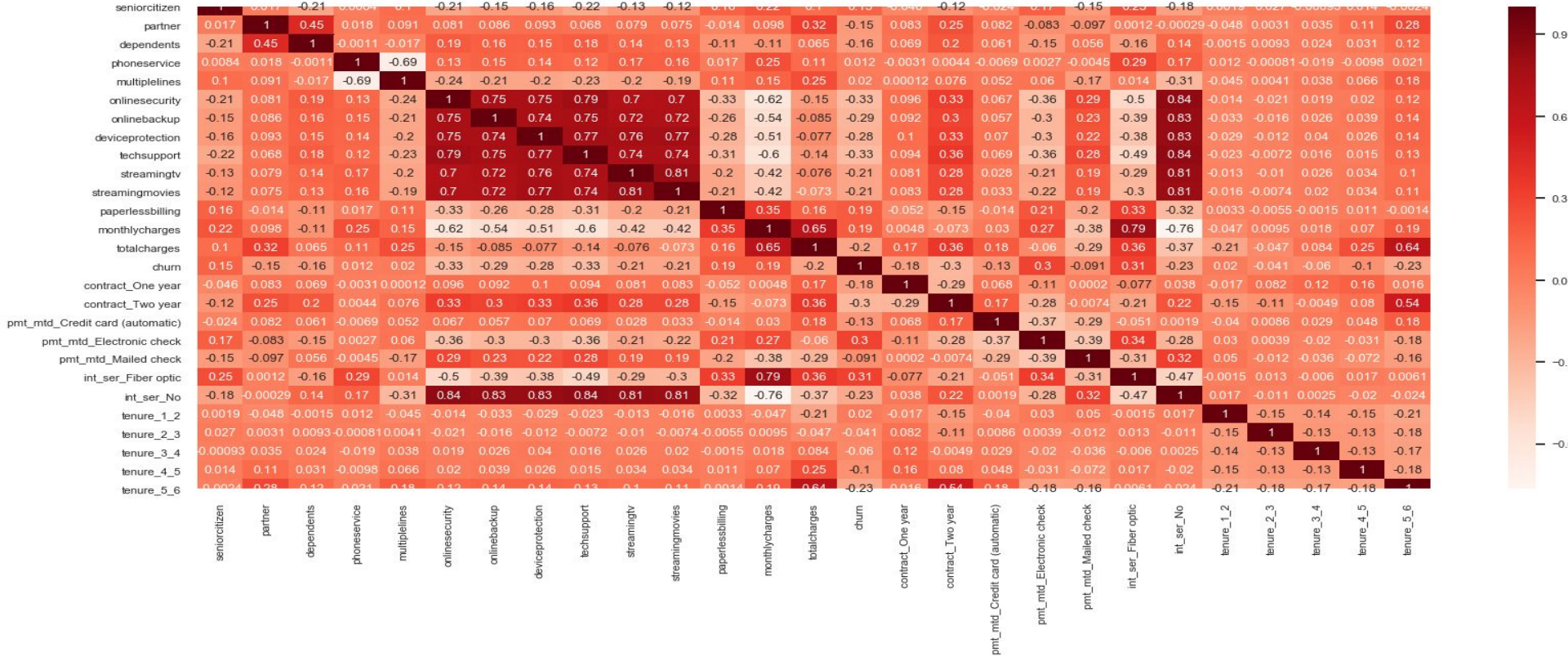
One Hot Encoding
Phoneservice
Partner
Dependents
Paperlessbilling
Churn

Binning
Tenure

Scaling
Monthly Charges
Total Charges

Correlation Analysis

Removing highly correlated columns: **streamingtv, streamingmovies, techsupport, deviceprotection, onlinebackup**.



Models

KNN	• Accuracy Score: 77%	• F1 Score: 55.7%
XG Boost(Grid Search)	• Accuracy Score: 76%	• F1 Score: 49%
Logistic regression(L1)	• Accuracy Score: 79%	• F1 Score: 52%
Logistic regression with recursive feature	• Accuracy Score: 79%	• F1 Score: 53%
Decision Tree	• Accuracy Score: 75%	• F1 Score: 59%
SMOTE	• Accuracy Score: 74%	• F1 Score: 61%
Random Forest	• **Accuracy Score: 74%**	• **F1 Score: 62%**

Best Model Parameters

Random Forest Classifier

bootstrap=True,
ccp_alpha=0.0,
class_weight='balanced',
criterion='gini',
max_depth=10,
min_samples_leaf=5,
min_samples_split=2,
n_estimators=200,
n_jobs=None,
random_state=23

Summary

Using Random Forest Model we get the best predictor for Churn.

The important features are:

- One Year Contract
- Two Year Contract
- Online Security
- Tenure