Customer Churn Prediction Project Phase 4 - Documentation

Phase 4: Development Part 2

Churn prediction using logistic regression is a common application in machine learning and data analysis. Churn prediction is about identifying and predicting when customers are likely to leave or "churn" from a service, such as a subscription, product, or platform. Logistic regression is a simple and effective method for binary classification problems like churn prediction, where you want to predict whether a customer will churn (1) or not (0).

Here are the general steps to perform churn prediction using logistic regression:

Data Collection and Preparation:

- Collect historical data on customer interactions, behavior, and characteristics.
- Preprocess and clean the data, handling missing values, and converting categorical variables into numerical format.

Feature Selection and Engineering:

- Identify and select relevant features (independent variables) that are likely to influence churn.
- Create new features if necessary, like customer tenure, usage patterns, etc.

Data Splitting:

- Split your dataset into training and testing sets. The training set will be used to train the model, and the testing set will be used to evaluate its performance.

Logistic Regression Model:

- Build a logistic regression model. You can use libraries like scikit-learn in Python for this purpose.
- Fit the model to the training data.

Model Evaluation:

- Evaluate the model's performance on the testing dataset using metrics such as accuracy, precision, recall, F1-score, and the ROC curve.
- Tune hyperparameters if necessary to improve model performance.

Interpret the Model:

- Analyze the model coefficients to understand which features are the most influential in predicting churn.

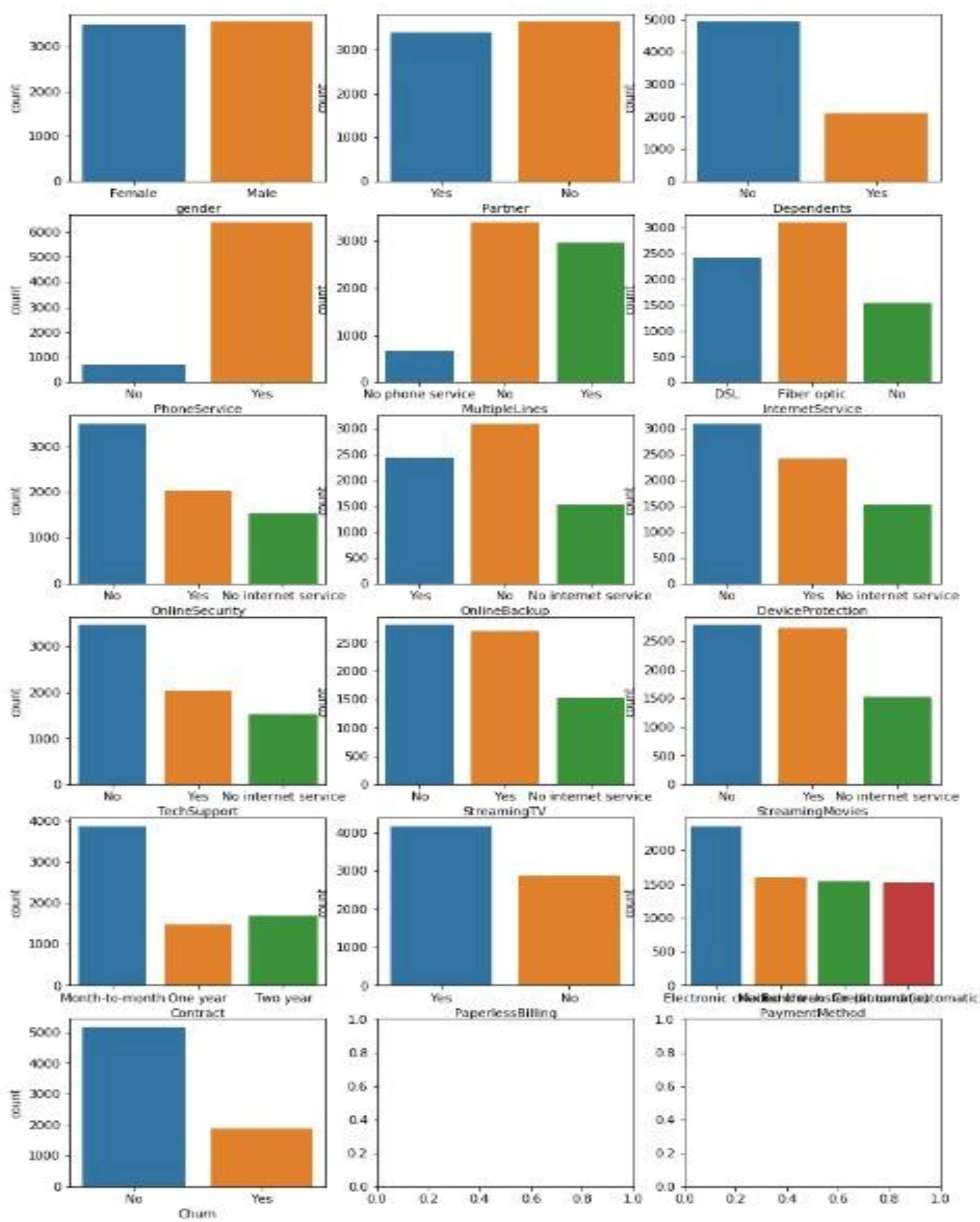
Predict Churn:

- Use the trained model to predict churn for new or existing customers.

Customer Retention Strategies:

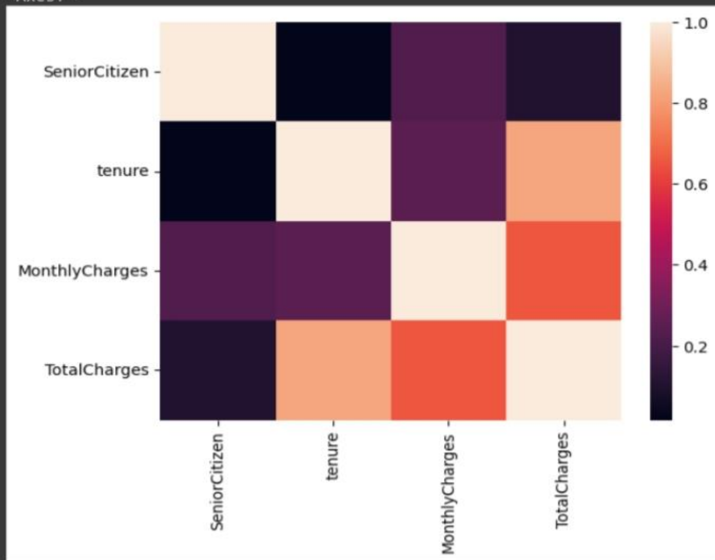
- Develop and implement customer retention strategies based on the insights from your model.Monitor and track the impact of these strategies on reducing churn.

```
/opt/conda/lib/python3.6/site-packages/matplotlib/figure.py:457: UserWarning: matplotlib is
ently using a non-GUI backend, so cannot show the figure
  "matplotlib is currently using a non-GUI backend, "
```



```
corr = data[continuous_var].corr()
sns.heatmap(corr)
```

<Axes: >

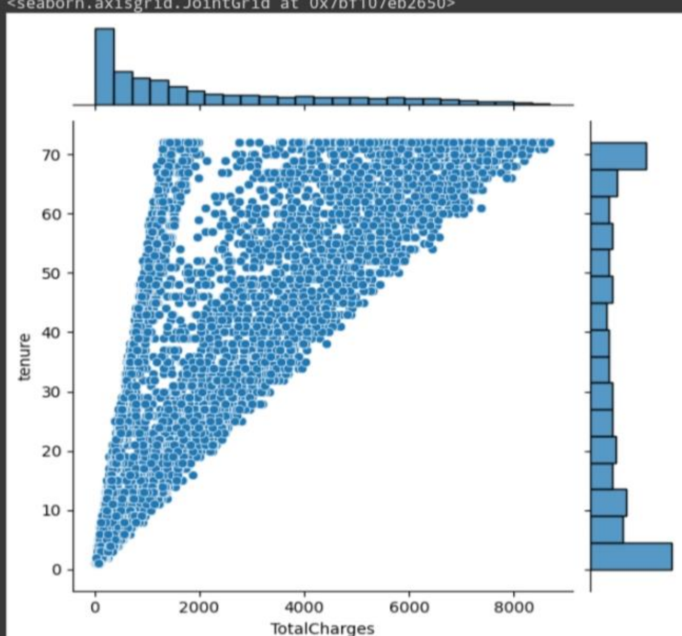



```
print (corr['TotalCharges'].sort_values(ascending=False), '\n')
```

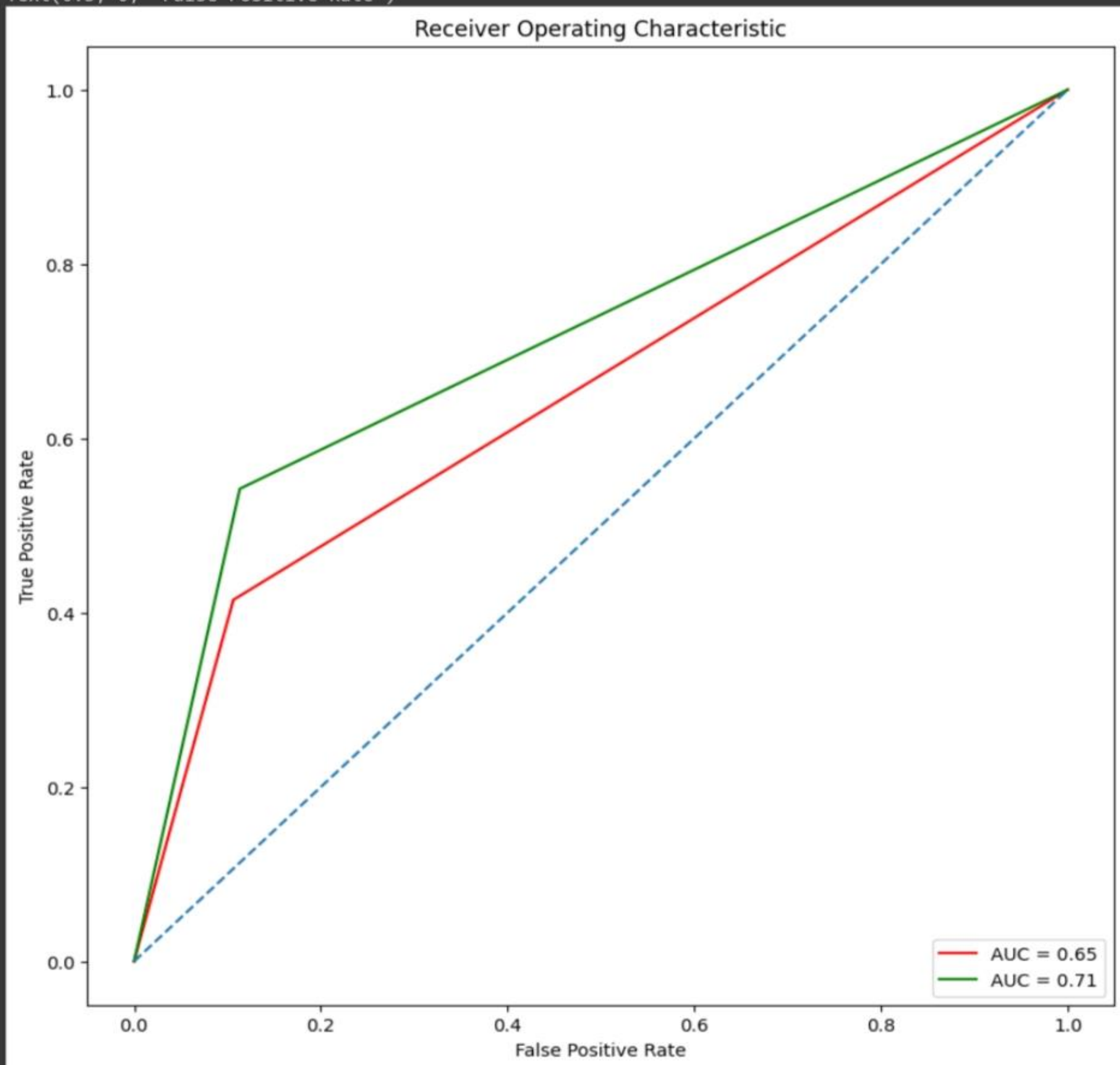
```
TotalCharges      1.000000
tenure            0.825880
MonthlyCharges    0.651065
SeniorCitizen     0.102411
Name: TotalCharges, dtype: float
```

```
sns.jointplot(x=data['TotalCharges'], y=data['tenure'])
```

<seaborn.axisgrid.JointGrid at 0x7bf107eb2650>



Text(0.5, 0, 'False Positive Rate')



DashBoard

Total Charges

16.1M

Tenure

228K

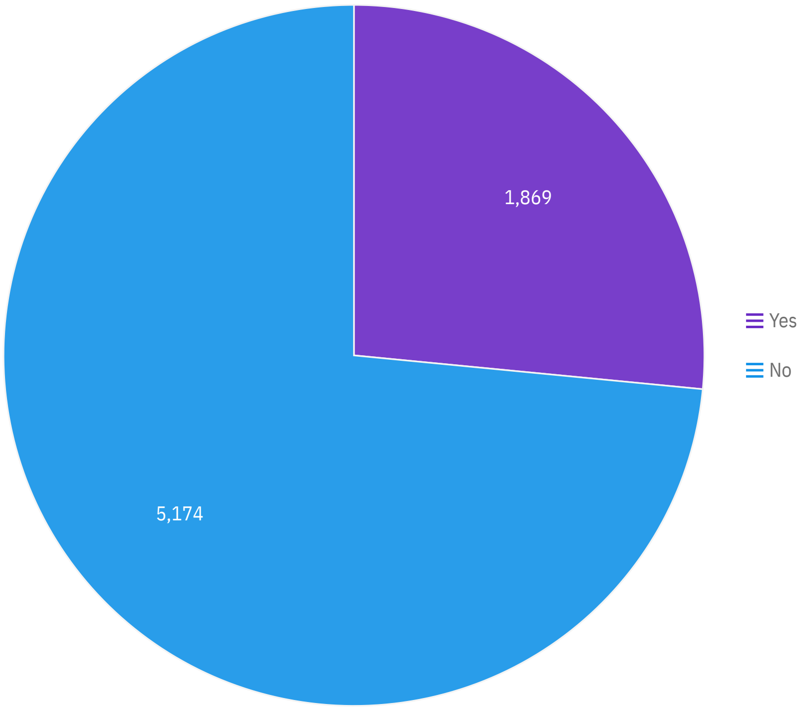
Sr. Citizen Users

1142

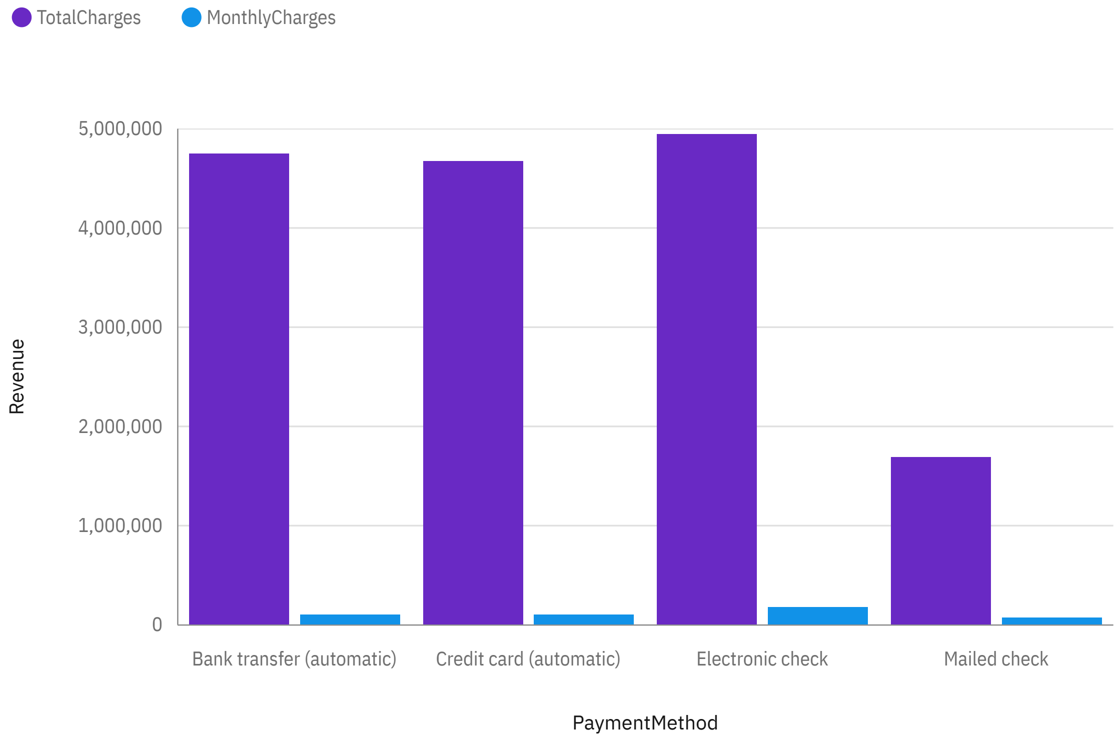
Churn 1

1.87K

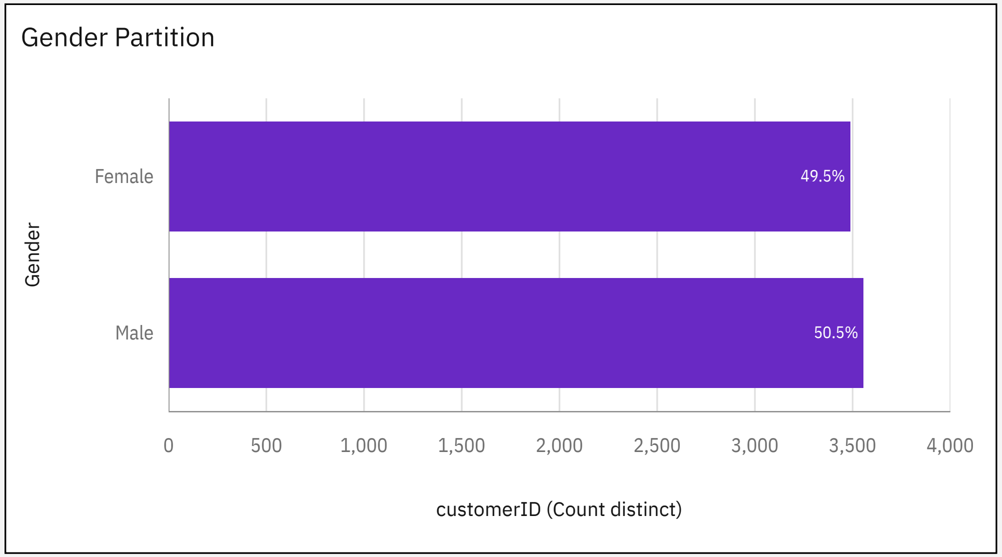
Churn Ratio

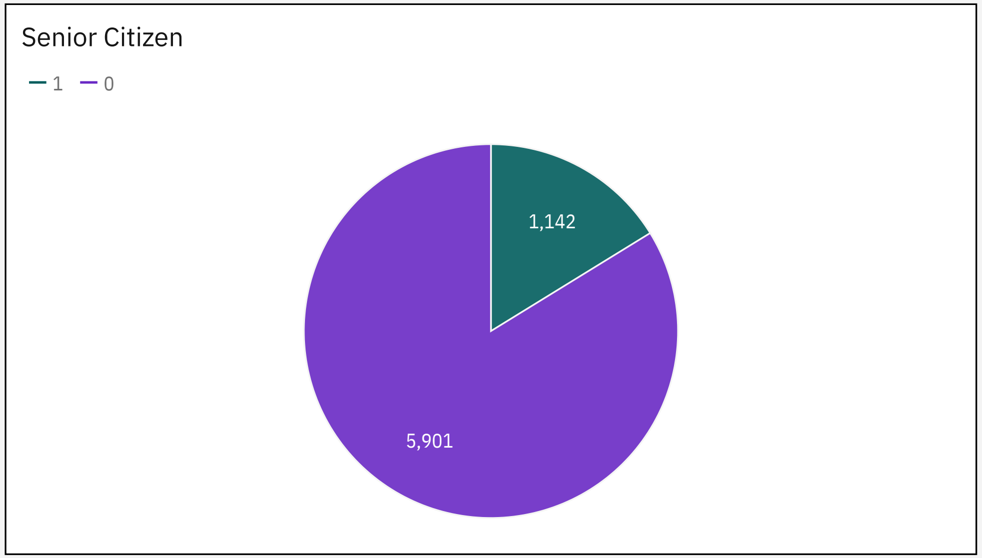


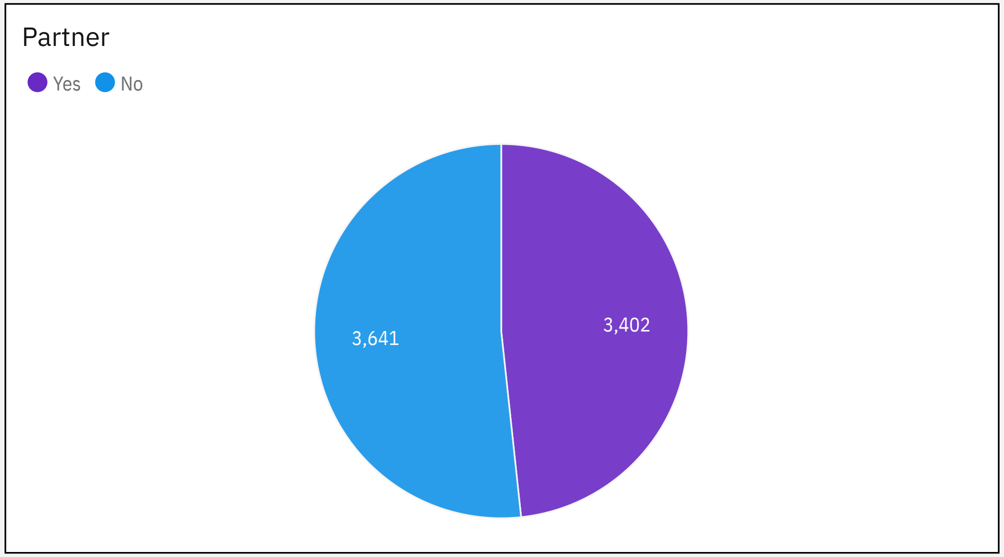
Charges Vs Payment Method

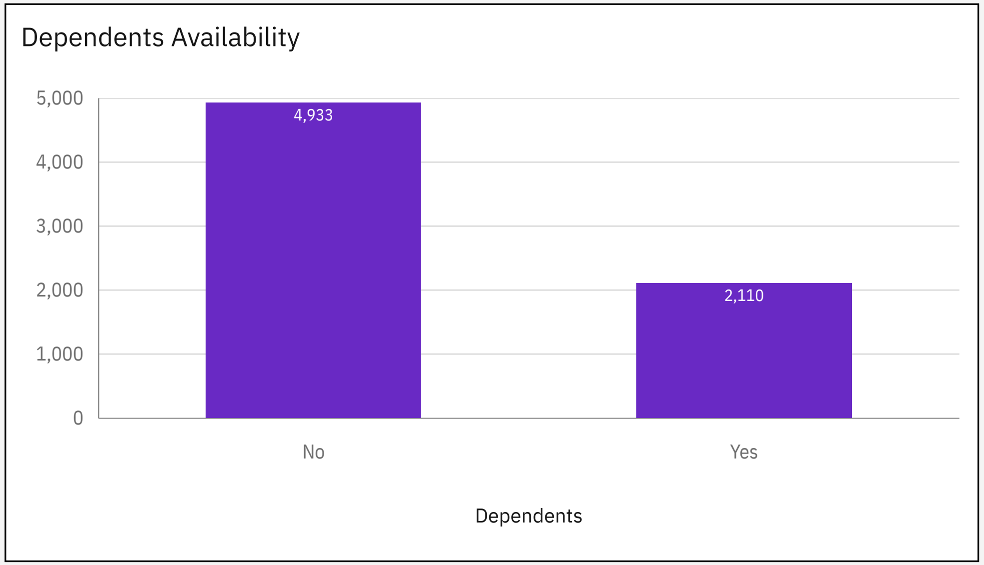


Analytics_1

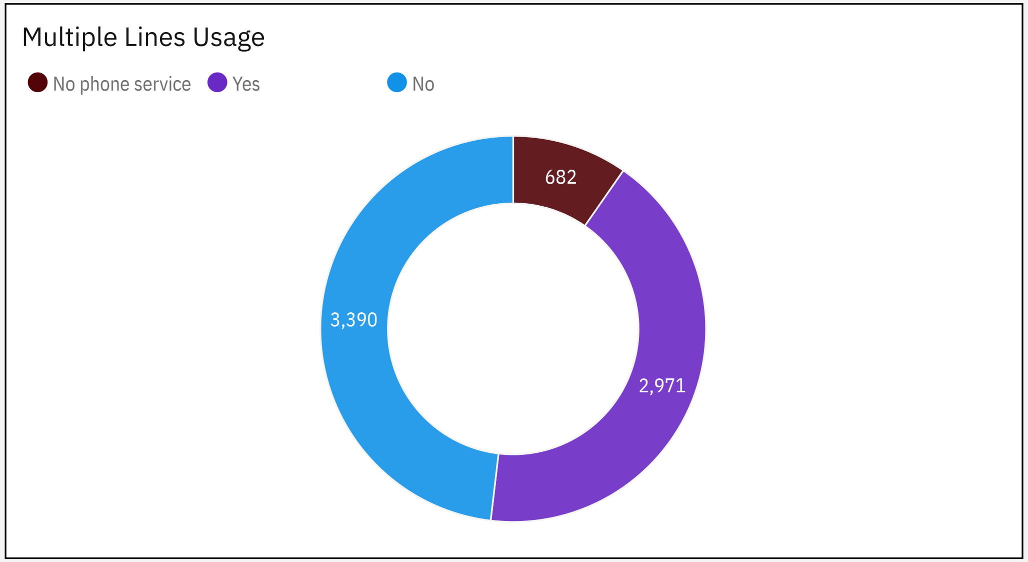

Gender Partition
Gender
Female
Male
49.5%
50.5%
0
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
customerID (Count distinct)

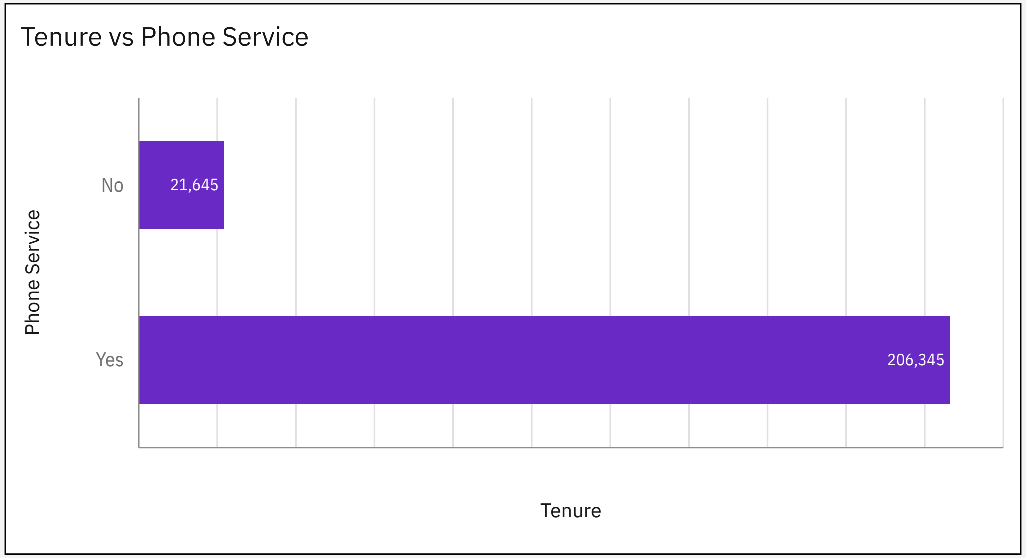

Senior Citizen
1
0
1,142
5,901

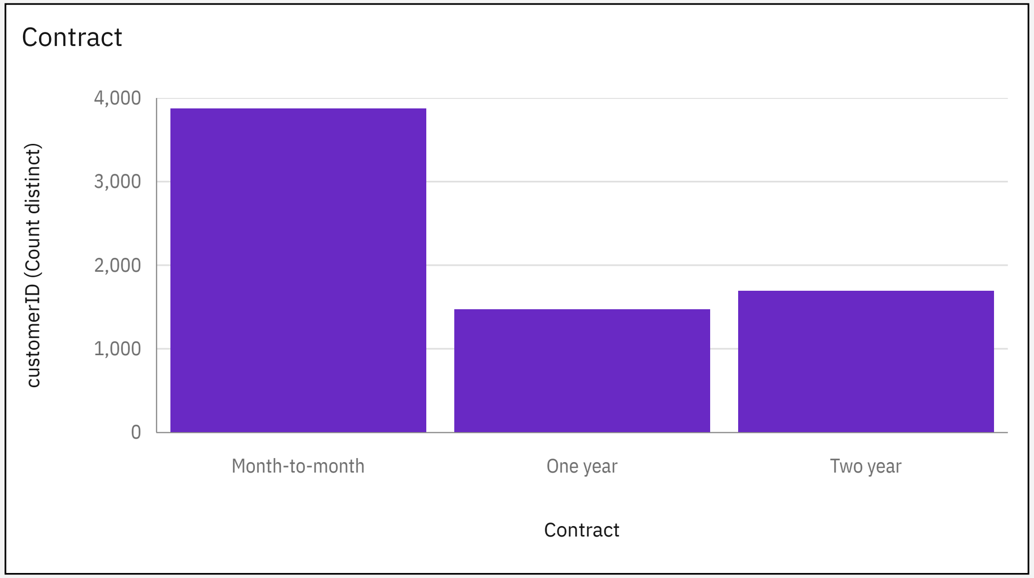

Partner
Yes
No
3,641
3,402

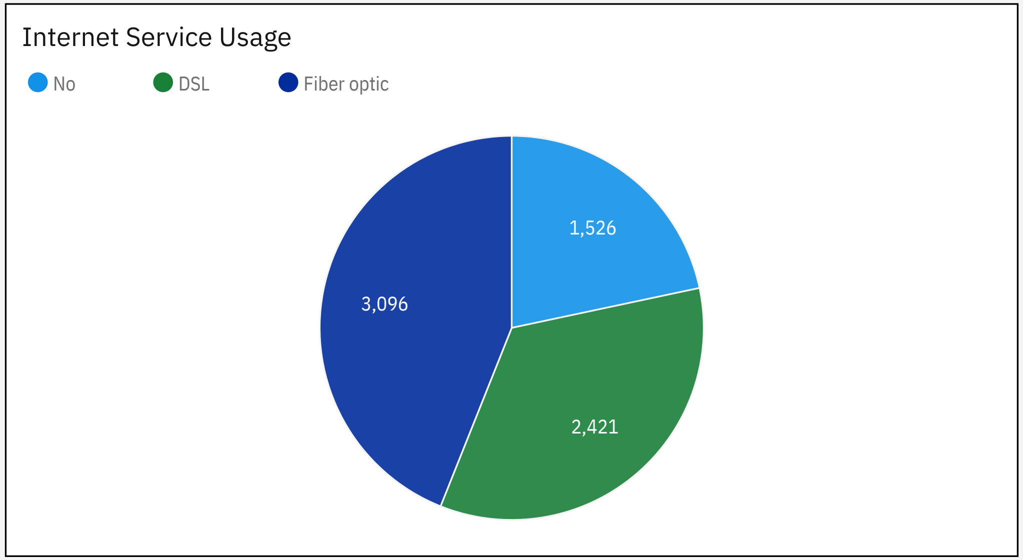

Dependents Availability
5,000
4,000
3,000
2,000
1,000
0
4,933
2,110
No
Yes
Dependents

Analytics_2

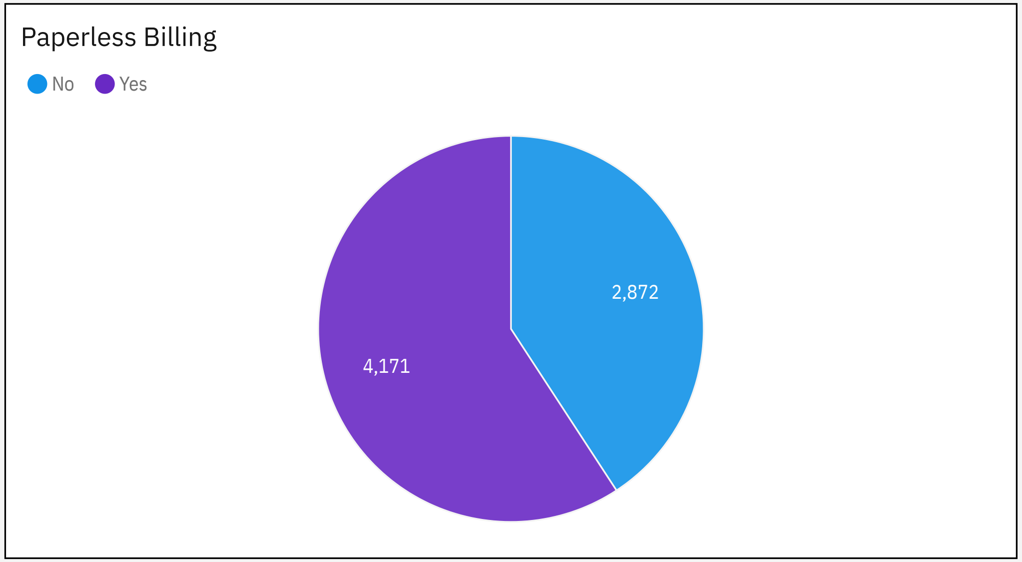

Multiple Lines Usage
No phone service
Yes
No
682
2,971
3,390

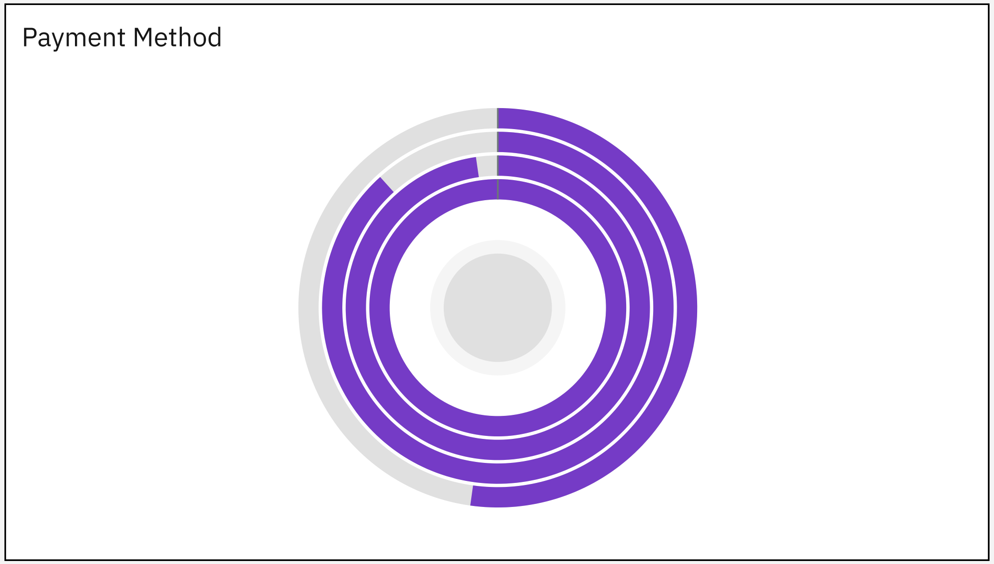

Tenure vs Phone Service
Phone Service
No
21,645
Yes
206,345
Tenure

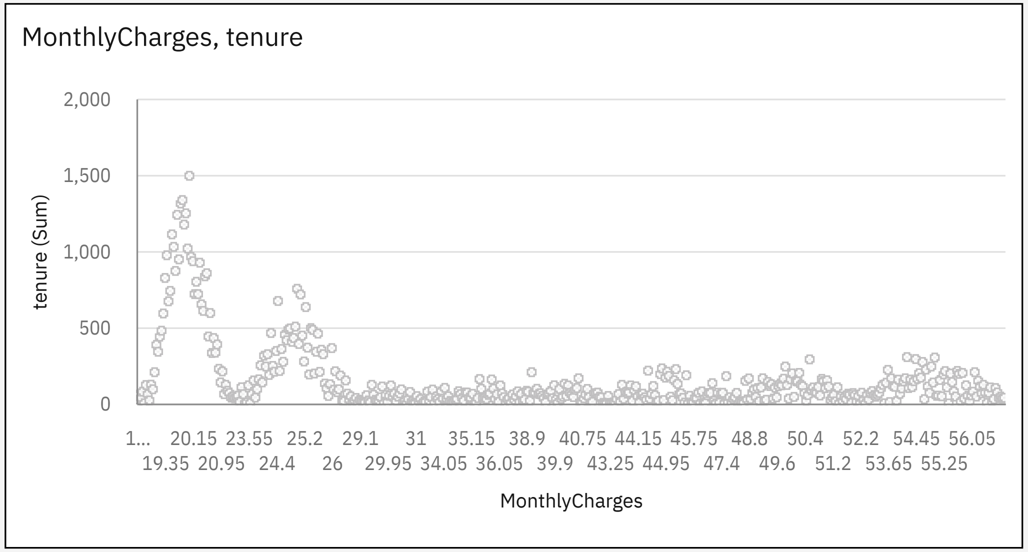

Contract
customerID (Count distinct)
4,000
3,000
2,000
1,000
0
Month-to-month
One year
Two year
Contract

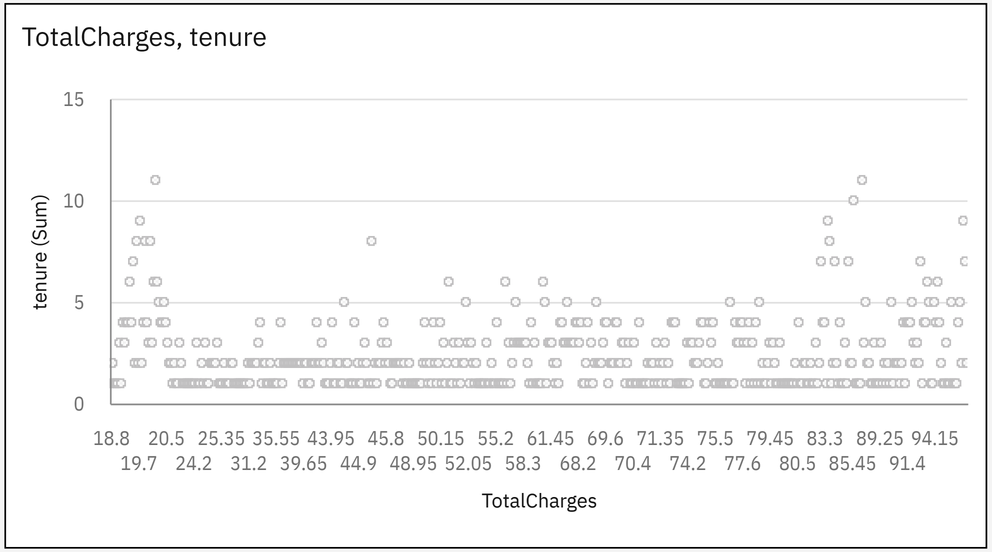

Internet Service Usage
No
DSL
Fiber optic
1,526
3,096
2,421

Analytics_3


Paperless Billing
No
Yes
2,872
4,171


Payment Method


MonthlyCharges, tenure
2,000
1,500
1,000
500
0
tenure (Sum)
1... 20.15 23.55 25.2 29.1 31 35.15 38.9 40.75 44.15 45.75 48.8 50.4 52.2 54.45 56.05
19.35 20.95 24.4 26 29.95 34.05 36.05 39.9 43.25 44.95 47.4 49.6 51.2 53.65 55.25
MonthlyCharges


TotalCharges, tenure
15
10
5
0
tenure (Sum)
18.8 20.5 25.35 35.55 43.95 45.8 50.15 55.2 61.45 69.6 71.35 75.5 79.45 83.3 89.25 94.15
19.7 24.2 31.2 39.65 44.9 48.95 52.05 58.3 68.2 70.4 74.2 77.6 80.5 85.45 91.4
TotalCharges