Customer Churn Analysis: Exploring Key Drivers of Attrition and Predictive Model Insights

An Interactive Visual Exploration of Factors Influencing Customer Churn, Behavioral Patterns, and Model Performance Using Statistical and Machine Learning Techniques

Shayma Remy 8/14/25 Springboard | August 2025 – Capstone 3

Table of Contents

Executive Summary

Customer churn represents one of the most critical challenges facing subscription-based businesses, with customer acquisition costs often exceeding five times the cost of retention. This comprehensive analysis examines customer churn patterns within a telecommunications dataset to identify key drivers and develop predictive models for proactive retention strategies.

Through advanced statistical analysis and machine learning techniques, **contract type emerges as the most significant predictor of customer churn, accounting for nearly 80% of the model's predictive power**. Payment method preferences, demographic factors, and service utilization patterns also demonstrate strong correlations with churn behavior. Two predictive models—**Logistic Regression and Decision Tree**—were developed, with the Logistic Regression model achieving superior performance with an **AUC of 0.836**.

The analysis provides actionable insights for reducing churn rates through targeted interventions, including incentivizing long-term contracts, promoting automatic payment methods, and addressing service quality issues that drive customer dissatisfaction.

Problem Statement and Business Context

Subscription businesses face significant revenue losses due to customer churn, with traditional manual approaches often failing to identify at-risk customer segments and their underlying drivers. This leads to inefficient retention campaigns that waste resources on low-risk customers while missing high-risk segments entirely.

This analysis addresses the critical need for data-driven churn prediction by developing an end-to-end analytical solution that identifies key drivers of customer attrition, builds predictive models capable of identifying at-risk customers, and provides actionable insights through interactive visualizations. Every percentage point reduction in monthly churn rates can translate into millions of dollars in retained subscription revenue, enabling retention managers to transition from reactive to proactive customer retention strategies.

Data Overview and Methodology

Dataset Description: The analysis utilizes the Telco Customer Churn Dataset from Kaggle, containing 7,043 unique customer records with 21 columns including demographics (gender, age, family structure), service information (internet types, phone services, streaming subscriptions), account data (contract terms, payment methods, billing preferences), and financial metrics (tenure, monthly charges, total charges).

Analytical Approach: The methodology follows a comprehensive data science workflow emphasizing interpretable machine learning techniques. After thorough data cleaning and feature engineering, two complementary algorithms were implemented: Logistic Regression for clear coefficient interpretations and probability estimates, and Decision Trees to capture non-linear relationships and feature interactions. The approach prioritizes practical business implementation over pure predictive performance.

Tools and Technologies: Python served as the primary programming language with Pandas, NumPy, and Scikit-learn for data processing and machine learning. Tableau Public provided interactive dashboard development, while Matplotlib and Seaborn supported exploratory data analysis.

Key Findings: Understanding Customer Behavior

Demographic and Relationship Factors

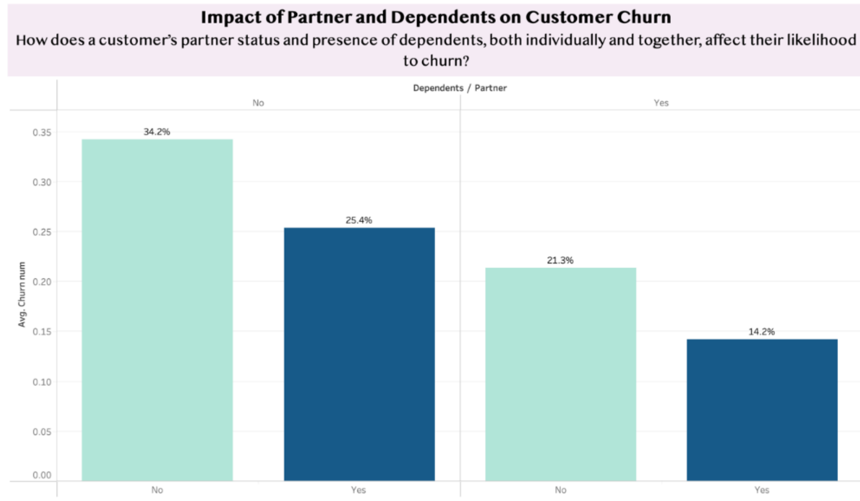

Figure 1: Partner and Dependents Impact on Churn Dashboard

<u>Family Structure Impact</u>

Customers with neither partners nor dependents exhibit the highest churn rate at **34.2%.** The presence of a partner reduces churn to 25.4%, while customers with both partners and dependents demonstrate the lowest churn rate at just **14.2%**. This suggests family responsibilities create stronger ties to service providers due to inconvenience and costs of service disruptions for multiple household members.

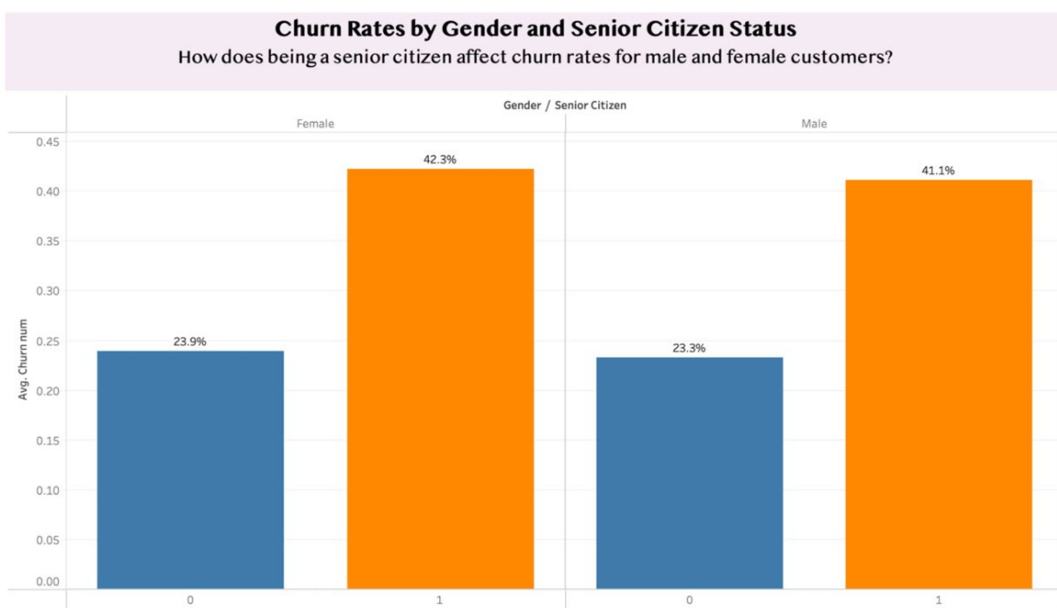

Figure 2: Age and Gender Demographics Dashboard

Age and Gender Dynamics

Senior citizens demonstrate dramatically higher churn rates **(~42%)** compared to non-senior customers (~24%), representing a nearly 75% increase in attrition risk. Gender shows minimal independent effect on churn rates.

Financial and Payment Behavior

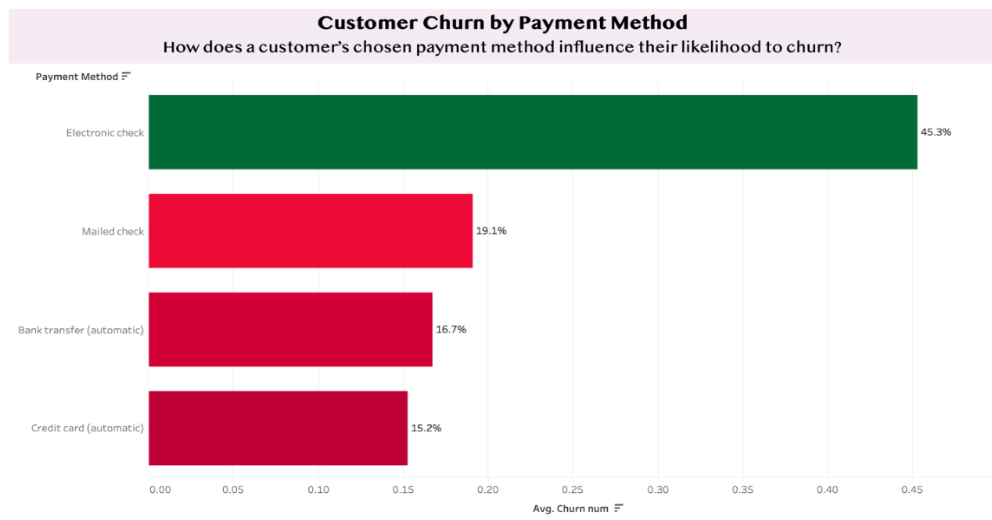

Figure 3: Payment Method Churn Analysis Dashboard

Payment Method as Key Predictor

Electronic check users exhibit an alarming **45.3% churn rate**, nearly three times higher than automatic payment users. The progression shows electronic check (45.3%), mailed check (19.1%), bank transfer automatic (16.7%), and credit card automatic (15.2%). This pattern reflects convenience factors and customer engagement levels, where automatic payment setup demonstrates higher service commitment.

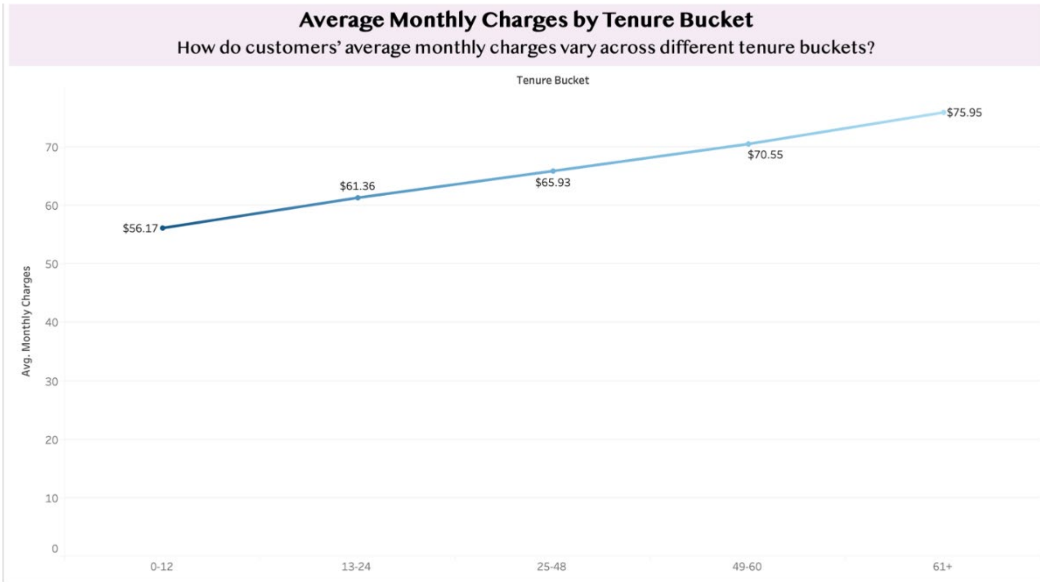

Figure 4: Tenure and Monthly Charges Relationship Dashboard

Tenure and Pricing Paradox

Customer tenure shows the expected inverse relationship with churn risk. However, longer-tenured customers pay higher monthly charges on average, ranging from $56.17 for 0-12 months tenure to $75.95 for 61+ months, suggesting either service upgrades over time or price increases that loyal customers accept.

Contract and Service Factors

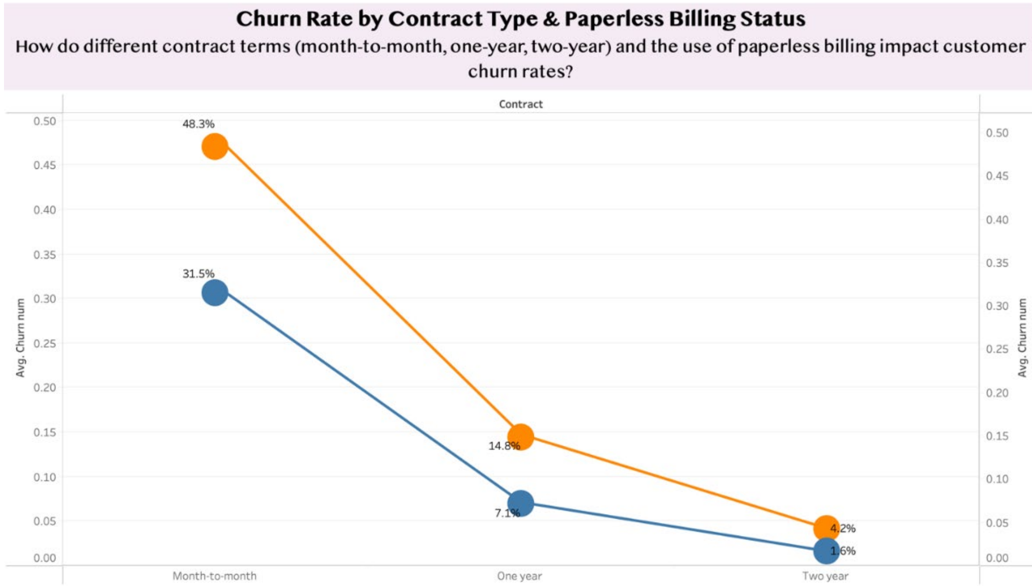

Figure 5: Contract Type and Paperless Billing Dashboard

Contract Type as the Primary Churn Driver

Contract terms demonstrate the strongest relationship with churn behavior. Month-to-month customers show dramatically higher churn rates (31.5% to 48.3%) compared to one-year contract holders (7.1% to 14.8%) and two-year contract customers (1.6% to 4.2%). Additionally, paperless billing consistently correlates with higher churn rates across all contract types, showing approximately 1.5 times higher churn rates.

Predictive Model Development and Performance

Model Architecture and Performance

The analysis implements two complementary machine learning algorithms chosen for their interpretability and business applicability. Logistic regression serves as the primary modeling approach due to its ability to provide clear coefficient interpretations and probability estimates that directly translate to business risk assessments. This linear approach enables stakeholders to understand exactly how each factor contributes to churn probability, facilitating targeted intervention strategies. The decision tree algorithm provides a secondary perspective that captures non-linear relationships and feature interactions that might be missed by linear approaches.

Both models undergo rigorous feature engineering processes including categorical variable encoding, numerical feature scaling, and creation of derived business variables such as charges ratios and tenure groupings. The modeling approach prioritizes practical business implementation over pure predictive performance, ensuring that model outputs can be readily integrated into existing customer relationship management systems and retention workflows.

Model Performance Analysis

Confusion Matrix - Logistic Regression Model

Actual / LogReg Predicted	0 (No Churn)	1 (Churn)
0 (No Churn)	739	296
1 (Churn)	73	301

Confusion Matrix - Decision Tree Model

Actual / Tree Predicted	0 (No Churn)	1 (Churn)
0 (No Churn)	591	444
1 (Churn)	45	329

Figure 6: Confusion Matrix Comparison for Both Models

Logistic Regression Results

The Logistic Regression model demonstrates exceptional performance with an ROC-AUC of 0.836, indicating strong discriminative ability between customers likely to churn and those likely to remain. The confusion matrix analysis reveals 301 true negatives and 739 true positives, alongside 296 false positives and 73 false negatives, reflecting the model's ability to identify actual churners with a recall rate of 91.0%. While the model generates some false positives, the high recall rate proves particularly valuable for business applications where early identification of potential churners enables proactive intervention strategies.

The model's probability scoring capability provides additional business value by enabling risk-based customer segmentation and tiered retention strategies. Rather than treating all customers equally, organizations can allocate retention resources proportionally based on individual churn probability scores, optimizing both intervention effectiveness and cost efficiency.

Decision Tree Performance

The Decision Tree model exhibits different performance characteristics, showing excellent recall capabilities at 92.9% in identifying potential churners while generating significantly more false positives than the Logistic Regression approach. The confusion matrix displays 329 true negatives and 591 true positives but also reveals 444 false positives and 45 false negatives, resulting in lower precision and reduced overall accuracy compared to the logistic regression model. This performance profile suggests that the decision tree adopts a more aggressive churn prediction strategy, prioritizing the identification of all potential churners even at the cost of increased false alarms. While this approach ensures minimal missed churn opportunities, the higher false positive rate may lead to inefficient allocation of retention resources and potentially unnecessary customer interventions.

Model Comparison and Selection

While both models demonstrate high recall in identifying potential churners, the Logistic Regression model provides better overall balance between precision and recall, making it more suitable for business implementation where false positives incur retention costs, but the primary goal remains identifying as many potential churners as possible for proactive intervention.

Feature Importance Analysis

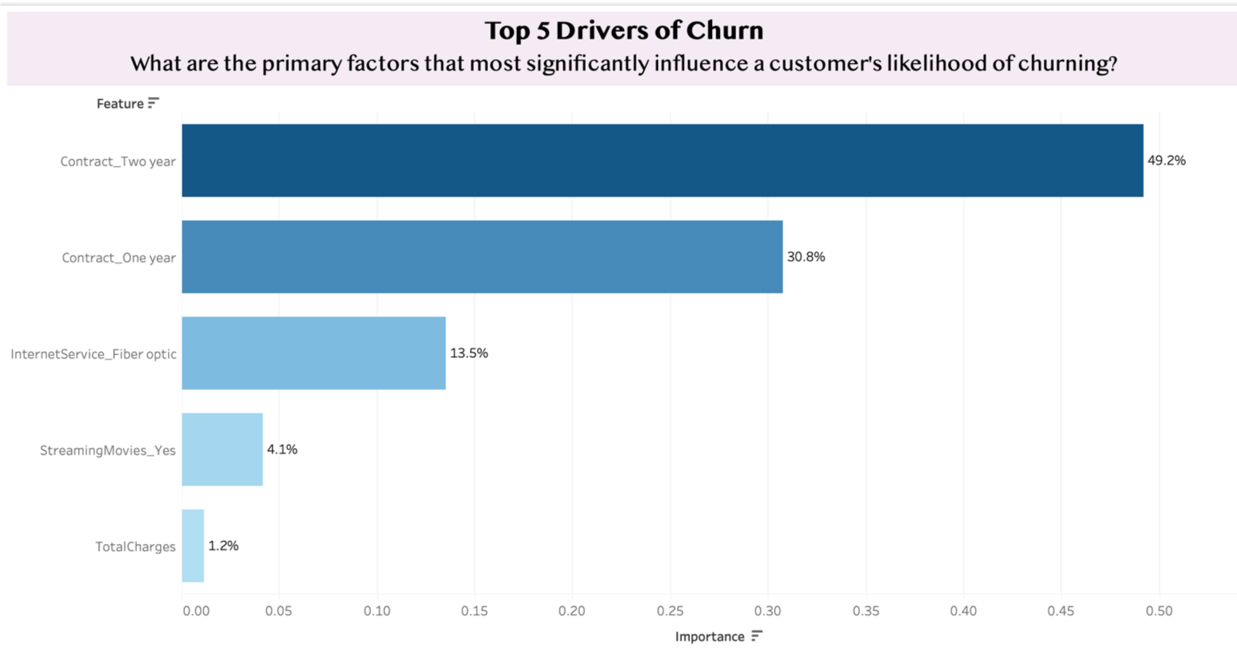

Figure 7: Feature Importance Analysis

The feature importance analysis reveals clear hierarchies in the factors driving customer churn decisions, providing actionable insights for retention strategy development. Contract terms emerge as the dominant predictive factors, with two-year contracts contributing 49.2% of the model's predictive power and one-year contracts adding another 30.8%. This combined 80% contribution from contract-related features demonstrates that customer commitment mechanisms far outweigh other factors in determining churn probability.

Fiber optic internet service appears as the third most important factor at 13.5% importance, suggesting that customers using premium fiber services exhibit higher churn tendencies despite typically paying higher rates. This counterintuitive finding indicates potential service quality issues, unmet performance expectations, or competitive pressures specifically affecting fiber customers. Streaming movie services contribute 4.1% to the model's predictive capability, while total charges account for merely 1.2%, reinforcing that service structure and contract factors significantly outweigh pure pricing considerations in churn prediction.

These insights fundamentally reshape traditional retention thinking by demonstrating that contract-based interventions and service quality improvements should take priority over price-based retention strategies. The minimal impact of total charges suggests that customers make churn decisions based on relationship factors and service satisfaction rather than cost optimization alone.

Interactive Dashboard and Business Intelligence

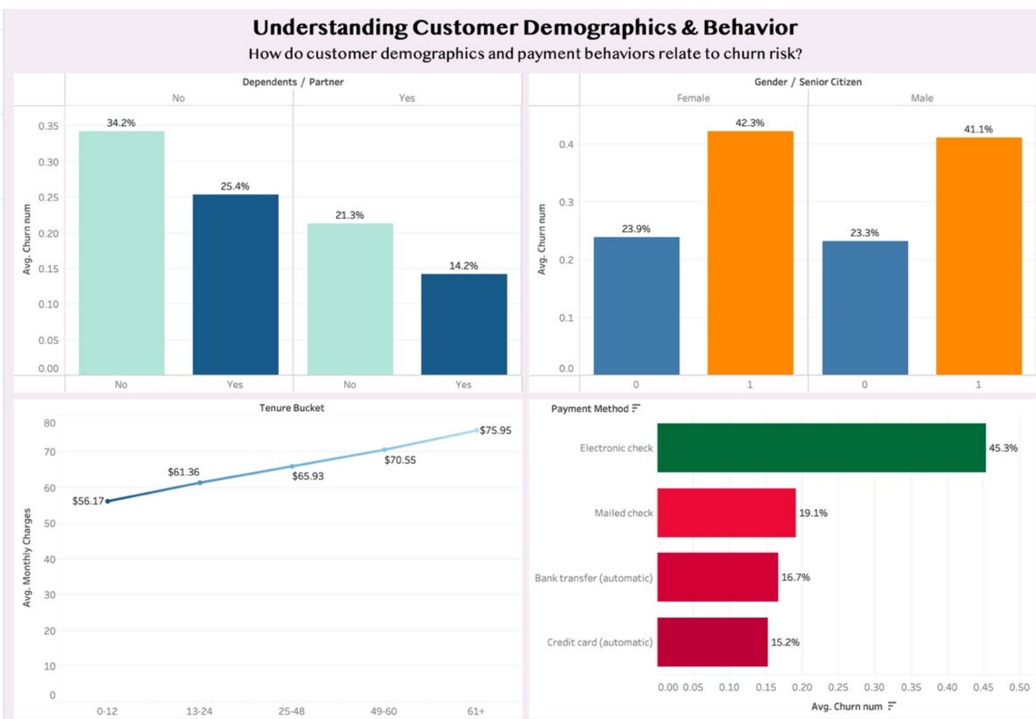

Figure 8: Customer Demographics and Churn Patterns Dashboard

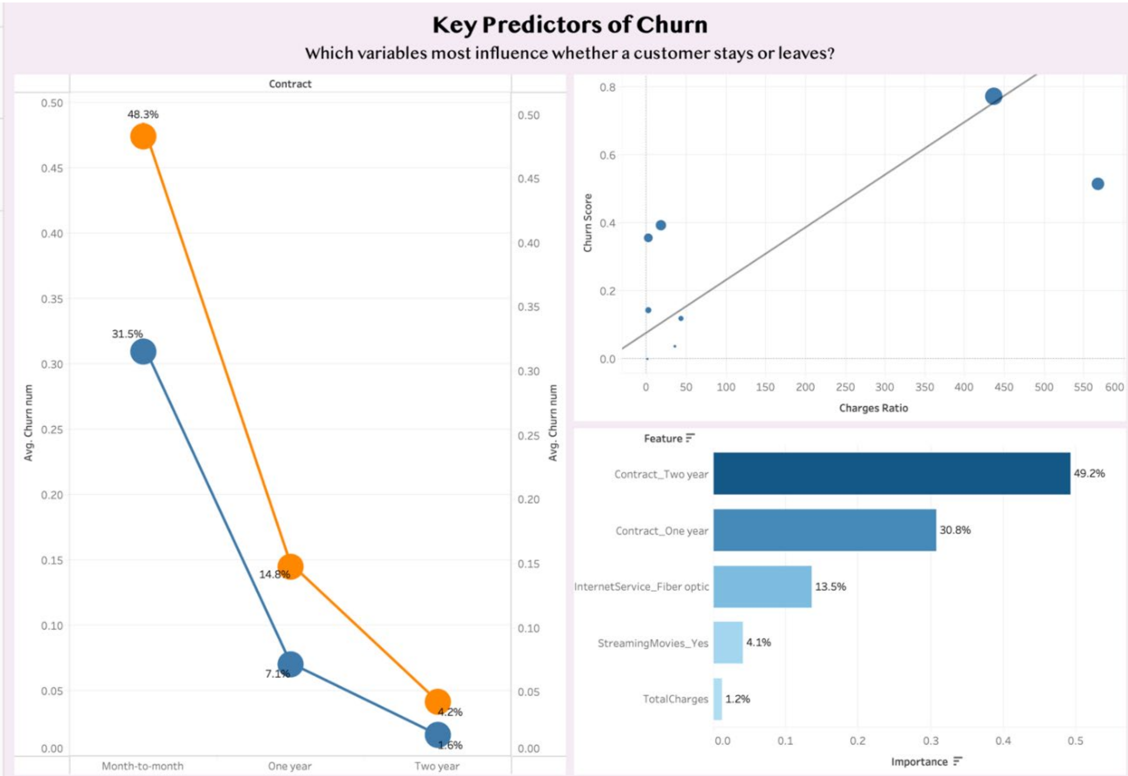

Figure 9: Payment and Contract Analysis Dashboard

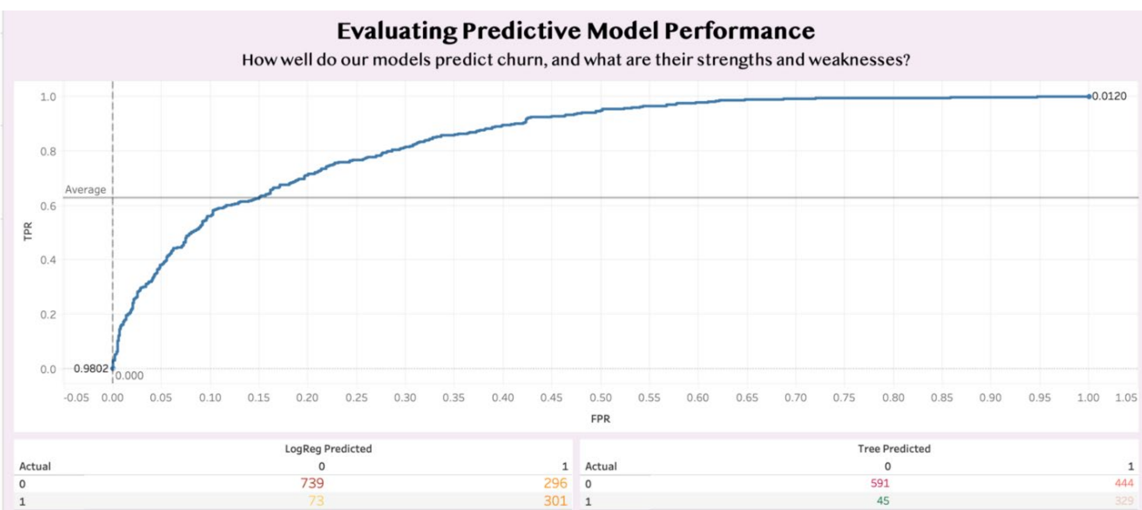

Figure 10: Model Performance and Feature Importance

The Tableau dashboard provides a comprehensive interactive environment for exploring customer churn patterns across multiple dimensions. Users can dynamically filter data by churn risk scores, tenure segments, contract types, and payment methods to identify specific customer cohorts requiring attention.

High-Risk Segment Identification

Through dashboard filtering, **month-to-month fiber customers using electronic check payments with less than 12 months tenure** emerge as the highest-risk segment, exhibiting **churn probabilities of 45-50%**. Additional vulnerable populations include senior citizens, customers without family ties, and paperless billing users.

Business Impact and Strategic Implications

Financial Impact

High recall rates enable efficient targeting of at-risk customers, reducing wasted retention spending. The dramatic difference in churn rates between contract types suggests significant ROI from contract incentive programs. Converting electronic check users to automatic payments could yield substantial retention improvements.

Operational Considerations

Higher churn among fiber internet users suggests potential service quality issues requiring investigation. Elevated churn rates among senior citizens indicate the need for targeted retention strategies, while the paperless billing paradox requires careful analysis of digital customer engagement strategies.

Limitations and Model Considerations

Data Limitations

The dataset represents simulated telecommunications data which may not capture all real-world complexities. The analysis provides a snapshot without considering seasonal variations, competitive landscape, economic conditions, or regulatory changes.

Model Limitations

The model is limited to available features and may miss important churn drivers. The Decision Tree model shows signs of potential overfitting with high false positive rates. Optimal threshold selection requires business context regarding relative costs of false positives versus false negatives.

Recommendations for Implementation

Immediate Action Items

Contract Incentive Programs: The analysis reveals contract terms as the highest-impact opportunity, given that they account for nearly 80% of the model's predictive power. Organizations should develop tiered discount structures that encourage customers to transition from month-to-month arrangements to annual or bi-annual contracts, offering 10-15% discounts on monthly rates for one-year commitments and 15-20% discounts for two-year commitments. These programs should include additional value-added perks such as free installation, equipment upgrades, or premium customer support to enhance the perceived value of longer-term commitments. Based on the analysis, this intervention could reduce churn rates from over 40% to under 15% for successfully converted customers, representing a dramatic improvement in retention performance.

Payment Method Optimization: The stark difference between electronic check users (45.3% churn rate) and automatic payment users (15-17% churn rate) indicates substantial potential for improvement through payment method migration programs. Organizations should implement incentive programs offering monthly bill credits of $5-10 for customers switching to automatic payment methods, combined with one-time signup bonuses to encourage enrollment. Simplifying the automatic payment setup process through guided assistance and removing friction from the enrollment experience will further enhance adoption rates. Converting electronic check users to automatic payments could reduce churn by approximately 60-65% among this particularly vulnerable segment.

Predictive Churn Scoring Implementation: Deploy the Logistic Regression model within existing CRM systems to generate real-time churn risk scores. This system should integrate model outputs with customer service workflows, enabling representatives to identify high-risk customers during routine interactions and proactively address concerns before they escalate to churn. Automated trigger systems should activate targeted retention programs based on churn probability thresholds, while tiered intervention strategies can be customized based on individual risk profiles and customer value. Proactive identification of at-risk customers through predictive scoring enables targeted interventions that can potentially reduce overall churn rates by 20-30% while optimizing the efficiency of retention spending.

Strategic Initiatives

Fiber Service Quality Improvement: Beyond immediate tactical interventions, the analysis reveals the need for strategic service quality improvements, particularly addressing the elevated churn rates observed among fiber internet service customers. A comprehensive fiber service satisfaction initiative should begin with targeted customer satisfaction surveys specifically designed for fiber users to understand pain points and service quality perceptions. Technical support ticket analysis can reveal patterns in service issues and identify systemic problems affecting customer experience. Based on these insights, organizations should implement targeted service quality improvements, which may include infrastructure upgrades, revised service level agreements, or adjusted pricing models that better reflect the value proposition for fiber services. Improving fiber service satisfaction represents an opportunity to reduce churn in this typically high-value customer segment while enhancing overall customer lifetime value and competitive positioning in premium service markets.

Future Research Directions

Advanced Modeling Opportunities

Future research should explore ensemble methods including random forest, gradient boosting, and neural network approaches that may capture more complex patterns and improve prediction accuracy beyond the current baseline. Time series analysis represents another valuable direction, incorporating temporal patterns and seasonality effects that could reveal cyclical churn behaviors and improve prediction timing. Customer lifetime value integration would enable the development of models that consider both churn probability and customer value simultaneously, allowing for optimized retention spending that prioritizes high-value customers while managing intervention costs effectively.

Additional Data Integration

Expanding the analytical foundation through additional data sources could significantly enhance model performance and business insights. Customer service interaction data including support tickets, call center logs, and satisfaction surveys would provide early warning signals of customer dissatisfaction before churn occurs. Actual service usage patterns and consumption metrics could reveal behavioral changes that precede churn decisions, enabling more precise intervention timing. Competitive intelligence data incorporating market conditions, competitor pricing, and promotional activities would help contextualize churn patterns within broader market dynamics and improve prediction accuracy during competitive pressures.

A/B Testing Framework

Developing controlled experimental frameworks would enable rigorous testing of retention strategy effectiveness and continuous improvement of intervention approaches. Contract incentive effectiveness testing could quantify the optimal discount levels and terms for different customer segments while measuring incremental retention lift. Payment method migration program experiments would determine the most effective incentive structures and communication strategies for encouraging automatic payment adoption. Targeted customer service intervention testing could optimize the timing, messaging, and channel preferences for different risk segments, while personalized pricing strategy experiments could explore dynamic pricing approaches that balance churn risk with revenue optimization objectives.

Conclusion

This comprehensive customer churn analysis successfully identified key drivers of customer attrition and developed predictive models capable of supporting proactive retention strategies. **Contract terms emerge as the dominant factor, accounting for nearly 80% of predictive power**, while payment methods and demographic factors play crucial supporting roles.

The **Logistic Regression model (AUC: 0.836)** provides a robust foundation for implementing data-driven retention strategies. Targeted interventions focusing on contract incentives, payment method optimization, and service quality improvements could significantly reduce churn rates and improve customer lifetime value.

Key Business Impact: This analysis transforms customer churn from a reactive business challenge into a predictable and manageable process. By implementing the recommended strategies, organizations can build sustainable competitive advantages through superior customer retention, improved customer satisfaction, optimized resource allocation, and enhanced profitability.

Project Repository: GitHub - Customer Churn Analysis
Interactive Dashboard: Tableau Public - Customer Churn Dashboard

---

This report represents the culmination of comprehensive data analysis combining statistical insights, machine learning techniques, and business strategy to address one of the most critical challenges in subscription-based businesses.