

Customer Churn Exploratory Data Analysis (EDA)

1. Introduction

This document presents an Exploratory Data Analysis (EDA) performed on a telecom customer churn dataset. The analysis focuses on understanding customer behavior patterns and identifying factors that contribute to customer churn using visual and statistical exploration.

2. Project Objectives

The main goals of this analysis are listed below:

1. Analyze overall churn behavior in the telecom dataset
 2. Study the impact of demographic factors such as gender and senior citizen status
 3. Understand how service usage and billing methods influence churn
 4. Compare customer behavior across different SIM operators
 5. Generate insights that can support churn reduction strategies
-

3. Tools and Environment

The analysis was carried out using the following tools and technologies:

- **Python programming language** for end-to-end implementation
 - **Pandas and NumPy** for efficient data manipulation and numerical analysis
 - **Matplotlib and Seaborn** for exploratory data visualization
 - **PyCharm IDE** for structured development using OOP, logging, and exception handling
 - CSV-based dataset (**Customer_Churn_Updated.csv**) as the data source
-

4. Dataset Overview

The dataset consists of customer-level records from a telecom company. Each record represents an individual customer and contains information related to demographics, subscribed services, billing preferences, tenure, and churn status.

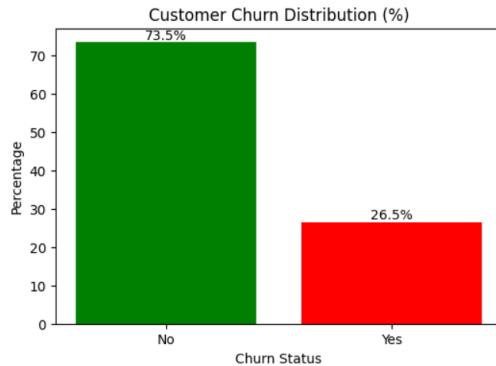
- **Key Columns:**
 - **Churn:** Whether the customer left the service (Yes / No)
 - **gender:** Customer gender (Male / Female)
 - **SeniorCitizen:** Senior citizen status (0 = No, 1 = Yes)
 - **SIM:** SIM operator used by the customer (Airtel / BSNL / Jio / Vi)
 - **InternetService:** DSL / Fiber optic / No
 - **PhoneService:** Whether phone service is active (Yes / No)
 - **MultipleLines:** Whether multiple phone lines are used (Yes / No)
 - **OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies:** Service subscriptions
 - **Contract:** Month-to-month / One year / Two year
 - **PaperlessBilling:** Paperless billing status (Yes / No)
 - **PaymentMethod:** Customer payment method
 - **tenure:** Number of months the customer stayed with the company
 - **MonthlyCharges:** Monthly bill amount
 - **Region:** Customer location

5. Visualizations, Graph Used & Understanding

5.1 Churn Distribution

Graph Used: **Bar Chart**

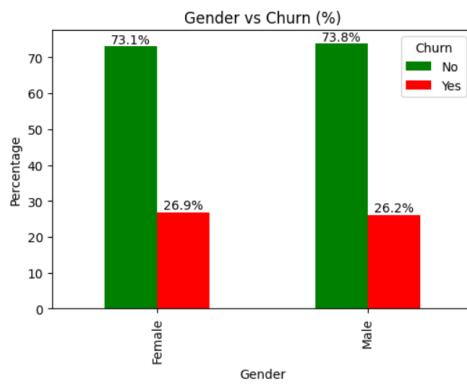
Understanding: Shows the overall proportion of churned and non-churned customers, giving a high-level view of churn severity.



5.2 Churn Distribution by Gender

Graph Used: **Bar Chart**

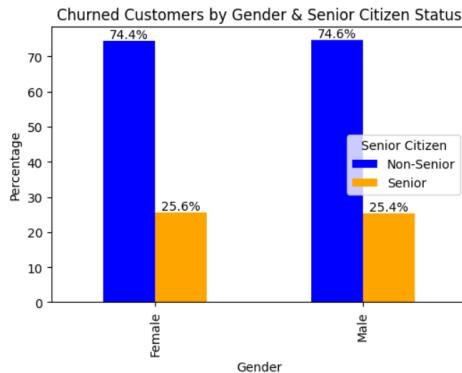
Understanding: Compares churn behavior between male and female customers to identify gender-based differences.



5.3 Churned Customers by Gender & Senior Citizen

Graph Used: **Bar Chart**

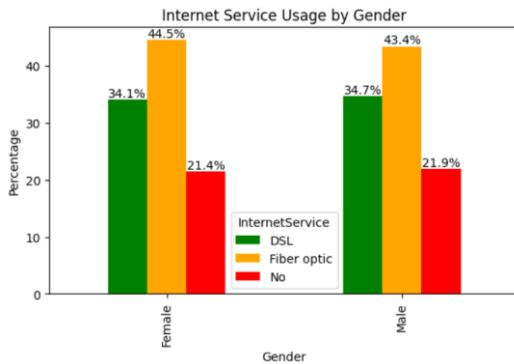
Understanding: Highlights how churned customers are distributed across gender and senior citizen categories.



5.4 Internet Service Usage by Gender

Graph Used: **Bar Chart**

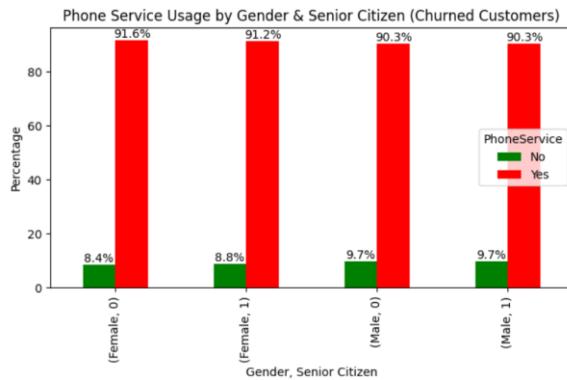
Understanding: Shows how different internet services are used by male and female customers.



5.5 Phone Service Usage by Gender & Senior Citizen (Churned Customers)

Graph Used: **Bar Chart**

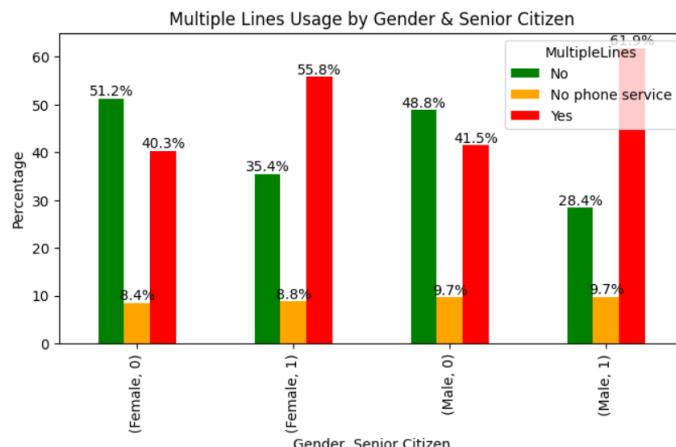
Understanding: Analyzes phone service usage among churned customers based on gender and senior citizen status.



5.6 Multiple Lines Usage by Gender & Senior Citizen

Graph Used: **Bar Chart**

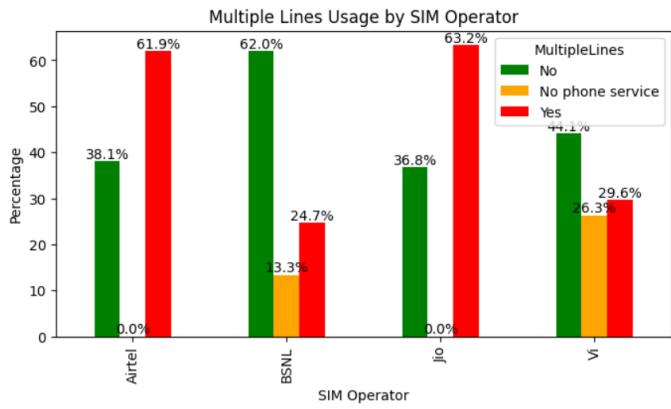
Understanding: Explains variation in multiple line subscriptions across demographic groups.



5.7 Multiple Lines by SIM Operator

Graph Used: **Bar Chart**

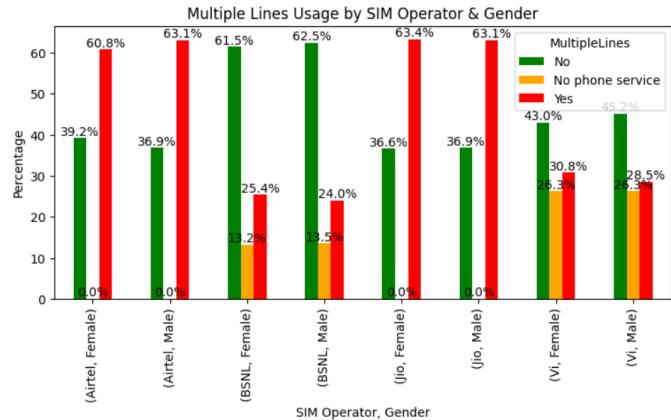
Understanding: Compares multiple line usage across different SIM operators.



5.8 Multiple Lines Usage by SIM Operator & Gender

Graph Used: **Bar Chart**

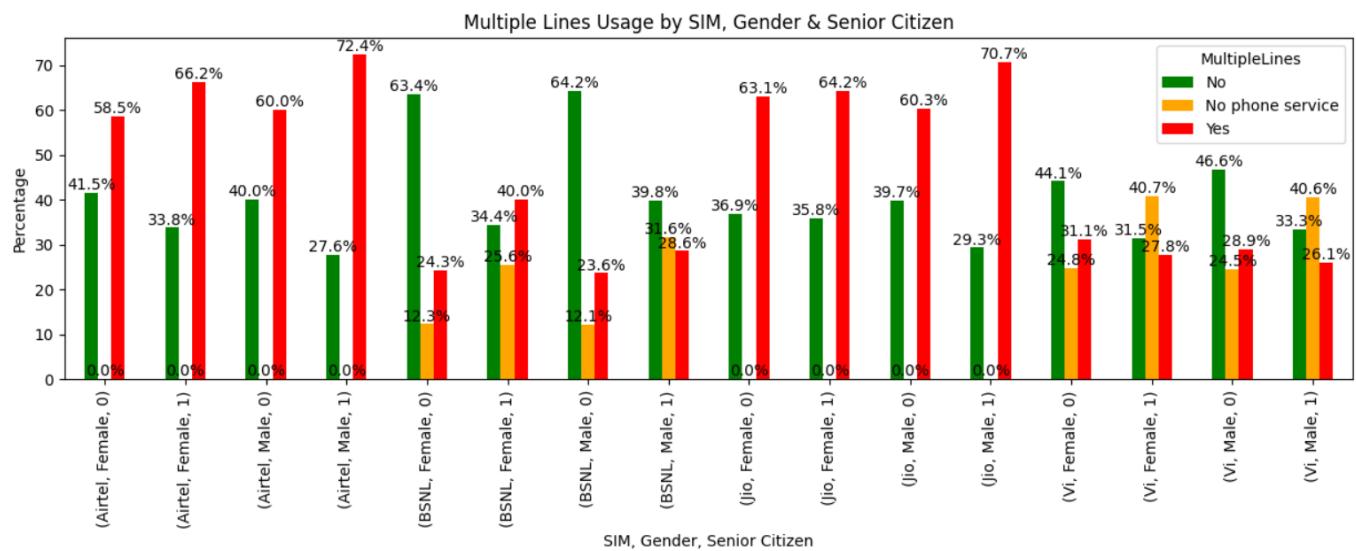
Understanding: Breaks down multiple line usage by SIM operator and gender.



5.9 Multiple Lines Usage by SIM, Gender & Senior Citizen

Graph Used: **Bar Chart**

Understanding: Provides a deeper demographic breakdown of multiple line usage.

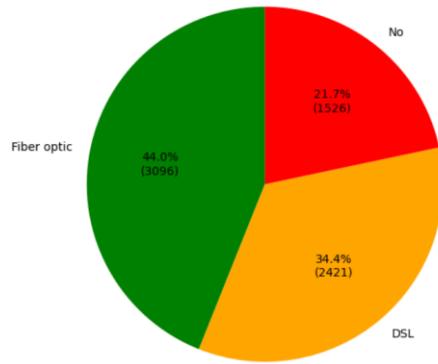


5.10 Total Customers by Internet Service

Graph Used: **Pie Chart**

Understanding: Represents the customer distribution by internet service type.

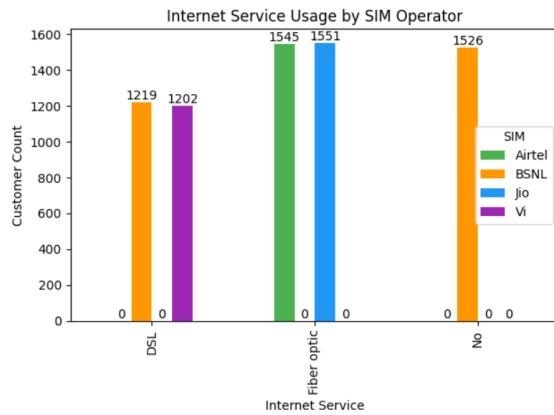
Total Customers by Internet Service Distribution



5.11 Internet Service Usage by SIM Operator

Graph Used: **Bar Chart**

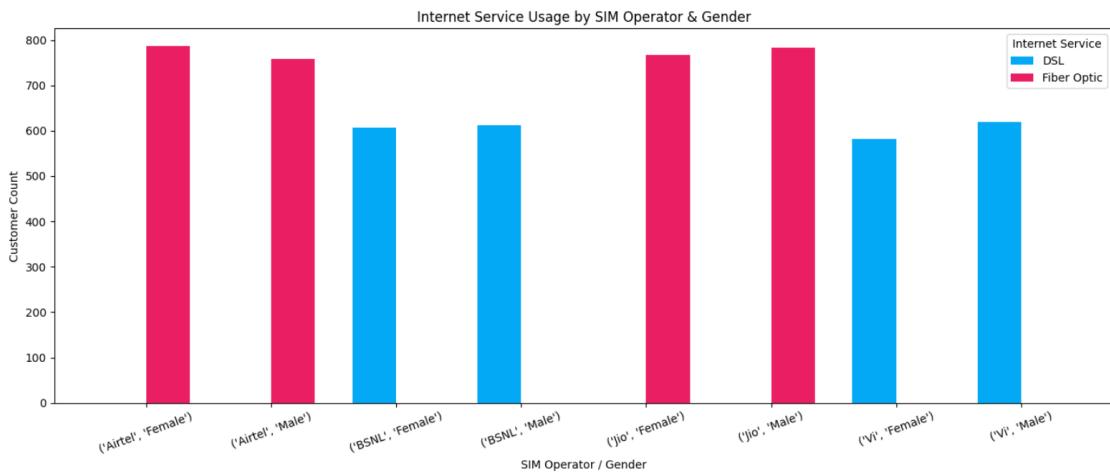
Understanding: **Shows how internet services are distributed across SIM operators.**



5.12 Internet Service Usage by SIM Operator & Gender

Graph Used: **Bar Chart**

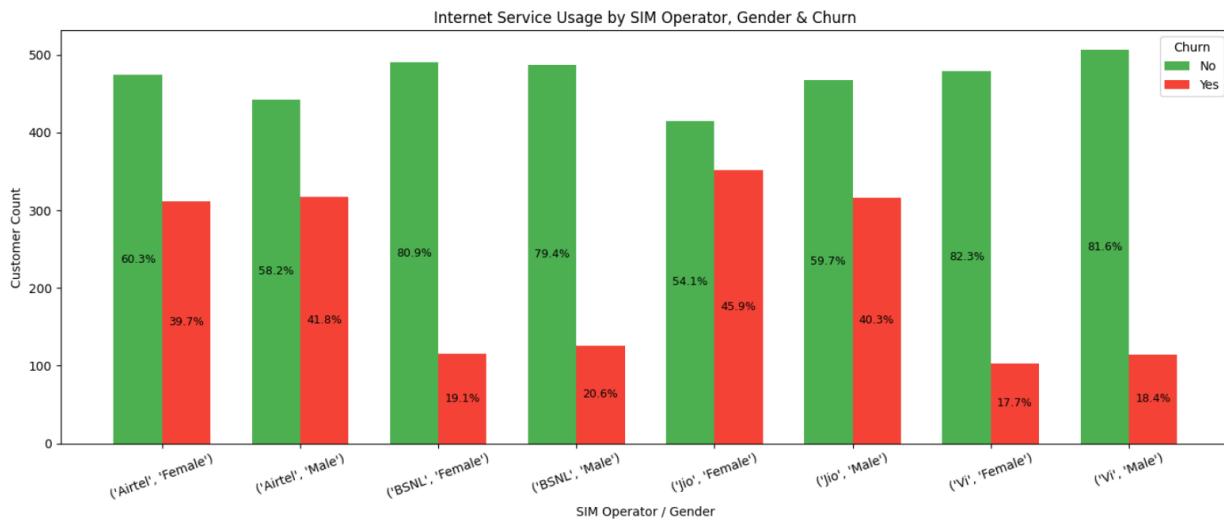
Understanding: **Compares DSL and Fiber Optic usage across SIM operators and gender.**



5.13 Internet Service Usage by SIM Operator, Gender & Churn

Graph Used: **Bar Chart**

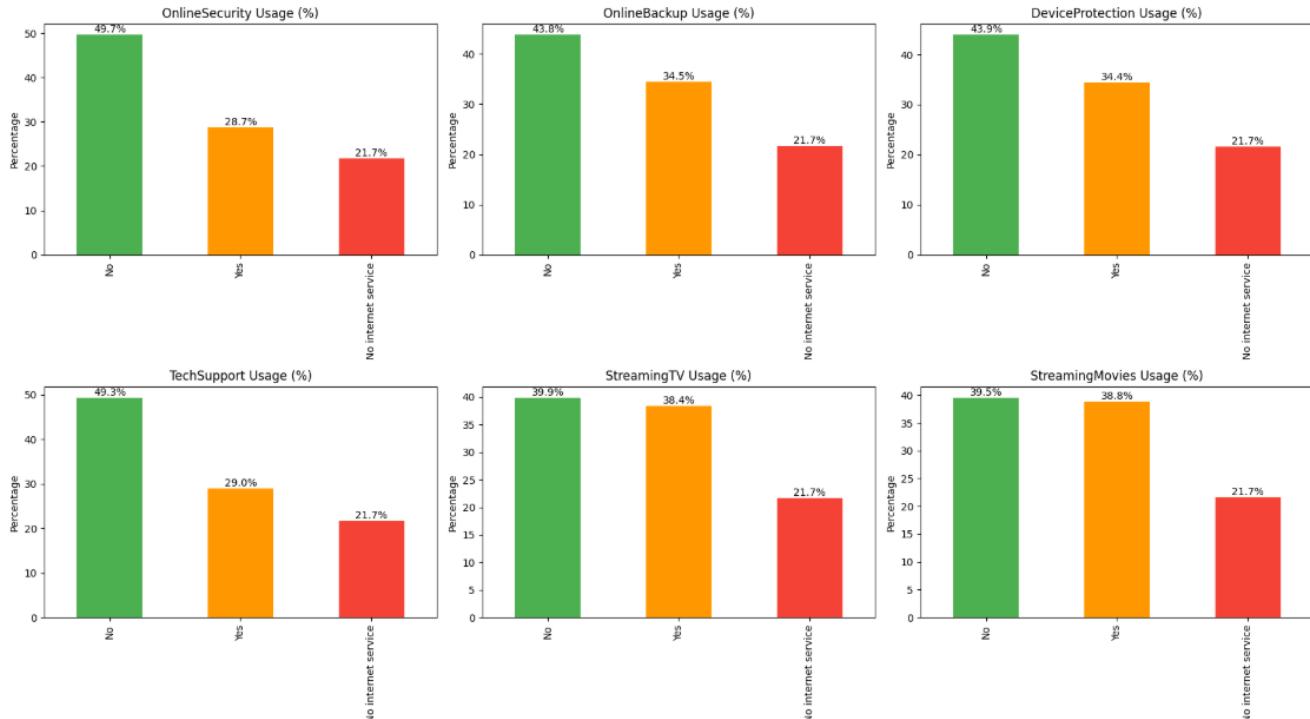
Understanding: **Analyzes churn behavior across SIM operators and gender for internet services.**



5.14 Service Feature Usage Analysis

Graph Used: **Multiple Bar Charts**

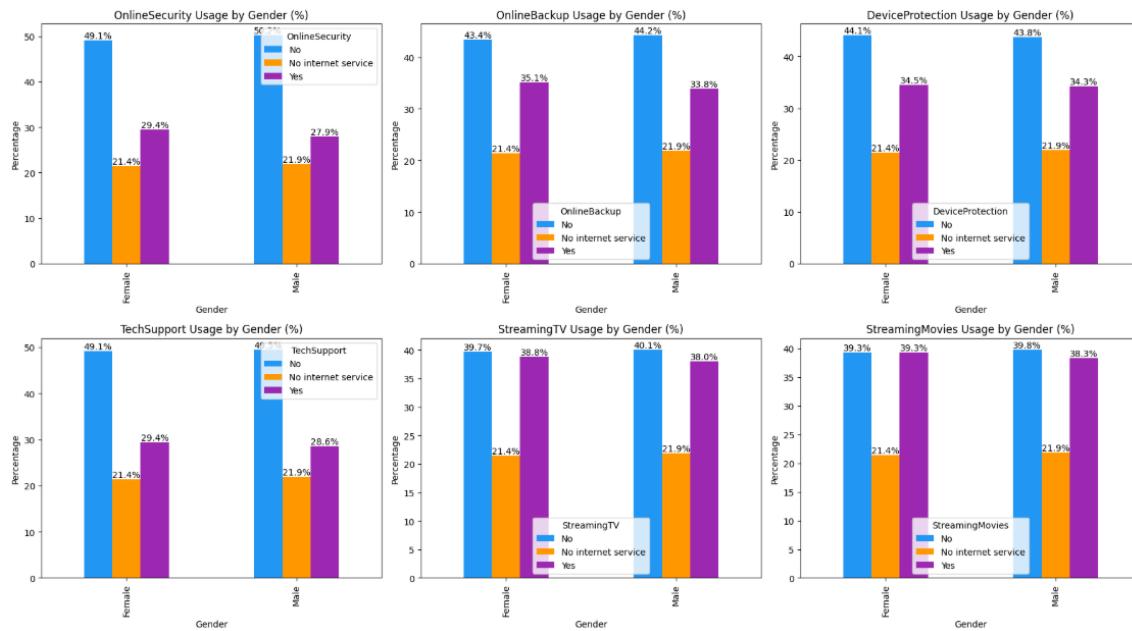
Understanding: Analyzes usage of service features such as Online Security, Streaming, and Tech Support overall and by gender.



5.14.1 Service Usage by Gender (%)

Graph Used: **Multiple Bar Charts**

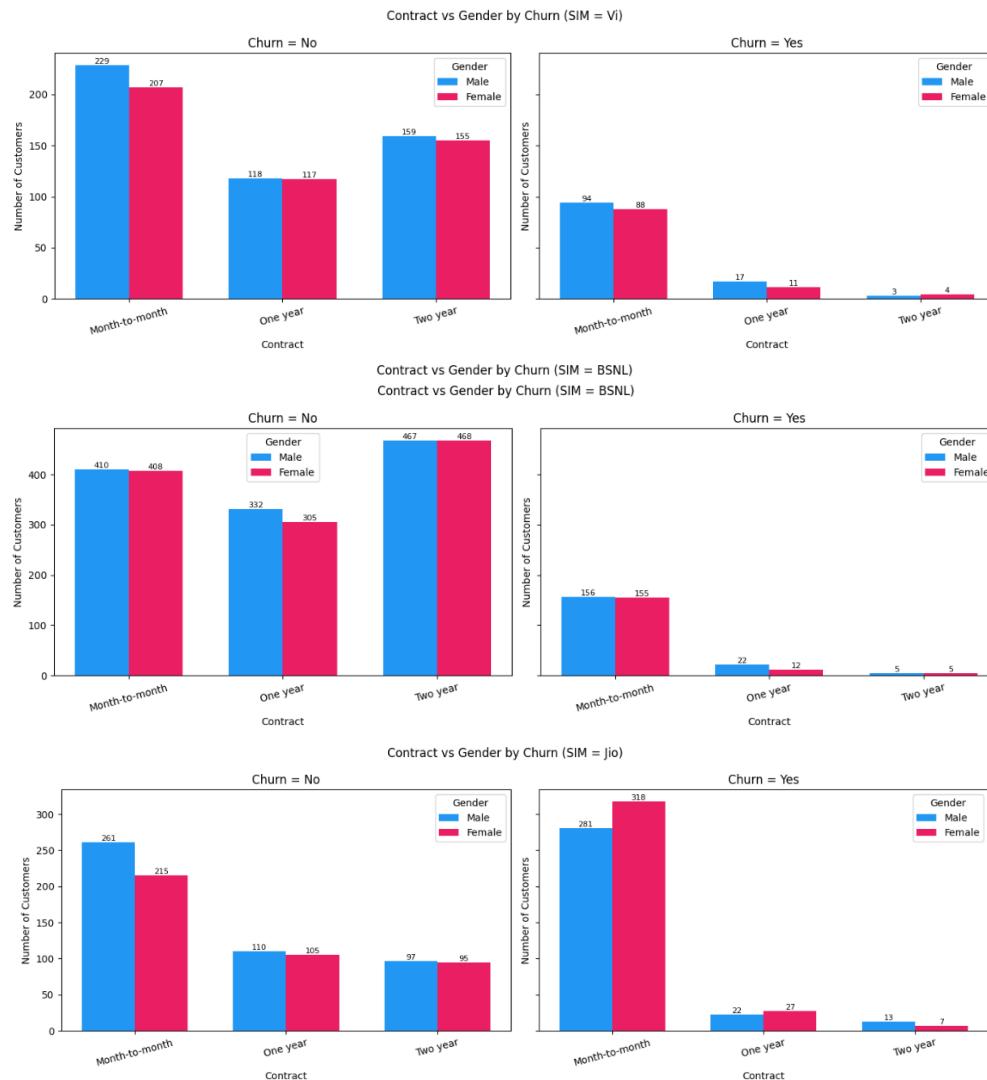
Understanding: Compares service feature usage percentages between male and female customers across multiple service categories.

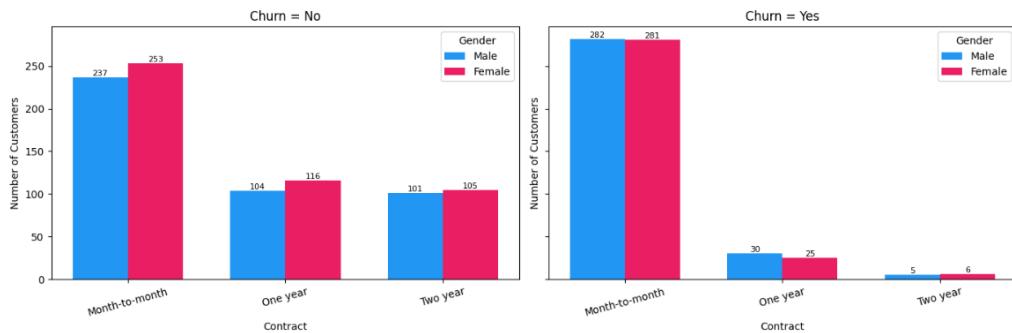


5.15 Contract Distribution Analysis

Graph Used: Bar Charts

Understanding: Examines how contract types vary by gender, churn status, senior citizen status, and SIM operator.

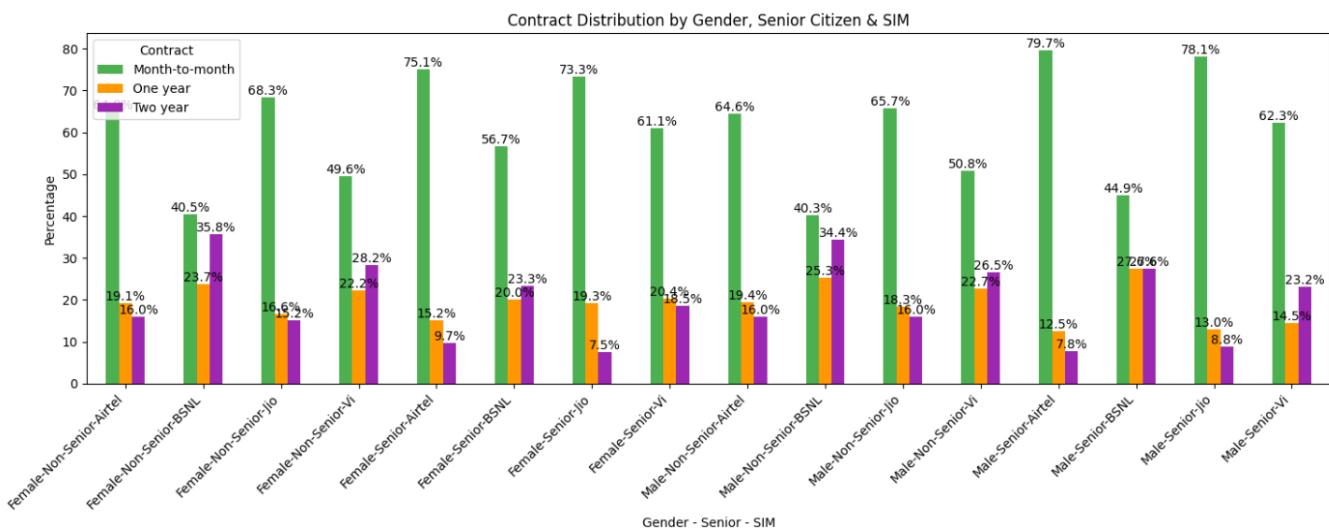




5.15.1 Contract Distribution by Gender, Senior Citizen & SIM

Graph Used: Bar Charts

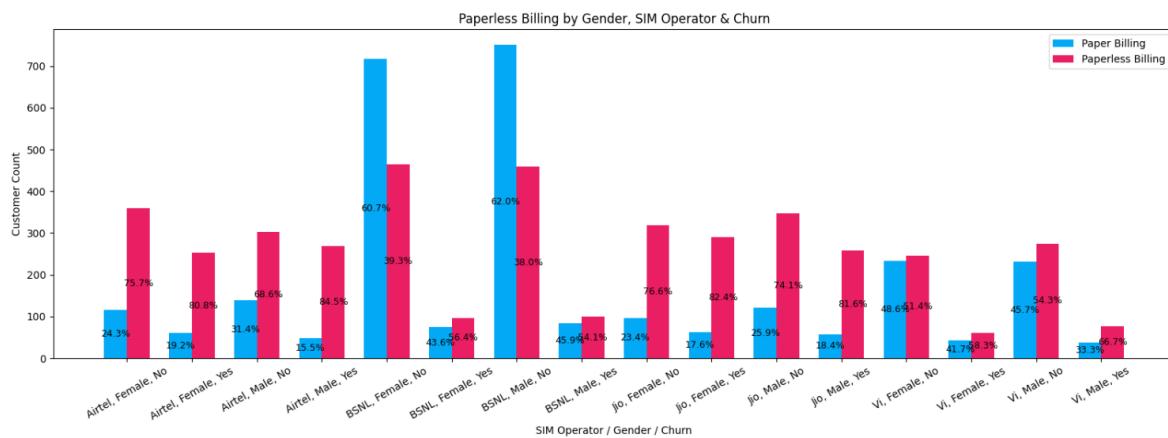
Understanding: Shows how contract type percentages differ across gender, senior citizen status, and SIM operator combinations.



5.16 Paperless Billing by Gender, SIM Operator & Churn

Graph Used: Bar Chart

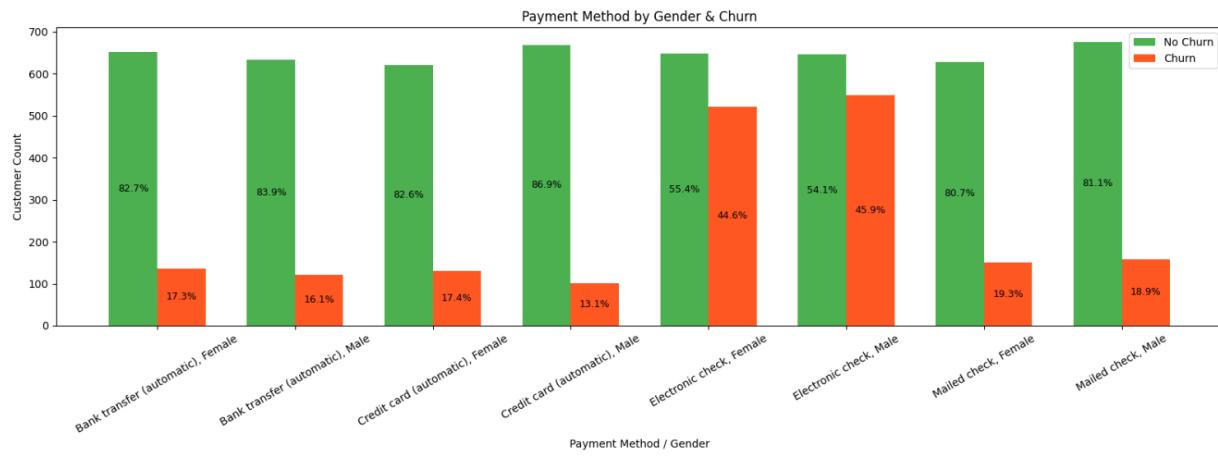
Understanding: Compares paperless and paper billing usage across different customer groups.



5.17 Payment Method by Gender & Churn

Graph Used: Bar Chart

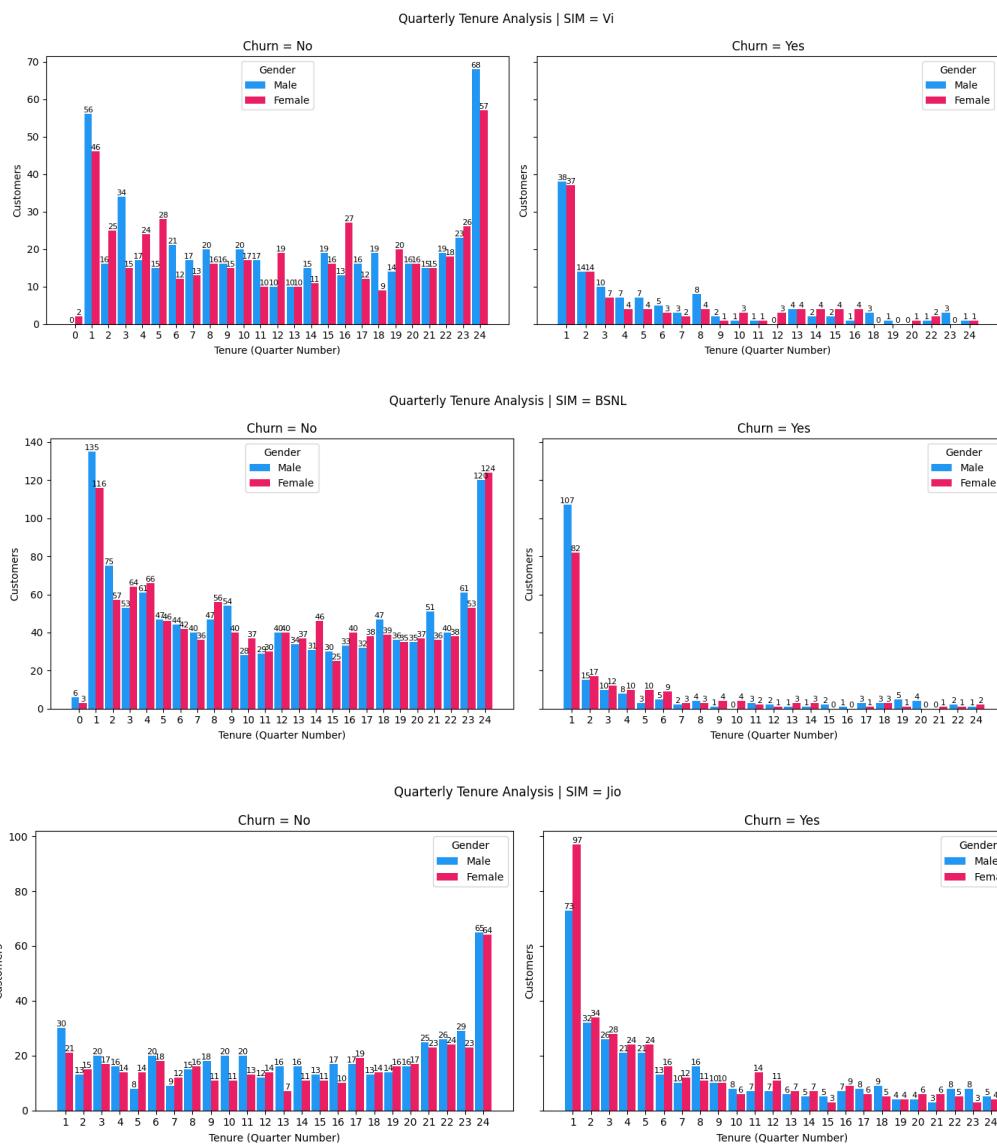
Understanding: Shows how payment methods relate to churn behavior across genders.

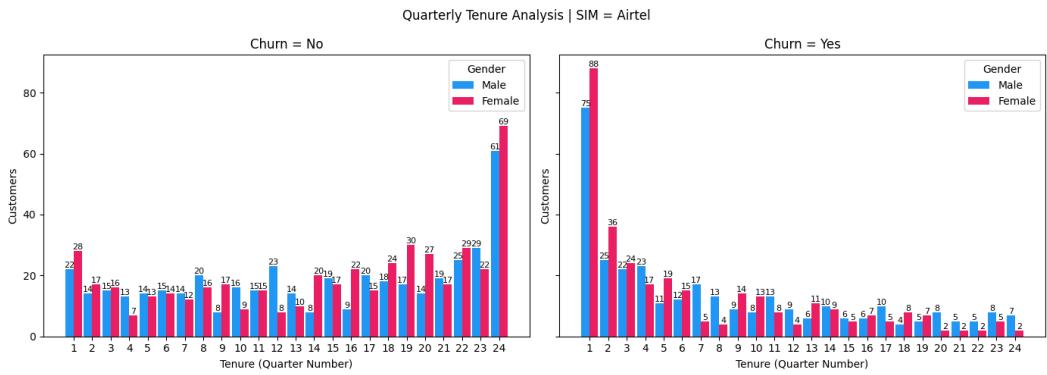


5.18 Quarterly Tenure Analysis with SIM

Graph Used: Bar Charts

Understanding: Groups tenure into quarters to understand customer lifecycle patterns by SIM operator.

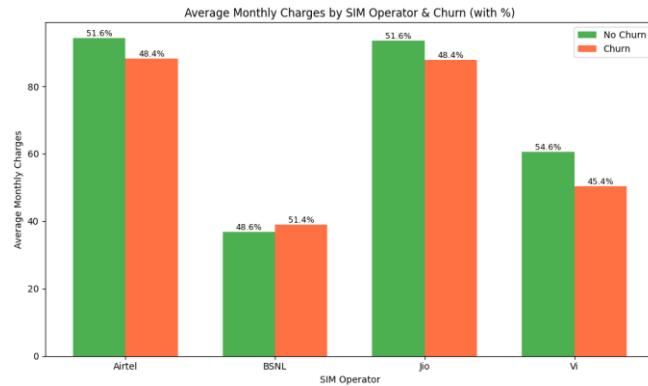




5.19 Monthly Charges vs SIM Operator by Churn

Graph Used: **Bar Chart**

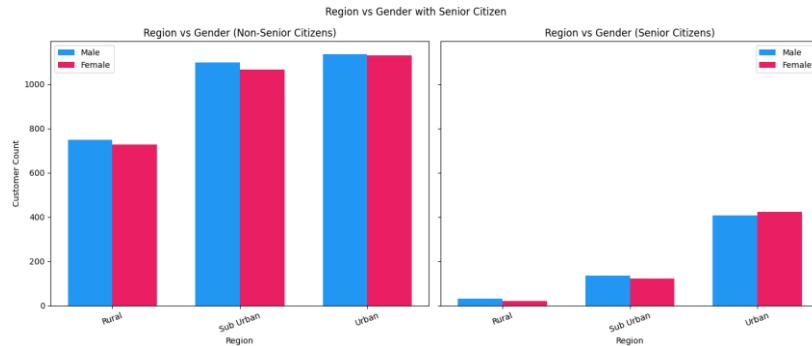
Understanding: **Compares average monthly charges of churned and non-churned customers across SIM operators.**



5.20 Region vs Gender with Senior Citizen

Graph Used: **Bar Charts**

Understanding: **Compares gender distribution across regions for senior and non-senior customers.**



6. Handling TotalCharges Column

The TotalCharges column was originally stored as a text (object) data type due to the presence of blank and non-numeric values. To ensure accurate numerical analysis, this column was converted into a numeric (float) data type.

During the conversion process, missing or invalid values were identified and handled by replacing them with 0. This ensured dataset consistency and avoided dropping customer records.

This preprocessing step was independently implemented and was not part of the reference implementation, making the analysis more robust and reliable.

7. Overall Business Insights

- Month-to-month contracts show higher churn compared to long-term contracts.
- Senior citizens are more likely to churn than non-senior customers.
- Fiber optic users exhibit higher churn compared to DSL users.
- Higher monthly charges are associated with increased churn.
- Customers with shorter tenure are more prone to churn.

SIM operator-specific strategies are required for effective retention.

8. Conclusion

This Exploratory Data Analysis provides a comprehensive understanding of factors influencing customer churn. The findings from this analysis can be used by telecom companies to design effective customer retention strategies and to support future predictive modeling efforts.