📊 Customer Churn Analysis: Predictive Modeling for Retention Strategy

A comprehensive data science capstone project

Shayma Remy | Springboard - Data Science Track | August 2025

Table of Contents

📈 Executive Summary

> Key Insight: Contract type emerges as the most significant predictor of customer churn, accounting for nearly 80% of the model's predictive power.

Project Overview

- **Dataset**: 7,043 customer records from telecommunications industry
- **Best Model**: Logistic Regression with **AUC of 0.836**
- **Primary Finding**: Contract terms dominate churn prediction over pricing factors
- **Business Impact**: Actionable strategies to reduce churn through targeted interventions

Critical Success Metrics

Metric	Value	Impact
Model AUC	0.836	Strong predictive capability
Contract Importance	80%	Primary intervention target
High-Risk Segment	45-50% churn rate	Month-to-month fiber customers

---

🎯 Problem Statement and Business Context

The Challenge

> Customer acquisition costs often exceed 5x the cost of retention in subscription businesses

Core Problems:

- Traditional manual approaches fail to identify at-risk segments
- Inefficient retention campaigns waste resources on low-risk customers
- Reactive rather than proactive retention strategies
- Lack of data-driven insights into churn drivers

Business Impact

- Every percentage point reduction in monthly churn = millions in retained revenue

- Need for transition from reactive to proactive retention strategies
- Requirement for actionable insights through predictive analytics

---

🔬 Data Overview and Methodology

Dataset Specifications

Source: Telco Customer Churn Dataset (Kaggle)

- **Records:** 7,043 unique customers
- **Features:** 21 columns covering demographics, services, and financials
- **Target Variable:** Customer churn (binary)

Feature Categories

Category	Examples	Business Relevance
Demographics	Age, Gender, Family Structure	Customer profile segmentation
Services	Internet Type, Phone, Streaming	Service utilization patterns
Account Data	Contract Terms, Payment Methods	Relationship strength indicators
Financial	Tenure, Monthly Charges, Total Charges	Economic value metrics

Analytical Approach

Philosophy: Interpretable machine learning for business implementation

Key Steps:

1. **Data Cleaning & EDA** - Understanding patterns and relationships
2. **Feature Engineering** - Creating business-relevant derived variables
3. **Model Development** - Logistic Regression + Decision Tree approaches
4. **Performance Evaluation** - Focus on business-relevant metrics
5. **Insight Generation** - Actionable recommendations for retention

Tools & Technologies

- **Python:** Pandas, NumPy, Scikit-learn
- **Visualization:** Matplotlib, Seaborn, Tableau
- **Dashboard:** Tableau Public for interactive exploration

Analytical Approach

Philosophy: Interpretable machine learning for business implementation

Key Steps:

1. **Data Cleaning & EDA** - Understanding patterns and relationships
2. **Feature Engineering** - Creating business-relevant derived variables
3. **Model Development** - Logistic Regression + Decision Tree approaches
4. **Performance Evaluation** - Focus on business-relevant metrics
5. **Insight Generation** - Actionable recommendations for retention

Tools & Technologies

- **Python:** Pandas, NumPy, Scikit-learn
- **Visualization:** Matplotlib, Seaborn, Tableau
- **Dashboard:** Tableau Public for interactive exploration

🔍 Key Findings: Understanding Customer Behavior

👥 Demographic Insights

Family Structure Impact

> Finding: Customers with neither partners nor dependents show highest churn (34.2%)

Family Status	Churn Rate	Insight
No Partner, No Dependents	34.2%	Highest risk segment
Partner Only	25.4%	Moderate risk reduction

Family Status	Churn Rate	Insight
Partner + Dependents	14.2%	Lowest risk - family ties create stability

Business Implication: Family responsibilities create stronger service provider ties due to disruption costs for multiple household members.

Age Demographics

> Critical Finding: Senior citizens show ~42% churn vs ~24% for non-seniors (75% higher risk)

- **Gender Impact:** Minimal independent effect on churn rates
- **Age Priority:** Senior citizen segment requires targeted retention strategies

---

💳 Financial & Payment Behavior

Payment Method as Predictor

> Shocking Discovery: Electronic check users have 45.3% churn rate - nearly 3x higher than automatic payment users

Payment Method	Churn Rate	Risk Level
Electronic Check	45.3%	**Critical**
Mailed Check	19.1%	Moderate
Bank Transfer (Auto)	16.7%	Low
Credit Card (Auto)	15.2%	**Lowest**

Key Insight: Automatic payment setup demonstrates higher service commitment and convenience preference.

Tenure & Pricing Relationship

- **Pattern:** Longer-tenured customers pay higher monthly charges
- **Range:** $56.17 (0-12 months) to $75.95 (61+ months)

- **Implication:** Either service upgrades over time or accepted price increases by loyal customers

---

📋 Contract & Service Factors

Contract Type: The Primary Driver

> Major Finding: Month-to-month customers show 31.5-48.3% churn vs 1.6-4.2% for two-year contracts

Contract Type	Churn Rate Range	Strategic Priority
Month-to-Month	31.5% - 48.3%	**Immediate intervention target**
One-Year	7.1% - 14.8%	Moderate risk management
Two-Year	1.6% - 4.2%	Retention success model

Paperless Billing Paradox

- **Pattern:** Paperless billing correlates with ~1.5x higher churn across all contract types
- **Consideration:** Digital engagement strategies need careful analysis

Service Quality Issues

- **Fiber Internet Users:** Show elevated churn despite premium pricing
- **Implication:** Potential service quality or expectation management issues

---

🤖 Model Performance & Results

Model Architecture Comparison

Logistic Regression (Primary Model)

> Performance: AUC = 0.836 | Status: Recommended for deployment

Strengths:

- Clear coefficient interpretations for business stakeholders
- Probability estimates enable risk-based segmentation
- Superior precision-recall balance
- Ready integration with CRM systems

Confusion Matrix Results:

Predicted	Actual No Churn	Actual Churn
No churn	739 ✅	296 ❌
Churn	73 ❌	301 ✅

Key Metrics:

- **Recall:** 91.0% (excellent at identifying churners)
- **Precision:** Balanced approach minimizing false positives
- **Business Value:** Enables tiered retention strategies

Decision Tree (Secondary Model)

> Performance: Higher recall (92.9%) but more false positives

Confusion Matrix Results:

Predicted	**Actual No Churn**	**Actual Churn**
No churn	591 ✅	444 ❌
Churn	45 ❌	329 ✅

Trade-offs:

- **Advantage:** Captures non-linear relationships and interactions
- **Challenge:** Higher false positive rate (444 vs 296)
- **Use Case:** Aggressive churn identification when intervention costs are low
- **Recall Rate:** 92.9% (excellent at catching all potential churners)

- **Precision Trade-off:** More aggressive prediction strategy with higher false alarms

Feature Importance Hierarchy

> Critical Insight: Contract-related features dominate with 80% combined importance

Feature	Importance	Strategic Focus
Two-Year Contract	49.2%	**Primary intervention target**
One-Year Contract	30.8%	**Secondary retention strategy**
Fiber Internet Service	13.5%	Service quality investigation
Streaming Movies	4.1%	Service bundle optimization
Total Charges	1.2%	Price is not the primary driver

Strategic Reframe: Contract-based interventions should take priority over price-based retention strategies.

📊 Interactive Dashboard and Business Intelligence

Dashboard Features

- **Dynamic Filtering:** Churn risk scores, tenure segments, contract types
- **Segment Identification:** Interactive exploration of customer cohorts
- **Real-time Insights:** Drill-down capabilities for targeted analysis

High-Risk Segment Discovery

> Identified: Month-to-month fiber customers using electronic check payments with <12 months tenure

Risk Profile:

- **Churn Probability:** 45-50%

- **Characteristics:**
  - Short tenure (vulnerability period)
  - Premium service (high expectations)
  - Inconvenient payment method
  - No commitment mechanism

Additional Vulnerable Segments:

- Senior citizens without family ties
- Paperless billing users with month-to-month contracts
- New customers in first 12 months

Dashboard Access

🔗 **Interactive Dashboard:** Tableau Public Link

💼 Business Recommendations & Implementation

🚀 Immediate Action Items (0-3 Months)

1. Contract Incentive Program

Objective: Convert month-to-month customers to annual/bi-annual contracts

Strategy:

- **One-Year Incentive:** 10-15% discount on monthly rates
- **Two-Year Incentive:** 15-20% discount + value-added perks
- **Additional Benefits:** Free installation, equipment upgrades, premium support

Expected Impact: Reduce churn from 40%+ to under 15% for converted customers

2. Payment Method Migration Campaign

Objective: Convert electronic check users to automatic payments

Tactics:

- **Monthly Credits:** $5-10 bill credit for automatic payment setup
- **One-time Bonus:** Signup incentive for immediate conversion
- **Process Simplification:** Guided assistance and friction removal

Expected Impact: 60-65% churn reduction among electronic check users

3. Predictive Scoring Deployment

Objective: Integrate model into CRM for proactive interventions

Implementation:

- **Real-time Scoring:** Generate churn probability during customer interactions
- **Automated Triggers:** Risk-based retention program activation
- **Tiered Interventions:** Customized strategies based on individual risk profiles

Expected Impact: 20-30% overall churn reduction through proactive identification

---

📋 Strategic Initiatives (3-12 Months)

Fiber Service Quality Enhancement

Challenge: Elevated churn among high-value fiber customers

Action Plan:

1. **Customer Satisfaction Surveys:** Targeted fiber user feedback collection
2. **Support Ticket Analysis:** Identify systematic service issues
3. **Infrastructure Assessment:** Technical performance evaluation
4. **Service Level Agreement Review:** Align expectations with delivery

Outcome: Improved retention in premium service segment

Senior Citizen Retention Program

Target: Address 75% higher churn risk among seniors

Specialized Approach:

- Dedicated support channels with enhanced assistance

- Simplified service packages with clear value proposition
- Community engagement programs
- Family plan incentives for multi-generational households

🎯 Success Metrics & KPIs

Metric	Current Baseline	Target (6 months)	Target (12 months)
Overall Churn Rate	~27%	<22%	<20%
Month-to-Month Churn	40%+	<30%	<25%
Electronic Check Churn	45.3%	<30%	<25%
Contract Conversion Rate	-	15%	25%
Payment Method Migration	-	20%	35%

🔬 Future Research Directions

Advanced Modeling Opportunities

- **Ensemble Methods:** Random Forest, Gradient Boosting, Neural Networks
- **Time Series Analysis:** Seasonal patterns and temporal churn behaviors
- **Customer Lifetime Value Integration:** Value-weighted churn prediction

Enhanced Data Integration

- **Customer Service Data:** Support tickets, satisfaction surveys, call logs
- **Usage Patterns:** Service consumption metrics and behavioral changes
- **Competitive Intelligence:** Market conditions and competitor analysis

Experimental Framework

- **A/B Testing:** Contract incentive effectiveness measurement
- **Payment Migration Experiments:** Optimal incentive structure identification
- **Personalized Intervention Testing:** Timing and messaging optimization

---

🎯 Conclusion & Business Impact

Transformation Achievement

This analysis successfully transforms customer churn from a **reactive business challenge** into a **predictable and manageable process**.

Key Success Factors

1. **Data-Driven Insights:** Contract terms identified as 80% of predictive power
2. **Actionable Intelligence:** Clear intervention priorities with expected ROI
3. **Implementation Ready:** Models designed for business integration
4. **Strategic Framework:** Comprehensive retention strategy development

Competitive Advantages

- **Superior Customer Retention** through predictive intervention
- **Optimized Resource Allocation** via risk-based targeting
- **Enhanced Customer Satisfaction** through proactive service
- **Improved Profitability** from reduced acquisition costs

Final Business Impact

> Expected Outcome: 20-30% reduction in overall churn rates through systematic implementation of contract incentives, payment optimization, and predictive intervention strategies.

📚 Project Resources

Code Repository

🔗 **GitHub:** Customer Churn Analysis - Complete Code

Interactive Exploration

🔗 **Dashboard:** Tableau Public - Customer Churn Dashboard

Technical Documentation

- **Data Dictionary:** Variable definitions and business meanings
- **Model Documentation:** Algorithm specifications and parameter tuning
- **Validation Results:** Cross-validation and performance metrics

---

This capstone project represents a comprehensive approach to customer churn analysis, combining advanced data science techniques with practical business strategy to address critical subscription business challenges.