Presentation on Churn Analysis & Exploratory Data Analysis

Navpreet kaur

Business Understanding &overview

churn analysis Exploratory Data Analysis (EDA) is to gain a deeper understanding of the data related to customer churn. EDA is a crucial step in the churn analysis process as it helps to identify patterns, relationships, and potential factors that may influence customer churn.

The main objectives of EDA in churn analysis are as follows:

1.Data Understanding: EDA helps in understanding the structure and characteristics of the data related to customer churn. It involves examining the data types, distributions, and summary statistics of the variables.

2.Identifying Patterns and Trends: EDA helps in identifying patterns and trends in the data that may be indicative of customer churn. This includes examining the distribution of churned vs. non-churned customers across different variables, such as demographics, usage patterns, and transaction history.

3.Exploring Relationships: EDA helps in exploring relationships between different variables and customer churn. This includes examining correlations between variables, identifying potential predictors of churn, and understanding how different variables may interact with each other.

4.Detecting Outliers and Anomalies: EDA helps in detecting outliers and anomalies in the data that may be indicative of errors or unusual behavior. This includes examining the distribution of variables, identifying extreme values, and understanding the potential impact of outliers on the analysis.

5.Identifying Data Quality Issues: EDA helps in identifying data quality issues that may affect the accuracy and reliability of the churn analysis. This includes examining missing values, inconsistencies, and data entry errors.

Understanding The Data

Exploratory Data Analysis: EDA is an important step in data analysis pr data science project. EDA is the process of investigating the data set to discover patterns, and outliers and form hypothesis based on our understanding of the data set.

EDA involves generating summary statistics for numerical data in the dataset and creating various graphical representations to understand the data better. We will use python language for this purpose.

Some steps in EDA:

1.Data Sourcing
2.Data Cleaning
3.Categorical Analysis
4.Numerical Analysis
5.derived metrices

Types Of Analysis: There are three types of data analysis

1.univariate Analysis
2.Bivariate Analysis
3.Multivariate Analysis

Importing Libraries:

we will start by importing the libraries we will require for performing EDA.these include Numpy,Pandas,Metaplotlib and Seaborn.

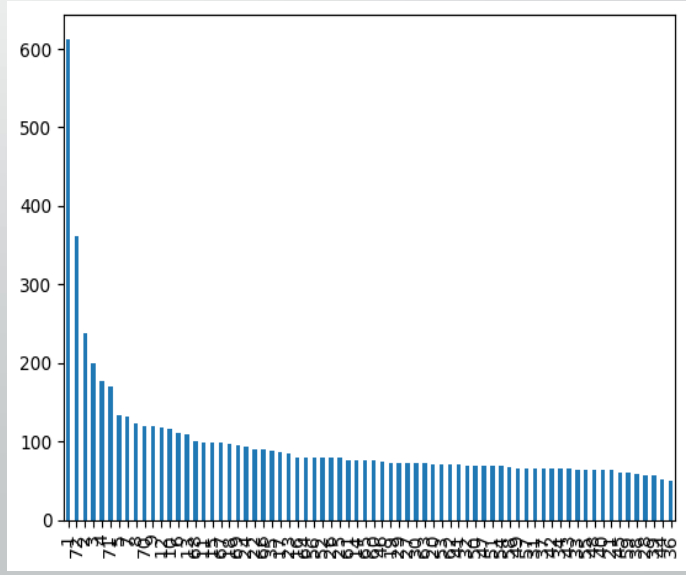
. Import NumPy as np
. import pandas as pd
. import matplotlib.pyplot as plt
.import seaborn as sns

Data Cleaning- some steps involving in data cleaning

1. Handle missing value
2. standardization of the data
3. outlier treatment
4. handle invalid value

missing value:

Missing value churn analysis refers to the process of analyzing the impact of missing data on churn prediction or customer attrition analysis. When dealing with real-world data, it's common to encounter missing values, which can arise due to various reasons such as data entry errors, incomplete surveys, or customers choosing not to provide certain information



- Dropping Unnecessary columns

. After analysing the data frame, we have numerical columns like senior citizen,Tenure,monthly charges and total charges

.The senior citizen data looks like a numerical, but it treated as a categorical.

. We will drop the unnecessary columns like customer ID and tenure

Types of Analysis:

There are three types pf analysis

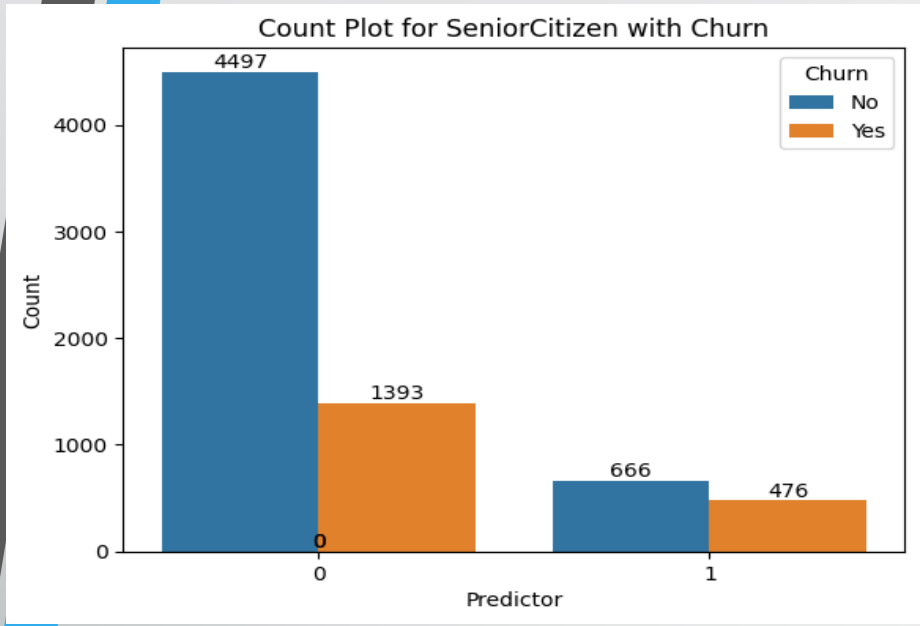
1.univariate analysis

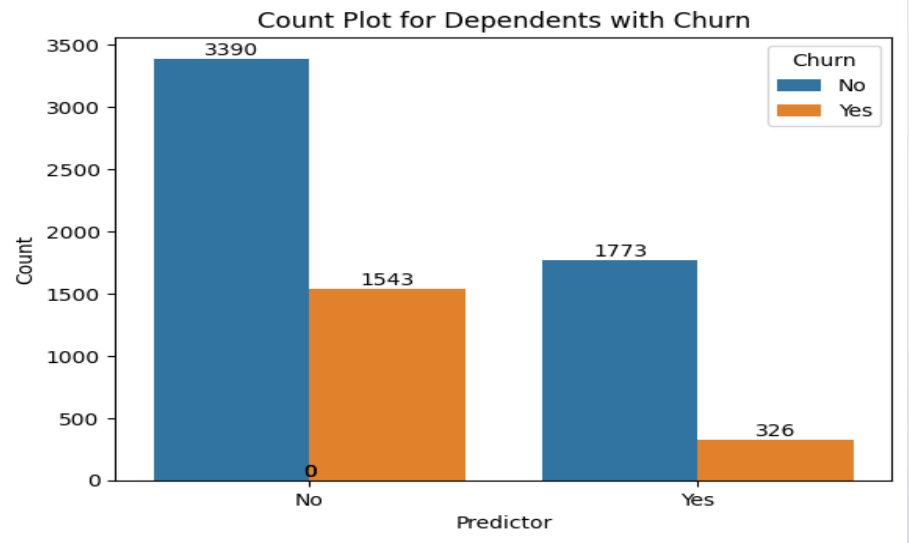
2.Bivariate analysis

3. Multivariate analysis

Univariate Analysis:univariate analysis is the simplest form of analyzing data.uni means one ,so in other words your data has only one variable

This shows senior citizen &dependent with churn

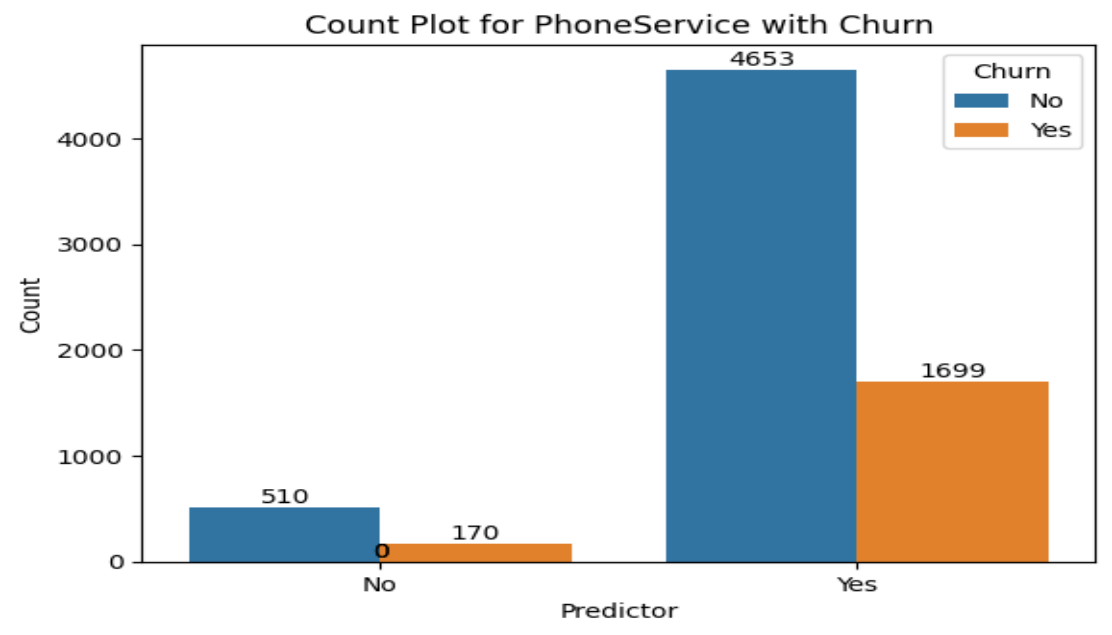


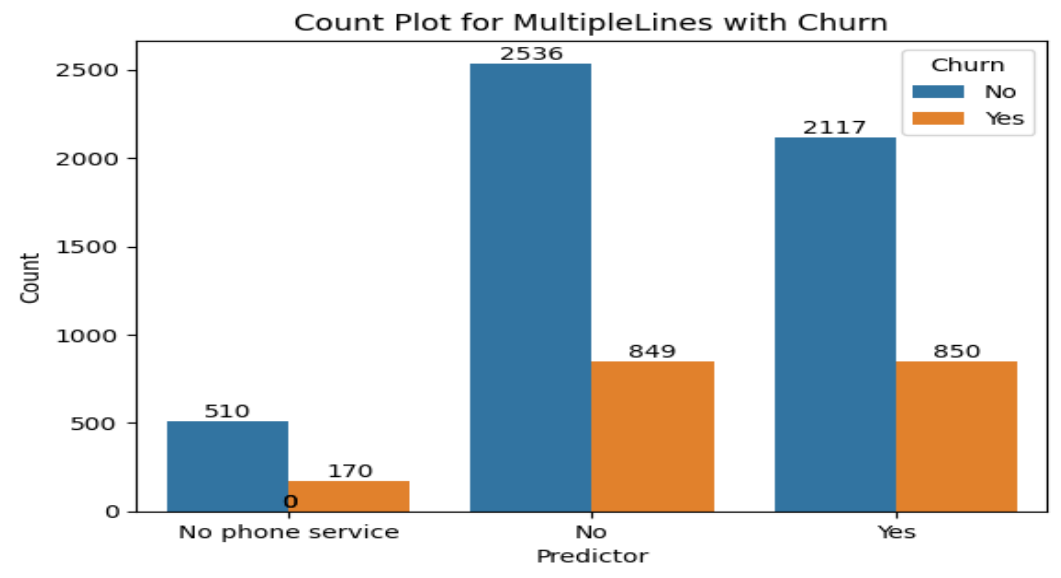


Churn rate of senior citizen=23.65%
churn rate of youngster=41.68%
youngster more like to churn it shows peak insight

1543/1543+3390/100=31.27%
326/326+1773*100=15.53%
31,27% people more dependent with the churn

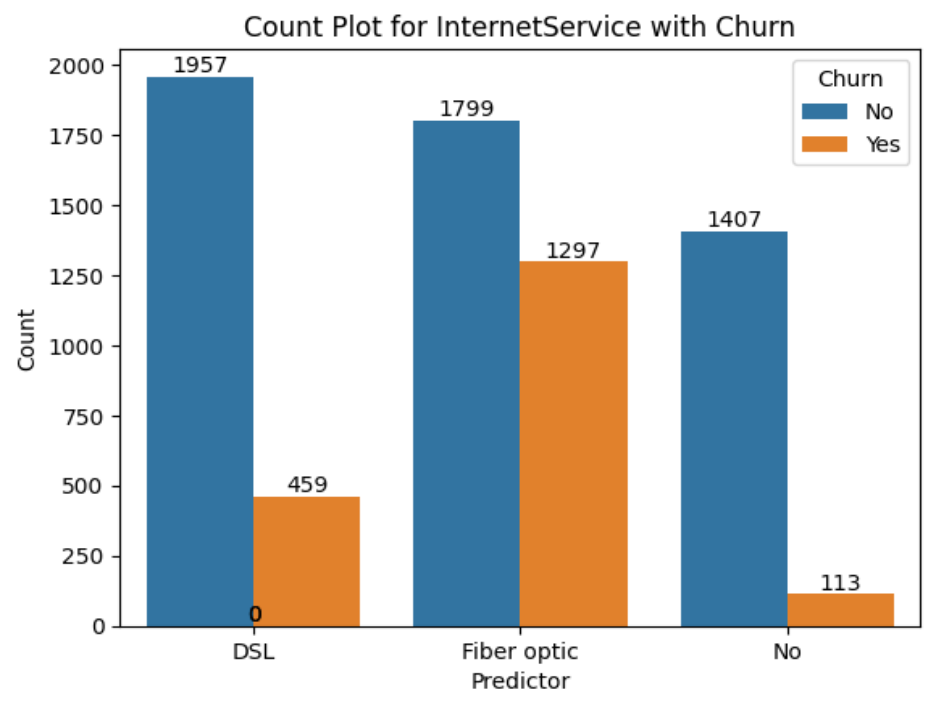
This graph show multiple lines and phone service with churn

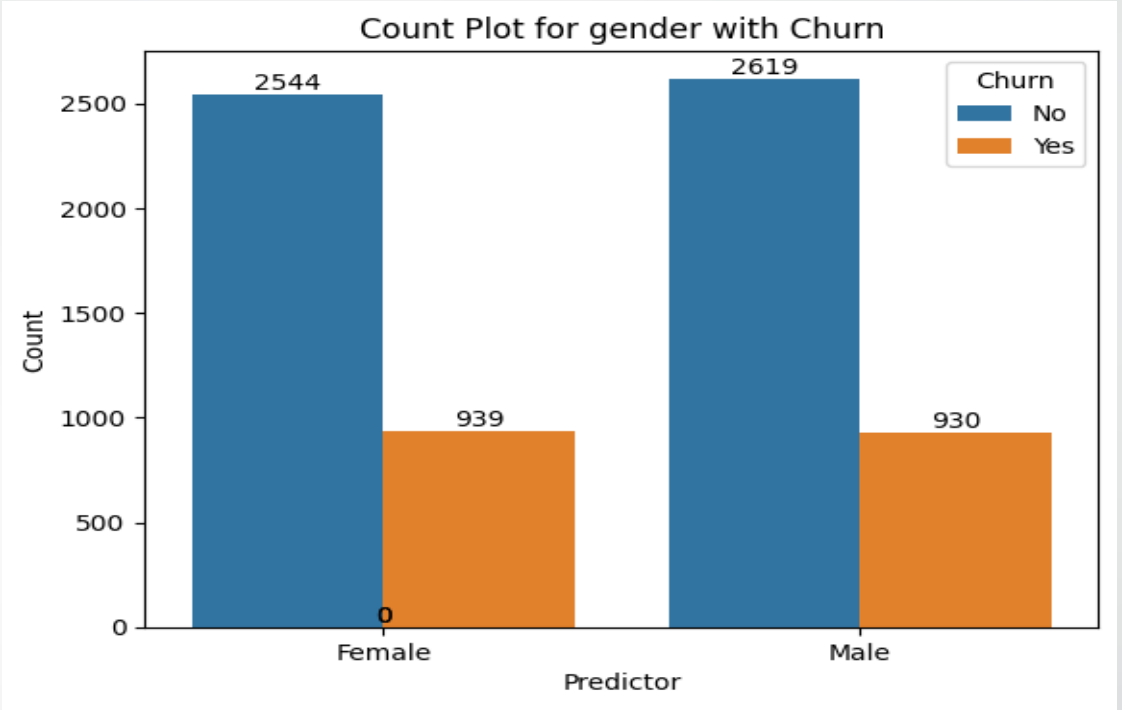




170/510+170/100=25%
1699/1699+4653*100=26.74
26.74% people are using phone service with churn

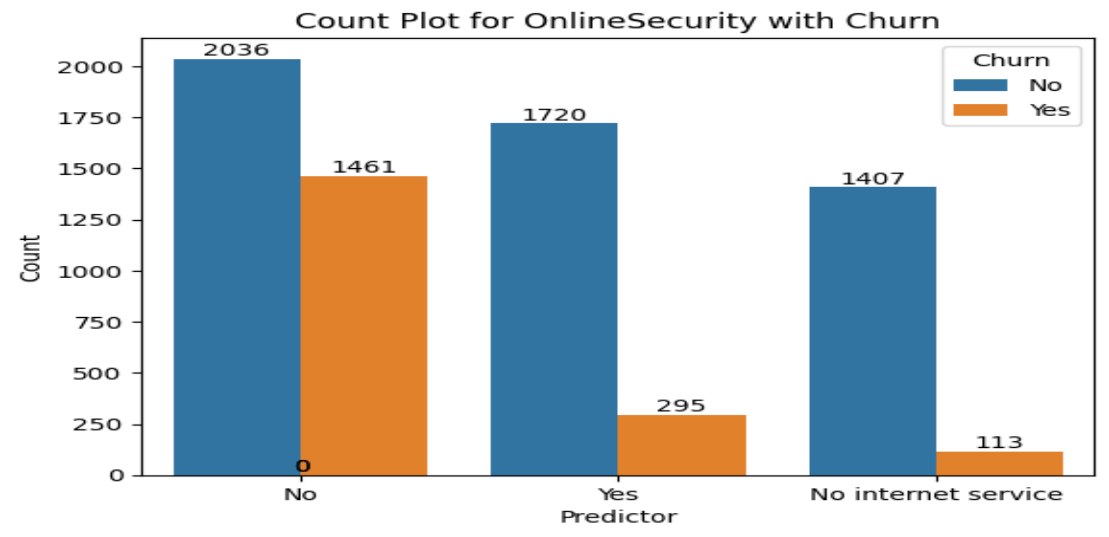
No phone service=25%
no multiple lines with churn=25.08
yes, multiple lines with churn=28.64%
mostly people are using multiple lines with churn as compared to others.

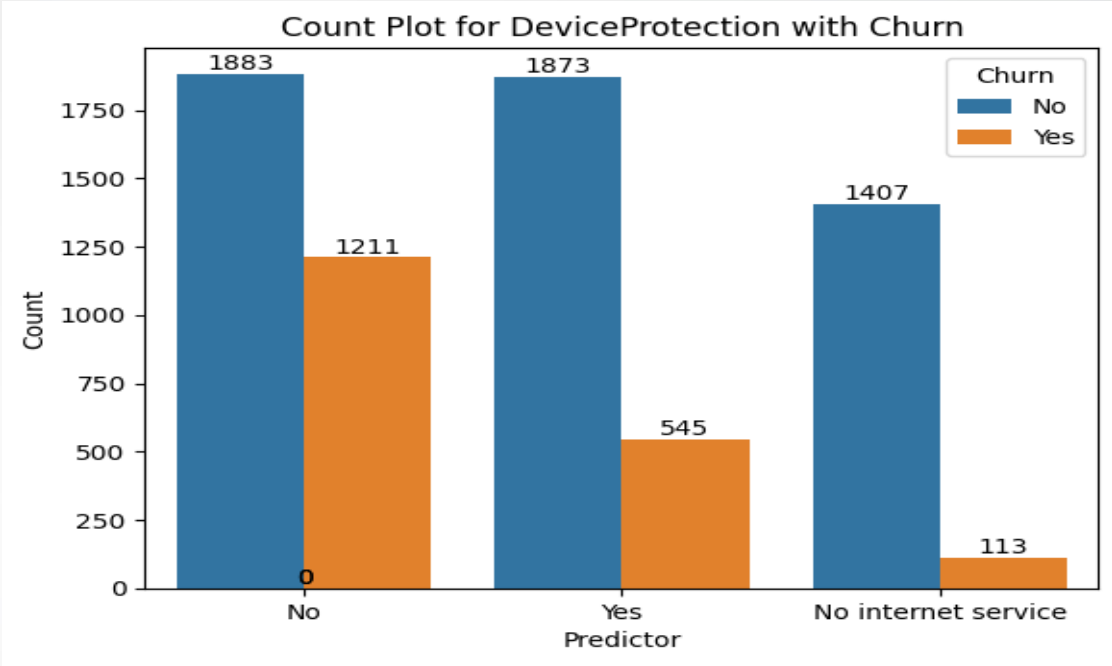




DSL=18.99%
Fiber optic=41.89%
no internet service=7.43%
most of the people prefer to fiber optic internet service as compared to other

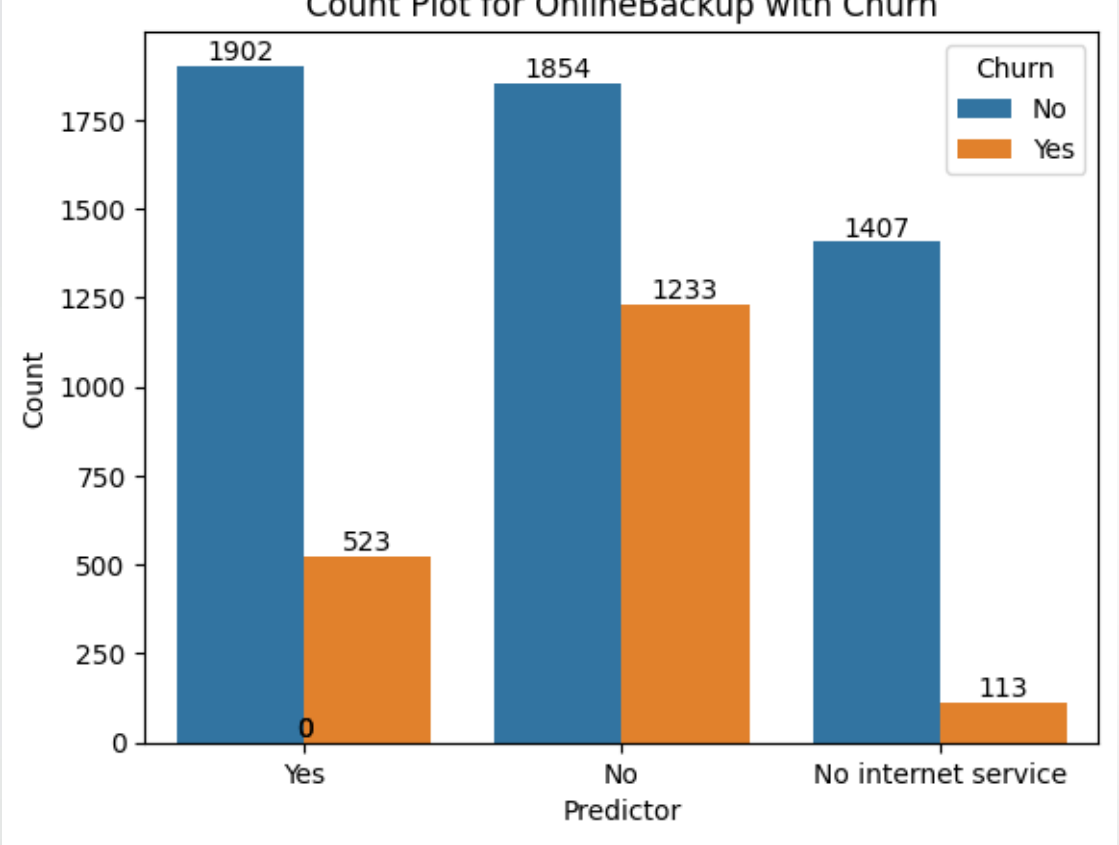
Churn rate of female=26.95%
churn rate of male=20.76%
females are more likely to churn as compared to male.





No online security with churn=41.77%
yes, online security with churn=14.64%
no internet service with churn=7.43%
highest people using no online security with churn 41.77%

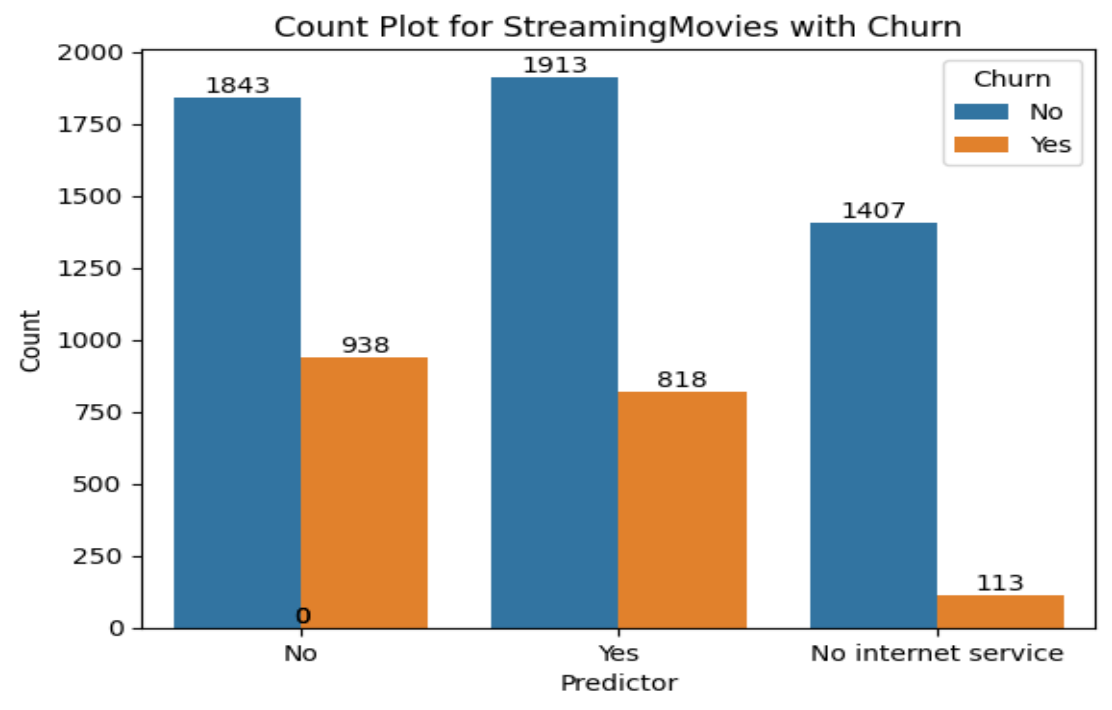
No=1211/1211+1883*100=39.14%
yes=545/545+1873*100=22.53%
no internet service=113/113+1407*100=7.43%
device protection with no churn rate is higher in no internet service as compare two others

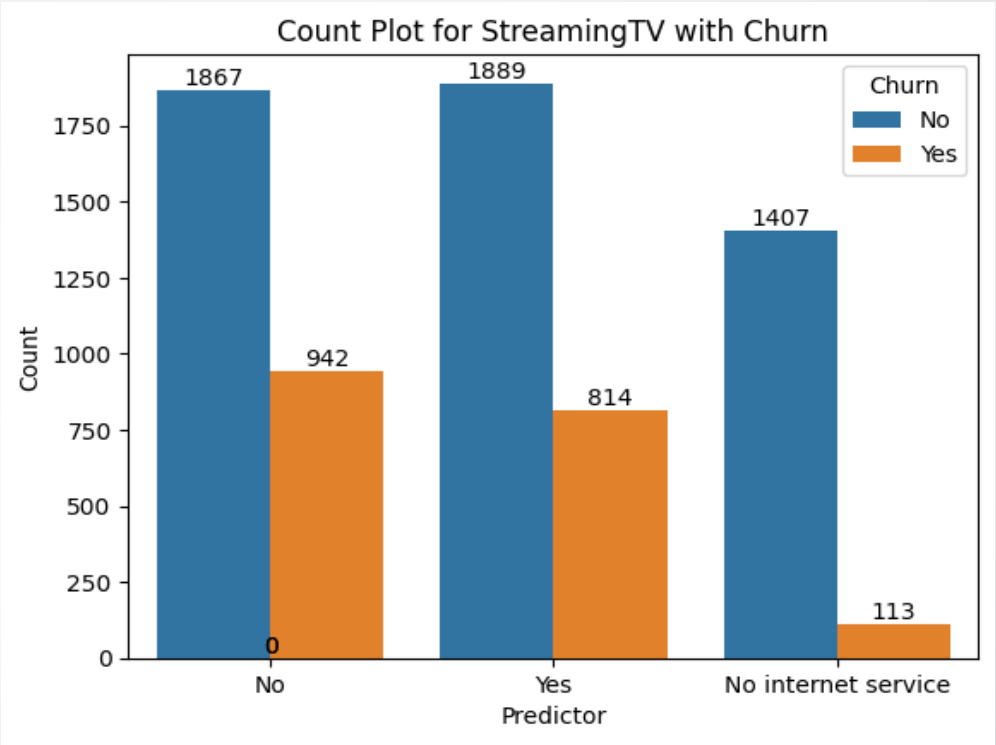


Yes, churn rate=523/523+1902*100=21.55%

no churn rate=1233/1233+1854*100=39.94%

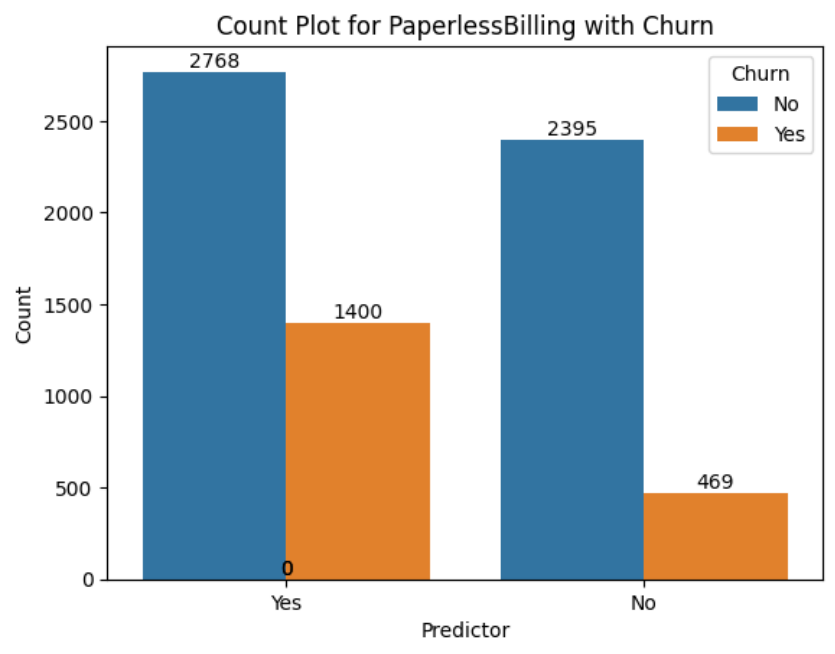
no internet service=113/113+1407*100=7.43%

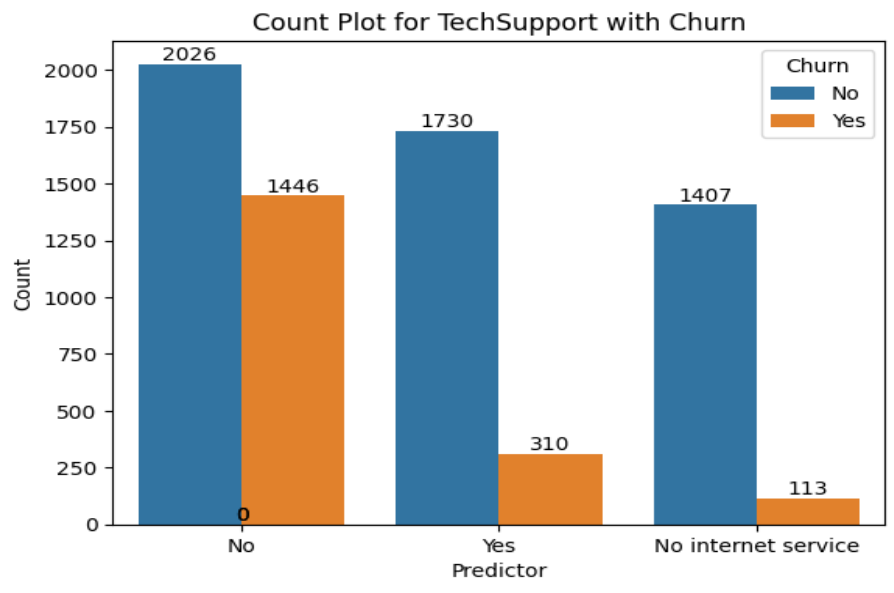




no=938/938+1843*100=33.72%
yes=818/818+1913*100=29.95%
no internet service=113/113+1407*100=7.43%

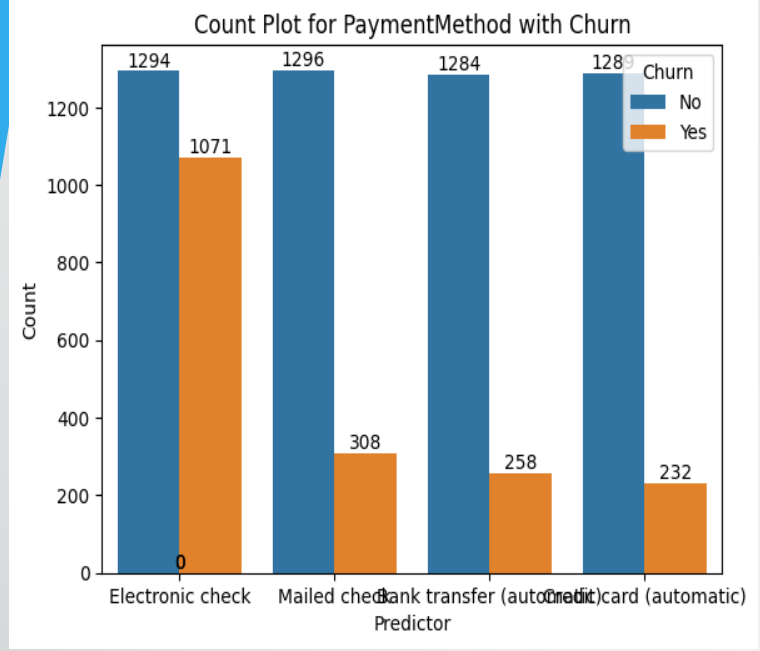
no==33.53%
yes=30.11%
no internet service==7.43%

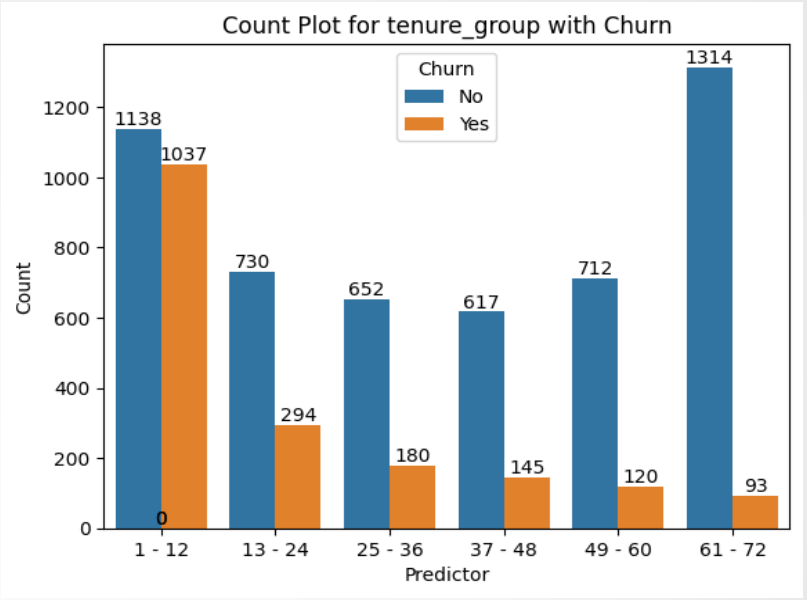




Churn rate Yes=33.58%
churn rate no=16.37%
most of the people using paperless billing with churn33.58%

No=1446/1446+2026*100=41.64%
yes=310/310+1730*100=15.19%
no internet service=113/113+1407*100=7.43%
without tech support maximum people are more likely to churn





Electronic check=1071/1071+1294*100=45.28

Mail check=308/308+1296*100=19.20%

Bank transfer=258/258+1284*100=16.73%

Credit card=232+1289*100=15.25%

1-12=1037/1037+1138*100=47.67%

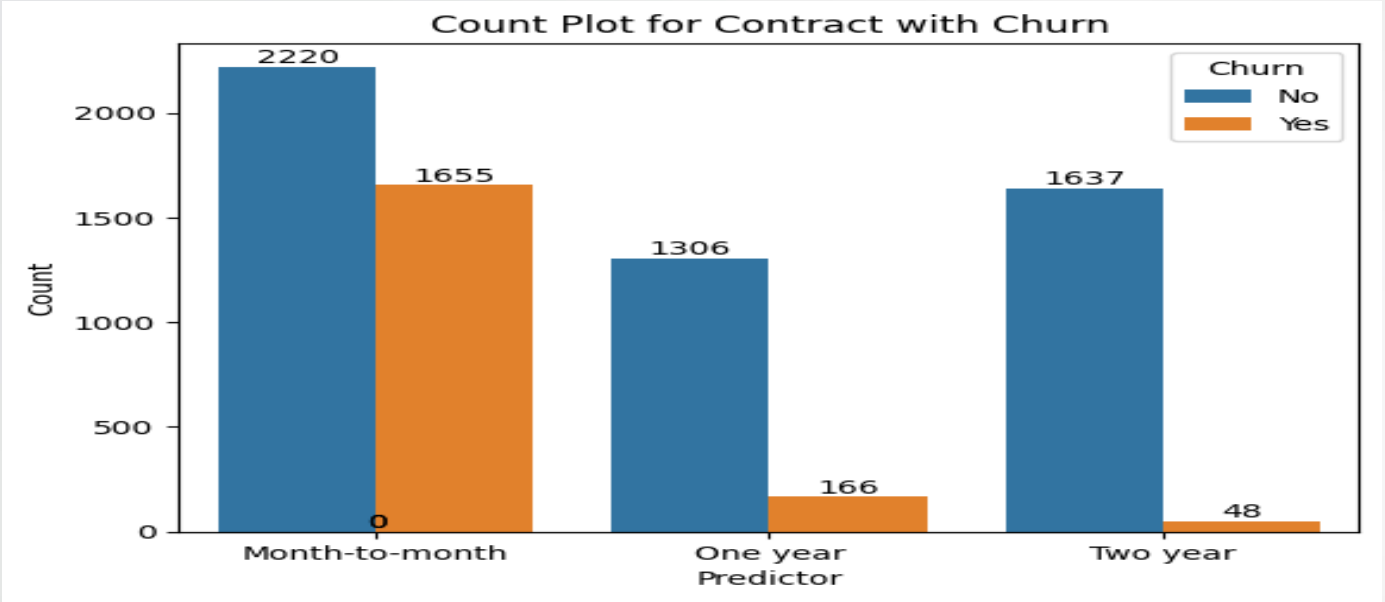
13-24=294/294+730*100=28.7%

25-36=180/180+652*100=21.63

37-48=145/145+617*100=19%

49-60=120/120+712*100=14.42%

61-72=93/93+1314*100=6.6%



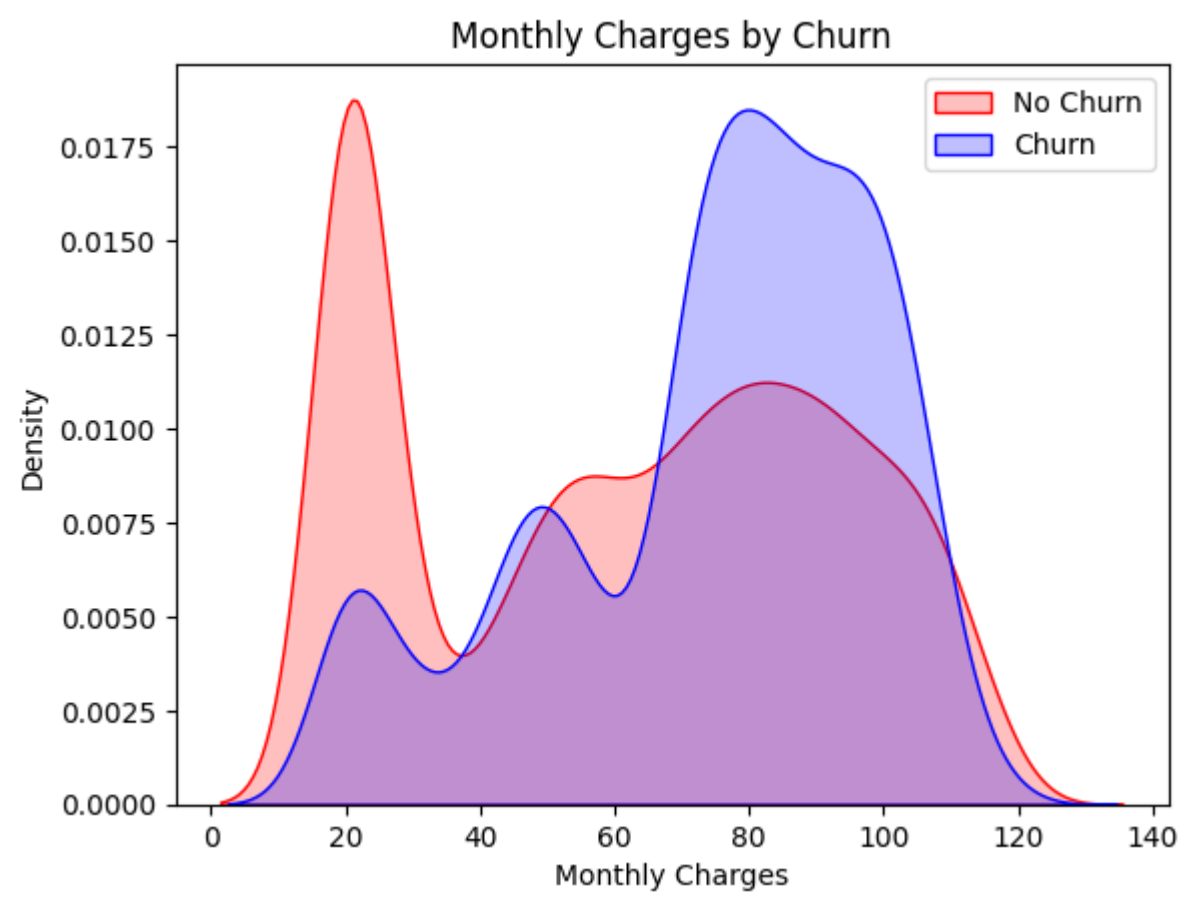
Month to month=1655/1655+2220*100=42.70%

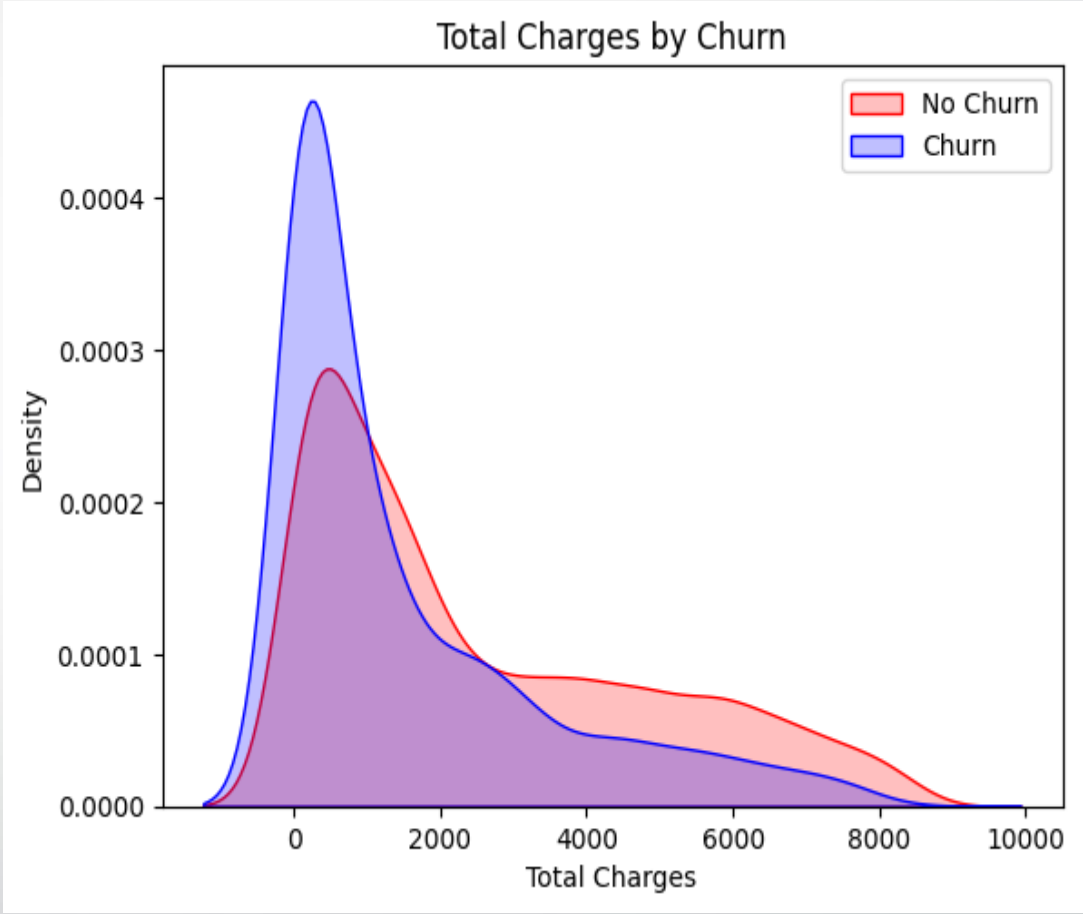
one year=166/166+1306*100=11.27%

two year=48/48+1637*100=2.8%

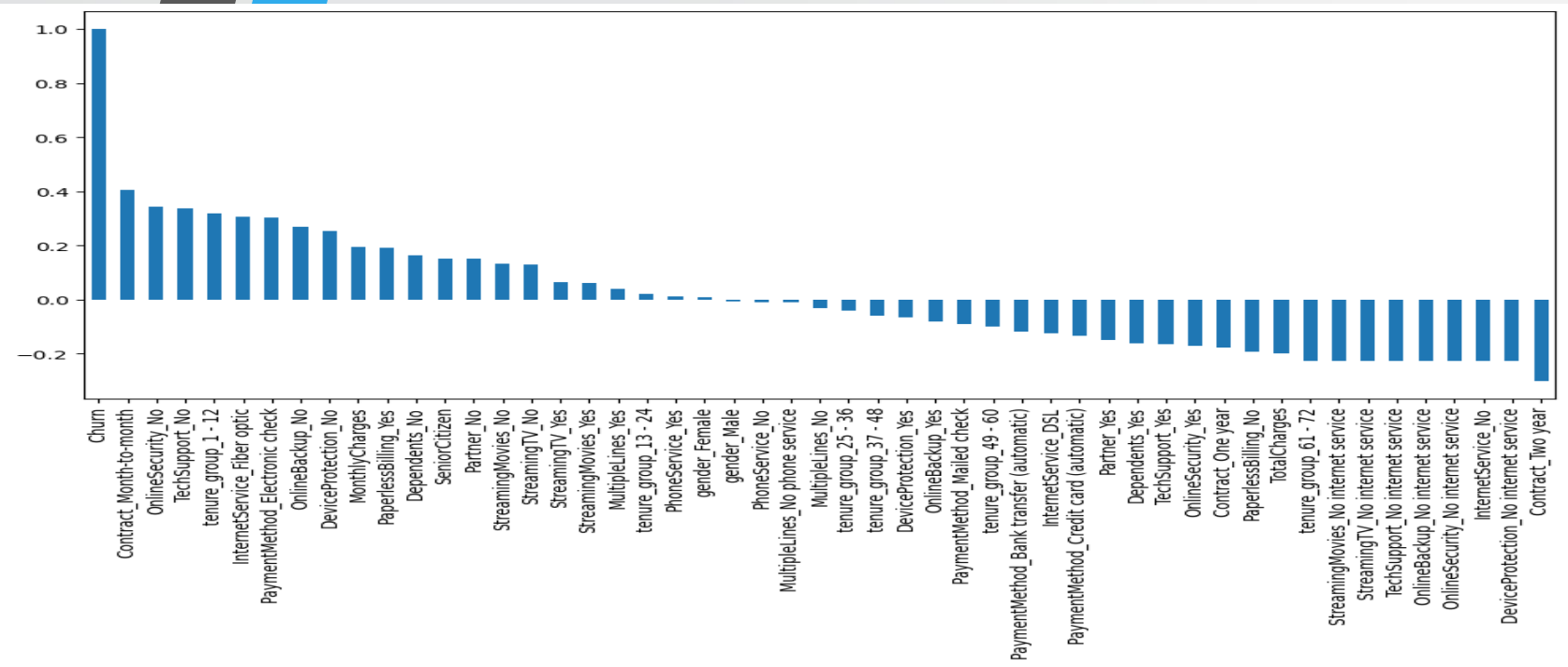
monthly customer are more likely to churn because they are free customer.

Numerical univariate analysis





Insight: churn is high when monthly charges are high
monthly charges and total charges are positively correlated



Insight:

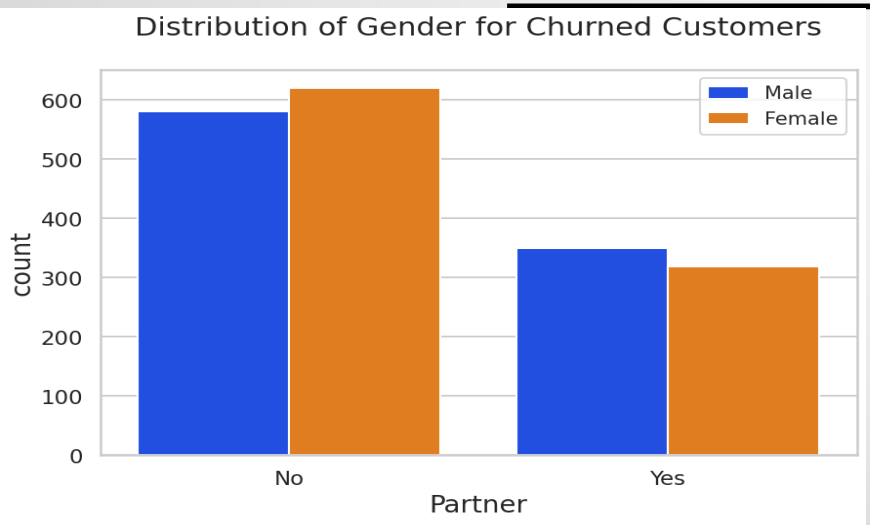
high churn seen in case of month-to-month contract, no online security, no tech support first year of subscription and fibre optic internet

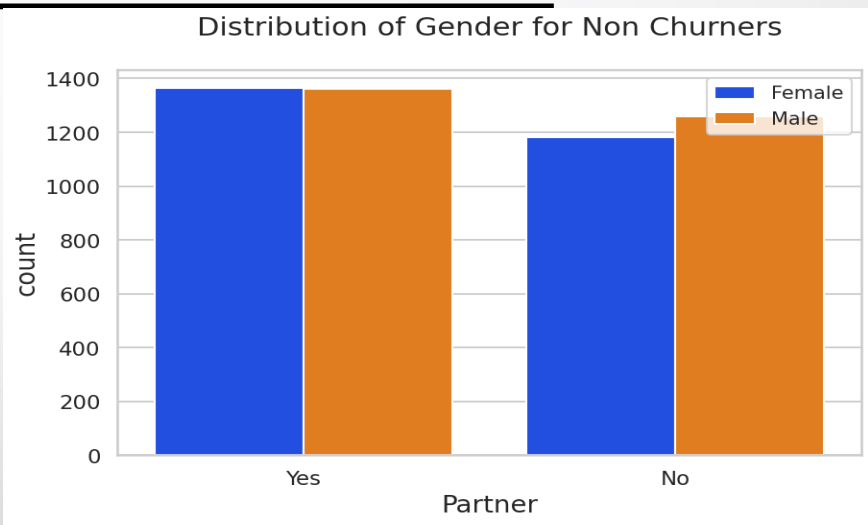
low churn is seen in case of long-term contracts, subscription without internet service and the customer engaged for 5+ years.

Factors like gender,availiabiltyof phone service and of multiples lines have almost no impact on churn.

Bivariate Analysis: Bivariate analysis means to the analysis of two variables to determine relationship between them.

Gender for churn vs non churn





- insight: Most of the people who are female and having no partner are more likely to churner

Conclusion

There are some of the quick insight from this exercise:

1Senior citizen are more likely to churn

2.People with no partner are more likely to churn

3.Monthly customer are more likely to churn because they are free customers.

4.People who are pay via electronic check are more likely to churn.

5.Monthly charges & total charges are positively corelated

6.Churn is high when monthly charges are high.

7.No online security,no tech support category are high churners.

THANKS