

PYTHON DATA ANALYSIS PROJECT

# TELECOM CUSTOMER CHURN ANALYSIS



**PWC  
CONSULTING  
CASE STUDY**

Prepared by Taara Nguyen

<https://github.com/taaranguyen/PwC-Customer-Churn-Analysis.git>

# BACKGROUND

PwC's Data & Analytics Consulting team has been engaged by a leading telecommunications provider facing elevated levels of customer churn. The company has sought PwC's expertise to uncover churn patterns, identify the primary factors driving attrition, and develop data-driven strategies to enhance customer retention. The objective is to generate actionable insights that will help reduce churn and strengthen loyalty initiatives. In a highly competitive telecom market, where customers often switch providers for better pricing, service quality, or user experience, this company is experiencing substantial monthly customer losses, directly affecting its revenue and profitability. The CEO is seeking clarity on key issues, including which customer segments are most at risk, what's driving churn, and how predictive strategies can help mitigate it.

## EXECUTIVE SUMMARY

**Objective:** The project involves conducting an exploratory data analysis (EDA) on customer demographics and behaviors, followed by the development of a churn prediction model. This model will identify the key risk factors contributing to customer churn, providing valuable insights to inform retention strategies and reduce customer attrition.

**Key insights:** Churn impacts 26.6% of the customer base, with elevated rates among individuals who incur higher monthly charges and frequently engage with technical support services. Demographically, churned customers are predominantly younger—outnumbering senior citizens by nearly three to one—although senior customers exhibit a notably high churn rate relative to their overall population. Gender shows no significant effect on churn, whereas customers without partners or dependents are more inclined to discontinue their services. Attrition is most prevalent during the early stages of the customer lifecycle, particularly within the first 15 months, with the highest rate occurring in the initial month. Service usage patterns also reveal churn tendencies. A large proportion of churned customers had phone services and fiber optic internet, while TV service usage did not show a significant difference between those who stayed and those who left. Billing and payment behavior further correlates with churn. A significant portion of churned customers opted for paperless billing and used electronic checks, indicating a potential connection to higher attrition. Additionally, the majority of churned customers were on month-to-month contracts, with lower churn rates observed among those with longer-term contracts, suggesting that longer commitments contribute to better retention.

**Recommendation:**

- Improve technical support and service quality
- Offer long-term contracts
- Ensure quality and reliability of fiber optic internet service
- Enhance customer engagement and experience for new customers issue

# CHURN ANALYSIS

## 1.CHURN PERCENTAGE

Based on the provided dataset, which includes service information for **7,043 customers**, the **overall churn rate** has been calculated to be **26.58%** (Figure 1).

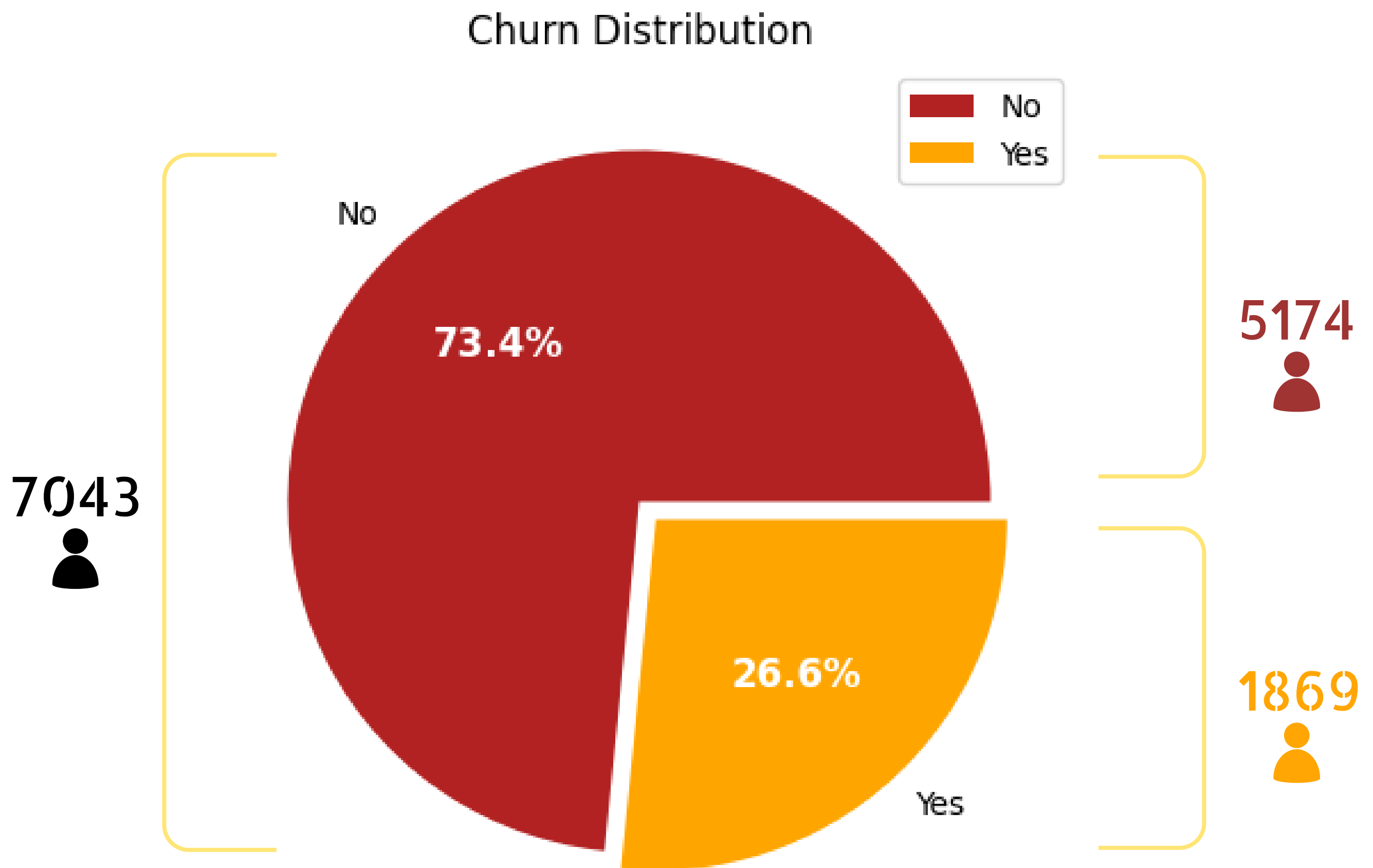


Figure 1. Churn Distribution

## 2.CHURN DISTRIBUTION IN NUMERICAL VARIABLES

(Monthly Charges, Total Charges, Num of Admin Tickets & Num of Tech Tickets)

The customer dataset contains a wide range of service information, including both numerical and categorical variables. Initially, the numerical variables were examined to gain an overview of the churn distribution across different customer segments.

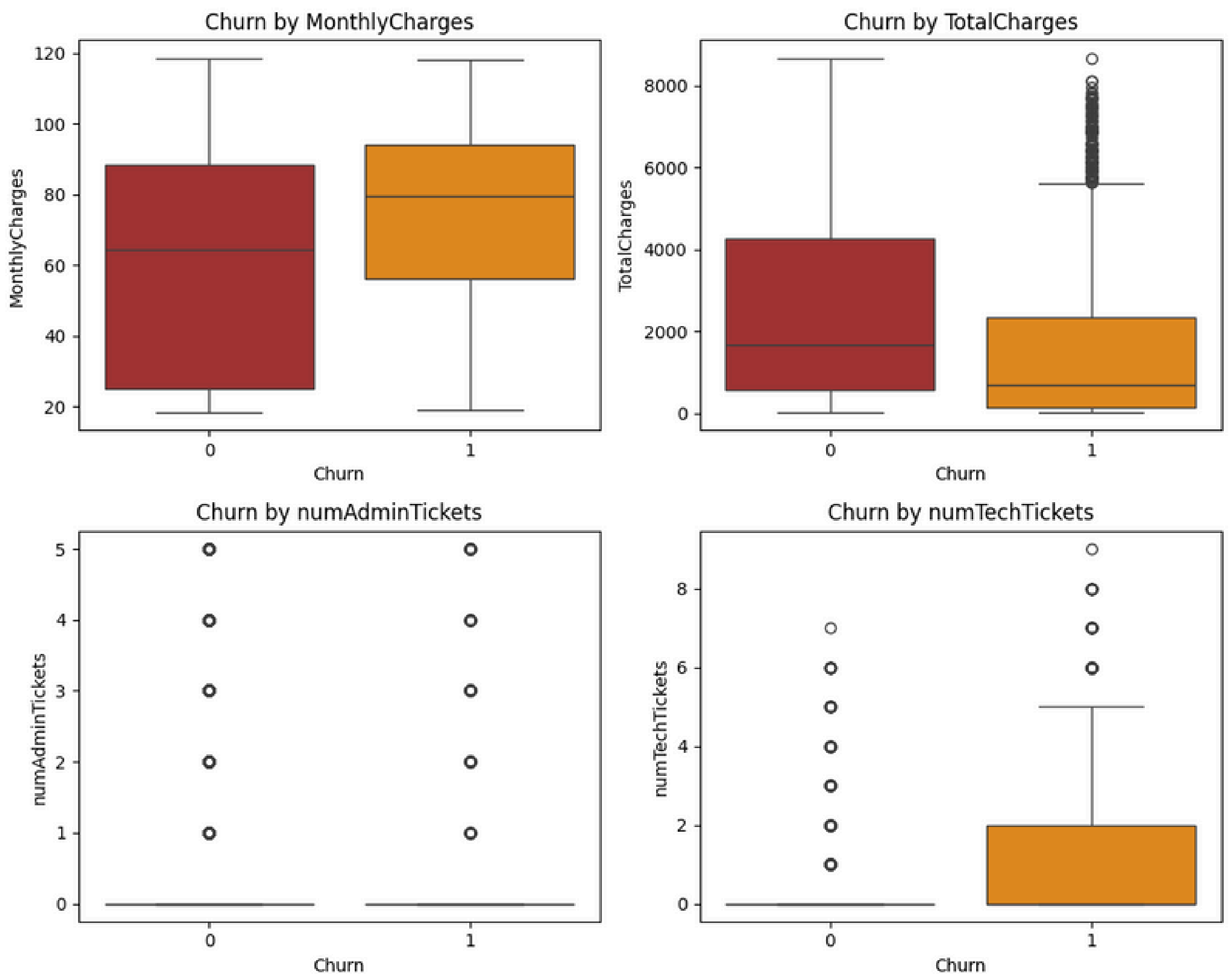


Figure 2. Churn By Monthly Charges, Total Charges, Number of Admin Tickets and Number of Tech Tickets

The charts suggest that customers who **churn** generally exhibit **higher monthly charges**, **lower total charges**, and a **greater number of technical support tickets** (figure 2).

Further analysis of the churn rate in relation to the number of technical support tickets reveals that customers with a **higher frequency of support interactions** are more likely to **churn**. This indicates that repeated service issues or dissatisfaction with the technical support experience may significantly contribute to customer attrition. The data shows a clear **correlation** between an **increasing number of tech tickets** and a **higher probability of churn**, highlighting this as a key area for targeted interventions aimed at enhancing customer retention (figure 3).

Num Tech Tickets	Count	Churn Rate (%)
9	1	100
8	11	100
7	29	96.55
6	72	81.94
5	116	75
4	133	69.17
3	151	66.89
2	201	62.69
1	256	65.63

Figure 3. Table of Churn Rate by Num Tech Tickets

### 3. CUSTOMER DEMOGRAPHIC ANALYSIS

#### 3.1.CHURN DISTRIBUTION IN CUSTOMER SEGMENTS

An analysis of customer demographics, including factors such as gender, age group, partner status, and dependents, was conducted to gain deeper insights into which customer segments are most susceptible to churn. This analysis aimed to identify patterns and characteristics that may correlate with higher churn rates, enabling the company to better understand the specific groups at greater risk of attrition.

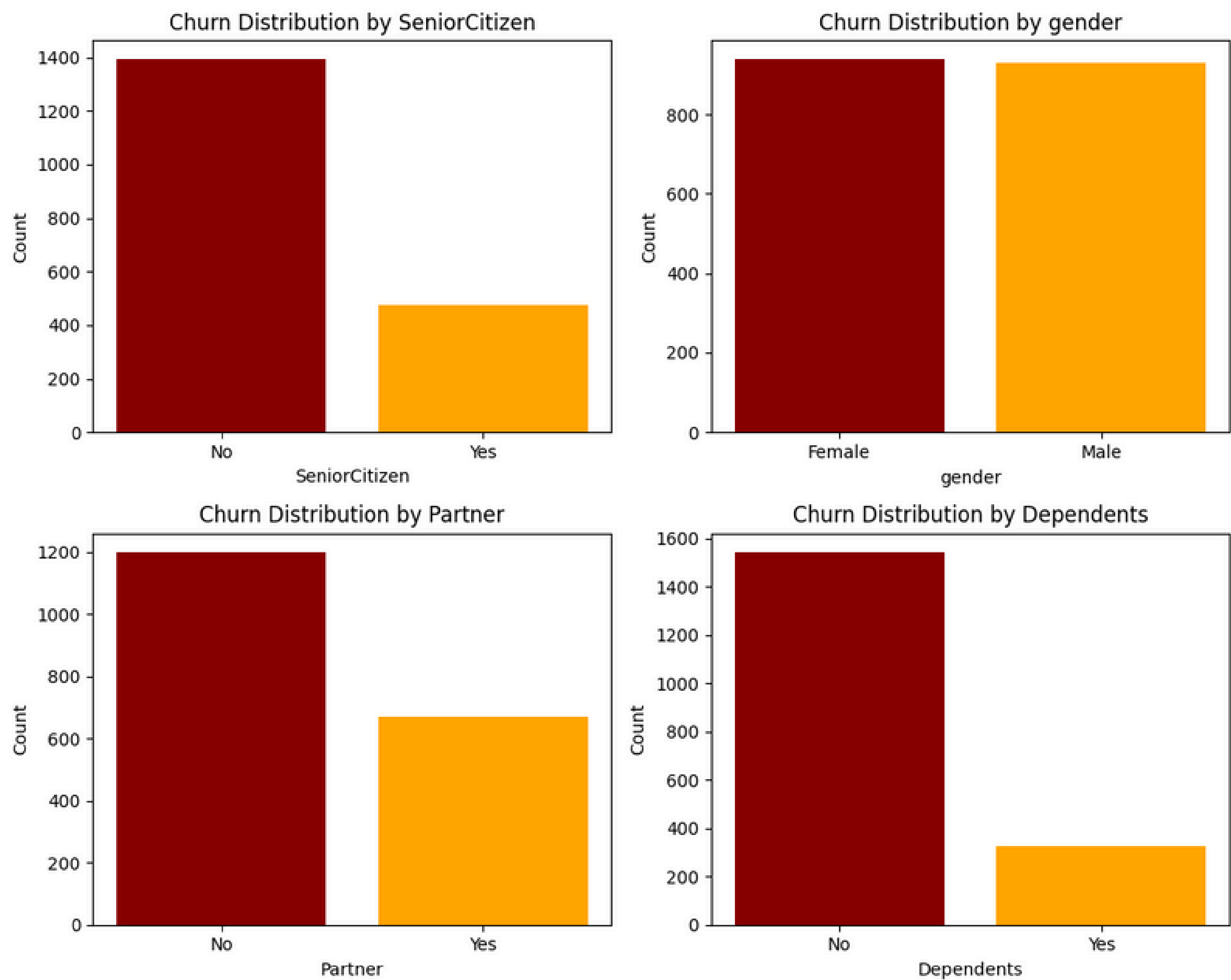


Figure 4. Churn Distribution by Age Groups, Gender, Partner Status and Dependents

The analysis indicates that **churned customers** are generally less likely to be senior citizens, with the number of **younger** customers being three times greater than that of senior citizens. Furthermore, customers **without partners or dependents** exhibit a higher likelihood of attrition, representing 17.07% and 21.94% of the total customer base, respectively. Gender, however, does not show a significant impact on churn rates between male and female customers (figure 4 and figure 5).



Variable	Category	Count	Churn Rate (%)
Senior Citizen	No	1393	19.81
	Yes	476	6.77
Gender	Female	939	13.35
	Male	930	13.23
Partner	No	1200	17.07
	Yes	669	9.51
Dependents	No	1543	21.94
	Yes	326	4.64

Figure 5. Table of Churn Distribution by Age Groups, Gender, Parter and Dependents

### 3.2.CHURN DISTRIBUTION IN CUSTOMER TENURE

Customer tenure is also a significant factor in determining churn probability, with **tenure** ranging from **1 month to 72 months** (figure 6).

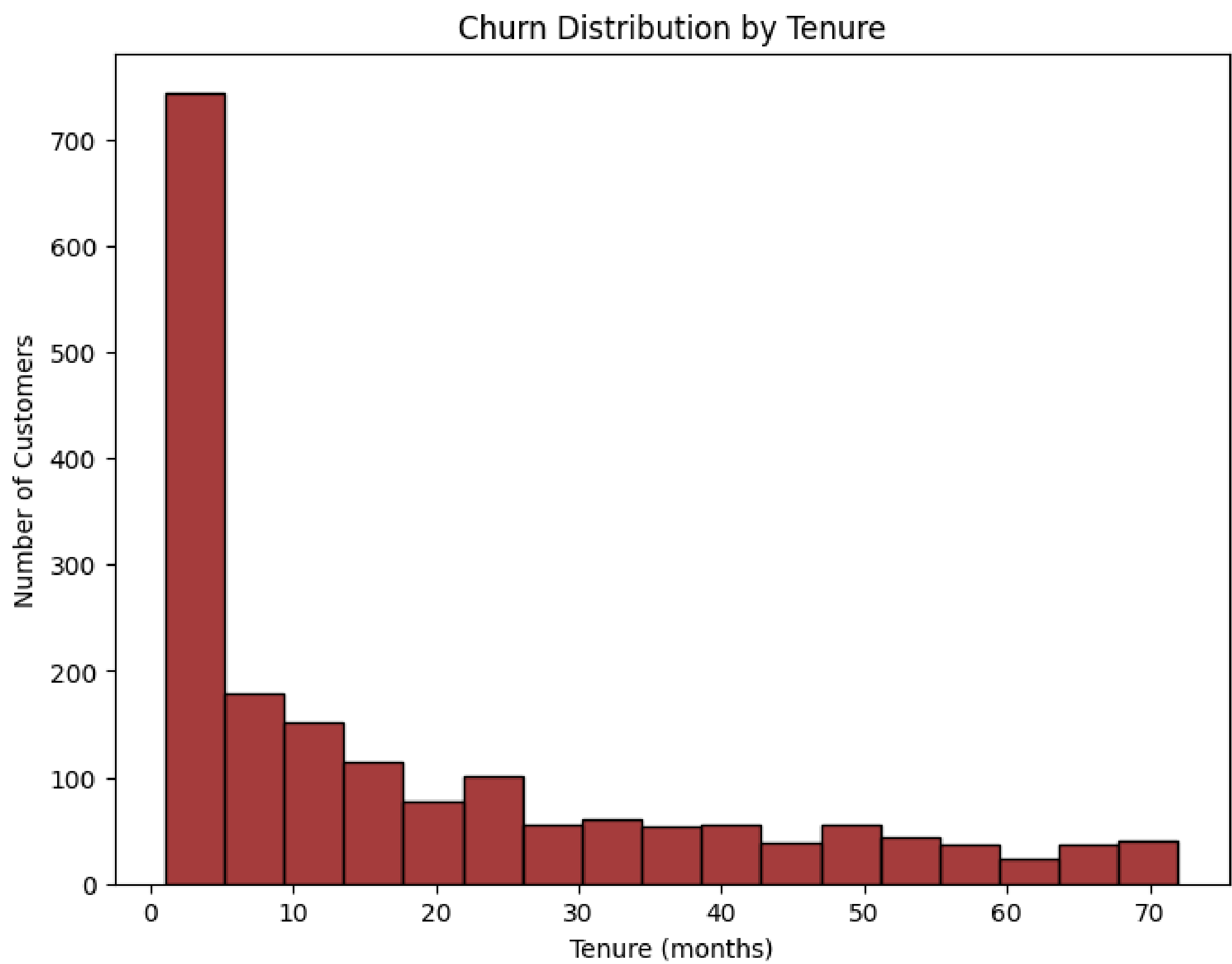


Figure 6. Churn Distribution by Tenure

A detailed examination of **churned customers** reveals that a **large majority** had relatively **short tenure lengths**, indicating a tendency for customers to discontinue their services within the earlier stages of their engagement.(figure 7).

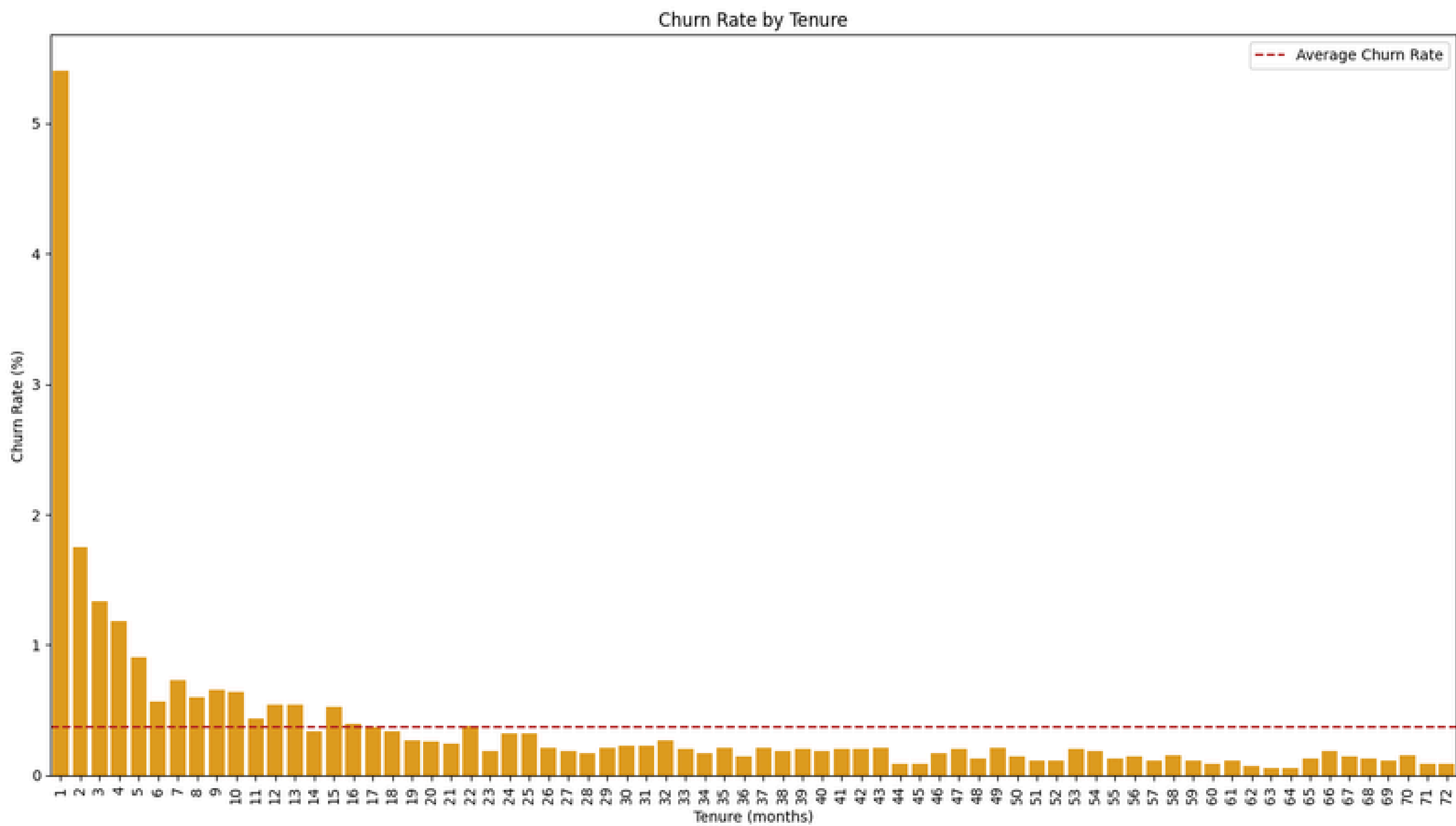


Figure 7. Churn Rate by Tenure

Churn rates are particularly elevated within the **first 15 months of tenure**, ranging from **0.53%** to **5.40%**. This indicates a significantly higher likelihood of customers discontinuing their services during the early stages of their engagement, with the **risk** being especially pronounced in the **first month**. (figure 8).

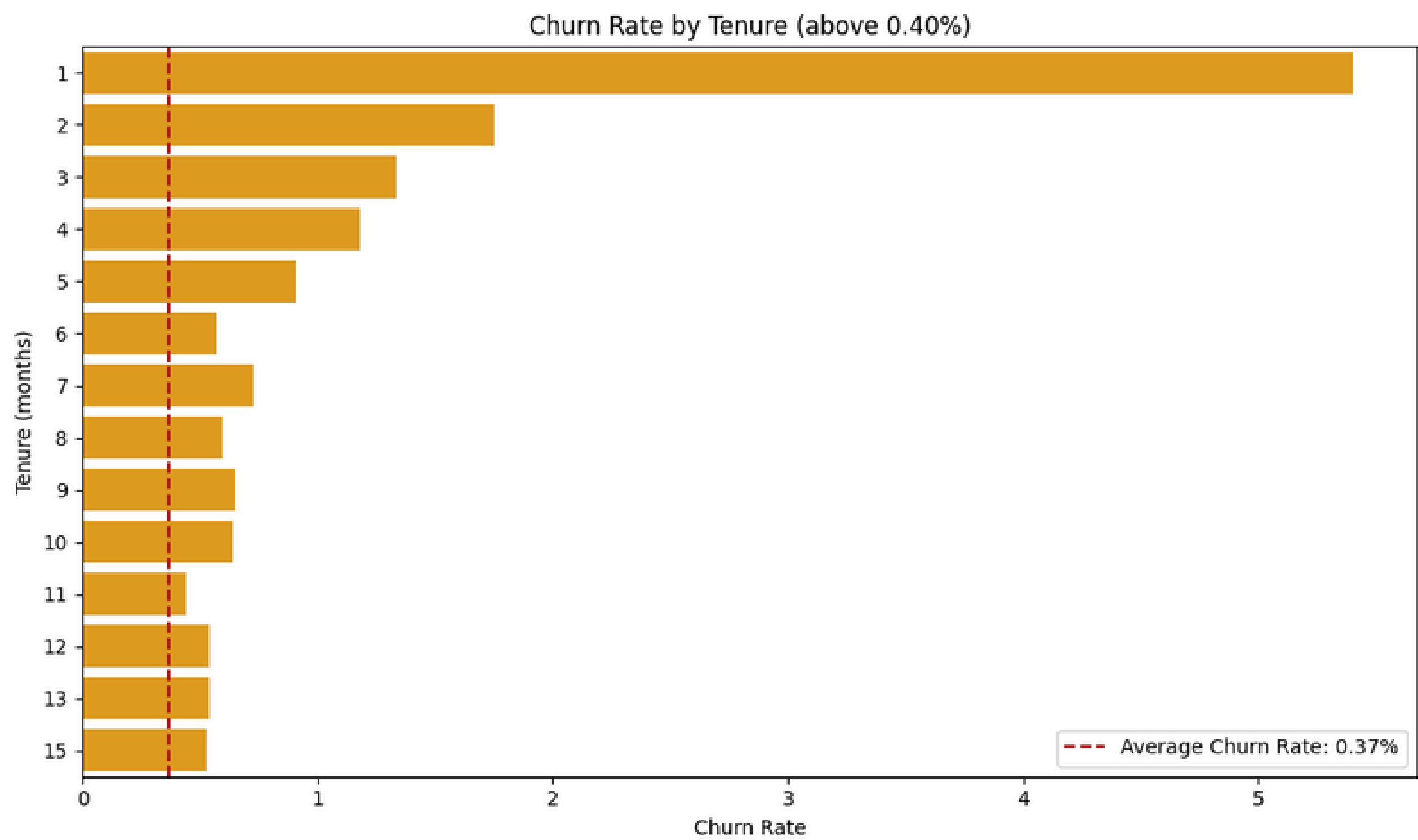


Figure 8. Churn Rate by Tenure (above 0.40%)

### 3.3.CHURN DISTRIBUTION IN SERVICES USED

A breakdown of the services, including Phone, Internet, and TV (Streaming TV), which were used by churned customers, was conducted with the objective of identifying any service-specific trends or correlations with customer attrition.

Customers who **canceled** their subscriptions are **predominantly subscribed to phone services (24.16%)** and **fiber optic internet (18.44%)**. In contrast, **TV service** usage does **not** appear to **significantly influence** the likelihood of churn. (figure 9 & 10).

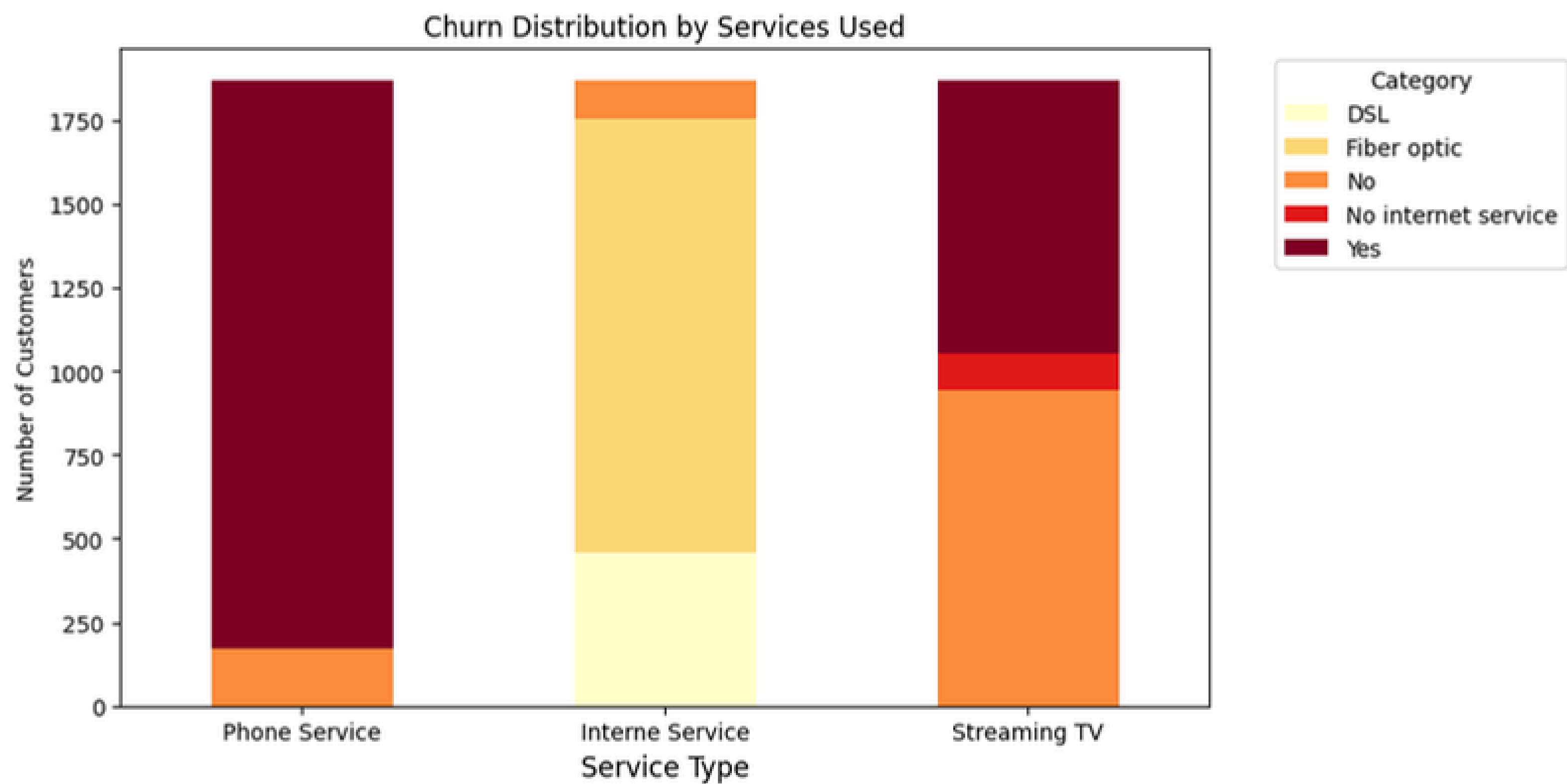


Figure 9. Churn Distribution by Phone Service, Internet Service and Streaming TV

Service	Category	Count	Churn Rate (%)
Phone Service	Yes	1699	24.16
	No	170	2.42
Internet Service	Fiber optic	1297	18.44
	DSL	459	6.53
	No	113	1.61
Streaming TV	No	942	13.40
	Yes	814	11.58
	No internet service	113	1.61

Figure 10. Table of Churn Distribution by Phone Service, Internet Service and Steaming TV



### 3.4.CHURN DISTRIBUTION IN BILLING AND PAYMENT METHODS

An analysis of billing and payment methods was also conducted to identify patterns linked to customer churn. This investigation examined variables such as the adoption of paperless billing and preferred payment options—including electronic checks, mailed checks, bank transfers, and credit cards—to uncover potential correlations with service termination.

Out of the total **1,869 customers** who **churned**, **1,400** selected **paperless billing** as their preferred method of billing. Furthermore, **1,071** of them chose **electronic checks** as their primary payment method, indicating a clear pattern in the billing and payment preferences of those who discontinued their services. (figure 11 & 12).

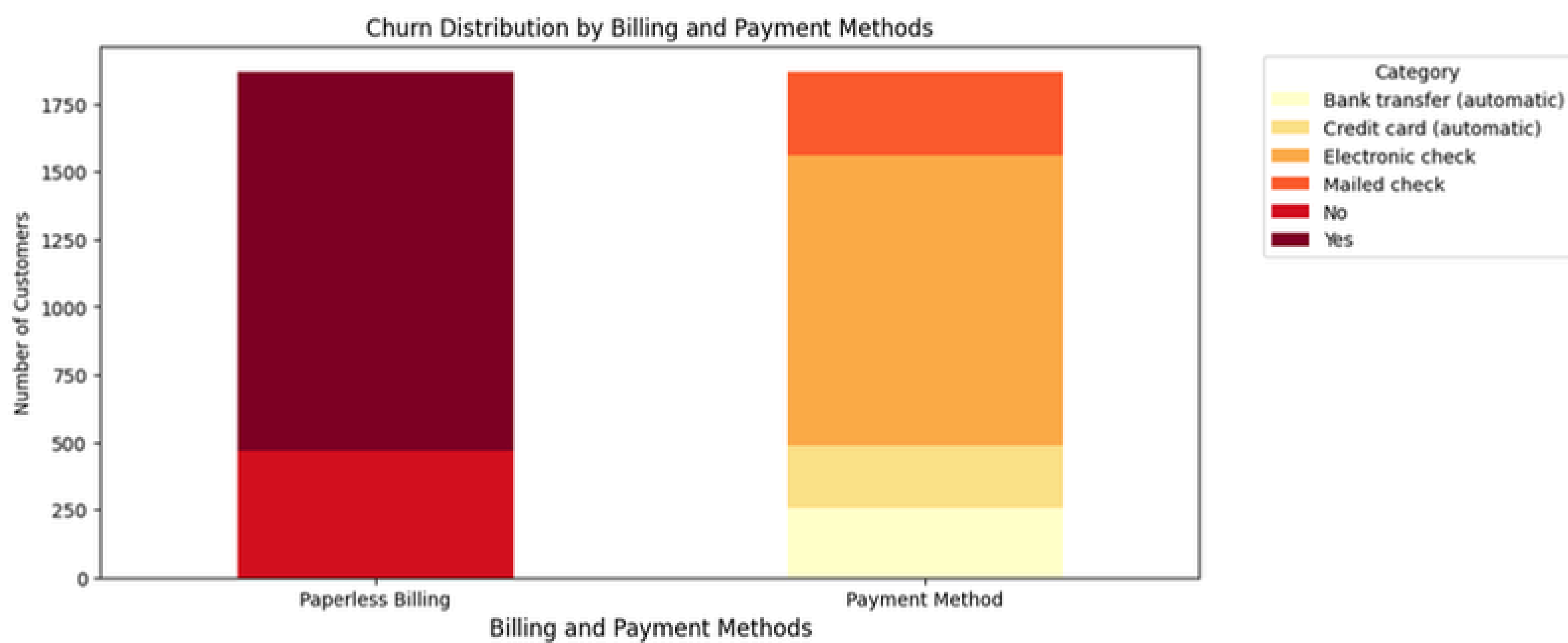


Figure 11. Churn Distribution by Billing and Payment Methods

Service	Category	Count	Churn Rate (%)
Paperless Billing	Yes	1400	19.91
	No	469	6.67
Payment Method	Electronic check	1071	15.23
	Mailed check	308	4.38
	Bank transfer	258	3.67
	Credit card	232	3.30

Figure 12. Table of Churn Distribution by Billing and Payment Methods

### 3.5.CHURN DISTRIBUTION IN CONTRACT TYPES

Given the correlation between short customer tenure and an increased risk of churn, an analysis of contract types was conducted. The contract types examined include month-to-month, one-year, and two-year contracts, to assess their impact on customer retention.

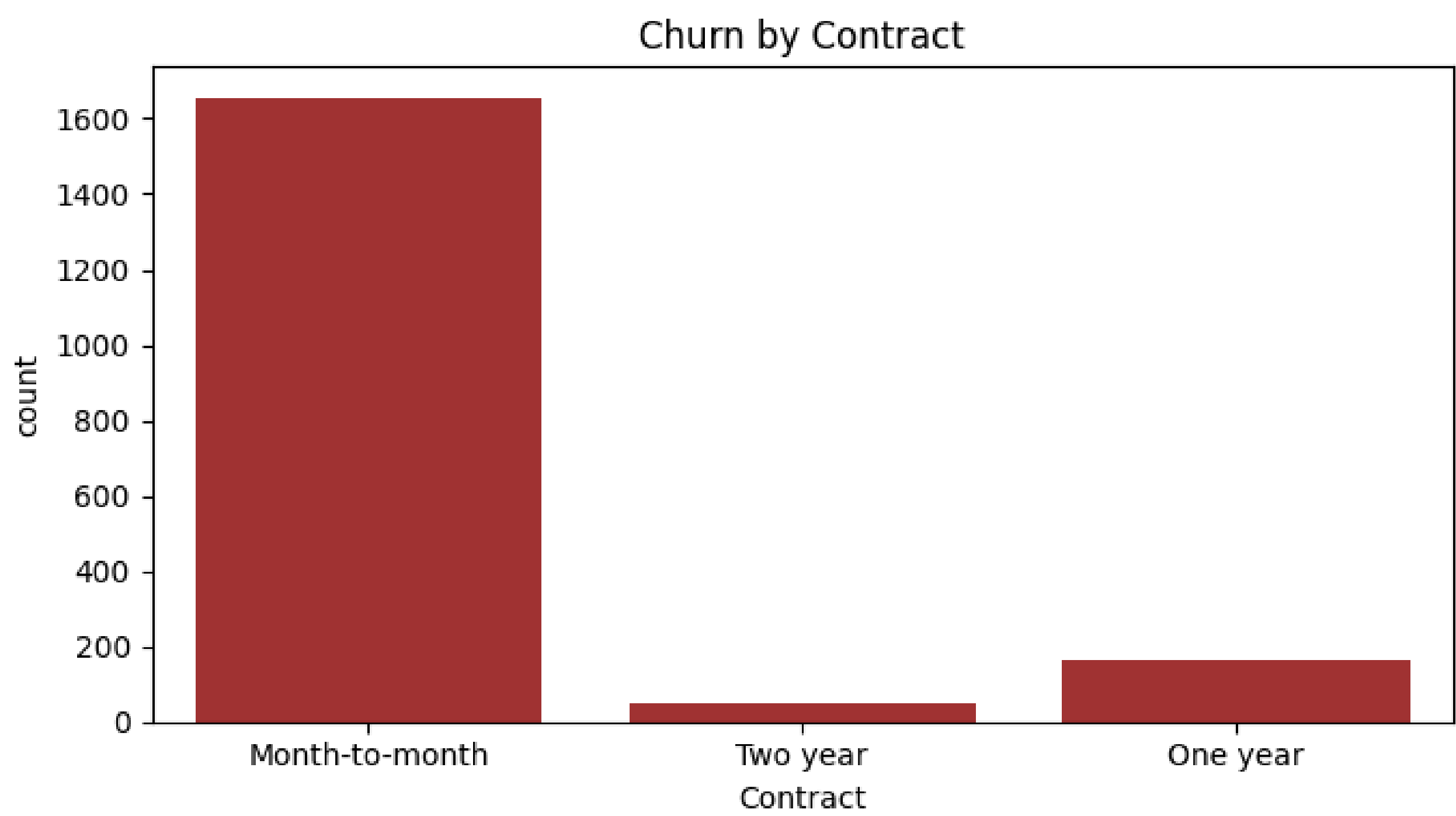


Figure 13. Churn Distribution by Contract Type

The **majority** of **churned customers** were on **month-to-month** contracts, with fewer customers opting out of one- or two-year contracts. This suggests that longer contract durations are linked to a decrease in churn rates. (figure 13 & 14).

Variable	Category	Count	Churn Rate (%)
Contract	Month-to-month	1655	23.54
	One year	166	2.36
	Two year	48	0.68

Figure 14. Table of Churn Distribution by Contract Type

# CUSTOMER CHURN PREDICTIVE MODEL

## 1.MODEL’S RESULTS

To develop a customer churn prediction model, **multiple algorithms** were utilized for testing and training, including **Logistic Regression**, **Decision Tree**, **Random Forest**, and **Support Vector Machine (SVM)**. The performance of each model is outlined below:

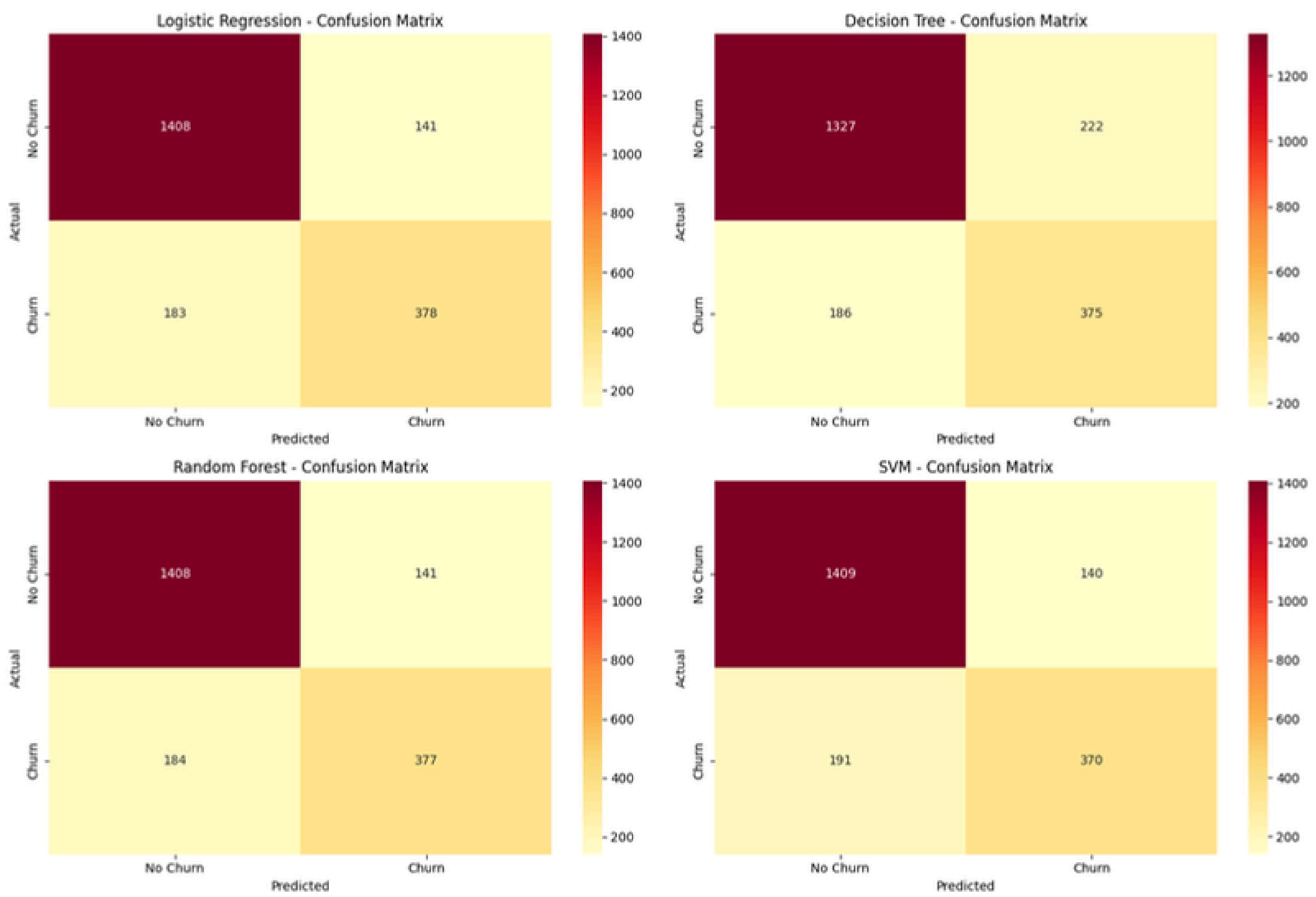


Figure 15. Models’ Confusion Matrix

Model	Accuracy	Precision (Churn)	Recall (Churn)	F1-Score (Churn)
Logistic Regression	0.85	0.73	0.67	0.70
Decision Tree	0.81	0.63	0.67	0.65
Random Forest	0.85	0.73	0.67	0.70
SVM	0.84	0.73	0.66	0.69

Figure 16. Models’ Classification Report

Among these models, **Logistic Regression** emerged as the **most effective**, achieving the **highest accuracy of 85.99%** and accurately predicting the **largest number of churns**, totaling **378** (figure 15, 16 & 17).

Model	Cross-Validation Accuracy
Logistic Regression	0.8599 (+/- 0.0152)
Decision Tree	0.8066 (+/- 0.0171)
Random Forest	0.8497 (+/- 0.0134)
SVM	0.8564 (+/- 0.0114)

Figure 17. Models’ Cross Validation Accuracy

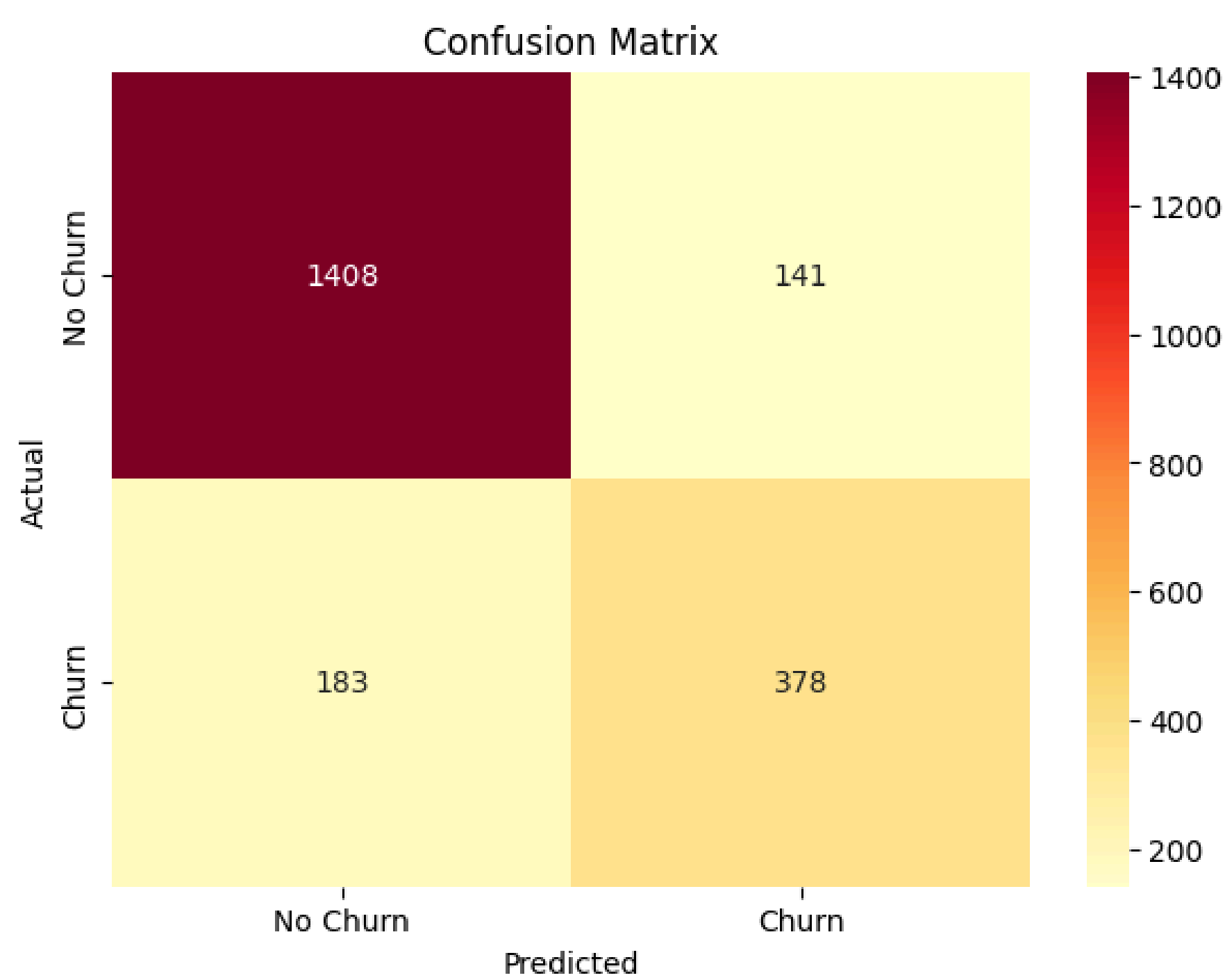


Figure 18. Logistic Regression Confusion Matrix

	precision	recall	f1-score	support
0	0.88	0.91	0.90	1549
1	0.73	0.67	0.70	561
accuracy			0.85	2110
macro avg	0.81	0.79	0.80	2110
weighted avg	0.84	0.85	0.84	2110

Figure 19.. Registic Regression Classification Report

However, several **limitations** should be taken into account before deploying the model in a business setting:

- The model’s relatively low recall for churned customers (67%) (figure 18) implies that approximately one-third of actual churn cases may go undetected. This poses a risk of missing the opportunity to proactively retain a significant portion of at-risk customers, potentially leading to substantial revenue loss.
- Moreover, the model misclassifies 141 loyal customers as potential churners, which could result in the misallocation of retention efforts and resources—such as unnecessary incentives or outreach—toward customers who were not at risk of leaving.
- Finally, because Logistic Regression relies on linear assumptions, it may not effectively capture complex or non-linear relationships among variables that influence churn behavior. This oversimplification can reduce predictive accuracy, especially in environments where customer behavior is dynamic or influenced by subtle patterns.

## 2.FEATURE IMPORTANCE ANALYSIS

A feature importance analysis was conducted using the aforementioned models, revealing that certain features consistently ranked at the top in most analyses. These key features include number of tech tickets, month-to-month contract, fiber optic internet service and tenure.

Below is the feature importance derived from the Logistic Regression model, which supports the findings mentioned above. The top factors identified are:

- **Number of Tech Tickets** (1.25): A high volume of technical support tickets is strongly correlated with churn, suggesting potential service issues or customer dissatisfaction.
- **Month-to-month Contract** (1.15): This is the most significant predictor of churn, indicating that customers with month-to-month contracts are at a higher risk of cancellation.
- **Fiber Optic Internet Service** (0.85): The presence of fiber optic internet service has a notable impact on churn, with customers using this service more likely to churn.
- **Customer Tenure** (-1.5): A shorter tenure is strongly associated with an increased probability of churn, suggesting that customers with less time on their contracts are more likely to leave.

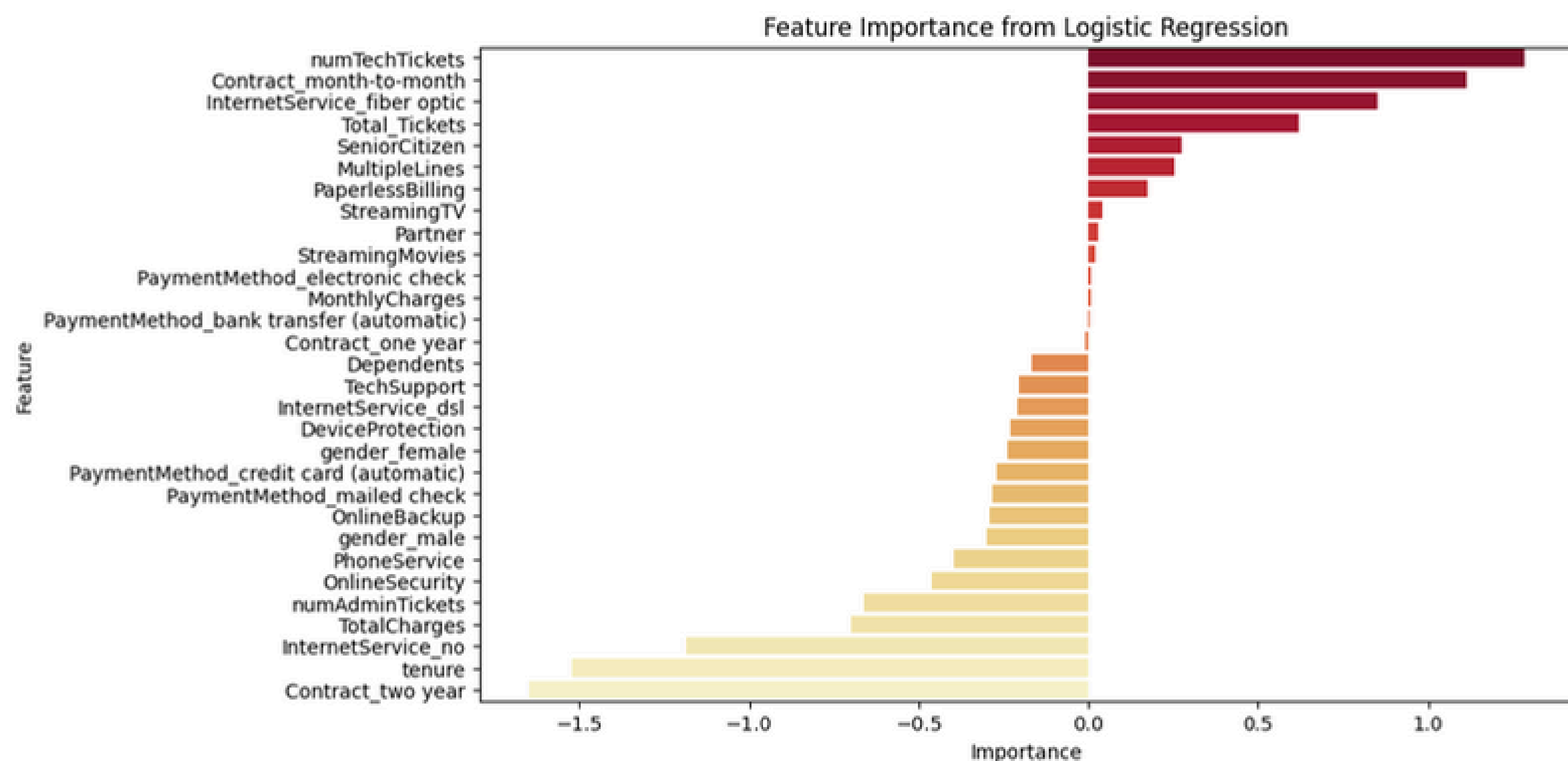


Figure 19. Feature Importance from Logistic Regression



# INSIGHTS AND RECOMMENDATIONS



## Improve Technical Support and Service Quality

**Issue:** *A high number of technical support tickets is linked to churn, indicating service issues or dissatisfaction.*

**Recommendation:** Invest in improving the quality of customer support and technical service. This could include:

- Training customer service representatives for faster resolution times.
- Implementing a more effective troubleshooting process to address common issues before they escalate.
- Proactively reaching out to customers who are experiencing frequent technical issues to resolve them before they decide to churn.
- Create a feedback system to gain insights into the causes of customer dissatisfaction. Conduct regular surveys to assess customer experiences, pinpoint areas of concern, and take action based on the feedback. Make it easy for customers to report problems and ensure their voices are heard.



## Offer Long-Term Contracts

**Issue:** *Customers with month-to-month contracts are at a higher risk of churn.*

**Recommendation:** Encourage customers to switch to longer-term contracts by offering attractive incentives, such as discounts or added benefits for committing to longer terms (e.g., 1-year or 2-year contracts). This would help increase customer retention and reduce churn rates



## Ensure Quality and Reliability of Fiber Optic Internet Service

**Issue:** *The presence of fiber optic internet service affects churn likelihood.*

**Recommendation:** Invest in ensuring the quality and reliability of fiber optic services, as it significantly influences customer retention. If fiber optic is a premium service, emphasize its advantages in marketing materials, and offer targeted promotions to retain fiber optic customers by addressing any network or service issues proactively.



## Enhance Customer Engagement and Experience for New Customers

**Issue:** *Shorter tenure correlates with increased churn probability.*

**Recommendation:** Focus on increasing engagement with new customers through:

- Personalized onboarding experiences.
- Regular follow-up to ensure customers are satisfied and fully utilizing their services.
- Offering incentives or rewards for customers who stay for a longer period (e.g., free upgrades or service bundles after a certain number of months).