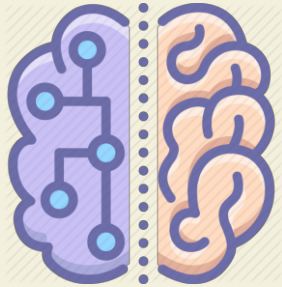
# Telco Analysis

Churn Prediction Model

Phatpicha Y. in 2020

# OVERVIEW OF ANALYSIS



**Data Cleaning**

Understand the structure of dataset and clean data

**Data Exploration**

Find significant patterns and trends

**Predictive Modeling**

Construct models to predict and forecast

# DATA CLEANING

# DATASET DESCRIPTION

telco.csv

```
df.head()
```

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

```
df.describe()
```

| | SeniorCitizen | tenure | MonthlyCharges |
|---|---|---|---|
| count | 7043.000000 | 7043.000000 | 7043.000000 |
| mean | 0.162147 | 32.371149 | 64.761692 |
| std | 0.368612 | 24.559481 | 30.090047 |
| min | 0.000000 | 0.000000 | 18.250000 |
| 25% | 0.000000 | 9.000000 | 35.500000 |
| 50% | 0.000000 | 29.000000 | 70.350000 |
| 75% | 0.000000 | 55.000000 | 89.850000 |
| max | 1.000000 | 72.000000 | 118.750000 |

```
df.describe(include=["O"])
```

| | customerID | Region | gender | Partner | Dependents | PhoneService | MultipleLines |
|---|---|---|---|---|---|---|---|
| count | 7043 | 7043 | 7043 | 7043 | 7043 | 7043 | 7043 |
| unique | 7043 | 3 | 2 | 2 | 2 | 2 | 3 |
| top | 7041-TXQJH | Germany | Male | No | No | Yes | No |
| freq | 1 | 6761 | 3555 | 3641 | 4933 | 6361 | 3390 |

# DATASET DESCRIPTION

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 22 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customerID        7043 non-null   object
 1   Region            7043 non-null   object
 2   gender            7043 non-null   object
 3   SeniorCitizen     7043 non-null   int64
 4   Partner           7043 non-null   object
 5   Dependents        7043 non-null   object
 6   tenure            7043 non-null   int64
 7   PhoneService      7043 non-null   object
 8   MultipleLines     7043 non-null   object
 9   InternetService   7043 non-null   object
 10  OnlineSecurity    7043 non-null   object
 11  OnlineBackup      7043 non-null   object
 12  DeviceProtection  7043 non-null   object
 13  TechSupport       7043 non-null   object
 14  StreamingTV       7043 non-null   object
 15  StreamingMovies   7043 non-null   object
 16  Contract          7043 non-null   object
 17  PaperlessBilling  7043 non-null   object
 18  PaymentMethod     7043 non-null   object
 19  MonthlyCharges    7043 non-null   float64
 20  TotalCharges      7043 non-null   object
 21  Churn             7043 non-null   object
dtypes: float64(1), int64(2), object(19)
memory usage: 1.2+ MB
```

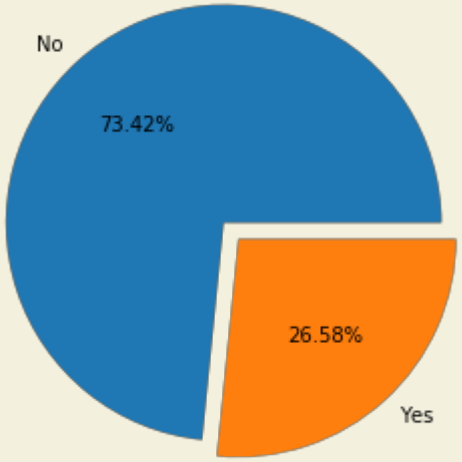
```
df.info()
```

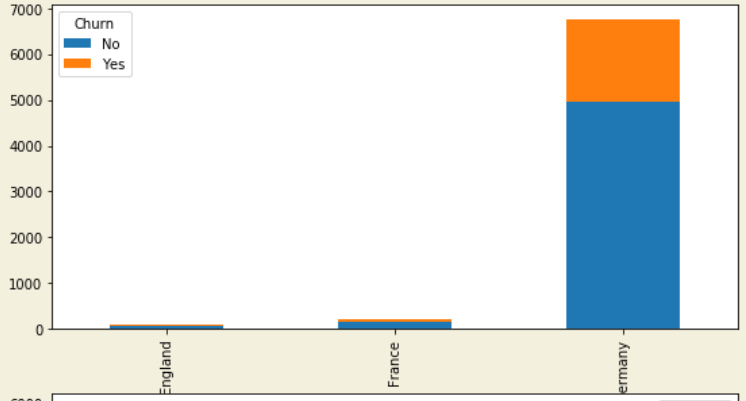
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 7032 entries, 0 to 7042
Data columns (total 22 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customerID        7032 non-null   object
 1   Region            7032 non-null   object
 2   gender            7032 non-null   object
 3   SeniorCitizen     7032 non-null   int64
 4   Partner           7032 non-null   object
 5   Dependents        7032 non-null   object
 6   tenure            7032 non-null   int64
 7   PhoneService      7032 non-null   object
 8   MultipleLines     7032 non-null   object
 9   InternetService   7032 non-null   object
 10  OnlineSecurity    7032 non-null   object
 11  OnlineBackup      7032 non-null   object
 12  DeviceProtection  7032 non-null   object
 13  TechSupport       7032 non-null   object
 14  StreamingTV       7032 non-null   object
 15  StreamingMovies   7032 non-null   object
 16  Contract          7032 non-null   object
 17  PaperlessBilling  7032 non-null   object
 18  PaymentMethod     7032 non-null   object
 19  MonthlyCharges    7032 non-null   float64
 20  TotalCharges      7032 non-null   float64
 21  Churn             7032 non-null   object
dtypes: float64(2), int64(2), object(18)
memory usage: 1.2+ MB
```

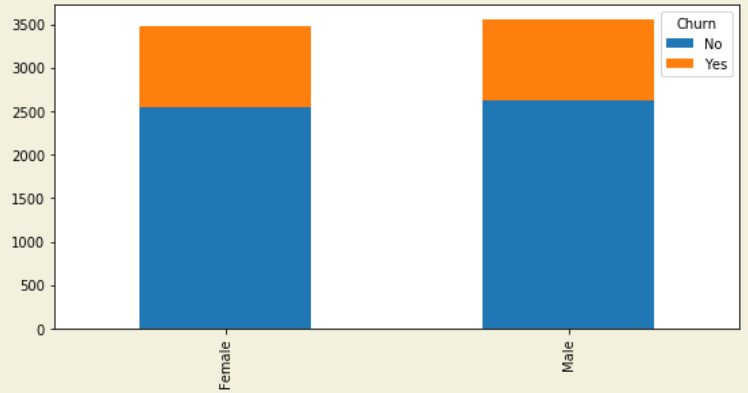
```
i = df[df["TotalCharges"] == " "].index
df = df.drop(i)
```

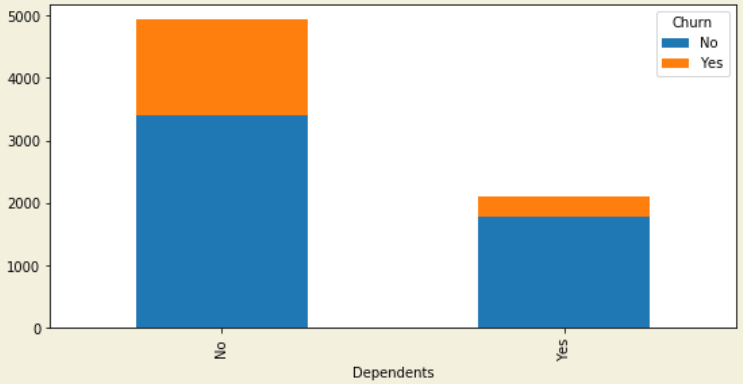
# EXPLORATORY DATA ANALYSIS

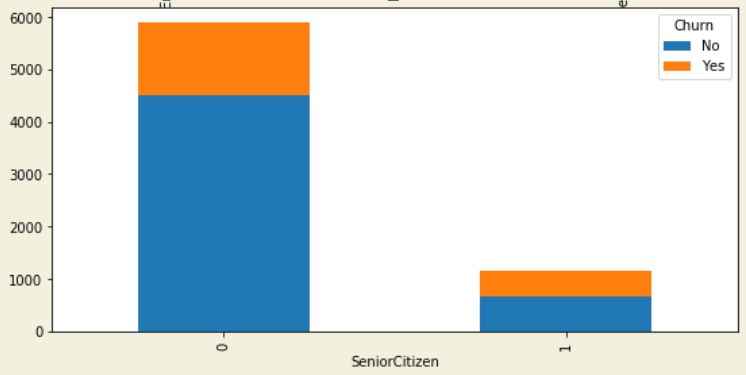
# Categorical Features

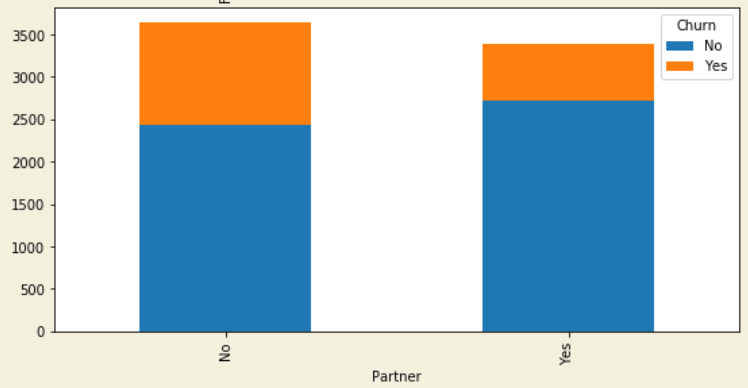


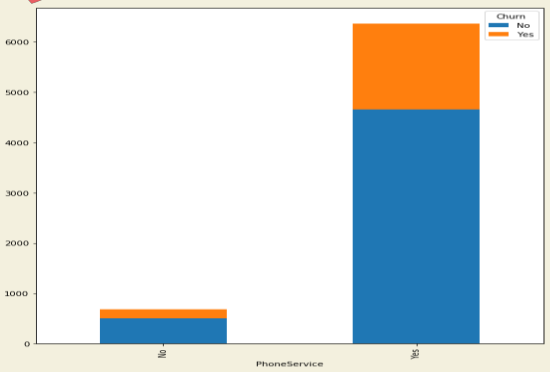



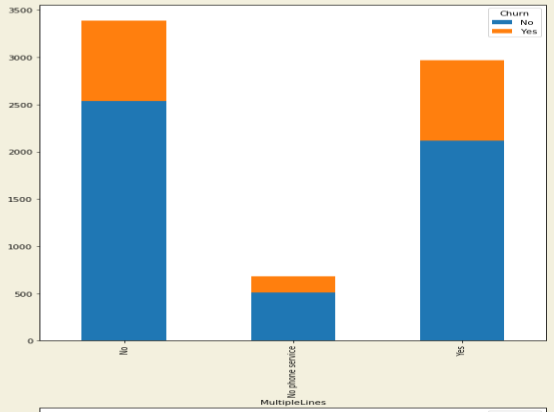



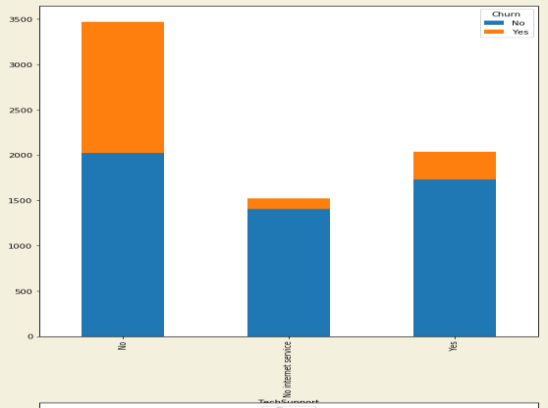



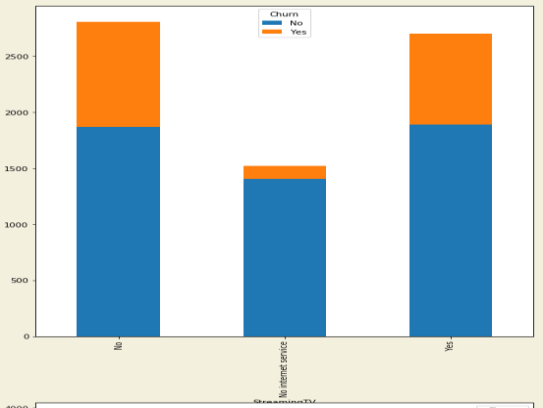



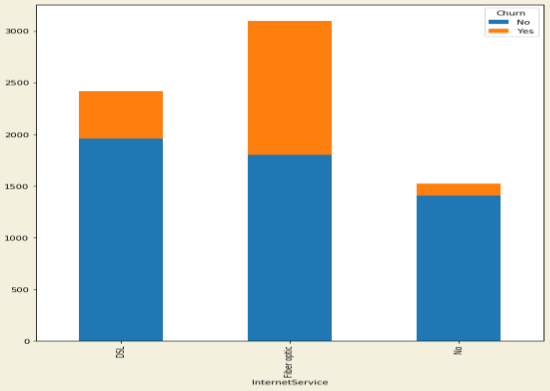
# Categorical Features

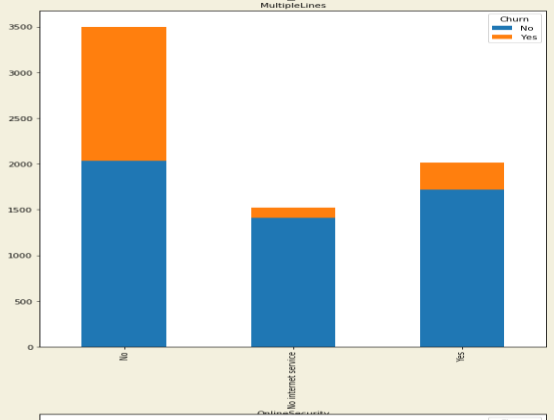


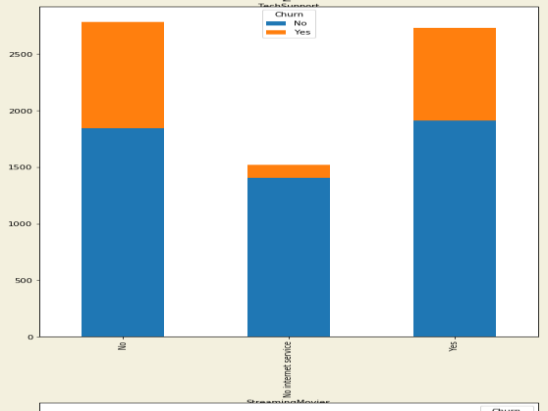



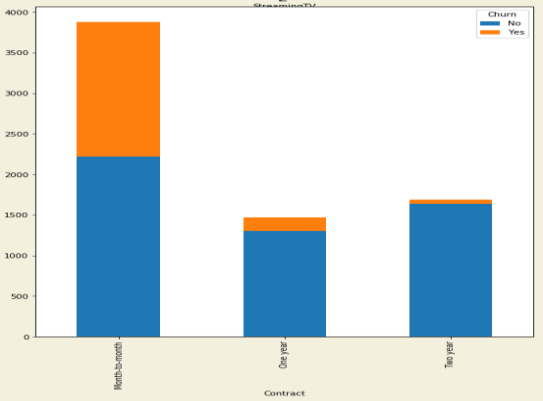



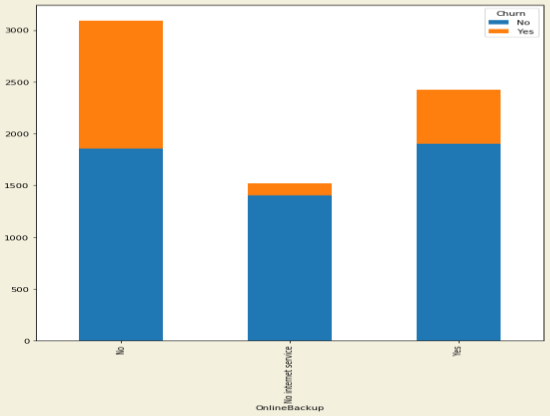



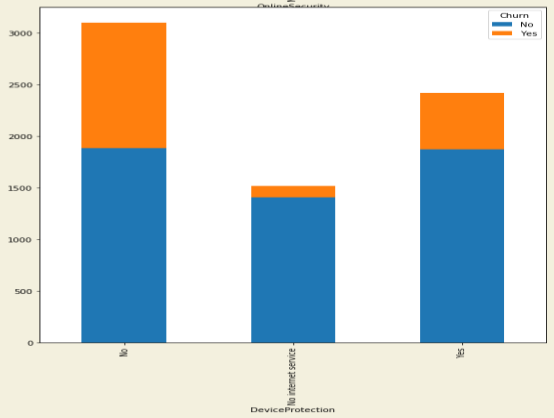



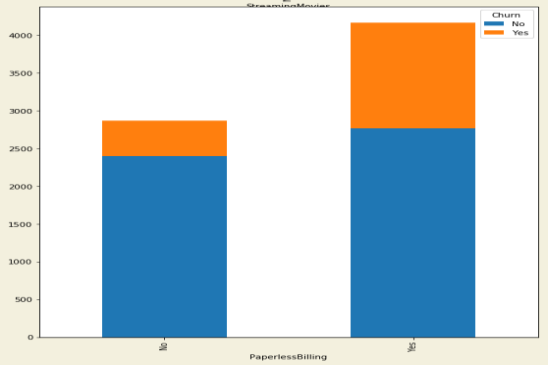



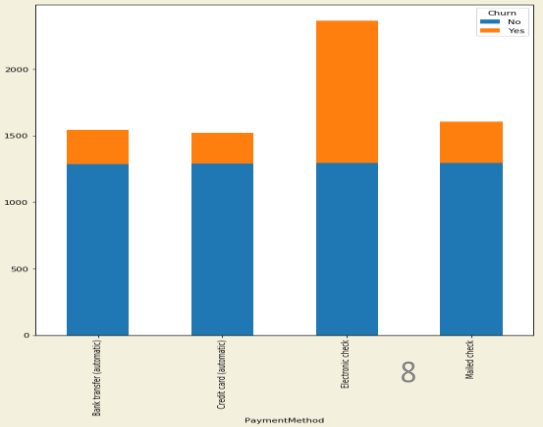



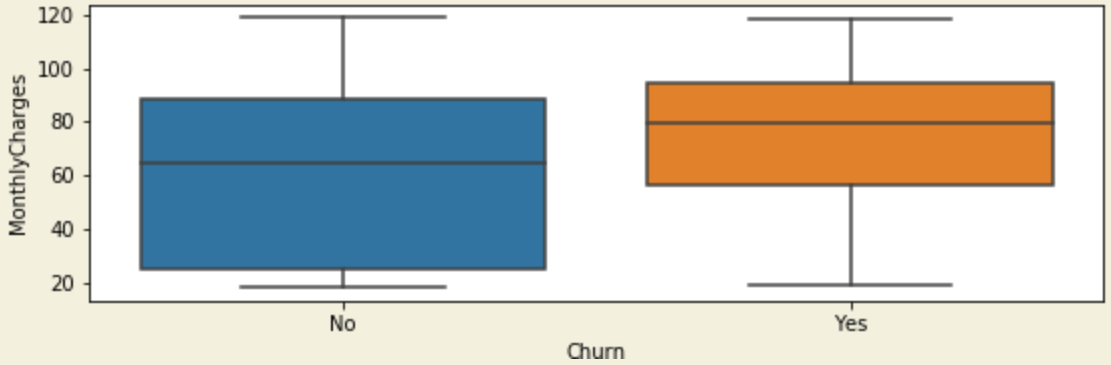



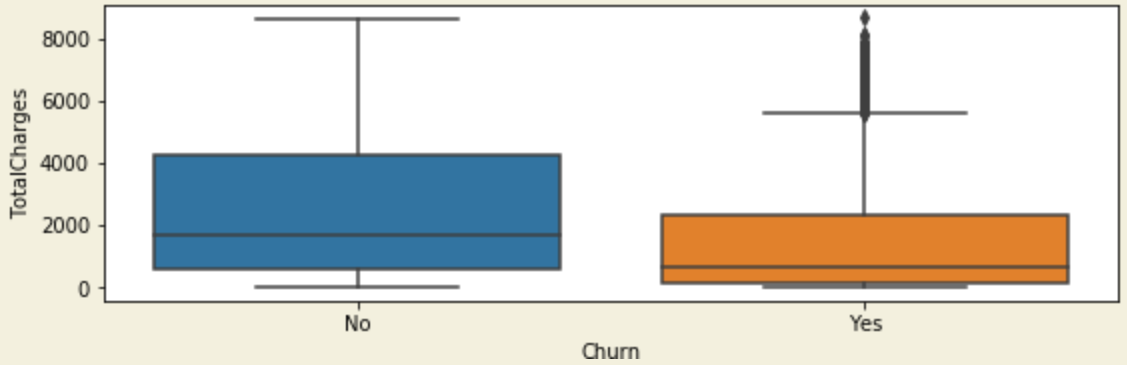



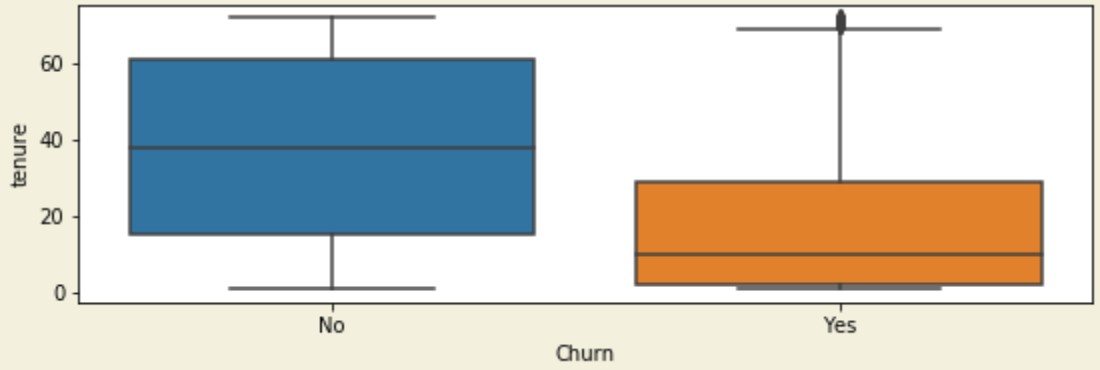



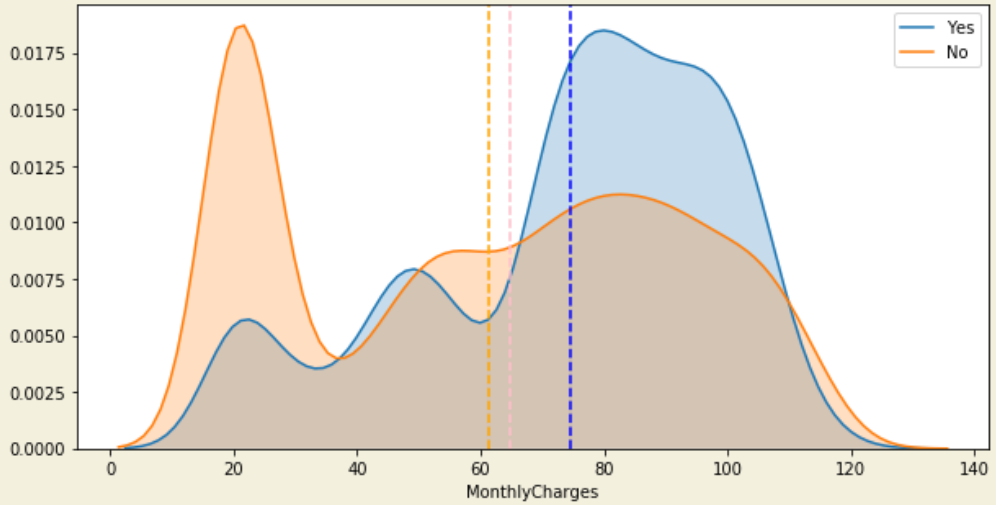
# Numeric Features

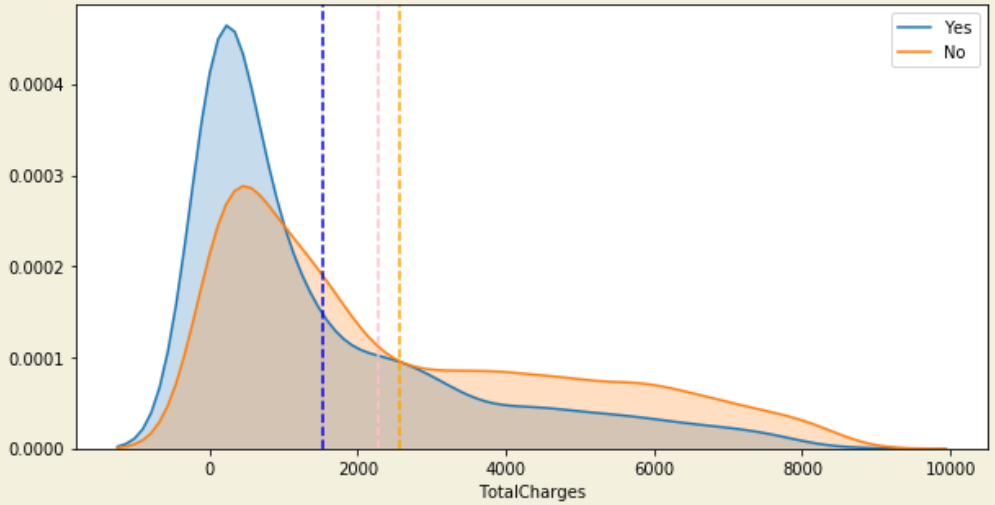


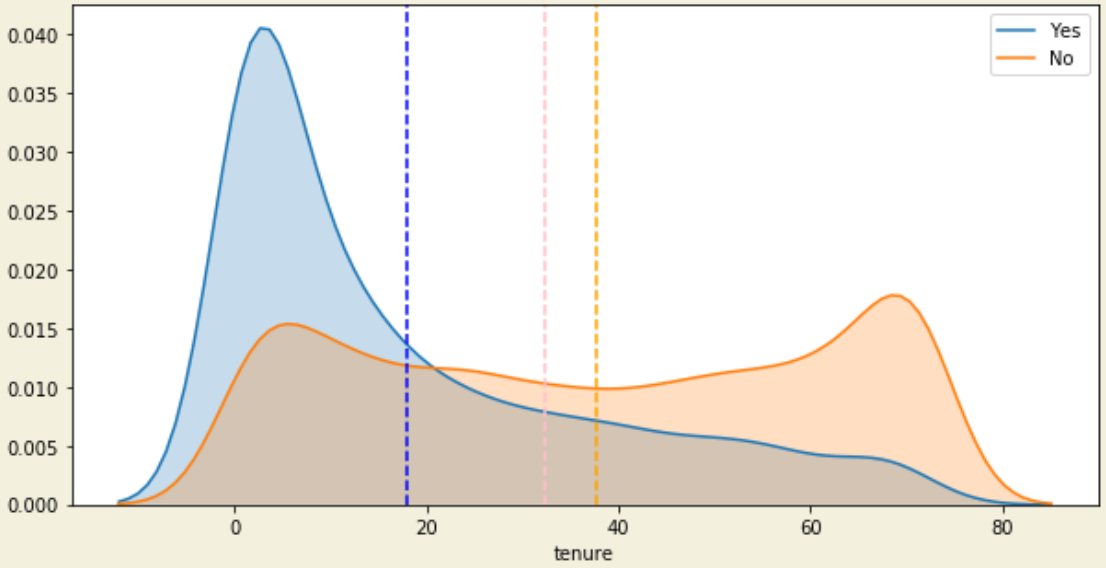



# Numeric Features

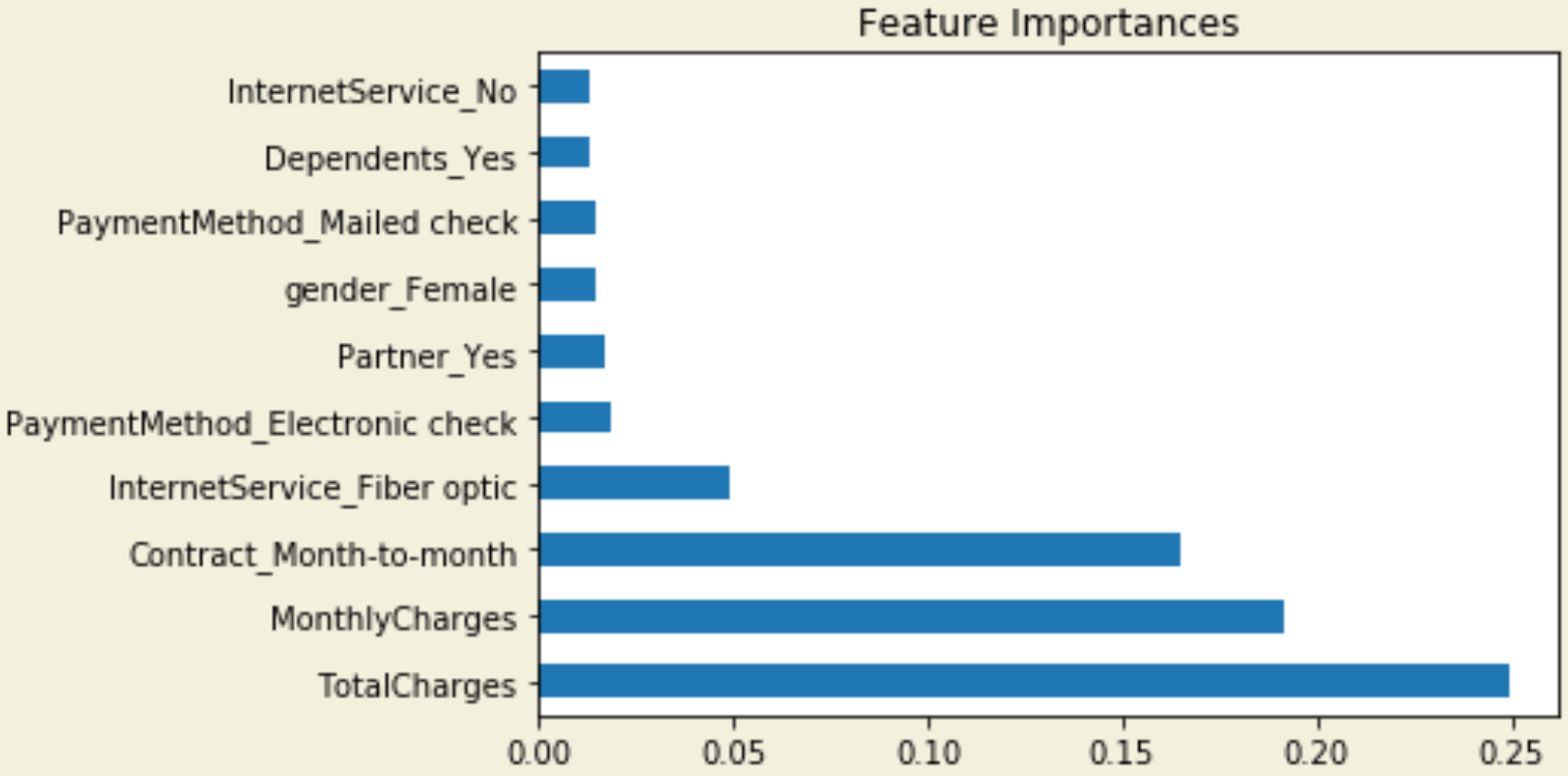


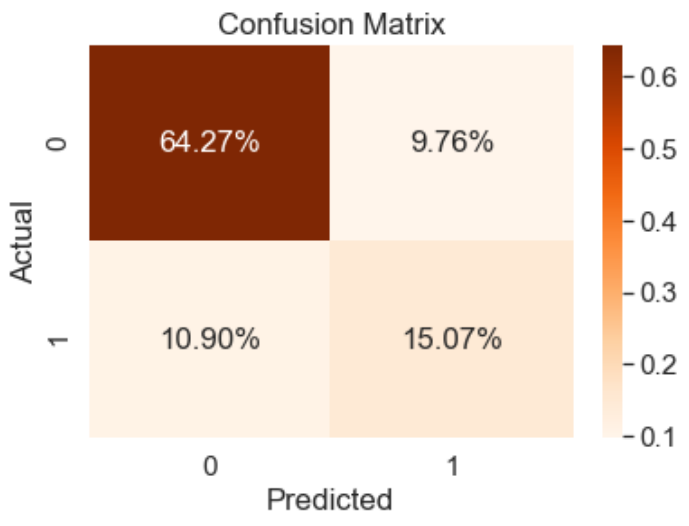
# PREDICTIVE MODELING

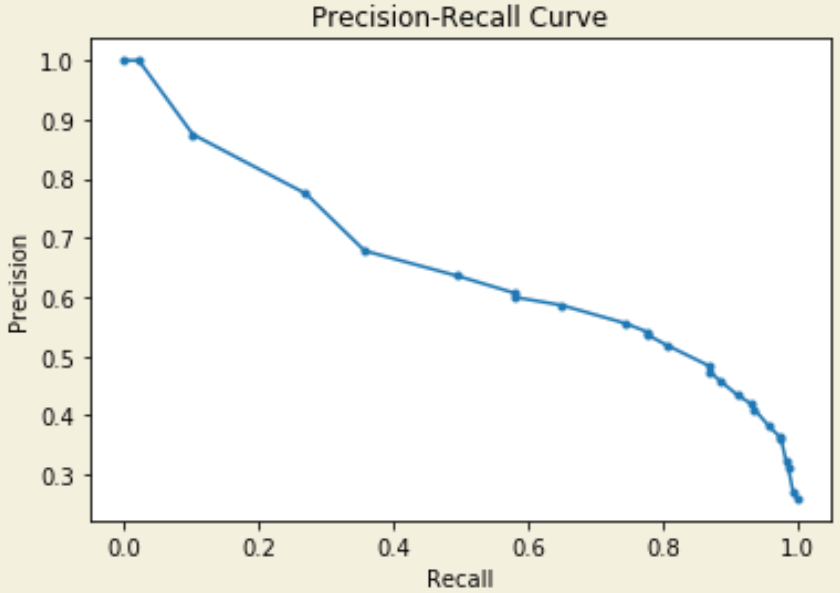
# Feature Engineering

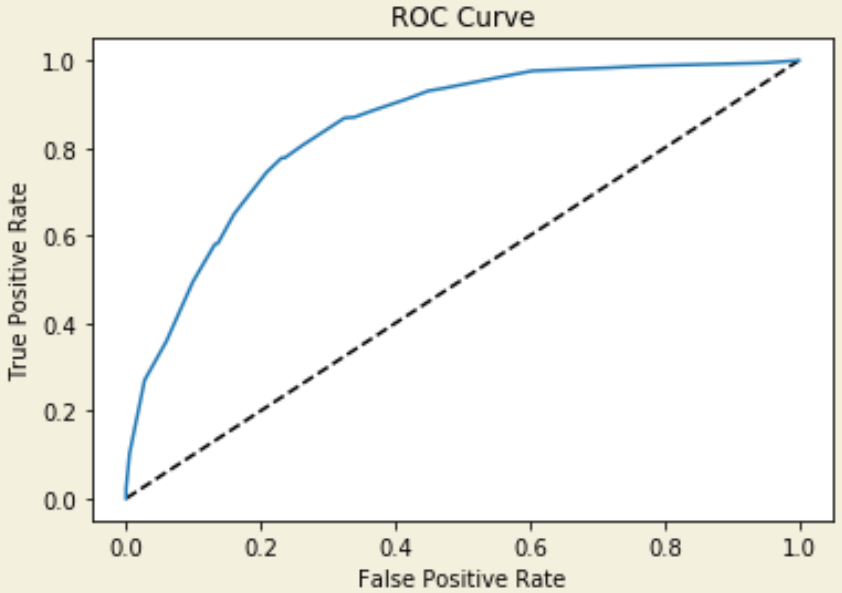


# Model

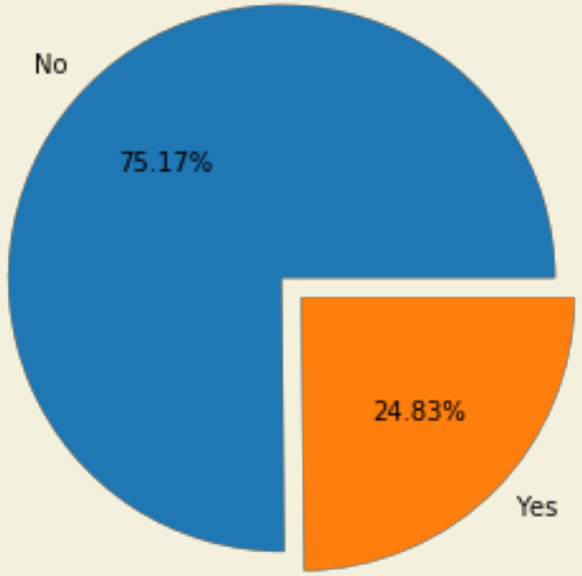
| | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| **Decision Tree** | 79.34% | 0.61 | 0.58 | 0.59 |



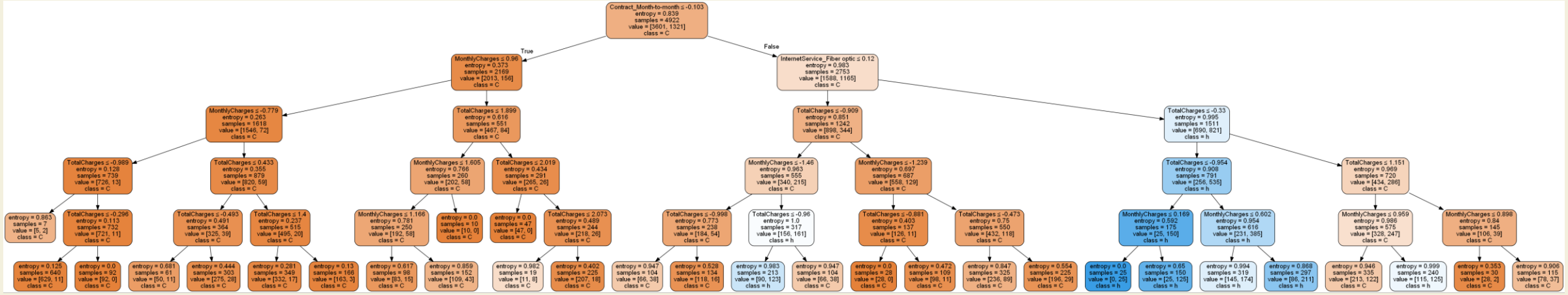



# Predict




y_test, y_predict



# THANK YOU