Telecommunications Industry

Customer Churn Analysis Report

Neha Jhakra

May 2025

A Data-Driven Approach to Understanding Customer Retention

Submitted for Public Sharing on GitHub and LinkedIn

Abstract

This report presents a comprehensive analysis of customer churn for a telecommunications company, utilizing a dataset of 7,043 customers. The project employs Python for data cleaning and exploratory data analysis (EDA), SQL for querying insights, and visualizations to identify factors influencing churn. Key findings highlight the impact of contract type, tenure, monthly charges, and payment methods on churn rates, providing actionable insights for improving customer retention.

Contents

Executive Summary

This project, conducted by Neha Jhakra, analyzes customer churn in a telecommunications company to identify factors driving customer attrition and propose retention strategies. The analysis leverages a dataset of 7,043 customers, combining Python-based data processing with SQL queries to extract actionable insights.

Objectives:

- Clean and preprocess the Telco Customer Churn dataset.
- Conduct exploratory data analysis (EDA) to uncover churn patterns.
- Query specific customer segments using SQL.
- Recommend data-driven strategies to reduce churn.

Methodology:

- Data Cleaning: Converted data types, handled missing values, and saved a cleaned dataset (7,032 records).
- EDA: Used Pandas, Seaborn, and Matplotlib to analyze churn distributions, correlations, and trends.
- SQL Analysis: Imported data into MySQL to query churn metrics, contract types, and service impacts.
- Visualizations: Created count plots, box plots, histograms, and heatmaps to visualize findings.

Key Findings:

- Month-to-month contract customers had the highest churn rate (42.7%).
- Fiber optic internet users churned at a high rate (69.4% of churned customers).
- Electronic check users were more likely to churn (45.3%).
- Customers with shorter tenure (0–12 months) and higher monthly charges were at greater risk of churning.

Recommendations:

- Promote long-term contracts with incentives to increase tenure.
- Investigate fiber optic service issues and offer bundled discounts.
- Streamline electronic check billing and promote automatic payments.
- Target new customers (0–12 months) with loyalty programs.

This analysis provides a foundation for reducing churn through targeted retention strategies, with potential for future predictive modeling.

1 Introduction

The telecommunications industry faces significant challenges with customer churn, where customers discontinue services. This project, conducted by Neha Jhakra, analyzes the Telco Customer Churn dataset to uncover patterns and predictors of churn. The analysis combines Python-based data processing and visualization with SQL queries to extract meaningful insights. The objectives are to:

- Clean and preprocess the dataset for analysis.
- Perform exploratory data analysis to identify churn trends.
- Use SQL to query specific customer segments and metrics.
- Provide recommendations for reducing churn based on findings.

2 Dataset Overview

The dataset, sourced from a telecommunications company, contains 7,043 records with 21 attributes, including customer demographics, services, billing details, and churn status. Key columns include:

- customerID: Unique customer identifier.
- gender: Male or Female.
- SeniorCitizen: Whether the customer is a senior citizen (0 or 1).
- tenure: Months of service subscription.
- MonthlyCharges: Monthly billing amount.
- TotalCharges: Total amount billed.
- Churn: Whether the customer churned (Yes or No).
- Service-related columns (e.g., InternetService, OnlineSecurity, Contract).

3 Data Cleaning

Preparing the dataset for analysis involved several steps:

- Initial Inspection: The dataset was loaded using Pandas, revealing 7,043 rows and 21 columns. No missing values were initially detected.
- Type Conversion: The `TotalCharges` column, stored as strings, was converted to

numeric values, introducing 11 missing values due to invalid entries (e.g., empty strings).

- Handling Missing Data: Rows with missing `TotalCharges` were dropped, reducing the dataset to 7,032 records.
- Output: The cleaned dataset was saved as `cleaned_telco_data.csv` for further analysis and SQL import.

4 Exploratory Data Analysis

EDA was conducted using Python libraries (Pandas, Seaborn, Matplotlib) to uncover churn patterns.

4.1 Univariate Analysis

- Churn Distribution: Approximately 26.54% of customers churned (1,869 out of 7,032), indicating a significant retention challenge.
- Gender: Churn rates were similar across genders (Male: 26.2%, Female: 26.9%).
- Payment Method: Electronic check users had the highest churn rate (45.3% of churned customers).
- Total Charges: The distribution was right-skewed, with most customers having lower total charges.

4.2 Bivariate Analysis

- Monthly Charges vs. Churn: Churned customers had higher median monthly charges ($79.65) compared to non-churned ($61.35).
- Tenure vs. Churn: Churned customers had shorter tenure (median: 10 months) than non-churned (median: 38 months).
- Correlation Heatmap: Numeric variables showed weak correlations, with tenure and total charges having a moderate positive correlation (0.83).

4.3 Visualizations

Key visualizations included:

- Count plots for churn and gender distributions.
- Box plots for monthly charges and tenure by churn status.
- Histograms for total and monthly charges.
- A correlation heatmap for numeric variables.

5 SQL Analysis

The cleaned dataset was imported into a MySQL database (`telco_db`) for advanced querying. Key queries included:

- Churn Counts: Confirmed 1,869 churned and 5,163 non-churned customers.
- Average Monthly Charges by Contract: Month-to-month contracts had the highest average charges ($66.40), followed by one-year ($61.65) and two-year ($61.06).
- Churn by Internet Service: Fiber optic users had the highest churn count (1,297 out of 1,869 churned customers).
- Tenure Groups: Most customers (2,406) had 0–12 months tenure, with higher churn rates in this group.
- Window Functions: Ranked customers by monthly charges and calculated running totals of total charges.

6 Key Findings

1. High Churn in Short-Term Contracts: Month-to-month contract customers exhibited the highest churn rate (42.7% of churned customers).
2. Fiber Optic Churn: Customers with fiber optic internet churned at a higher rate (69.4% of churned customers), possibly due to higher costs.
3. Payment Method Impact: Electronic check users were more likely to churn, suggesting dissatisfaction or billing issues.
4. Tenure Effect: Customers with shorter tenure (0–12 months) were significantly more likely to churn.
5. Charge Sensitivity: Higher monthly charges correlated with increased churn risk.

7 Interesting Fact

An unexpected finding was the high churn rate among fiber optic users, despite the service being a premium offering. This may indicate pricing sensitivity or service quality issues, as fiber optic plans had higher monthly charges (median: $91.30) compared to DSL ($58.10) or no internet ($20.15).

8 Recommendations

Based on the analysis, the following strategies could reduce churn:

- Promote Long-Term Contracts: Offer incentives for one-year or two-year contracts to increase tenure and reduce churn.

- Address Fiber Optic Churn: Investigate service quality or pricing issues for fiber optic plans and offer bundled discounts.
- Improve Billing Experience: Streamline electronic check processes or promote automatic payment methods to reduce friction.
- Target New Customers: Implement retention programs for customers with 0–12 months tenure, such as loyalty discounts or personalized support.

9 Conclusion

This project, conducted by Neha Jhakra, successfully analyzed customer churn using Python and SQL, identifying key drivers such as contract type, tenure, monthly charges, and payment methods. The insights provide a foundation for data-driven retention strategies. Future work could involve predictive modeling to forecast churn probability and deeper segmentation of customer profiles.

10 References

- Telco Customer Churn Dataset.
- Python Libraries: Pandas, Seaborn, Matplotlib.
- MySQL for SQL querying.