Telco Customer Churn

Project description

- Our project aims to analyze customer churn in the telecommunications industry and provide insights that could help businesses reduce their churn rates and improve customer retention, ultimately leading to increased profitability and customer satisfaction. Through exploratory data analysis and machine learning techniques, we seek to identify patterns and trends in customer behavior that can be utilized to predict and prevent churn.

- The project centers on analyzing the phenomenon of customer churn in the telecommunications industry, a ubiquitous problem faced by companies in this sector. Leveraging the Telcom Customer Churn dataset, our analysis aims to uncover valuable insights that can aid businesses in reducing their churn rates and enhancing customer retention, ultimately contributing to improved profitability and customer satisfaction.

- To this end, we will conduct a comprehensive exploratory data analysis to identify intricate patterns and relationships within the data. We will utilize various machine learning techniques, including logistic regression, support vector machines, k-nearest neighbors, decision trees, and random forests, to develop and evaluate predictive models that enable us to ascertain factors that contribute to customer churn in the telecommunications industry. Through the analysis of this dataset, we hope to provide actionable recommendations that empower telecom companies to proactively retain their customers and boost their bottom line.

Research Questions?

- What factors have the most significant correlation with customer churn in the telecommunications industry, and how precisely can these factors be utilized to predict which customers are most susceptible to churn in the future?

- To address this question, we will conduct an extensive exploratory data analysis to identify pertinent patterns and correlations within the dataset. This analysis will enable us to uncover the most salient factors driving customer churn, such as customer demographics or services used, which we will then leverage to develop and evaluate predictive models. Through assessing the accuracy of these models, we can ascertain the potential effectiveness of using these factors to predict customer churn, which can ultimately support businesses in targeting high-risk customers to improve retention rates.

Project steps:

Project Overview

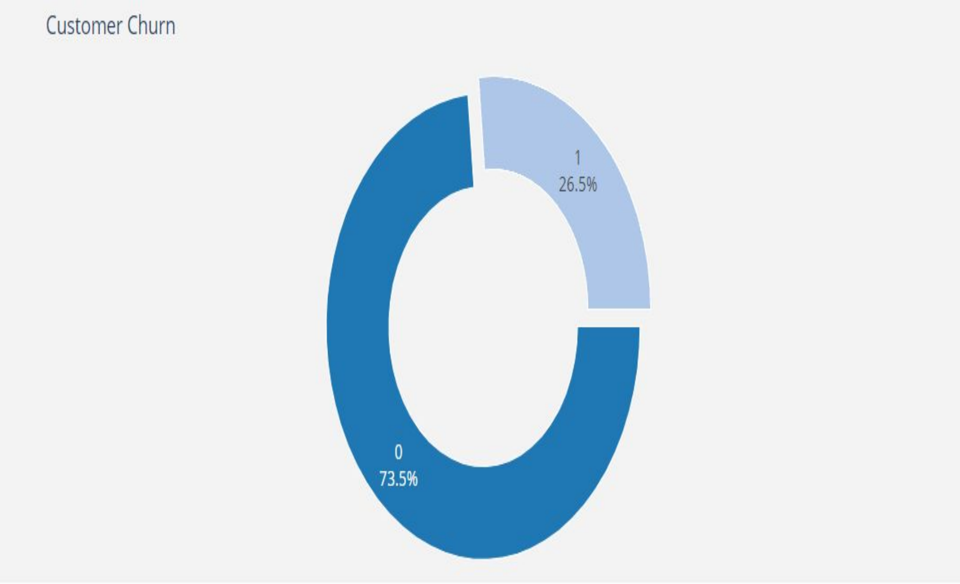
Research Questions

Customer Churn Analysis

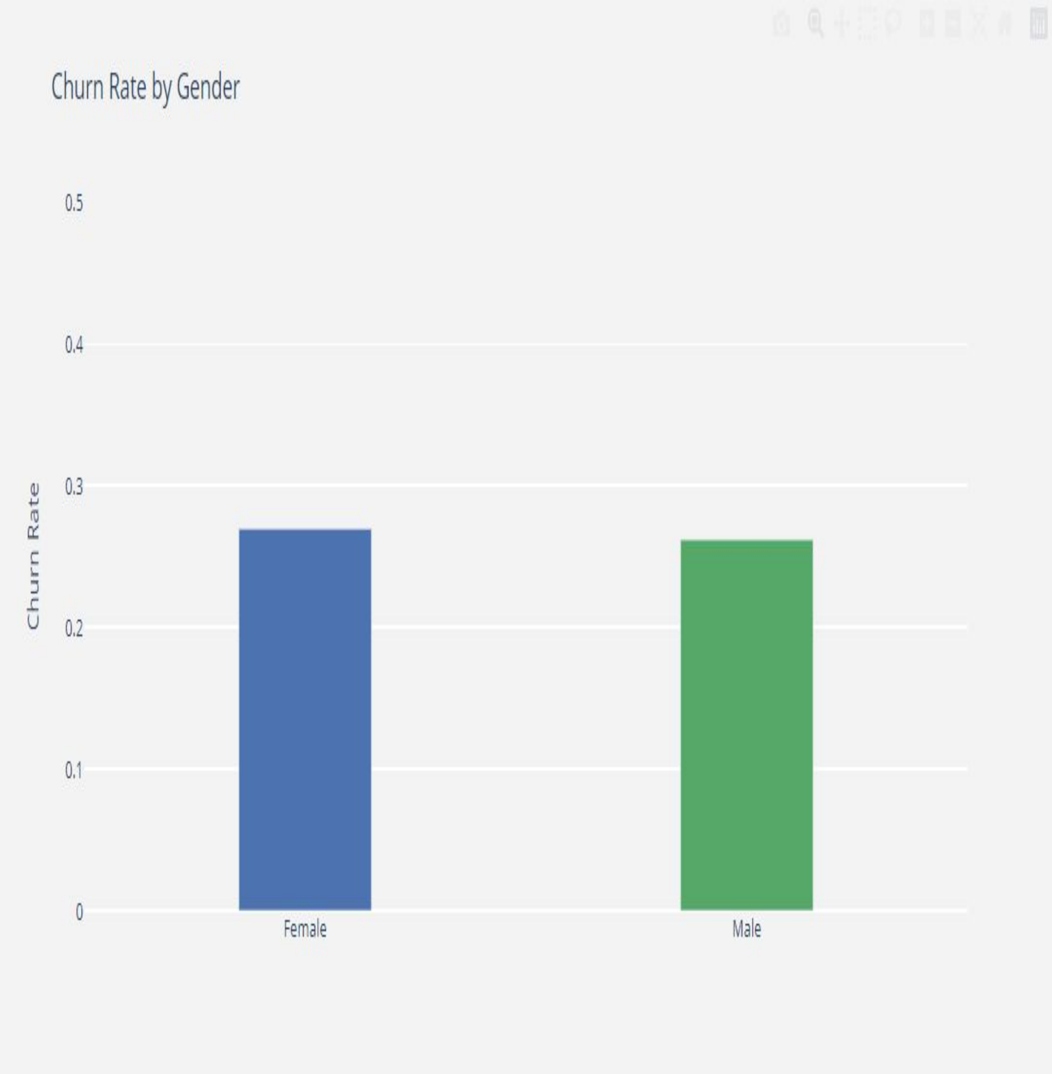
Data Exploration /Analysis

Analysis:

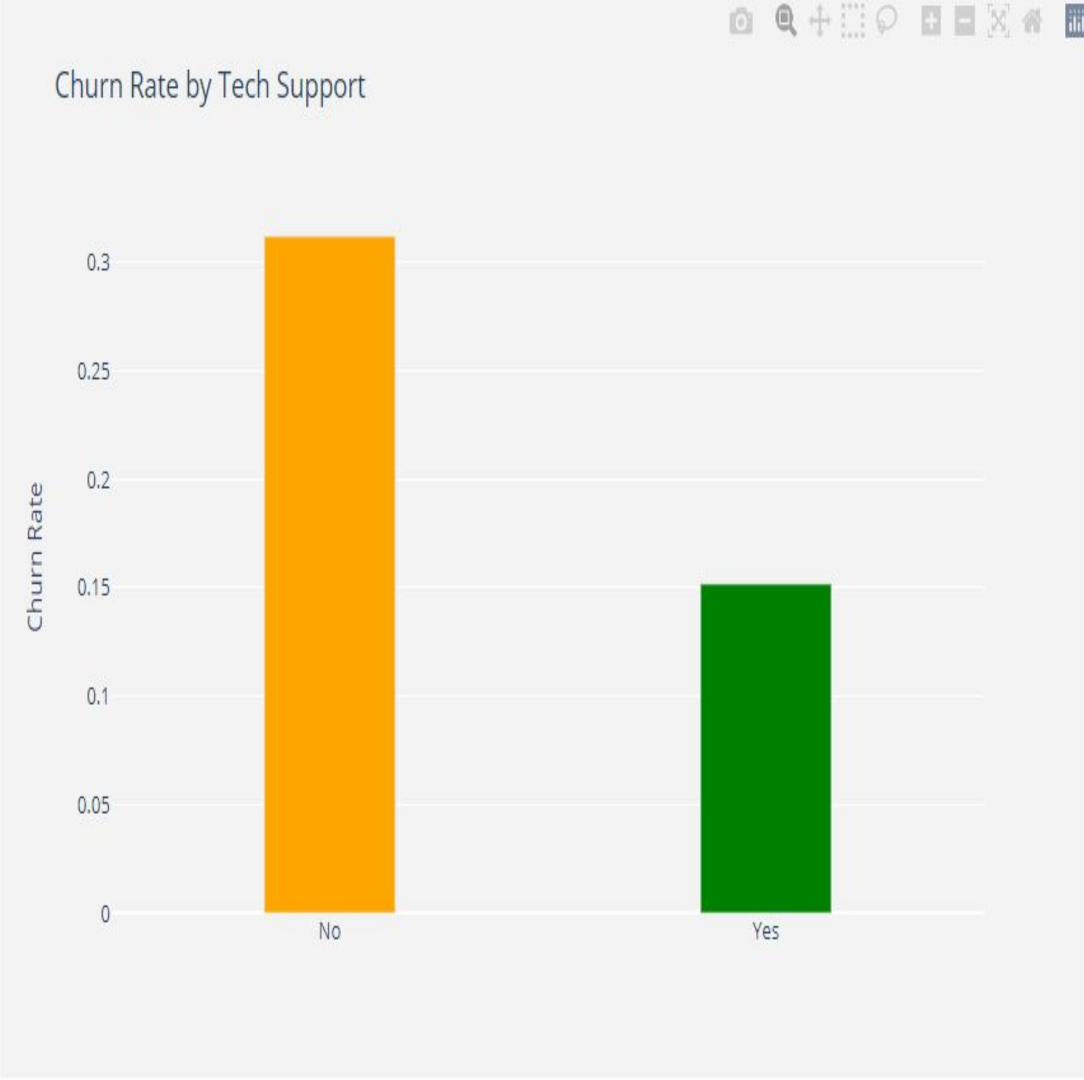
Our analysis includes several visualizations aimed at effectively conveying our findings, by using several machine learning models on the dataset, including logistic regression, support vector machines, k-nearest neighbors, decision trees, and random forests. We compared the models based on their accuracy, and the logistic regression model was selected as the best performer.



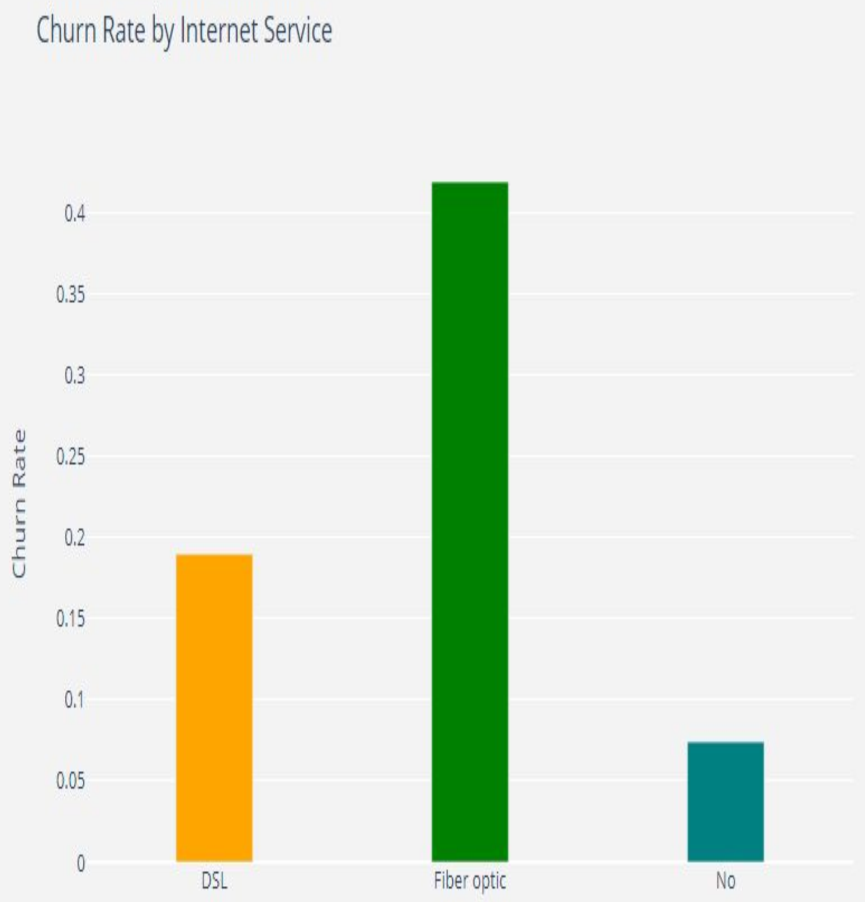
Churn Rate by Gender:



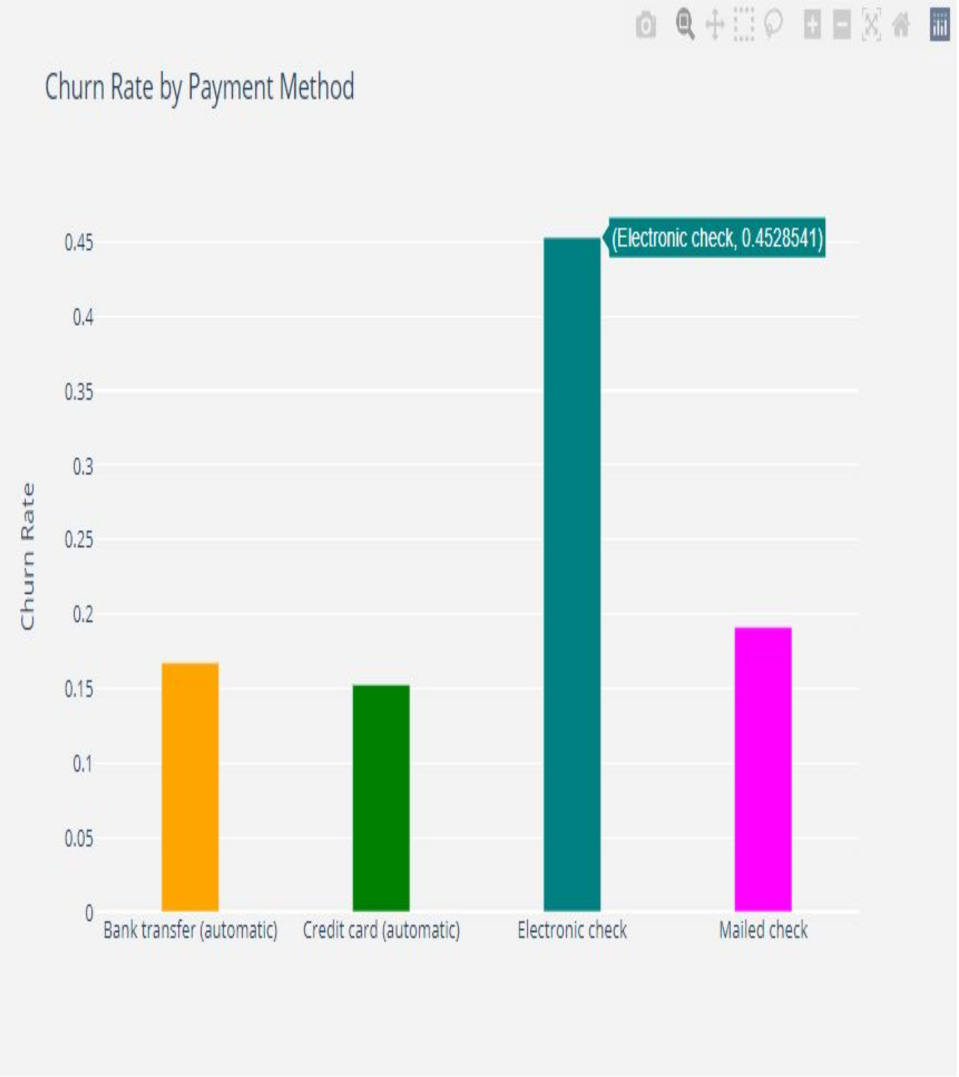
Churn Rate by Tech Support Bar Graph:



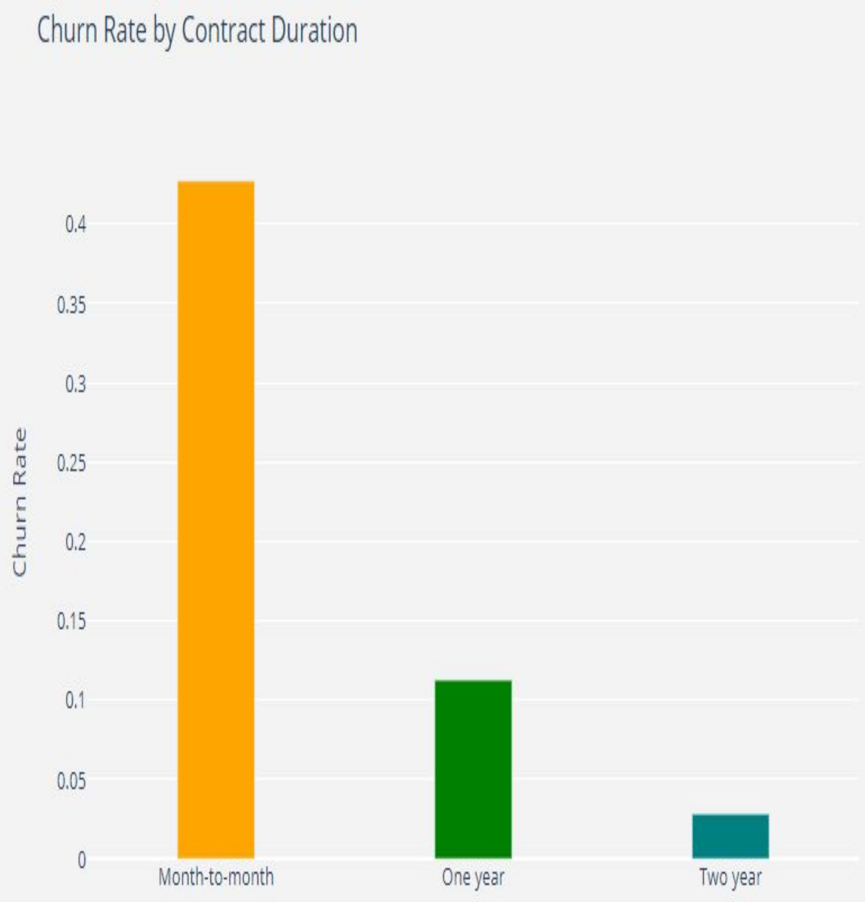
Churn Rate by Internet Service:



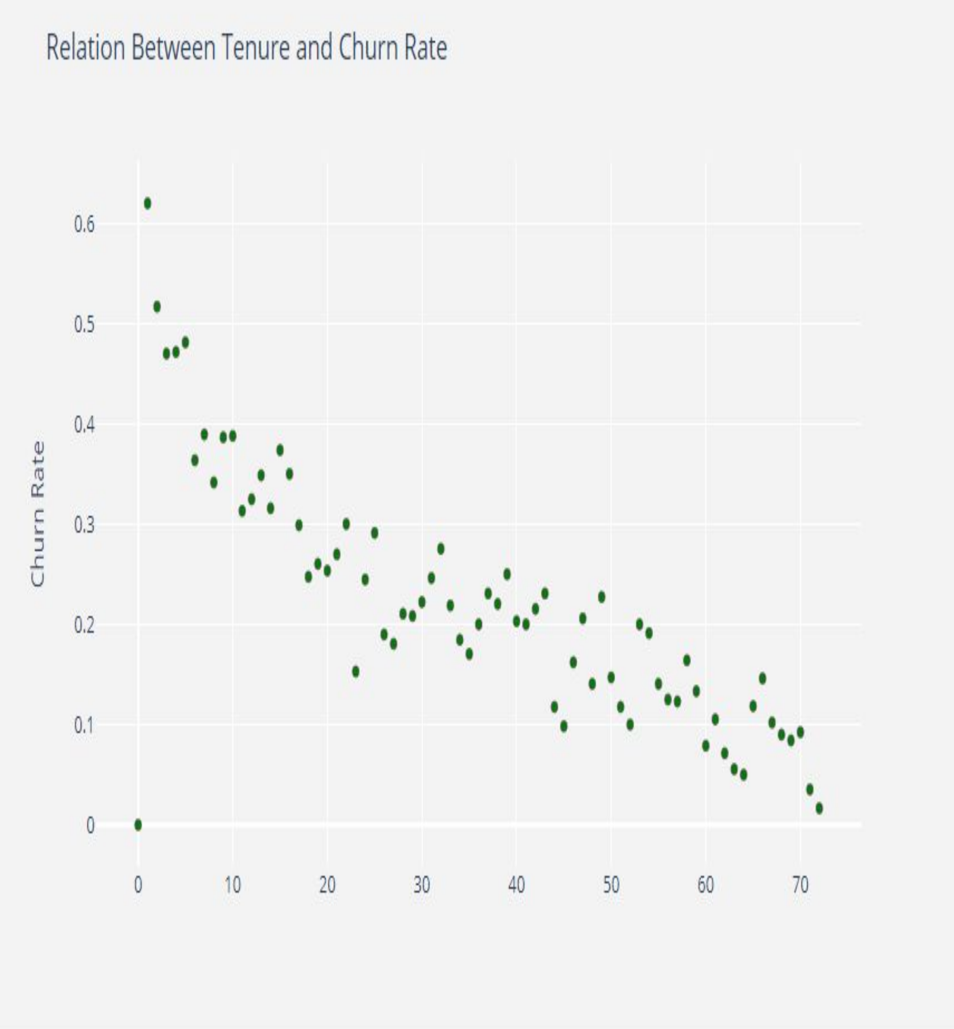
Churn by Payment Method Bar Graph:



Churn Rate by Contract Duration:



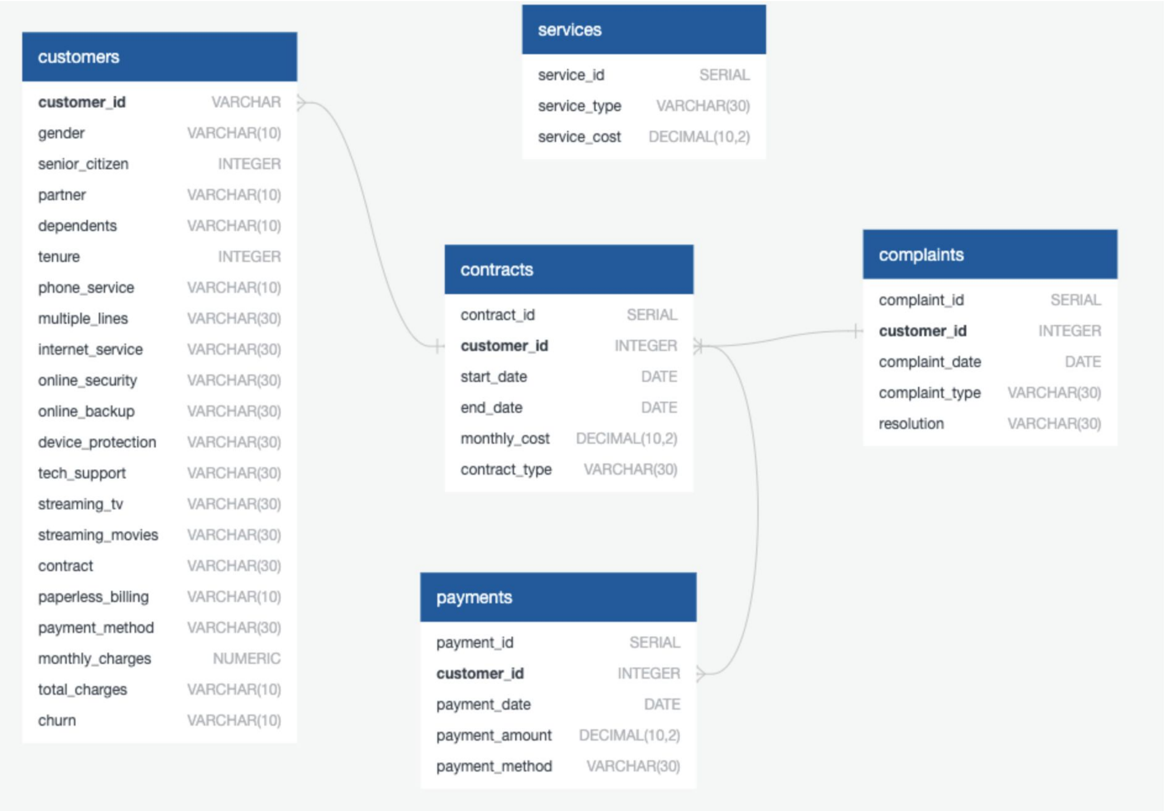
Relation Between Tenure and Churn Rate



Questions to discuss:

In this project, we addressed the following questions:

- What is the overall churn rate in the dataset?
- How does churn rate vary by gender?
- How does churn rate vary by tech support usage?
- How does churn rate vary by internet service?
- How does churn rate vary by payment method?
- How does churn rate vary by contract duration?
- Is there a relationship between tenure and churn rate?



Technologies Used

- We explored the dataset using various visualizations, including pie charts, bar graphs, and scatter plots. We analyzed the distribution of the features and cleaned any missing data. We also used several technologies in our analysis, including Python, Jupyter Notebook, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn, one-hot encoding, feature scaling, logistic regression, support vector machines, k-nearest neighbors, decision trees, and random forests

Resources:

We utilized the Telco Customer Churn dataset available on Kaggle (https://www.kaggle.com/datasets/blastchar/telco-customer-churn).

Team members:

Issa Olmedo, Marley Amisial, Kim Chung, and Christian Perez