

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/blatchar/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.

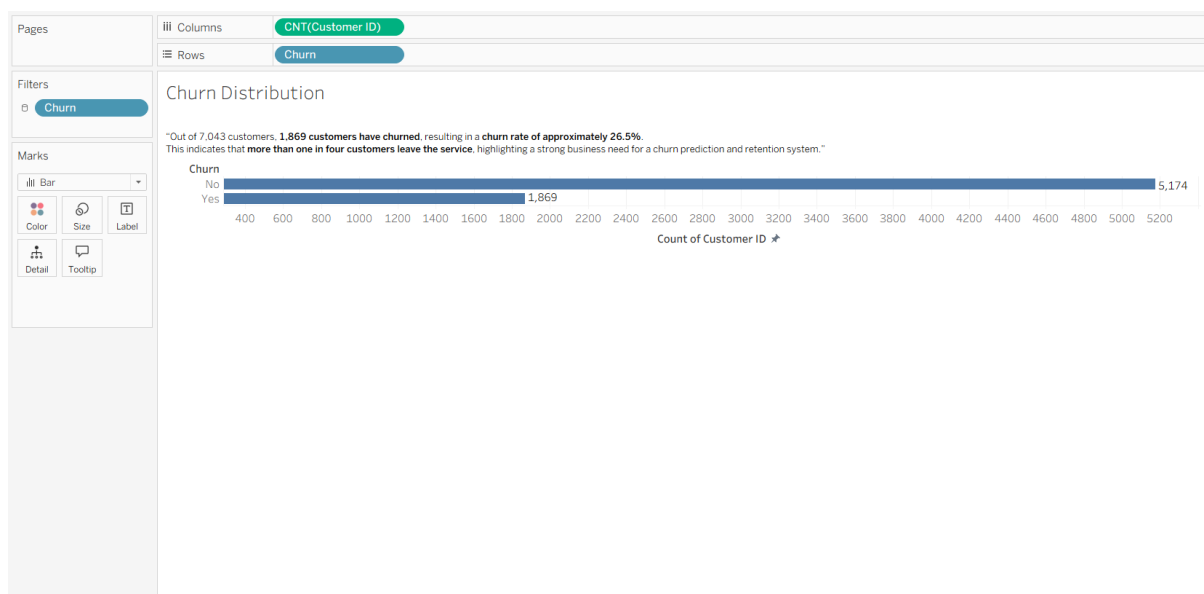
Purpose of Using Tableau:

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

6. Visualizations Created

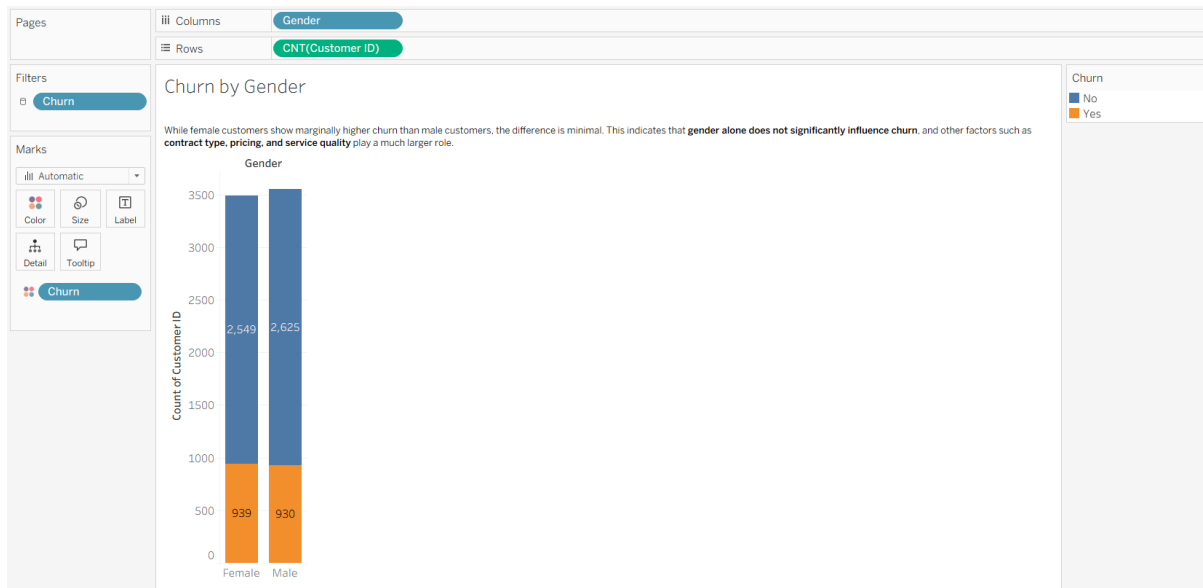
The following five core visualizations were created to answer key business questions:

6.1 Churn Distribution



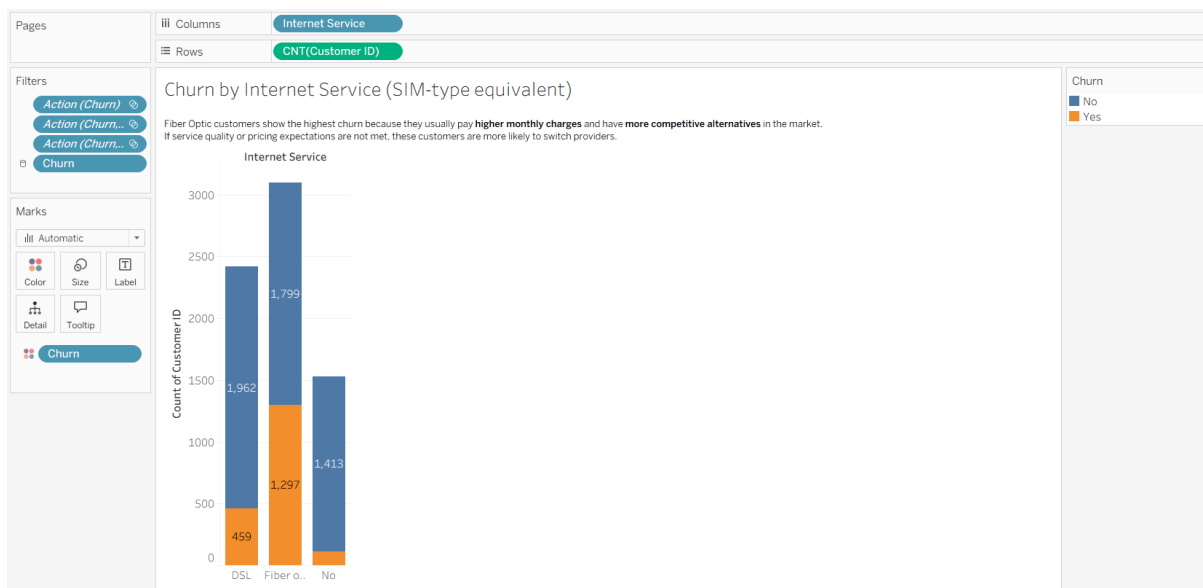
- Shows the proportion of churned vs retained customers
- Helps understand the scale of the churn problem

6.2 Churn by Gender



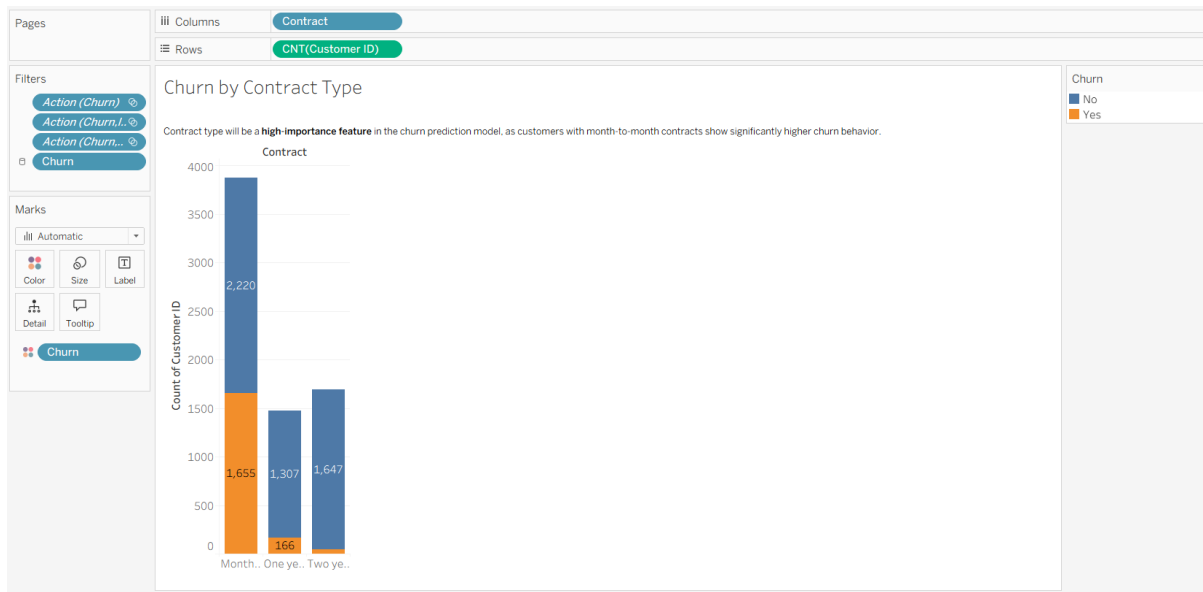
- Compares churn rates between male and female customers
- Used to check demographic influence

6.3 Churn by Internet Service



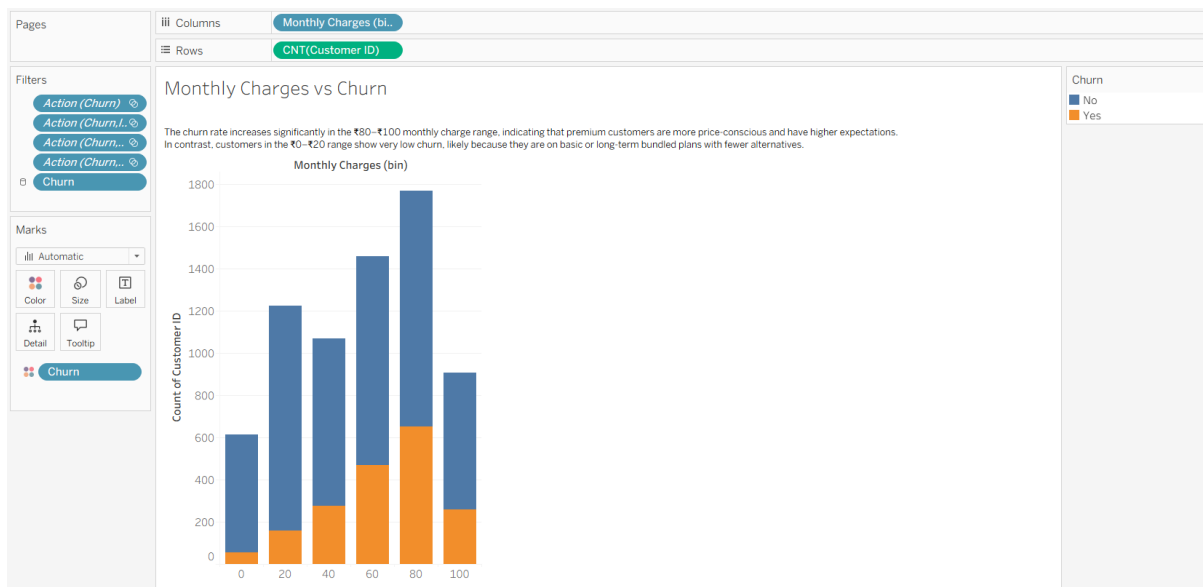
- Compares churn across DSL, Fiber optic, and No internet
- Identifies service-related dissatisfaction

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

8. Additional Data Enrichment & Derived Fields

After initial analysis, additional derived fields were introduced to support deeper segmentation and churn understanding:

Derived Fields Added:

- **AgeBand** — simulated based on SeniorCitizen (20–64 vs 65+)
- **SIM_Network** — simulated telecom provider (Jio, Airtel, Vodafone, BSNL)
- **Tenure_Quarter** — mapping customer tenure into quarterly lifecycle buckets (Q1–Q8)

Purpose of Derived Fields:

- Allow multidimensional churn analysis
- Evaluate churn across demographic, temporal, and product dimensions
- Support business segmentation logic

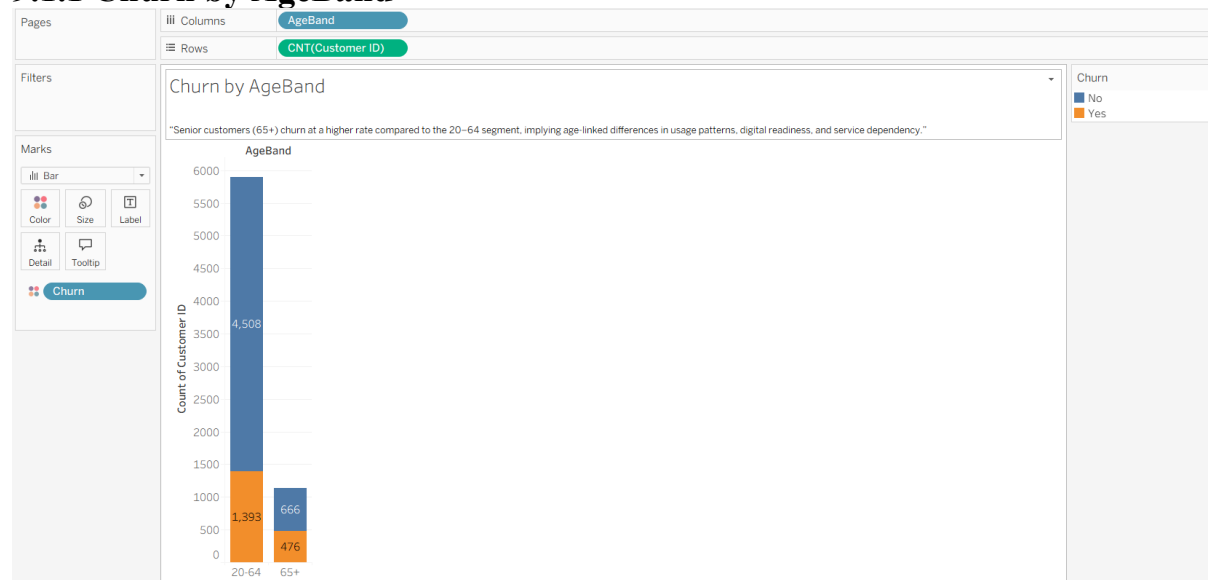
These enriched fields enabled advanced churn behavior insights beyond basic demographic comparisons.

9. Advanced Visualizations Created

Beyond the core 5 charts, additional Tableau worksheets were built to explore churn across demographics, product offerings, and customer lifecycle dimensions.

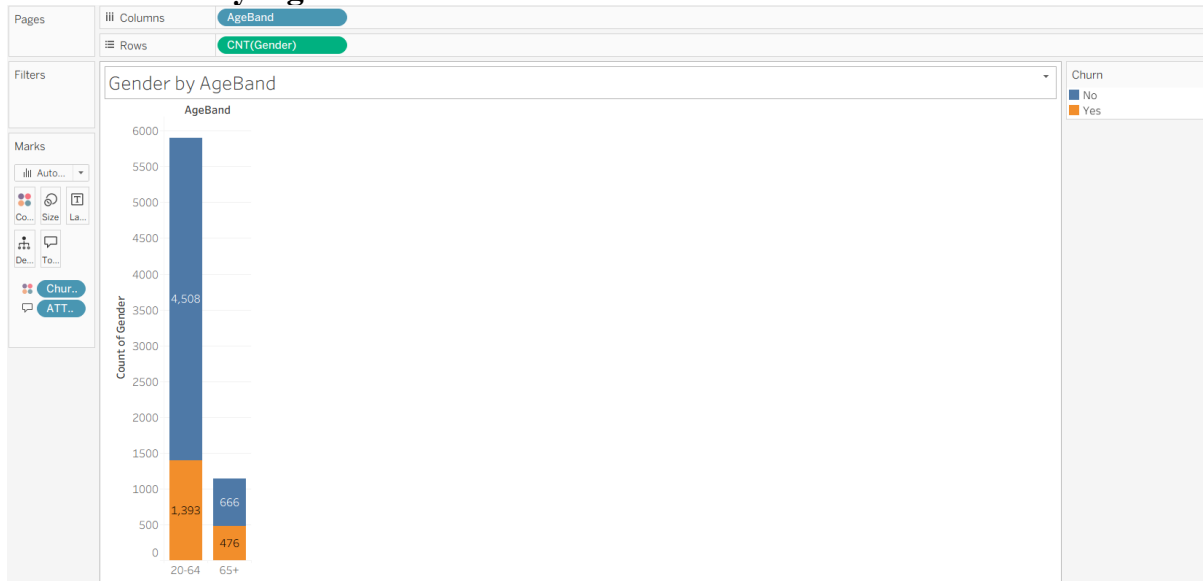
9.1 Demographic Visualizations

9.1.1 Churn by AgeBand



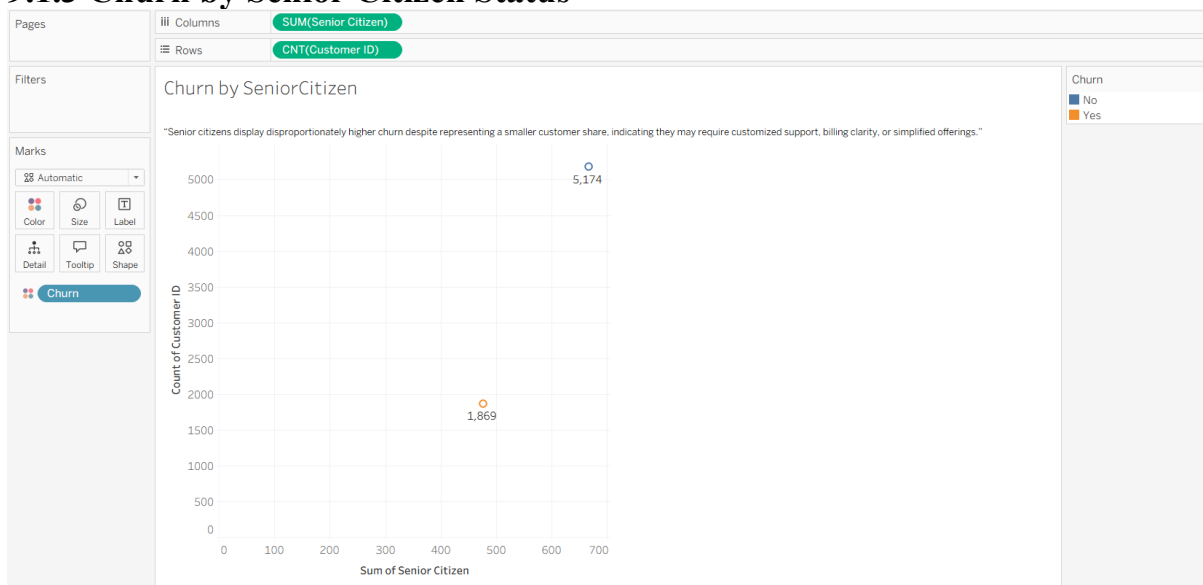
- Segments churn into 20–64 vs 65+
- Identifies age-linked churn sensitivity

9.1.2 Gender by AgeBand



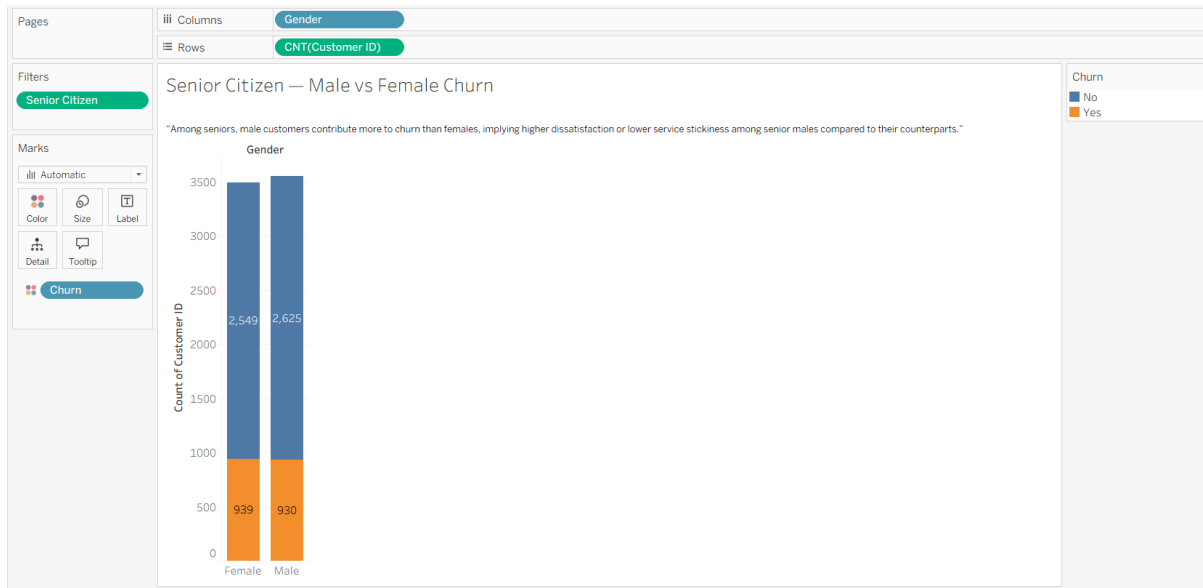
- Shows how gender distribution varies within age groups
- Used to validate demographic neutrality

9.1.3 Churn by Senior Citizen Status



- Compares churn rates between senior and non-senior customers
- Useful for segmentation-based retention strategies

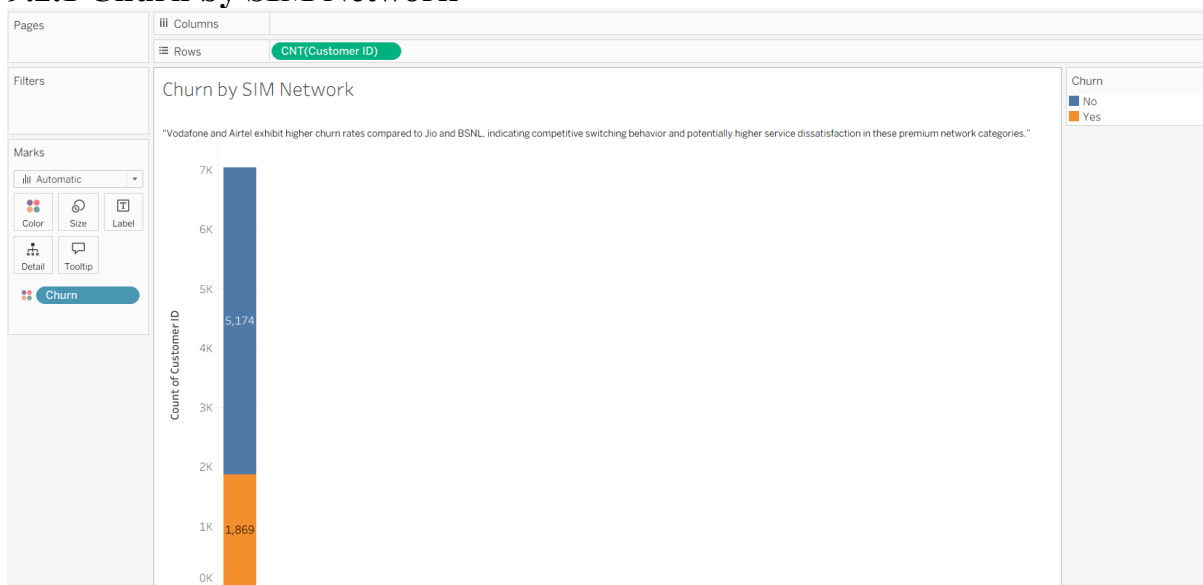
9.1.4 Senior Citizen Male vs Female Churn



- Further splits the senior group by gender
- Detects asymmetric churn drivers among older users

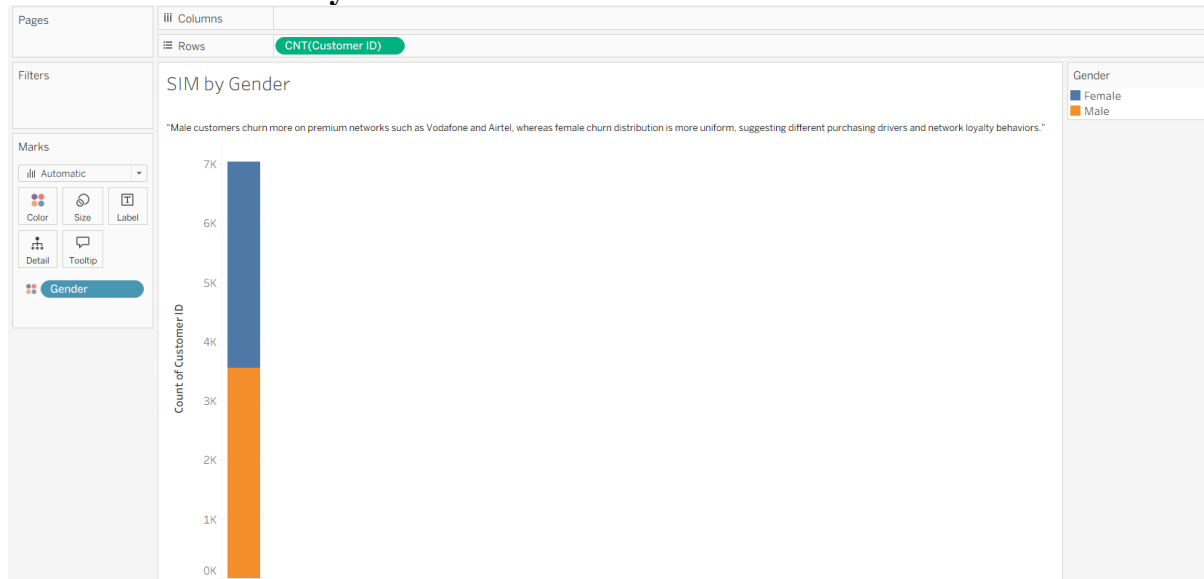
9.2 Product & SIM-Based Visualizations

9.2.1 Churn by SIM Network



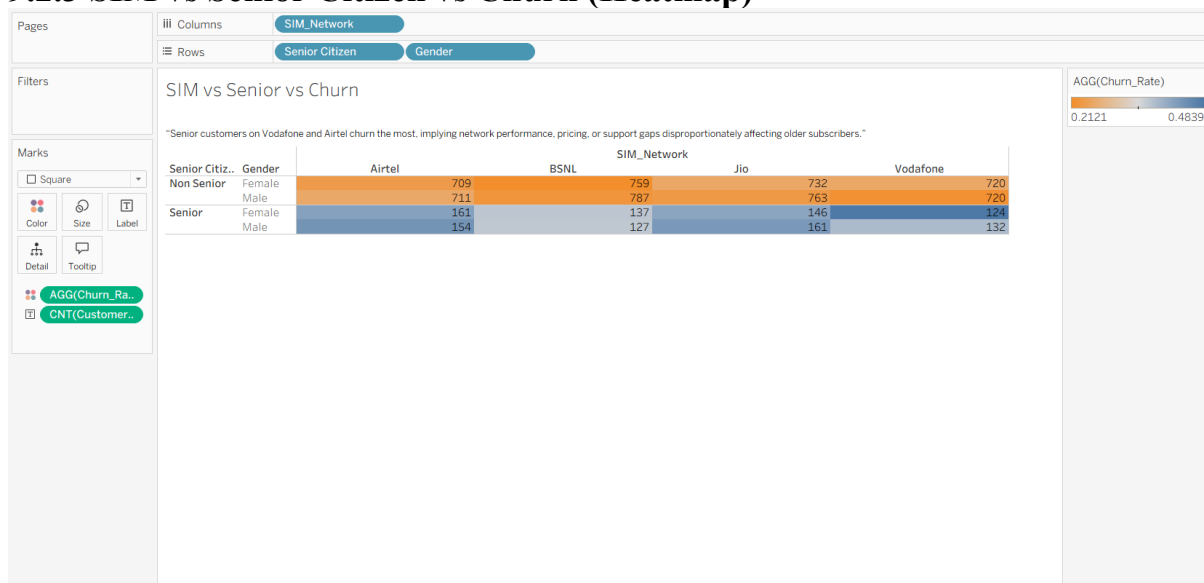
- Compares churn across SIM providers (Jio, Airtel, Vodafone, BSNL)
- Simulates competitive switching behavior

9.2.2 SIM Network by Gender



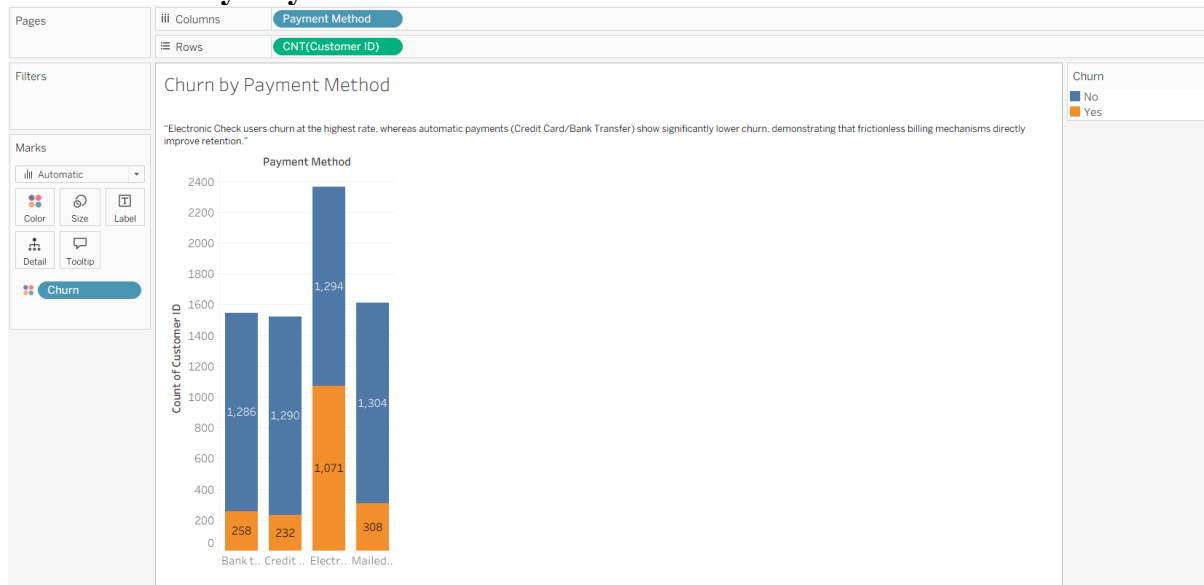
- Identifies gender preferences across SIM providers
- Useful for marketing campaign alignment

9.2.3 SIM vs Senior Citizen vs Churn (Heatmap)



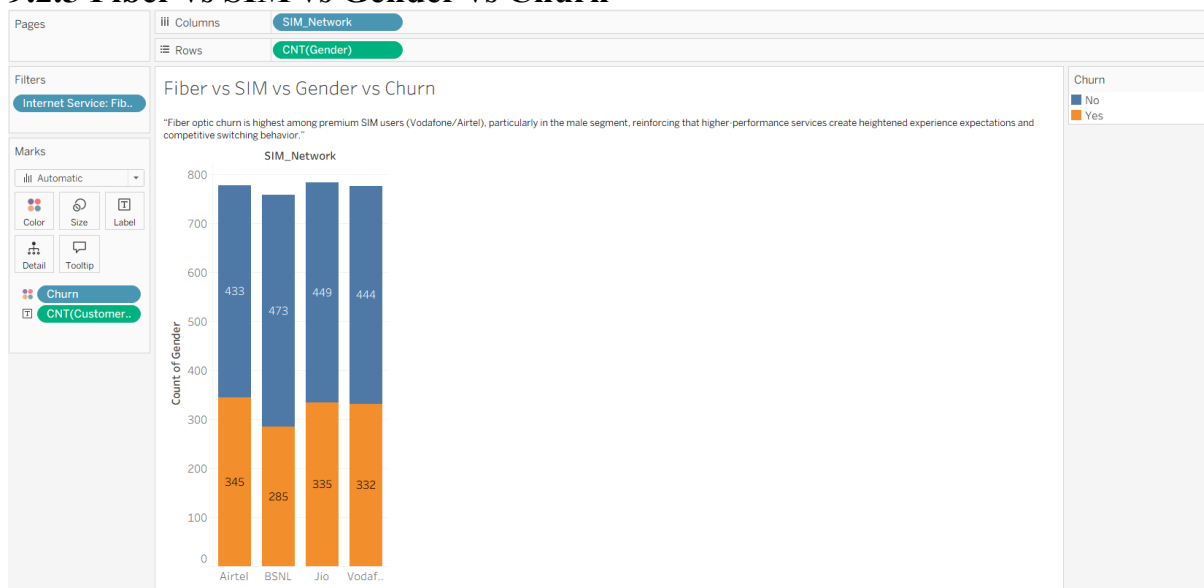
- 3-D relationship: Product × Demographic × Churn
- Enables high-risk cohort identification

9.2.4 Churn by Payment Method



- Shows churn differences across Electronic Check, Credit Card, etc.
- Highlights churn due to billing friction

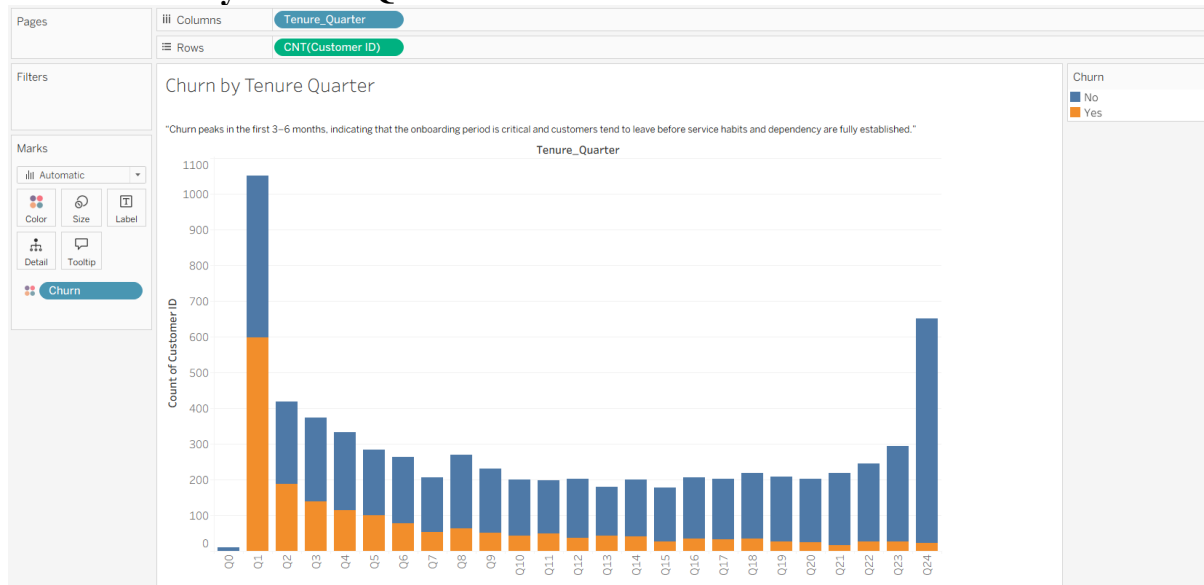
9.2.5 Fiber vs SIM vs Gender vs Churn



- Multi-factor comparison of premium services and network providers
- Reveals switching dynamics among different groups

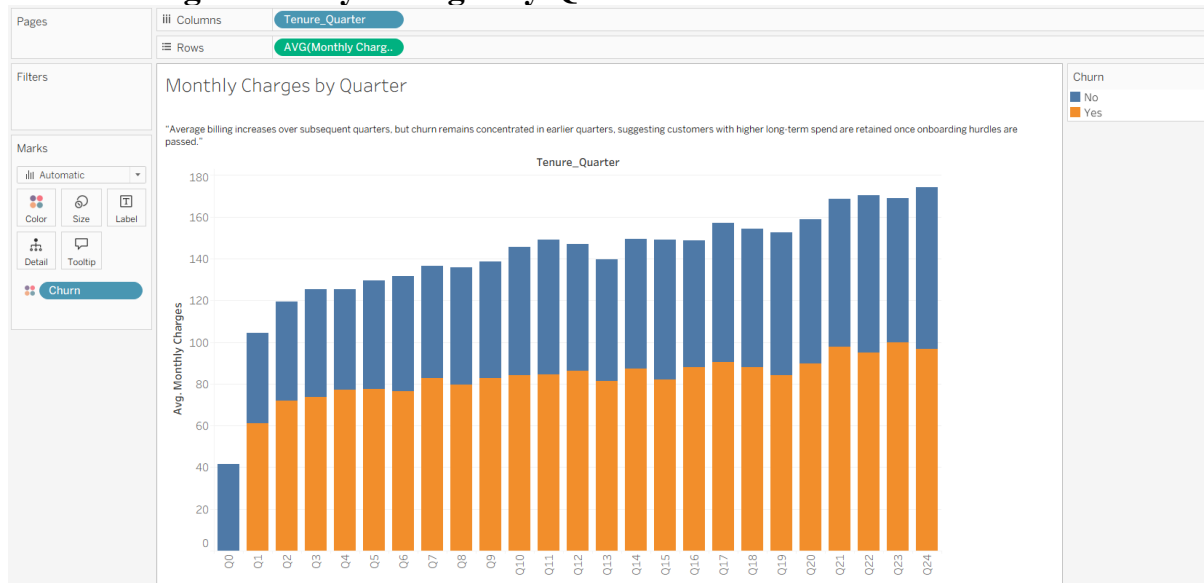
9.3 Lifecycle & Tenure-Based Visualizations

9.3.1 Churn by Tenure Quarter



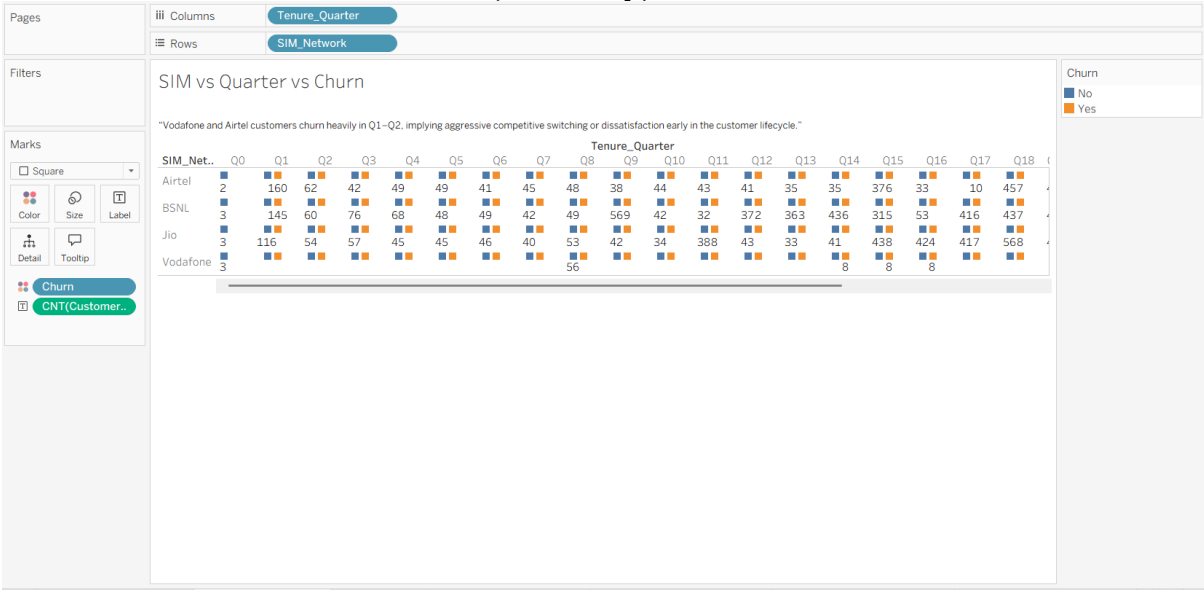
- Shows churn behavior by early vs matured customer lifecycle
- Identifies onboarding phase churn risk

9.3.2 Average Monthly Charges by Quarter



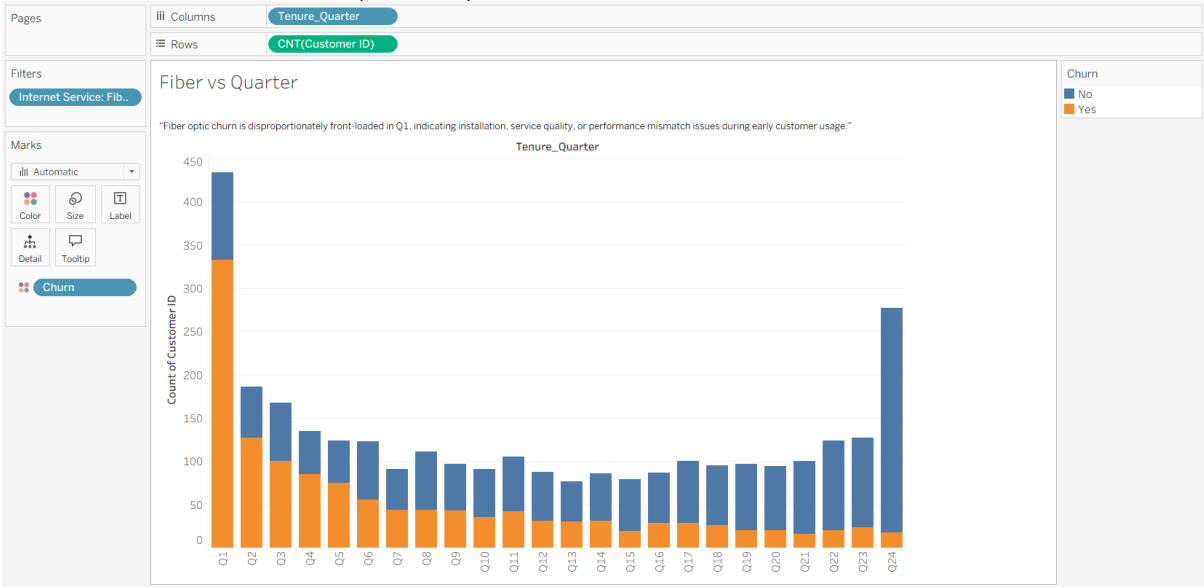
- Connects pricing evolution with retention dynamics

9.3.3 SIM vs Quarter vs Churn (Heatmap)



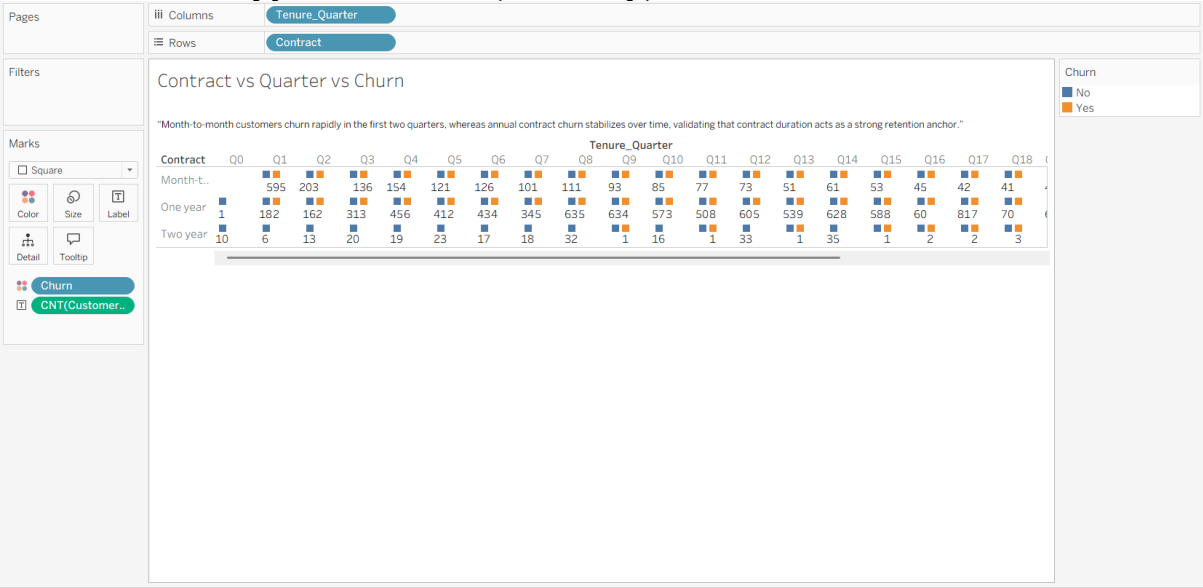
- Tracks network switching behavior over time
- Useful for competitor strategy inference

9.3.4 Fiber vs Quarter (Churn)



- Shows churn timing among high-speed internet users
- Validates dissatisfaction during early months

9.3.5 Contract Type vs Quarter (Heatmap)

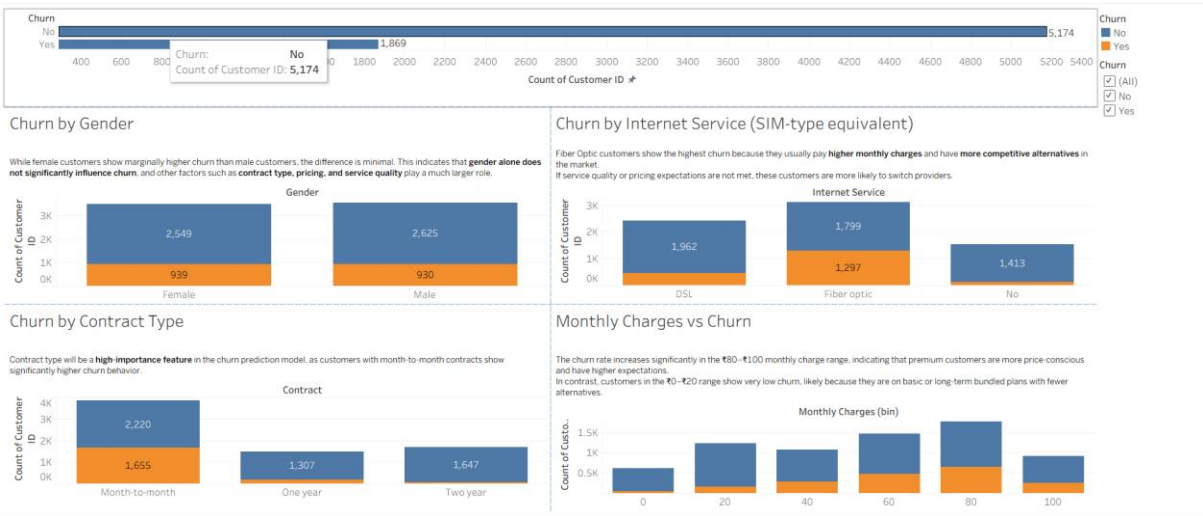


- Connects lifecycle progression with retention mechanisms
- Demonstrates impact of contract lock-in terms

10. Dashboard Architecture

To ensure clarity and avoid visual clutter, visualizations were structured into **four dashboards**, each answering a distinct business question:

10.1 Dashboard 1 — Churn Overview Dashboard

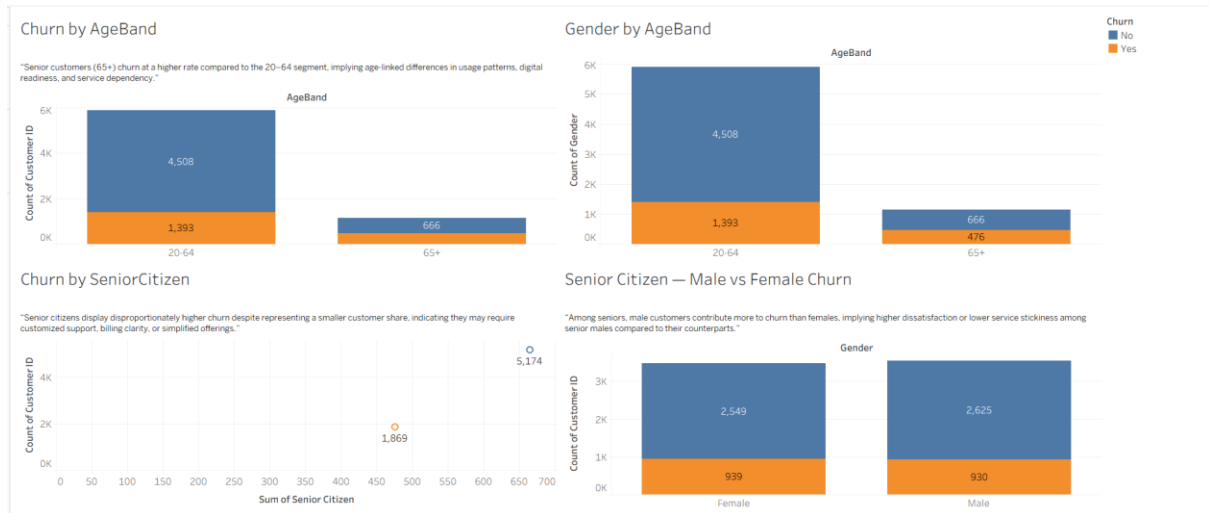


- Contains:
- Churn Distribution
 - Churn by Gender

- Churn by Internet Service
- Churn by Contract Type
- Monthly Charges vs Churn

Purpose:

→ Understand high-level churn drivers



10.2 Dashboard 2 — Demographic Churn Dashboard

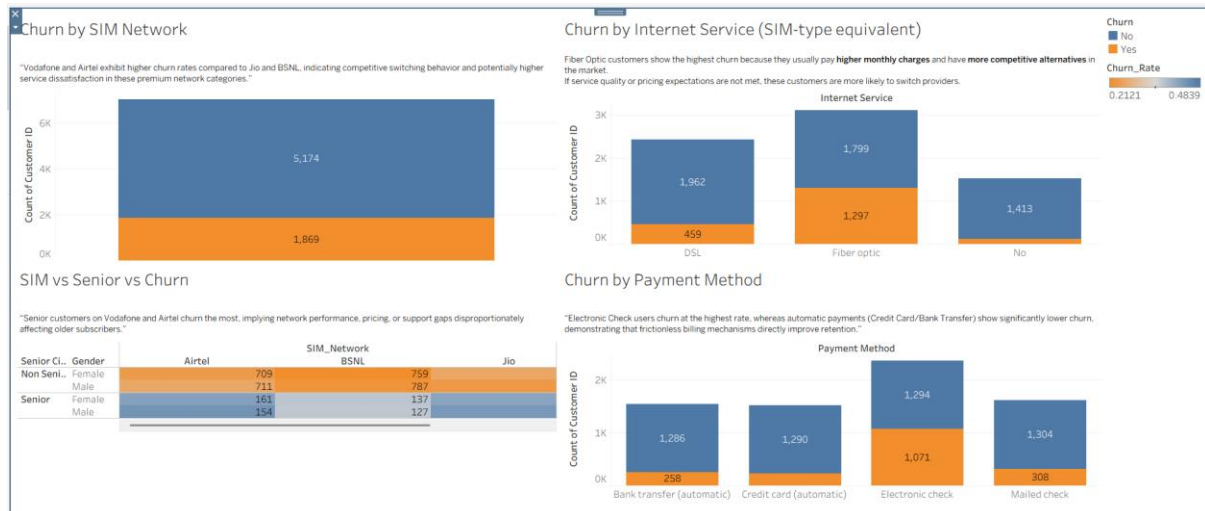
Contains:

- Churn by AgeBand
- Gender by AgeBand
- Churn by Senior Citizen Status
- Senior Citizen Male vs Female Churn

Purpose:

→ Identify churn patterns across customer demographics

10.3 Dashboard 3 — Product & SIM Churn Dashboard



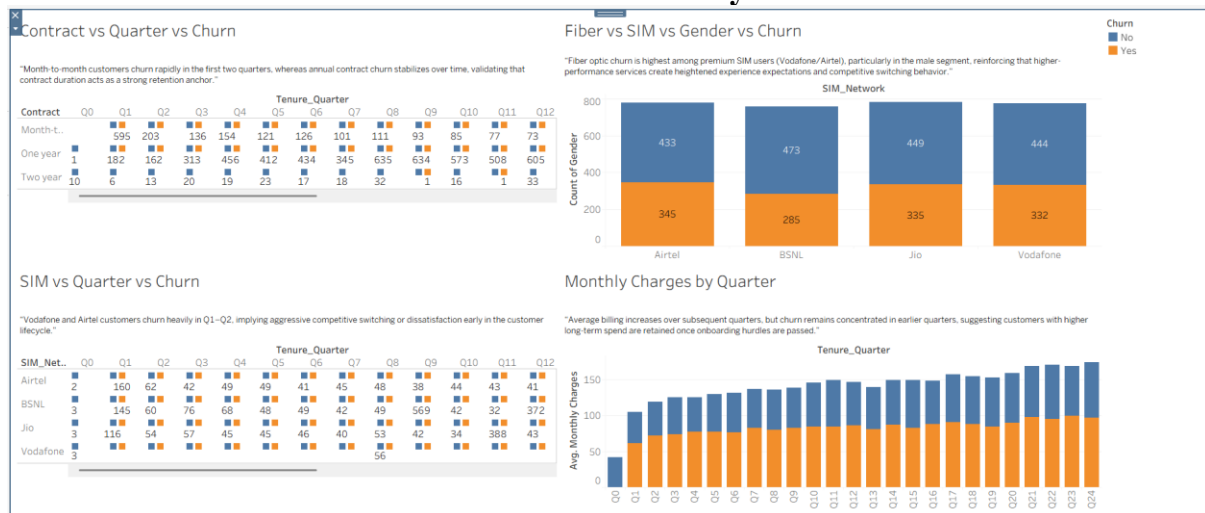
Contains:

- Churn by SIM Network
- SIM vs Gender
- SIM vs Senior vs Churn (Heatmap)
- Churn by Payment Method
- Fiber vs SIM vs Gender vs Churn

Purpose:

→ Analyze churn by product category, network, and billing preferences

10.4 Dashboard 4 — Tenure & Customer Lifecycle Dashboard



Contains:

- Churn by Tenure Quarter
- Average Charges by Quarter
- SIM vs Quarter (Heatmap)
- Fiber vs Quarter
- Contract Type vs Quarter (Heatmap)

Purpose:

→ Understand when churn occurs during customer lifecycle

10.5.1 Dashboard Features:

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

10.5.2 Purpose of Dashboard:

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

11. Cross-Dashboard Insights

From the combined dashboards, multiple major patterns emerged:

Key Insight Themes:

- ✓ Onboarding churn is high (first 1–2 quarters)
 - ✓ Premium users churn more (Fiber + higher billing)
 - ✓ Contract duration strongly influences retention
 - ✓ Age and seniority matter more than gender
 - ✓ SIM/network switching behavior exists
 - ✓ Payment friction increases churn
 - ✓ Not all churn is price-driven—service expectations matter
-

12. Final Conclusion of Step 1 (Extended)

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

The extended EDA and visualization phase demonstrated that customer churn is:

- **Segmentable** (specific high-risk cohorts exist)

- **Predictable** (drivers show clear patterns)
- **Actionable** (interventions can be targeted)

These insights validate that machine learning is meaningful for this use case and directly support feature engineering and model design in Step 2.

13. 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**
-

14. 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.
