# CUSTOMER CHURN ANALYSIS

AUTHORS:
Anh Tung Le (Tony)
Code link – Press here

## BACKGROUND

A leading telecom company is experiencing high customer churn, impacting revenue and profitability. Customers frequently switch providers due to better pricing, service quality, or enhanced contract experiences. As competition intensifies, understanding and mitigating churn has become crucial for business sustainability.

PwC's Data & Analytics Consulting Division has been engaged to analyze customer churn patterns, identify key churn drivers, and develop a predictive model to help the company proactively retain customers. By leveraging data-driven insights, this project aims to support strategic decision-making and improve overall customer loyalty.

## TOOLS

matplotlib

pandas

statsmodels

scikit learn

## EXECUTIVE SUMMARY

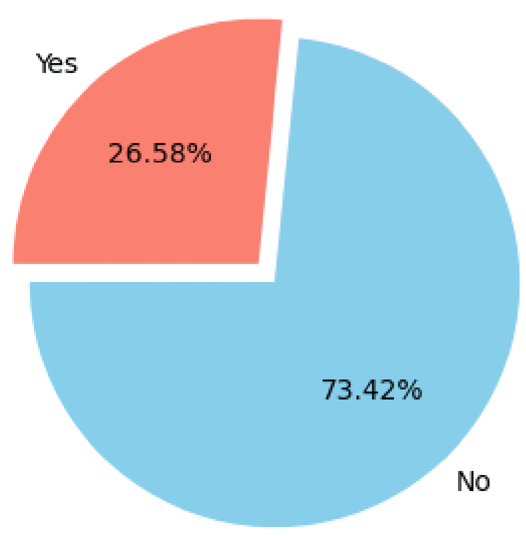
Customer churn presents a major challenge for telecom companies, affecting revenue and long-term growth. **This project, led by PwC's Data & Analytics Consulting Division, applies machine learning to analyze churn patterns, identify key drivers, predict churn, and guide retention strategies.**

Through **customer segmentation, service usage patterns, and support engagement analysis**, several models were tested, with **Tuned XGBoost selected as the best-performing model**. It achieved 85% accuracy, 74% precision, and the highest recall (70%), making it the most effective at identifying true churners. Unlike single decision trees, XGBoost builds multiple trees that correct earlier mistakes, improving overall prediction quality.

**Key recommendations** include implementing **loyalty programs, long-term contract incentives, and proactive support interventions** to reduce churn and improve customer satisfaction.

## ANALYSIS

The analysis reveals that **approximately 27% of customers have churned**, indicating a notable retention issue for the telecom provider. While the majority of customers remain, this **churn rate represents a substantial revenue risk** that needs to be addressed. This study will primarily **focus on identifying the key drivers behind customer attrition and** developing data-driven strategies to **enhance retention, improve customer satisfaction, and reduce future churn**.



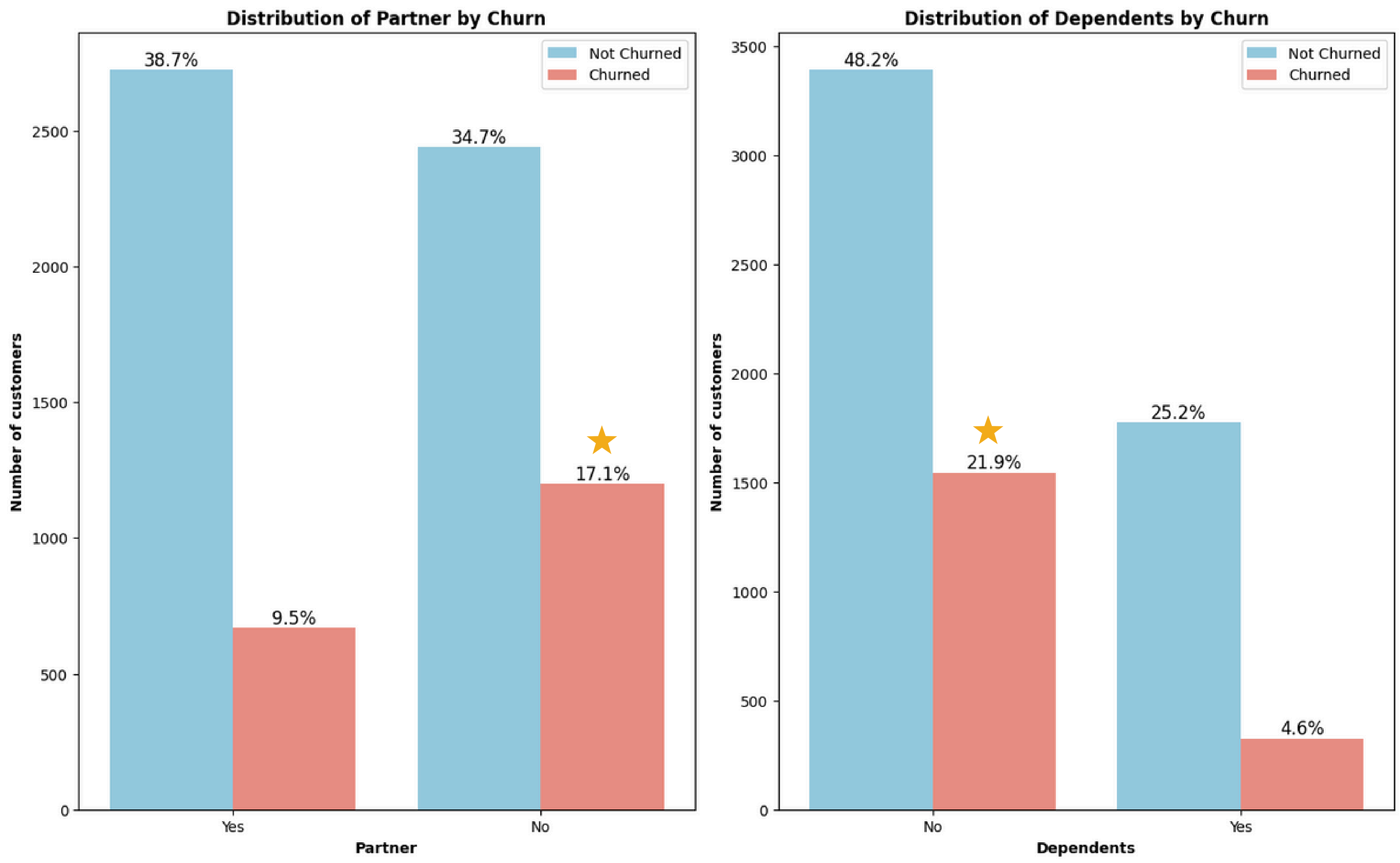

## RECOMMENDATION

In conclusion, the analysis reveals that customer churn is largely **influenced by short tenure, premium service dissatisfaction, flexible contract types, and limited support engagement** rather than demographic factors. New customers are especially vulnerable within their first 10 months, while premium users often churn due to unmet value expectations. Additionally, customers on month-to-month contracts and those who rarely interact with support services are more prone to leaving. To effectively reduce churn, businesses should **focus on strengthening early-stage engagement, enhancing the perceived value of premium offerings, incentivizing long-term contracts, and encouraging proactive use of support channels to resolve issues before dissatisfaction escalates**.

# DESCIPTIVE ANALYSIS

## 1. CUSTOMER SEGMENT

**Figure 1: Distribution of Customer's relationship status by Churn**




+) Higher churn among singles:

- **Customers without a partner have a higher churn rate** (27.1%) compared to those with a partner (9.5%).
- This suggests that single customers are more likely to switch providers, possibly due to fewer shared commitments or financial constraints.

+) Dependents impact churn:

- Customers without dependents have a churn rate of 21.1%, significantly higher than those with dependents (4.6%).
- This indicates that customers with family responsibilities may be more stable and loyal, possibly due to bundled family plans or shared usage benefits.

**48.06% (males) & 50.69% (females) churn in the first ten months**

- This suggests that customers leave early, likely due to dissatisfaction with service, pricing, or unmet expectations.

+) Churn decreases with tenure

- As **tenure increases, churn gradually declines**, with significantly lower churn rates beyond 30 months.
- Customers who stay longer tend to be more loyal and less likely to leave.

+) Gender distribution is relatively balanced

- The churn rate is similar for both males and females across different tenure periods.
- Slightly more females churn than males in the early tenure (0-10 months), but the gap narrows over time.

**Figure 2: Population Pyramid of Churn by Tenure and Gender**

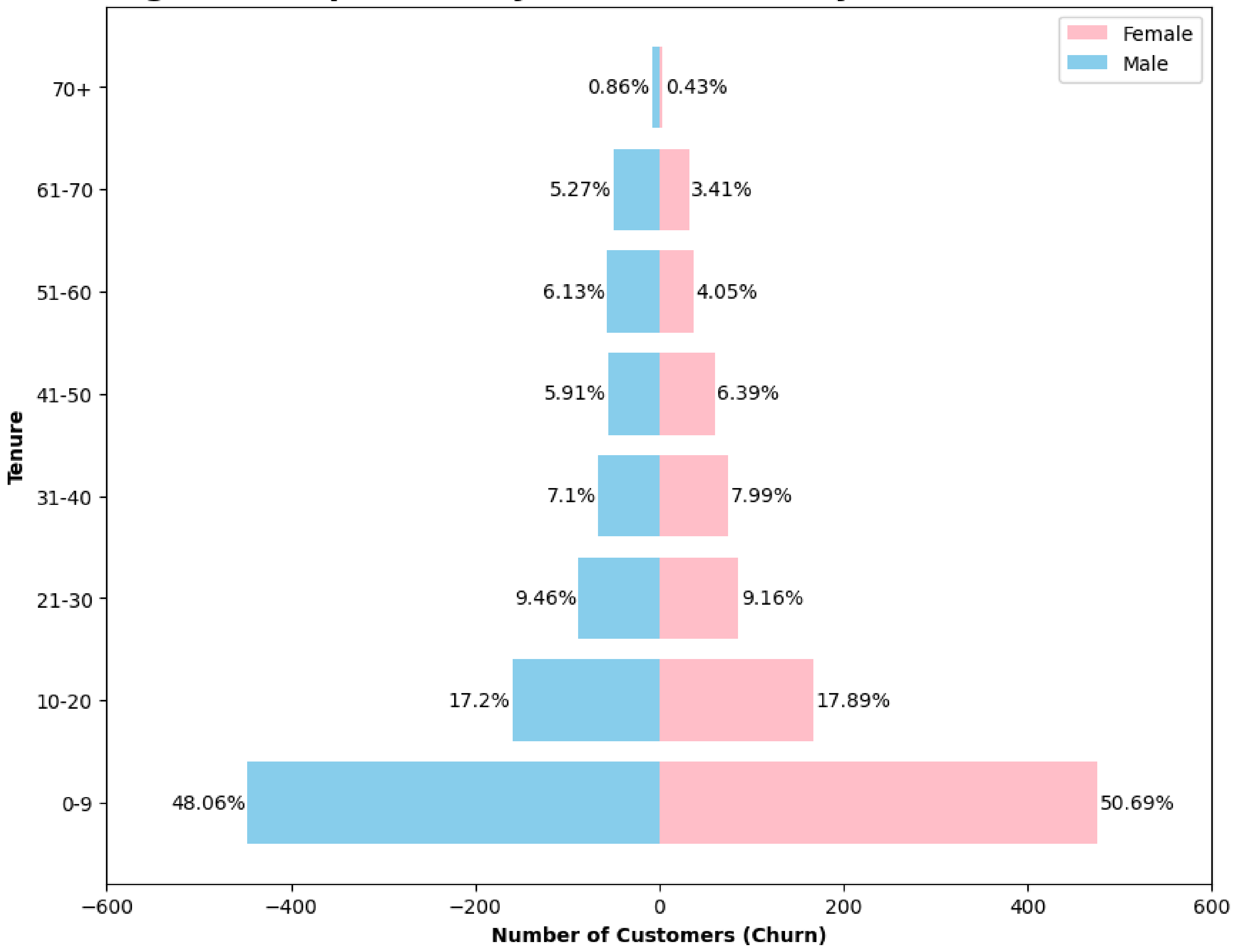
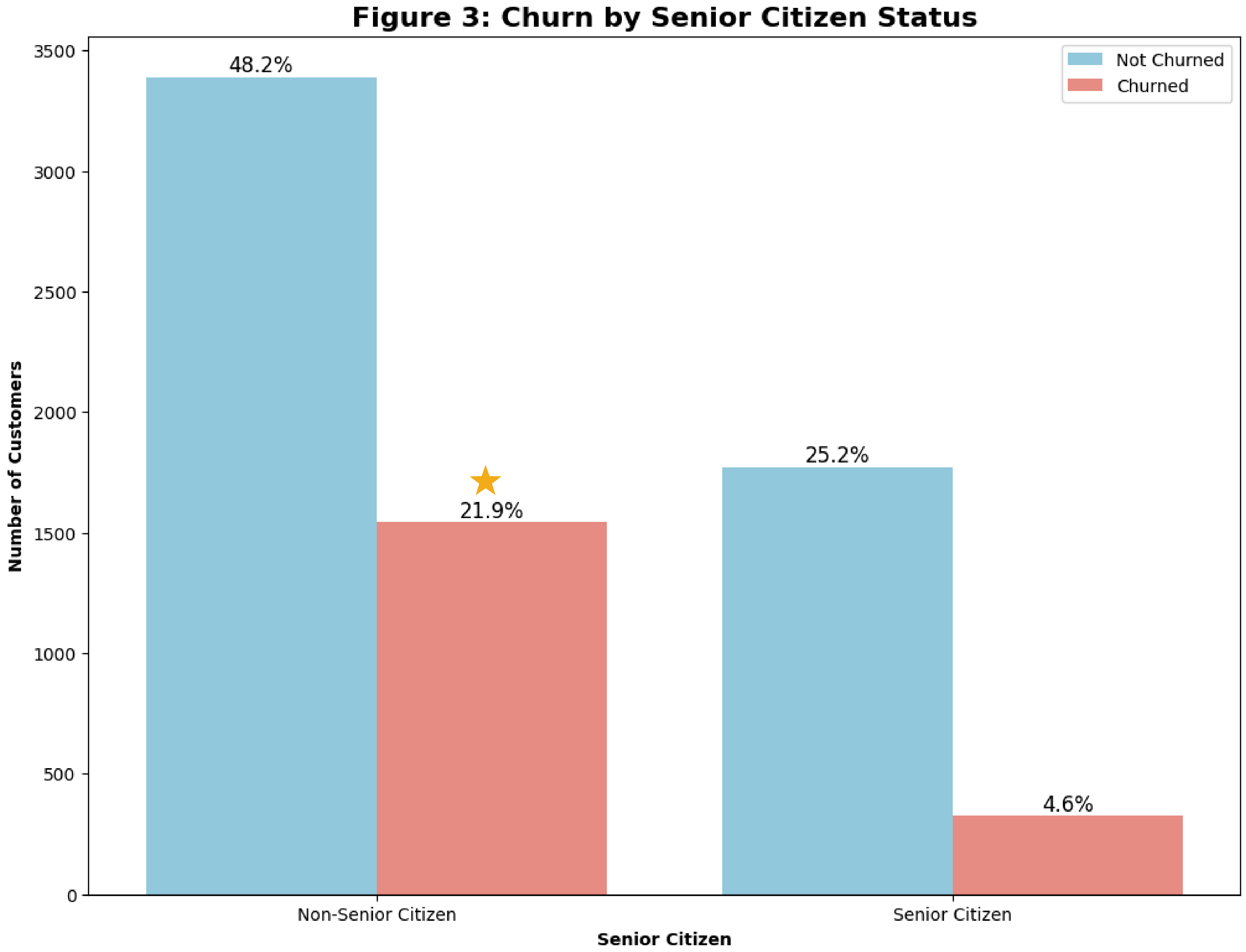



**Figure 3: Churn by Senior Citizen Status**




+) Senior citizens & Young people

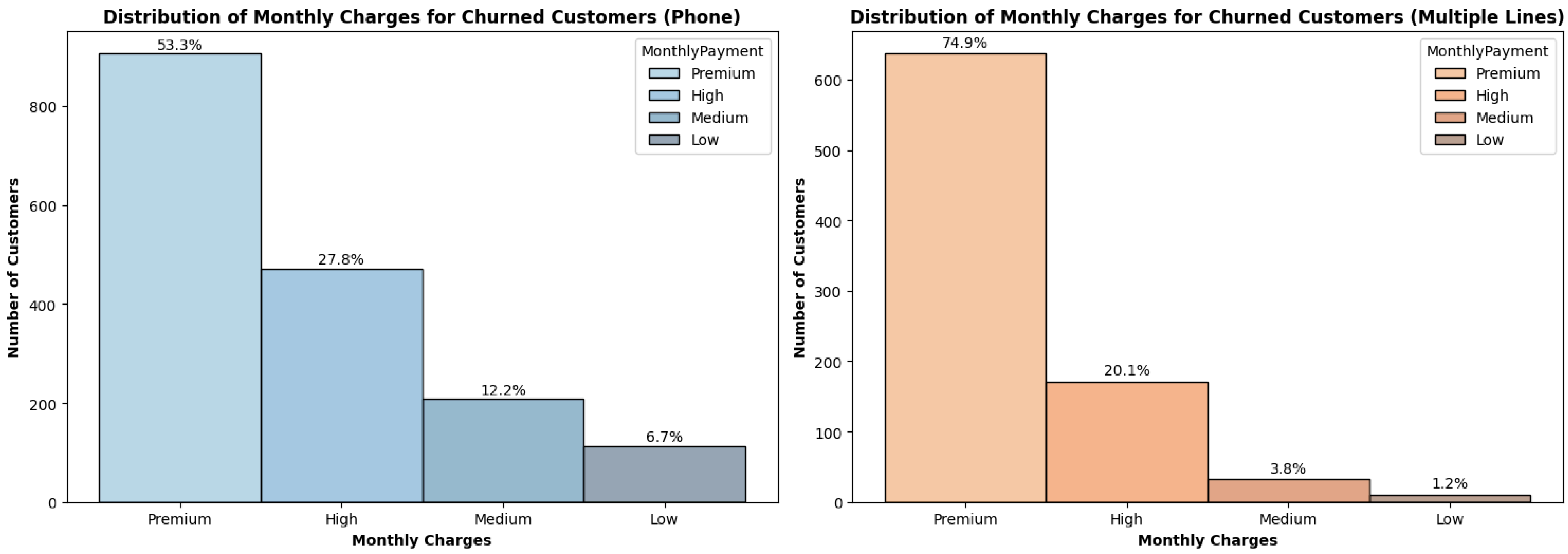
- 21.9% of non-senior citizens have churned, while 48.2% remain active customers.
- Only 4.6% of senior citizens have churned, compared to 25.2% who stayed with the service.
- This indicates that **senior citizens are less likely to leave their telecom provider**.

+) Churn behavior varies by age group:

- The data suggests that **younger customers are more likely to churn**, potentially due to greater willingness to explore other providers or dissatisfaction with current services.

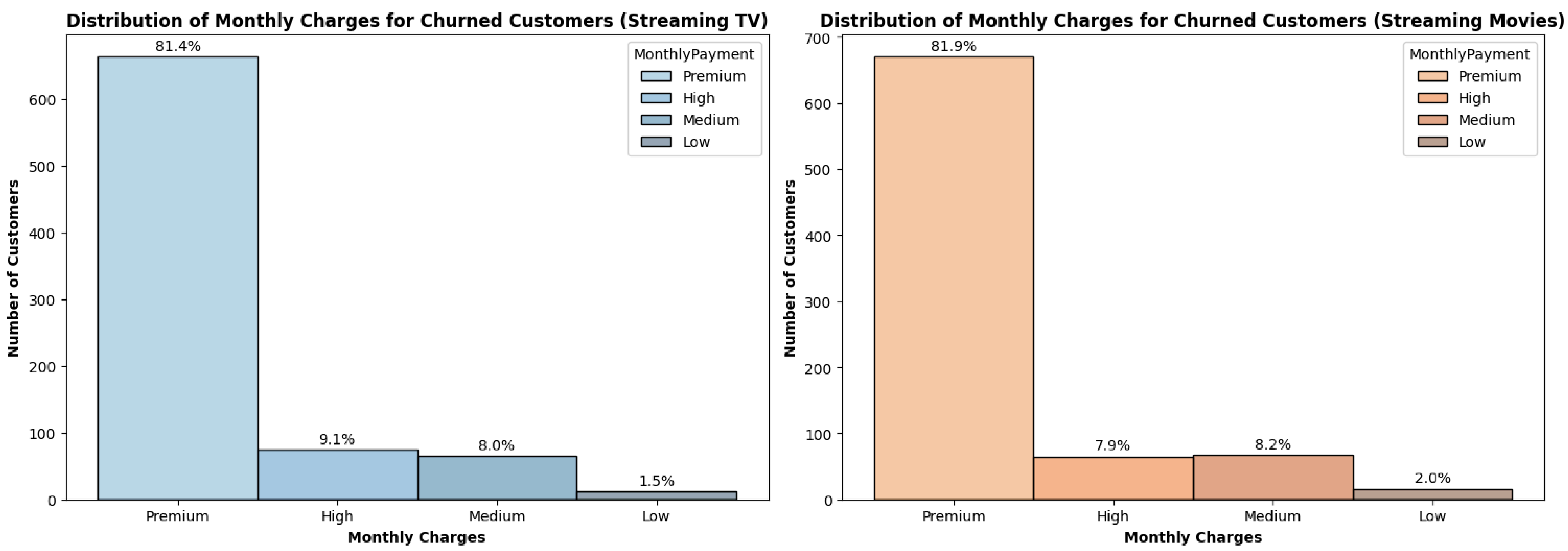
## 2. SERVICE USED

**Figure 4: Distribution of Services Used by Churn**



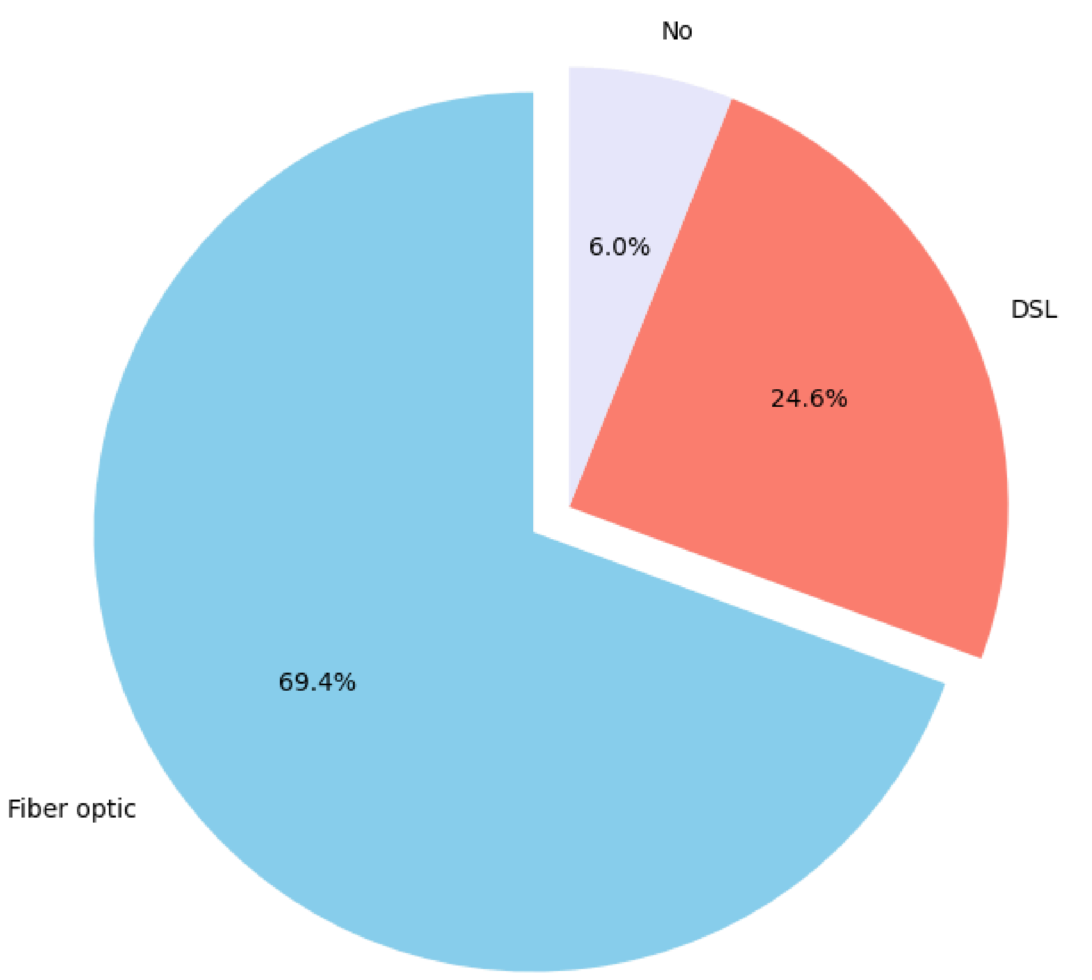

In **the Premium payment tier are the most likely to leave**, particularly among customers using Multiple Lines, where a striking 74.9% of churned users fall into this category. Similarly, among churned customers with Phone service, 53.3% are also in the Premium tier. The proportion of churned customers decreases notably as monthly charges decline—dropping to 27.8% (High), 12.2% (Medium), and 6.7% (Low) for Phone users, and 20.1%, 3.8%, and 1.2%, respectively, for those with Multiple Lines. This consistent pattern across both service types suggests that higher-paying customers are more prone to churn, potentially due to greater service expectations, dissatisfaction with value, or heightened price sensitivity. The data indicates that **premium segments may require closer attention, as they not only represent a significant portion of revenue but also a higher risk of attrition.**

**Figure 5: Distribution of Services Used by Churn**




**The largest proportion of churned users belongs to the Premium payment tier**. In Streaming TV, 81.4% of churned customers fall into this category, with significantly fewer in the High (9.1%), Medium (8.0%), and Low (1.5%) tiers. A comparable trend is evident in Streaming Movies, where 81.9% of churned users are Premium subscribers, followed by 7.9% in High, 8.2% in Medium, and just 2.0% in Low. These figures indicate that customers paying the highest monthly charges are far more likely to churn, especially in the context of streaming services. **The parallel patterns across both TV and movie services suggest that premium users may have higher expectations, and are thus more sensitive to perceived value, content quality, or pricing, which can influence their decision to cancel**.

**Figure 6: Customer Churn by Internet Service**




+) Higher churn among fiber optic services users

- **Fiber optic services, likely referring to NBN (National Broadband Network)**, account for the majority of churned customers (69.4%). Despite offering **faster and more advanced internet, fiber optic users still experience the highest churn rate**, suggesting that speed alone may not satisfy customer expectations.
- DSL users make up 24.6% of churned customers, indicating a moderate level of dissatisfaction, possibly due to outdated or slower service.
- The data implies that **premium infrastructure like NBN does not guarantee loyalty**, and other factors such as pricing, reliability, or customer service may drive churn even among users of high-speed internet.
- Moreover, **internet stability may be a key issue**, as even customers with fast connections might experience frequent dropouts or inconsistent service quality, leading them to switch providers.

# 3. PAYMENT METHOD

Figure 7: PaperlessBilling by Churn

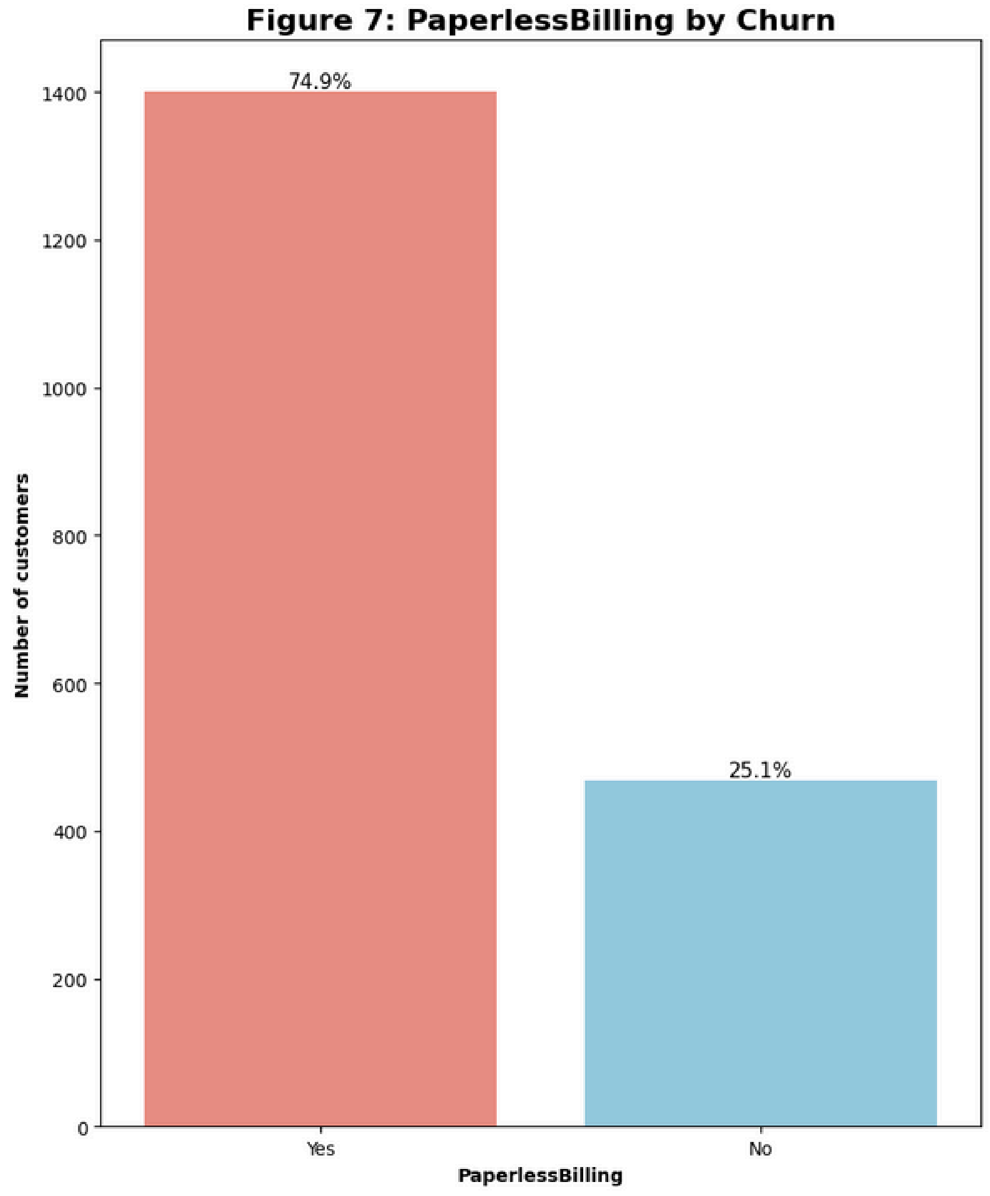
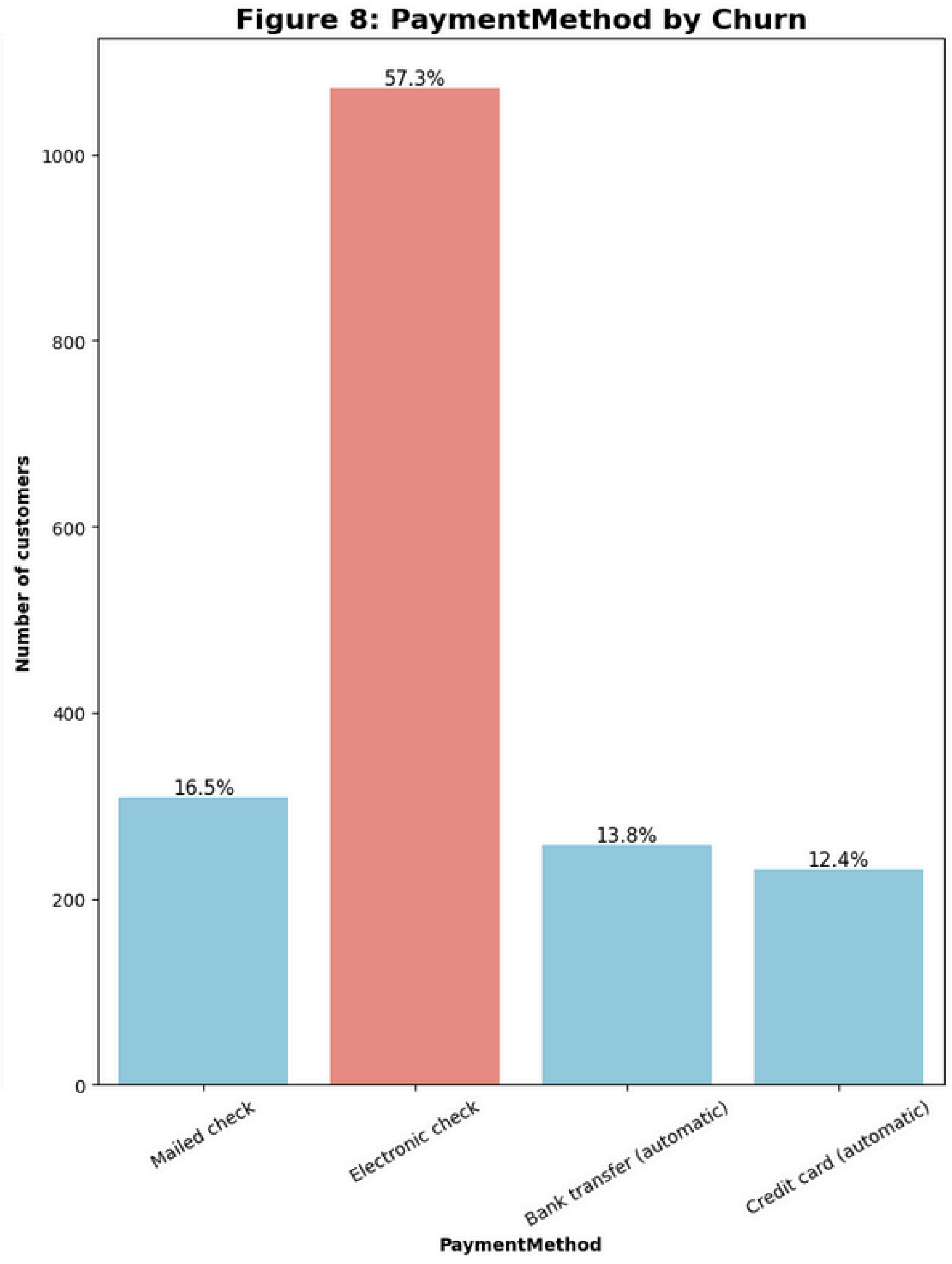



Figure 8: PaymentMethod by Churn




<u>+) Billing and payment method impacting churn</u>

- A **significant 74.9% of churned customers used paperless billing**, **suggesting that digitally engaged users may be more likely to switch**, possibly due to higher expectations or service awareness.

- **Electronic check users make up 57.3%** of churned customers, the highest among all payment methods, **indicating a strong link between this method and higher churn**.

- In contrast, automatic payment users—via bank transfer (13.8%) and credit card (12.4%)—exhibit much lower churn rates, hinting that payment convenience contribute to customer retention.

# 4. CONTRACT TYPE BREAKDOWN

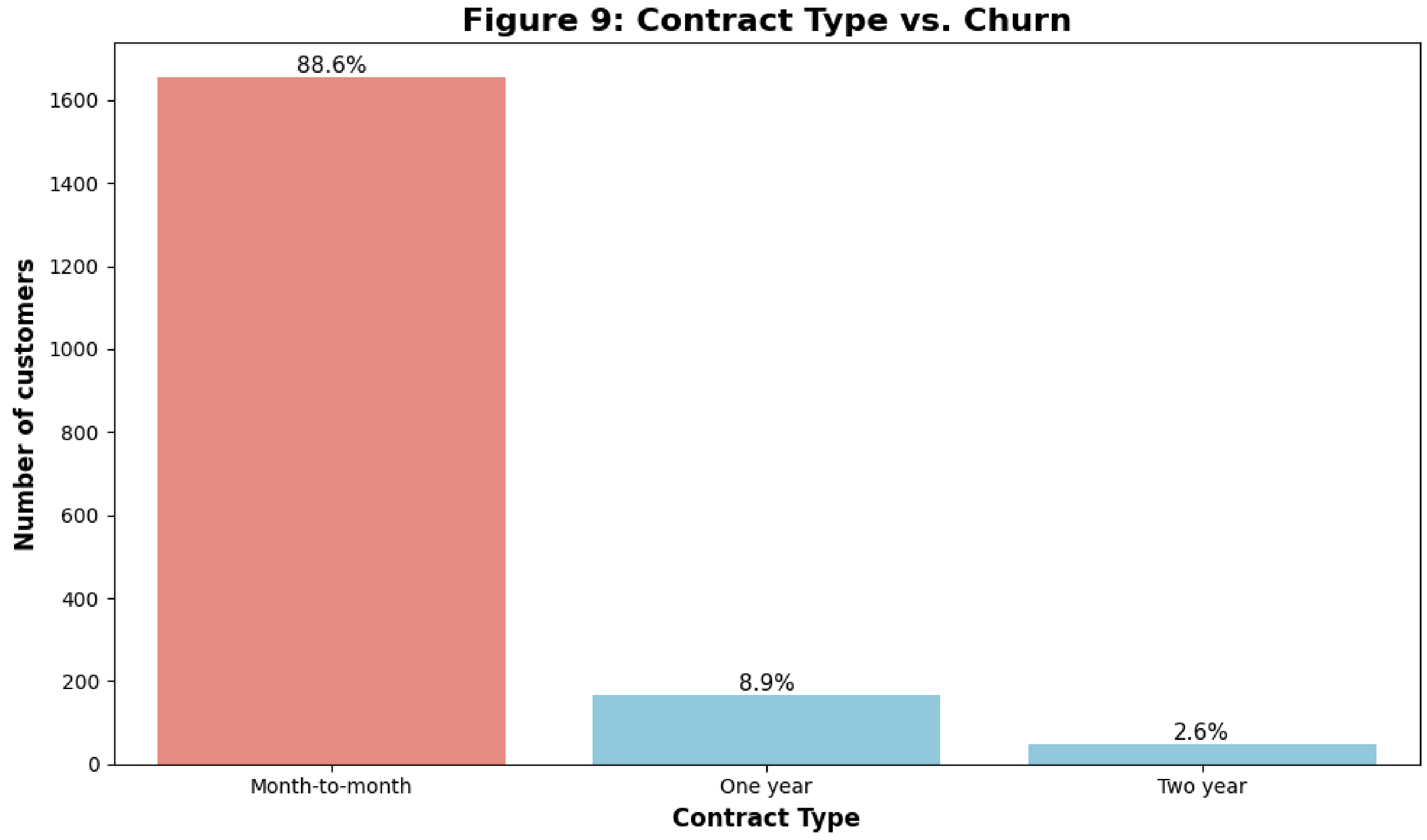
**The vast majority of churned customers (88.6%) are on month-to-month contracts**

- This group analysis indicating that **short-term agreements are strongly linked to higher churn rates**.

- In contrast, churn drops significantly among customers with longer commitments—only 8.9% have a one-year contract, and a minimal 2.6% are on two-year contracts.

- This trend suggests that customers with long-term contracts are more stable and less likely to leave, likely due to higher switching costs or stronger service commitment.

Figure 9: Contract Type vs. Churn




# 5. CUSTOMER SUPPORT

Figure 10: Admin Tickets (Churned Customers)

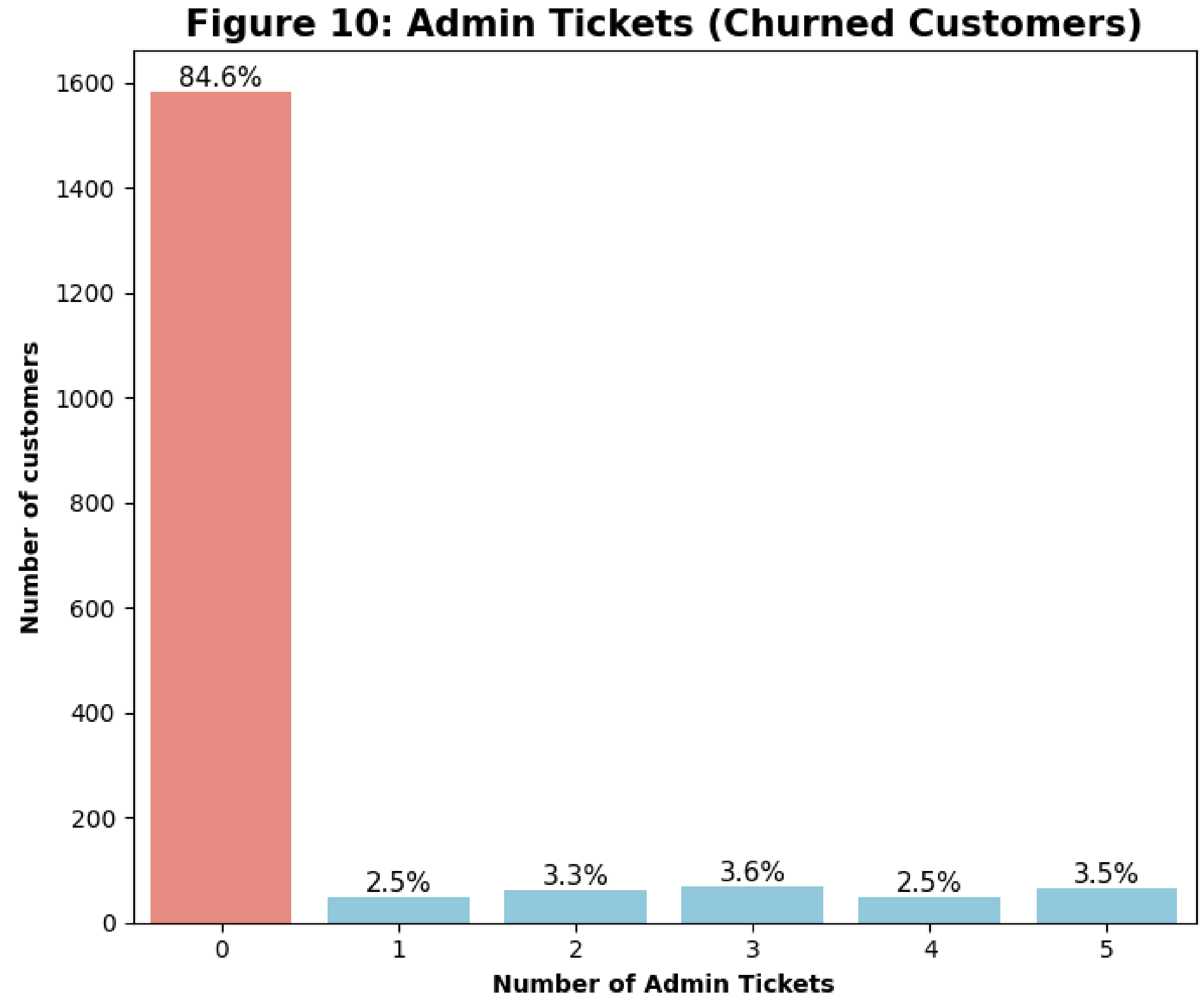
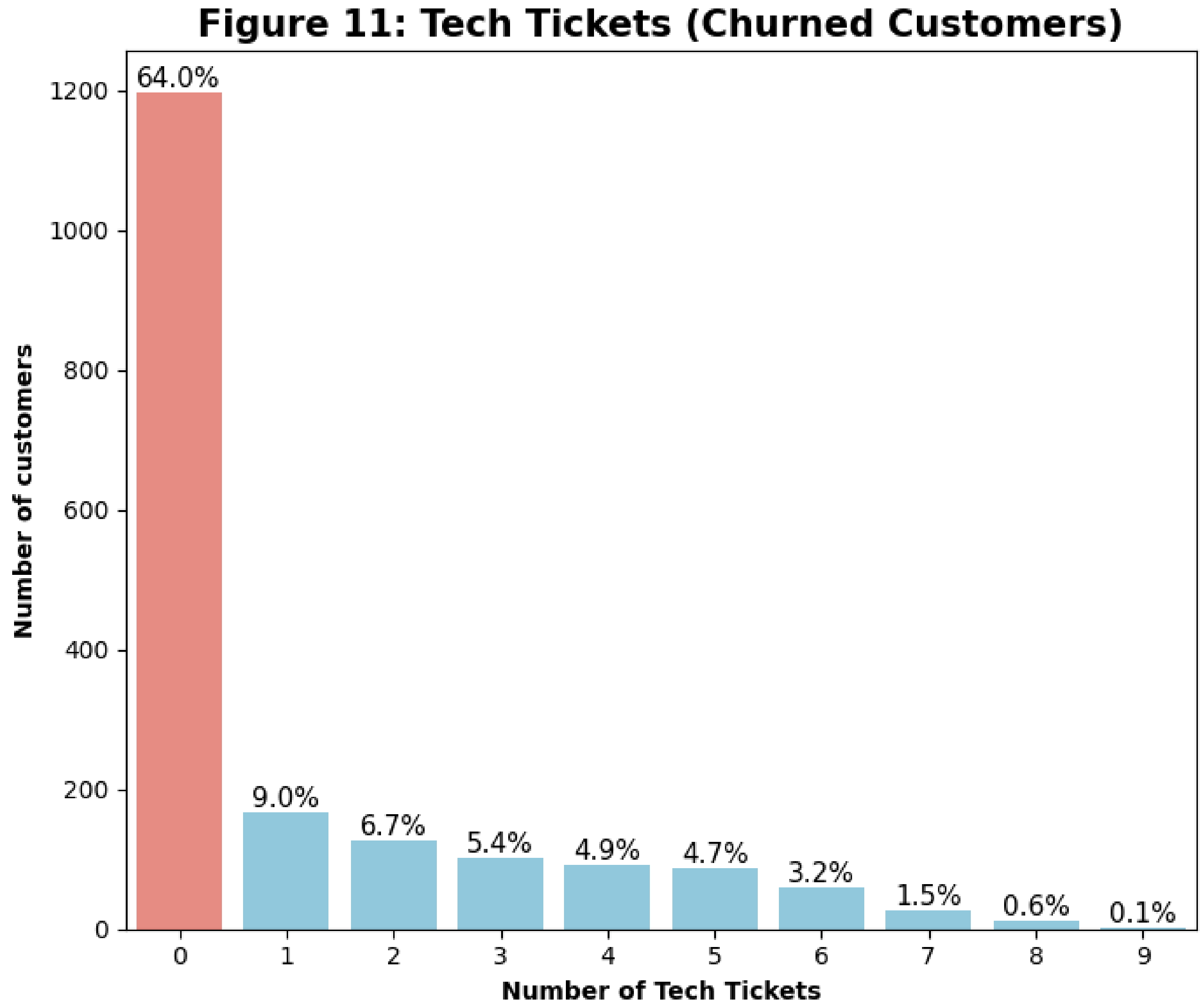



Figure 11: Tech Tickets (Churned Customers)

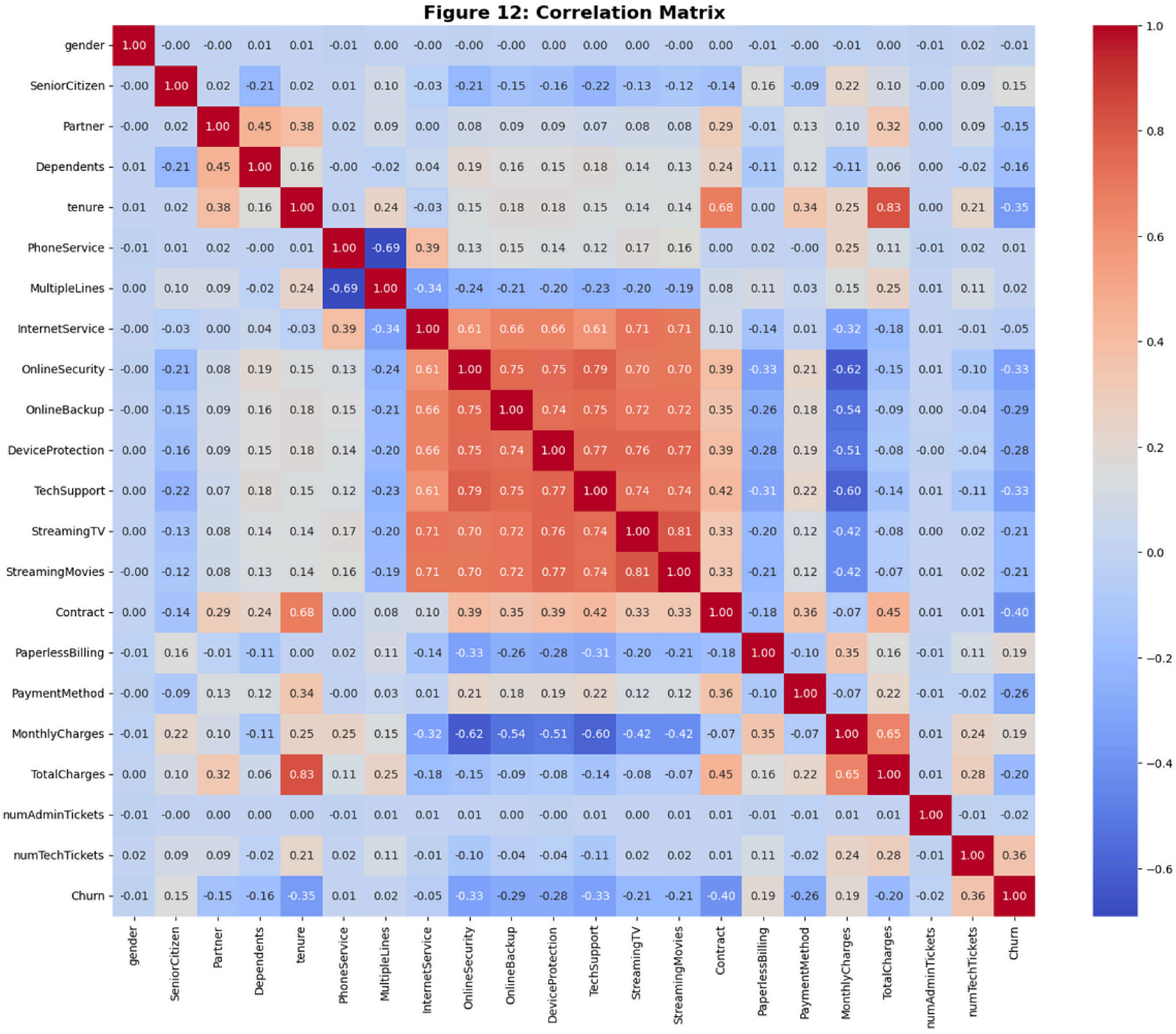



The analysis reveals that a **significant number of churned customers had minimal or no interaction with support services before leaving**. Specifically, 84.6% of churned users did not submit any admin tickets, and 66.0% did not raise any technical support tickets. This indicates that many customers either chose not to engage with support channels or possibly lacked confidence in their effectiveness. Among those who did seek technical help, most submitted only one or two tickets, showing limited interaction overall. These patterns suggest that **low support engagement may be an early sign of customer dissatisfaction, and that some users may prefer to churn rather than invest time in resolving their issues through existing support systems**.

# PREDICTIVE ANALYSIS

## 1. CORRELATION MATRIX

Figure 12: Correlation Matrix




- Contract (-0.39), TechSupport (-0.35), and OnlineSecurity (-0.35) show the strongest negative correlations with churn, meaning customers with longer contracts or using support services are less likely to churn.

- PaperlessBilling (0.19) and MonthlyCharges (0.19) have positive correlations, indicating a slightly higher likelihood of churn among customers with digital billing and higher bills.

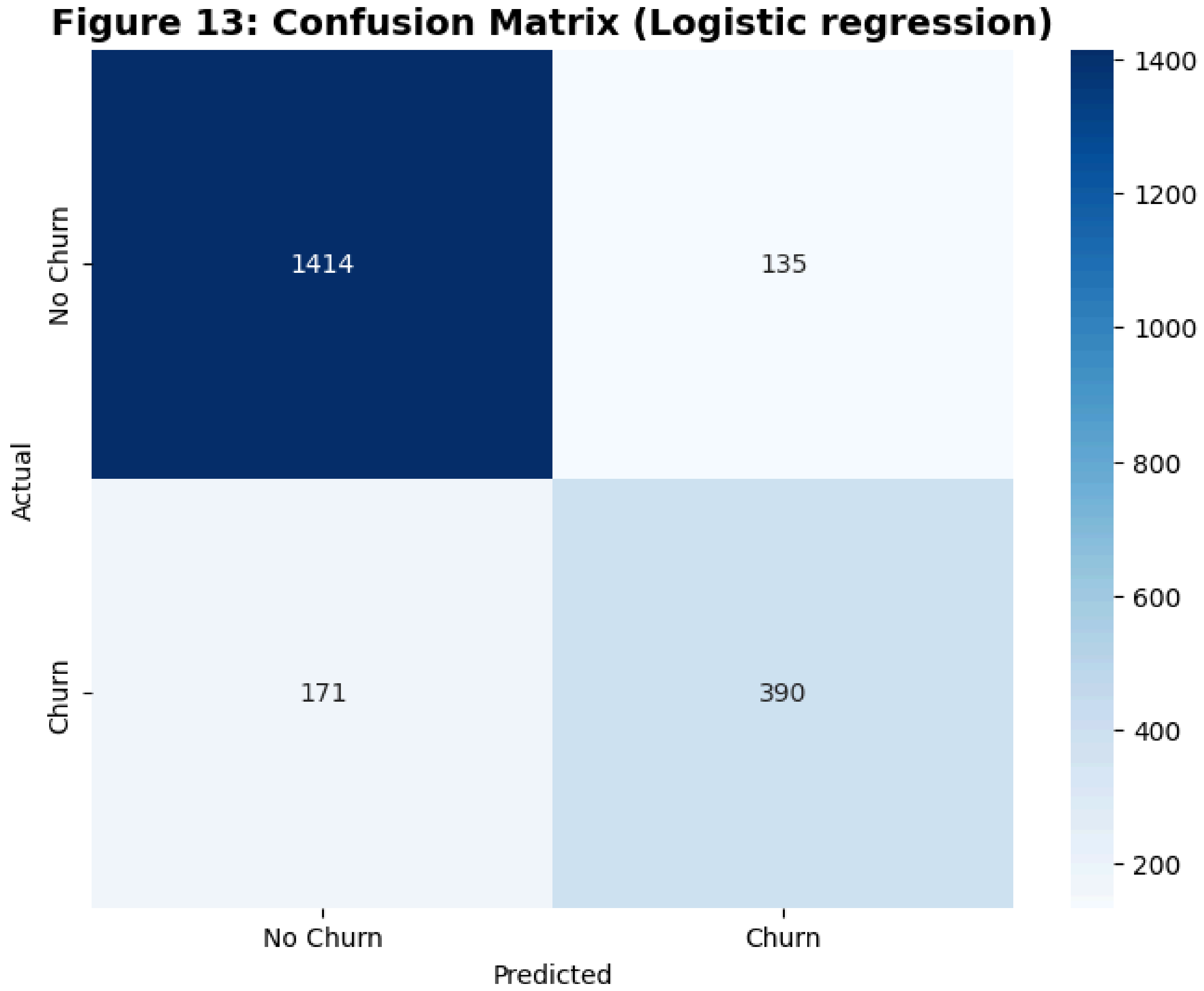
- StreamingTV (0.14) and StreamingMovies (0.13) also show weak positive correlations, suggesting customers using these services are a bit more prone to churn.

- Most demographic features like gender (0.01), PhoneService (0.01), and SeniorCitizen (0.15) have minimal correlation, implying limited impact on churn behavior.

★ **<u>Key data insights:</u>** While several variables show some relationship with churn, most correlations are weak, indicating that no single factor strongly predicts churn on its own, and churn is likely influenced by a combination of multiple factors.

## 2. LOGISTIC REGRESSION MODEL

**Figure 13: Confusion Matrix (Logistic regression)**



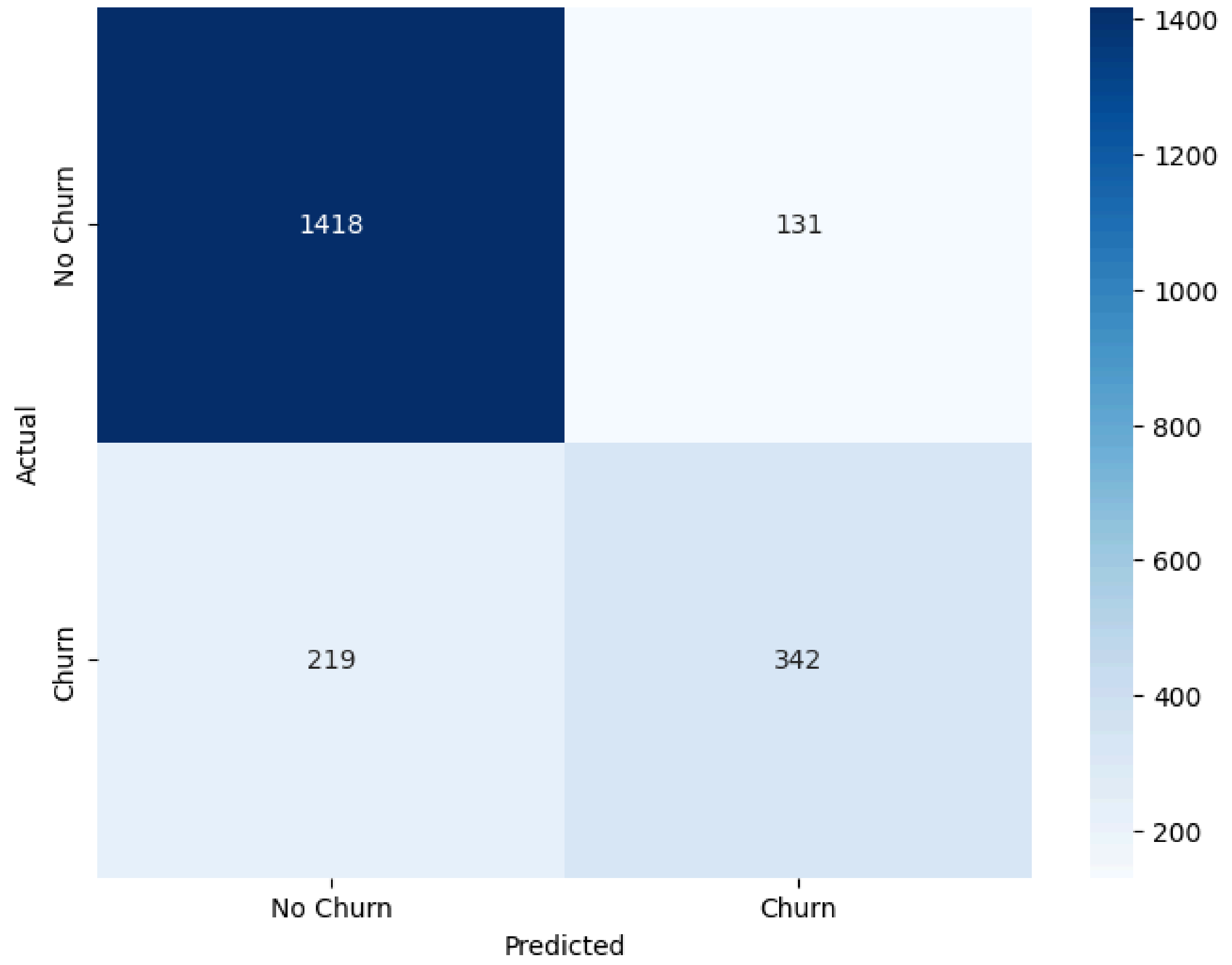

| | Precision | Recall | f1-score | Support |
|---|---|---|---|---|
| **0** | 0.89 | 0.91 | 0.90 | 1549 |
| **1** | 0.74 | 0.68 | 0.71 | 561 |
| Accuracy: 85 | | | | |

**<u>The performance of logistic regression on Test set ( 30% of dataset)</u>**

The logistic regression model shows **a strong overall accuracy of 85%**, indicating that it correctly predicts customer churn status in a majority of cases. However, diving deeper into key performance metrics reveals areas for refinement. **The positive predictive value (precision) for churners is 0.74**, meaning that when the model predicts a customer will churn, **it's correct 74% of the time**—suggesting relatively good confidence in churn predictions. On the other hand, **the true positive rate (recall) for churn is 0.68**, which means the **model successfully identifies 68%** of actual churners but misses 32% of them, potentially leading to lost opportunities for customer retention.

# 3. TUNED DECISION TREE

**Figrure 14: Confusion Matrix (Tuned Decision Tree)**

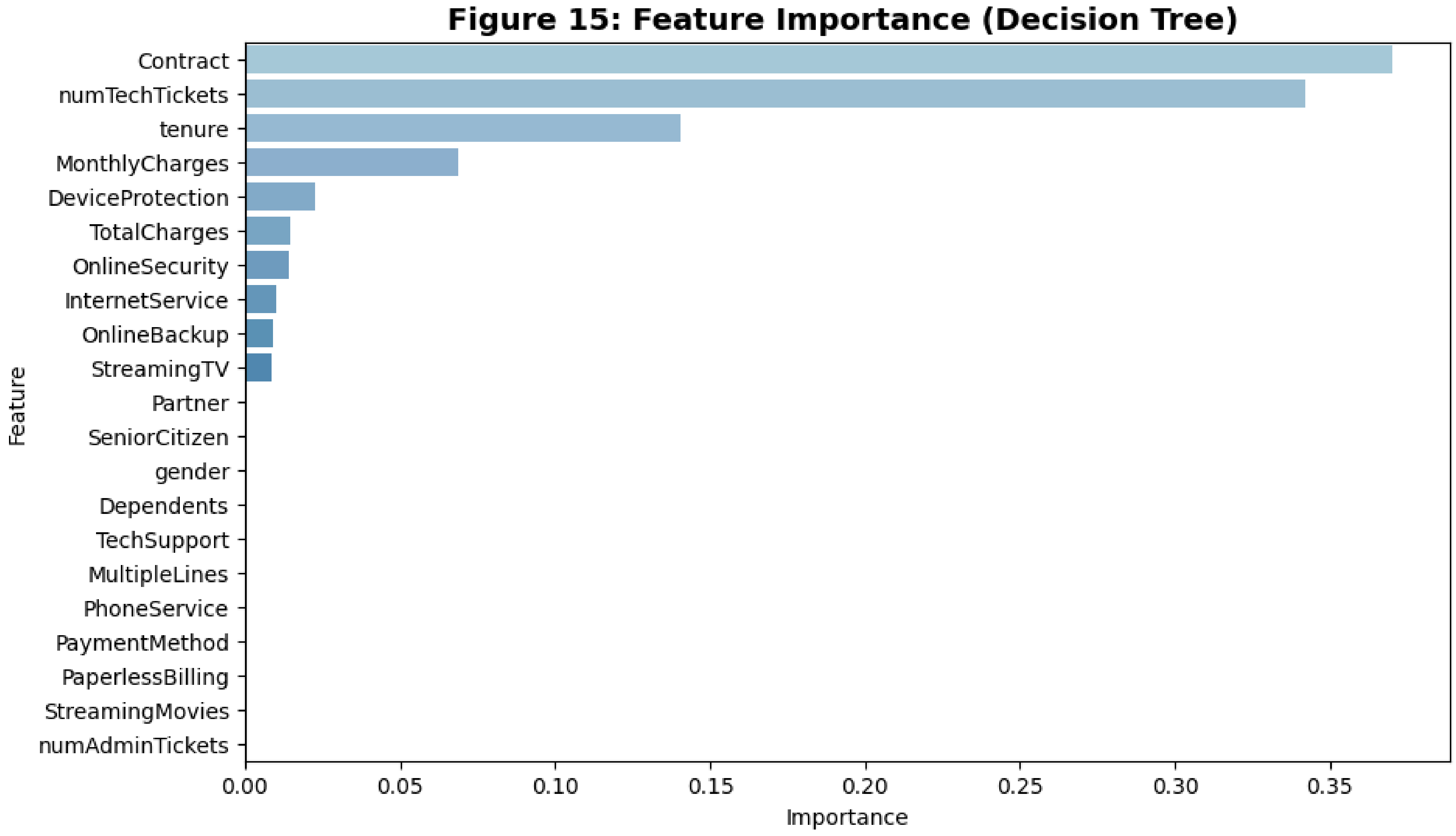



| | Precision | Recall | f1-score | Support |
|---|---|---|---|---|
| **0** | 0.87 | 0.92 | 0.89 | 1549 |
| **1** | 0.72 | 0.61 | 0.66 | 561 |
| Accuracy: 83 | | | | |

## The performance of Tuned Decision Tree on Test set ( 30% of dataset)

The tuned decision tree **achieves an overall accuracy of 83%**, marking a solid improvement over the baseline model and demonstrating that even simple models can perform well when optimized. **The model's precision for churn is 0.72, meaning that 72% of customers predicted to churn actually did** — indicating good reliability in its positive predictions. Its **recall for churn reaches 0.61, showing it correctly identifies over half of the actual churners**. While not as high as more complex models like XGBoost, this tuned tree offers a valuable trade-off: improved interpretability and low false positive rates, making it a practical choice for business teams prioritizing transparency and targeted retention actions.

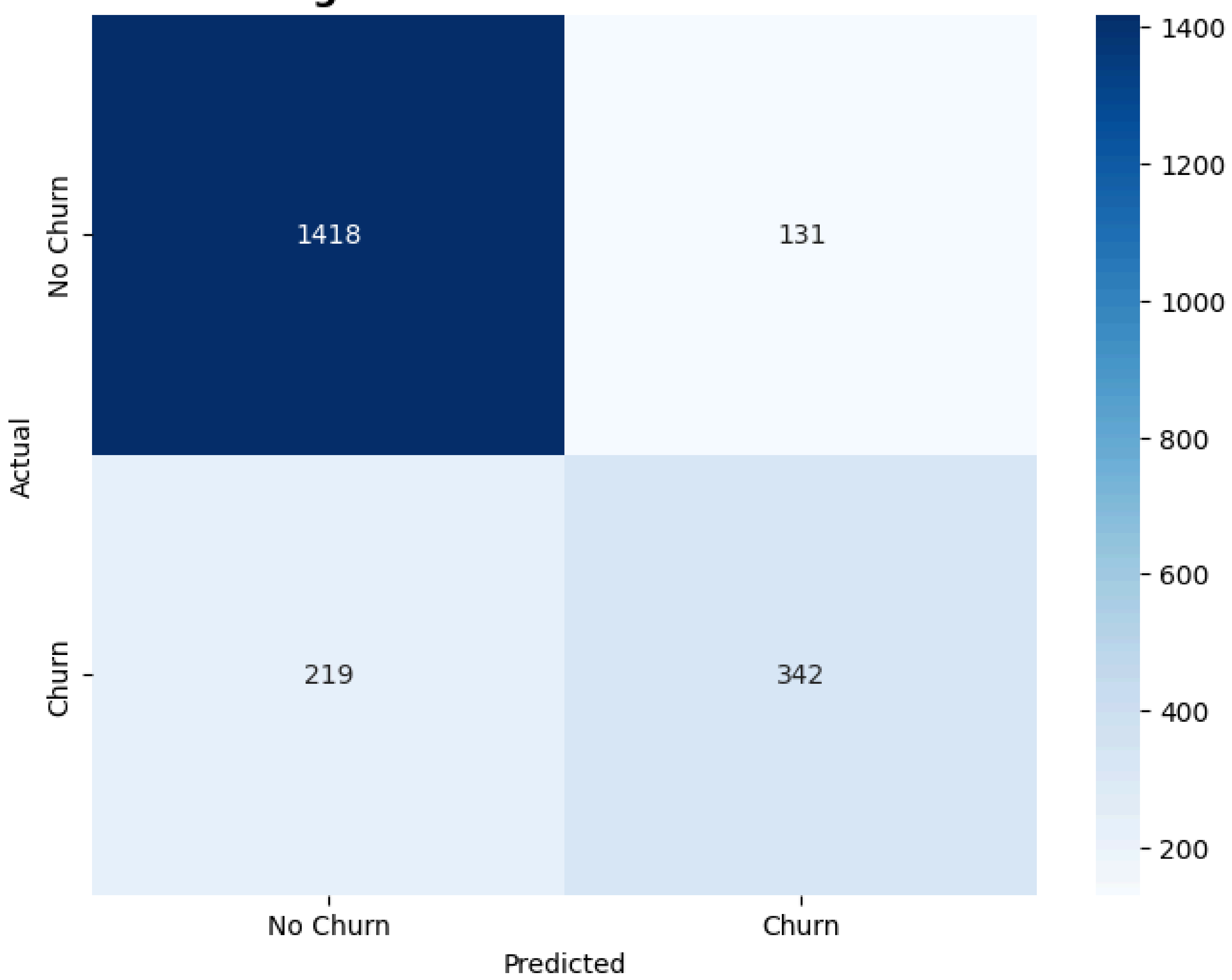
**Figure 15: Feature Importance (Decision Tree)**




The tuned Decision Tree model highlights **Contract type as the strongest indicator of churn**, showing that **customers on month-to-month plans are significantly more likely to cancel** compared to those on longer-term commitments such as one-year or two-year contracts. This emphasizes the value of encouraging contractual stability to reduce churn risk. The second most important factor is the **number of technical support tickets**, where a lower count is surprisingly associated with higher churn. This suggests that **customers who experience issues but choose not to seek help are more likely to leave**, possibly due to frustration, lack of trust in support, or disengagement from the service. In addition, tenure plays a key role—**newer customers are more prone to churn**, reinforcing the importance of a strong onboarding experience and proactive engagement during the early stages of the customer lifecycle. **Together, these findings point to churn being driven by lack of long-term commitment, disengagement, and missed opportunities to resolve pain points**.

# 4. TUNED XG BOOST

Figrue 16: Confusion Matrix

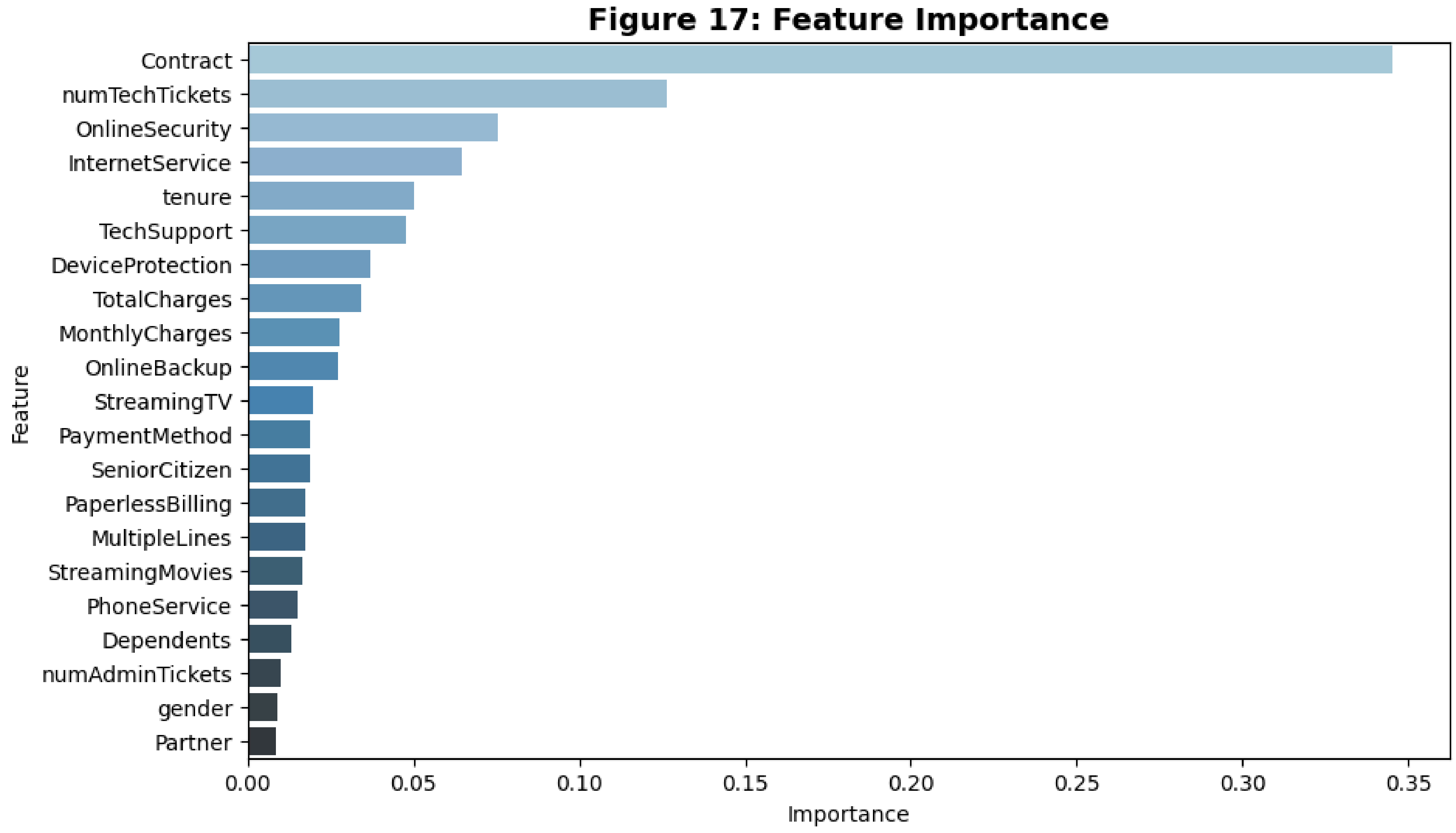



| | Precision | Recall | f1-score | Support |
|---|---|---|---|---|
| **0** | 0.89 | 0.91 | 0.90 | 1549 |
| **1** | 0.74 | 0.70 | 0.72 | 561 |
| Accuracy: 85 | | | | |

**The performance of Tuned XG Boost on Test set ( 30% of dataset)**

The model achieves an **overall accuracy of 85%**, consistent with the previous logistic regression model and decision tree, indicating strong overall performance. However, XGBoost shows notable gains in recall for churners and slightly better precision. The **positive predictive value (precision) for the churn class is 0.74, meaning 74% of customers predicted to churn actually did**—reflecting solid reliability in churn predictions. More importantly, the **true positive rate (recall) for churn has increased to 0.70**, suggesting that **the model now captures more actual churners while reducing the number of false negatives**. This enhancement reflects the power of XGBoost and **the benefits of tuning in handling complex patterns and class imbalance more effectively**. With these improvements, the model becomes more actionable for churn prevention strategies, catching more at-risk customers without a major sacrifice in precision.

Figure 17: Feature Importance




The Decision Tree model identifies **Contract type as the most critical predictor of churn**, contributing more to the model's decision-making than any other variable. Customers on **month-to-month contracts face a much higher likelihood of churn** than those with longer-term agreements, highlighting the importance of promoting contractual commitment to improve retention. The second most influential factor is **the number of technical support tickets**, but unlike traditional assumptions, the model suggests that **customers who submit fewer or no tickets are more likely to churn**. This may imply that these customers either silently experience issues without seeking help or feel disconnected from available support channels. The third key factor is **tenure, where shorter customer lifespans correlate strongly with higher churn risk**, reinforcing the need to strengthen early-stage engagement.

★ **Key data insights:** These findings reveal that churn is heavily influenced by contract flexibility, passive dissatisfaction, and customer lifecycle stage rather than basic demographics. Companies aiming to reduce churn should focus on promoting longer-term contracts, encouraging support interaction, and investing in personalized onboarding strategies to retain new and silent-risk customers.

## 5. COMPARE MODELS

| | Accuracy | Precision | Recall | f1-score |
|---|---|---|---|---|
| **Logistic Regression Model** | 0.85 | 0.74 | 0.68 | 0.71 |
| **Tuned Decision Tree** | 0.83 | 0.72 | 0.61 | 0.66 |
| **Tuned XG Boost** | **0.85** | **0.74** | **0.70** | **0.72** |

## Why Tuned XG Boost win ?

**85% Accuacy**

- Tuned XGBoost achieves the **same accuracy as Logistic Regression (85%) and outperforms the Tuned Decision Tree (83%)**. This shows that XGBoost is just as reliable at getting predictions right overall — with no compromise in general performance.

**74% PPV**

- All three models—**XGBoost, Logistic Regression, and Decision Tree**—**deliver similar precision when identifying churners**. However, **XGBoost matches the highest score** (74%), ensuring that when it predicts a customer will churn, it's just as likely to be correct while avoiding false alarms.

**70% TPR**

- This is **where Tuned XGBoost clearly outperforms**. With a recall of 70%, it **catches more real churners than Logistic Regression (68%) and significantly more than Decision Tree (61%)**. This means fewer missed opportunities to save at-risk customers.

**72% f1-score**

- XGBoost achieves the **highest F1-score** (0.72), reflecting the best trade-off between precision and recall. **Compared to Logistic Regression (0.71) and Decision Tree (0.66)**, XGBoost is the most balanced model for churn detection.

## ★ Key data insights

Tuned **XGBoost** stands out by **developing a series of decision trees where each new tree is built to correct the errors made by the previous ones, resulting in a model that becomes progressively stronger with each iteration**. Unlike a single Decision Tree that learns in isolation and may overfit or underperform, XGBoost uses boosting to combine many shallow trees, each focusing on the mistakes of the last, ultimately creating a more accurate and resilient ensemble. This process **allows XGBoost to capture complex patterns that Logistic Regression might miss and address the weaknesses of traditional Decision Trees, such as low recall**. With its ability to refine itself through each round of learning, **Tuned XGBoost achieves the highest balance of performance**—matching Logistic Regression in precision (0.74), significantly improving recall (0.70), and delivering the best F1-score (0.72)—**making it the most robust and adaptive model for predicting customer churn**.

# RECOMMENDATION

### +) Engage and retain new customers (Tenure: 0-10 months)

- Create a personalized onboarding program to educate customers about service features and usage tips within the first month.
- Set up automated follow-ups during the first 3, 6, and 9 months to address concerns before dissatisfaction sets in.
- Offer exclusive retention deals or service upgrades at the 6-month mark to encourage continued engagement.

### +) Address Premium Tier User Dissatisfaction

- Conduct surveys or interviews to understand what premium users expect versus what they're receiving.
- Enhance premium plans by bundling exclusive content, better bandwidth, or priority support.
- Highlight premium plan benefits clearly and continuously in customer communications (e.g., emails, app notifications).

### +) Improve Customer Support Engagement

- Deploy in-app or messaging when users show signs of struggle (ex: service drop, complaints) to encourage ticket submission.
- Enhance training to ensure that support teams resolve issues quickly and empathetically, restoring customer trust.
- Monitor and analyze ticket logs to identify recurring pain points and proactively solve them across the user base.
- Set alerts for customers who never reach out to support; flag them for personalized re-engagement efforts.

### +) Incentivize Longer-Term Contracts

- Offer discounts or value-added services (ex: free month, extra bandwidth) for customers switching to 12- or 24-month plans.
- Use tenure and usage data to identify customers likely to churn and offer them personalized long-term deals.
- Educate customers on the cost-saving and benefit advantages of long-term plans during onboarding and billing cycles.