CUSTOMER CHURN ANALYSIS

EDA CASE STUDY

SUREKHA BERLIN

BUSINESS UNDERSTANDING & OVERVIEW

- Telecom Industry is a highly competitive industry. So, customer retention is very important. In the telecom industry, customers can easily choose from multiple service providers and actively switch from one operator to another. In this highly competitive market, the telecommunications industry experiences a significant annual churn rate.

- Customer Churn means whether the customer has stopped using the product and switched to a competitor product.

- In this project I will be performing Exploratory Data Analysis to analyze the customer level data of a leading telecom industry to produce actionable insights into the main indicators of the churn

OBJECTIVES

- The objective is to obtain a data-driven solution that will allow us to reduce churn and, therefore, to increase customer retention.
- To predict how likely a customer will churn by analyzing its characteristics:

(1) Demographic information

(2) Account information

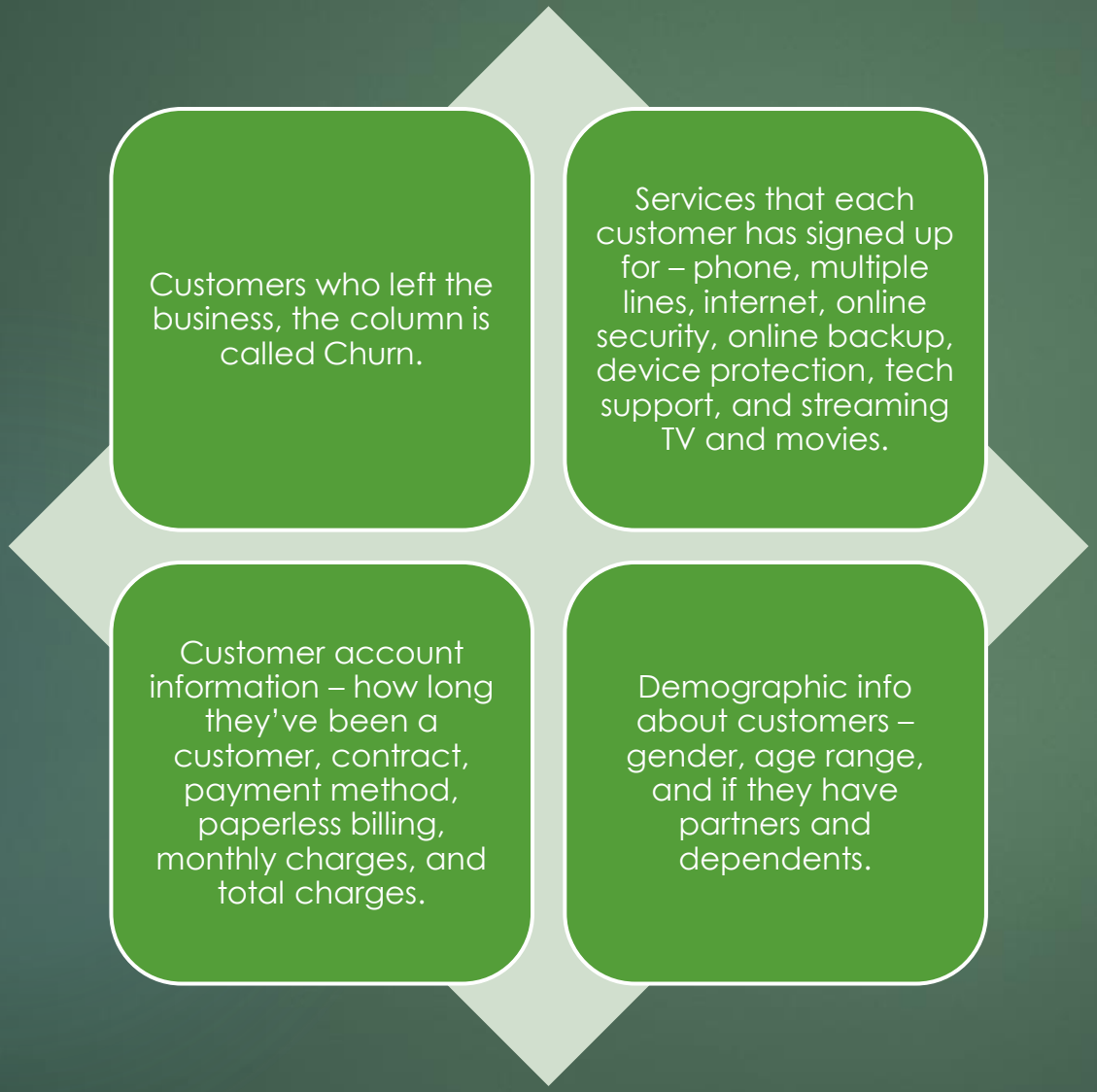
(3) Services information.

- To Compare the demographic makeup of churned and non-churned consumers. Visualizations, such as stacked bar charts, are used to display the findings.
- To prevent customers from churning, we can use data analysis to identify potential features and design strategies to retain customers from leaving.
- Recommend product strategies to business team based on the analysis of product offerings that will help in retaining the customer based on available data

RESOURCES & TOOLS REQUIRED

- A dataset, the dataset I am using is for the purpose of doing this project. The name of the dataset is (Customer-Churn.csv)
- Google Colab/Jupyter Notebook
- Import Libraries :Pandas, NumPy, Matplolib, Seaborn, and other machine learning libraries are available.

INFORMATION INSIDE THE DATASET



DATA DESCRIPTION

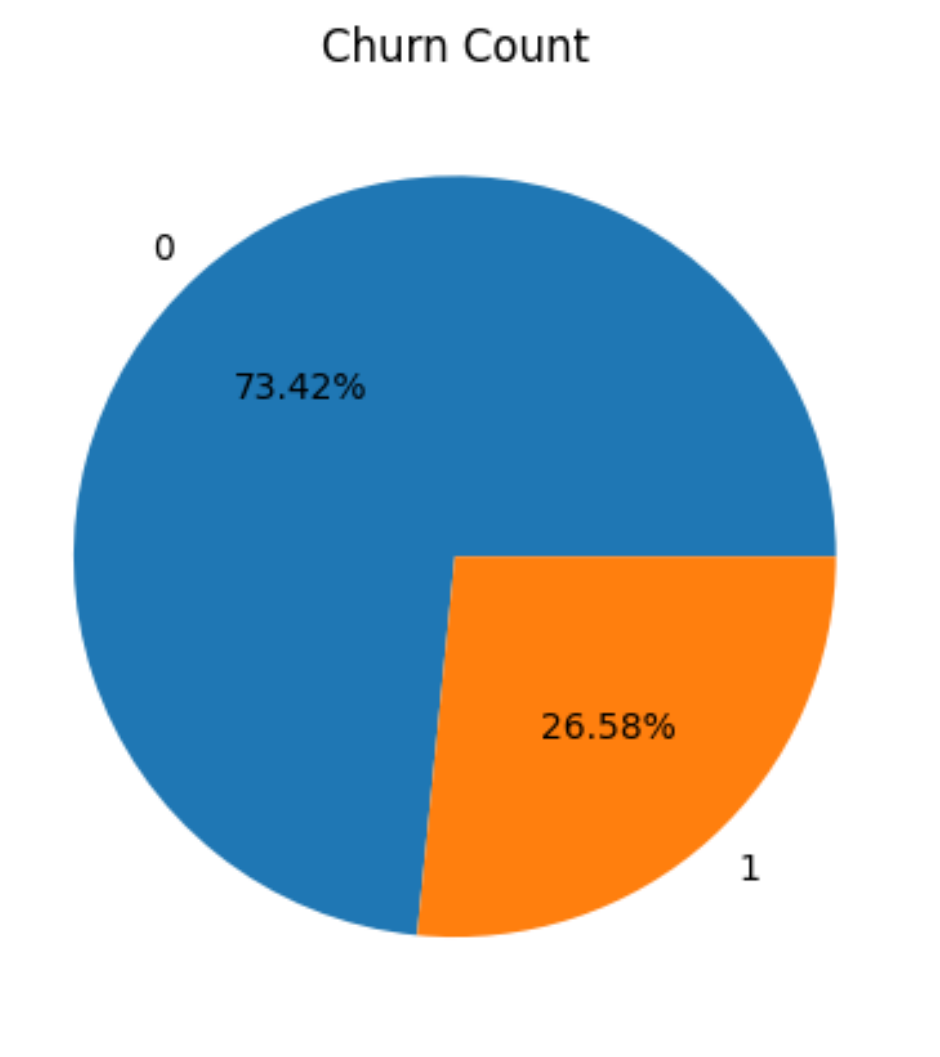
- Data consists of 7043 fictional customers who belong to various demographics (single; with dependents; senior citizen) and subscribe to different products offerings (internet service; phone line; streaming TV; streaming movies; online security) from a telecom company.

- Independent variables: 17 Categorical and 3 Continuous

- Dependent Target variable: “Churn”

TARGET VARIABLE

- Dependent Target variable: “Churn”
- The Churn Rate is 26.58 %
- The dataset is highly imbalanced i.e., there are a lot of loyal customers who are not leaving (the ‘No’ category customers) the telecom service provider to switch to the competitor product. So, we analyze the data with other features while taking the target values separately to get some insights.
- The plot of the target variable with count is as shown

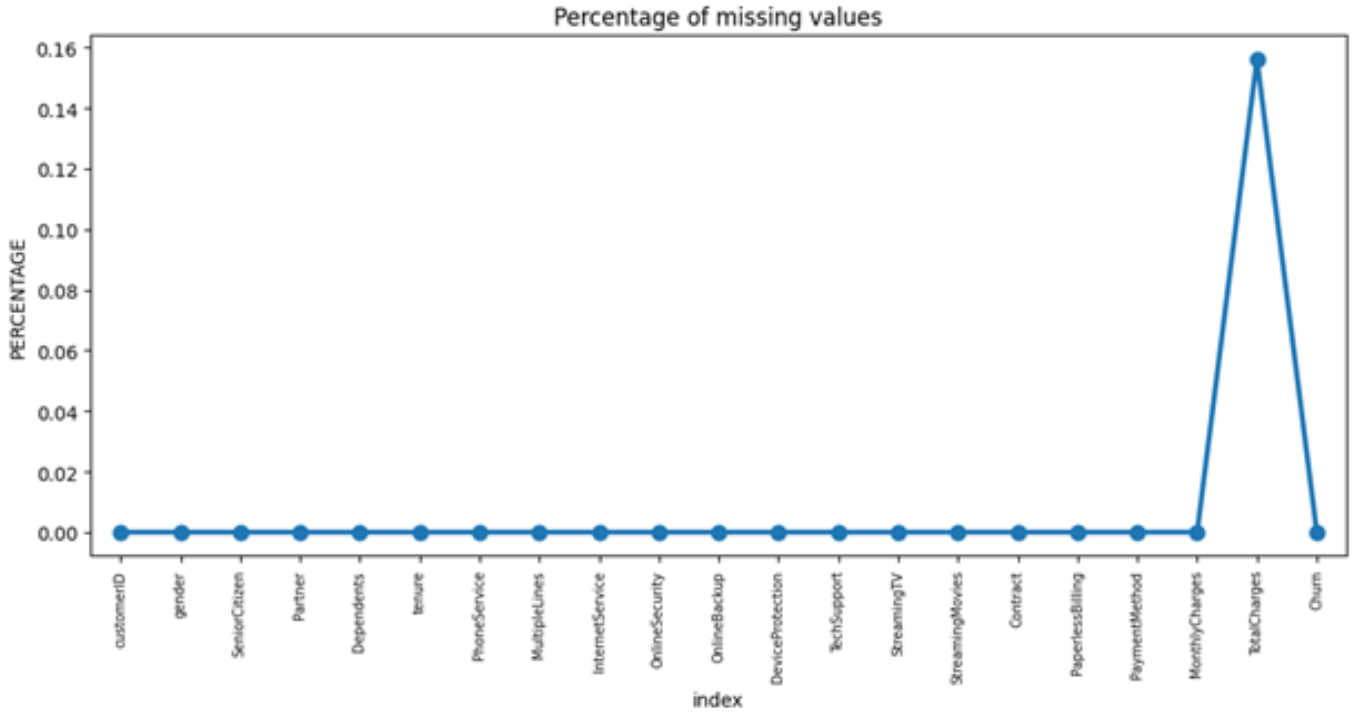


MISSING VARIABLE

- In Total charges there are 11 null values among the 7043 data.
- Missing variable plot is as shown

- **Missing Value Treatment**

Since the percentage of these records compared to the total dataset is very low i.e., 0.15% it is safe to ignore them or drop them from further processing



DATA DESCRIPTION

The data set contains 19 independent variables, which can be classified into 3 groups:

(1) **Demographic Information**

- gender: Whether the client is a female or a male (Female, Male).
- SeniorCitizen: Whether the client is a senior citizen or not (0, 1).
- Partner: Whether the client has a partner or not (Yes, No).
- Dependents: Whether the client has dependents or not (Yes, No).

(2) **Customer Account Information**

- tenure: Number of months the customer has stayed with the company (Multiple different numeric values).
- Contract: Indicates the customer’s current contract type (Month-to-Month, One year, Two year).
- PaperlessBilling: Whether the client has paperless billing or not (Yes, No).
- PaymentMethod: The customer’s payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit Card (automatic)).
- MontlyCharges: The amount charged to the customer monthly (Multiple different numeric values).
- TotalCharges: The total amount charged to the customer (Multiple different numeric values)

- **(3) Services Information**

- PhoneService: Whether the client has a phone service or not (Yes, No).
- MultipleLines: Whether the client has multiple lines or not (No phone service, No, Yes).
- InternetServices: Whether the client is subscribed to Internet service with the company (DSL, Fiber optic, No)
- OnlineSecurity: Whether the client has online security or not (No internet service, No, Yes).
- OnlineBackup: Whether the client has online backup or not (No internet service, No, Yes).
- DeviceProtection: Whether the client has device protection or not (No internet service, No, Yes).
- TechSupport: Whether the client has tech support or not (No internet service, No, Yes).
- StreamingTV: Whether the client has streaming TV or not (No internet service, No, Yes).
- StreamingMovies: Whether the client has streaming movies or not (No internet service, No, Yes).

STEPS INVOLVED IN THE PROCESS

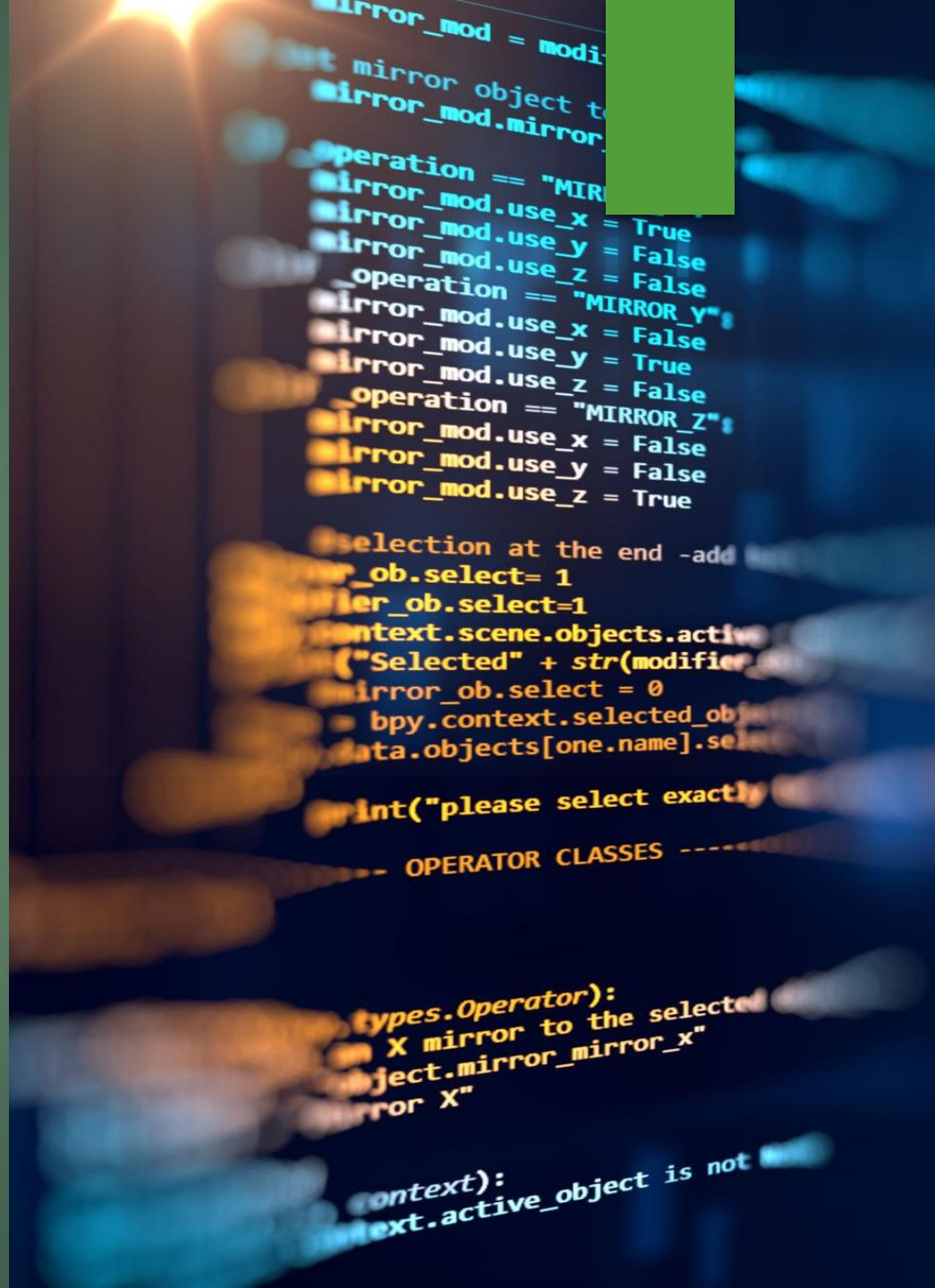
- Initial setup
- Data Inspection/Reading
- Data Cleaning
- Data preparation
- EDA (Exploratory Data Analysis)
- Data Visualization
- Univariate Analysis
- Bivariate Analysis
- Make Decisions
- Inference and Conclusions

INITIAL SETUP:

1. I have used Python programming language inside Google colab to perform the project.
2. Importing the required libraries.

DATA INSPECTION / READING

1. The first step of the analysis consists of reading and storing the data in a Pandas data frame using the pandas.read_csv function.
2. Visualize column names
3. Check the descriptive statistics of numeric variables
4. Data Wrangling
5. Check the distribution of the target variable now
6. Summary of the data frame



DATA CLEANING

1. changing the datatype
2. Check whether there are any missing values present in the data frame.
3. Check whether there are any missing values present in the data frame.
4. Length of the dataset before removing the missing values
5. Removing null/missing values(Since the percentage of these records compared to the total dataset is very low i.e., 0.15% it is safe to ignore them or drop them from further processing.

- DATA PREPARATION

- •Divide the Tenure

(Divide the data into bins based on tenure for e.g. tenure<12months, assume a tenure group of 0-12 and then 12-24 and so on....*)

- •Group the tenure to bins of 12 Months

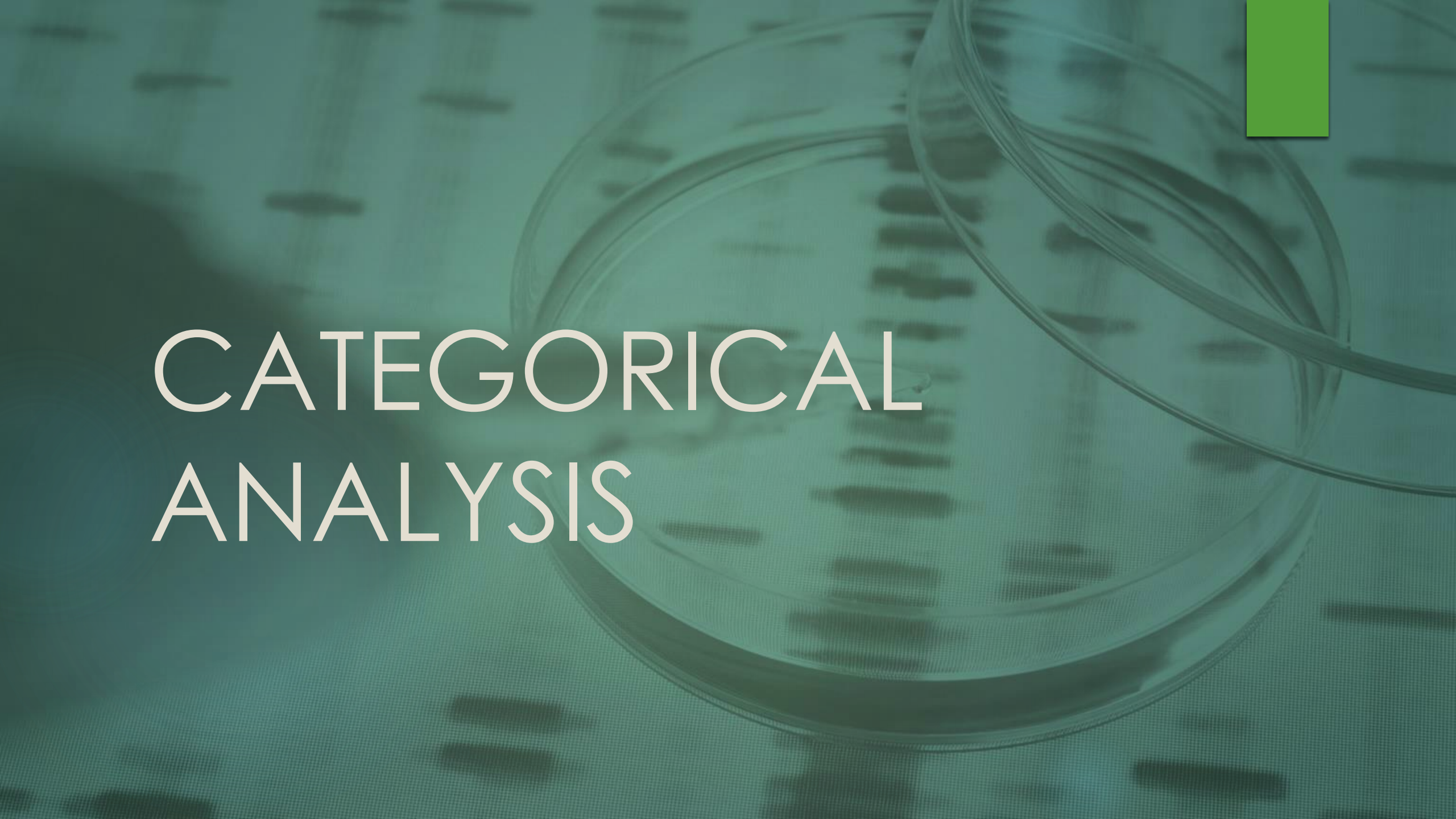
(We divide customers into bins based on tenure e.g., for tenure < 12 months: assign a tenure group of 1-12, for tenure between 1 to 2 Yrs., tenure group of 13-24; so on.)

- •Remove unnecessary Variables

(A clean data set is very important so; it is better to remove irrelevant variables like tenure (we divided them into bins) and customerID. These operations help us to create a clean dataset.)

- •Payment method denominations

(Some payment method denominations contain in parenthesis the word automatic. These denominations are too long to be used as tick labels in further visualizations. Therefore, we remove this clarification in parenthesis from the entries of the PaymentMethod column.)



CATEGORICAL ANALYSIS

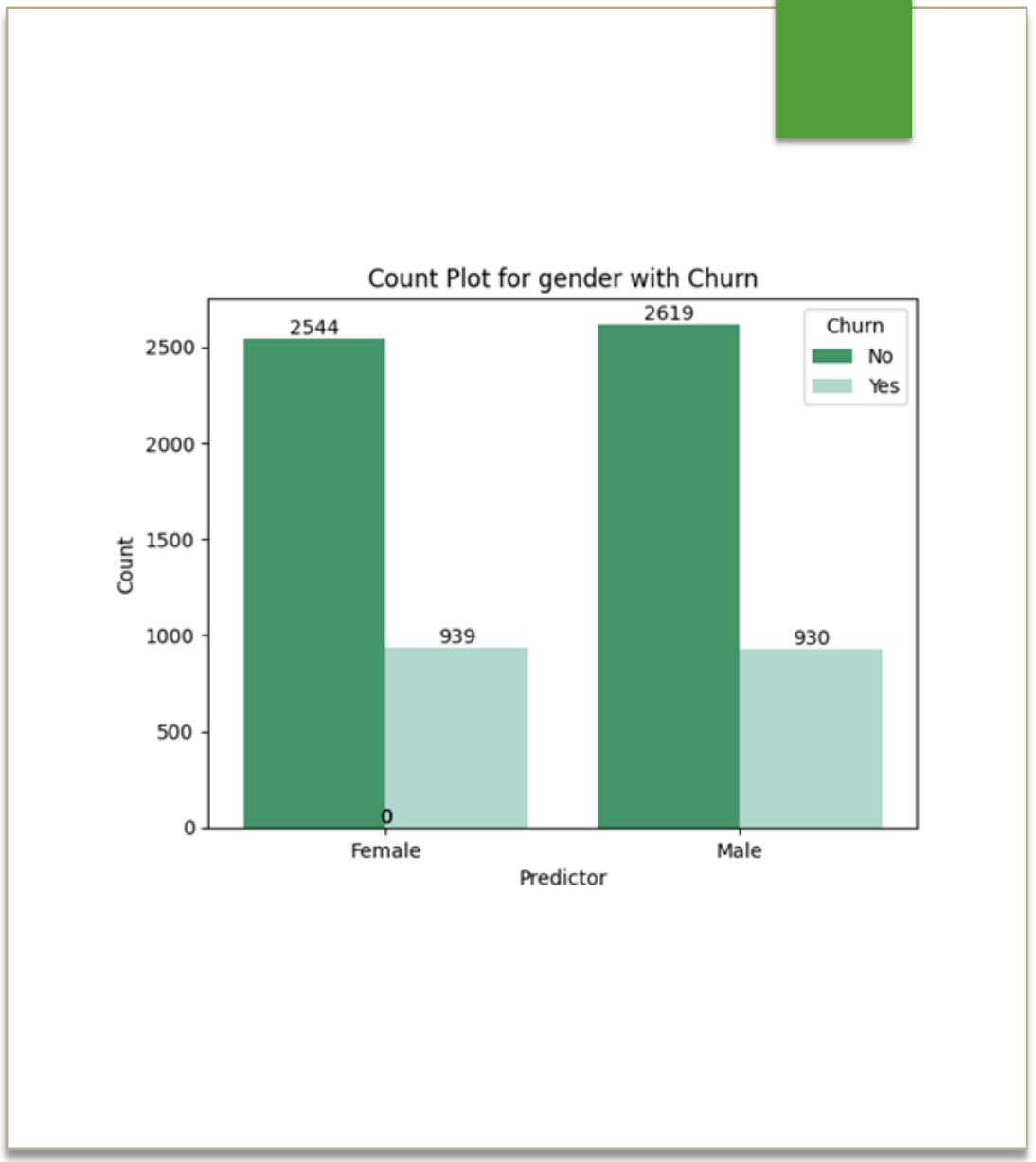
UNIVARIATE ANALYSIS

1.**Count plot for gender with Churn**

The gender variable is balanced in our dataset ,there are exactly 3549 male and 3483 female in 7032 ,looking at the figure above we can see that churn is equally present with in female as in males. The Churn Rate is as shown below

Female: 26.95%

Male : 26.20%

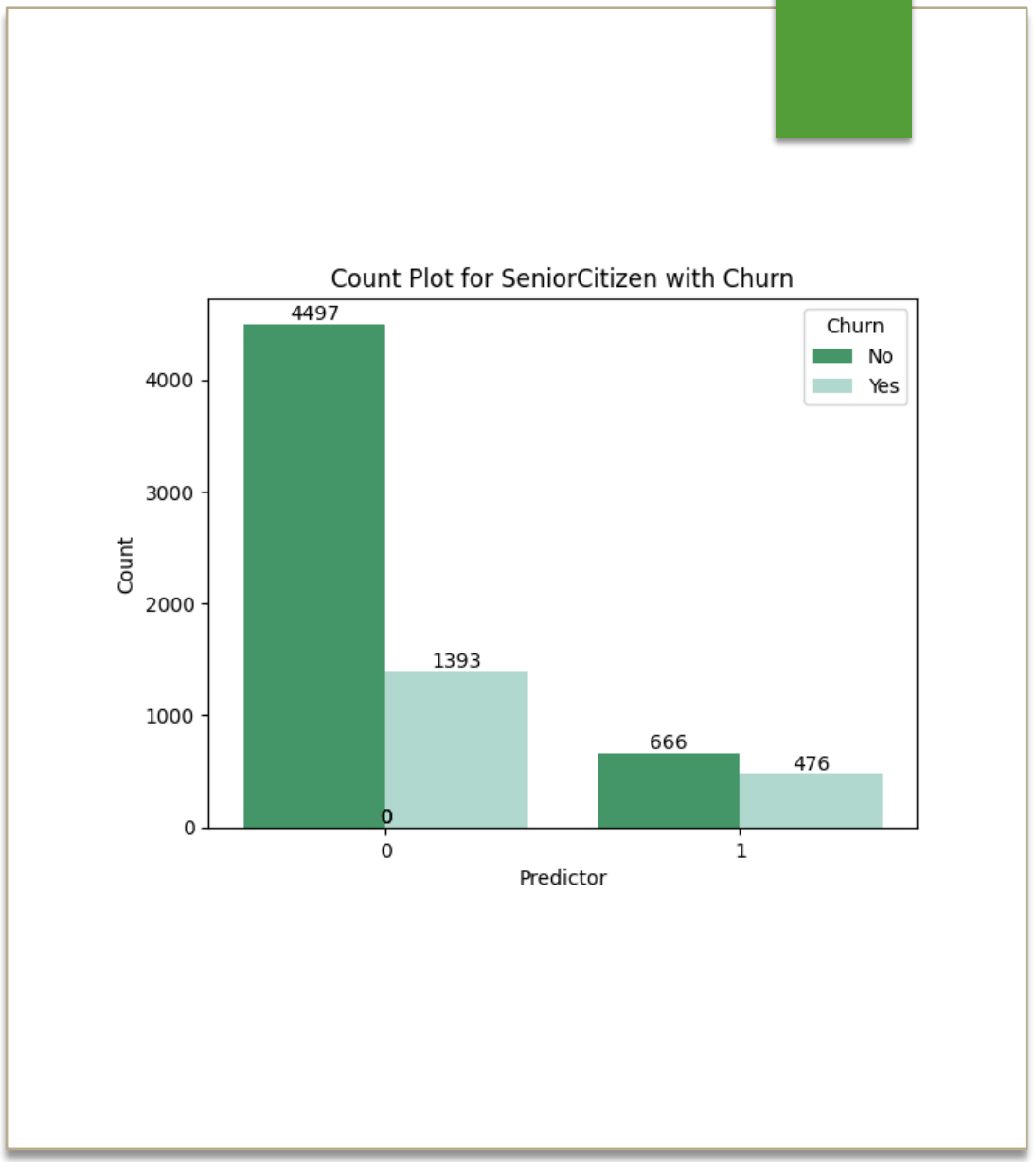


2. Count plot for Senior Citizen with Churn

Senior Citizen a binary feature that indicates whether the customer is an elderly person or not. Looking at the figure below we can see that the dataset contains fewer senior citizens (1142 from the total 7043 customers). The churn rate of seniors is almost double that of young customers.

Young Citizens: 23.65%

Senior Citizen : 41.68%

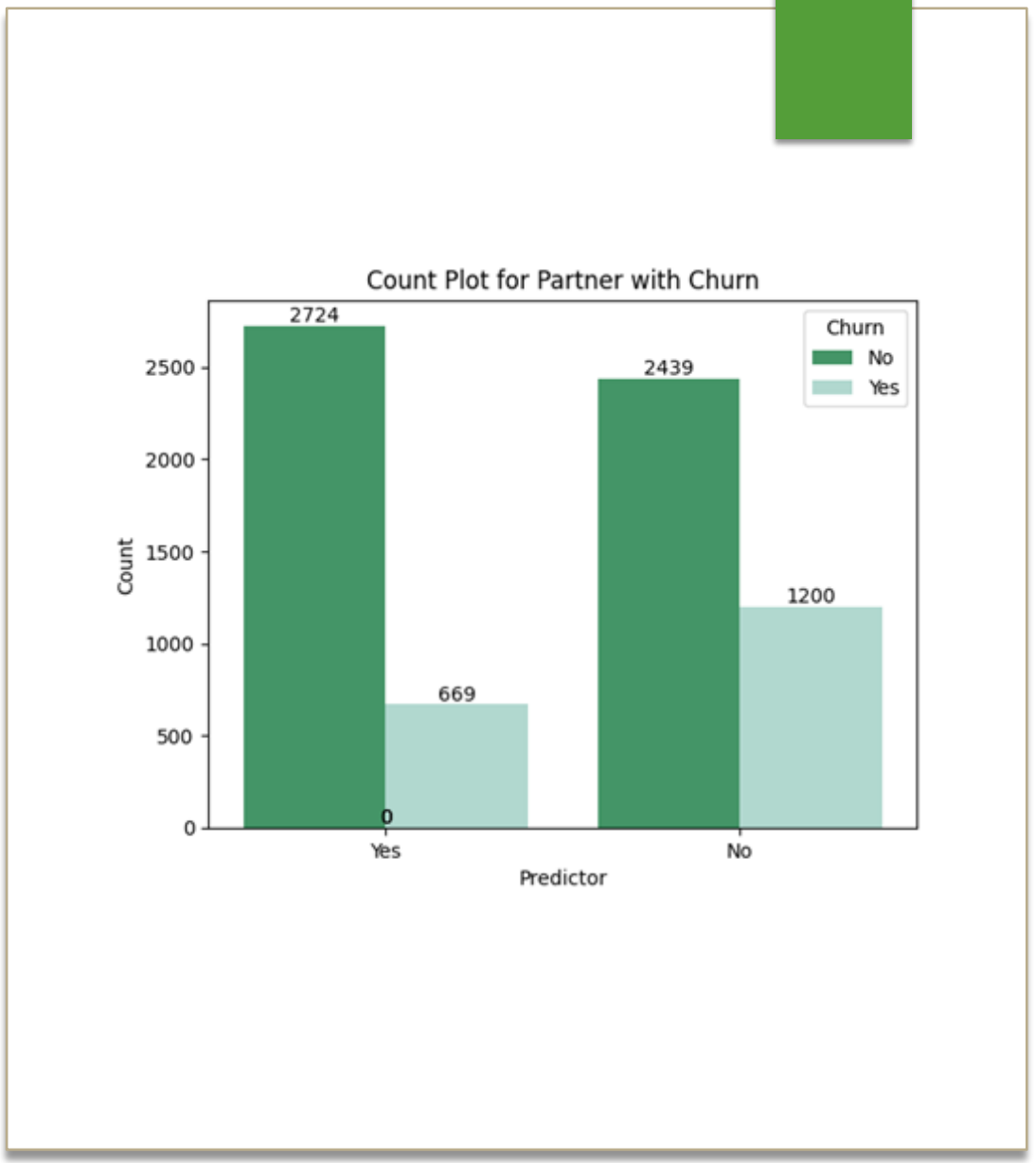


3. **Count plot for Partner with Churn**

The partner feature is a binary feature that indicates whether the customer has a partner or not. The figures below show how this variable is present in our data and how it relates to churn. The Churn Rate is more in customers with out partners. Customers without partners are churning more in comparison to customers with partners.

With Partner: 19.7%

Without Partner: 32.97%

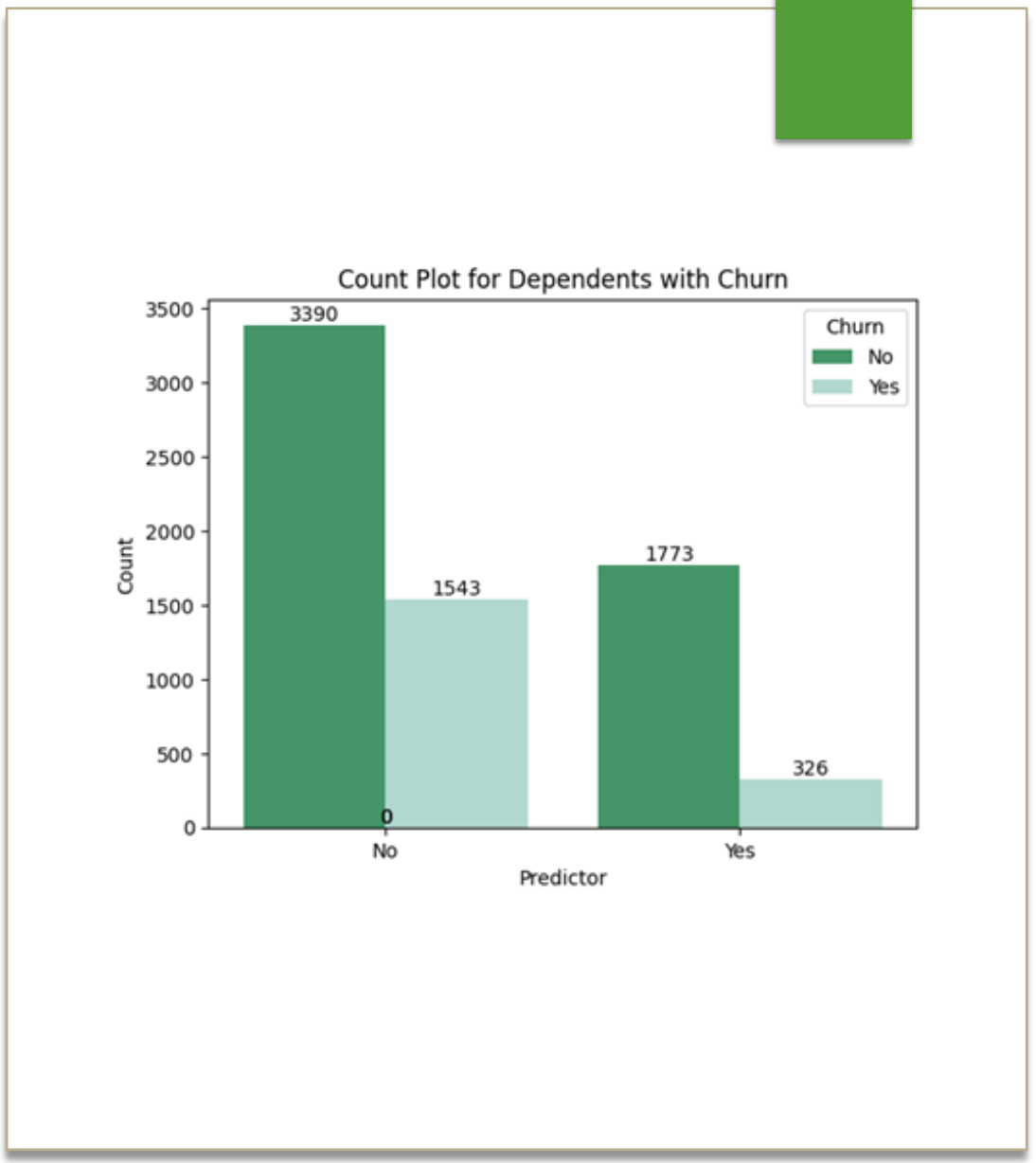


4. Count plot for Dependents with Churn

The Dependents feature is a binary variable that tells whether the customer is dependent or not. And we can see that most customers aren't dependent while churn is fairly present in both types. Dependent customers are less churners when compared to the non-dependents

Non-Dependent: 31.28%

Dependent: 15.53%

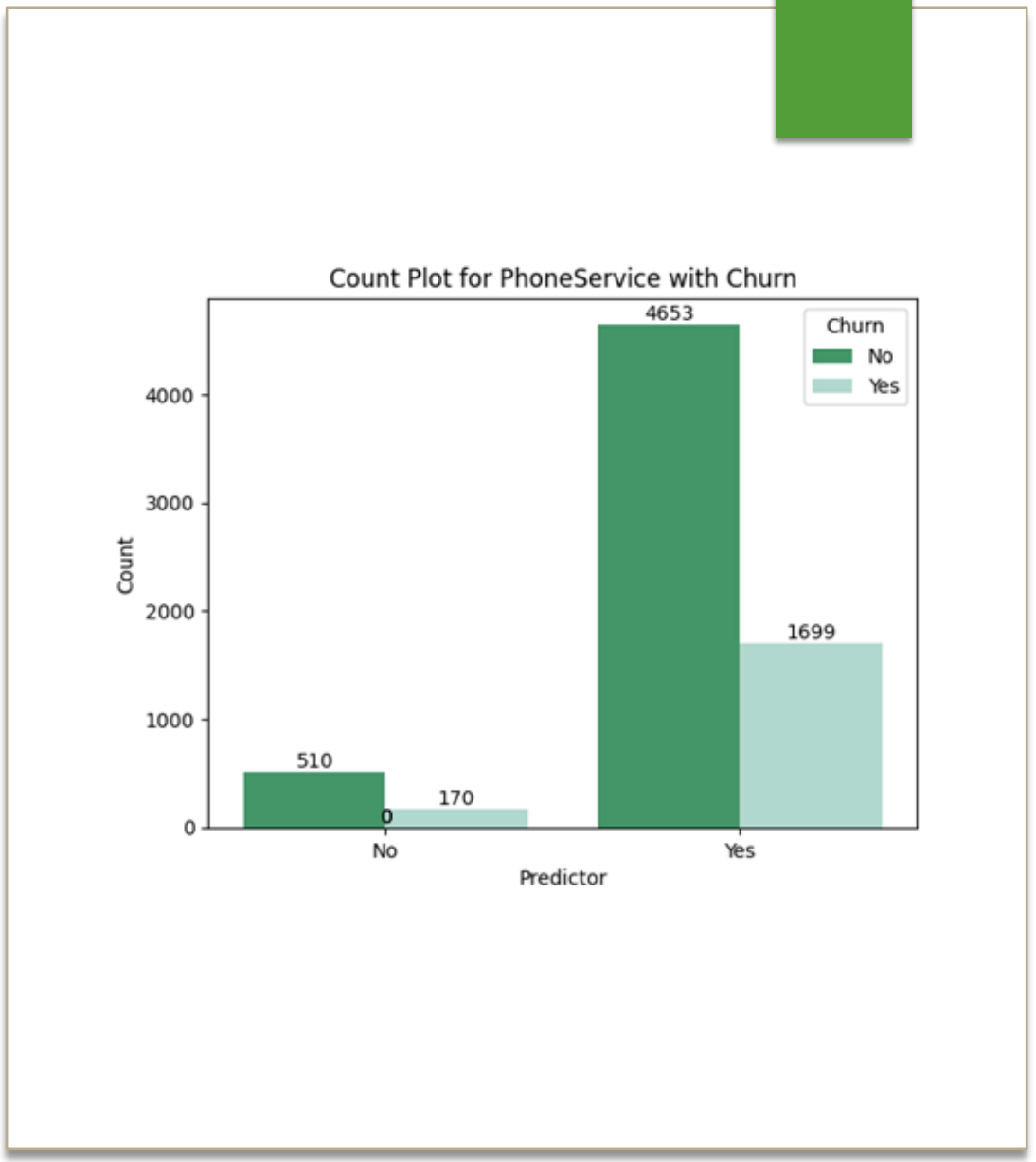


5. Count plot for Phone Service with Churn

The count plot for phone service is as shown below, it's evident that the people with more phone service are like to churn. The churn rate is slightly more for people with phone service

Without phone service :25%

With phone service: 26.74%



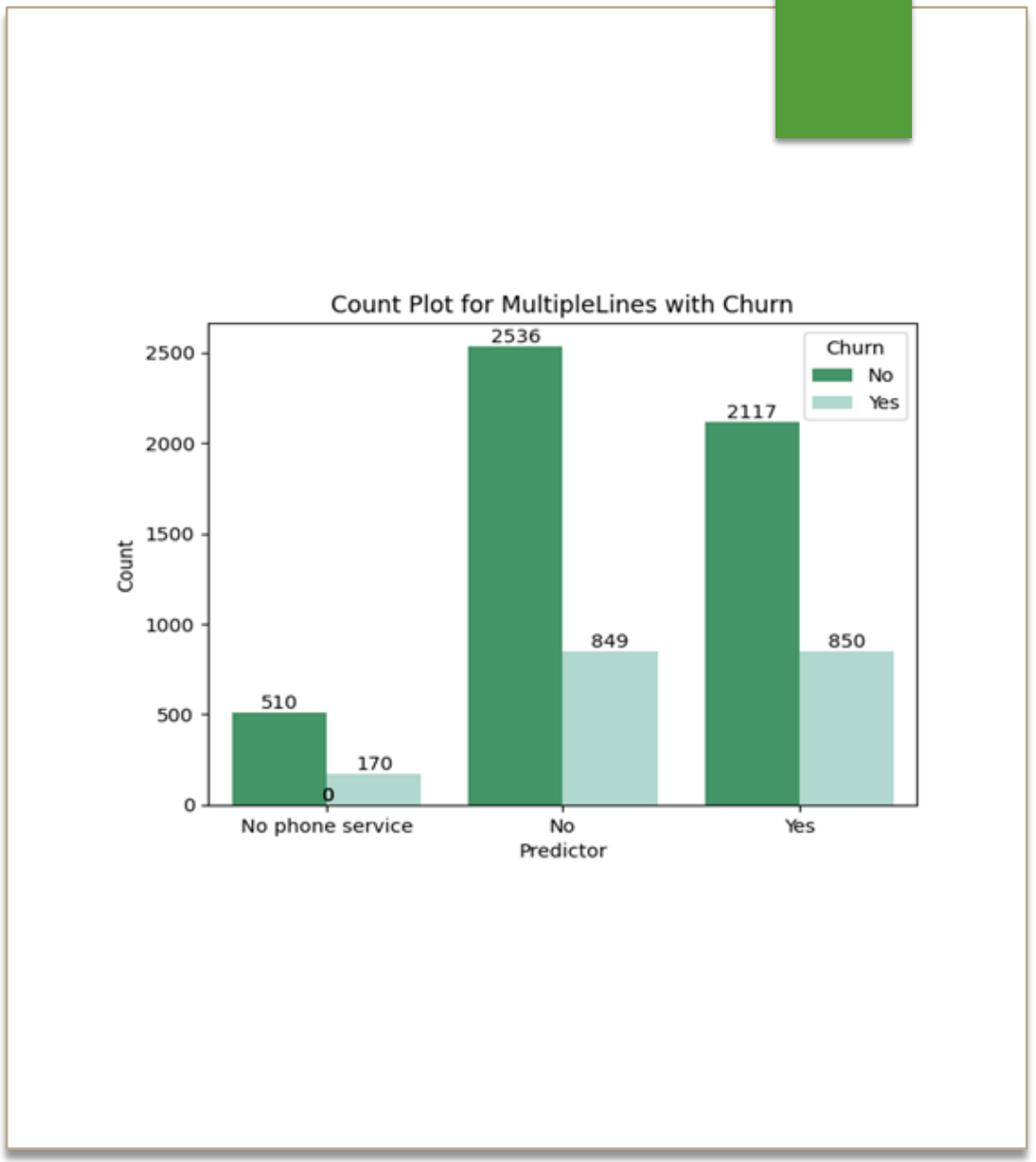
6. Count plot for Multiple Lines with Churn

We do not expect phone attributes (PhoneService and MultipleLines) to have significant predictive power. The churn rate for customers with

No Phone service: 25%

Without Multiple Lines: 25.08%

With Multiple Lines: 28.65%



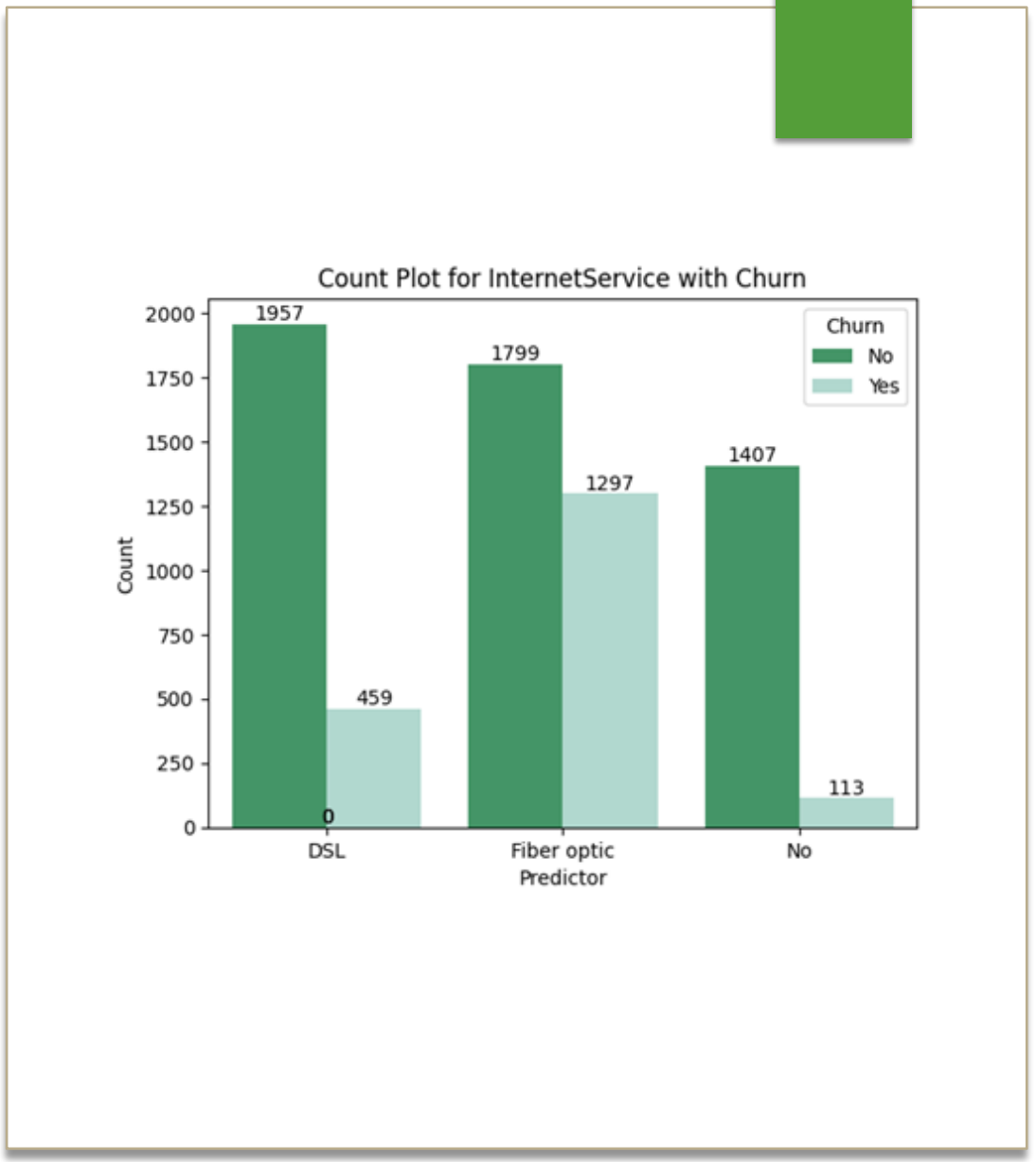
7. Count plot for Internet Service with Churn

The customers with Fiber optic internet service are having a high churning rate. The churn rate for customers with

No Internet: 7.4%

With DSL: 18.99%

With Fiber optics: 41.89%



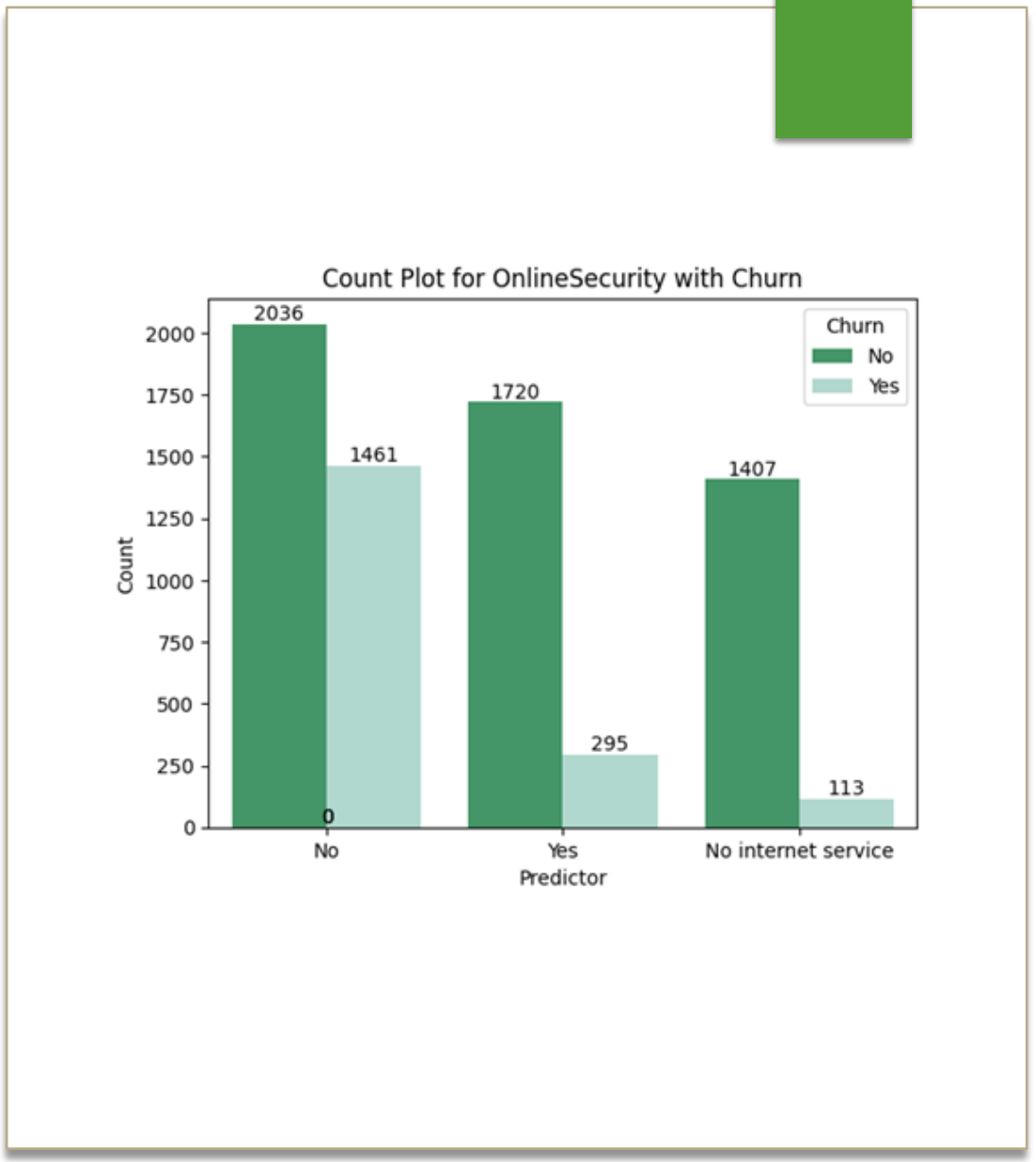
8. Count plot for Online Security with Churn

The churn rate for customers without online security is very high.

No Internet: 7.4%

With Online Security:14.6 %

Without Online Security:41.79%



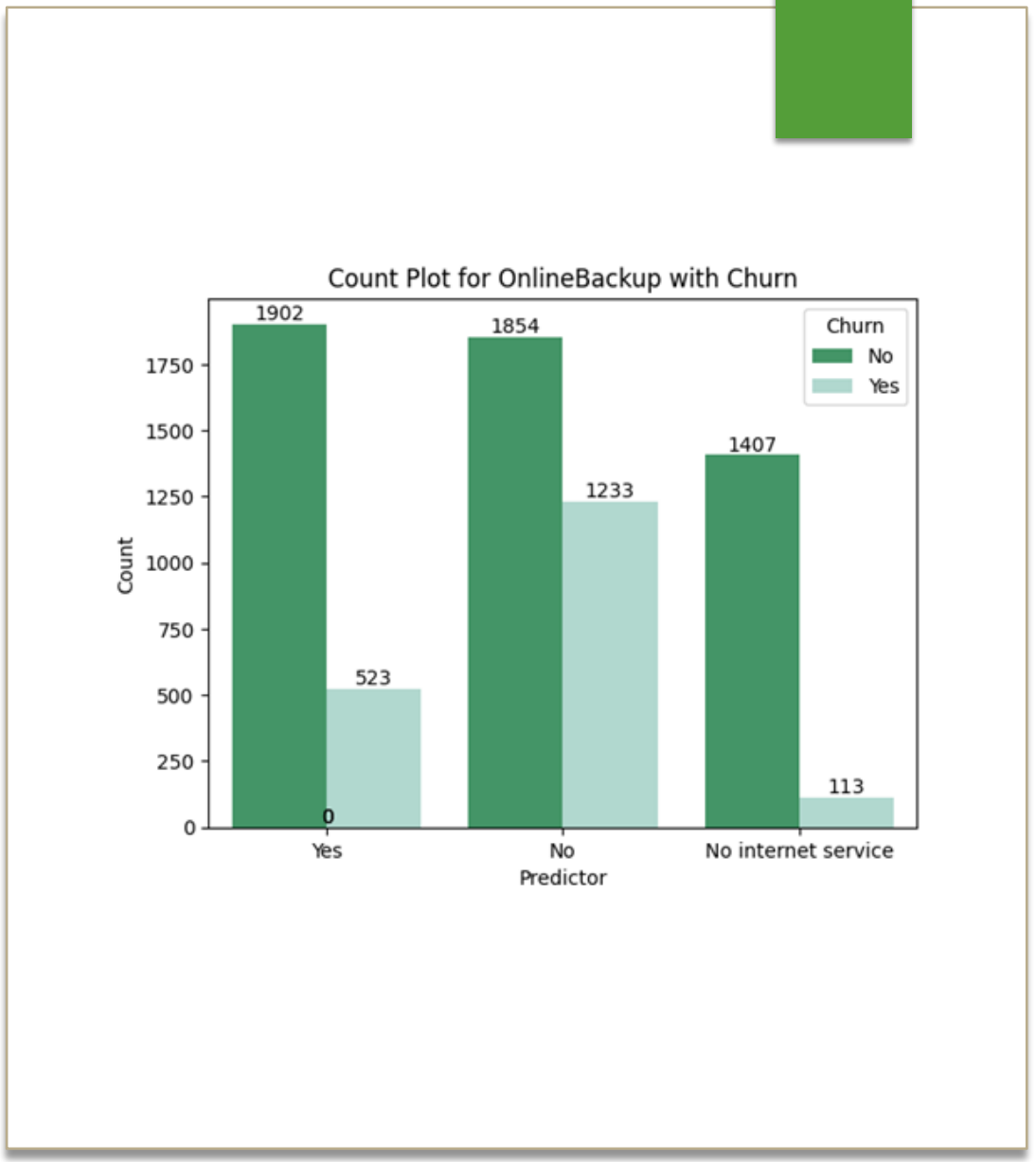
9. Count plot for Online Backup with Churn

The customers with no internet service are churning the least. The churn rate for customers without Online Backup is high.

No Internet: 7.4%

With Online Backup: 21.57%

Without Online Backup: 39.94%



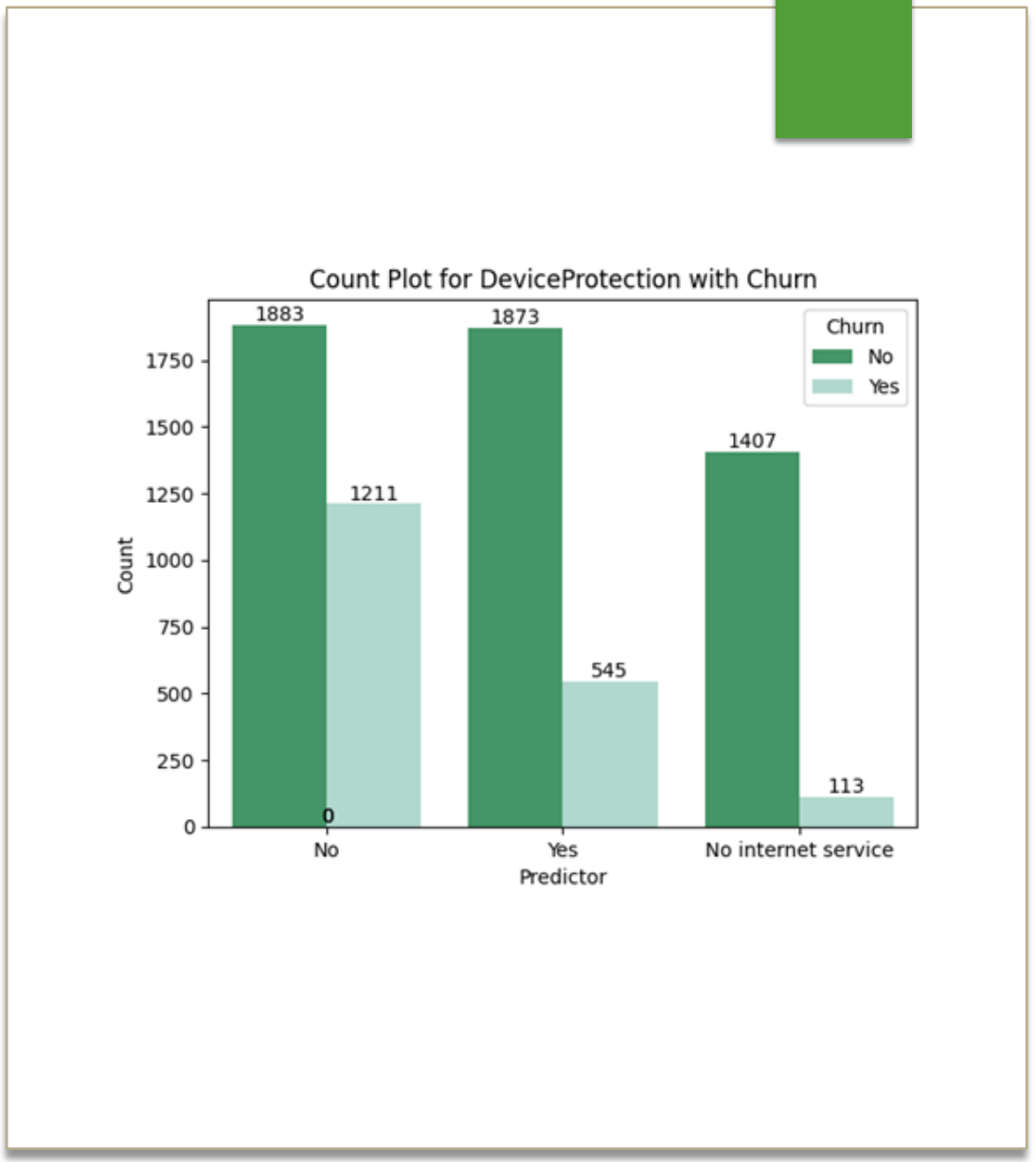
10. Count plot for Device protection with Churn

The churn rate for customers Without Device Protection is high compared with customers with Device protection.

No Internet: 7.4%

With Device Protection: 21.57%

Without Device Protection: 39.94%



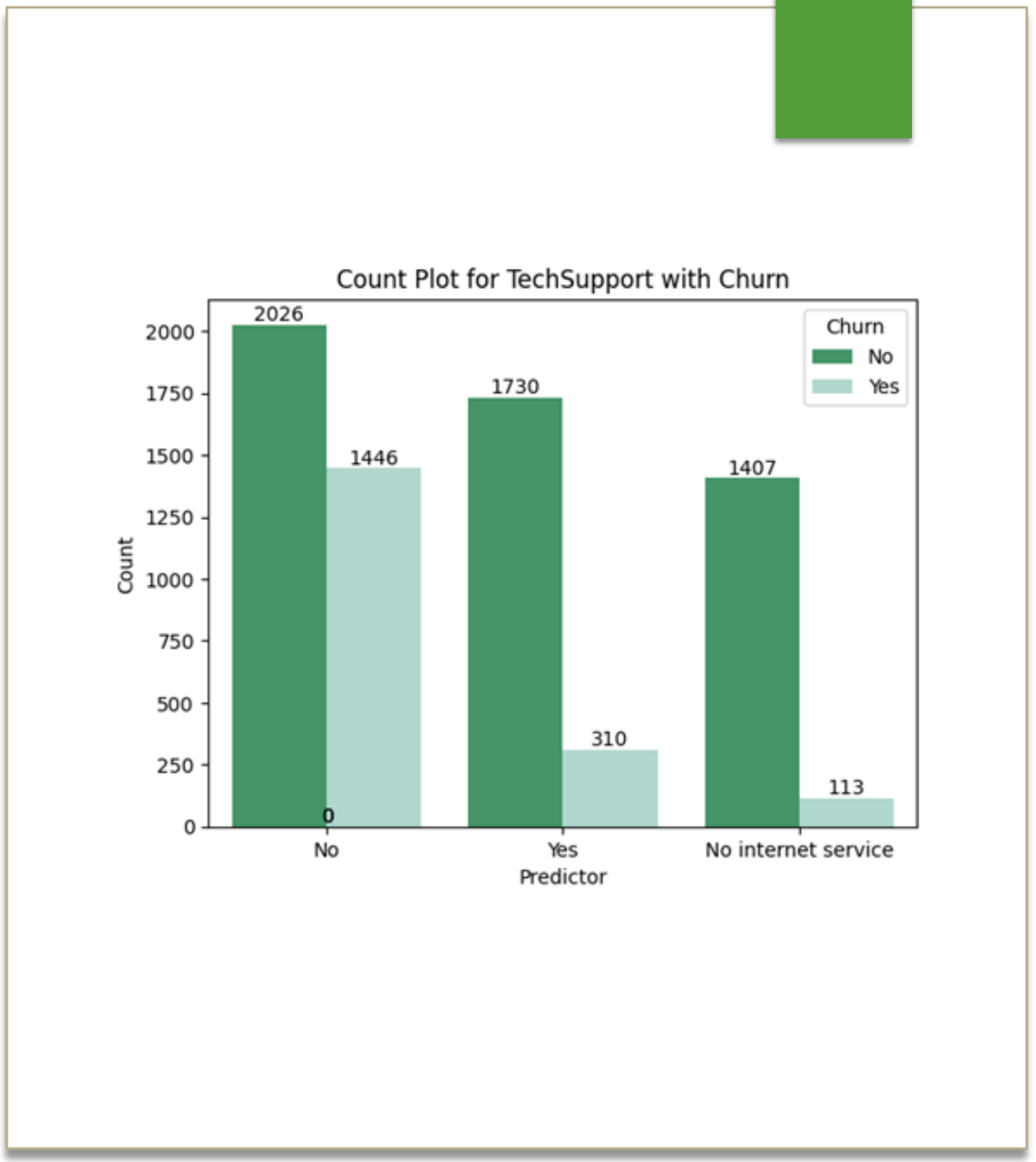
11. Count plot for Tech Support with Churn

Customers with no tech support are churning more. The churn rate for customers with

No Internet: 7.4%

With Tech Support: 15.20%

Without Tech Support: 41.65%



12. Count plot for Streaming TV with Churn

Customers with no Streaming TV are churning more. The churn rate for customers with

No Internet: 7.4%

With Streaming TV: 30.11%

Without Streaming TV: 33.54%

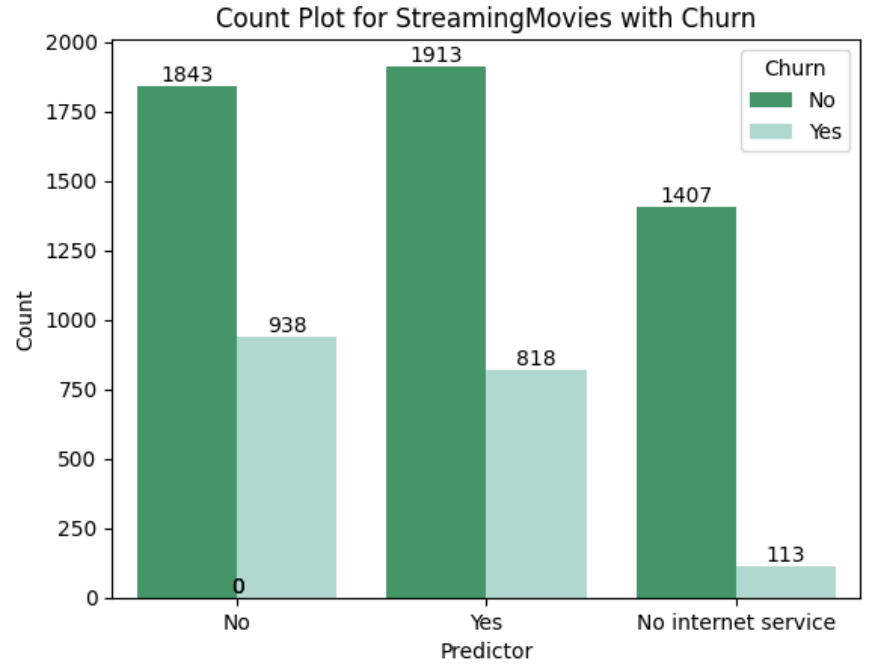
13. Count plot for Streaming Movies with Churn

Customers with no Streaming movies are churning more. The churn rate for customers with

No Internet: 7.4%

With Streaming Movies: 29.95%

Without Streaming Movies: 33.73%



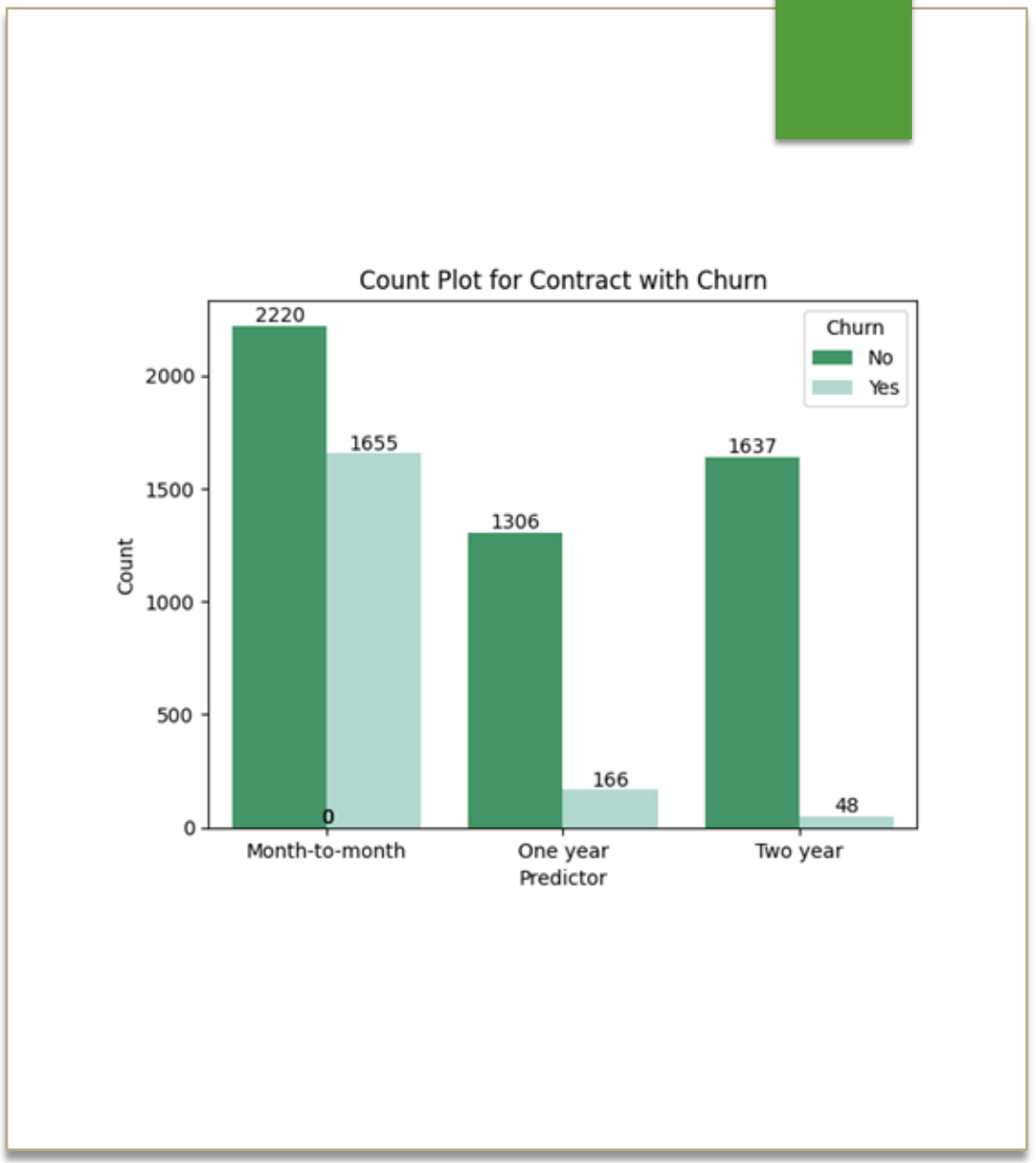
14. Count plot for Contract with Churn

Customers with Month-to-Month contract are churning more. The churn rate for customers with

Month - Month: 42.71%

One-Year Contract: 11.28%

Two-Year Contract: 2.85%



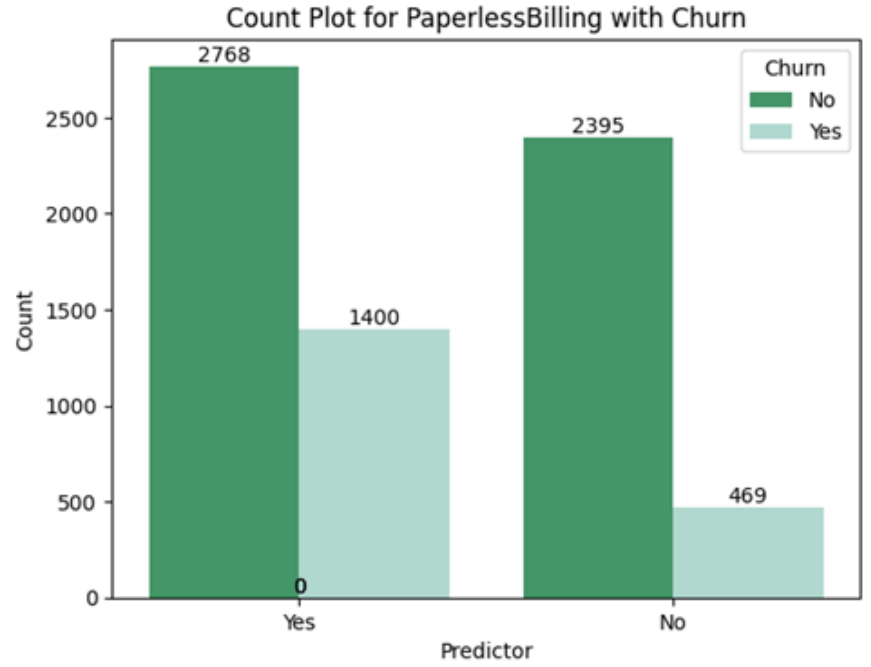
15. Count plot for Paperless Billing with Churn

Customers who use Paperless Billing are churning more in comparison to their counterparts.

The churn rate for customers with

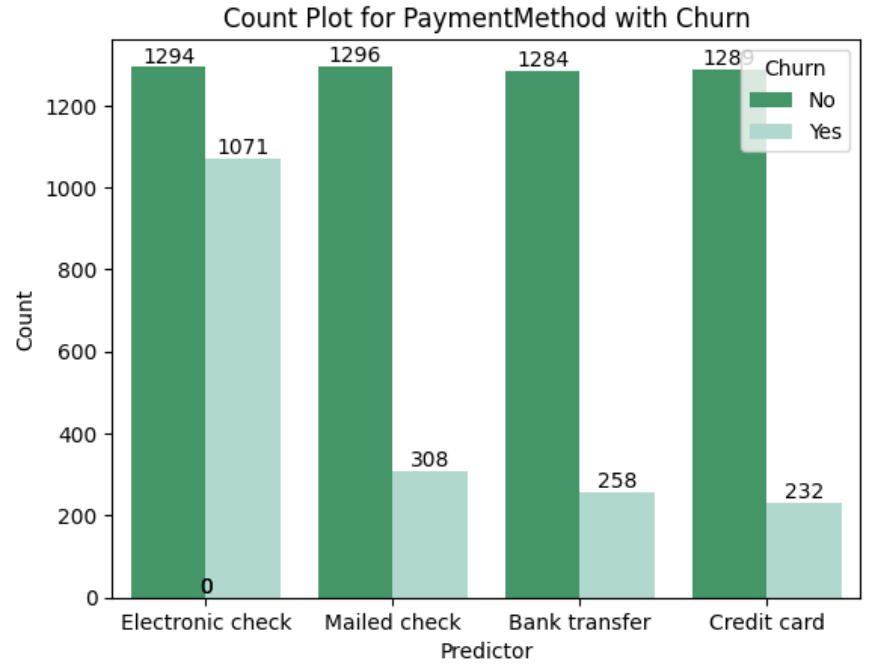
PaperlessBilling: 33.59 %

Paper Billing: 16.38%



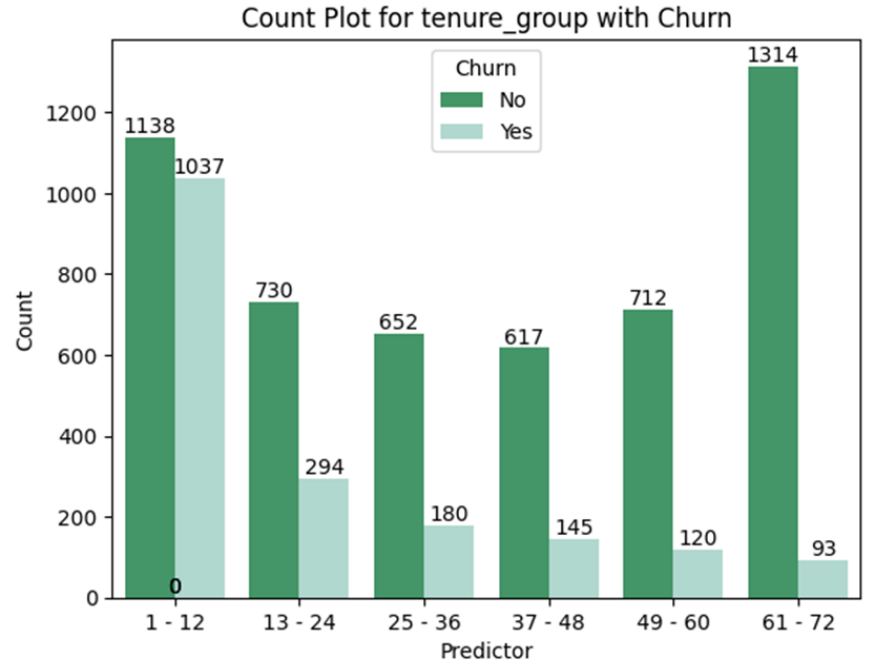
16. Count plot for Payment Method with Churn

- PaymentMethod feature tells what's the method that customer use to pay his bills , there are four methods of payment ; Electronic check , mailed check , bank transfer and credit card.
- The churn rate for customers with
  - Electronic Check: 45.29%
  - Mailed Check: 19.20 %
  - Bank Transfer: 16.73%
  - Credit Card: 15.25%
- From the graph and from the churn analyses it’s clear that the customer opting for Electronic Check are churning more when compared to the customer using other payment methods.



17. Count plot for Tenure group with Churn

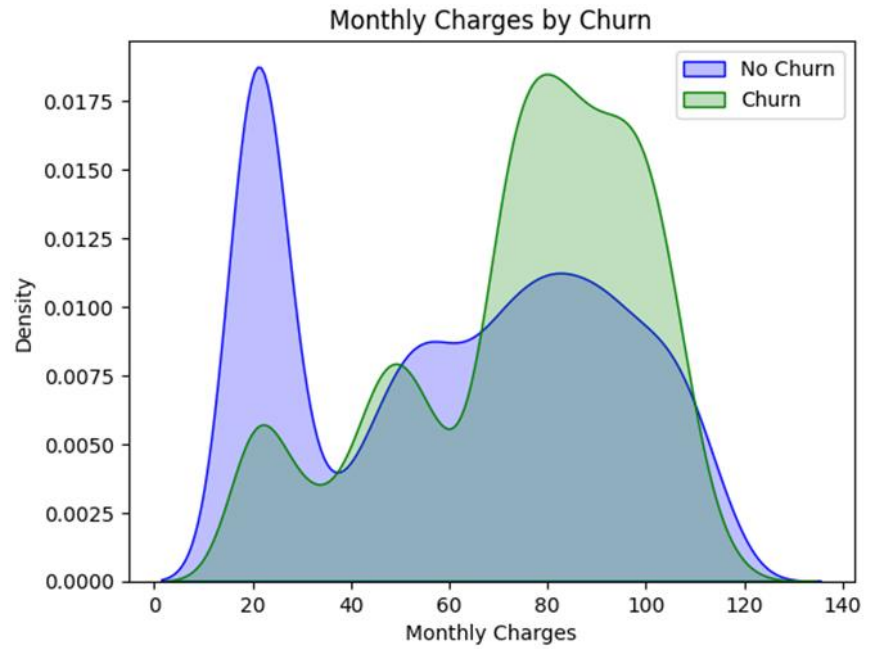
- Here we are analyzing customers with which tenure are churning more in comparison to their counterparts. The churn rate for customers with each tenure is as shown below.
- From the graph and also from the churn rate it is evident that the customer with less tenure is more likely to churn. The least churners are there in the tenure group of 61 – 72 Months and the top churners are there in the tenure group of 1 - 12 Months.
- **Tenure-Group**
  - 1 - 12: 47.68%
  - 13 - 24: 28.71%
  - 25 – 36: 21.63%
  - 37 – 48: 19.03%
  - 49 – 60: 14.42%
  - 61 – 72: 6.60%



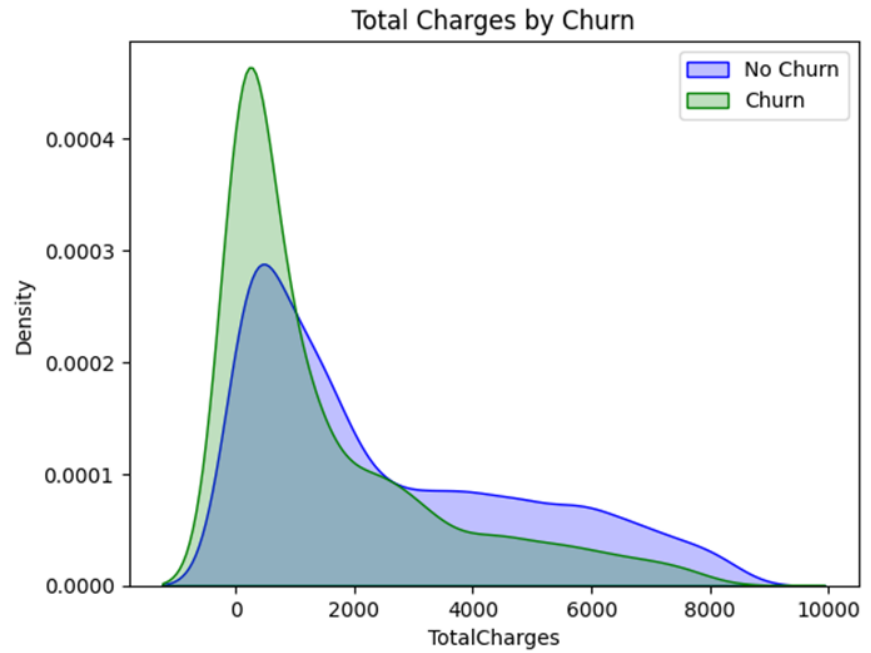
NUMERICAL ANALYSIS

Data Visualization will give better insights into the relationship between variables, distribution of the certain variables on the graph, etc. I have tried to use correlation plots, univariate, and bivariate plots.

Inference : From the graph it is clear that if the monthly charges are more the churn rate is higher

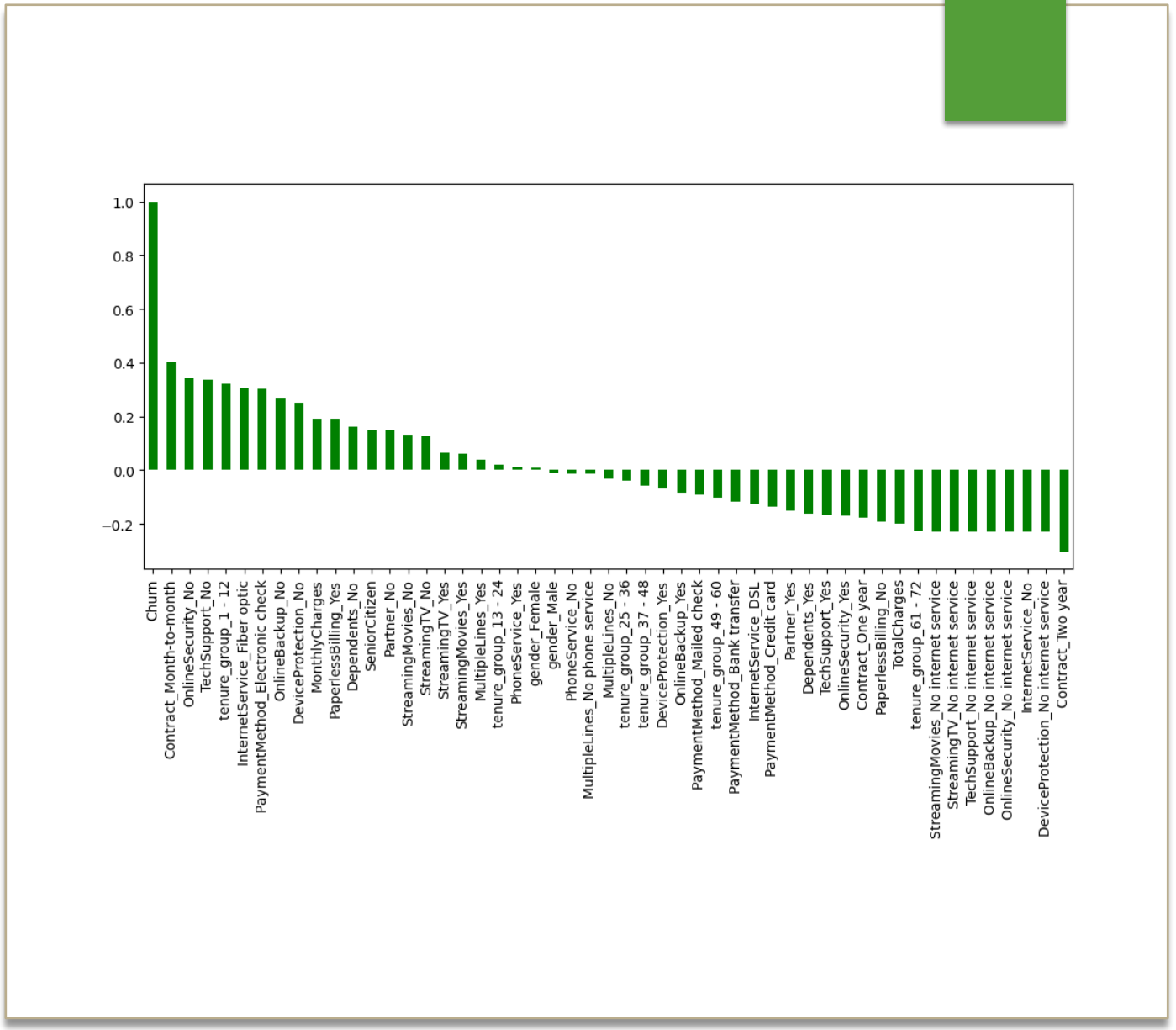


Inference : From the graph we can see that if the total charges are less the churn rate is high



CORRELATION PLOT

Plotting the correlation plot of the entire data w.r.t Churn

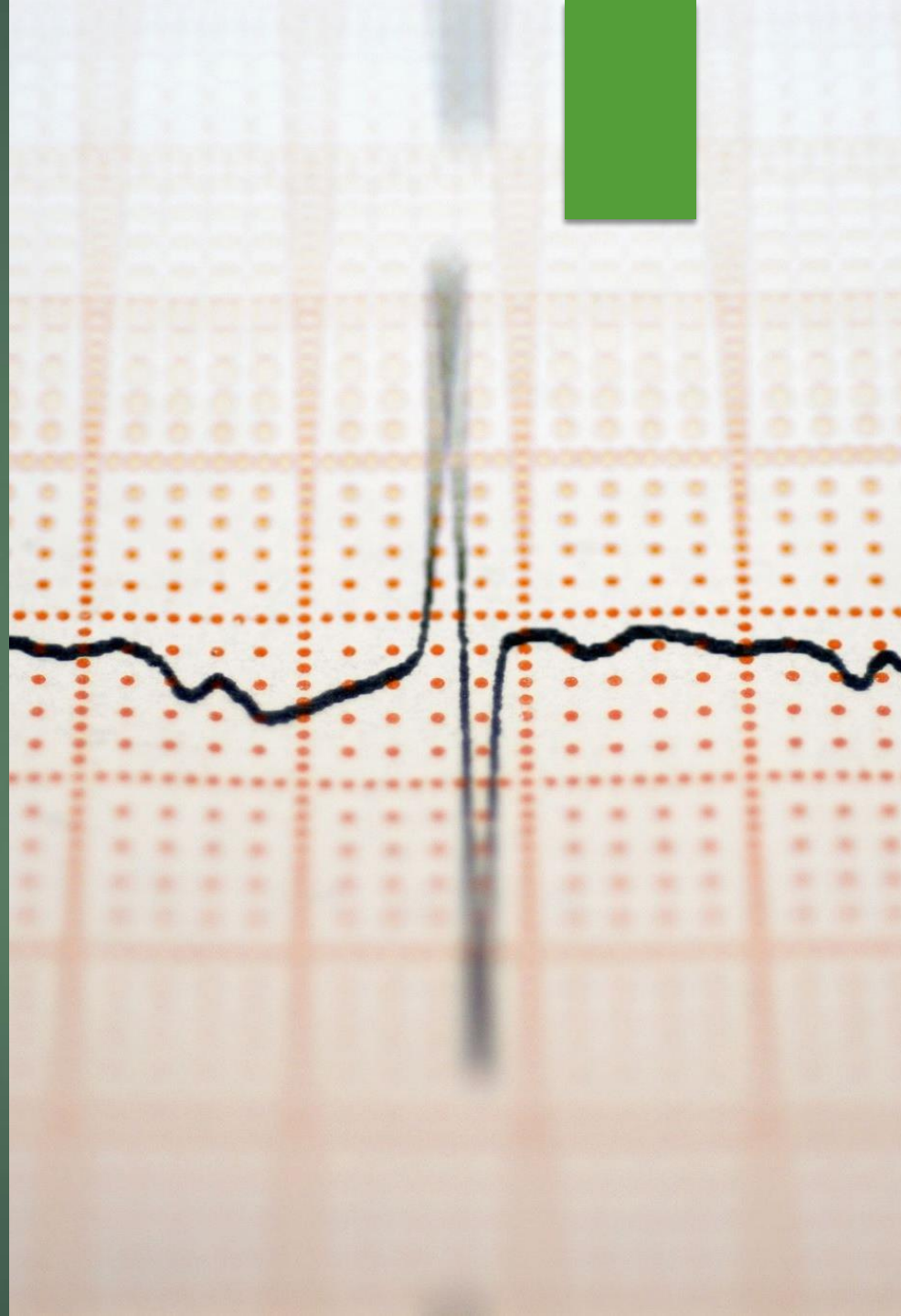


Initial inference from the correlation plot

- High Churn is seen in the case of Senior Citizens, Customers without partners, Month-to-month contracts, No online security, No Tech support, First year of subscription, and Fibre Optics Internet
- Low Churn is seen in the case of long-term contracts, Subscriptions without internet service, and the customers engaged for more years
- Factors like Gender, Availability of PhoneService, and number of multiple lines have almost no impact on Churn

BIVARIATE ANALYSIS

- There are two variables here, as bi means two and variate implies variable. The study is focused on the root of the problem as well as the relationship between the two variables.
- Bivariate analysis is classified into three types.
  - Bivariate Analysis of two Numerical Variables (Numerical-Numerical) we do Scatter plot, Correlation.
  - Bivariate Analysis of one numerical and one categorical variable (Numerical-Categorical), we do ANOVA
  - Bivariate Analysis of two categorical Variables (Categorical-Categorical), we do Chi2 Test.



1. Distribution of Gender for Churners and Non-Churners

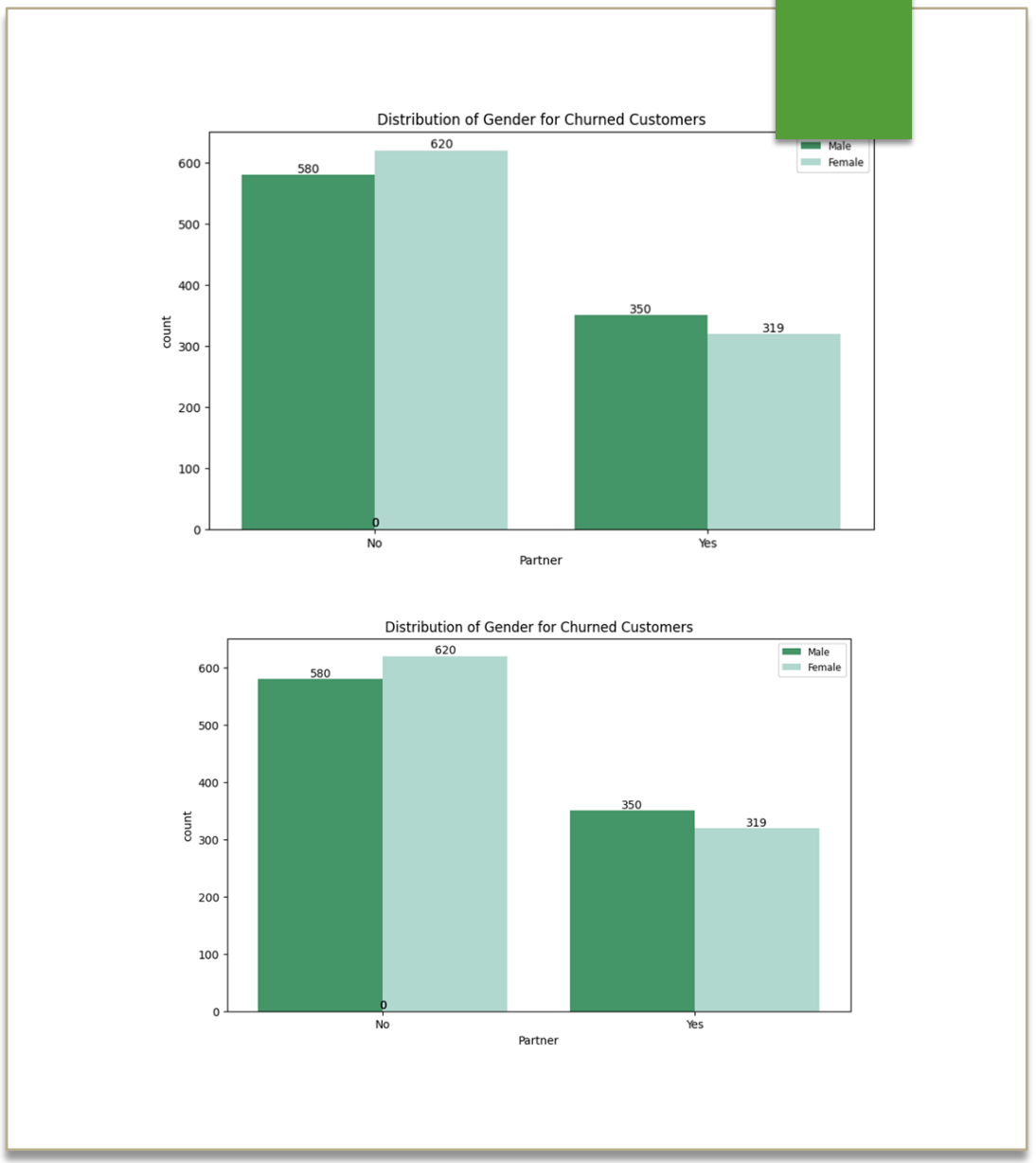
From this visualization we can see that gender along with another parameter called partner is giving more insights compared to gender alone plot. Both females and males who do not have a partner are more likely to churn. Female with partners are non-churners. The retain rate for customers with partner and no partner is almost the same. But if we look, the churn rate has increased in females with no partners when compared to females with partners.

M/P = 20.4%

M/NoP = 31.54%

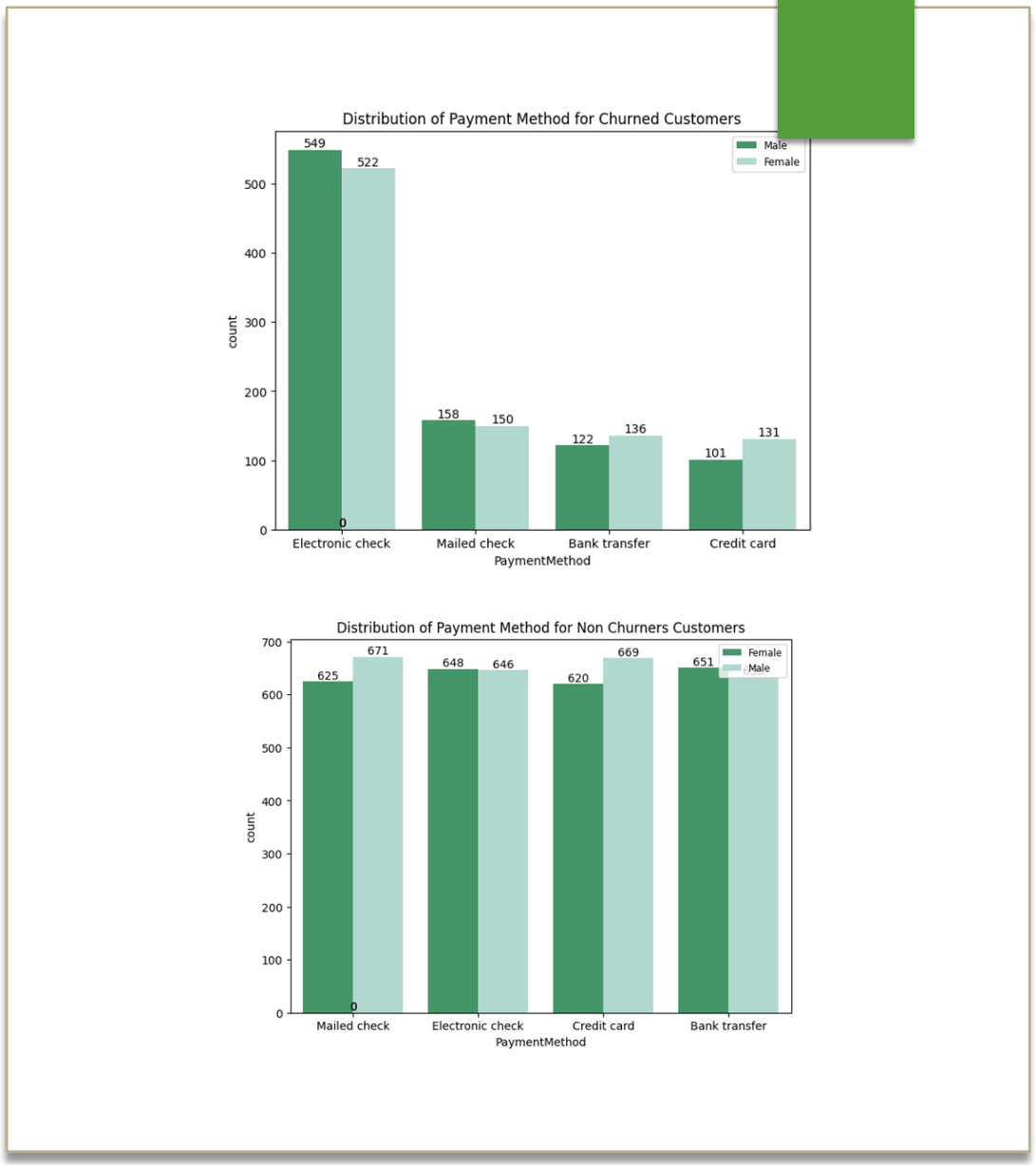
F/P = 18.95%

F/Nop = 34.44%



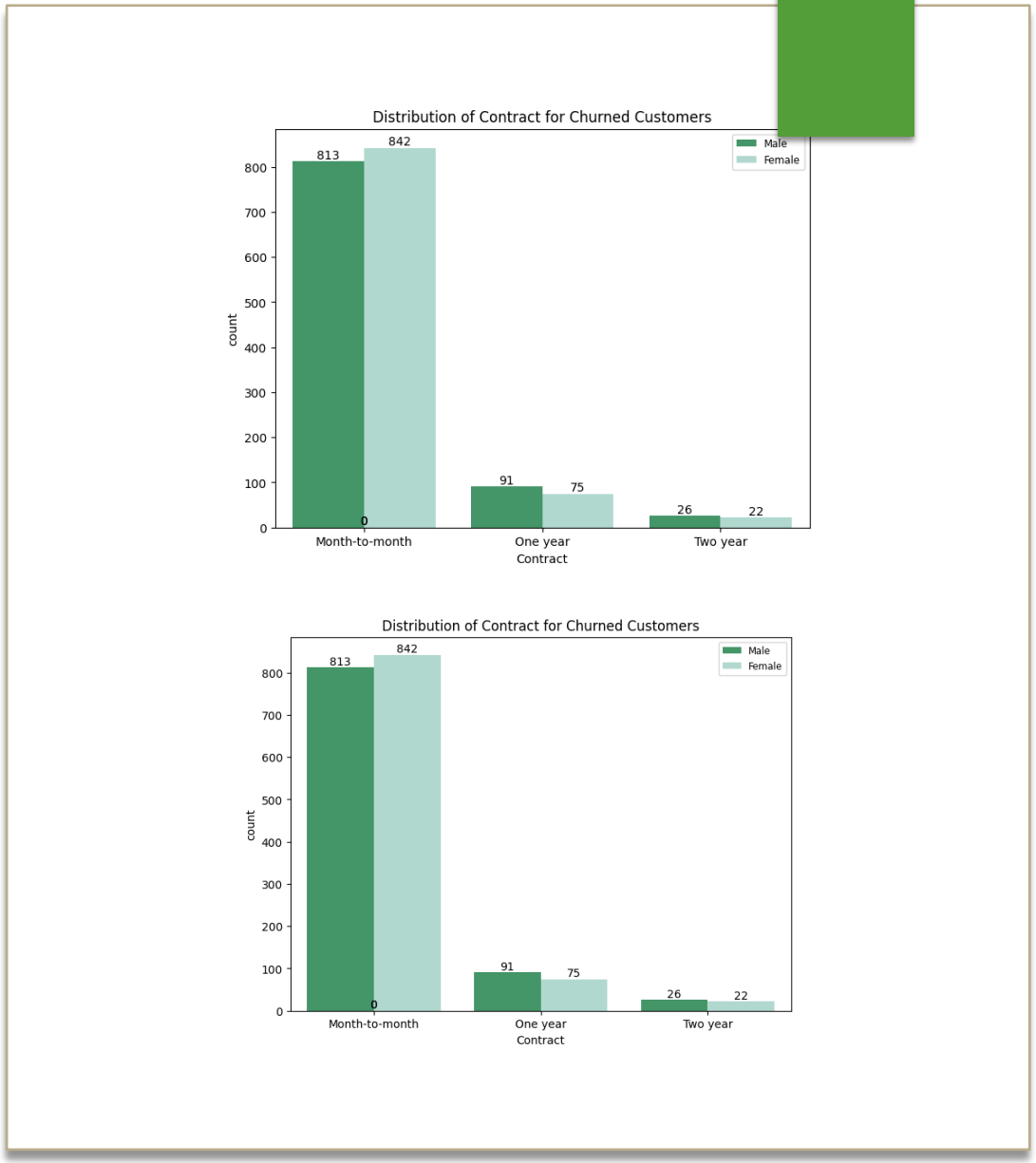
2. Distribution of Payment Method for Churners and Non-Churners

From this visualization we can see that the churn rate for customers with payment option electronic check has more churners in that the male and female churners are almost the same but slightly male customers are churning more when compared to the female. But customers opting the payment method credit card has more female churners when compared to the male customers.



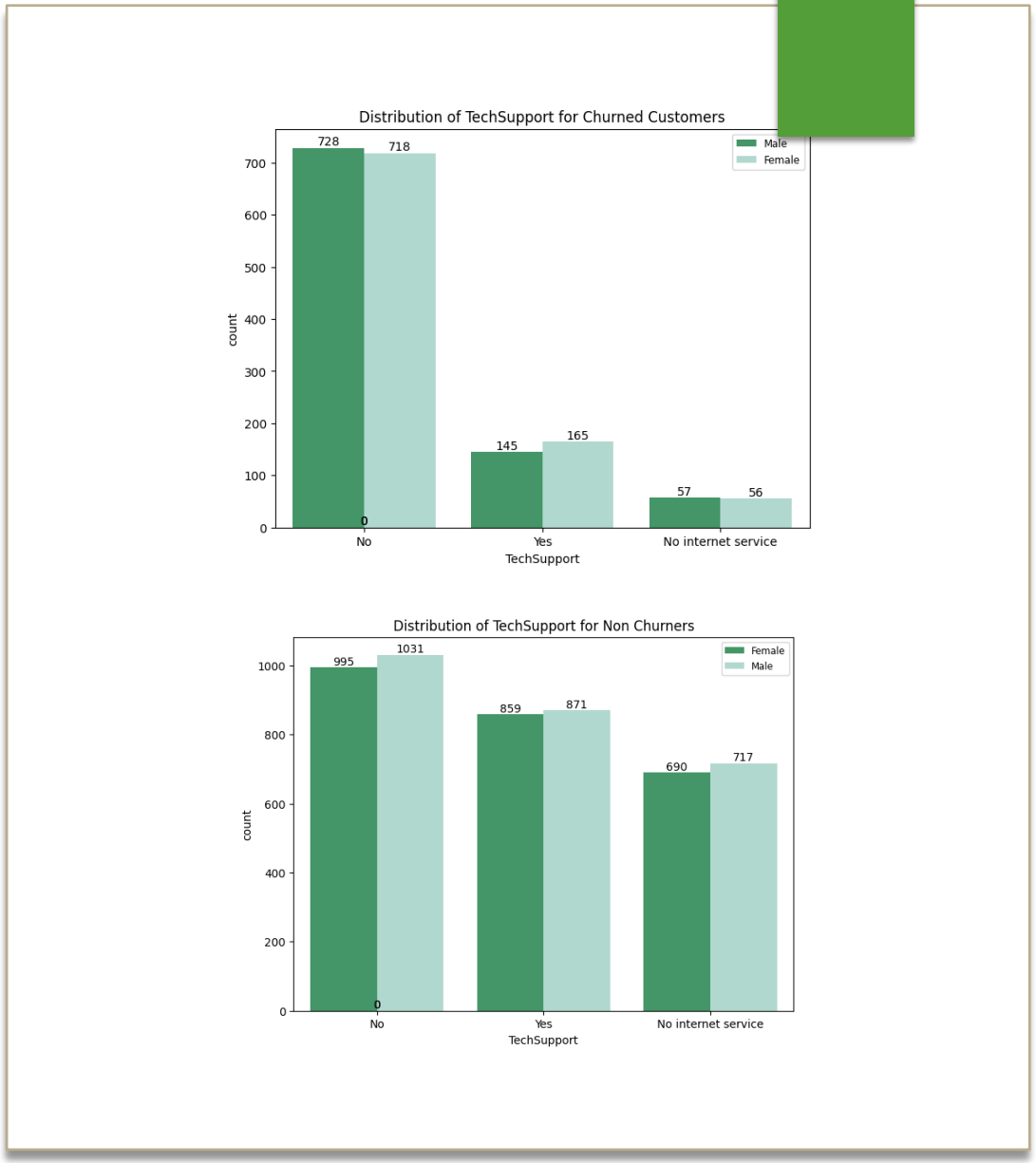
3. Distribution of Contract for Churners and Non-Churners

From this visualization we can see that we can see that the churn rate for customers with month – month contract has more female churners and the churn rate is more for males with one year contract.



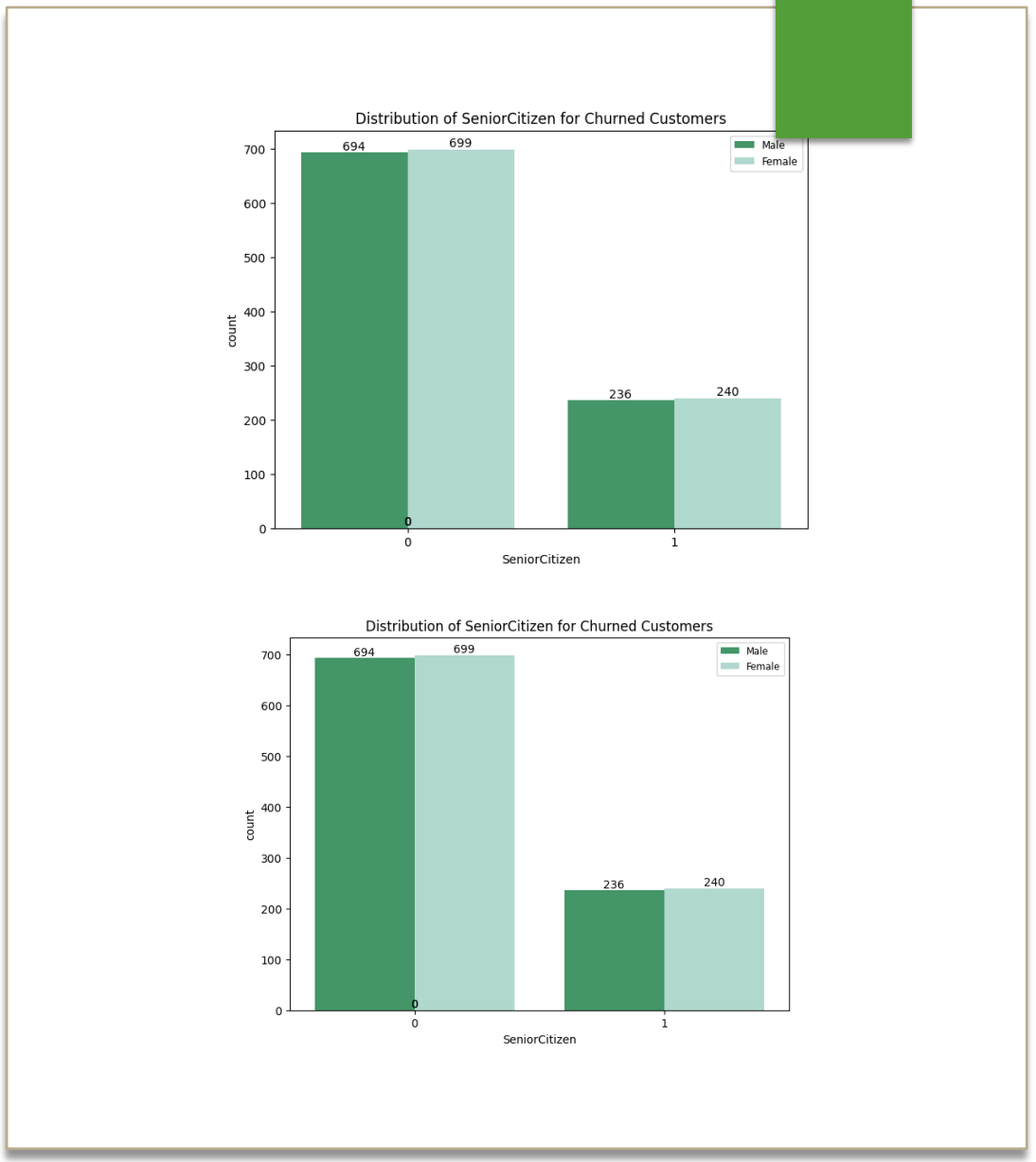
4. Distribution of TechSupport for Churners and Non-Churners

From the visualization above, we can see that the churn rate for male and female customers with tech support are almost the same but the churn rate for females with tech support has a higher value when compared to the males with tech support



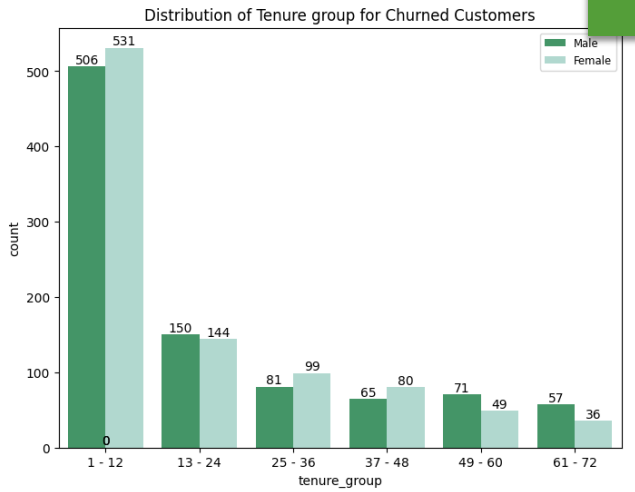
5. Distribution of Senior Citizen for Churners and Non-Churners

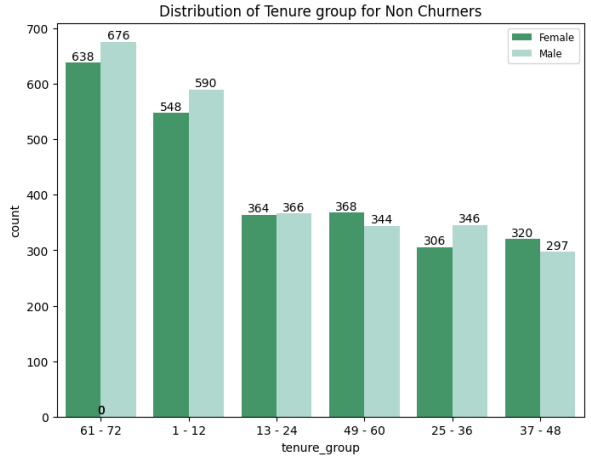
From the visualization above, we can see that the churn and retain rate for senior citizens, both male, and female is almost the same. But if we look, the churn rate has increased in senior citizens when compared to young citizens.



6. Distribution of Tenure group for Churners and Non-Churners

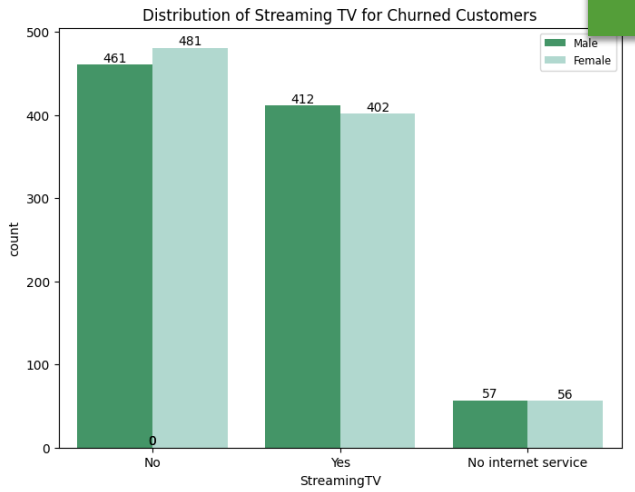
From the visualization above, we can see that the churn ratio for female customers within 1 – 12 tenure group if more when compared to the male in the same group. The female customers after the tenure group 49 – 60 are less churners when compared to the male customers of the same group.

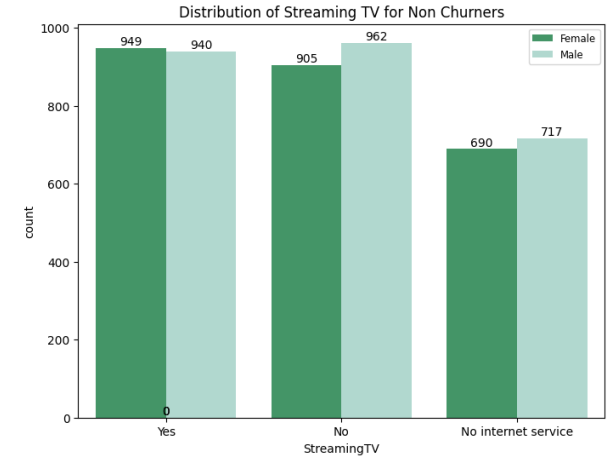




7. Distribution of Streaming TV for Churners and Non-Churners

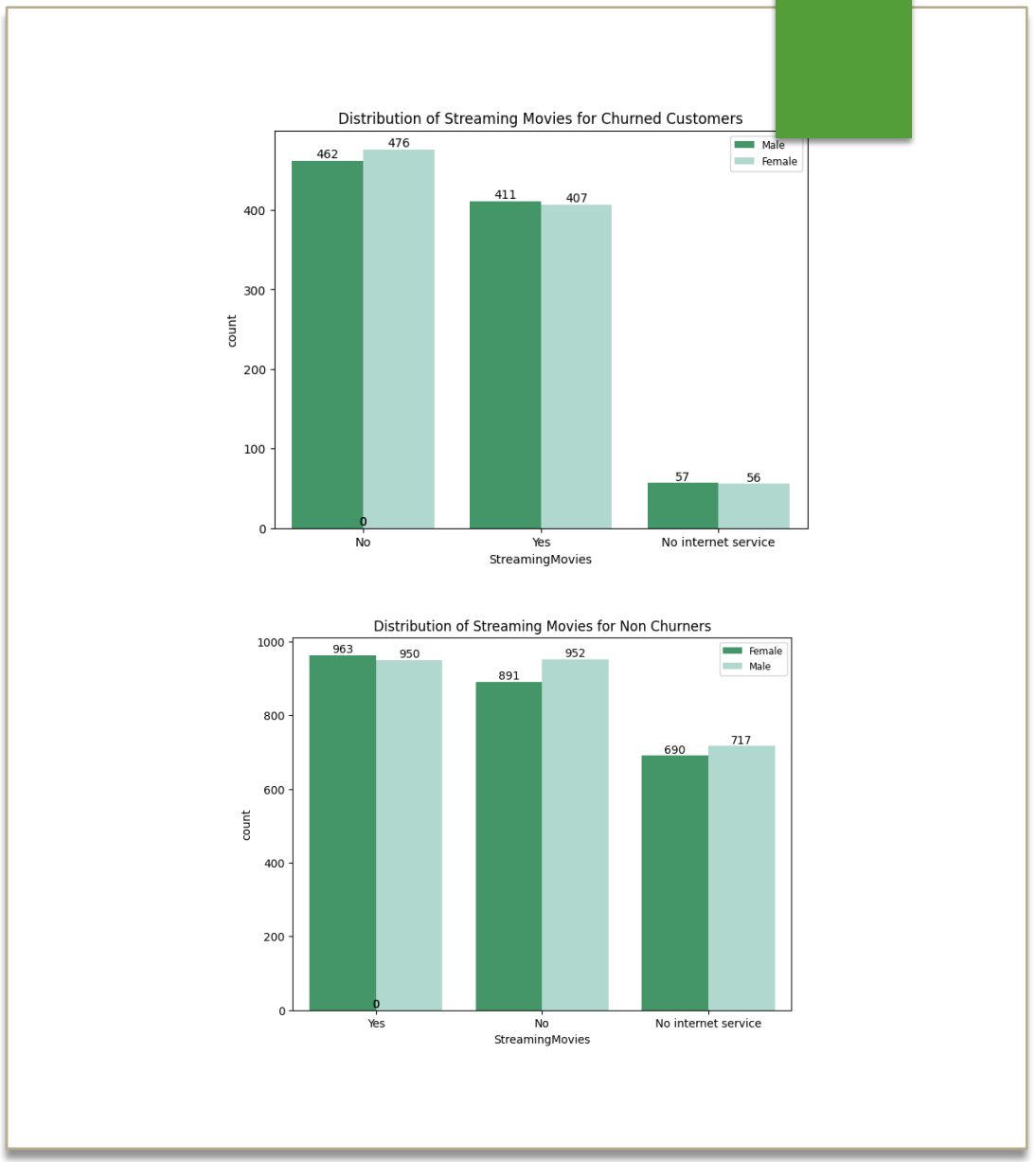
From the visualization above, we can see that the churn rate for male and female customers with Streaming TV are almost the same, but the males with no streaming TV are slightly less churning compared to the females.





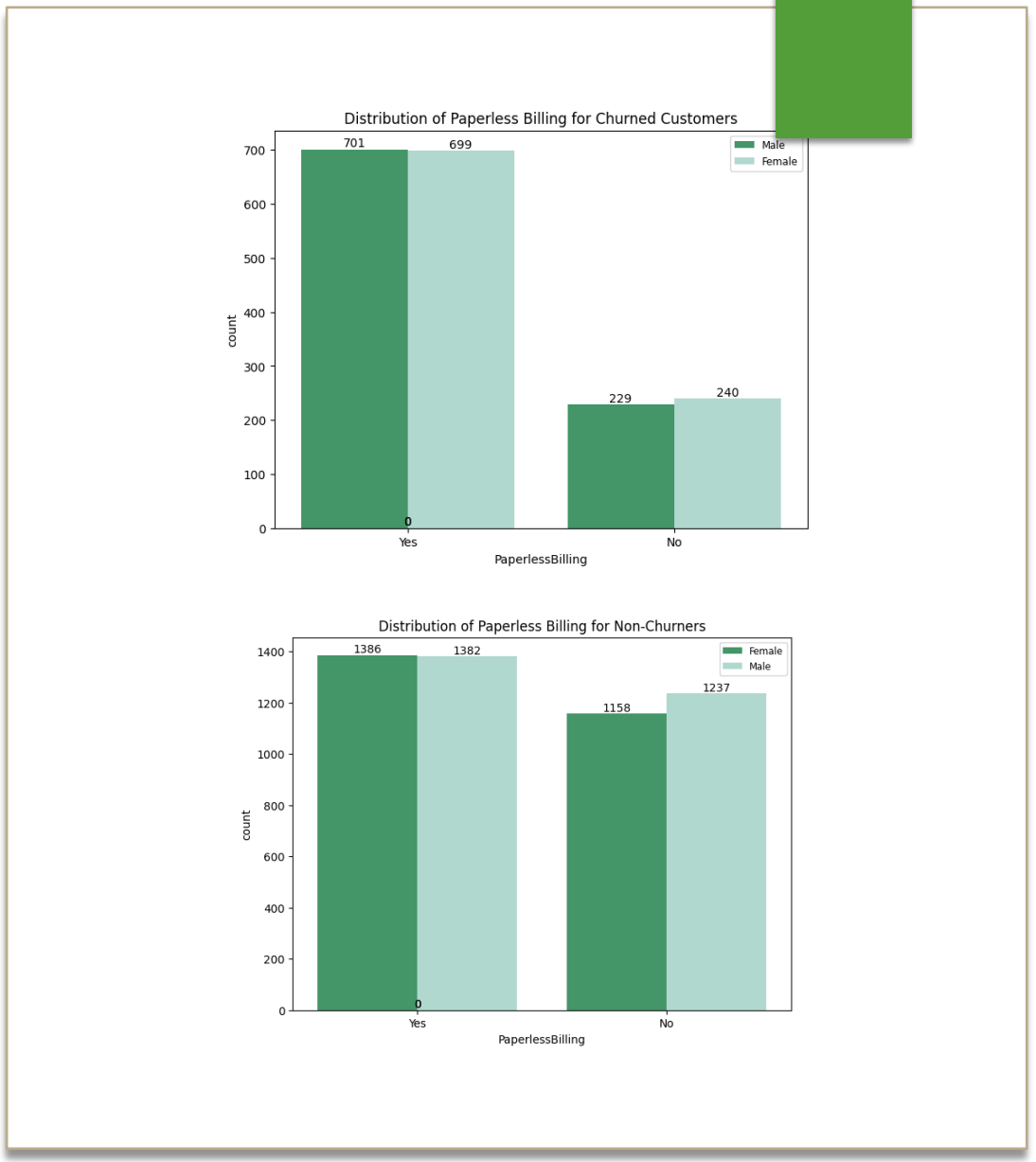
8. Distribution of Streaming Movies for Churners and Non-Churners

From the visualization above, we can see that the churn rate for male and female customers with Streaming Movies are almost the same, but the churn rate for males with no streaming Movies are less churning compared to the female with no streaming movies.



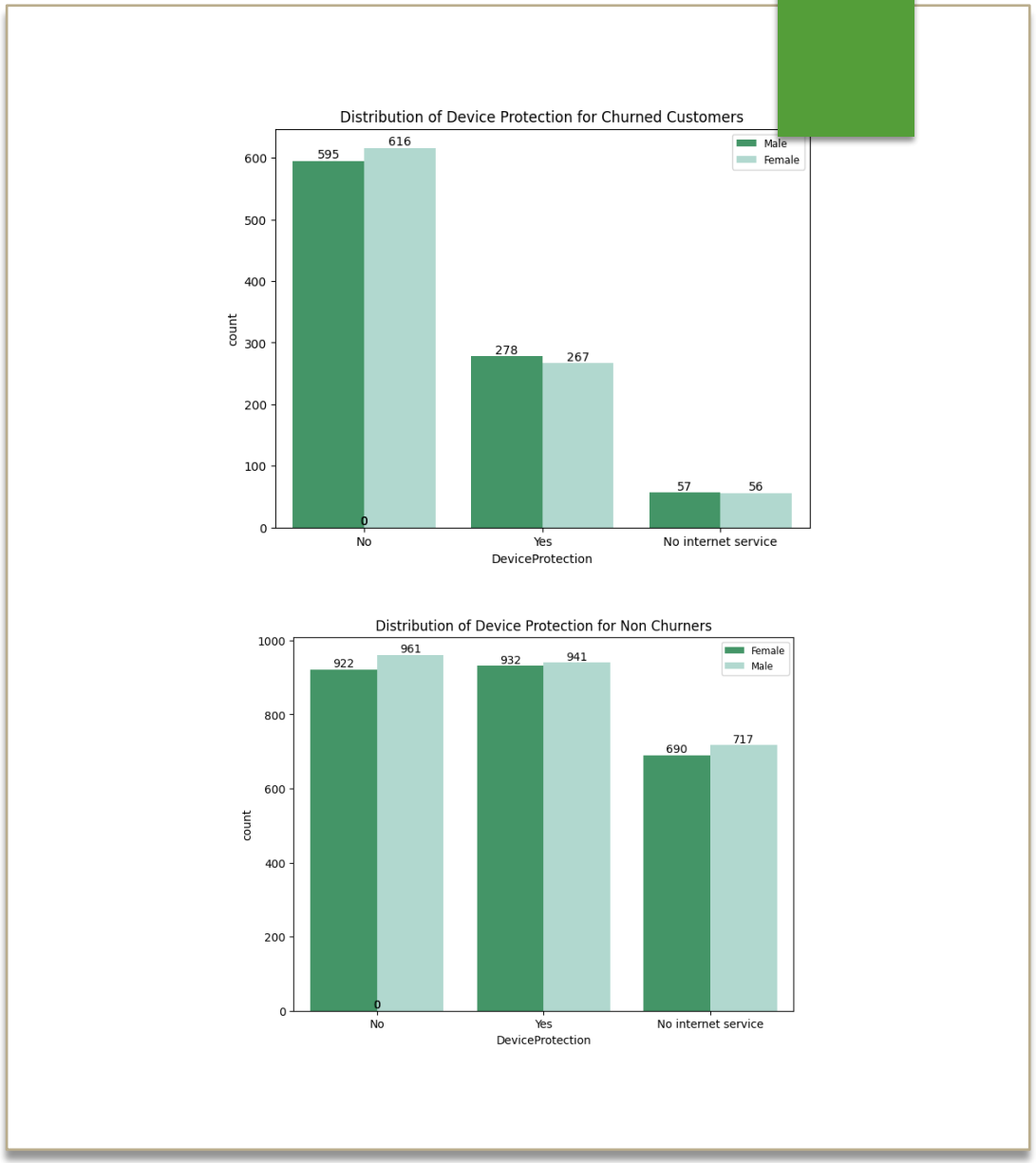
9. Distribution of Paperless Billing for Churned and Non-Churners

From the visualization above, we can see that the churn rate for male and female customers with paperless billings and paper billings are almost the same. A slight increase in the churn rate with male with paper billing can be noticed from this visualization.



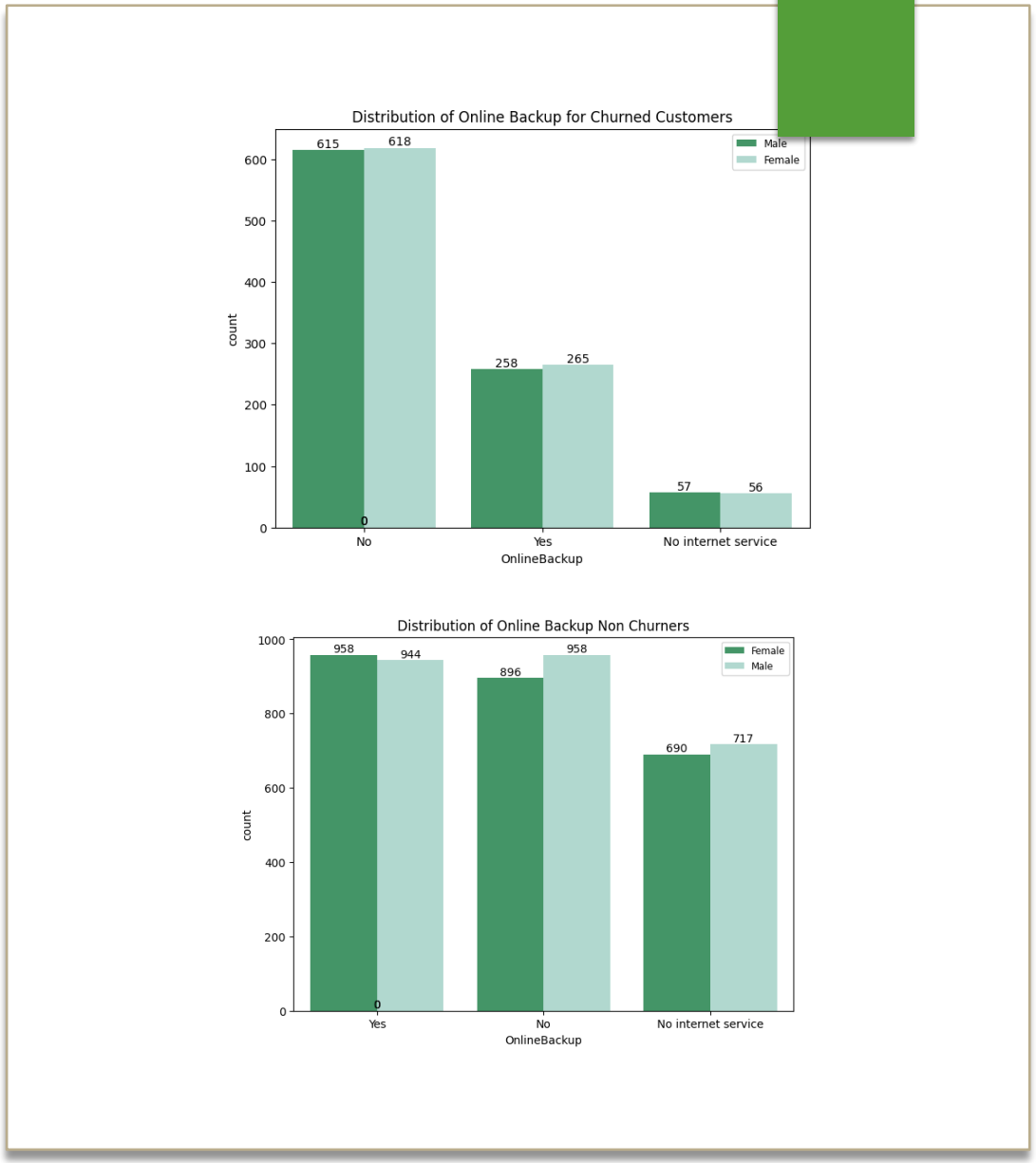
10. Distribution of Device Protection for Churned and Non-Churners

From the visualization above, we can see that the churn rate for female customers with no device protection is high compared to the churn rate of male customers without device protection



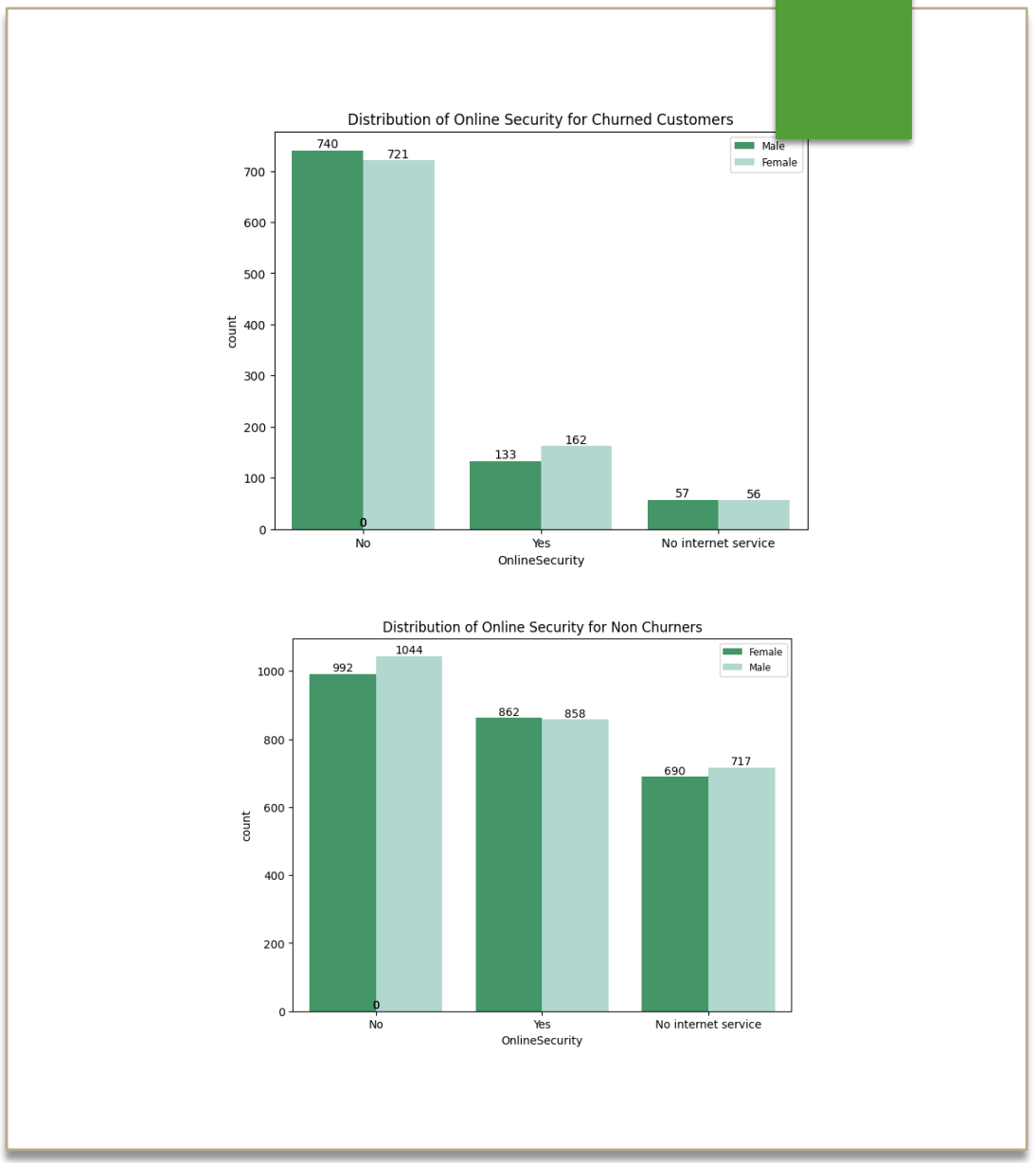
11. Distribution of Online Backup for Churned and Non-Churners

From the visualization above, we can see that the churn rate for female and male customers with no online backup is almost the same.



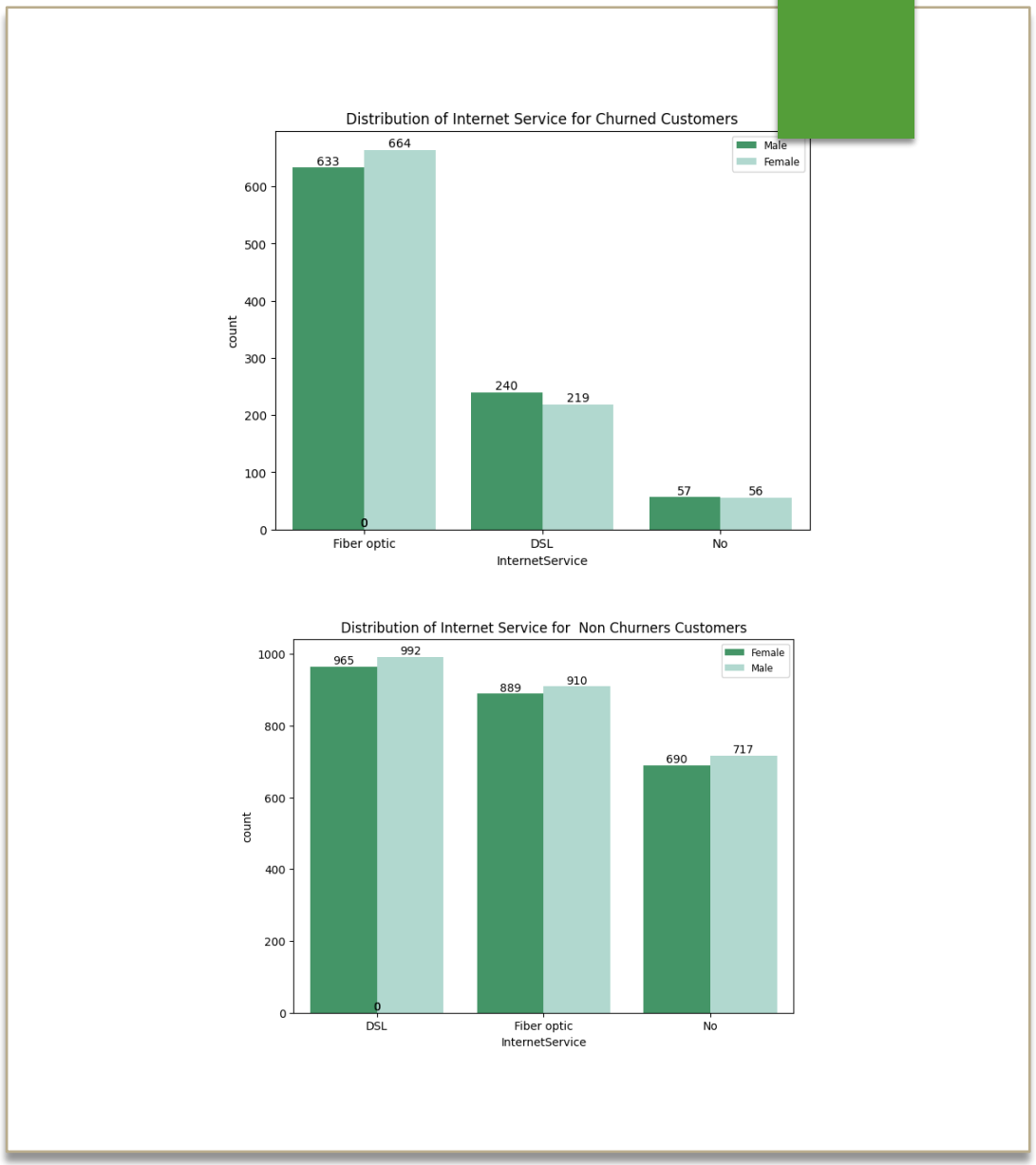
12. Distribution of Online Security for Churners and Non-Churners

From the visualization above, we can see that the churn rate for female customers with online security is high compared to the male customers.



13. Distribution of Internet Service for Churned and Non-Churners

From the visualization above, we can see that the churn ratio for female customers with fiber optic internet service is high compared to the male customers. Also, the retain ratio for female customers with DSL internet service is less compared to the male customers.



FINAL THOUGHTS

1. The churn rate of female and male customers is almost the same. So, from this analysis no insights can be concluded. The churn rate of Female is 26.95% and Male is 26.20%
2. The churn rate of senior citizens is almost double that of young citizens. The churn rate of seniors is almost double that of young customers. The rates for Young Citizens are 23.65% and for Senior Citizen is 41.68%
3. Customers with a partner churn less than customers with no partner. Customers without partners are churning more in comparison to customers with partners. The churn rate of customers with Partner is 19.7% and without Partner is 32.97%
4. Dependent customers are less churners when compared to the non-dependents. The churn rates for Non-Dependent customers are 31.28% and the churn rate of Dependent is 15.53%
5. The customers with Fiber optic internet service are having a high churning rate. The churn rate for customers with DSL internet service is 18.99% and the churn rate for customers with Fiber optics internet service is 41.89%

7. The churn rate for customers without Online Backup is high. The churn rate of customers without Online Backup is 39.94%
8. The churn rate for customers Without Device Protection is high compared with customers with Device protection. The churn rate of customers without Device Protection is 39.94%
9. Customers with no tech support are churning more. The churn rate for customers without Tech Support is 41.65%
10. Customers with month-to-month contracts have higher churn rates compared to clients with yearly contracts. The churn rate for customers with Month - Month contracts is 42.71%.
11. Customers opted for paperless billing churn more than the counterpart. The churn rate for customers opted for Paperless Billing is 33.59 %
12. Customers who opted for an electronic check as a payment method are more likely to churn. The churn rate for customers with Electronic Check is 45.29%

13. The least churners are there in the tenure group of 61 – 72 Months and the top churners are there in the tenure group of 1 - 12 Months. The churn rate of the tenure group of 1 - 12 is 47.68% and the churn rate of the tenure group 61 – 72 is 6.60% .
14. New customers (low tenure) are more likely to churn.
15. The churn rate tends to be larger when monthly charges are high.
16. Customers with high total charges are less likely to leave the company.
17. The churn ratio for female customers with fiber optic internet service is high compared to the male customers.
18. Customers with Partners and Dependents have lower churn rate as compared to those who don't have partners & Dependents.
19. Customers with no internet service have a lower churn rate.
20. Churn rate is much higher in the case of Fiber Optic InternetServices. Customers who do not have services like OnlineSecurity, OnlineBackup, and TechSupport have left the platform in the past month.
21. Contract duration (strong): Month-to-month contracts churn significantly more than one- or two-year contracts.

22. Internet services (moderate): Internet Fiber or DSL services churned at a higher pace than other services.]
23. Dependents (Moderate): Customers who support dependents churned less.
24. (Low) Paperless billing: Customers who opted for paperless billing had a higher attrition rate.
25. (Low) Streams: Indicates whether the consumer is utilizing the Internet service to stream TV or movies.
26. (Low) Automatic payment: Customers who had automatic payments set up were less likely to churn.
27. (Low) Partner: Customers who were married had a lower churn rate.
28. High Churn is seen in the case of Senior Citizens, Customers without partners, Month-to-month contracts, No online security, No Tech support, First year of subscription, and Fibre Optics Internet
29. Low Churn is seen in the case of long-term contracts, Subscriptions without internet service, and the customers engaged for more years
30. Factors like Gender, Availability of PhoneService, and number of multiple lines have almost no impact on Churn if considered alone
31. Female with partners are non-churners. The retain rate for customers with partner and no partner is almost the same. But if we look, the churn rate has increased in females with no partners when compared to females with partners.

THANK YOU

SUREKHA BERLIN