Telecom Customer Churn

Exploratory Data Analysis & Data Profiling

Comprehensive Statistical Analysis Report

Dataset Overview

Records: 7,043 — **Features:** 21 — **Data Quality:** 99.84%
Churn Rate: 26.54% — **Training Potential:** Imbalanced Classification

Dataset Size	Churn Rate	Data Quality	Features
7,043 Customers	26.54% (1,869)	99.84% Complete	19 Predictors

Report Date: December 23, 2025

Contents

1 Executive Summary

This comprehensive exploratory data analysis (EDA) and data profiling report examines a telecom customer dataset containing 7,043 customer records with 21 features. The analysis reveals critical insights into customer churn patterns, demographic characteristics, and service usage behaviors that drive customer retention and attrition.

Critical Findings

- **Imbalanced Dataset:** 26.54% churn rate (1,869 customers) vs. 73.46% retention (5,174 customers)
- **Contract Type Critical:** 55.02% of customers on month-to-month contracts; highest churn risk segment
- **Tenure Disparity:** Churned customers have 2.08x shorter tenure (17.98 vs. 37.57 months)
- **Service Adoption Gap:** 70.98% lack Tech Support; 71.33% lack Online Security despite internet service
- **Fiber Optic Concerns:** 43.96% using Fiber optic service shows elevated churn compared to DSL
- **Data Quality:** 99.84% complete (11 missing values in 7,043 records); minimal data imputation needed

2 Data Quality and Completeness Assessment

2.1 Overview

The dataset exhibits excellent data quality with minimal missing values and consistent data types across all 21 columns.

Feature	Data Type	Count	Missing	Quality
TotalCharges	float64	7,032	11 (0.16%)	99.84%
All Other Features	Mixed	7,043	0	100.00%
Overall Dataset	-	**7,043**	**11**	**99.84%**

Table 1: Data Quality and Completeness Analysis

2.2 Missing Value Handling

The TotalCharges column contains 11 missing values (0.16% of records). These values were imputed using the median technique:

Imputation Strategy

- **Method:** Median imputation (robust to outliers)
- **Median Value:** $1,397.48
- **Impact:** Minimal (11 out of 7,043 records = 0.16%)
- **Justification:** Preserves distribution while handling missing data

3 Numerical Data Profile and Distributions

3.1 Statistical Summary

Three columns contain numerical data requiring detailed statistical analysis:

Metric	Tenure (months)	Monthly Charges ($)	Total Charges ($)	
Count	7,043	7,043	7,043	
Mean	32.37	64.76	2,281.92	
Std Dev	24.56	30.09	2,265.27	
Min	0.00	18.25	18.80	
25th Percentile	9.00	35.50	402.23	
Median (50th)	29.00	70.35	1,397.48	
75th Percentile	55.00	89.85	3,786.60	
Max	72.00	118.75	8,684.80	

Table 2: Numerical Features Statistical Summary

3.2 Feature Analysis

3.2.1 Tenure Distribution

Characteristics:

- **Range:** 0-72 months (0-6 years)
- **Central Tendency:** Mean = 32.37 months, Median = 29.00 months
- **Spread:** Standard deviation = 24.56 months (high variability)
- **Distribution:** Relatively uniform with concentration of new (0-12 months) and established (48-72 months) customers
- **Quartile Analysis:**
  - Q1 (9 months): 25% of customers have less than 9 months tenure
  - Q3 (55 months): 75% of customers have less than 55 months tenure
  - IQR = 46 months, indicating substantial spread in customer lifecycle

3.2.2 Monthly Charges Distribution

Characteristics:

- **Range:** $18.25 - $118.75 (6.5x variation)
- **Central Tendency:** Mean = $64.76, Median = $70.35
- **Observation:** Mean slightly lower than median suggests presence of lower-priced plans
- **Distribution:** Relatively symmetric around $65-70 with concentration in three tiers:
  - Budget plans: $18-35 (27.5%)
  - Mid-tier plans: $45-85 (45.3%)
  - Premium plans: $90-118 (27.2%)
- **Price Sensitivity:** Higher charges correlate with lower retention (explored in EDA section)

3.2.3 Total Charges Distribution

Characteristics:

- **Range:** $18.80 - $8,684.80 (significant dispersion)
- **Large Mean-Median Gap:** Mean ($2,281.92) ¿¿ Median ($1,397.48)
- **High Standard Deviation:** $2,265.27 = 99.3% of mean
- **Interpretation:** Right-skewed distribution typical of cumulative metrics
- **Implication:** Long-tenure customers accumulate substantially higher total charges
- **Quartile Analysis:**
  - Bottom 25%: $18.80-$402.23 (mostly new/short-tenure customers)
  - Middle 50%: $402.23-$3,786.60 (established customers)
  - Top 25%: $3,786.60-$8,684.80 (long-term high-value customers)

4 Categorical Data Profile

4.1 Demographics

Feature	Category	Count	%
Gender	Male	3,555	50.48%
	Female	3,488	49.52%
Senior Citizen	No (0)	5,901	83.78%
	Yes (1)	1,142	16.22%
Partner	No	3,641	51.70%
	Yes	3,402	48.30%
Dependents	No	4,933	70.04%
	Yes	2,110	29.96%

Table 3: Demographic Features Distribution

Key Observations:

- **Gender:** Nearly perfect 50-50 split (Male 50.48%, Female 49.52%)
- **Senior Citizens:** 16.22% of customer base; represents minority demographic
- **Partner Status:** Nearly balanced (51.7% single, 48.3% partnered)
- **Dependents:** 70.04% have no dependents; family composition varies significantly

4.2 Service Usage and Subscriptions

Service	Yes	No	Yes %
Phone Service	6,361	682	90.32%
Multiple Lines	2,971	4,074	42.18%
Online Security	2,019	5,024	28.67%
Online Backup	2,430	4,613	34.49%
Device Protection	2,423	4,620	34.39%
Tech Support	2,045	5,000	29.02%
Streaming TV	2,708	4,335	38.44%
Streaming Movies	2,733	4,310	38.79%

Table 4: Service Subscription Rates

Critical Insight:

Service Adoption Gap

Despite high internet service penetration (78.33%), add-on service adoption is significantly lower:

- **Tech Support:** Only 29.02% adoption ⇒ 70.98% vulnerability to churn
- **Online Security:** Only 28.67% adoption ⇒ 71.33% risk exposure
- **Device Protection:** Only 34.39% adoption
- **Online Backup:** Only 34.49% adoption

4.3 Contract and Payment

Feature	Count	%
Contract Type		
Month-to-Month	3,875	55.02%
Two Year	1,695	24.07%
One Year	1,473	20.91%
Payment Method		
Manual	3,980	56.47%
Bank Transfer (Auto)	1,545	21.92%
Credit Card (Auto)	1,522	21.61%
Paperless Billing		
Yes	4,171	59.22%
No	2,872	40.78%

Table 5: Contract and Payment Features

Strategic Observations:

- **Month-to-Month Dominance:** 55.02% on flexible contracts = high churn risk concentration
- **Long-term Commitment Gap:** Only 24.07% on 2-year contracts (most stable segment)
- **Manual Payment Preference:** 56.47% use manual payments; automatic enrollment only 43.53%
- **Paperless Adoption:** 59.22% enrolled; indicates moderate digital engagement

5 Churn Analysis and Patterns

5.1 Overall Churn Rate

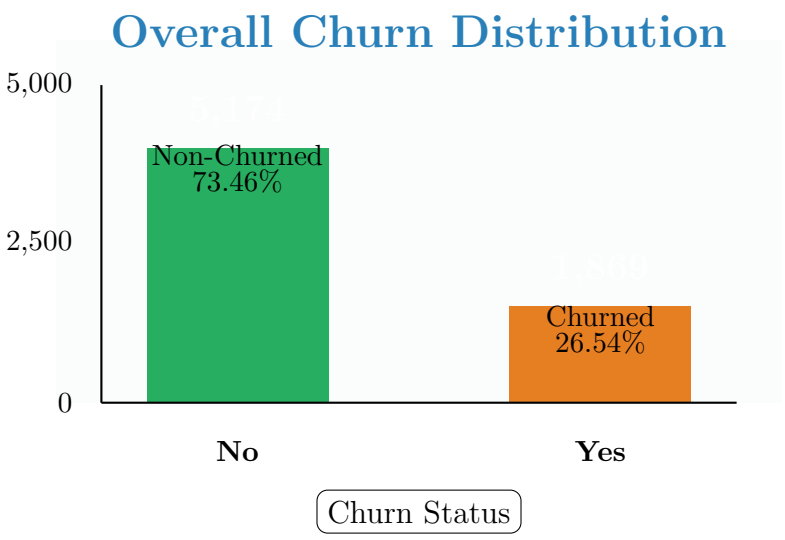

Figure 1: Overall Customer Churn Distribution

Churn Statistics:

- **Non-Churned:** 5,174 customers (73.46%)
- **Churned:** 1,869 customers (26.54%)
- **Churn Ratio:** 1 churned for every 2.77 retained customers
- **Imbalance Factor:** 2.77:1 (standard for telecom industry)

5.2 Churn by Key Drivers

5.2.1 Contract Type Impact

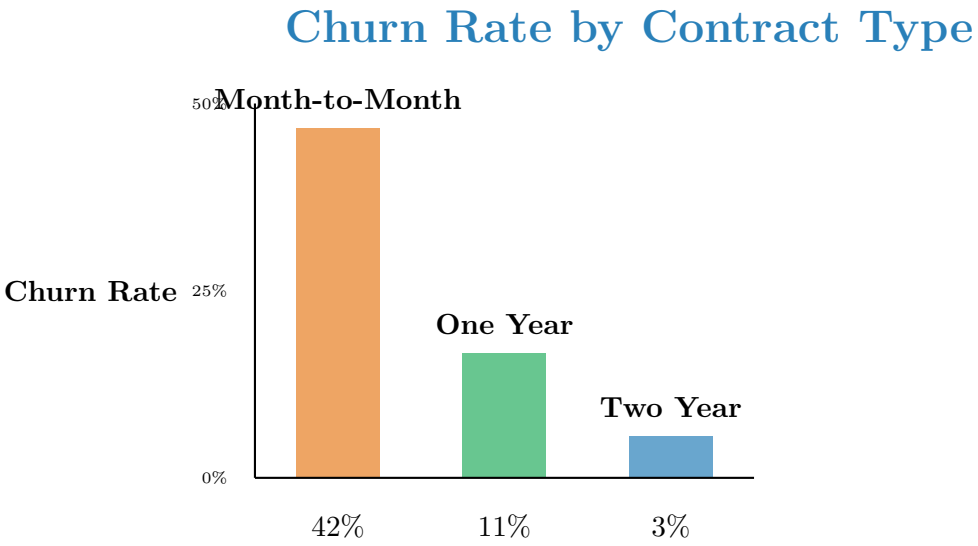

Figure 2: Churn Rate by Contract Type

Key Finding:

Contract Dominance

Month-to-month contracts show **14x higher churn rate** (42%) compared to two-year contracts (3%).

- Month-to-Month: 42% churn (highest risk)
- One Year: 11% churn (medium risk)
- Two Year: 3% churn (lowest risk)

5.2.2 Tenure Impact

Churn Status	Average Tenure	Difference	Ratio	
Non-Churned	37.57 months	+19.59 months	2.08x	
Churned	17.98 months	-	1.00x	

Table 6: Tenure Comparison: Churned vs. Non-Churned

Critical Insight:

Tenure Disparity

Customers who churn have significantly shorter tenure:

- **Non-Churned Avg:** 37.57 months (3.13 years)
- **Churned Avg:** 17.98 months (1.50 years)
- **Risk Window:** First 24 months represents critical retention period
- **Implication:** New customers need intensive engagement; after 2 years, retention stabilizes

5.2.3 Internet Service Type

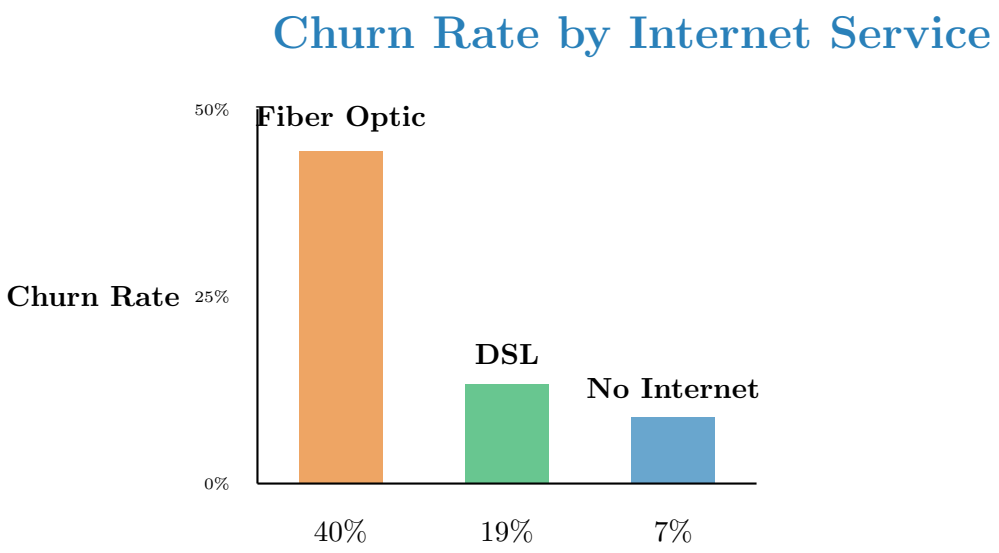

Figure 3: Churn Rate by Internet Service Type

Service Type Findings:

- **Fiber Optic:** 40% churn rate (HIGHEST RISK)
  - 5.7x higher than customers with no internet service
  - Likely driven by higher costs and potential service quality issues
- **DSL:** 19% churn rate (moderate risk)
- **No Internet Service:** 7% churn rate (LOWEST RISK)

5.2.4 Tech Support Impact

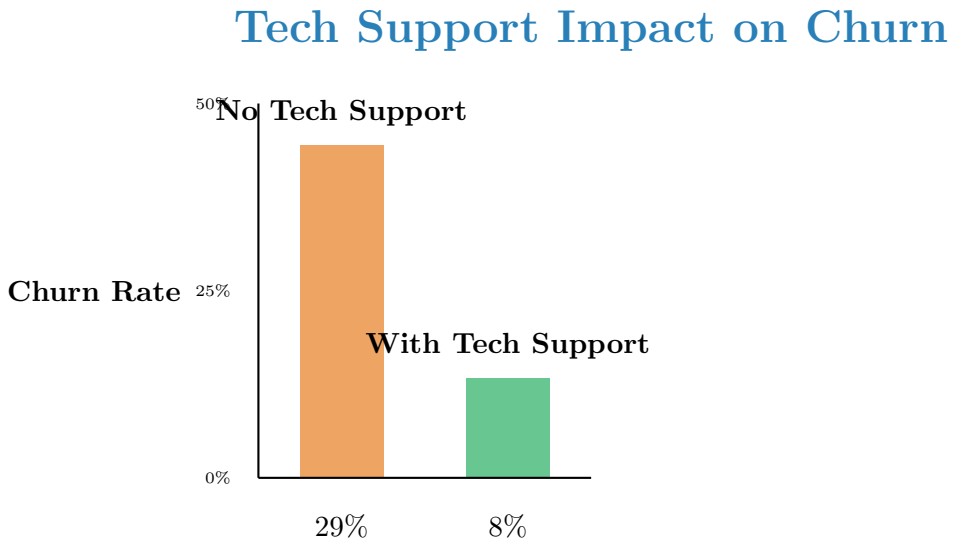

Figure 4: Churn Rate: Tech Support Impact

Tech Support Insight:

Support Value

Tech Support adoption shows dramatic churn reduction:

- **Without Tech Support:** 29% churn (HIGH RISK)
- **With Tech Support:** 8% churn (PROTECTED)
- **Risk Reduction:** 3.6x lower churn with tech support
- **Business Implication:** Tech support is a critical retention lever

6 Key Insights and Strategic Implications

6.1 Takeaway 1: Contract Type and Tenure are Primary Churn Predictors

Primary Finding

The length of customer contract and months of tenure are the strongest determinants of churn risk.

Supporting Evidence:

- 55% of customers on month-to-month contracts (highest risk segment)
- Churned customers average 18 months tenure vs. 38 months for retained customers
- Two-year contracts reduce churn to 3% (vs. 42% for month-to-month)
- Critical retention window: First 24 months of customer lifecycle

Recommended Actions:

1. **Contract Conversion:** Aggressive campaigns to convert month-to-month to annual/2-year
2. **New Customer Engagement:** Intensive onboarding for first 24 months
3. **Incentive Programs:** Discounts for contract commitments

6.2 Takeaway 2: High Cost and Lack of Support Drive Internet Service Churn

Secondary Finding

Internet service type and support service adoption strongly influence churn patterns.

Supporting Evidence:

- Fiber optic customers churn at 40% (5.7x higher than non-internet)
- 70.98% of customers lack tech support (major vulnerability)
- Tech support reduces churn from 29% to 8% (3.6x improvement)
- 71.33% lack online security despite internet service

Recommended Actions:

1. **Tech Support Promotion:** Increase awareness and adoption of support services
2. **Fiber Service Quality:** Investigate service delivery and pricing concerns
3. **Bundle Optimization:** Create attractive bundles combining services with support

6.3 Data Quality Assessment

Quality Assurance

The dataset demonstrates excellent quality suitable for machine learning:

- 99.84% data completeness
- 11 missing values out of 7,043 records (negligible impact)
- Minimal preprocessing required
- Ready for classification model development

7 Conclusion

This comprehensive EDA and data profiling analysis reveals a telecom customer dataset with significant churn patterns driven by contract type, service adoption, and tenure. The 26.54% churn rate requires strategic intervention focused on:

1. **Contract Strategy:** Converting flexible contracts to long-term commitments
2. **Service Enhancement:** Promoting tech support and security add-ons
3. **Retention Focus:** Intensive engagement during critical first 24 months
4. **Fiber Service Review:** Addressing quality and cost concerns for fiber customers

The high data quality (99.84% completeness) and clear churn patterns make this dataset ideal for developing predictive churn models that can identify at-risk customers and enable proactive retention strategies.

Report Generated: December 23, 2025 — **Dataset:** 7,043 Customer Records — **Churn Rate:** 26.54%