# Churn Analytics

Agenda

- Objective

- Data Insight

- Model selection

- Deployment Steps

- Library used in source code

Objective

- To analyse the customer churn data and build a machine learning models to predict the customer churn .

Data Insight

- Data Shape: (7032, 20)

- Data in class imbalanced: 73% of data of one class.

- Senior Citizen has very high proportion of churn.

- Two year contract has very low churn.
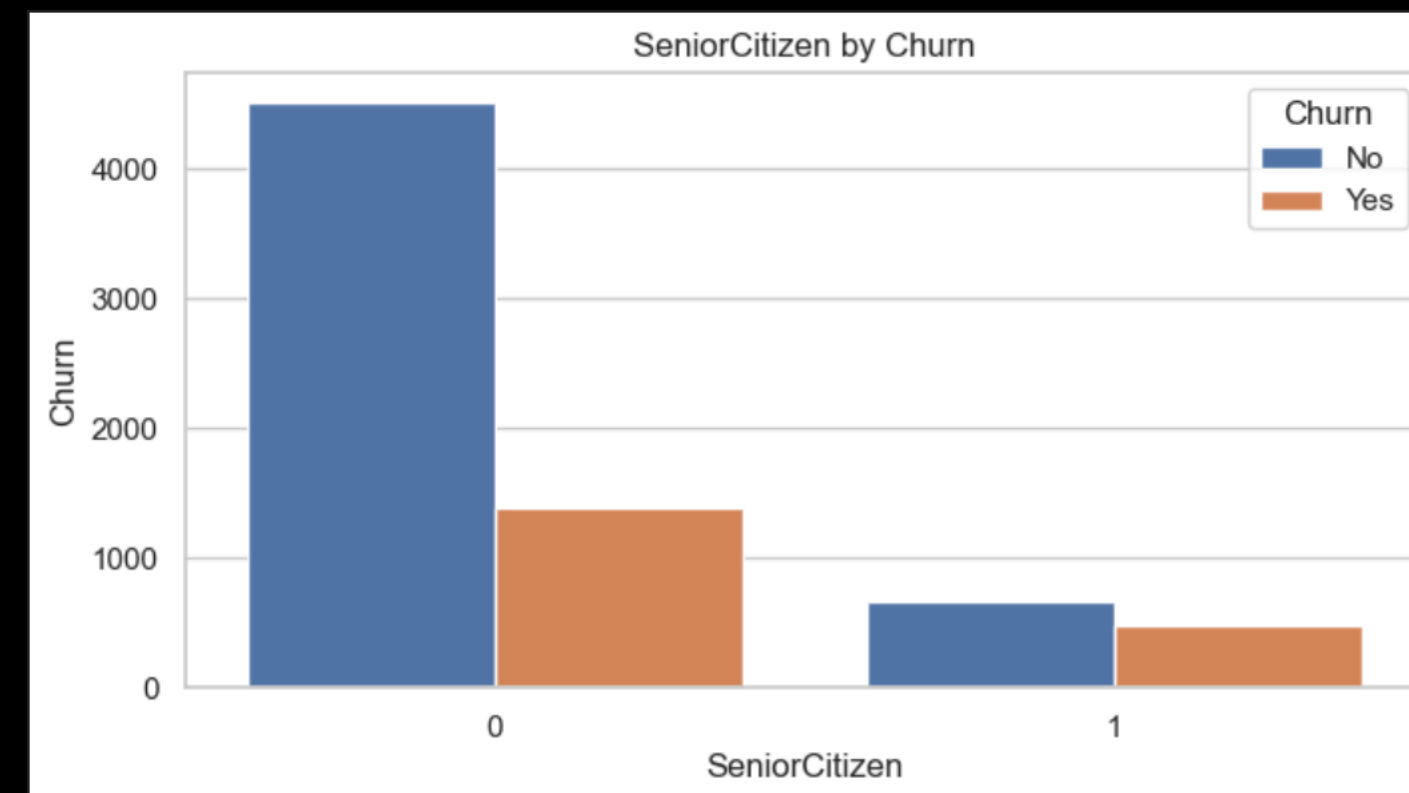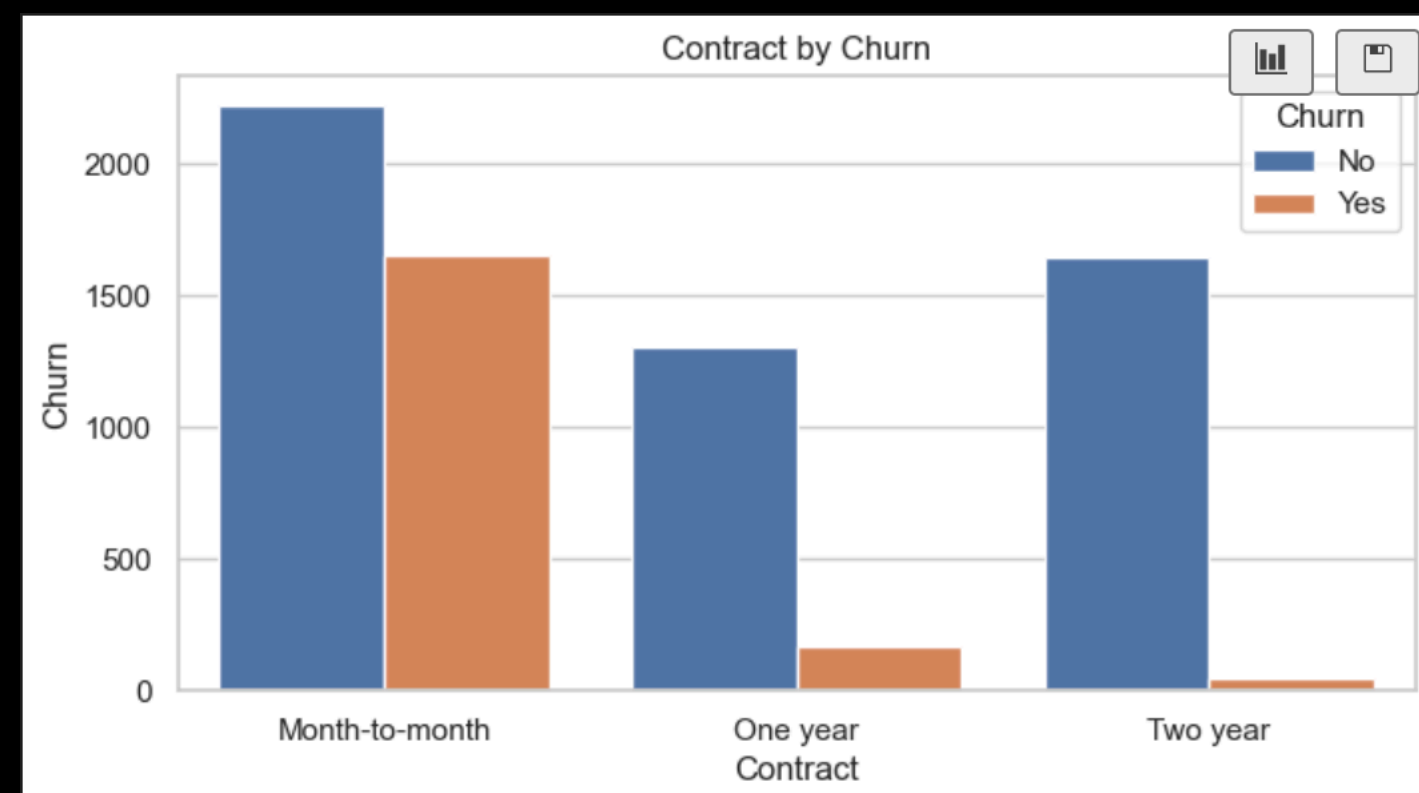
Model Selection

- Label is categorical so model should be classifier.

- Trained two models logistic regression and random forest model.

- Fine Tuned the models with hyper parameters.

- Balanced the data with synthetic data generation of minority class.

Model Evaluation

- Precision and Recall for model evaluation

- Random forest with balanced data worked best:

- Best model scores:

  - Recall : 0.93

  - Precision:0.65

- Random forest model worked well because data has categorical features.

- Random forest models can build non linear boundary.

Deployment Steps

- After training save the model in pickle file

- Upload model in S3 bucket.

- Create a inference script and configuration file.

- Deploy method will give us a end point for inference for real time.

Library used in source code

```python
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
import plotly.express as px
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import precision_score, recall_score
from imblearn.over_sampling import SMOTE
from sklearn.ensemble import RandomForestClassifier
```

Thank You