CONNECTTEL CUSTOMER CHURN PREDICTION AND IMPROVING RETENTION

USING SUPERVISED MACHINE LEARNING TECHNIQUES

PRESENTED

BY

UDOCHI OGBONNA

INTRODUCTION

- ConnectTel is a leading telecommunications company at the forefront of innovation and connectivity solutions.
- With a strong presence in the global market, ConnectTel has established itself as a trusted provider of reliable voice, data, and Internet services. Offering a comprehensive range of telecommunications solutions, including mobile networks, broadband connections, and enterprise solutions, ConnectTel caters to both individual and corporate customers, they are committed to providing exceptional customer service and cutting-edge technology.
- ConnectTel ensures seamless communication experiences for millions of users worldwide. Through strategic partnerships and a customer-centric approach, ConnectTel continues to revolutionize the telecom industry, empowering individuals and businesses to stay connected and thrive in the digital age.

PROBLEM OVERVIEW

- **ConnectTel Telecom Company** faces the pressing need to address customer churn, which poses a significant threat to its business sustainability and growth.

- The company's current customer retention strategies lack precision and effectiveness, resulting in the loss of valuable customers to competitors.

- To overcome this challenge, **ConnectTel** aims to develop a robust customer churn prediction system by leveraging advanced analytics and machine learning techniques on available customer data, the company seeks to accurately forecast customer churn and implement targeted retention initiatives.

- This proactive approach will enable **ConnectTel** to reduce customer attrition, enhance customer loyalty, and maintain a competitive edge in the highly dynamic and competitive telecommunications industry.

APPROACH

- Data Collection & Pre-processing
- Exploratory Data Analysis
- Feature Selection & Engineering .
- Model Development.
- Hyper Parameter Optimization.
- Evaluation of the Model, Comparison & Selection
- Potential Benefits/Recommendation.

Data Collection & Pre-processing

..

- The data set contains a total of 21 columns, 7,043 entries with the Total Charges in the object format.

- The numerical features indicates tenure has mean value as 32.4, minimum as 0 and maximum as 72. Monthly Charges has mean 64.8, minimum as 18.3 and maximum as 118.8. Total charges has mean of 2280, minimum as 18.8, maximum as 8684.8 respectively. The Total Charges have 7,032 entries instead of 7,043. This shows that there are 11 missing values.

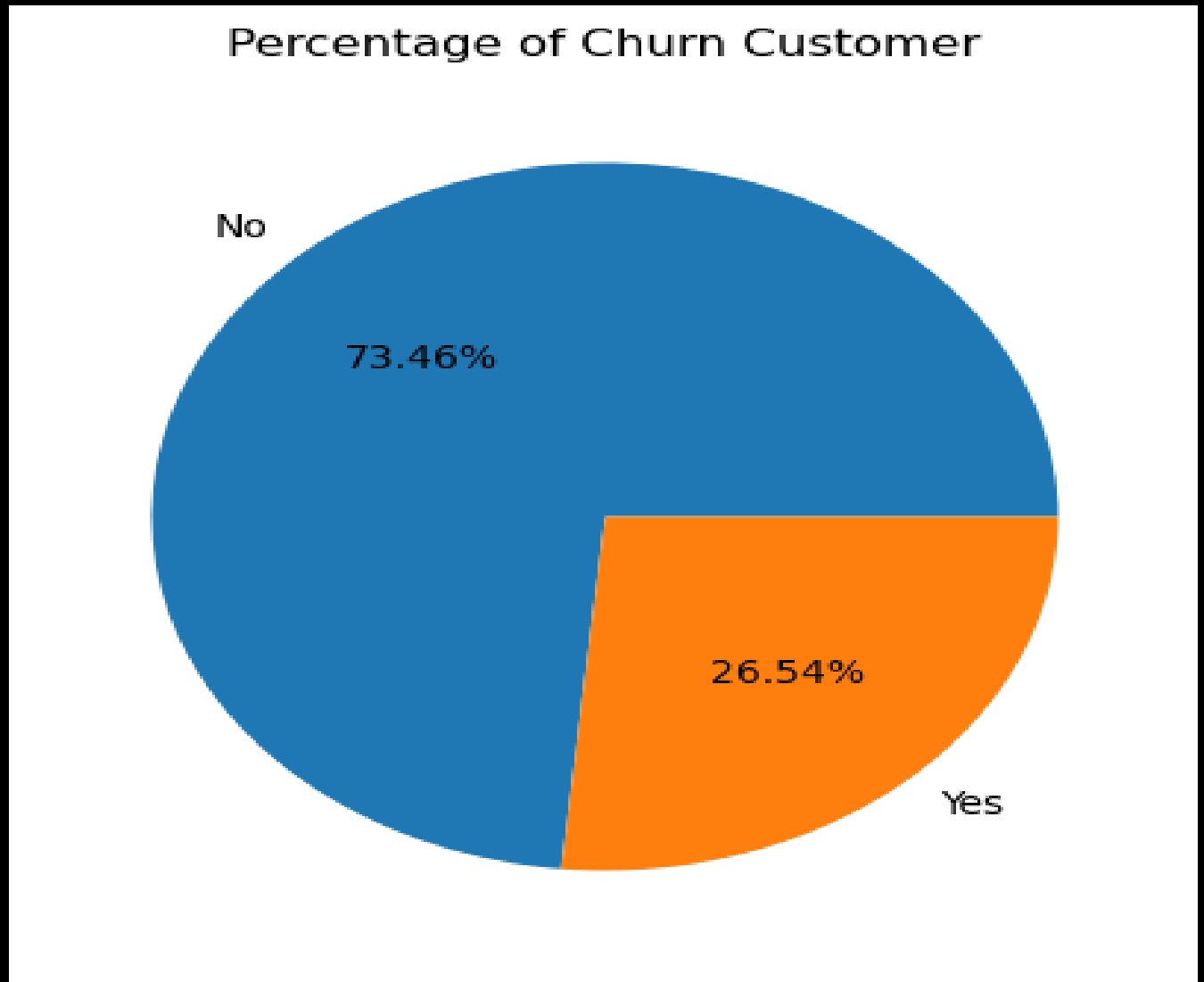
- The categorical features showed that the dataset has more males, fibre optics topping on the Internet service, month to month as the preferred contract term and the electronic check as the highest payment method. There are no duplicated values.

Exploratory Data Analysis:

Univariate Analysis

CUSTOMERS WHO CHURN

With a churn rate of 27%, the dataset exhibits a substantial class imbalance, where the majority of customers (73%) remain, while the minority (27%) churn, demanding careful handling to achieve accurate predictions.

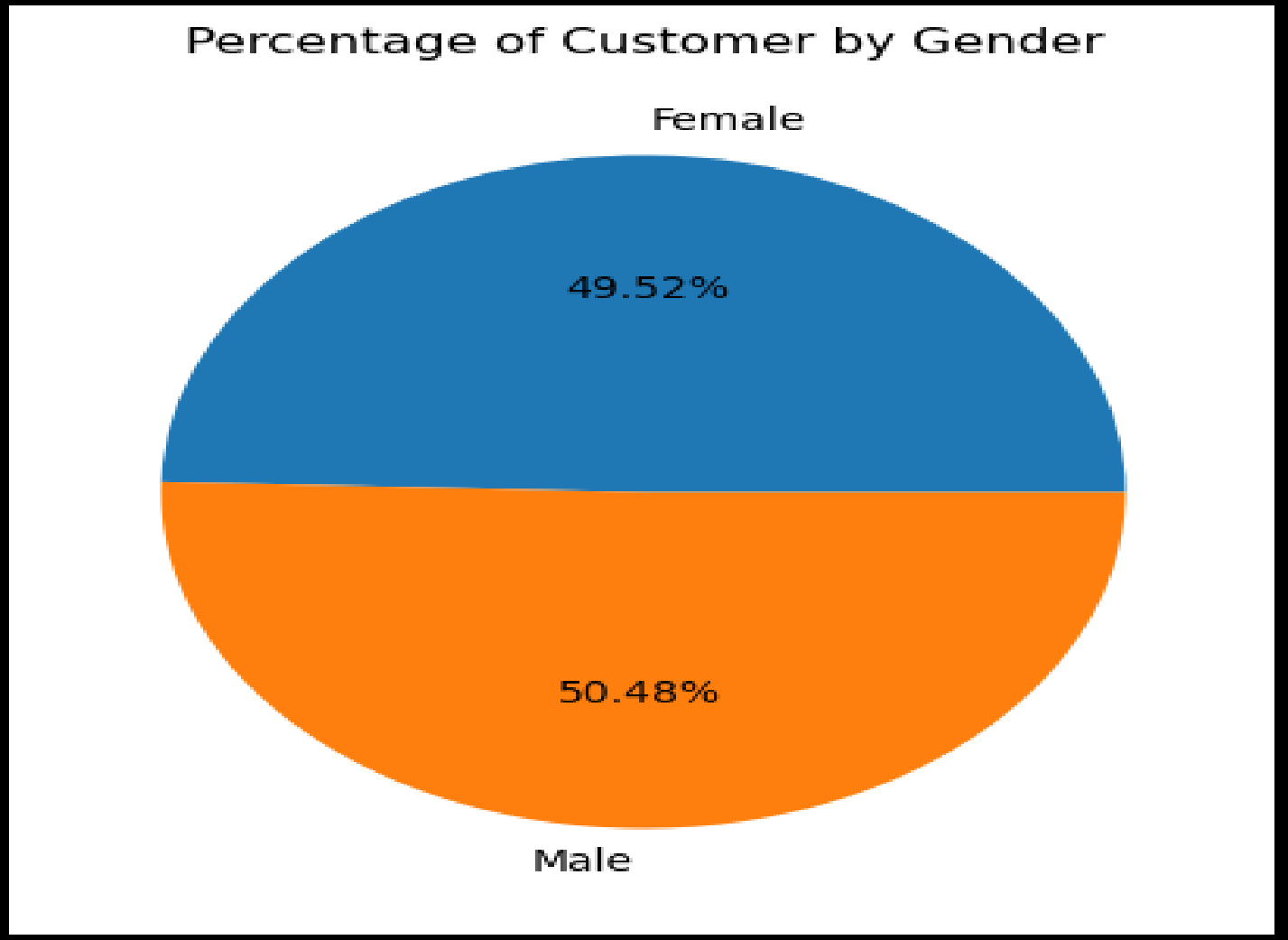


Exploratory Data Analysis:

Univariate Analysis

GENDER

The customer gender distribution is approximately equal.

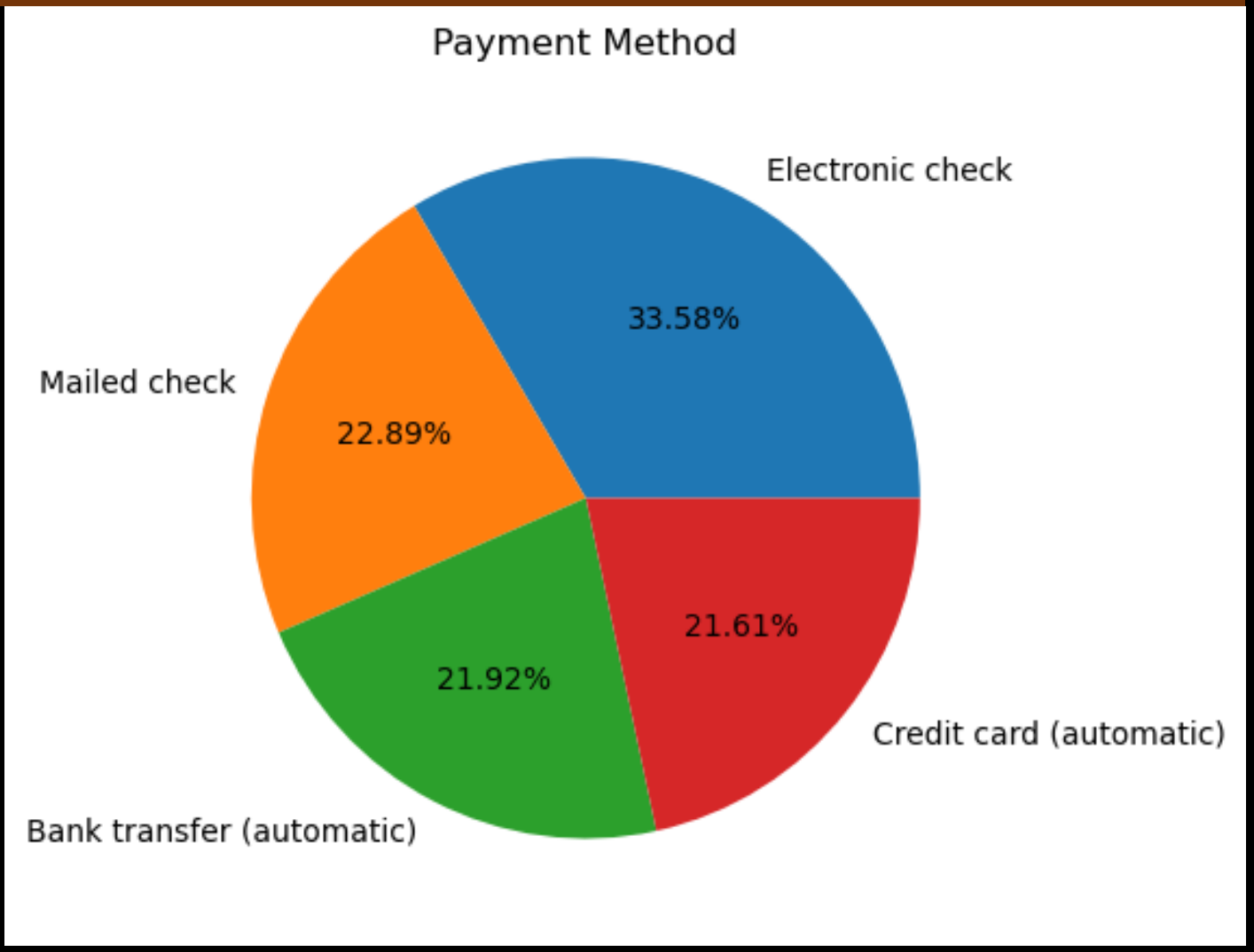


Exploratory Data Analysis:

Univariate Analysis

PAYMENT METHOD

The most preferred payment method is Electronic Check (34%), followed closely by Mail Check (23%), and then equally by Bank Transfer and Credit Card (both at 22%) respectively.

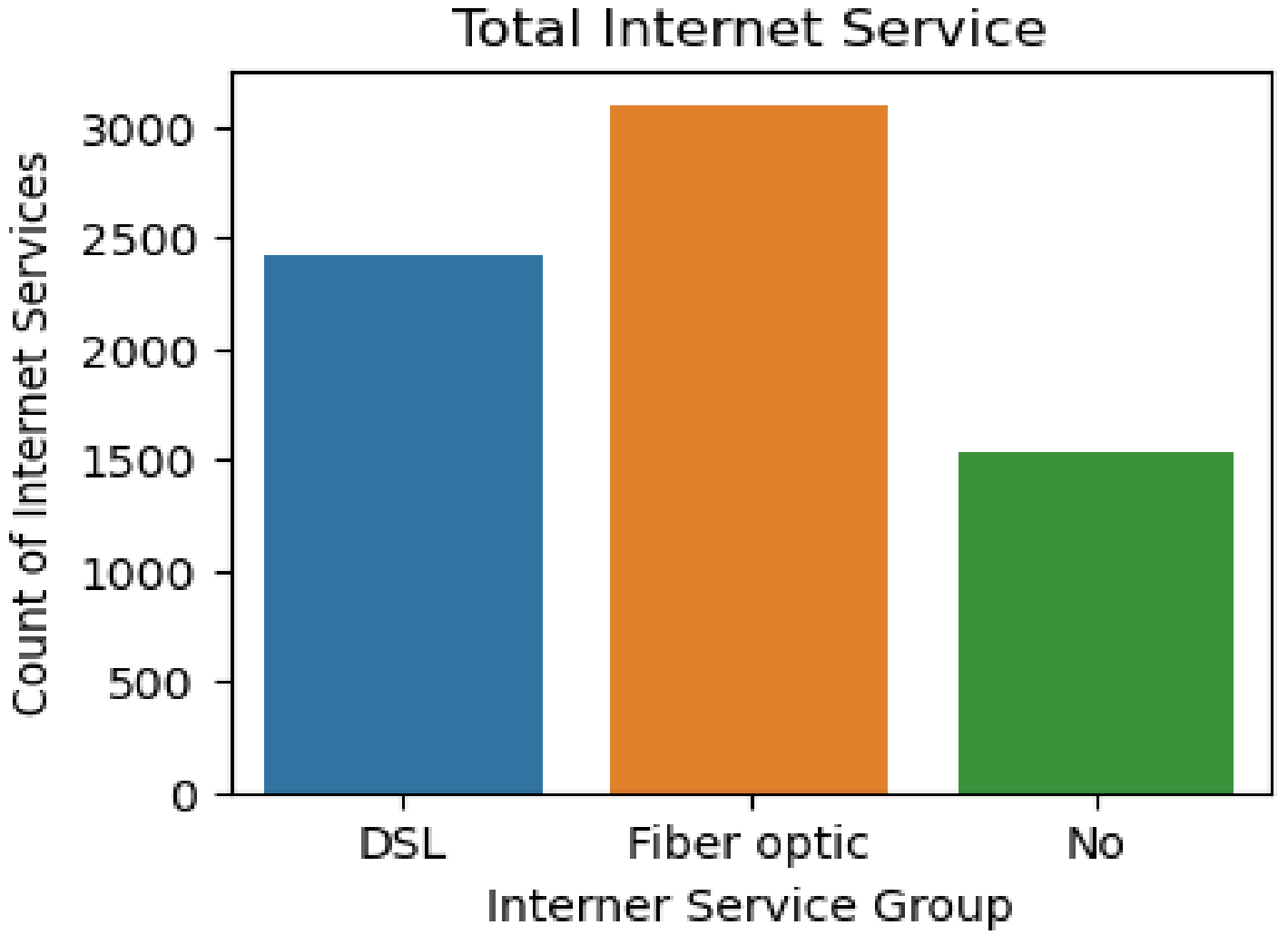


Exploratory Data Analysis:

Univariate Analysis

INTERNET SERVICE

Most customers use Fiber optic as their Internet service followed by DSL

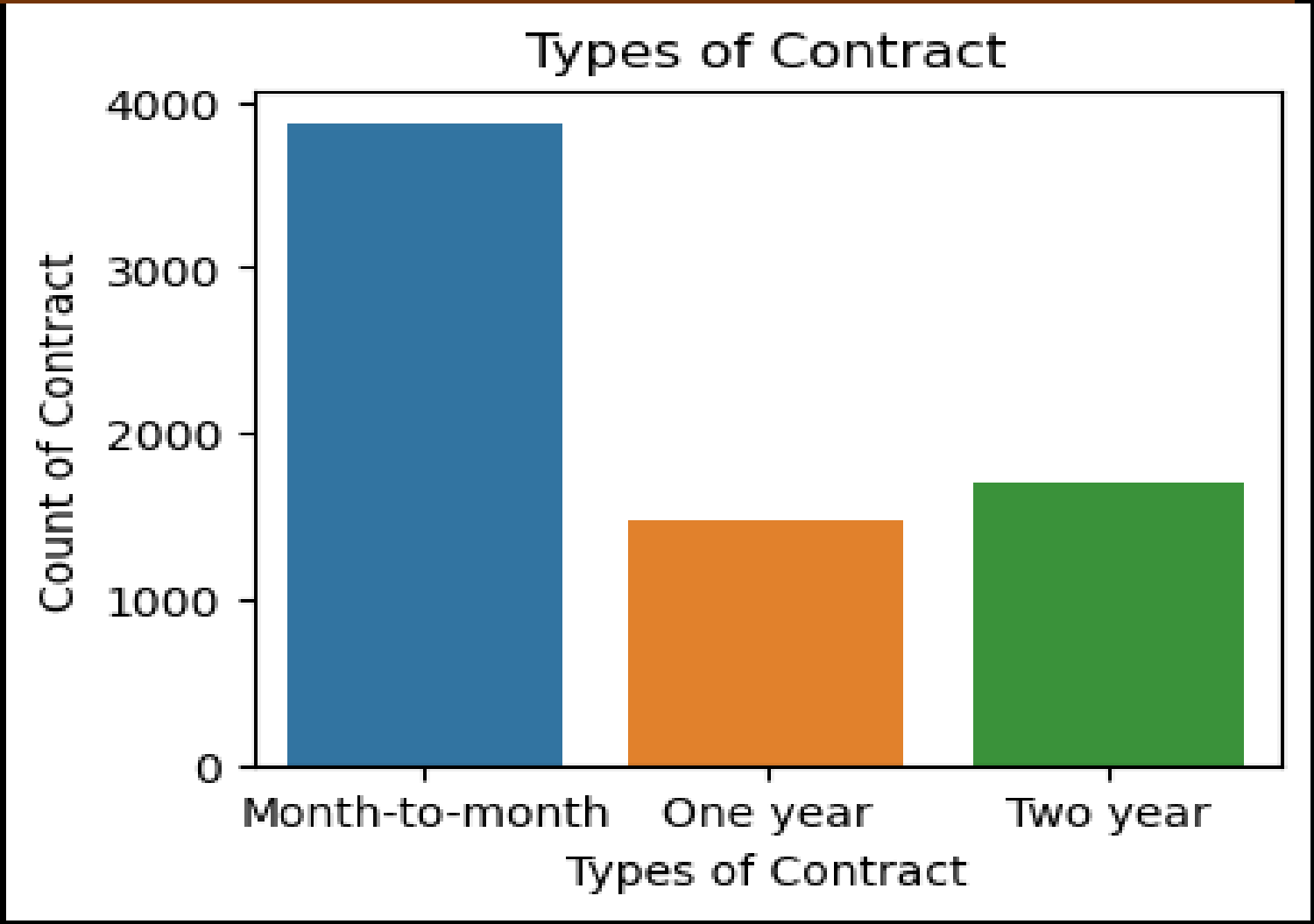


Exploratory Data Analysis:

Univariate Analysis

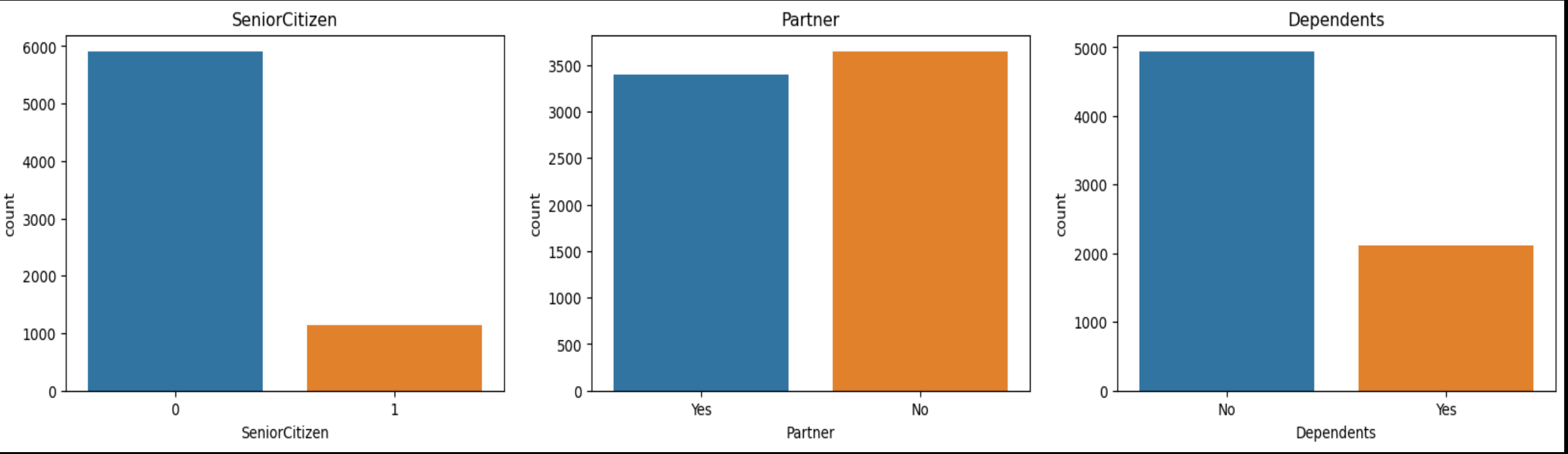
TYPES OF CONTRACT

The dataset reveals that the majority of customers are on a month-to-month subscription plan.



Exploratory Data Analysis:

Univariate Analysis



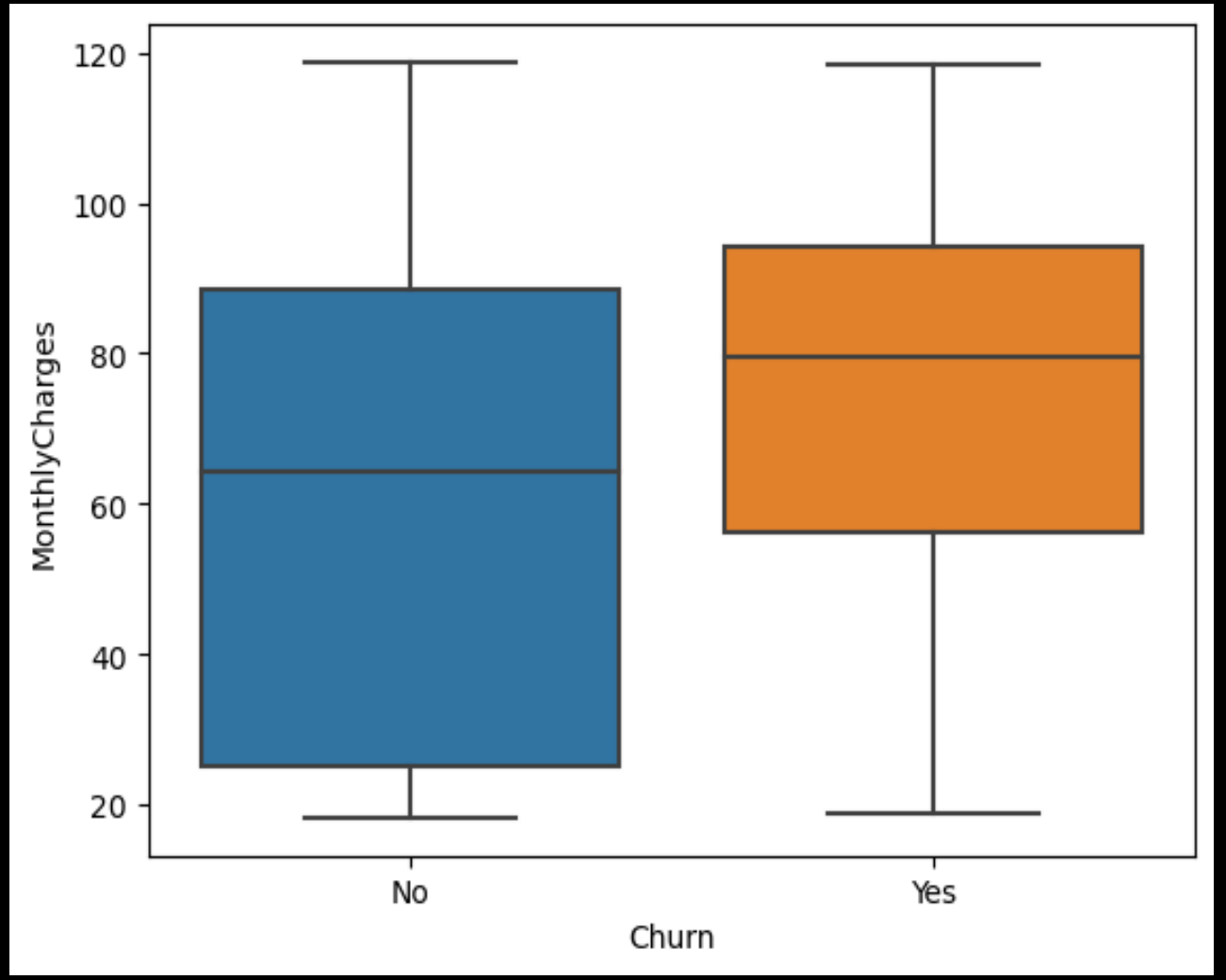
- The dataset reveals that most customers are senior citizens, with an equal distribution of marital status and no dependents

Exploratory Data Analysis:

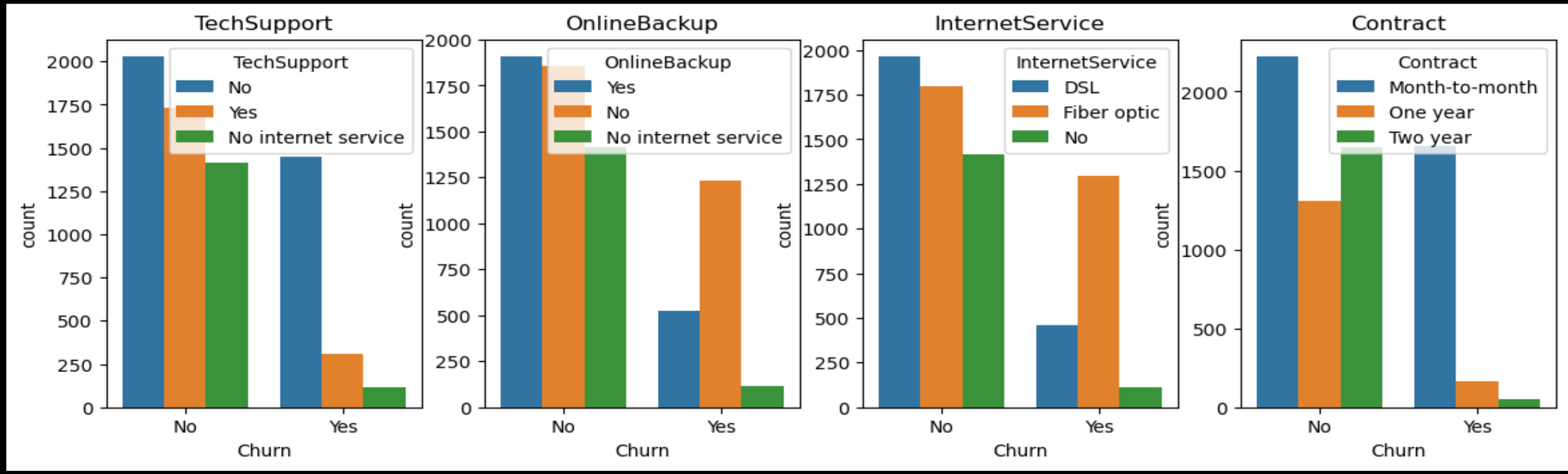
Bivariate Analysis

MONTHLY CHARGES VS CHURN

The data reveals that customers who churn have higher average monthly charges compared to those who renew their subscription. This suggests a strong correlation: as monthly charges increase, the likelihood of customers unsubscribing from the network also increases.



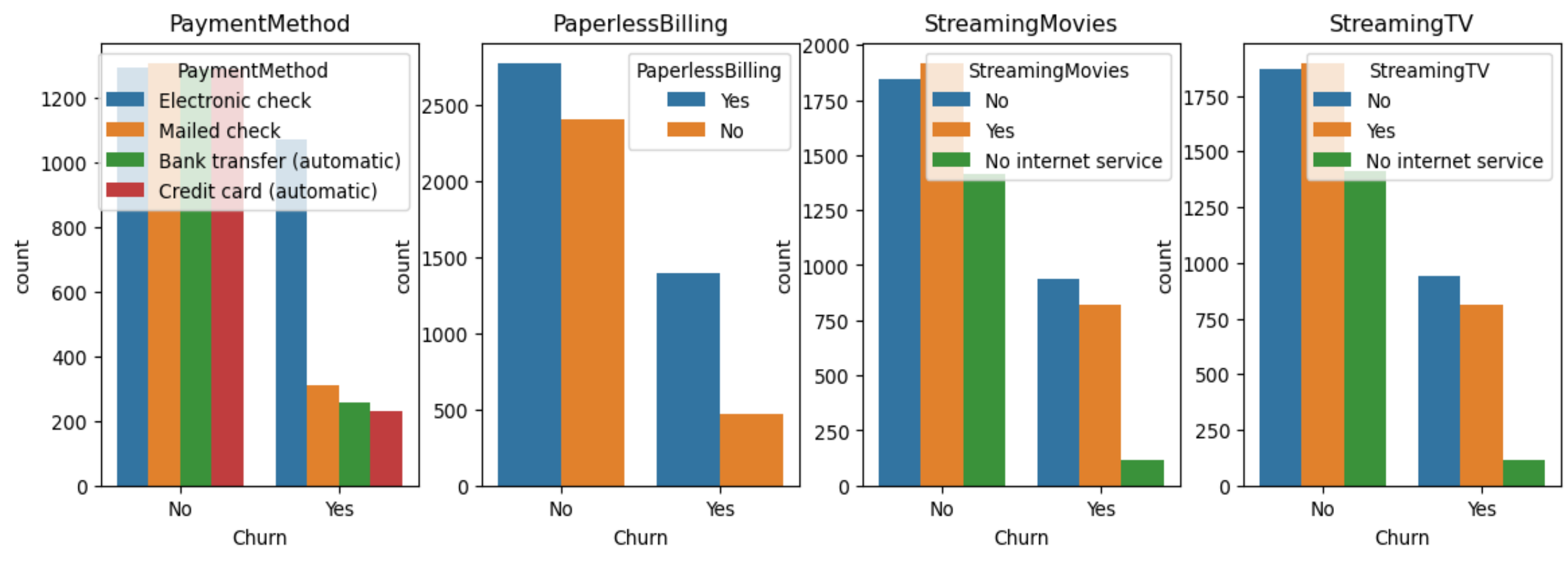
Exploratory Data Analysis: Bivariate Analysis



CHURN VS TECH SUPPORT, ONLINE BACKUP, INTERNET SERVICE, CONTRACT

The majority of customers who churned share certain characteristics: they were on a monthly contract plan, used fibre optic internet service, did not have online backup, and had never received technical support.

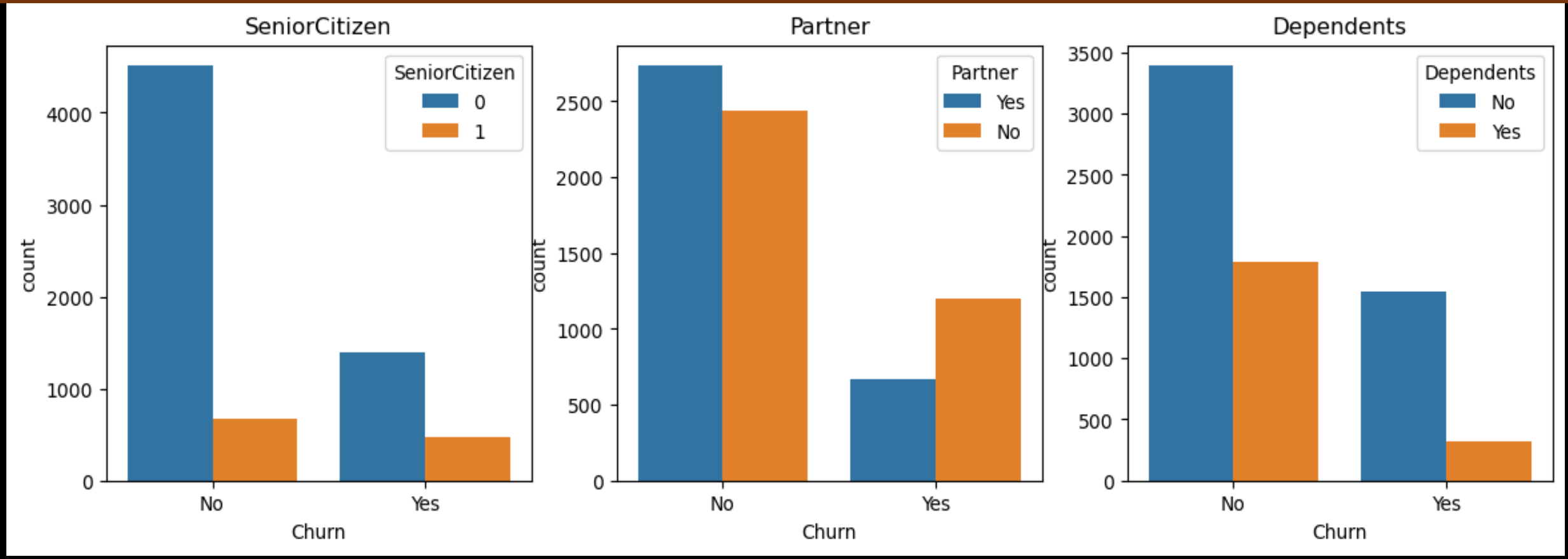
Exploratory Data Analysis: Bivariate Analysis



PAYMENT METHOD, PAPERLESS BILLING, STREAMING MOVIES AND TV

Customers who churned tend to share certain habits: they prefer Electronic Check payments, choose Paperless Billing, and regularly stream movies and TV content.

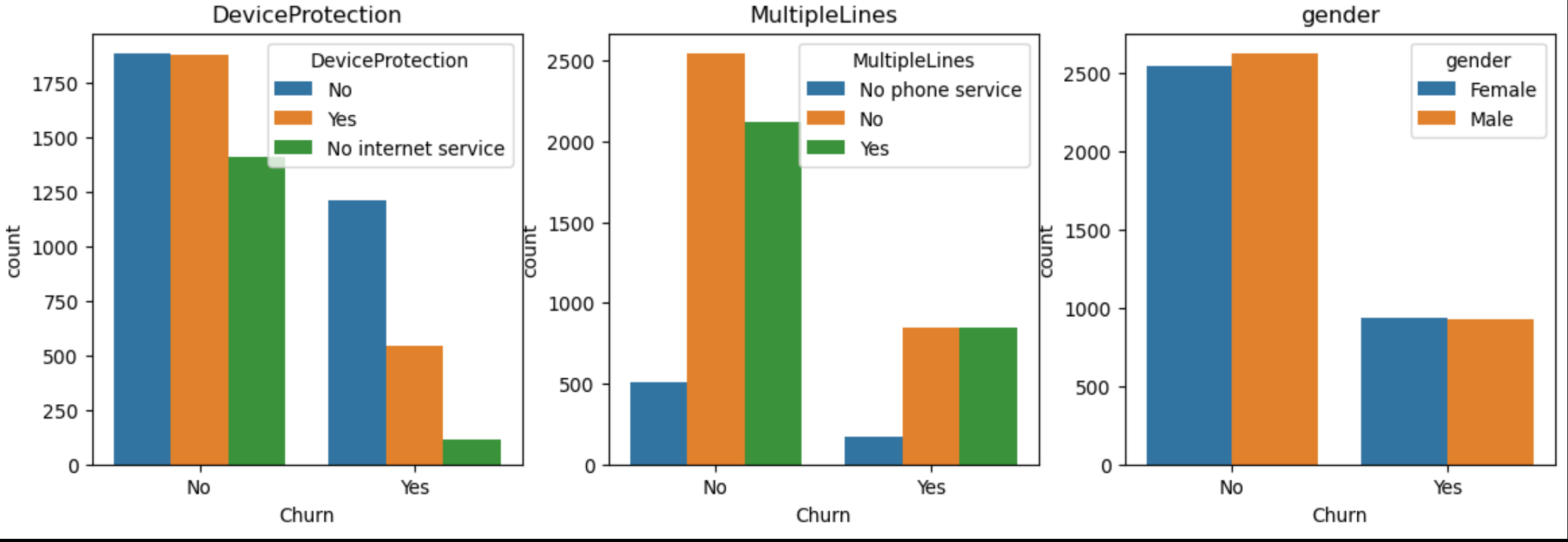
Exploratory Data Analysis: Bivariate Analysis



SENIOR CITIZEN, PARTNER AND DEPENDENTS

Customers with partners and dependents have a high churn rate. Additionally, a significant proportion of Senior citizens also churned.

Exploratory Data Analysis: Bivariate Analysis



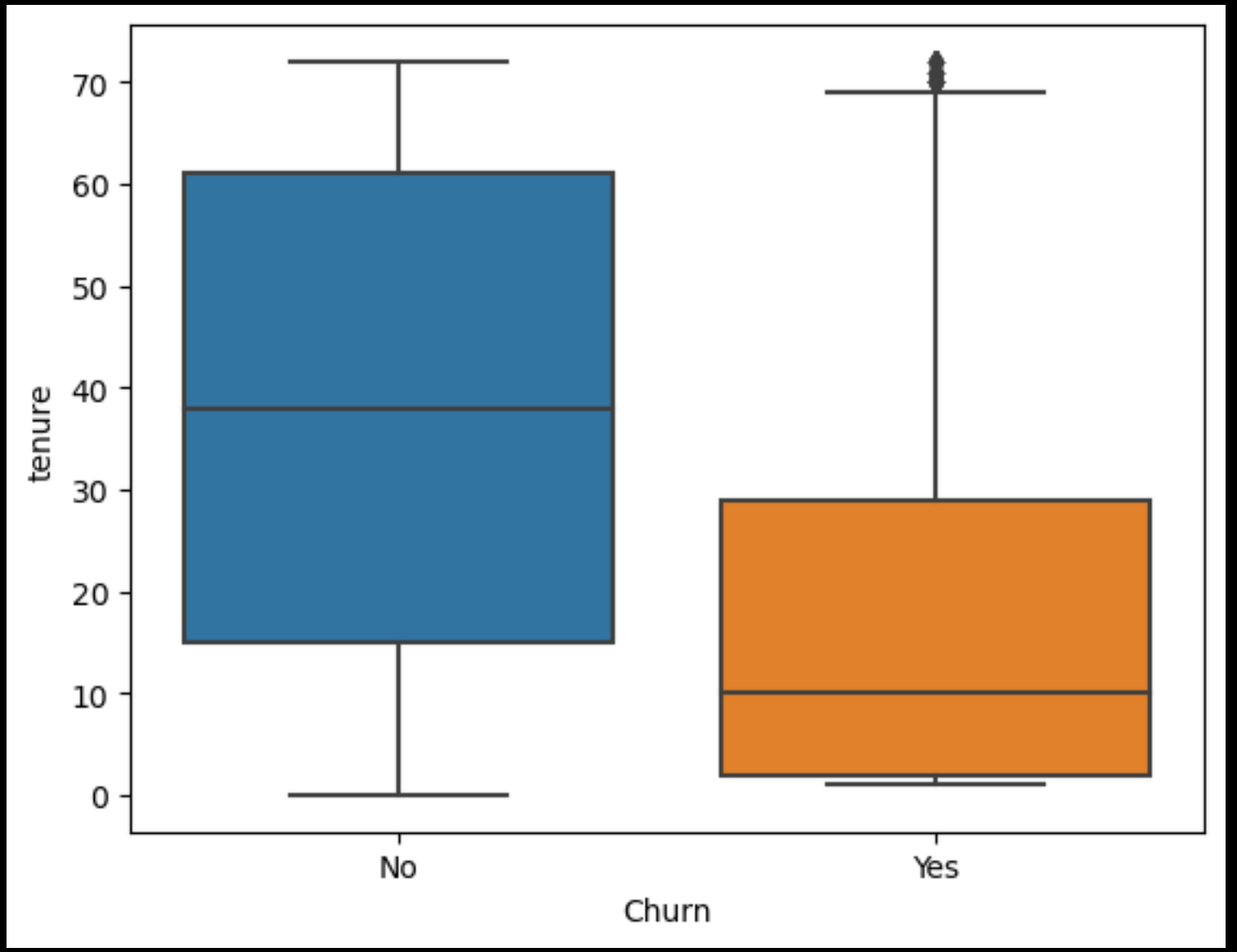
DEVICE PROTECTION, MULTIPLE LINES AND GENDER

Customers without Device Protection are more likely to churn, regardless of the number of lines and churn rate is equal across genders.

Exploratory Data Analysis: Bivariate Analysis

TENURE VS CHURN

Customers with short tenures (0-10) have a higher churn rate, and there are also outliers who churn despite having longer tenures.

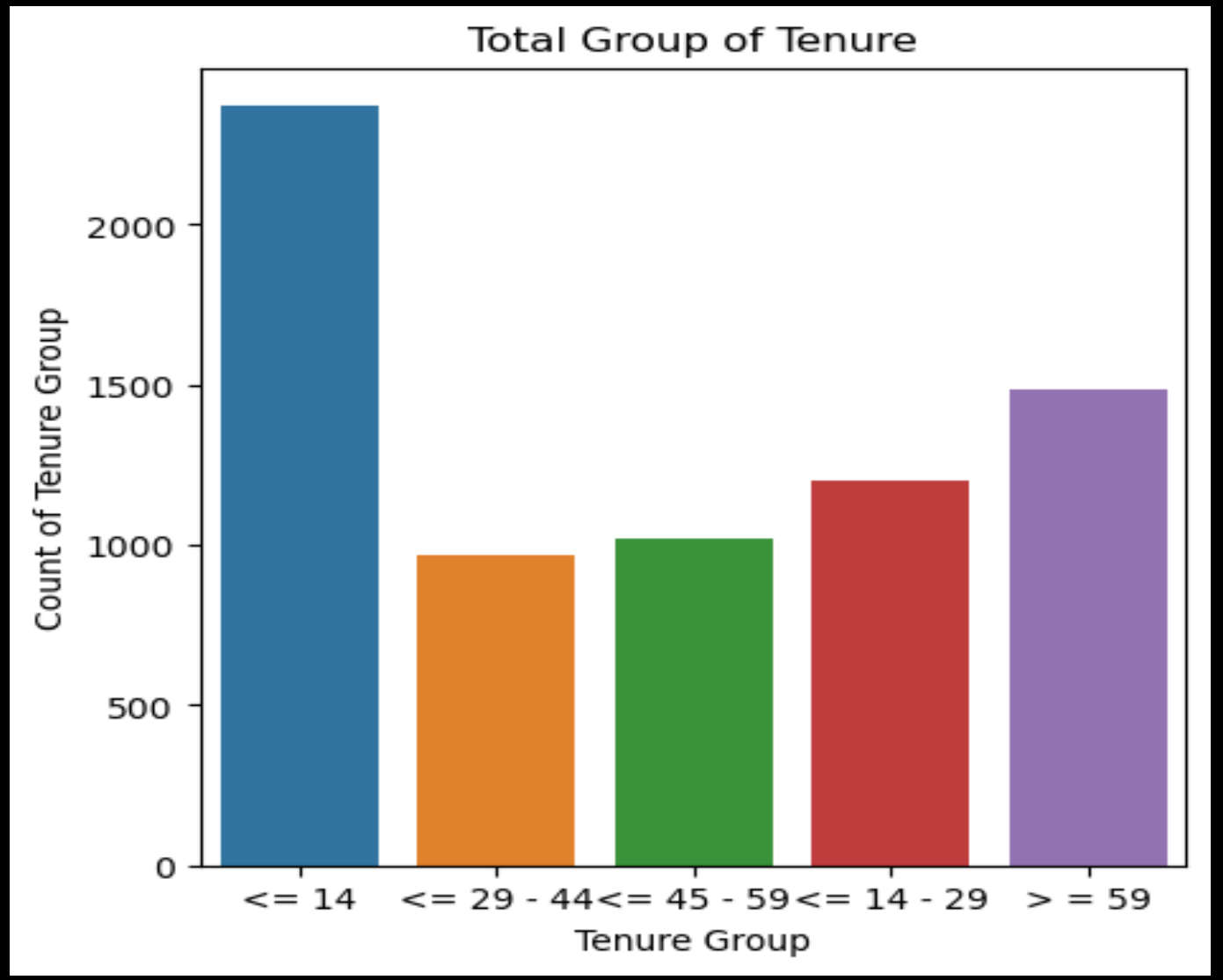


Exploratory Data Analysis:

FEATURE ENGINEERING

GROUPING THE TENURE INTO BUCKETS

Most customers are new (0 - 14) in the Telecom company and a substantial number who have stayed with the company.

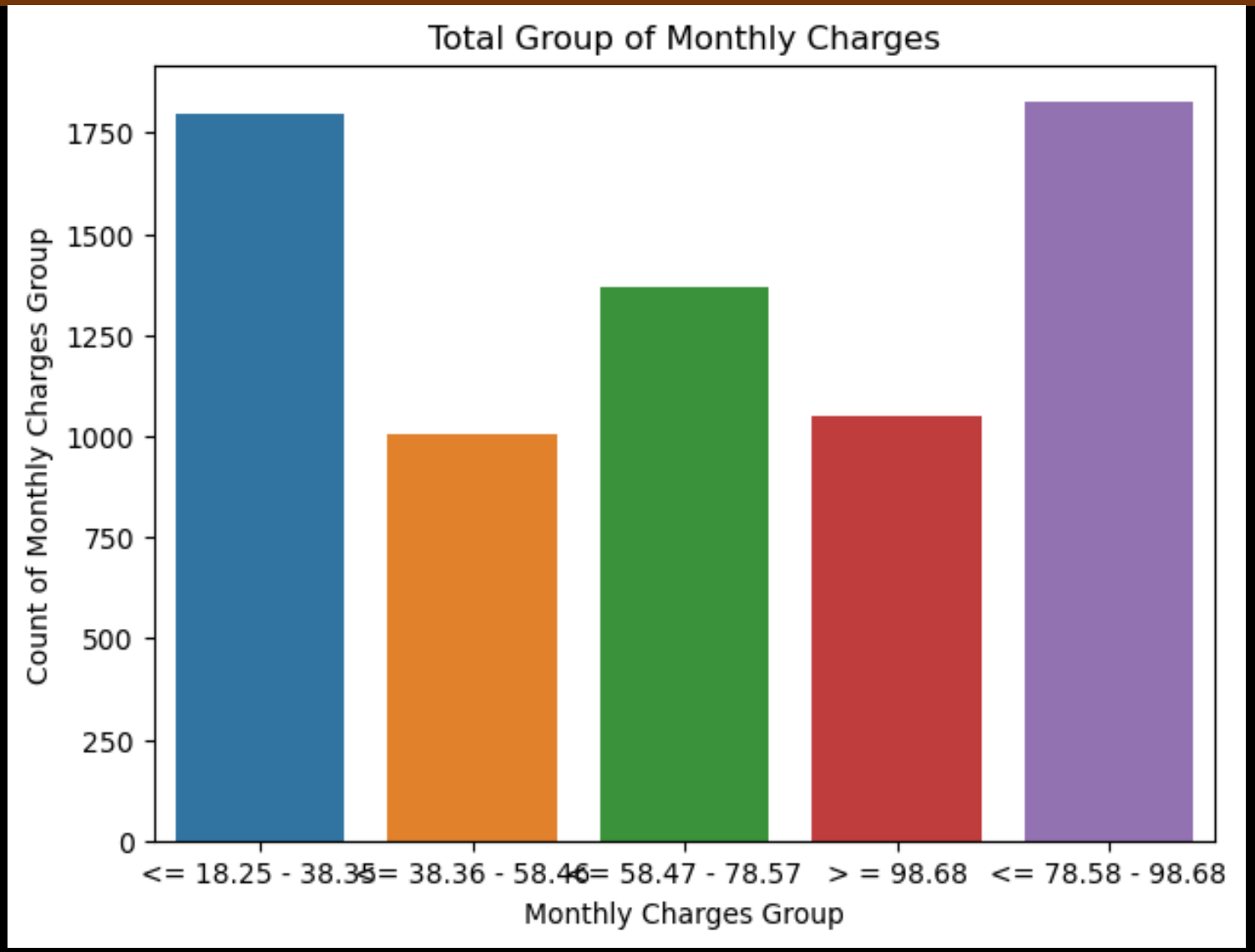


Exploratory Data Analysis:

FEATURE ENGINEERING

GROUPING THE MONTHLY CHARGES INTO BUCKETS

The dataset reveals that most customers that are joining (0 - 38) and those with longer tenure (78 - 98) are on monthly charges.

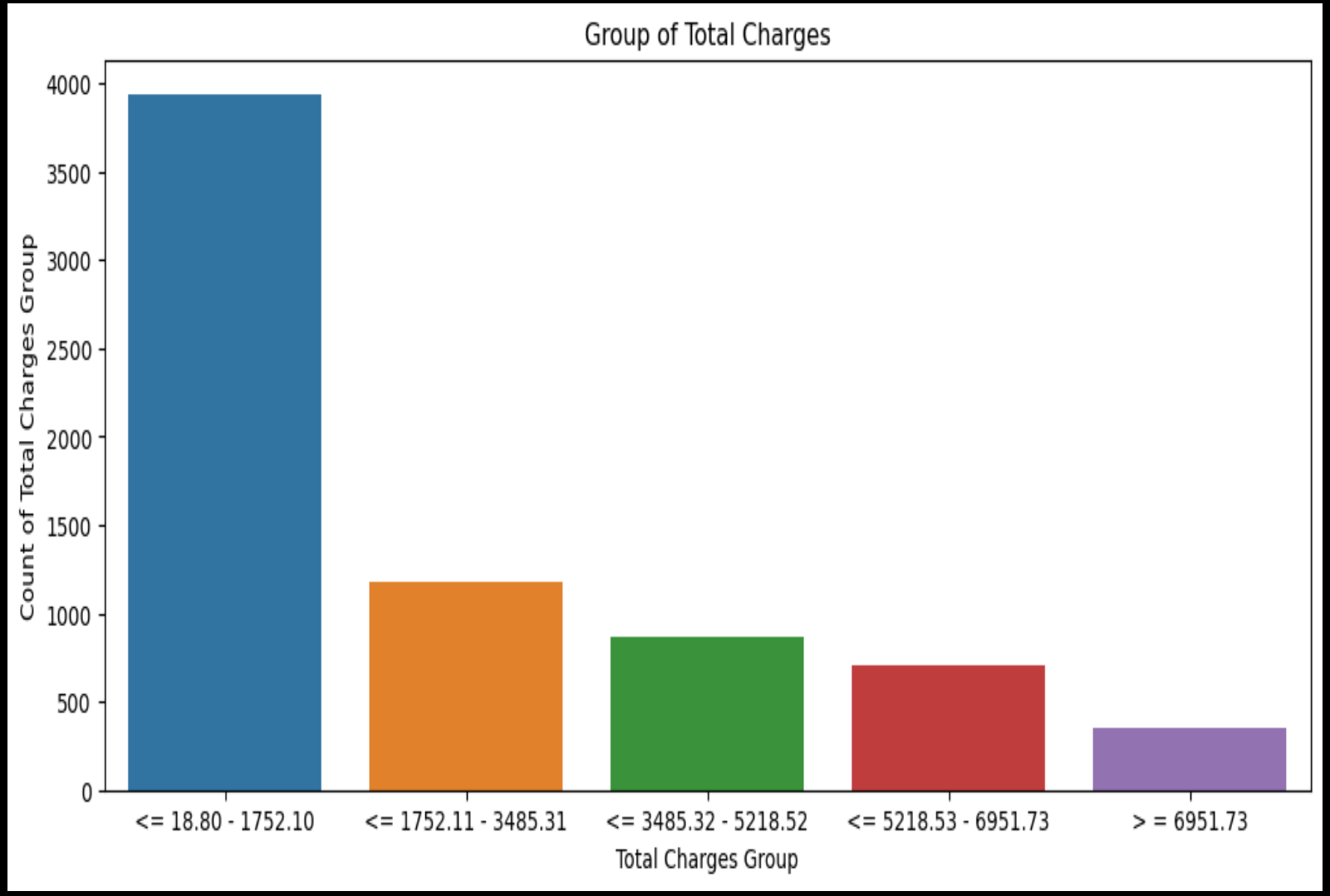


Exploratory Data Analysis:

FEATURE ENGINEERING

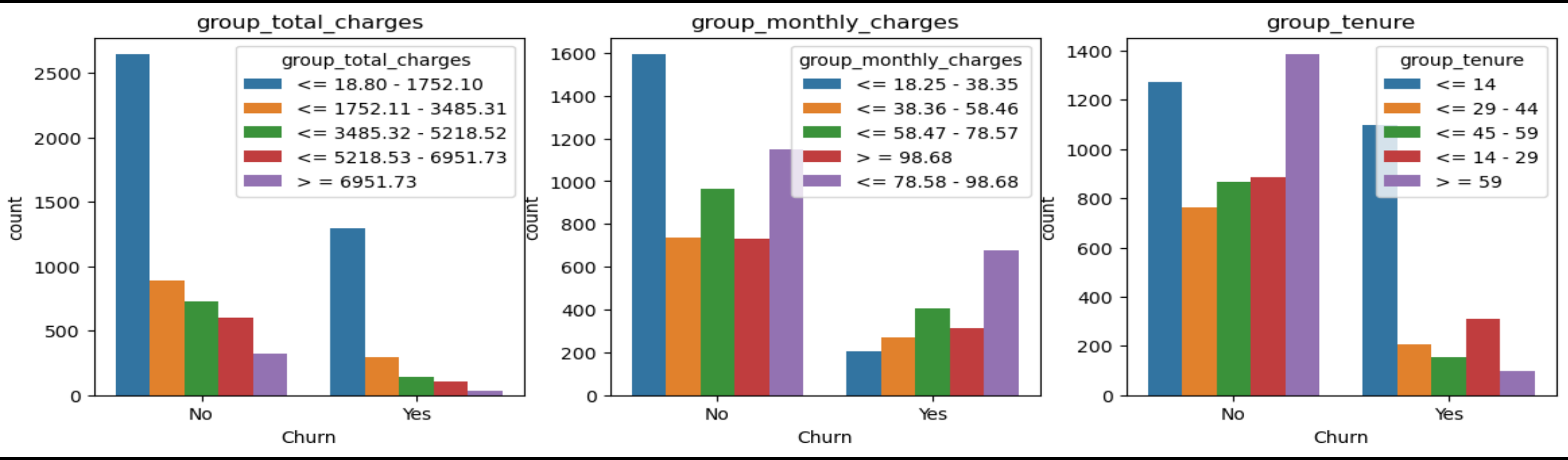
GROUPING THE TOTAL CHARGES INTO BUCKETS

Total billing amounts are generally high, between $18 and $1750



Exploratory Data Analysis:

FEATURE ENGINEERING

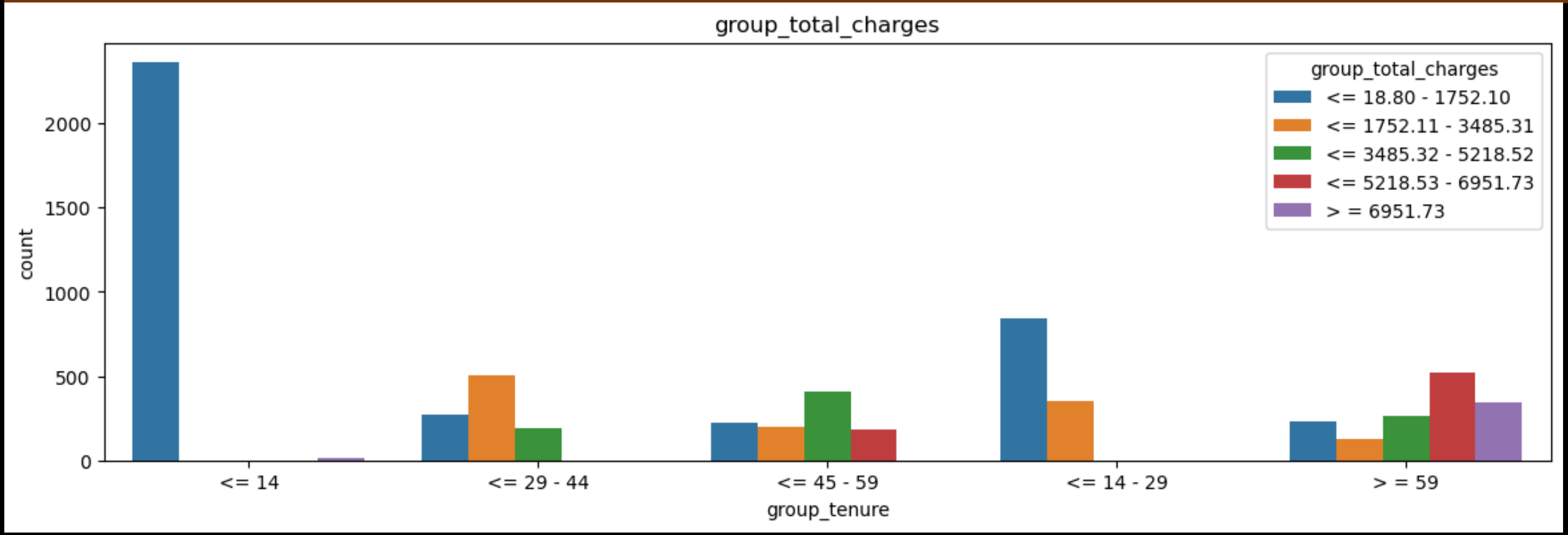


GROUP_TOTAL_CHARGES, GROUP_MONTHLY_CHARGES, GROUP_TENURE VS CHURN

High-billed customers (USD18 - USD1750) churn faster. - Customers on monthly charges especially (78 - 98) tend not to renew their subscription. - Most customers (0 - 14) churn immediately after registration.

Exploratory Data Analysis:

FEATURE ENGINEERING



GROUP_TOTAL_CHARGES VS TENURE.

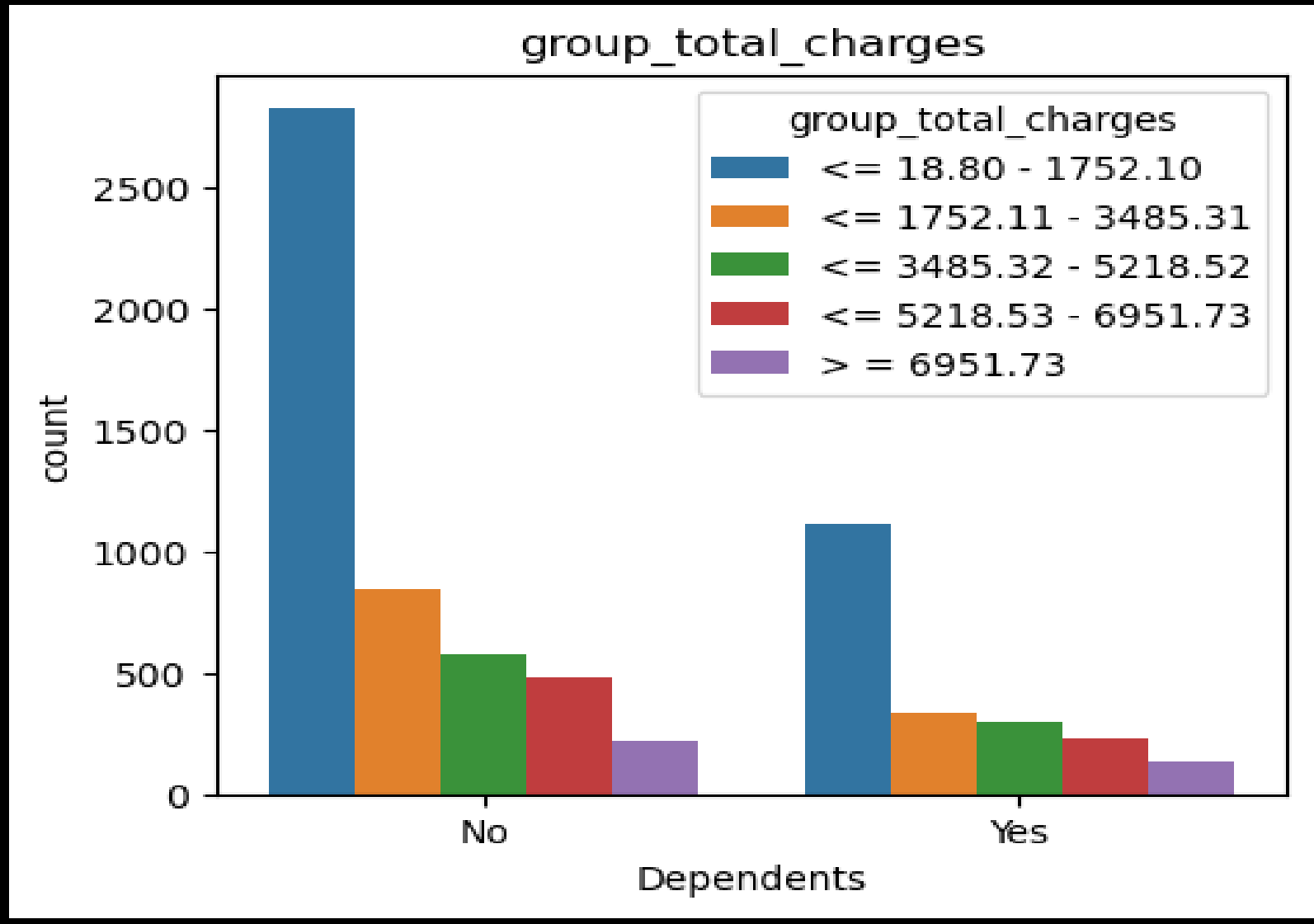
Customers billed highly (USD 2,500) tend to stay within 0-14 days.

Exploratory Data Analysis:

FEATURE ENGINEERING

GROUP_TOTAL_CHARGES VS DEPENDENTS.

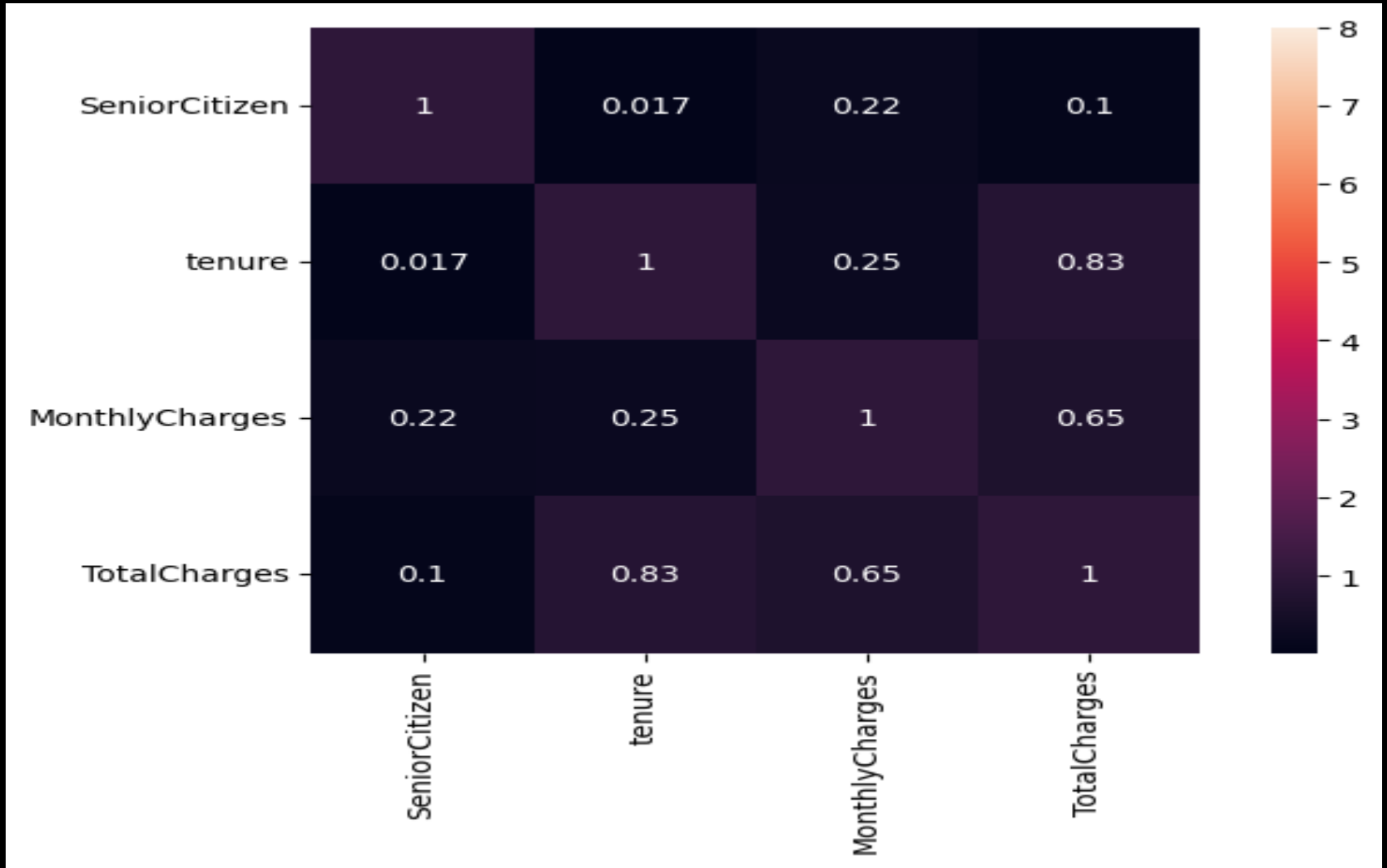
Most dependents on a customer's plan are highly billed, with costs reaching up to USD 2,500



Exploratory Data Analysis:

Multivariate Analysis

There is strong positive correlation between Tenure and Total/Monthly charges.



Data Pre-processing

Replacing the missing values with median in Total Charges and converting the objects to strings.

Encode the categorical features to a numerical features.

Merge both the numerical features and the encoded categorical features.

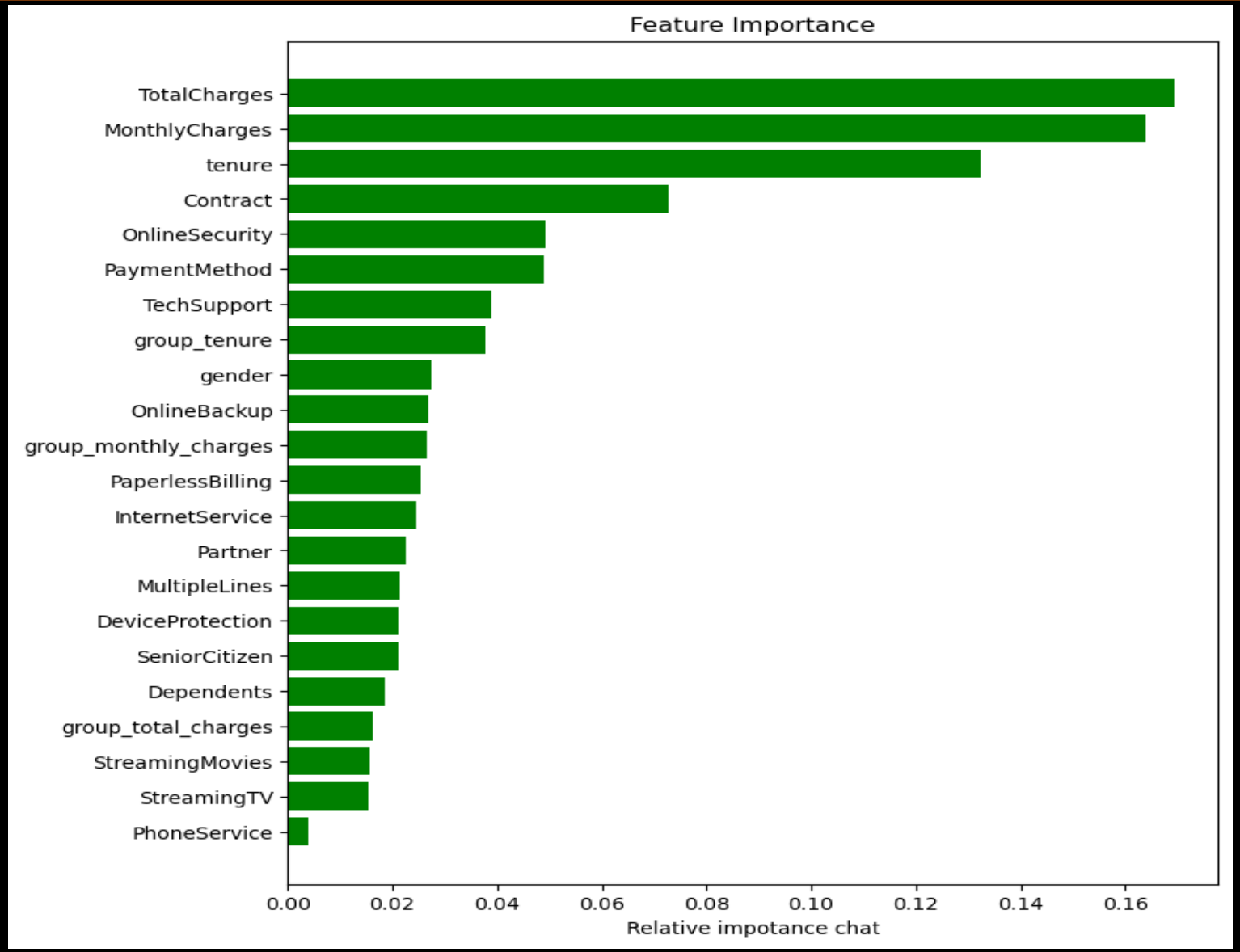
Dropping off redundant features.

Segment dataset into data and target label.

Scale dataset features using the MinMaxScaler by using the RandomForestClassifier.

IMPORTANT FEATURES INFLUECING CHURNING

1. Total charges
2. Monthly charges
3. Tenure
4. Contract type



MACHINE MODELING

- Splitting data into training and test dataset using test_size =0.33
- Oversample the training dataset using the SMOTE.
- Model building using 7 machine learning Algorithms and evaluation.
- Hyper Parameter Optimization Tunning of the 7 models.
- Metrics Evaluation

MODEL PERFORMANCE (BEFORE OPTIMIZATION)

MODEL	ACCURACY	PRECISION	F1-SCORE	AUC/ROC	RECALL
RandomForestClassifier	0.79	0.62	0.60	0.72	0.57
SVC	0.77	0.55	0.63	0.76	0.74
LogisticRegression	0.76	0.54	0.65	0.78	0.83
XGBClassifier	0.78	0.60	0.59	0.72	0.57
SGDClassifier	0.72	0.49	0.63	0.77	0.88
GaussianNB	0.74	0.52	0.63	0.77	0.82
DecisionTreeClassifier	0.72	0.49	0.52	0.67	0.55

HYPER PARAMETER OPTIMIZATION TUNNING

MODEL	ACCURACY	PRECISION	F1-SCORE	AUC/ROC	RECALL
RandomForestClassifier	0.79	0.87	0.85	0.91	0.84
SVC	0.73	0.83	0.81	0.83	0.73
LogisticRegression	0.76	0.92	0.81	0.88	0.72
XGBClassifier	0.79	0.87	0.85	0.91	0.84
SGDClassifier	0.74	0.93	0.80	0.83	0.70
GaussianNB	0.74	0.92	0.80	0.84	0.72
DecisionTreeClassifier	0.75	0.85	0.82	0.85	0.80

CONFUSION MATRIX OF EACH MODEL (AFTER OPTIMIZATION IN %)

Confusion Matrix for RFC:
61.33 (TP) 9.33 (FN)

11.66 (FP) 17.68 (TN)

Confusion Matrix for SVC:
|57.46 (TP) 11.48 (FN)

15.53 (FP) 15.53 (TN)

Confusion Matrix for LR:
52.90 (TP) 4.34 (FN)

20.09 (FP) 22.67 (TN)

Confusion Matrix for XGBC:
61.12 (TP) 9.12 (FN)

11.87 (FP) 17.89 (TN)

Confusion Matrix for SGDC:
50.97 (TP) 4.13 (FN)

22.02 (FP) 22.88 (TN)

Confusion Matrix for GNB:
52.34 (TP) 4.86 (FN)

20.65 (FP) 22.15 (TN)

Confusion Matrix for DTC:
58.58 (TP) 10.67 (FN)

14.41 (FP) 16.34 (TN)

MODEL SELECTION

MODEL	ACCURACY	PRECISION	F1-SCORE	AUC/ROC	RECALL
XGBClassifier	0.79	0.87	0.85	0.91	0.84
RandomForestClassifier	0.79	0.87	0.85	0.81	0.84
DecisionTreeClassifier	0.75	0.85	0.82	0.85	0.80

MODEL SELECTION

Based on the hyperparameter tuning results and metrics evaluation, the top-performing models for predicting churn using False Negatives (FN) in an imbalanced dataset are:

1. XGBClassifier (XGBC)
2. RandomForestClassifier (RFC)
3. DecisionTreeClassifier (DTC).

These models demonstrate a good balance of metrics, with:

- Low FN rates (9.12%, 9.33%, 10,67%) indicating better identification of true positives (churners).
- High recall rates (0.84, 0.80) indicating good identification of true positives (churners).
- High AUC/ROC scores (0.91, 0.85) indicating excellent discrimination between churners and non-churners.
- High F1-scores (0.85, 0.82) indicating a good balance between precision and recall.
- High precision rates (0.87, 0.85) indicating good identification of true positives without flagging too many false positives.

Overall, XGBC, RFC and DTC offer the best balance of metrics, making them suitable models for predicting churn.

Potential Benefits/Recommendation.

1. Upgrading internet and network quality
2. Enhancing customer service relationships
3. Implementing device protection
4. Offering competitive pricing and discounts
5. Introducing incentives and subscription plans
6. Implementing a loyalty program
7. Conducting regular customer feedback surveys
8. Offering personalized plans and services
9. Investing in employee training
10. Monitoring industry trends and competitors

By implementing these strategies, Connectel can:

- Improve customer satisfaction
- Reduce churn
- Increase retention
- Drive business growth and success

Merci beaucoup.