Telco Customer Churn Analysis

# Executive Summary

This project investigates customer churn within a telecommunications company using the Telco Customer Churn dataset. The goal was to identify key churn drivers through exploratory data analysis (EDA) and to build a visually insightful dashboard for stakeholders using Streamlit. Data cleaning, feature engineering, and visual storytelling were core components of the workflow. The final product allows business teams to explore churn patterns interactively and formulate strategic decisions to reduce customer attrition.

1. Introduction

Customer churn, the phenomenon of customers leaving a service provider, represents a significant threat to telecom companies due to the recurring revenue model. The ability to identify patterns and proactively reduce churn can result in substantial long-term financial gains. This project aims to provide a data-driven analysis of churn behavior using publicly available data, culminating in an interactive dashboard for business use.

2. Dataset Overview

The dataset contains 7043 customer records with 21 attributes including demographic information, account details, service subscriptions, and churn status. Features include 'gender', 'SeniorCitizen', 'tenure', 'Contract', 'MonthlyCharges', 'TotalCharges', and the target variable 'Churn'.

3. Data Cleaning Process

The data cleaning process began by identifying missing values, incorrect types, and irrelevant records. 'TotalCharges' had invalid strings due to blank inputs, which were handled using `pd.to\_numeric()` with coercion. Null values (11 records) were removed. Duplicates were checked and confirmed absent. Columns were standardized using lowercase formatting to avoid case-related bugs. 'customerID' was retained as a unique identifier. The cleaned data was saved as `cleaned\_telco\_churn.csv`. Later, the 'gender' column was re-merged using the original dataset via a left join on 'customerID'.

4. Exploratory Data Analysis (EDA)

EDA revealed that customers with month-to-month contracts, fiber optic internet, and electronic check payments were more likely to churn. Senior citizens and customers with high monthly charges also showed increased churn risk. Tenure analysis showed that most churners left early, within their first year. Correlation analysis showed weak linear relationships overall, suggesting non-linear modeling might be required for predictions.

5. Dashboard Implementation

A Streamlit dashboard was built to visualize churn patterns dynamically. Filters allow exploration by gender and internet service. KPIs like churn rate and customer count are displayed. Charts such as churn by contract, tenure distribution, and scatterplots of monthly vs. total charges give users business insight. A dark theme and layout structure similar to Power BI enhance readability.

6. Key Insights

- Month-to-month customers churn most frequently.  
- Tenure under 12 months correlates with higher churn.  
- Fiber optic internet service customers show elevated churn rates.  
- Senior citizens and electronic check payers are high-risk groups.  
- Customers with high monthly charges but short tenure are vulnerable.

7. Recommendations

- Introduce loyalty incentives for early-tenure customers.  
- Encourage migration to annual contracts through discounts.  
- Improve fiber service reliability or bundle with additional value.  
- Offer assistance or training to senior citizens.  
- Consider redesigning onboarding for the first 3-month experience.

8. Reflections & Learning

This project was a comprehensive journey into handling real-world datasets: from cleaning and merging to insight generation and dashboard delivery. I learned the importance of understanding the business context, asking the right analytical questions, and communicating results visually. Working through filtering issues, type mismatches, and aesthetic design deepened my Python and Streamlit skills.