Telecom Customer Churn Analysis & Prediction

1. Project Overview

Telecom Customer Churn Analysis & Risk Identification focuses on identifying customer churn patterns and high-risk customer segments using structured data analytics techniques. The project simulates a real-world telecom analytics workflow, combining SQL-driven data preparation, Python-based exploratory analysis (Jupyter Notebook), and interactive Power BI dashboards to support data-driven retention strategies.

2. Executive Summary

Customer churn is a major revenue challenge in the telecom industry, where acquiring a new customer costs **5–7× more** than retaining an existing one. This project analyzes historical telecom customer data to uncover churn drivers, quantify churn risk, and identify customers most likely to discontinue services.

The analysis reveals an **overall churn rate of 27%**, indicating a significant retention gap. Key churn drivers include **contract type, payment method, internet service quality, age group, and value-added service adoption**. Using pattern-based churn risk identification aligned with business analytics practices, **1,655 customers (~26%)** were classified as **high-risk churn candidates**.

The project enables telecom stakeholders to proactively target high-risk customers, potentially reducing churn-related revenue loss and improving customer lifetime value.

3. Business Problem

In highly competitive telecom markets, customer churn directly impacts:

- Revenue stability
- Customer lifetime value (CLV)
- Marketing and acquisition costs

Despite competitive pricing and service bundles, telecom providers continue to experience churn due to:

- Short-term contracts
- Payment friction
- Service dissatisfaction
- Low customer engagement

Most organizations react **after churn occurs**, resulting in lost revenue. This project addresses the need for **early churn risk identification**, enabling **proactive and targeted retention strategies**.

4. Project Objectives

The key objectives of this project are to:

- Quantify historical churn and identify behavioral patterns
- Analyze churn drivers across demographics, services, and billing behavior
- Segment customers based on churn risk indicators
- Identify high-risk customers for proactive retention planning
- Deliver executive-ready insights through interactive dashboards

5. Dataset Overview

- **Total Customers:** 6,418
- **Granularity:** One record per customer

Data Categories

- **Demographics:** Gender, age group, marital status, state
- **Service Details:** Contract type, internet service, phone service, value-added services
- **Billing & Payments:** Monthly charges, total charges, payment method
- **Target Variable:** Churn (Yes / No)

The binary churn variable makes the dataset suitable for churn analysis, segmentation, and risk identification.

6. Data Cleaning & Preprocessing

Comprehensive preprocessing was performed to ensure data accuracy and analytical reliability.

Key Steps

- Dataset inspection and schema validation
- Removal of duplicate customer records
- Treatment of missing values without data loss
- Conversion of billing fields into numeric formats
- Standardization of categorical variables

- Feature selection focused on churn-relevant attributes

SQL-Based Data Preparation

SQL was used to:

- Calculate churn metrics
- Segment customers by contract, service, and payment method
- Generate structured datasets for dashboard consumption

Outcome: A clean, consistent, and analysis-ready dataset suitable for EDA and churn risk identification.

7. Exploratory Data Analysis (EDA)

EDA was conducted to understand customer behavior and identify churn-related trends.

Key Insights

- **Overall churn rate:** 27%
- **Gender:** No significant churn difference across genders
- **Contract type:** Month-to-month contracts exhibit the highest churn; two-year contracts the lowest
- **Payment method:** Manual payment users churn significantly more than automated payment users
- **Internet service:** Fiber optic customers show higher churn than DSL and no-internet users
- **Age group:** Customers aged **50+** demonstrate higher churn risk
- **Value-added services:** Customers with multiple add-ons churn less frequently

These insights highlight clear behavioral and service-level churn drivers.

8. Dashboard Development

Two interactive Power BI dashboards were developed to support executive and operational decision-making.

Executive Summary Dashboard

Key KPIs

- Total Customers: 6,418
- New Customers: 411
- Total Churned Customers: 1,732
- Churn Rate: 27%
- Avg. Monthly Charges (Churned Customers): ₹73

Figure 1: Executive Summary Dashboard showing overall churn rate, total customers, and churn distribution across key customer segments

This dashboard provides a high-level view of churn distribution across states, contracts, payment methods, and services, enabling faster decision-making.

Churn Risk & High-Risk Insights Dashboard

- **High-Risk Customers Identified:** 1,655

High-risk customers are predominantly:

- On month-to-month contracts
- Using bank withdrawal or mailed check payments
- Subscribed to fiber optic internet
- Belonging to the 50+ age group

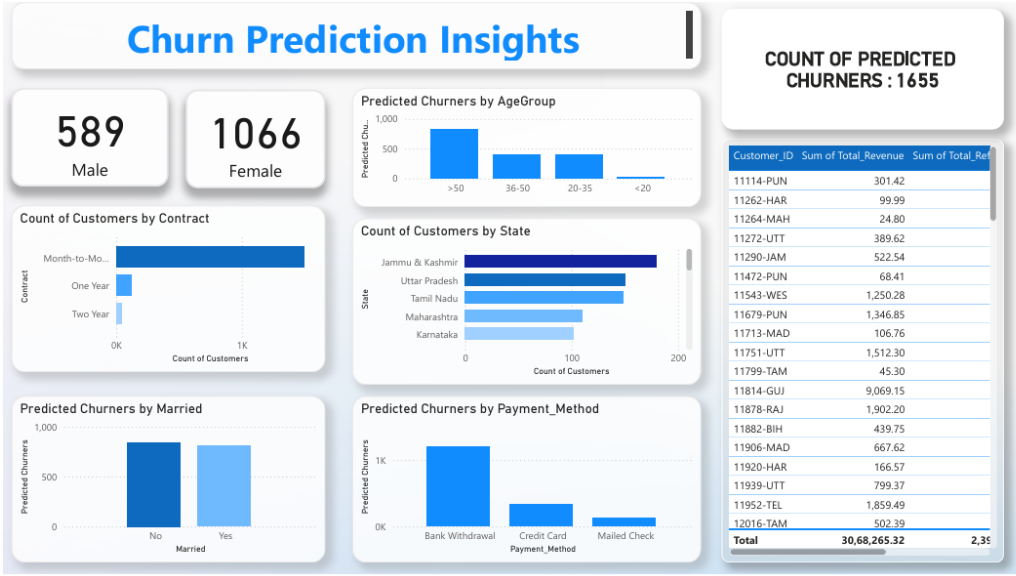

Figure 2: Churn Prediction Dashboard highlighting high-risk customers based on contract type, payment method, and service attributes.

Interactive filters allow stakeholders to drill down by demographic, service, and billing attributes.

9. Churn Risk Identification Approach

Instead of deploying a black-box machine learning model, this project applies **pattern-based churn risk identification**, reflecting real-world business analytics practices.

Methodology

- Identified churn-influencing features through EDA
- Extracted high-risk behavioral patterns from historical churn data
- Flagged customers exhibiting similar risk profiles

Outcome

- **1,655 customers** identified as high-risk churn candidates
- Enables focused retention campaigns rather than blanket discounts

10. Key Business Findings

- Short-term contracts are the strongest churn driver
- Manual payment methods significantly increase churn probability
- Internet service quality directly affects retention
- Value-added services improve customer engagement and loyalty

- Older customers require tailored retention strategies

11. Business Recommendations

- Incentivize long-term contract adoption through loyalty pricing
- Promote automated payment methods to reduce payment friction
- Improve service quality for high-churn internet segments
- Bundle value-added services to increase customer stickiness
- Design targeted retention campaigns for high-risk customer groups

A **2–3% reduction in churn** could potentially save **millions annually** for mid-scale telecom providers.

12. Conclusion

This project demonstrates how structured analytics, SQL-based data preparation, and interactive dashboards can solve a real-world telecom churn problem. By combining historical analysis with churn risk identification, the solution supports proactive retention planning and revenue protection.

13. Tools & Technologies

- **SQL:** Data extraction, aggregation, churn metrics
- **Python (Jupyter Notebook):** Exploratory analysis and pattern identification
- **Excel:** Data validation and preprocessing
- **Power BI:** Interactive dashboards and business reporting

14. Limitations & Future Scope

Limitations

- The analysis is based on **historical customer data**, which may not fully capture sudden market or policy changes.
- Churn risk identification is **pattern-based** and does not use real-time behavioral signals such as call drops, network latency, or customer support interactions.
- External factors like **competitor pricing, regional outages, and promotional campaigns** were not included due to data unavailability.
- Customer tenure was analyzed at an aggregate level; finer time-series behavior could improve churn sensitivity.

Future Enhancements

- Integrate **real-time usage and network performance data** for more accurate churn detection.
- Implement **machine learning models** (Logistic Regression / Random Forest) to predict churn probabilities.
- Add **financial impact modeling** to estimate revenue saved through churn reduction.
- Automate dashboard refresh using scheduled data pipelines.