Predicting Customer Churn: A Data-Driven Analysis

Leveraging Data Analytics to Drive Business Retention

September 2025

Executive Summary

Objective

Identify key drivers of customer churn and develop a predictive model to proactively flag at-risk customers.

Methodology

Comprehensive data analysis using SQL, Python visualisation libraries (Seaborn, Folium, Plotly), and machine learning classification models.

Business Impact

Achieved 82% accuracy, enabling targeted retention campaigns that could save significant revenue by reducing customer attrition.

The Customer Retention Challenge

The Problem

Customer churn poses a critical challenge for subscription-based businesses. Research shows that acquiring a new customer costs 5 to 25 times more than retaining an existing one.

Project Goals

- Analyse historical customer data to understand why customers leave

- Build a reliable predictive model to identify who is likely to leave

- Enable proactive retention strategies

ⓘ **Dataset Overview:** 7,043 customer records with 21 attributes including demographics, account information, services subscribed, and payment data.

Data Foundation & Preparation

01

Data Collection

Sourced comprehensive customer records from centralised SQL database containing demographics and service history.

02

Data Cleaning

Handled missing values in 'Total Charges' using tenure-based imputation and corrected data types for accurate calculations.

03

Feature Engineering

Created 'Tenure in Years' column and removed duplicate entries to ensure data integrity and improved interpretability.

Exploratory Data Analysis Methodology

Uncovering Patterns & Relationships

SQL Analysis

Complex data aggregation and querying directly from database for deep customer insights and churn pattern identification.

Statistical Visualisation

Matplotlib & Seaborn for static visualisations including bar charts, histograms, and correlation heatmaps.

Geographic Analysis

Folium interactive choropleth maps visualising churn rates across different geographic locations and regions.

Interactive Dashboards

Plotly & Dash for comprehensive, dynamic data exploration allowing stakeholders to filter and analyse in real-time.

Predictive Analysis Framework

Model Development Process

1

Data Preprocessing

One-hot encoding for categorical features and StandardScaler for numerical standardisation to prevent model bias.
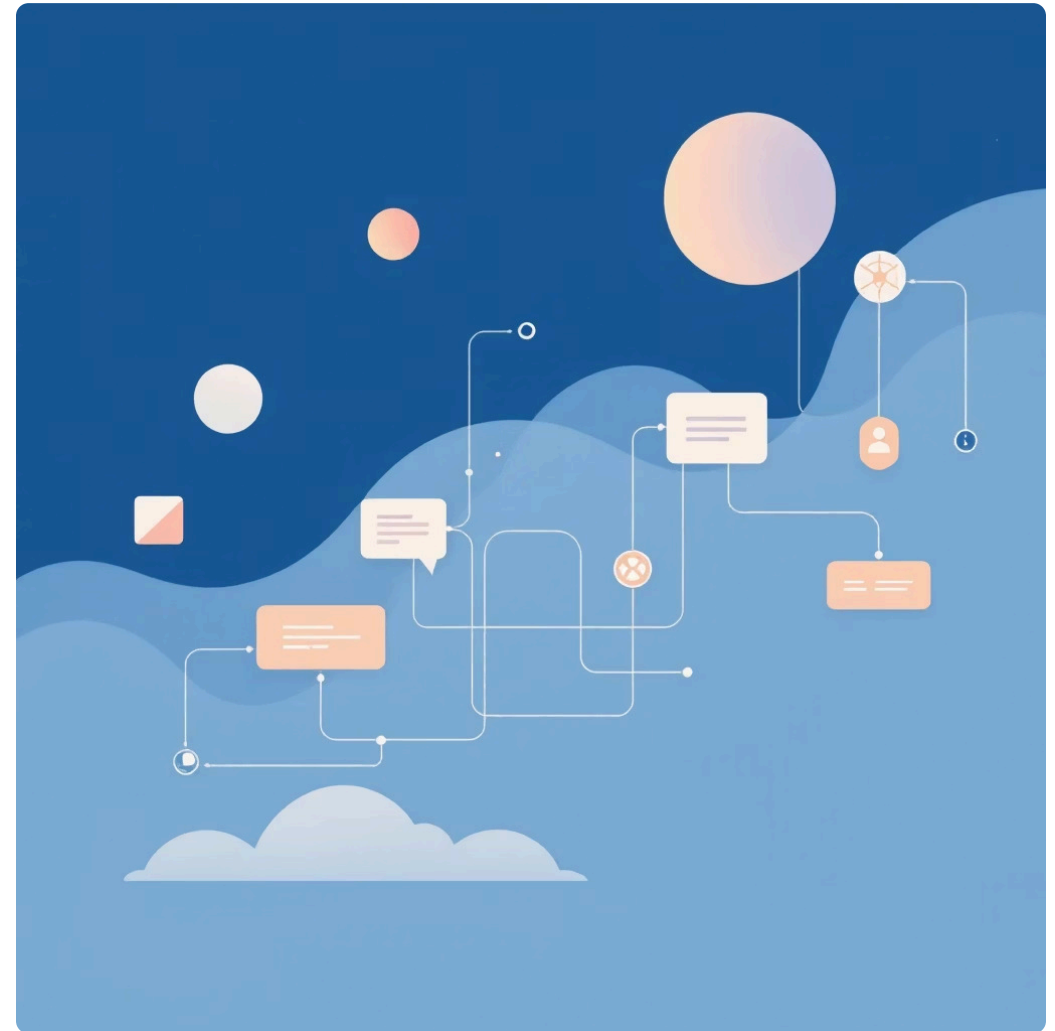
2

Model Selection

Evaluated Logistic Regression and Random Forest algorithms, selecting Random Forest for superior performance.
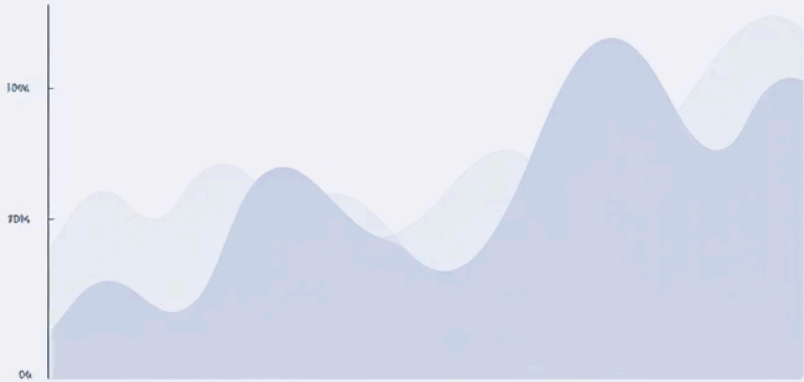
3

Performance Evaluation

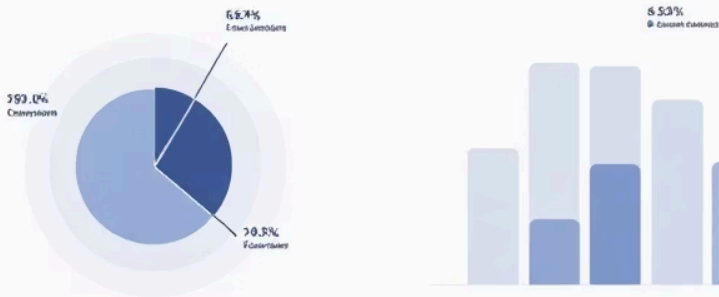Assessed using accuracy, precision, recall, and F1-score to account for class imbalance in churn data.

Key Customer Insights from Data Analysis

1

Contract Type Impact

Month-to-month contracts show significantly higher churn rates compared to one-year or two-year contracts, indicating commitment importance.

2

Tenure Patterns

Customers with tenure under 12 months demonstrate highest churn probability, highlighting critical early relationship period.

3

Pricing Correlation

Higher monthly charges strongly correlate with increased churn likelihood, suggesting price sensitivity amongst customer base.

SQL Analysis Reveals Critical Patterns

Internet Service Analysis

Fibre optic customers churn at higher rates, potentially due to increased costs or service stability issues requiring investigation.

Service Bundle Impact

Customers with fewer additional services show higher churn rates, demonstrating that bundled services significantly increase customer loyalty.

Payment Method Insights

Electronic cheque users exhibit highest churn rates, suggesting potential payment process friction requiring operational review.

Geographic Churn Distribution Analysis

Interactive Regional Mapping

Created comprehensive choropleth map visualising customer churn rates across different states and regions using Folium technology.

- Hover functionality displays specific churn rates and customer totals
- Identified regional hotspots with unusually high churn
- Revealed potential local network infrastructure issues
- Highlighted areas with increased competition impact

Geographic analysis enables targeted regional retention strategies and infrastructure investment decisions.

Predictive Model Success & Strategic Recommendations

82%
Model Accuracy
Successfully predicts customer churn status

68%
Precision Rate
Correct positive churn predictions

59%
Recall Rate
Actual churners successfully identified

Actionable Business Recommendations

1 **Contract Incentives**

Target month-to-month customers with promotions for longer-term contracts to reduce churn risk significantly.

2 **Enhanced Onboarding**

Improve first-year customer experience through dedicated support and engagement programmes to build early loyalty.

3 **Service Bundle Strategy**

Encourage adoption of additional services like tech support and online security to increase customer stickiness.