Churn analysis

Using Survival Curve and Interpretable Machine Learning Model

Kyra (Xiao) Quan

Introduction

Customer churn happens when customers stop doing business with the company. It's very important to understand why customers churn and implement incentive promotion plans to those high risk customers. Because getting new customers is way more expensive than earning the trust and loyalty of existing customers.

In this project, I took the sample dataset from a telecommunication company and applied survival analysis and interpretable machine learning models to analyze the churn problem.

Data Source

- ❑ IBM sample set for practicing data analysis on a real-world type of business problem
- ❑ 7043 observations and 22 variables that contain information about:
  - customer demographics
  - services they signed up for
  - account information
  - churn (target variable)

Preliminary Analysis

Key Points:

- ❑ Inspected dataset
- ❑ Changed datatype of 'TotalCharges’
- ❑ Filled missing values:

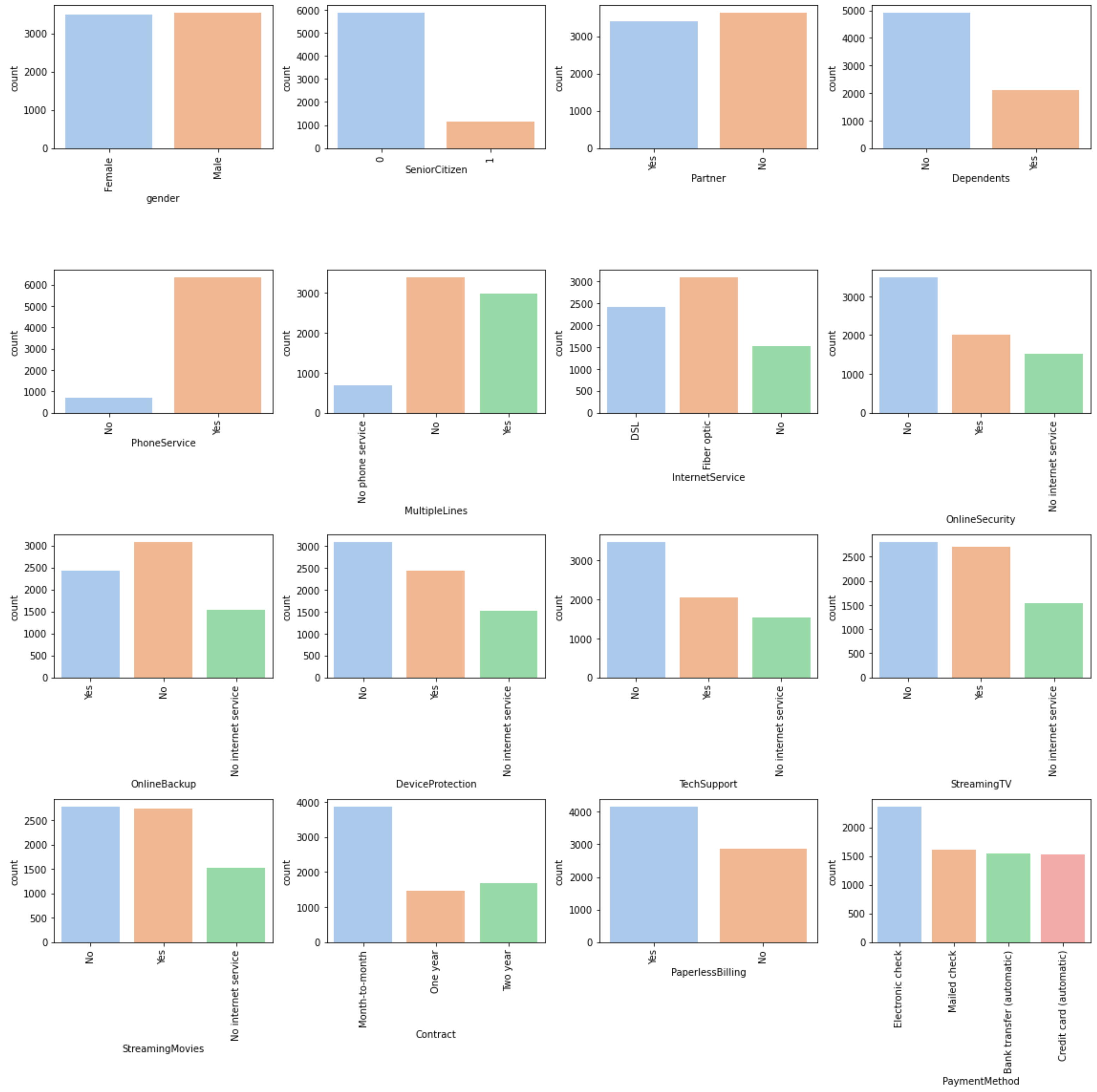
  The missing values in 'TotalCharges' column all have 0 in 'tenure' column, which means they are all new customers. So the missing value in 'TotalCharges' column were set equal to 'MonthlyCharges'.

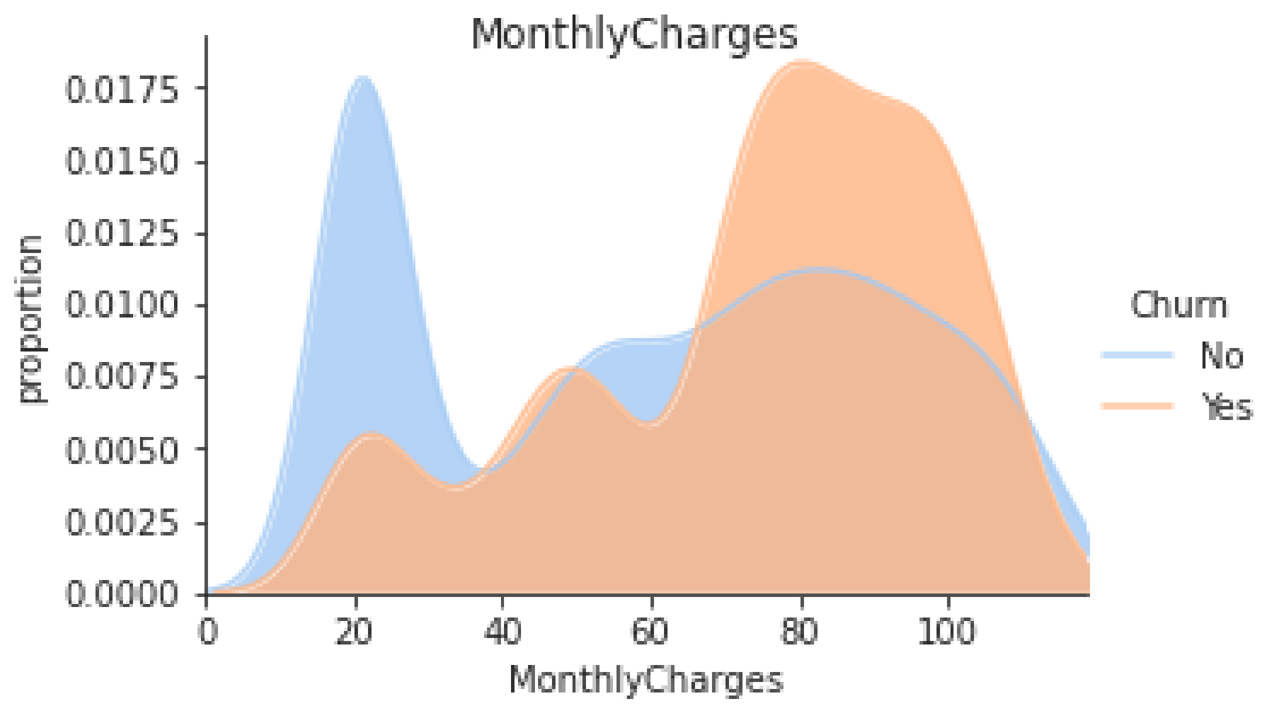
Exploratory Data Analysis

Key Points:

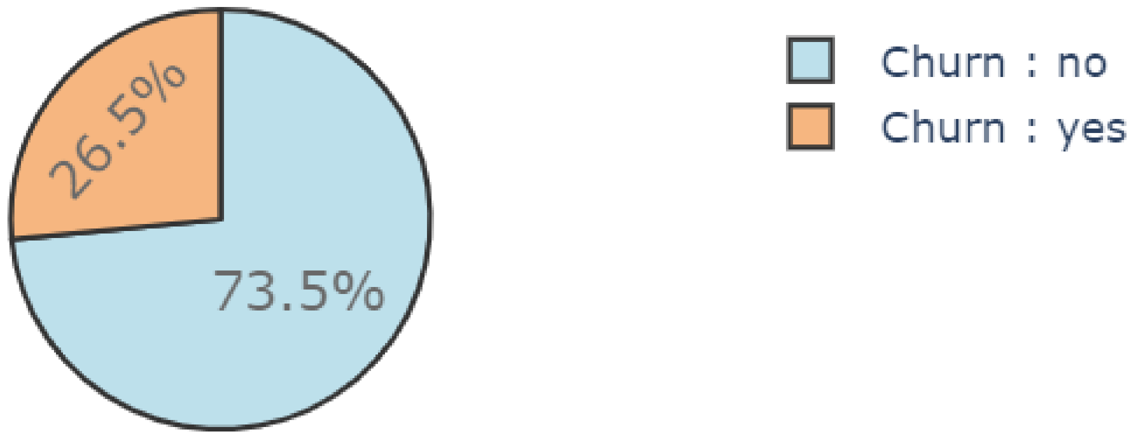
- ❑ Explored categorical and numerical columns
- ❑ Noticed the imbalanced distribution of target variable
- ❑ Transformed categorical columns using one hot encoding
- ❑ Removed collinear features
- ❑ Identified customer churn problem using survival analysis

Exploratory Data Analysis





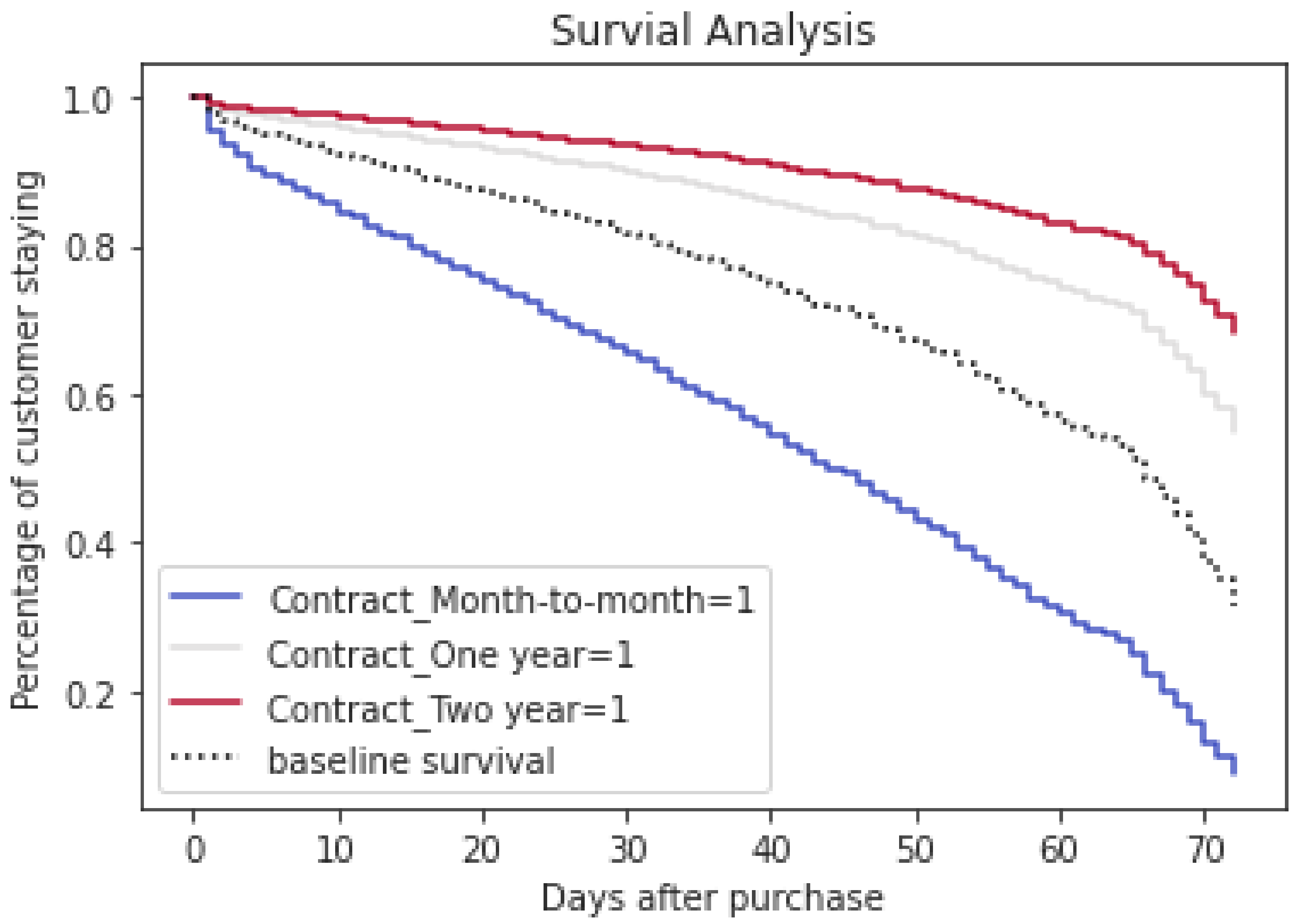
Distribution of target variable



After 50 days of purchase

more than 50%

Month-to-month contract customers churn



Exploratory Data Analysis

Why use survival analysis?

- Strength:
  - Survival Curves - communication tool that easy for business leaders to understand
  - Cox's proportional hazard model - used to incorporate multivariate analysis
- Weakness:
  - Not as high performance as Machine Learning
- Solution:
  - Build Machine Learning model to predict churn propensity

Model Building and Evaluation

Key Points:

- ❑ Resampled imbalanced class for training set
- ❑ Built models using GLM, GBM, Random Forest and autoML
- ❑ Tuned model hyper parameters using Grid Search
- ❑ Chose the right performance matrix to evaluate models

GLM

- Generalized Linear Model refers to a larger class of linear models, which in this case is a Binary Logistic Regression model
- It fits a sigmoid function to data
- Outputs probability which is in [0,1] range

Random Forest

- Constructs multiple decision trees and take the average prediction over all the trees to make a final prediction
- Randomly sample training data when building trees and randomly subset features when splitting nodes
- Avoid overfitting

GBM

- ❑ Gradient Boosting Machine is a forward learning ensemble method
- ❑ Consecutive decision trees with single split are constructed and each tree solves for the net loss of the prior trees.
- ❑ Directly optimize the cost function
- ❑ Tend to overfit

AutoML

- ❑ Automatic Machine Learning
- ❑ Automated stacking (ensembles), neural architecture search, pipeline optimization and feature engineering
- ❑ Improve the efficiency of finding optimal solutions to machine learning problems

Model Building and Evaluation

- Hyper parameters were tuned using grid search
- For imbalanced data, accuracy is not a good evaluation measurement
- The dataset has small positive class, so F1 score is a better measurement than AUC
- Random Forest is selected as the best model among all. Because it has the highest F1 score, and a moderate AUC

	Accuracy	F1	Precision	Recall	AUC
GLM	0.7988	0.6407	0.5786	0.7178	0.8541
Random Forest	0.7818	0.6510	0.5422	0.8143	0.8592
GBM	0.7998	0.6412	0.5806	0.7159	0.8617
AutoML	0.7893	0.6385	0.5590	0.7443	0.8623

Model Interpretation

Why Interpretation Matters?

- ❑ If we can explain churn, we can reduce churn
- ❑ Provide actionable insights to marketing team

Methodology:

- ❑ Local Interpretable Model-Agnostic Explanation (LIME)
- ❑ Partial Dependence Plot (PDP)

LIME

- Explain individual predictions
- Select observation of interest
- Probe the Black Box with permuted samples of training data
- Weight sample by proximity to sample of interest
- Train an interpretable model on weighted

PDP

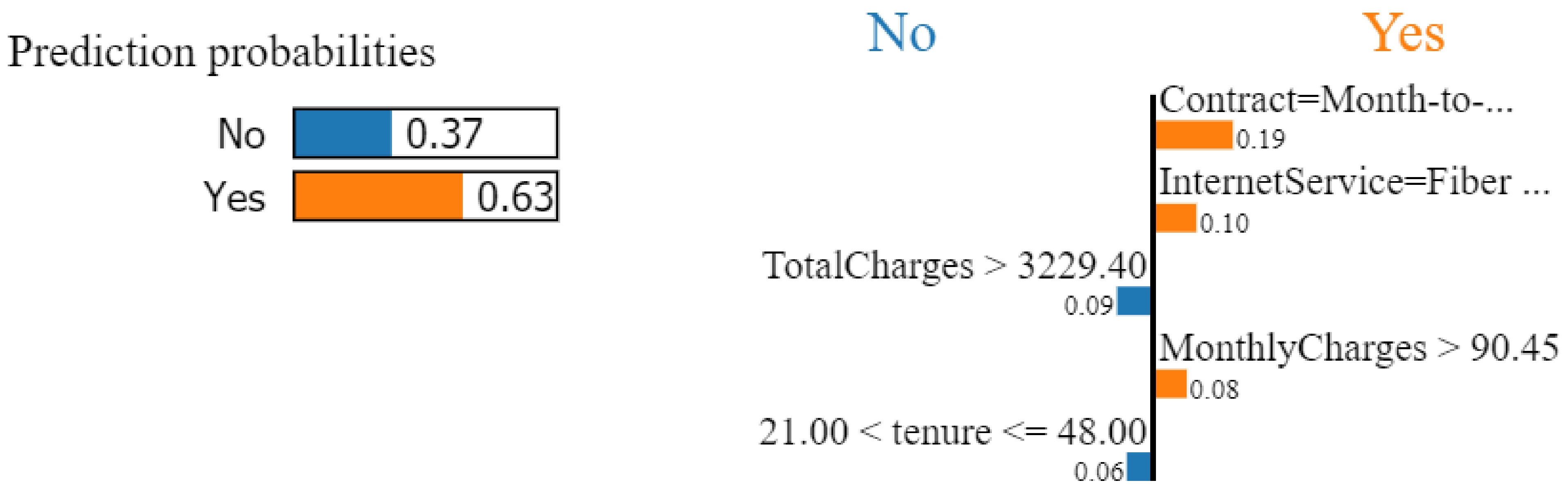
- Partial Dependence Plots
- Explain model on global level
- Show how the expected model response for random observations
- Hold all other features constant and vary feature of interest
- Then Average Results

Model Interpretation

Use Case – Customer A

Why does the model predict this customer is more likely to churn?

- ❑ We can see that it is attributed to his contract type, internet service type , total charges, monthly charges and tenure. The fact that he's paying 3741.85 of total charges and he's been a loyal tenure for 38 years has tried to pull down his likelihood of churn, but overall the model has decided that this person is 63% likely to churn.



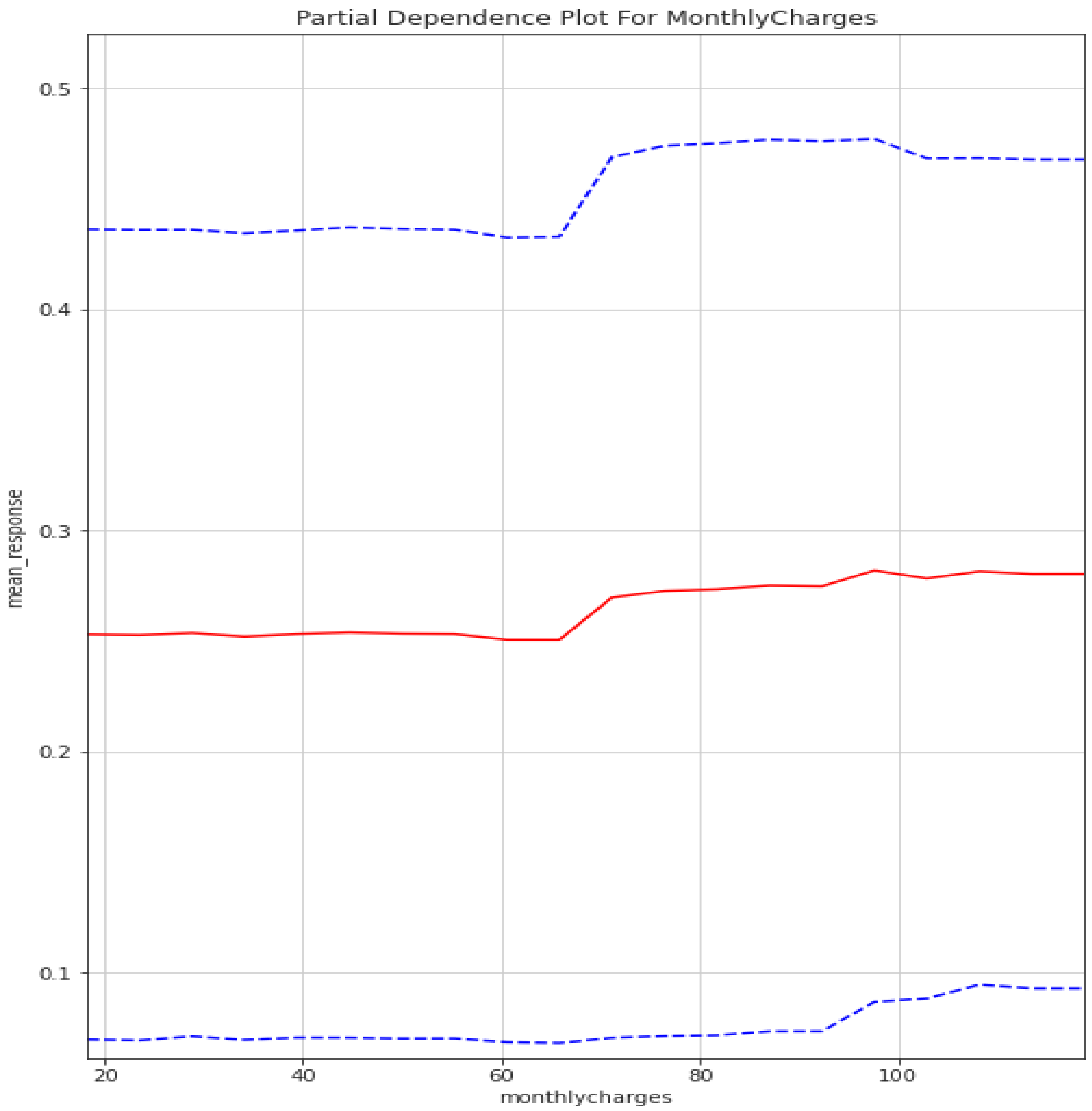
Feature	Value
Contract=Month-to-month	True
InternetService=Fiber optic	True
TotalCharges	3741.85
MonthlyCharges	101.15
tenure	38.00

Model Interpretation

Use Case – Customer A

What actions can we take to stop this customer from churning?

- ❑ Lime chart tells us $101.15 monthly charge is one of the reasons that make the customer leave
- ❑ PDP shows in general, our customers have a lower probability to churn when the monthly charge is below 70
- ❑ We could try to offer customized monthly discount to this customer



Summary

Insights:

- Customers with high monthly charges and low tenure are more likely to churn, different service types have different effect on Churn
- Month-to-month contract customers are churning fast

What to do next:

- Further investigate the model, apply hypotheses tests to identify the non-effective attributes.
- Investigate effects of different service type

Recommendations:

- Offer customized promotions to selected customers
- Apply A/B test on those promotion strategies
- Get customers sign up for services that have a negative effect on churn