PROJECT DOCUMENTATION

Step 1: Exploratory Data Analysis & Visualization

Project: AI-Powered Customer Retention (Churn Prediction) System
Dataset: Telco Customer Churn (Kaggle)
(https://www.kaggle.com/datasets/blastchar/telco-customer-churn)

---

1. Objective of Step 1

The objective of Step 1 is to **understand customer churn behavior** using exploratory data analysis (EDA) and visual analytics before building any machine learning model.

This step focuses on:

- Understanding the dataset structure and features
- Identifying key factors contributing to customer churn
- Gaining business insights through Excel analysis and Tableau visualizations
- Building an interactive dashboard to summarize findings

---

2. Dataset Overview

The dataset was downloaded from Kaggle and contains **7043 customer records** related to a telecom service provider.

Key Characteristics:

- Each row represents **one customer**
- Each column represents a **customer attribute**
- The target variable is **Churn (Yes / No)**

Data Sources Used:

- Customer demographics (gender, senior citizen, dependents)
- Service details (internet service, phone service, add-ons)
- Contract and billing information
- Payment methods
- Usage duration (tenure)

---

3. Initial Data Understanding Using Excel

The dataset was first opened and analyzed using **Microsoft Excel** to understand raw data behavior.

Actions Performed:

- Verified total number of records and columns
- Checked for missing and inconsistent values
- Identified data types for each column
- Converted `TotalCharges` from string to numeric format
- Applied filters to understand churn patterns

Key Observations:

- Some customers with very low tenure had missing `TotalCharges`
- Churn column is categorical with values “Yes” and “No”
- Dataset is moderately imbalanced, with fewer churned customers compared to retained customers

---

4. Exploratory Analysis Using Excel (Pivot Tables)

Pivot tables were created to analyze churn across multiple dimensions.

Analysis Performed:

- Churn by Gender
- Churn by Contract Type
- Churn by Internet Service
- Churn by Payment Method

Initial Insights:

- Customers on **month-to-month contracts** show significantly higher churn
- **Fiber optic** internet users churn more compared to DSL users
- Customers using **electronic check** payment method have higher churn
- Gender does not show a very strong influence on churn

---

5. Data Visualization Using Tableau

After Excel analysis, the dataset was imported into **Tableau** for advanced visual analysis and dashboard creation.

Purpose of Using Tableau:

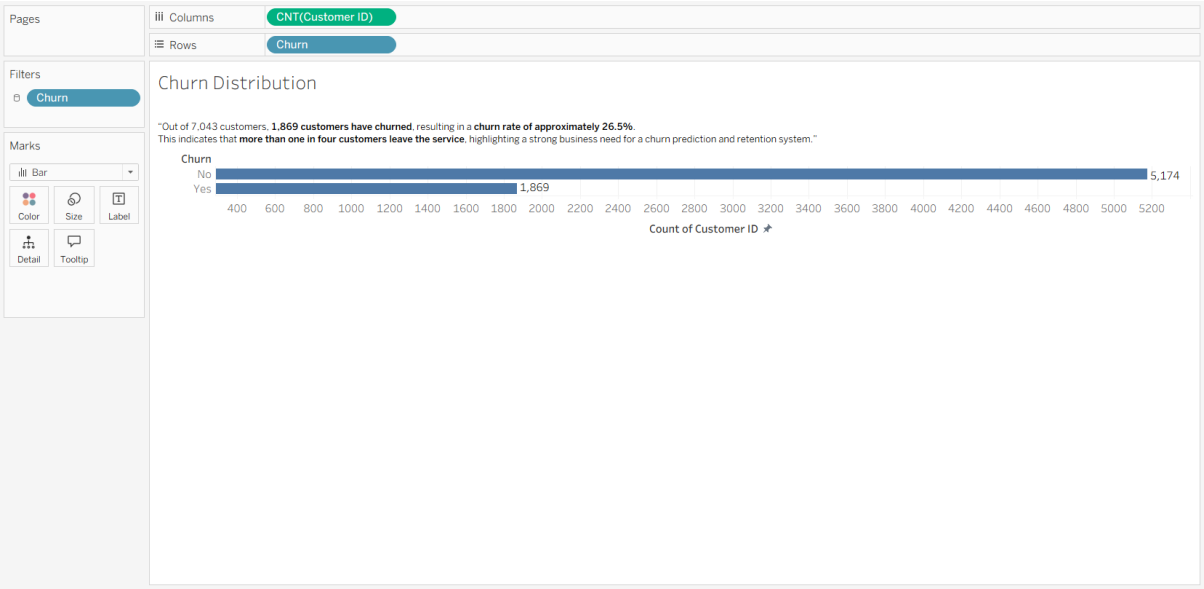
- Visual identification of churn patterns
- Interactive filtering for business users
- Clear communication of insights

---

6. Visualizations Created

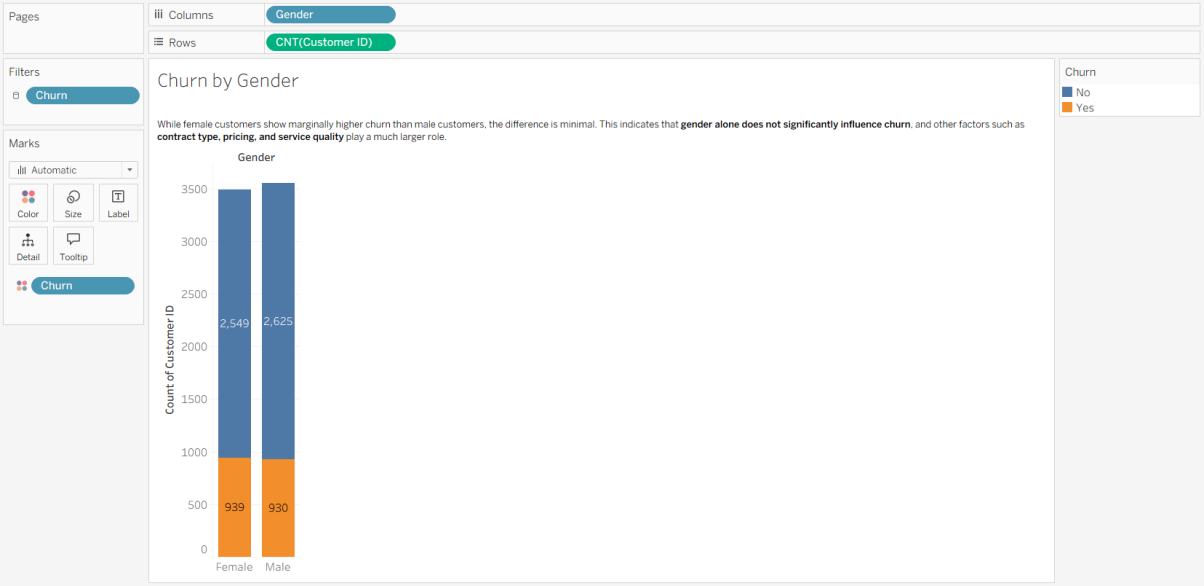
The following five core visualizations were created to answer key business questions:

6.1 Churn Distribution



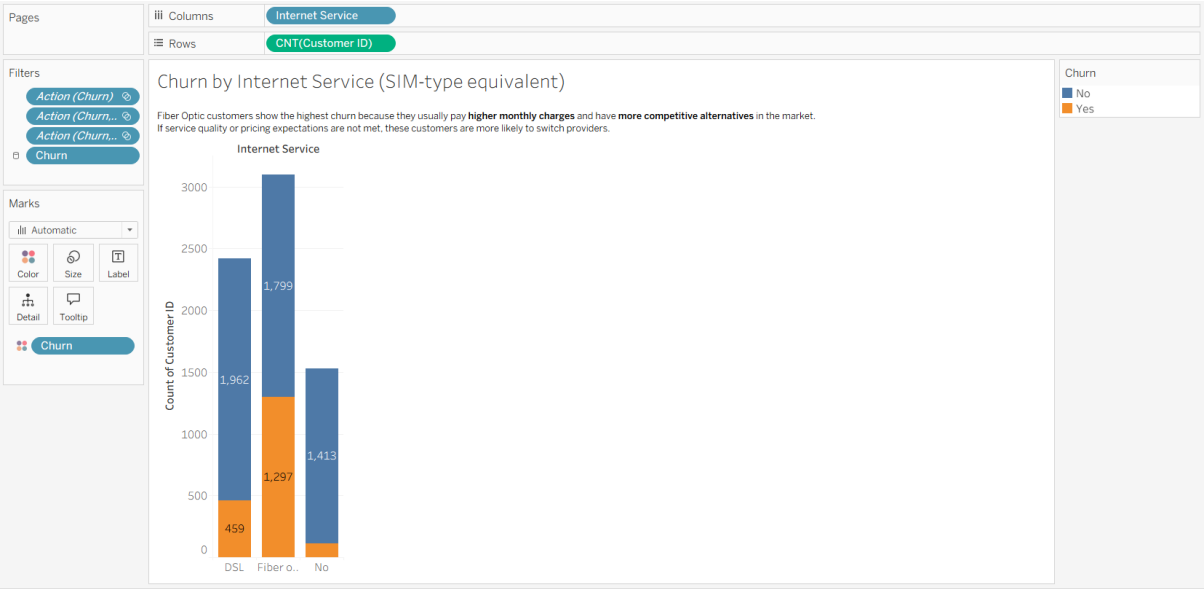
- Shows the proportion of churned vs retained customers
- Helps understand the scale of the churn problem

6.2 Churn by Gender



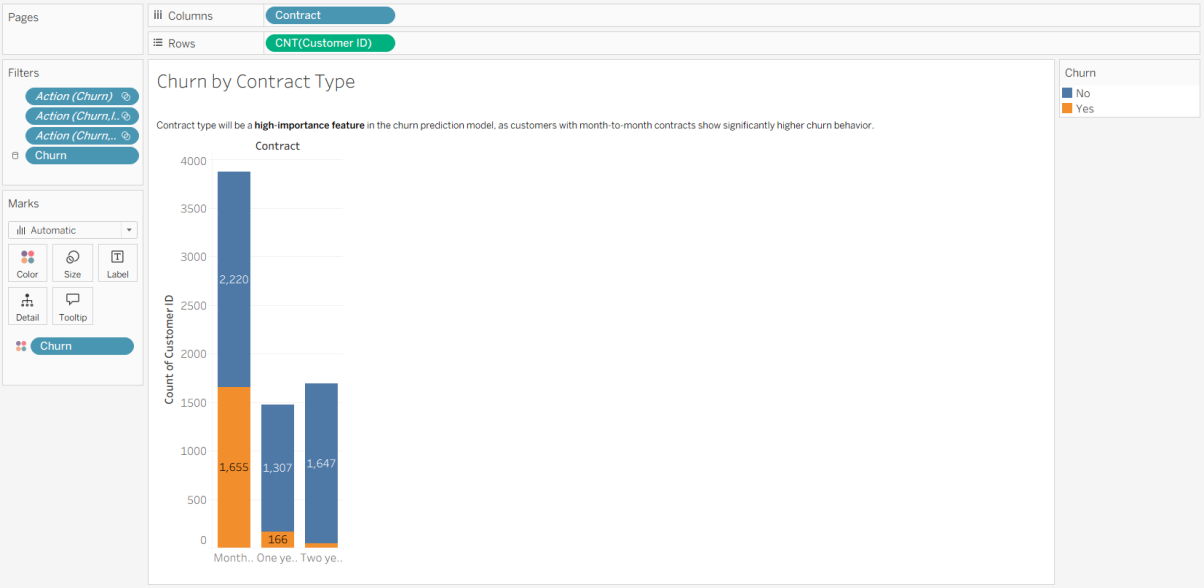
- Compares churn rates between male and female customers
- Used to check demographic influence

6.3 Churn by Internet Service



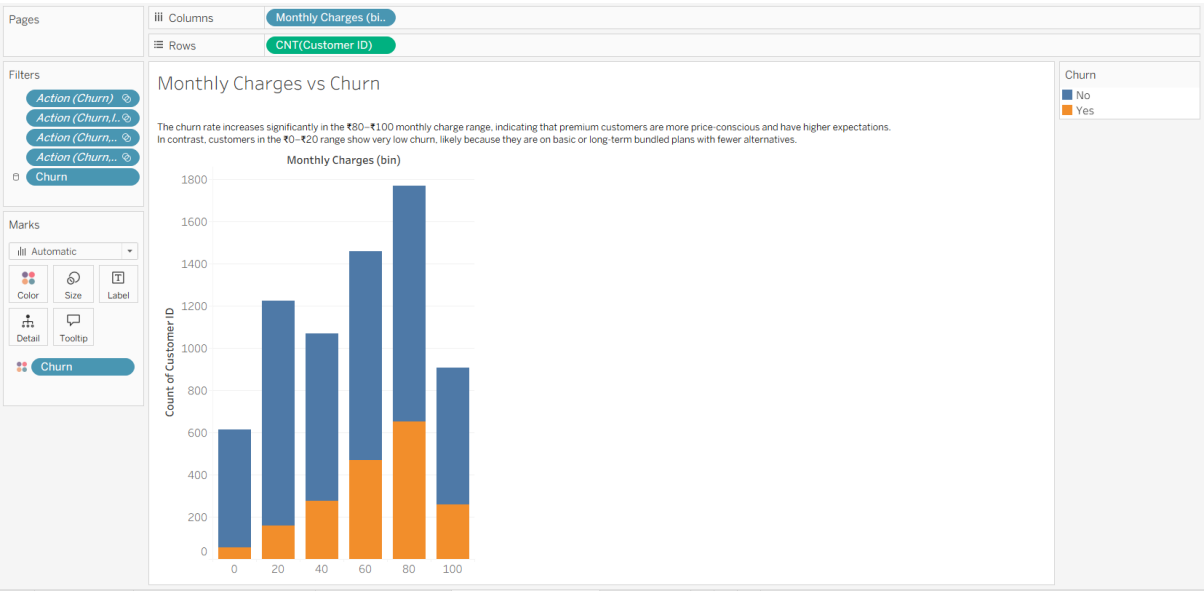
- Compares churn across DSL, Fiber optic, and No internet
- Identifies service-related dissatisfaction

6.4 Churn by Contract Type



- Analyzes churn across Month-to-month, One-year, and Two-year contracts
- Helps understand the role of customer commitment

6.5 Monthly Charges vs Churn

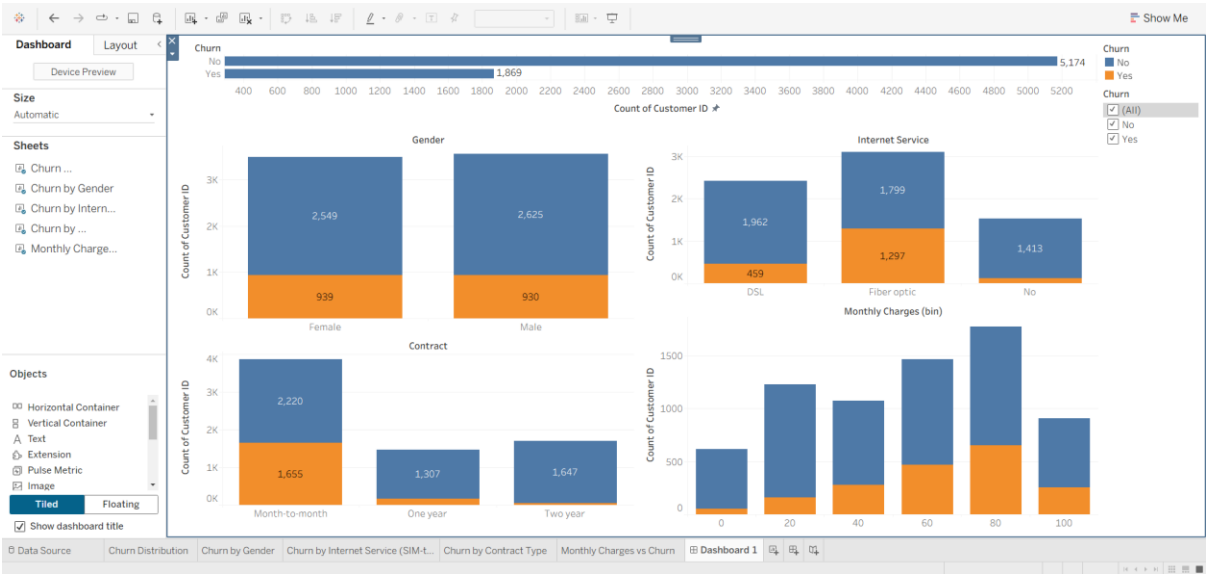


- Analyzes whether higher monthly charges increase churn probability
- Helps identify pricing sensitivity

---

7. Dashboard Creation

All five visualizations were combined into a **single interactive dashboard**.



Dashboard Features:

- Unified view of key churn drivers
- Interactive filters (e.g., Churn, Contract Type)
- Allows stakeholders to quickly explore high-risk customer segments

Purpose of Dashboard:

- Enable quick decision-making
- Provide a clear churn story at a glance
- Avoid the need to analyze individual charts separately

---

8. Key Business Insights Identified

From Excel analysis and Tableau dashboard, the following insights were observed:

- **Month-to-month contract customers churn the most**
- **Fiber optic internet users show higher churn rates**
- Customers with **higher monthly charges** are more likely to churn
- **Early-tenure customers** are at higher risk of leaving
- Payment method plays a role, with **electronic check users** showing higher churn
- Gender shows minimal influence on churn behavior

---

9. Importance of Step 1 for Machine Learning

This exploratory analysis is critical because:

- It helps select **relevant features** for model training
- It validates real-world business logic before ML implementation
- It ensures the ML model focuses on meaningful churn drivers
- It bridges the gap between **business understanding and data science**

---

10. Conclusion of Step 1

Step 1 successfully established a strong foundation for the churn prediction project by:

- Understanding customer behavior through data
- Identifying major churn drivers
- Visualizing patterns clearly using dashboards
- Preparing the ground for feature engineering and model building in subsequent steps

The insights derived from this step will directly guide **feature selection, model design, and retention strategy formulation** in the next phase of the project.
