Customer Churn Prediction Project Phase 3 - Documentation

Phase 3: Development Part 1

In Phase 3, we will begin building the Customer Churn Prediction project. This phase focuses on loading and preprocessing the dataset, defining analysis objectives, and using IBM Cognos for data visualization.

Step 1: Dataset Download

- Access the dataset from the provided Kaggle link: [Telco Customer Churn Dataset](https://www.kaggle.com/datasets/blastchar/telcocustomerchurn).
- Download the dataset to your local working directory or preferred location.

Step 2: Loading the Dataset

- Import the necessary Python libraries, including Pandas, for data analysis.
- Load the dataset into a Pandas DataFrame for further analysis.
- Display the first few rows of the dataset to inspect the data structure.

Step 3: Exploratory Data Analysis

- Perform an initial exploration of the dataset to understand its structure and features.
- Check for missing values, data types, and basic statistics.
- Visualize key features to gain insights into the data.
- Identify potential relationships or correlations between features and the target variable 'Churn.'

Step 4: Define Analysis Objectives

Define the objectives of the analysis for this phase, including:

- Churn Prediction: Develop a machine learning model to predict customer churn.
- Key Churn Drivers: Identify factors contributing to customer churn.
- Retention Strategies: Propose strategies to reduce churn.

Step 5: Data Cleaning and Preprocessing

Clean and preprocess the data to ensure its quality and suitability for analysis:

- Handle missing values by imputing or removing them based on data exploration.
- Encode categorical variables if necessary.
- Perform feature engineering (create new features) if required.

- Split the data into training and testing sets for model development and evaluation.

Step 6: IBM Cognos for Visualization

Utilize IBM Cognos for creating data visualizations that provide insights into the dataset:

- Create various visualizations, such as bar charts, line charts, heatmaps, and interactive dashboards.
- Visualize the distribution of 'Churn,' feature importance, and other relevant insights using IBM Cognos.

Step 7: Data Validation

Validate the processed data to ensure its quality and accuracy:

- Perform data validation checks to confirm data consistency and accuracy.
- Identify and address any potential data quality issues.

Step 8: Documentation

Maintain comprehensive documentation of the activities performed in this phase:

- Record the data preprocessing steps, including any transformations and cleaning.
- Document the objectives defined for this phase.
- Capture insights gained from data visualization using IBM Cognos.
- Document any observations, challenges, or discoveries made during this phase.

This documentation outlines the key steps and activities to be performed in Phase 3 of the Customer Churn Prediction project. It serves as a guide for the project's development, dataset preparation, and the use of IBM Cognos for data visualization.

DashBoard

Total Charges

16.1M

Tenure

228K

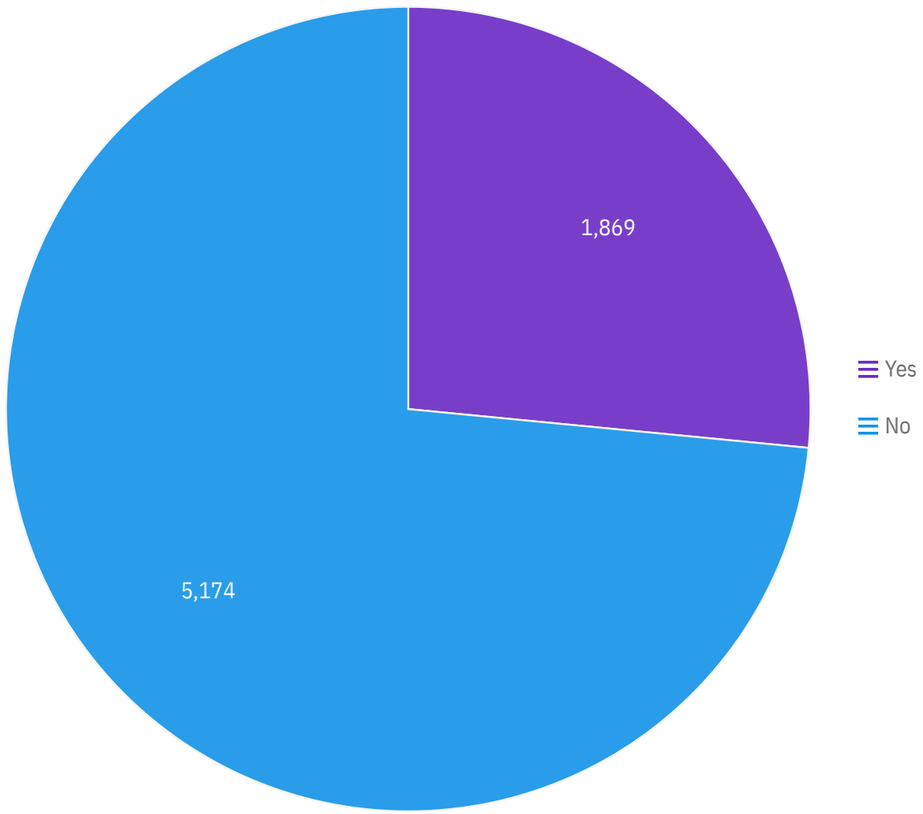
Sr. Citizen Users

1142

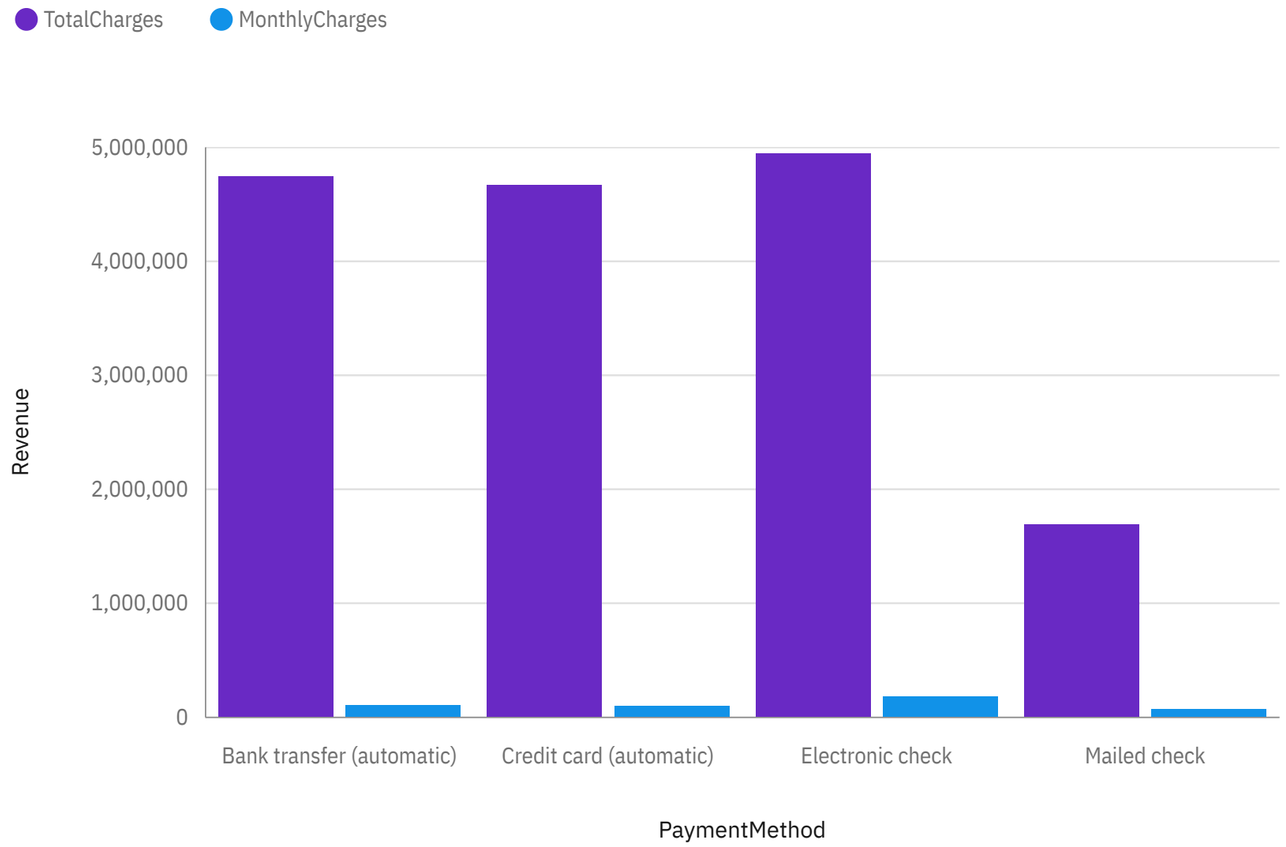
Churn 1

1.87K

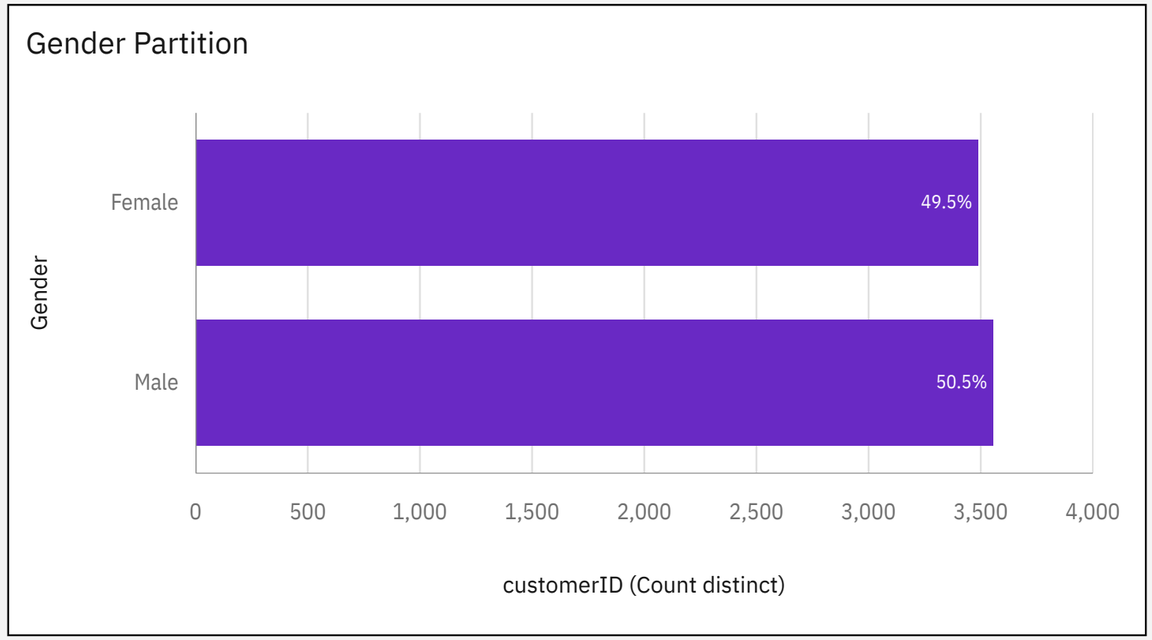
Churn Ratio

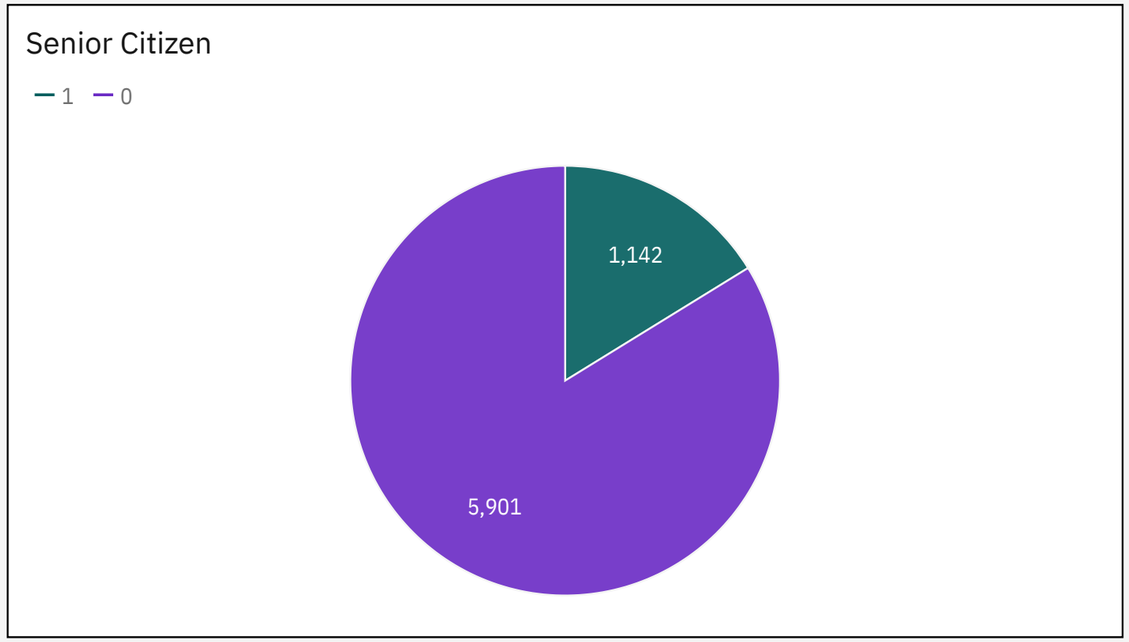


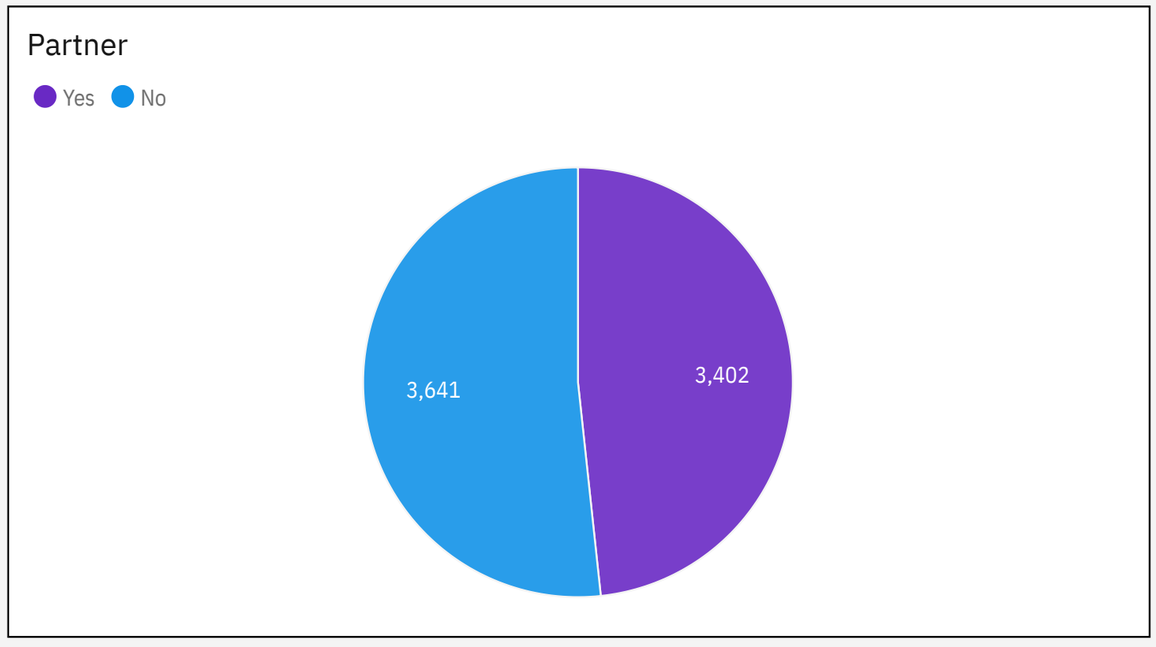
Charges Vs Payment Method

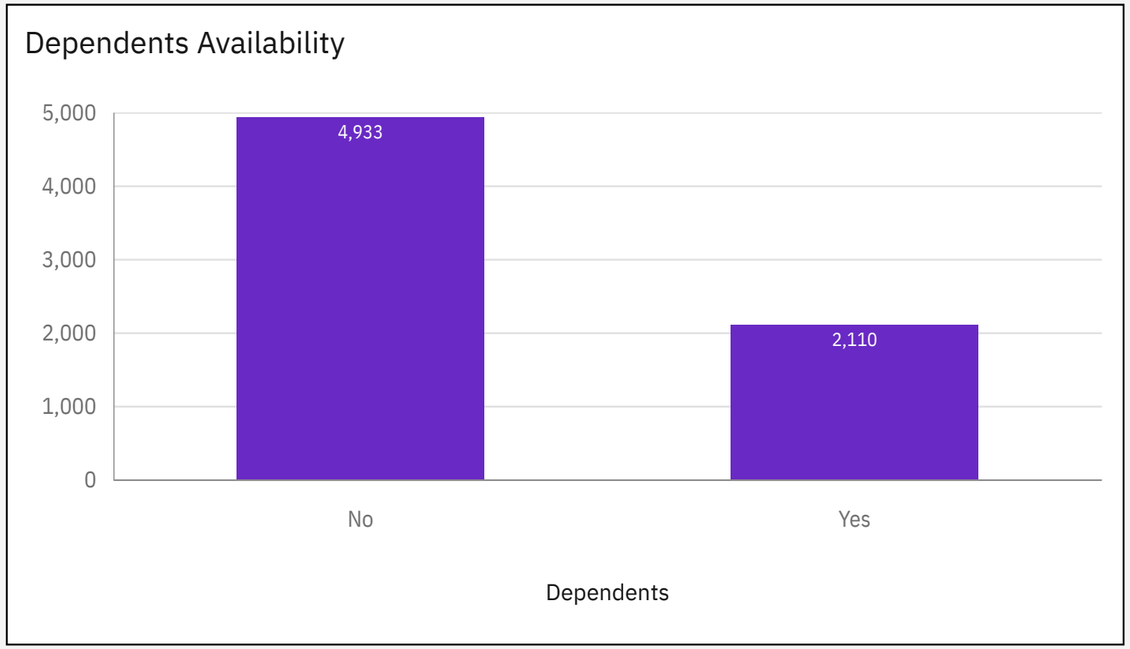


Analytics_1


Gender Partition
Female
Male
49.5%
50.5%
Gender
0
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
customerID (Count distinct)

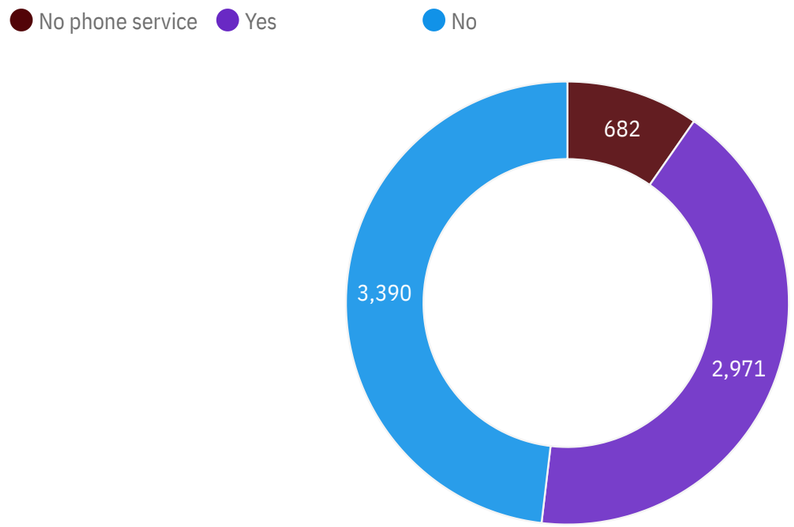

Senior Citizen
1
0
1,142
5,901


Partner
Yes
No
3,641
3,402

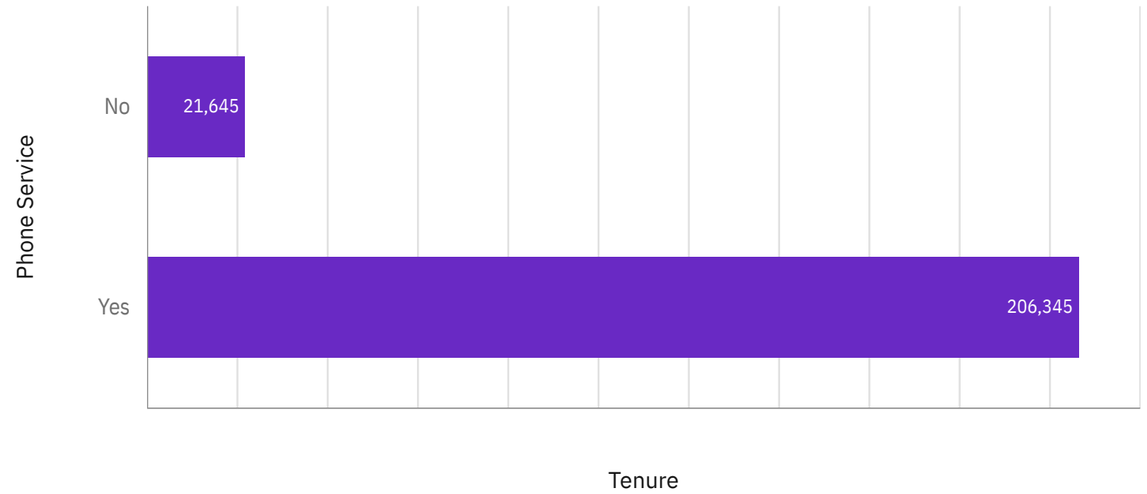

Dependents Availability
5,000
4,000
3,000
2,000
1,000
0
4,933
2,110
No
Yes
Dependents

Analytics_2

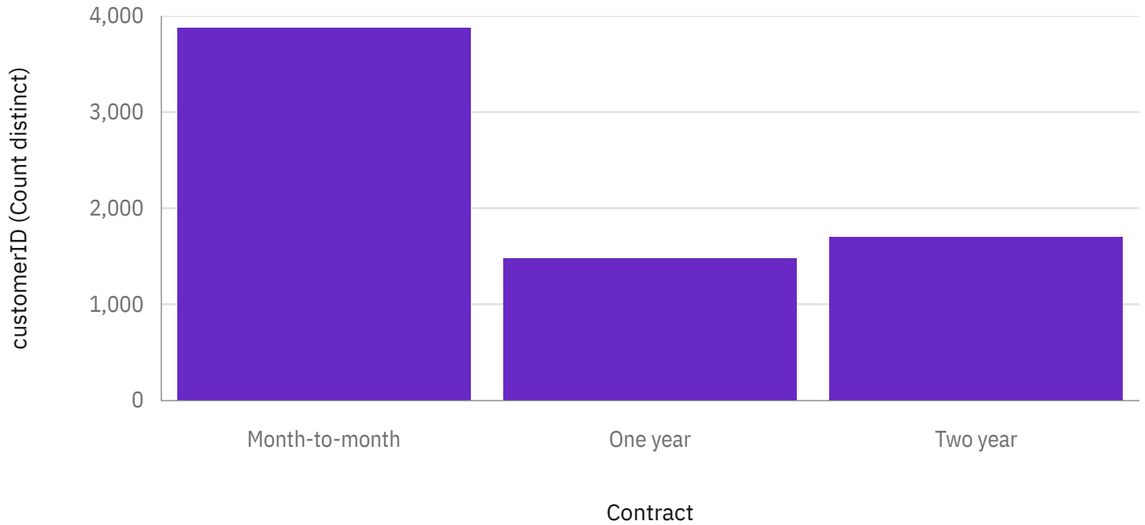
Multiple Lines Usage



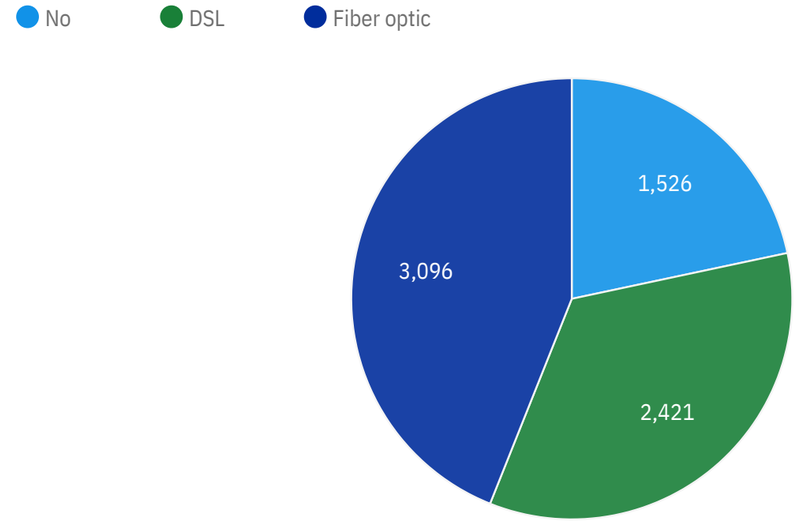
Tenure vs Phone Service



Contract

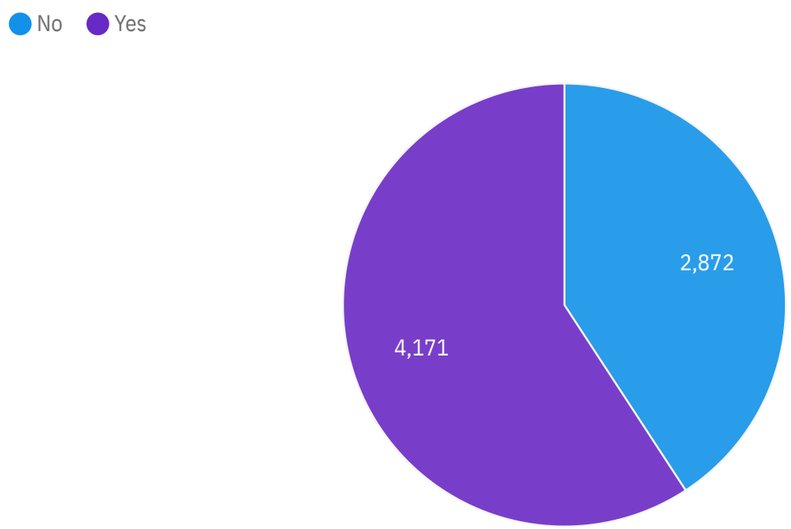


Internet Service Usage

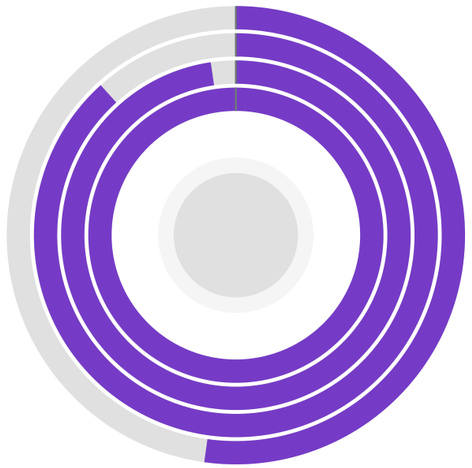


Analytics_3

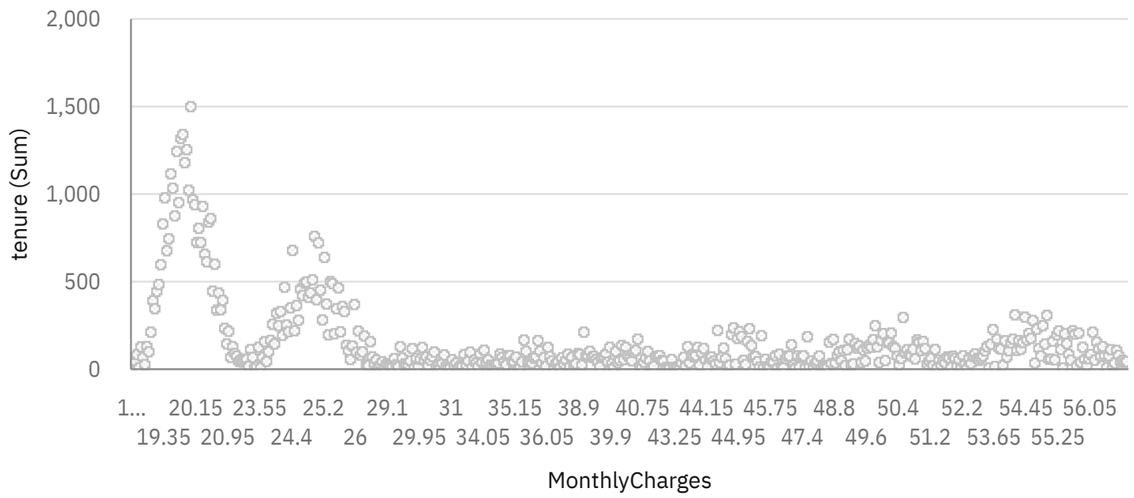
Paperless Billing



Payment Method



MonthlyCharges, tenure



TotalCharges, tenure