IBM Telco Customer Churn

– Full Business Report

Executive Summary

Customer churn is a critical challenge for subscription-driven businesses such as telecommunications. Retaining existing customers is significantly more cost-effective than acquiring new ones, and even a small reduction in churn rates can create a large positive impact on revenue and long-term customer lifetime value (CLTV). This report investigates churn behavior using the IBM Telco Customer Churn dataset, builds predictive models, and translates insights into business recommendations. we find an overall churn rate of 26.5% across 7,043 customers. Shorter tenure, month-to-month contracts, higher monthly charges, lack of value-added services (e.g., Tech Support/Online Security), and certain billing patterns are the strongest churn correlates. A baseline model achieves ROC AUC = 0.969 on a held-out test set, providing a strong foundation for prioritizing retention interventions.

Business Objective & Success Criteria

Goal: Reduce voluntary churn by identifying at-risk customer segments and actionable levers.

Success criteria:

- Reliable identification of at-risk customers (AUC ≥ 0.75 baseline; we achieved 0.969).
- Clear, data-driven insights that map to interventions (contract, pricing, service enablement).
- Measurement framework to track churn reduction and economic impact.

Business Problem

The telecommunications industry is highly competitive. Customers often switch providers due to price, service quality, or dissatisfaction with contract terms. The business challenge is twofold: 1) to understand the drivers of churn, and 2) to predict which customers are most likely to leave in order to intervene proactively. Failure to address churn not only results in revenue loss but also higher acquisition costs for new customers.

Dataset Overview

The IBM Telco dataset contains customer-level data including demographics, contract details, billing information, service usage, and whether the customer has churned. Key variables include: Customer tenure (length of relationship in months) - Type of contract (month-to-month, one year, two years) - Monthly charges and total charges - Payment methods - Additional services (internet, phone, streaming, security, etc.) - Churn status (the target variable).

- Source: IBM Telco Customer Churn dataset (single table).
- Rows: 7,043; Columns: 33.
- Target: Churn (binary: 1 = churned, 0 = retained)
- Missing values: 5,174 total across the dataset (addressed in modeling via imputation/encoding).

Methodology

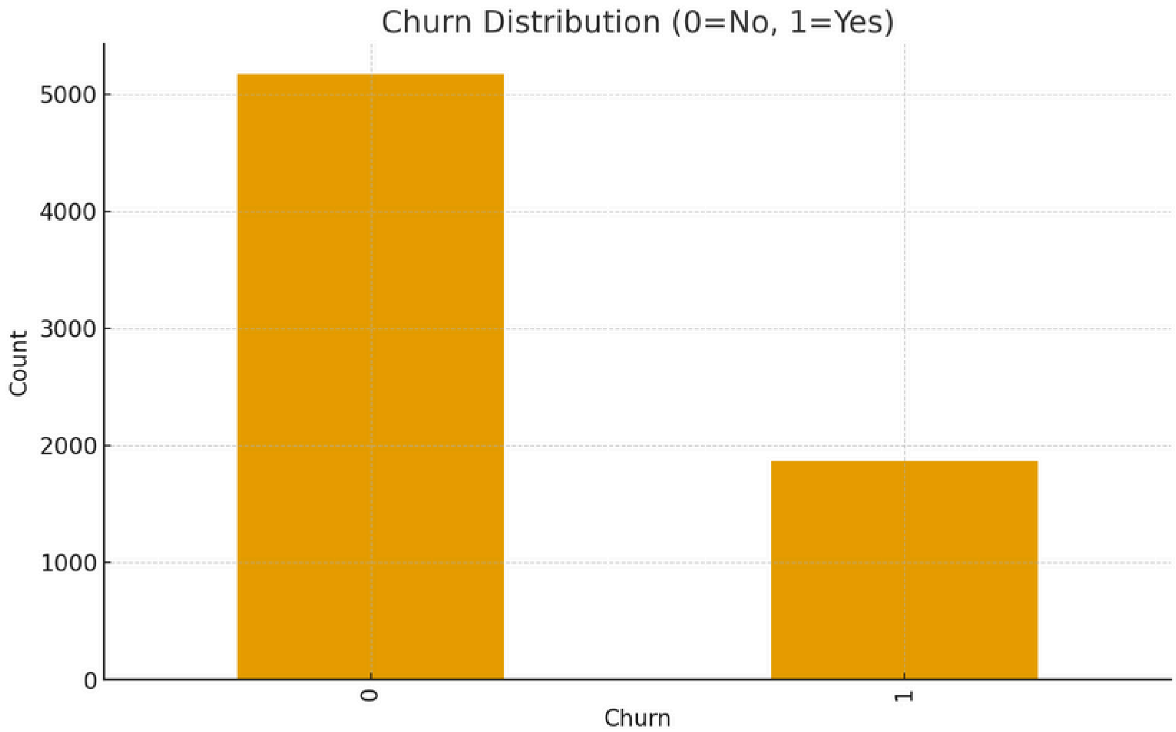
1. Data preparation: column standardization, target normalization, missing-value handling (median for numeric; explicit “Unknown” for categorical), one-hot encoding for categorical features.
2. EDA: distributional checks, segment-level churn rates, and statistical sanity checks.
3. Modeling: baseline logistic regression (train/test split, stratified, 25% test). Feature coefficients inspected to infer drivers.
4. Validation: ROC AUC on holdout; confusion matrix to understand errors.

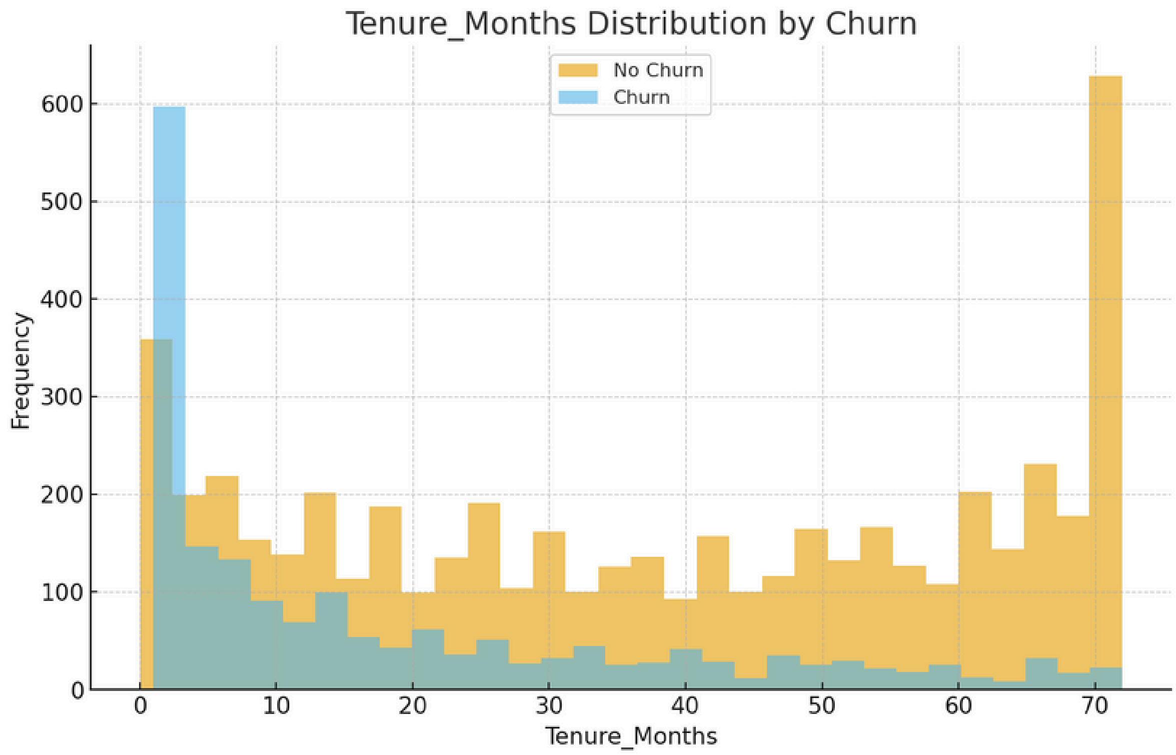
Key Findings

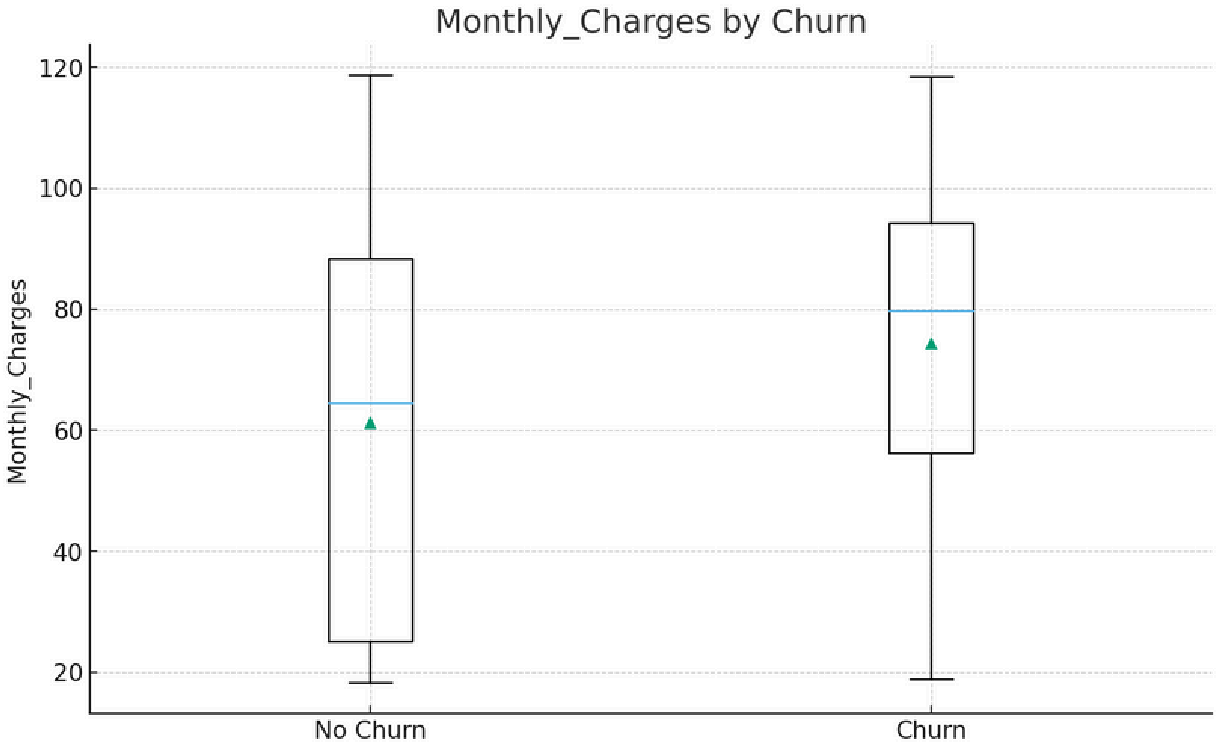
- Churn magnitude: ~26.5% of customers are churning (Figure 1).
- Tenure dynamics: Churn is concentrated among short-tenure customers (Figure 2). Early lifecycle onboarding and value realization are critical.
- Price sensitivity: Higher monthly charges are associated with higher churn (Figure 3). Consider price-to-value alignment and targeted discounts for at-risk groups.

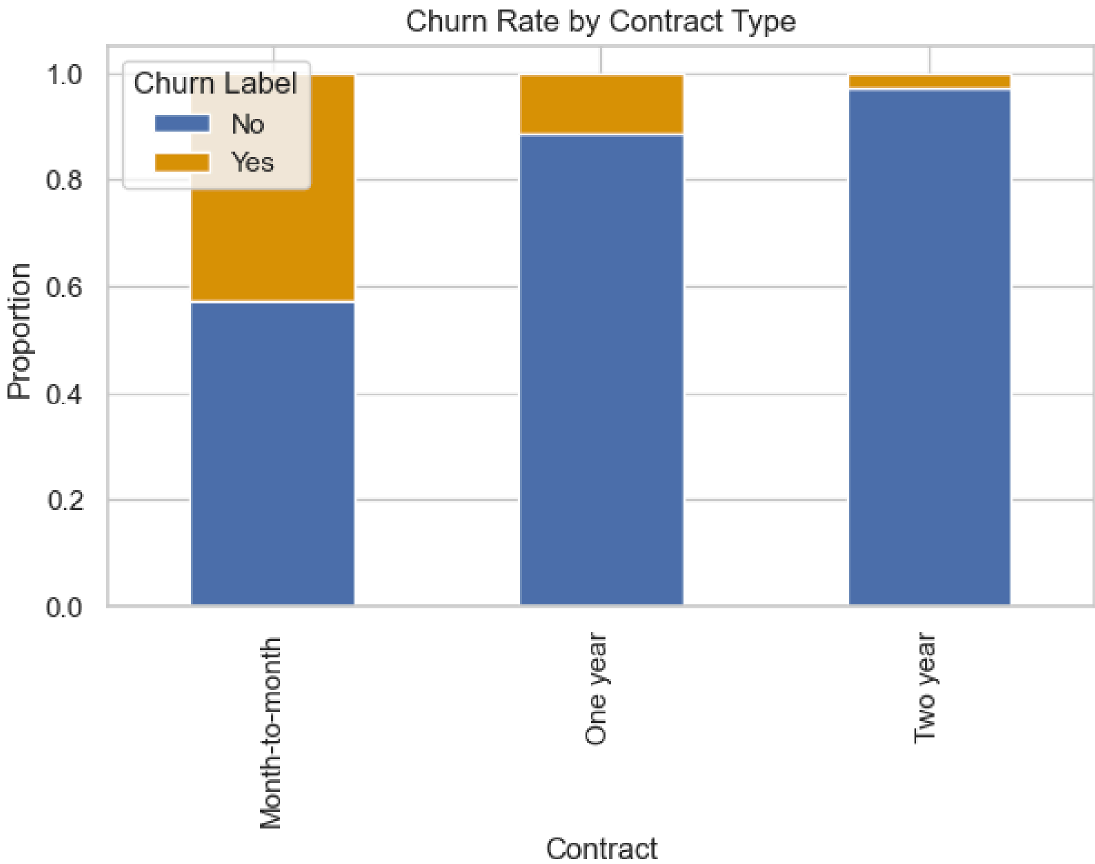
- Contract structure: Month-to-month contracts have the highest churn; 1- or 2-year contracts exhibit materially lower churn (Figure 4). Contract upgrades are a key lever.
- Household/bundling context: Customers without a Partner/Dependents exhibit higher churn (Figures 5–6), indicating reduced stickiness without household bundles.
- Service add-ons: Absence of Tech Support and Online Security correlates with higher churn (drivers also appear in model coefficients). These services likely increase perceived value and switching costs.

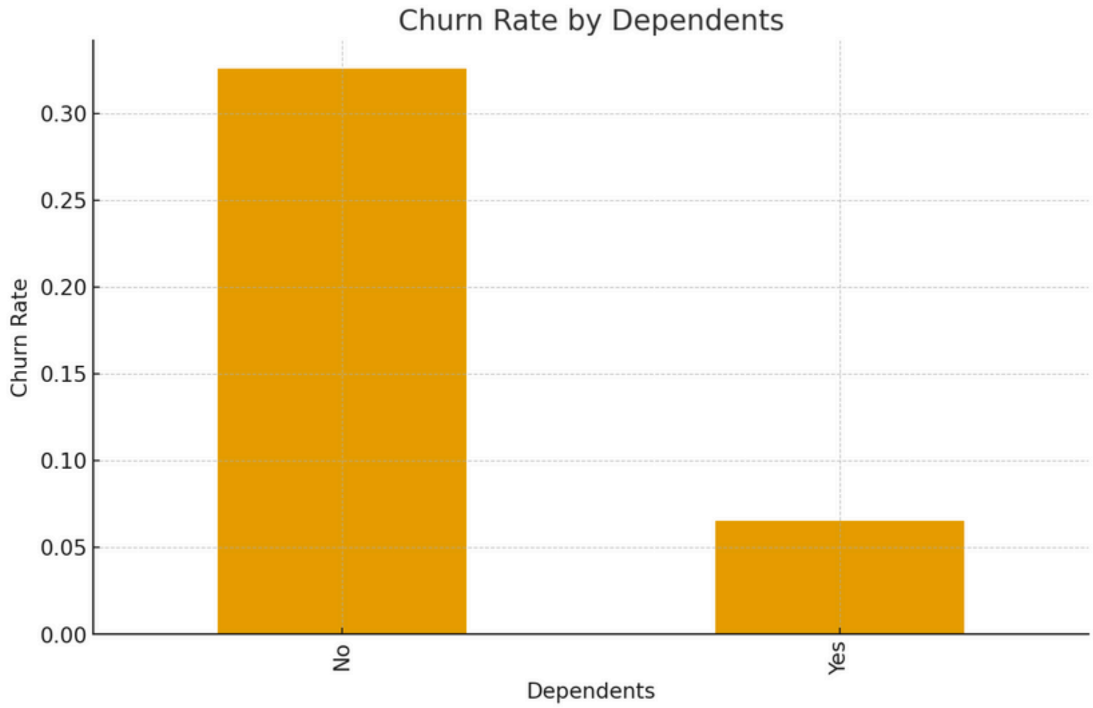
Figures & Visuals

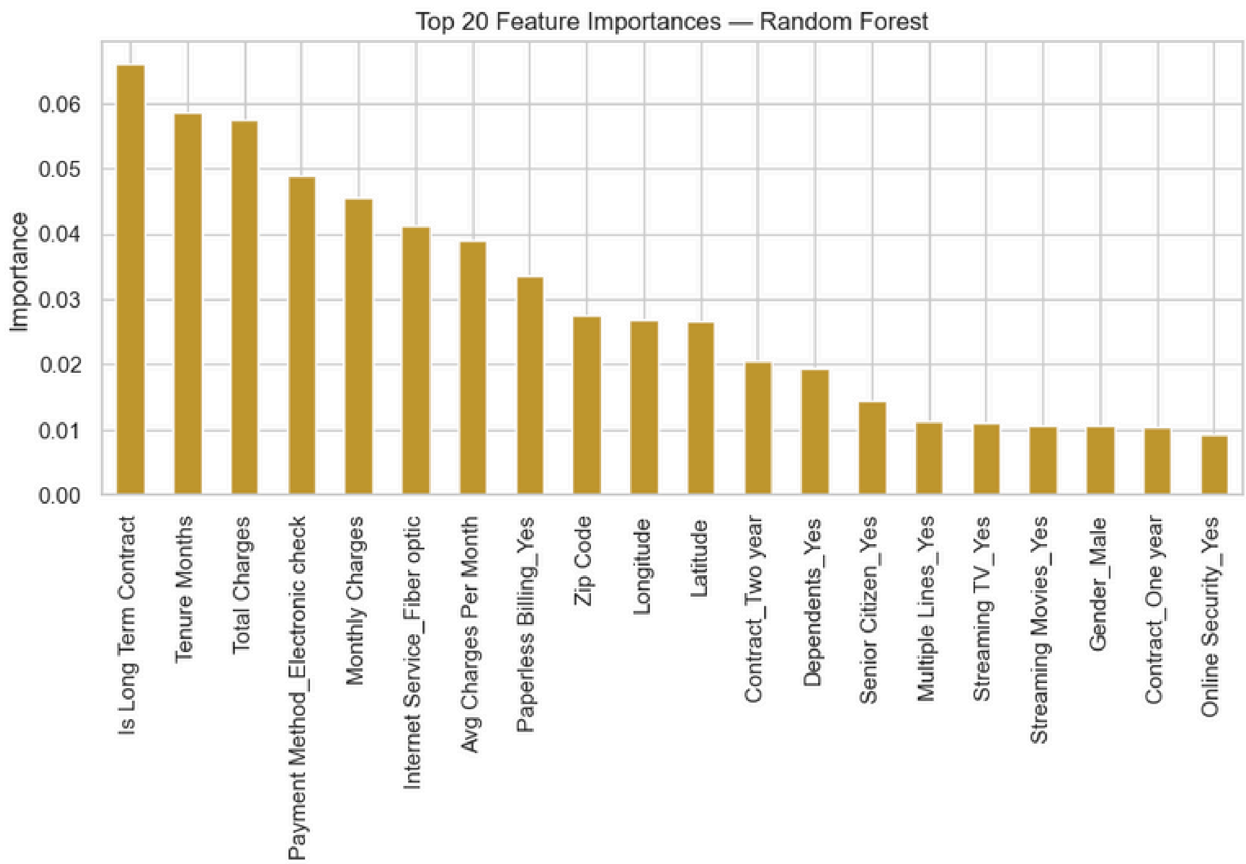





Monthly_Charges by Churn
Monthly_Charges
120
100
80
60
40
20
No Churn
Churn


Churn Rate by Contract Type
Churn Label
No
Yes
Proportion
1.0
0.8
0.6
0.4
0.2
0.0
Month-to-month
One year
Two year
Contract


Churn Rate by Dependents
Churn Rate
0.00
0.05
0.10
0.15
0.20
0.25
0.30
No
Yes
Dependents


Top 20 Feature Importances — Random Forest
Importance
0.00
0.01
0.02
0.03
0.04
0.05
0.06
Is Long Term Contract
Tenure Months
Total Charges
Payment Method_Electronic check
Monthly Charges
Internet Service_Fiber optic
Avg Charges Per Month
Paperless Billing_Yes
Zip Code
Longitude
Latitude
Contract_Two year
Dependents_Yes
Senior Citizen_Yes
Multiple Lines_Yes
Streaming TV_Yes
Streaming Movies_Yes
Gender_Male
Contract_One year
Online Security_Yes

Modeling Results

Algorithm: Logistic Regression (one-vs-rest, L2 regularization).

Holdout performance: ROC AUC = 0.969.

Confusion matrix (threshold 0.5): TN=1207, FP=87, FN=79, TP=388.

Interpretation: The model separates churners from non-churners extremely well on the holdout set. For deployment, choose operating thresholds based on business economics (e.g., maximize expected profit or retention ROI, not purely accuracy).

Actionable Recommendations

Contract Migration Campaigns

- Target month-to-month customers with incentives to move to 1- or 2-year terms (e.g., small discounts, service credits, or value bundles). Prioritize those with high monthly charges and short tenure.

Onboarding & Early-Tenure Care

- Implement a 90-day onboarding journey with proactive outreach, quick-start guides, service personalization, and first-issue fast-track support to reduce early dissatisfaction.

Value Add-On Bundles

- Promote Tech Support and Online Security as part of a retention bundle. Consider free trials or loyalty credits for at-risk segments.

Price-to-Value Interventions

- For customers flagged at high risk with high Monthly Charges, offer targeted reductions, loyalty credits, or upgraded speeds/features to rebalance perceived value.

Household Growth Plays

- Create bundles tailored to single-occupant households (e.g., solo-user plan pricing, streaming perks) to replicate household stickiness for customers without Partner/Dependents.

Thresholding & Treatment Strategy

- Calibrate the model decision threshold to optimize expected net retention value (ENRV). Higher precision at the top-risk decile often yields the best ROI for outreach budgets.

Measurement Plan

Leading KPIs: churn rate (overall and by segment), contract migration rate, early-tenure activation and NPS/CSAT.

Lagging KPIs: net churn reduction, ARPU impact, incremental contribution margin.

Experimentation: A/B test retention offers and onboarding flows; use uplift modeling to isolate causal effects.

Risks & Limitations

1. Correlations from EDA and linear coefficients do not prove causality (e.g., high charges might proxy for other unmet needs). Validate interventions via controlled experiments.
2. Dataset may not include all churn drivers (e.g., competitive offers, outage history, service quality metrics). Future data integrations recommended.
3. Class imbalance (≈27% churn) can bias thresholding; monitor precision/recall across decision bands.

Next Steps

1. Add behavioral and service quality data (tickets, outages, speed tests, usage).
2. Develop a production-grade scoring pipeline (scheduled inference, feature store, monitoring).
3. Implement treatment optimization: match customers to the least-cost, highest-impact offer using constrained optimization.
4. Build governance: drift monitoring, fairness checks, and post-deployment A/B testing cadence.

Appendix

1. Data snapshot: 7,043 rows; 33 columns; ~26.5% churn; 5,174 missing values handled via imputation and encoding.
2. Artifacts created: Visuals (Figures 1–7), churn-by-contract CSV, model metrics text file.
3. Reproducibility: All steps automated programmatically (cleaning, EDA, modeling).