Telecom Customer Churn Insights – SQL, Python & Tableau

Author: Haythem Abouda

Date: 21/03/2025

“"Analyzing customer churn patterns to uncover retention opportunities using SQL, Python, and Tableau."”

📌 **Tableau Dashboard:**
You can explore the **interactive visualizations** here.
Click through different views to analyze key insights!

📌 **SQL & Python Scripts:**
All queries, data cleaning, and analysis scripts are available in my **GitHub repository** here.
This includes raw SQL queries and Python scripts for data cleaning and preprocessing.

Executive Summary

This project explores customer churn patterns for a telecom company using SQL, Python, and Tableau.

The goal was to identify which customer segments are at highest risk of leaving, and what factors are driving that behavior.

Using SQL, we analyzed churn in relation to contract type, payment method, internet service, and tech support availability.

Python was used for additional data preparation and cleaning, and Tableau helped present the findings in a clear, interactive dashboard.

The analysis revealed that customers on month-to-month contracts, using electronic check payments, or lacking tech support had the highest churn rates. These insights provide a data-driven foundation for building targeted customer retention strategies.

Dataset Overview

The dataset used in this project is an open-source customer churn dataset provided by IBM, containing information on **7,043 telecom customers**. Each row represents a customer, and columns include demographic data, service subscriptions, billing information, and whether the customer churned.

🧾 Key Dataset Features:

- **Demographics**: Gender, Senior Citizen, Partner, Dependents
- **Services**: Phone, Internet, Online Security, Tech Support, Streaming
- **Billing**: Contract Type, Paperless Billing, Payment Method, Monthly & Total Charges
- **Target Variable**: Churn (Yes/No)

🧼 Data Cleaning Highlights:

- `TotalCharges` was stored as a string and contained missing values → converted to numeric
- All service-related columns were standardized (e.g., removing variations like “No internet service”)
- Missing or invalid entries were cleaned or excluded as needed during SQL analysis
- Columns were renamed for readability (e.g., `customerID` → `customer_id`, `SeniorCitizen` → `senior_citizen`)

This dataset provided a realistic business scenario for analyzing customer behavior and identifying churn risk factors.

SQL Analysis

The SQL portion of this project focused on answering key business questions using raw customer data from the database. We explored churn behavior across service types, contract lengths, and payment methods to help the company prioritize its retention strategy.

🧩 Business Questions Answered with SQL

- What is the **overall churn rate**?
- How does churn vary by **contract type**?
- Which **internet service types** have the highest churn?
- What is the churn rate for **electronic check vs. other payment methods**?
- Do customers with **tech support** churn less?

🧠 Key Results

Segment	Churn Rate
Overall	26.5%
Month-to-Month Contracts	42.7%
One-Year Contracts	11.3%
Two-Year Contracts	2.8%
Fiber Optic Users	41.3%
DSL Users	23.8%
No Internet Users	7.4%
Electronic Check	47.5%
Credit Card (Auto)	16.0%
Bank Transfer (Auto)	18.0%
Mailed Check	22.0%
Without Tech Support	36.9%
With Tech Support	14.0%

🛠️ Approach

- SQL queries were written to group and count customers by category (e.g., contract type, payment method)
- Churn rate was calculated as:

$$\text{Churn Rate} = \left(\frac{\text{Number of Churned Customers}}{\text{Total Customers in Segment}}\right) \times 100$$

- NULL values and data inconsistencies (e.g., missing `total_charges`) were handled using filtering and basic data cleaning in SQL

Python Preparation & Exploration

Python was used after SQL to prepare the dataset for Tableau and ensure all fields were clean and usable for visual analysis. While no advanced machine learning was used, Python played a key role in shaping the data into a presentable and consistent format.

🧼 Cleaning & Prep Steps

- Loaded raw CSV data using `pandas`
- Identified and handled **null values** in `total_charges`
- Removed invalid entries (e.g., rows where `total_charges` was blank or non-numeric)
- Converted `total_charges` and `monthly_charges` to **numeric types**
- Renamed columns to follow consistent **snake_case formatting** (e.g., `SeniorCitizen` → `senior_citizen`)
- Verified that all columns had correct data types (`int`, `float`, `object`)
- Exported the cleaned dataset for Tableau dashboard creation

---

🛠️ Key Tools Used

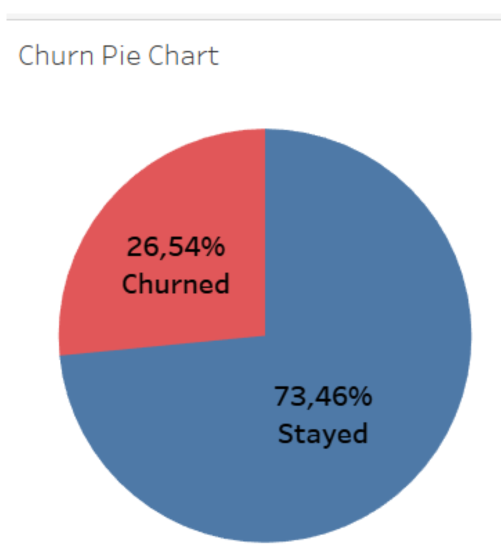
- **Pandas**: For loading, cleaning, and inspecting the dataset
- **Jupyter Notebook**: As a sandbox for testing and documenting steps
- **SQLAlchemy**: To connect to the PostgreSQL database from Python

Tableau Dashboard

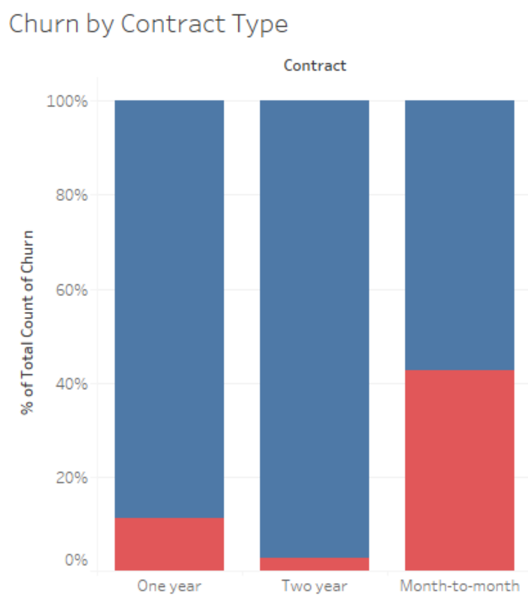
After preparing the cleaned dataset in Python, we used Tableau to build an interactive dashboard that highlights churn patterns across different customer segments. The dashboard was designed to be clean, executive-friendly, and focused on actionable insights.

📊 Dashboard Sections & Charts

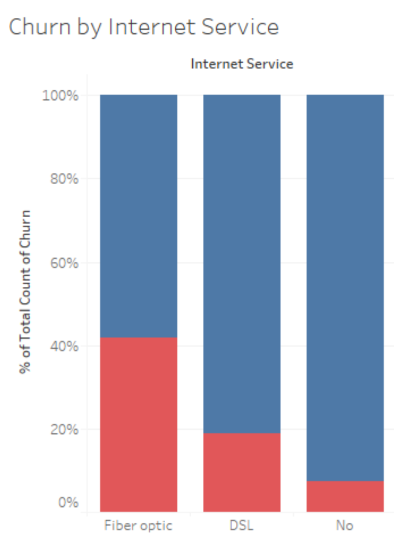
1. **Churn Overview**
    - A pie chart showing the overall churn rate (~26.5%)
    - Clear distinction between churned and retained customers



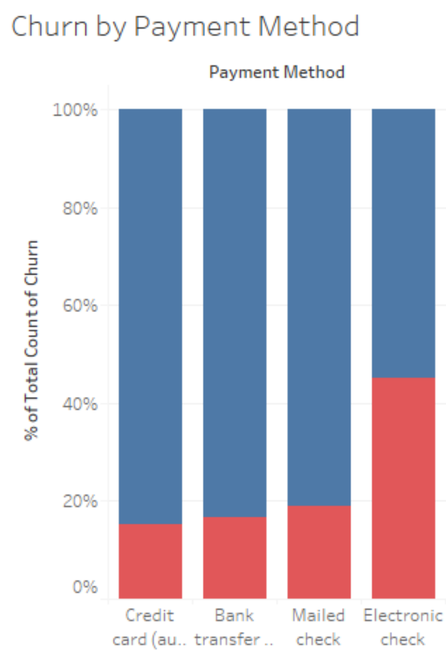
2. **Churn by Contract Type**
    - Shows that **month-to-month contracts** had the highest churn rate
    - One- and two-year contracts showed significantly lower churn



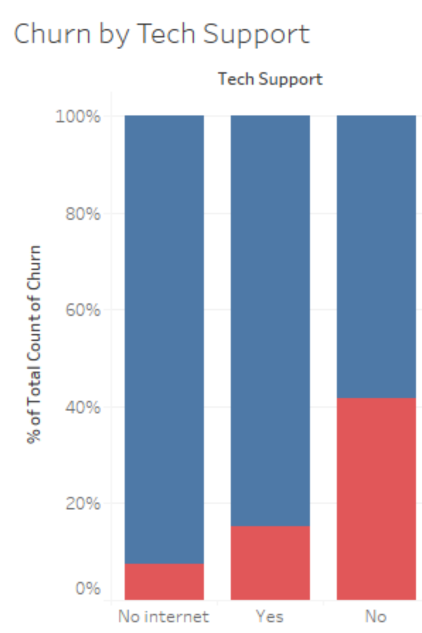
3. **Churn by Internet Service**
    - **Fiber optic users** churned more than DSL or no-internet users
    - Insight: Internet type correlates strongly with churn behavior



4. **Churn by Payment Method**
    - **Electronic check** customers had the highest churn (~47.5%)
    - Auto-payment methods had lower churn rates



5. **Churn by Tech Support**
    - Customers without tech support churned at **~36.9%**, vs. just **~14%** for those with support



📈 Design Notes

- Used **stacked bar charts** to show churn vs. stayed
- Maintained consistent color coding (e.g., red for churned)
- Dashboard layout allows for quick, visual comparison of churn drivers

Key Business Insights

The analysis surfaced several clear patterns that can directly inform customer retention strategies:

- 📉 **Churn is concentrated among customers with short-term, flexible contracts.**
  - Customers on month-to-month plans churned at **42.7%**, while two-year contracts had just **2.8%** churn.
- 💳 **Payment method strongly predicts churn.**
  - **Electronic check users churned at 47.5%**, significantly more than users with auto-pay methods.
- 🌐 **Internet service type matters.**
  - **Fiber optic users churned at 41.3%**, compared to just **7.4%** for those without internet service.
- 🛠️ **Tech support helps retention.**
  - Churn was **36.9% without tech support**, but only **14.0%** for those who had it.
- 🧭 **Combining risk factors increases churn.**
  - Customers with **month-to-month contracts**, **no tech support**, and **electronic check payments** had the highest churn probability

Conclusion & Next Steps

This churn analysis helped uncover the customer segments most at risk of leaving, enabling the telecom company to focus its retention efforts where they matter most.

By combining SQL, Python, and Tableau, we delivered a complete data pipeline, from raw data cleaning to executive-ready visual insights.

✅ Strategic Next Steps for the Business:

- Offer **incentives for customers on month-to-month contracts** to upgrade to longer-term plans
- Promote **auto-payment options** to reduce churn from electronic check users
- Bundle **tech support** with internet plans to improve retention
- Target **fiber optic customers** with specialized loyalty offers