Machine Learning (CS7620)

Project: System for forecasting customer attrition

Deepak kumar(m22ma004)

About the Dataset:

The dataset has been sourced from IBM Developer Platform which is a sample data of a non-existing telecommunication company. It has 21 features for 7043 customers. It includes target label classifying whether or not a customer left within the previous month and some dependent variables like customer information, location, services they are subscribed to, etc.

The feature set can be divided into 3 parts:

- Customer Account Information:
  - Tenure, Contract , PaperlessBilling , PaymentMethod , MonthlyCharges , TotalCharges
- Services that the customers has signed up for:
  - PhoneService , MultipleLines , InternetService , OnlineSecurity , OnlineBackup , DeviceProtection , TechSupport , StreamingTV , StreamingMovies,
- Customer Personal Information:
  - Gender , SeniorCitizen , Partner , Dependents
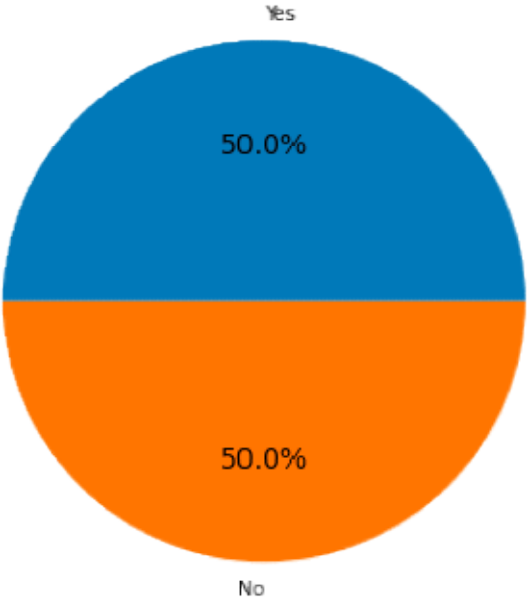
Exploratory Data Analysis
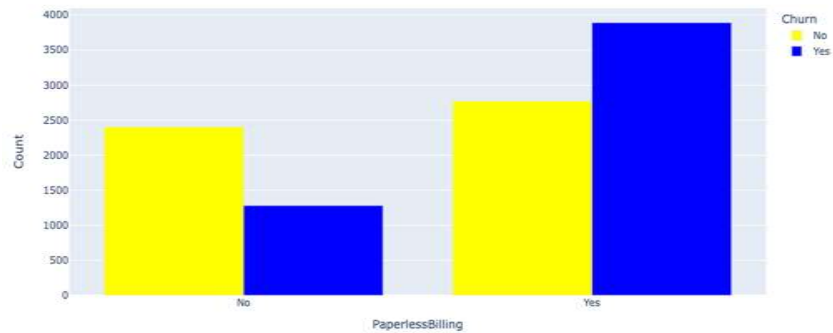
Original Distribution

Churn Distribution

No

73.5%

26.5%

Yes
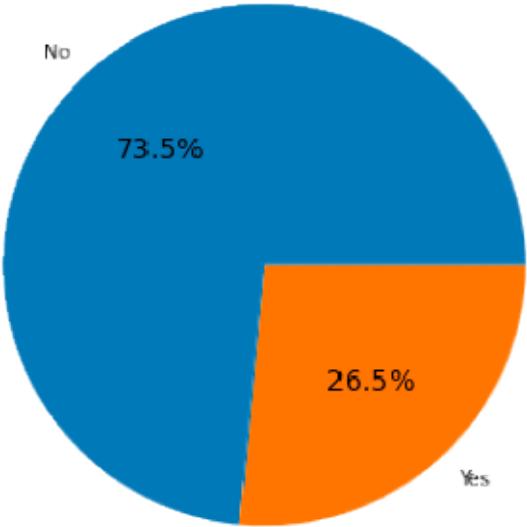
After Downsampling
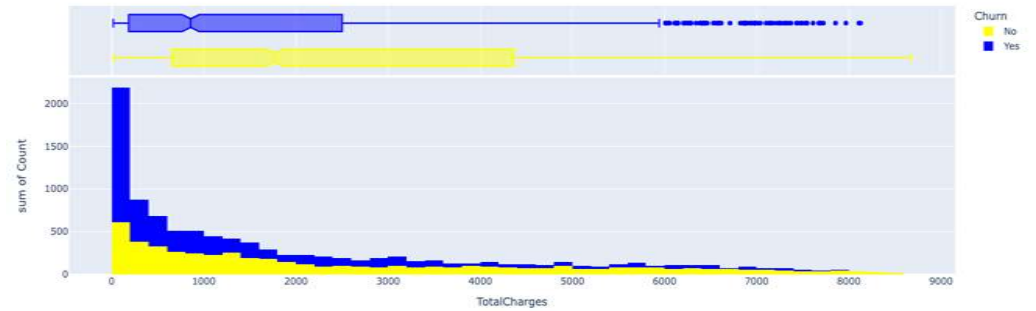
Churn Distribution

Yes

50.0%

50.0%

No

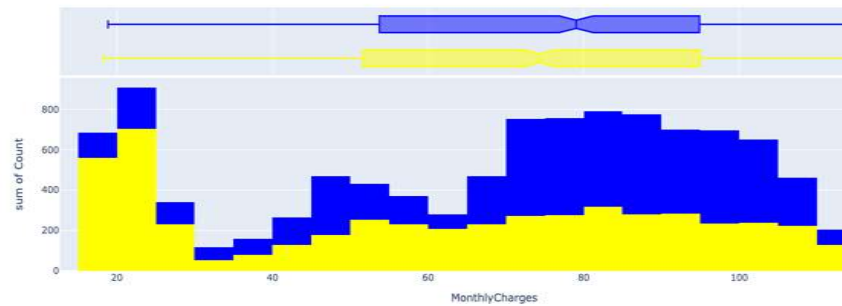# Customer Account Information
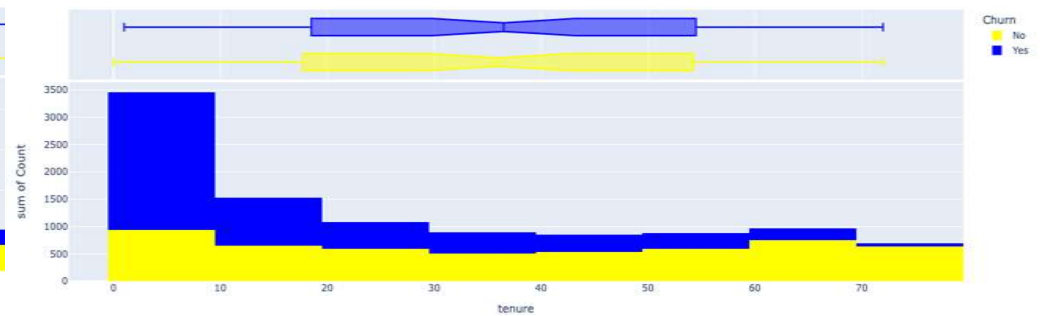
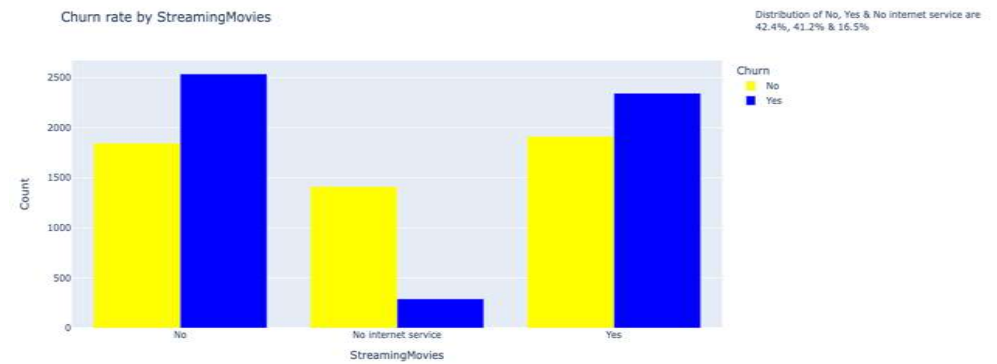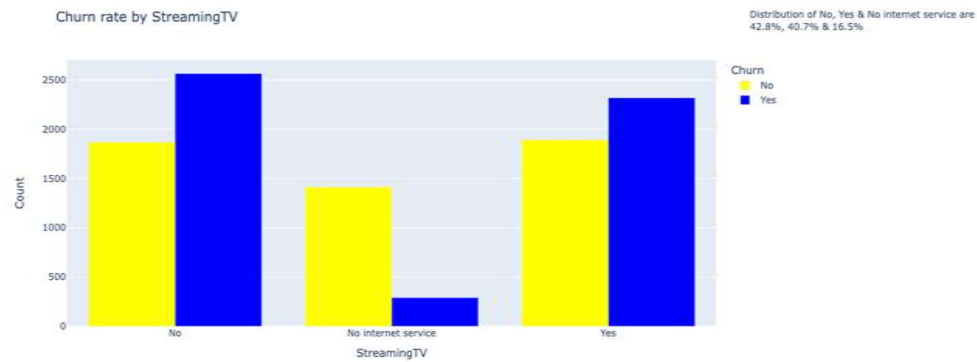### Churn rate by PaperlessBilling

Churn rate frequency to TotalCharges distribution
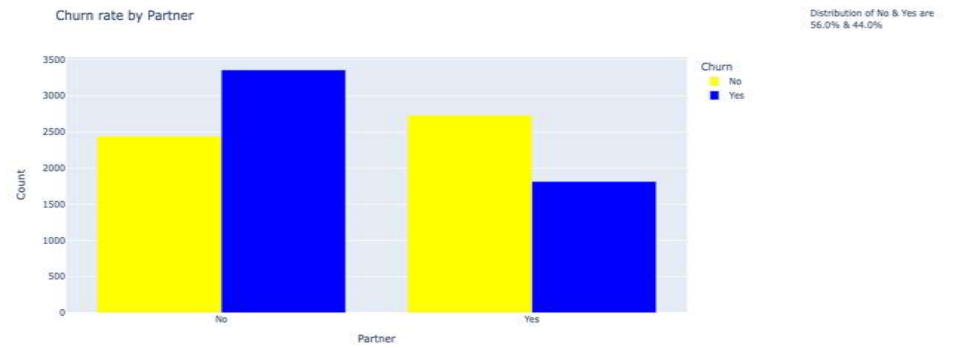

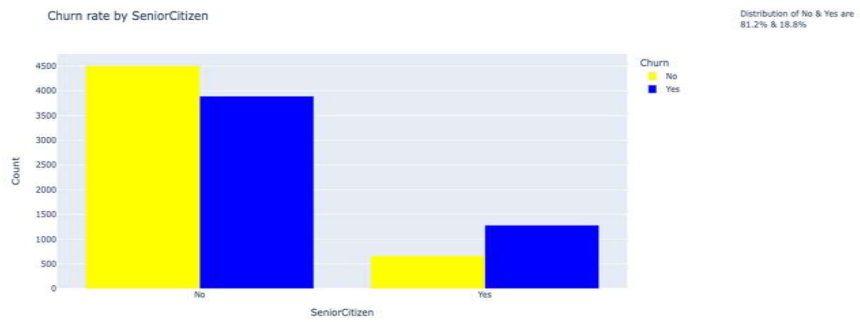### Churn rate frequency to MonthlyCharges distribution

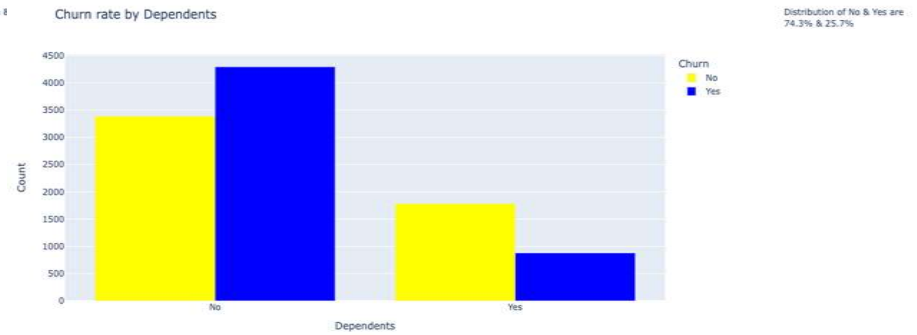### Churn rate frequency to tenure distribution

Services signed up for

Churn rate by StreamingTV

Distribution of No, Yes & No internet service are
42.8%, 40.7% & 16.5%

Churn rate by StreamingMovies

Distribution of No, Yes & No internet service are
42.4%, 41.2% & 16.5%

Personal Information

Churn rate by gender

Distribution of Female &
50.1% & 49.9%

Churn rate by Dependents

Distribution of No & Yes are
74.3% & 25.7%

Churn rate by SeniorCitizen

Distribution of No & Yes are
81.2% & 18.8%

Churn rate by Partner

Distribution of No & Yes are
56.0% & 44.0%

HEATMAP

RESULTS

	Accuracy	Precision	Recall	F1 score
Logistic Regression	0.7665056360708534	0.7418590882178804	0.8125810635538262	0.7660563149633365
Naive Bayes	0.7536231884057971	0.7254788160185722	0.8106355382619974	0.752881143986151
SVM	0.7838969404186795	0.7596899224806202	0.8261997405966277	0.7835515087538022
Decision Tree	0.8608695652173913	0.8153409090909091	0.930609597924773	0.860247303059138
Random Forest	0.8898550724637682	0.8500583430571762	0.9448767833981842	0.8895576491496541