Customer Churn Prediction using Business Analytics & Machine Learning
Independent Analytics Case Study

Author: Amiya Ghosh
Tools: Python, Machine Learning, Power BI
Year: 2025

❖ Executive Summary

Customer churn is a major challenge for subscription-based businesses, as it directly affects revenue stability and customer lifetime value. This project focuses on building a data-driven churn prediction framework that identifies customers at high risk of churn and translates analytical insights into actionable business decisions.

Using historical customer data, exploratory analysis and machine learning techniques were applied to estimate churn probability and segment customers into risk categories. Logistic Regression was selected as the final model due to its strong predictive performance and interpretability. The analytical results were operationalized through an interactive Power BI dashboard to support business teams in prioritizing retention efforts.

This case study demonstrates an end-to-end analytics workflow, from business problem definition to model-driven insights and decision support.

❖ Business Context

In highly competitive industries such as telecommunications and subscription-based services, retaining customers is often more cost-effective than acquiring new ones. However, many organizations rely on reactive retention strategies, addressing churn only after customers disengage.

Predictive analytics enables businesses to identify early churn signals and intervene proactively. By understanding which customers are most likely to churn and why, organizations can design targeted retention strategies and optimize resource allocation.

❖ Problem Definition

Many organizations lack a systematic, data-driven mechanism to identify customers who are likely to churn before the event occurs. As a result, retention strategies are often applied uniformly across customers, leading to inefficient use of marketing and customer support resources.

This project addresses the need for a predictive churn analytics solution that:

- Identifies customers at risk of churn in advance
- Explains the key drivers behind churn behavior
- Supports targeted and actionable retention decisions

❖ Objectives

The objectives of this analysis are to:

- Analyze customer behavior and identify factors influencing churn
- Build a predictive model to estimate churn probability
- Segment customers into churn risk categories
- Present insights through an interactive dashboard
- Provide actionable recommendations to reduce churn

❖ Dataset Overview

The analysis uses the **IBM Telco Customer Churn dataset**, which contains customer-level information from a telecommunications service provider.

➢ **Dataset summary:**

- Total customers: 7,043
- Churned customers: 1,869
- Overall churn rate: 27%

➢ **Key attributes include:**

- Customer tenure
- Contract type
- Internet and service subscriptions
- Monthly charges
- Payment method
- Churn indicator (1 / 0)

The dataset combines behavioral, contractual, and billing variables, making it suitable for churn prediction analysis.

❖ Analytical Approach

A structured analytics lifecycle was followed:

➢ **Data Preparation**

- Verified data structure, missing values, and duplicates
- Corrected data type inconsistencies (e.g., TotalCharges column)

➢ **Exploratory Data Analysis (EDA)**

- Analyzed churn patterns across tenure, contract type, and charges
- Identified key behavioral trends influencing churn

- **Feature Engineering**
    - Encoded categorical variables
    - Created tenure-based groupings
    - Prepared a model-ready dataset

- **Model Development & Evaluation**
    - Logistic Regression
    - Advanced machine learning model evaluated for comparison
    - Model performance assessed using ROC-AUC

- **Explainability & Segmentation**
    - Identified key churn drivers
    - Generated churn probabilities
    - Segmented customers into actionable risk groups

- **Visualization**
    - Developed a Power BI dashboard to communicate insights

Key Insights from Exploratory Analysis

The exploratory analysis revealed several important patterns:

- Customers with month-to-month contracts exhibit significantly higher churn
- Early-tenure customers (0-12 months) show the highest churn rates
- Higher monthly charges are associated with increased churn probability
- Contract duration and tenure are among the strongest churn indicators

These insights informed feature selection and model development.

Model Selection & Evaluation

Multiple models were evaluated to predict customer churn. Logistic Regression achieved the **highest ROC-AUC score** among the tested models and was selected as the final model.

In addition to strong performance, Logistic Regression offers high interpretability, making it suitable for explaining churn drivers to business stakeholders and supporting transparent decision-making.

Explainability & Risk Segmentation

Model explainability techniques were applied to identify the most influential churn drivers, including:

- Contract type
- Customer tenure

- Monthly charges
- Internet service type

Based on churn probability scores, customers were segmented into:

- High Risk
- Medium Risk
- Low Risk

This segmentation enables differentiated retention strategies instead of one-size-fits-all interventions.

❖ Dashboard & Business Action Layer

An interactive Power BI dashboard was developed to translate analytical outputs into business-friendly insights.

➢ **Dashboard highlights include:**

- Total customers: 7,043
- Churned customers: 1,869
- Overall churn rate: 27%
- Churn by contract type and tenure
- Risk segment distribution
- Customer-level churn probability table

➢ **Business-Action-Insight:**

Retention efforts should prioritize high-risk, early-tenure customers on month-to-month contracts, as this segment represents the most immediate opportunity to reduce churn.

❖ Recommendations

Based on the analysis, the following actions are recommended:

- High-Risk-Customers:
  Proactive retention campaigns, personalized offers, and priority customer support.

- Early-Tenure-Customers:
  Improved onboarding experience and early engagement initiatives.

- Month-to-Month-Contracts:
  Incentives to encourage migration to long-term contracts.

These targeted actions can help reduce churn and improve customer lifetime value.

❖ Conclusion

This case study demonstrates a complete, end-to-end customer churn prediction framework using business analytics and machine learning. By integrating data preparation, predictive modeling, explainability, and dashboard-driven insights, the project supports proactive and targeted retention strategies.

The use of Logistic Regression as the final model ensures strong predictive performance while maintaining interpretability, making the solution practical for real-world business applications.

❖ Future Enhancements

Potential extensions of this work include:

- Real-time churn prediction
- Survival analysis to estimate time-to-churn
- Integration with CRM systems
- Experimental testing of retention strategies

❖ Project Artifacts

This project is supported by the following artifacts:

- ✓ **Jupyter_Notebooks:**
  Five structured notebooks covering data preparation, exploratory analysis, feature engineering, model development, explainability, and business insights.
- ✓ **Power_BI_Dashboard:**
  An interactive dashboard visualizing churn KPIs, risk segmentation, and customer-level churn probability.
- ✓ **Source_Code_&_Data:**
  All code, processed datasets, and visual assets are available in the accompanying GitHub repository.

Link: https://github.com/AnalytixAI/Customer-Churn-Prediction