AI – Powered Customer Churn Prediction System

1. Project Overview

This project focuses on analyzing customer behavior within a telecommunications dataset to identify the key drivers of customer attrition (churn). By examining demographic information, subscription details, account tenure, and monthly charges, the analysis aims to provide actionable insights into which customer segments are most likely to leave. The ultimate goal is to aid in developing targeted retention strategies for different service providers and regions.

2. Dataset Description

The dataset contains **7,043** unique customer records and **24** features. It includes a mix of categorical (demographic, service details) and numerical (tenure, charges) variables.The dataset includes the following key columns:

- **customerID**: Unique identifier for each customer
- **gender**: Gender of the customer (Male / Female)
- **SeniorCitizen**: Indicates whether the customer is a senior citizen (0 = No, 1 = Yes)
- **Partner**: Whether the customer has a partner
- **Dependents**: Whether the customer has dependents
- **tenure**: Number of months the customer has stayed with the company
- **PhoneService**: Whether the customer has phone service
- **MultipleLines**: Whether the customer has multiple phone lines
- **InternetService**: Type of internet service (DSL, Fiber optic, No)
- **OnlineSecurity**: Whether the customer has online security service
- **OnlineBackup**: Whether the customer has online backup service
- **DeviceProtection**: Whether the customer has device protection service
- **TechSupport**: Whether the customer has technical support service
- **StreamingTV**: Whether the customer uses streaming TV
- **StreamingMovies**: Whether the customer uses streaming movies
- **Contract**: Contract type (Month-to-month, One year, Two year)
- **PaperlessBilling**: Whether the customer uses paperless billing
- **PaymentMethod**: Payment method used by the customer
- **MonthlyCharges**: Monthly amount charged to the customer
- **TotalCharges**: Total amount charged to the customer
- **Churn**: Target variable indicating whether the customer churned (Yes / No)
- **Service_provider**: Indicates the customer's telecom service provider (Jio, VI, Airtel, or BSNL), derived from internet service information.
- **Region_Type:** Classifies the customer's region as Urban, Suburban, or Rural based on internet service characteristics.
- **Device_Status:** Represents whether the customer is using a new or old device, inferred from device protection availability.

CATEGORIES:

- **Target Variable:** Churn (Indicates if a customer left the service).
- **Demographics:** gender, SeniorCitizen, Partner, Dependents, Region_Type.
- **Services:** Service_Provider (Jio, Airtel, Vi, BSNL), PhoneService, InternetService, etc.
- **Account Info:** tenure, Contract, PaymentMethod, MonthlyCharges, TotalCharges.

Summary for Dataset: Below is the distribution of the most critical columns used in this analysis:

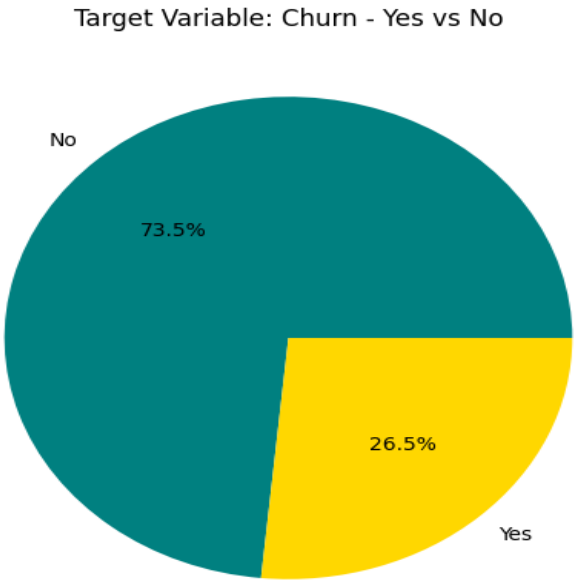
- **Total Customers:** 7,043
- **Churn Status:**
  - **No (Retained):** 5,174 (73.5%)
  - **Yes (Churned):** 1,869 (26.5%)
- **Gender Distribution:**
  - **Male:** 3,555
  - **Female:** 3,488
  - (The dataset is well-balanced regarding gender).
- **Service Providers:**
  - **Jio:** 2,858
  - **Airtel:** 2,113
  - **Vi:** 1,377
  - **BSNL:** 695

3. Visualization Analysis and Interpretation

Exploratory Data Analysis (EDA) was conducted using Matplotlib only, ensuring complete control over visualization design and clear interpretability of results. Each visualization was carefully selected based on its relevance, its ability to reveal meaningful patterns, and its usefulness in understanding customer churn behavior. The EDA focused on identifying key factors influencing churn, analyzing customer behavior across demographic, service, and billing attributes, and uncovering trends related to customer tenure and pricing. These insights guided feature understanding and provided a strong foundation for subsequent modelling and prediction tasks.

3.1. Churn Distribution (Pie Chart)

Description: This pie chart illustrates the distribution of the target variable, `Churn`. It visualizes the ratio of retained customers ("No") versus those who exited the service ("Yes"), providing a clear high-level view of the overall attrition rate within the dataset.

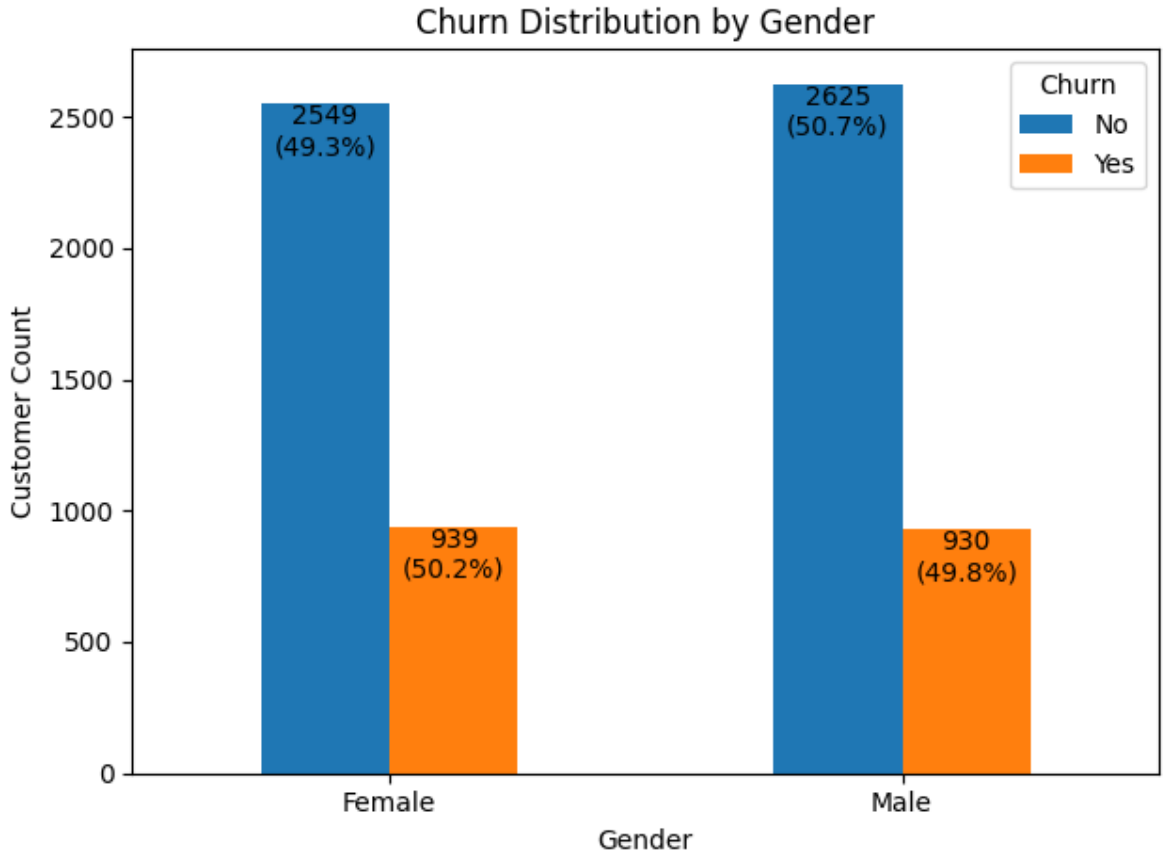


Summary & Insight:

- **Total churn users are 1,869 (26.5%)** (male=930, female=939)
- **Total trusted users are 5,174 (73.5%)** (male=2,625, female=2,549)
- **Imbalanced Distribution:** The dataset reveals a class imbalance, with a clear majority (**73.5%**) of customers remaining loyal to their service provider.
-
- The churn rate stands at **26.5%**. While this is the minority class, it implies that roughly **1 in 4 customers** are leaving.

3.2. Churn Distribution by Gender

This grouped bar chart analyzes the relationship between customer gender and churn status. The X-axis represents the gender categories (Male/Female), while the Y-axis displays the customer count. The bars are color-coded by churn status (No/Yes) to facilitate a direct comparison of attrition between the two demographic groups.



Summary & Insight:

- **Retained Customers (No Churn):**

  1.Female: **2,549** (49.3% of total retained)

  2.Male: **2,625** (50.7% of total retained

- **Churned Customers (Yes Churn):**

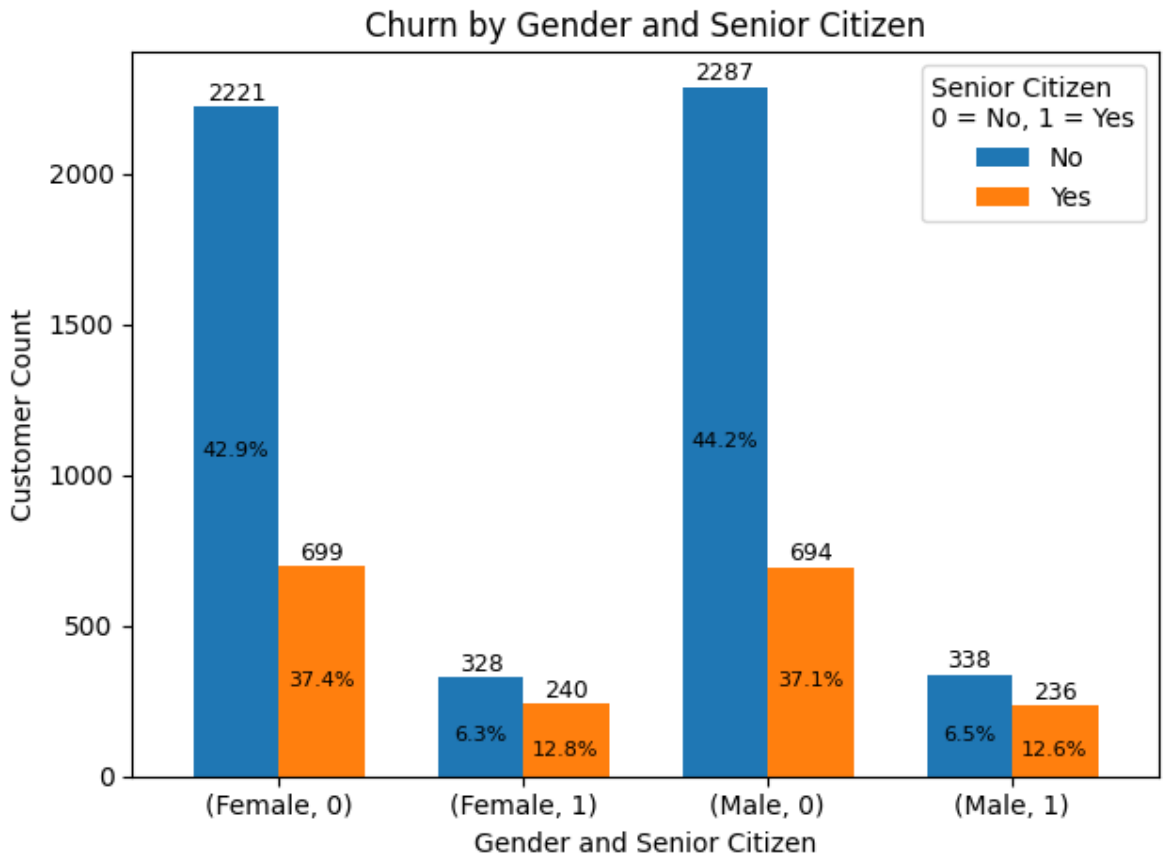
    1.Female: **939** (50.2% of total churned)

    2.Male: **930** (49.8% of total churned)

- Gender **Neutrality:** The churn behavior is nearly identical for both genders (approximately a 50/50 split in both groups). This indicates that **gender is not a driving factor** in customer attrition; men and women are equally likely (or unlikely) to leave the service.

3.3. Churn by Gender and Senior Citizen Status

Description: This multi-grouped bar chart segments the customer base by both Gender and Senior Citizen status (0 = Non-Senior, 1 = Senior) to analyze their combined effect on churn. The X-axis represents the four demographic combinations (e.g., Female Non-Senior, Female Senior), while the bars display the count of retained versus churned customers, annotated with specific percentages.

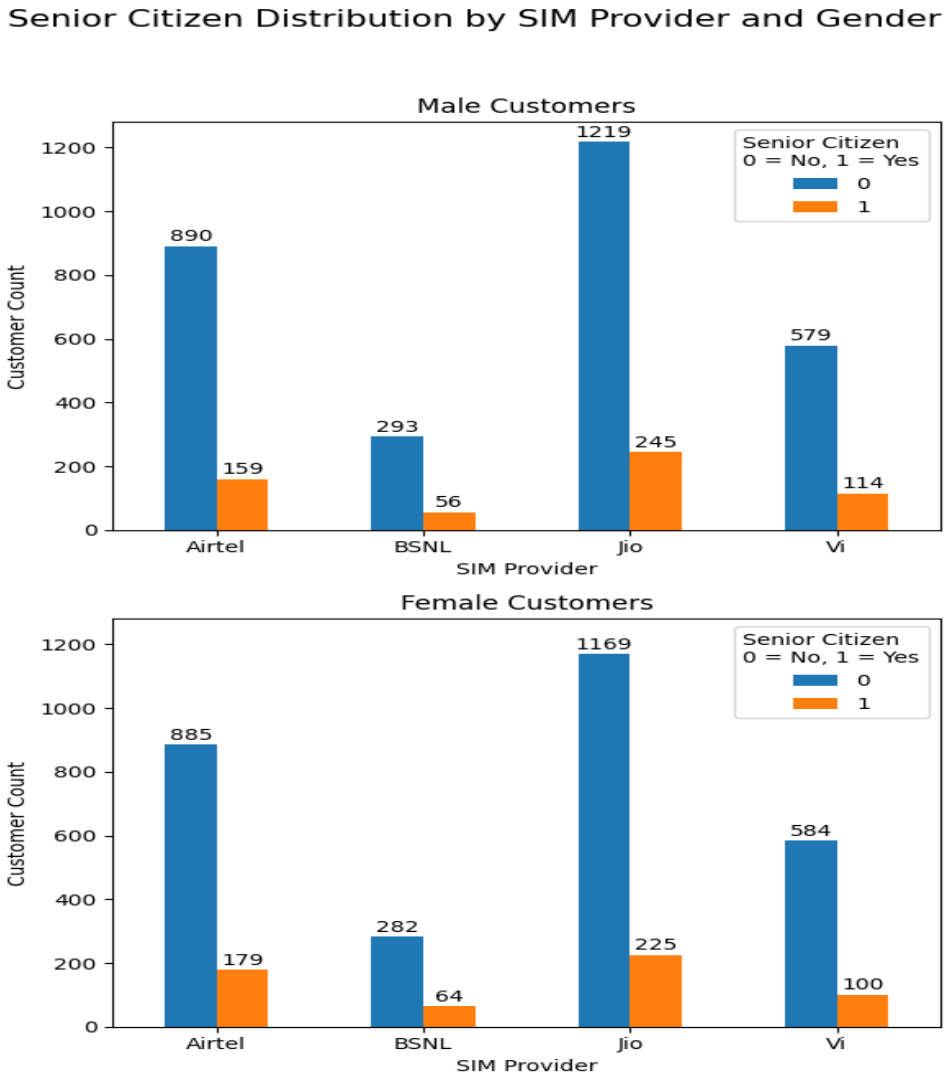


Summary & Insight:

- **Total trusted users are 5,174** (Male=2,625, Female=2,549)
- **Total churn users are 1,869** (Male=930, Female=939)
- **Total Male Senior Citizen are 574** (Churn=236, No Churn=338)
- **Total Male Non-Senior Citizen are 2,981** (Churn=694, No Churn=2,287)
- **Total Female Senior Citizen are 568** (Churn=240, No Churn=328)
- **Total Female Non-Senior Citizen are 2,920** (Churn=699, No Churn=2,221)
- **Age Over Gender:** Senior Citizens exhibit a significantly higher churn rate (~41%) compared to non-seniors (~23%), a trend that is consistent across both genders.
- The data confirms that **age is a critical predictor** of attrition while gender is not, identifying older customers as a high-priority group for retention efforts.

3.4. Senior Citizen Distribution by Service Provider & Gender

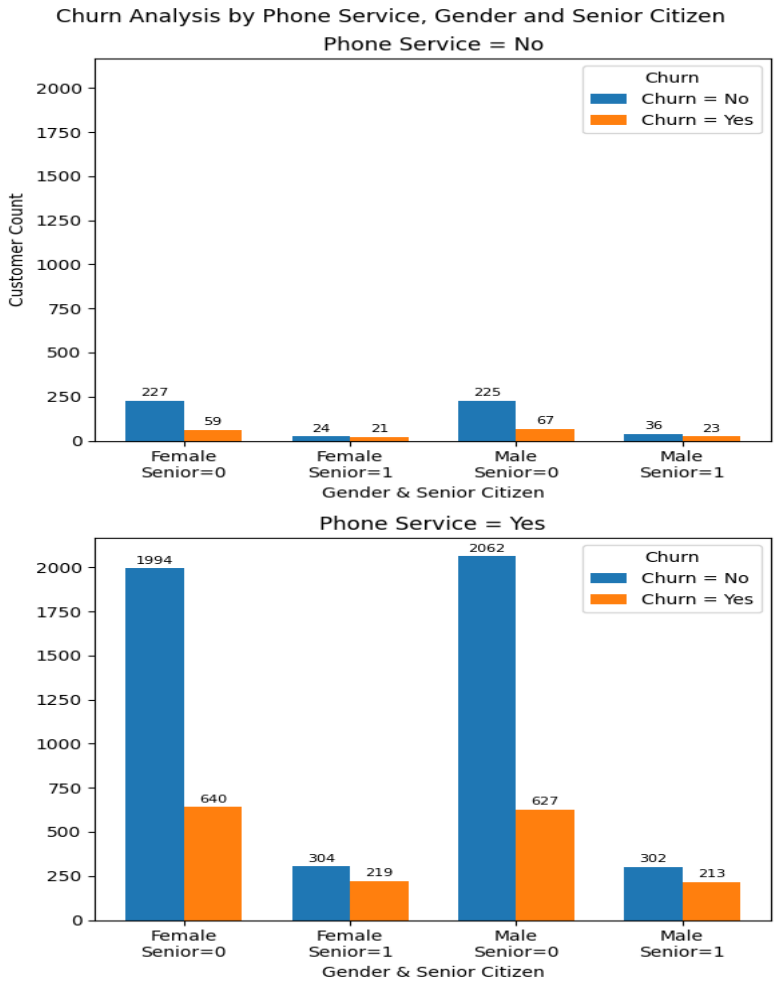
Description: This visualization utilizes vertically stacked subplots to analyze the distribution of Senior Citizens across different Service Providers, segmented by gender. The top chart displays data for Male customers, while the bottom chart represents Female customers. Each plot compares the volume of Senior vs. Non-Senior subscribers for every provider (Jio, Airtel, Vi, BSNL).



Summary & Insight: The pattern is the same for both men and women; neither gender shows a unique preference for a specific provider based on age.

3.5. Churn Analysis by Phone Service, Gender & Senior Citizen

Description: This chart splits customers into two main groups: the top graph shows those **without** Phone Service, and the bottom graph shows those **with** Phone Service. For each group, it compares churn counts based on Gender and Senior Citizen status (Senior=0 vs Senior=1).

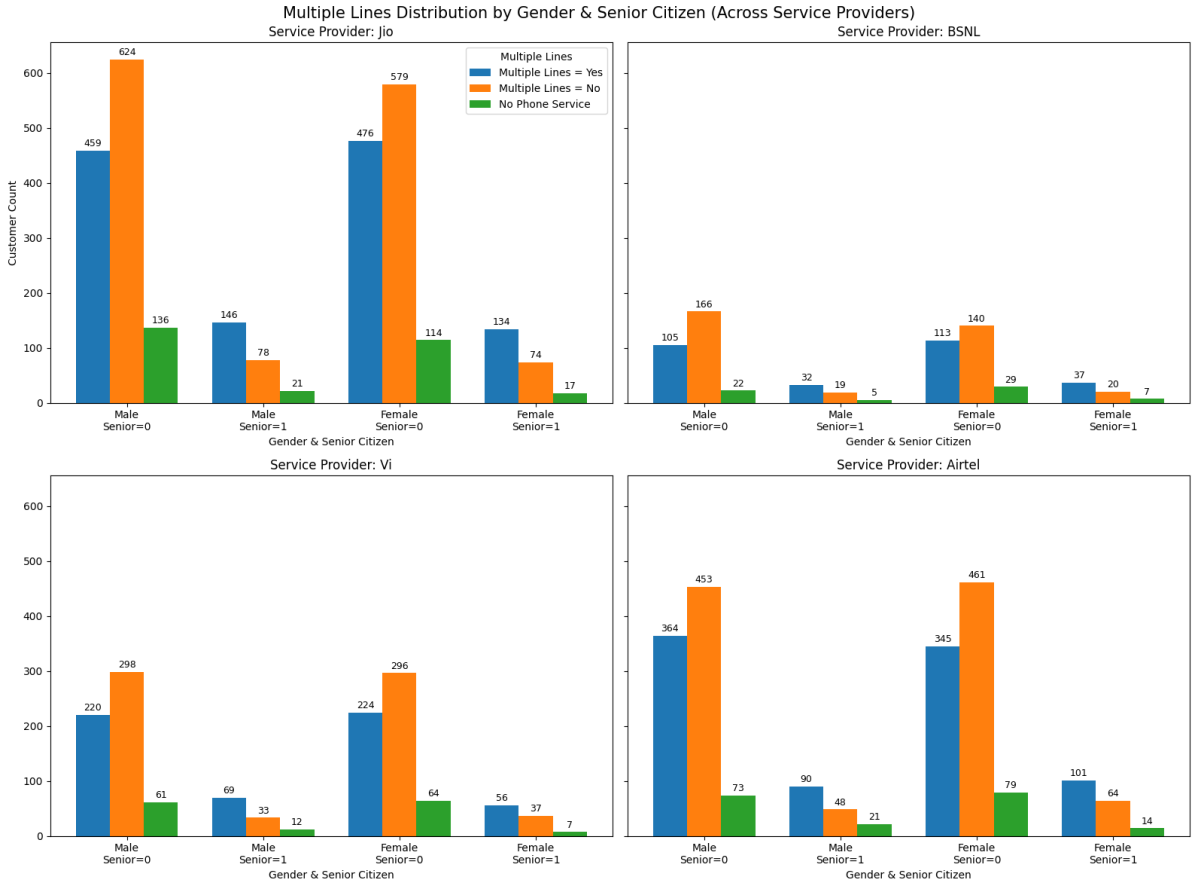


Summary & Insight:

- **Low Volume for "No Phone":** Very few customers do not have phone service (top chart has low numbers). Most customers subscribe to phone service.
- **High Senior Churn:** In both groups, Senior Citizens (Senior=1) show a higher tendency to leave compared to younger customers.
- **Gender Consistency:** Men and women behave almost exactly the same way. Whether they have phone service or not, gender does not affect the churn rate.

3.6. Multiple Lines -> Gender, Senior Citizen By Service Provider

This figure contains four charts, one for each Service Provider (Jio, Airtel, Vi, BSNL). It breaks down customers by Gender and Senior Citizen status to see their preference for phone lines. The colored bars show how many users have "Multiple Lines," "Single Lines" (No), or "No Phone Service."



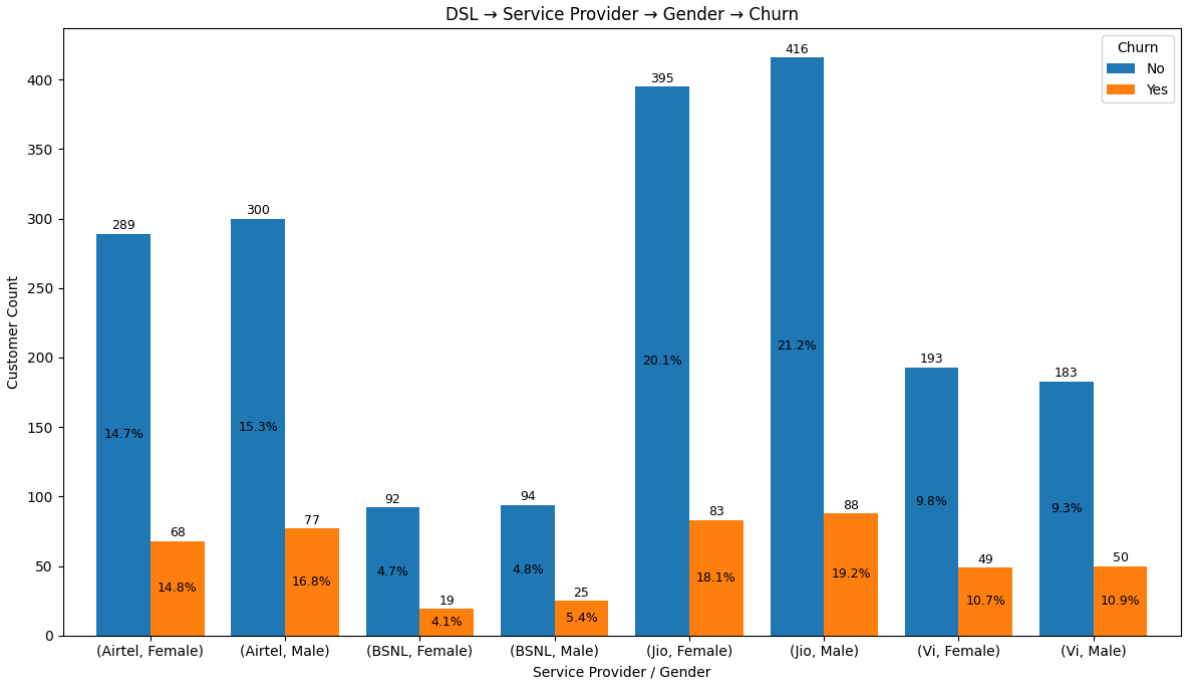
Summary & Insight:

- Across all providers and demographics, the majority of customers opt for a single line rather than paying for multiple lines.
- Senior Citizens generally subscribe to fewer services overall; their bars are significantly shorter than the non-senior groups.
- The behavior is remarkably similar across all Service Providers (Jio, Airtel, etc.). No specific provider has a unique advantage in selling multiple lines to specific demographic groups.

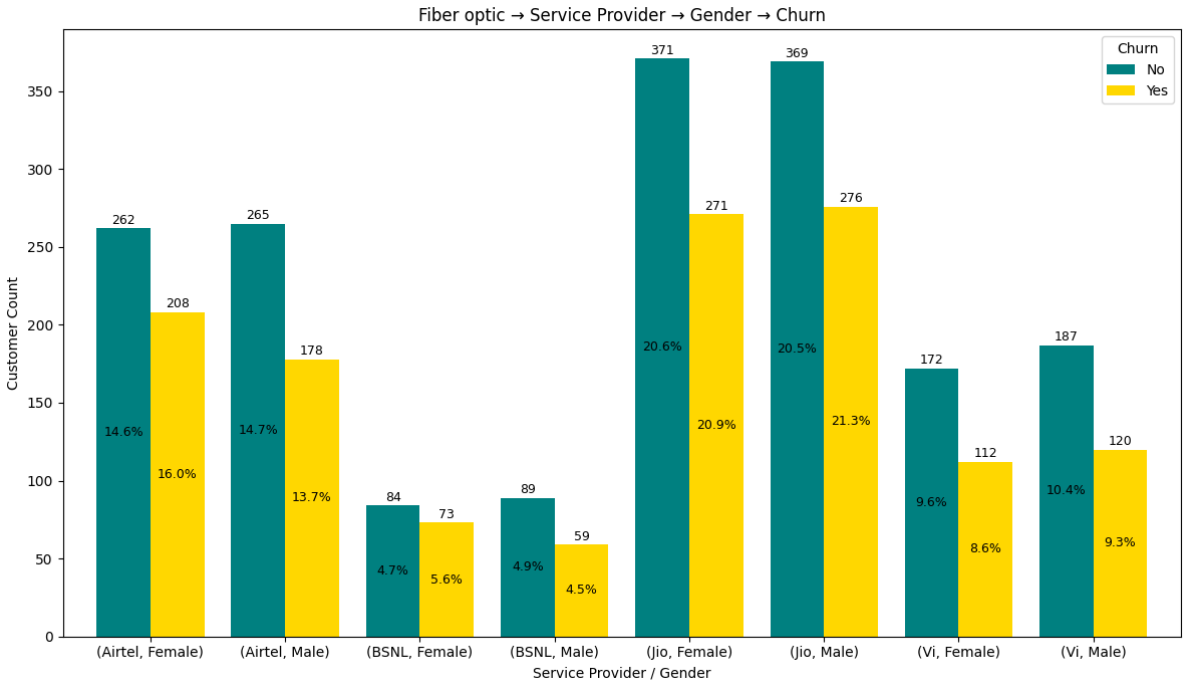
3.7. Churn Analysis by Internet Service Type

Description: This analysis consists of three separate bar charts, each focusing on a specific Internet Service type: **DSL**, **Fiber Optic**, and **No Internet**. For each type, the charts display the number of customers who stayed vs. those who left, further broken down by **Service Provider** and **Gender**. Percentages are shown inside the bars to indicate the exact churn rate for each group.

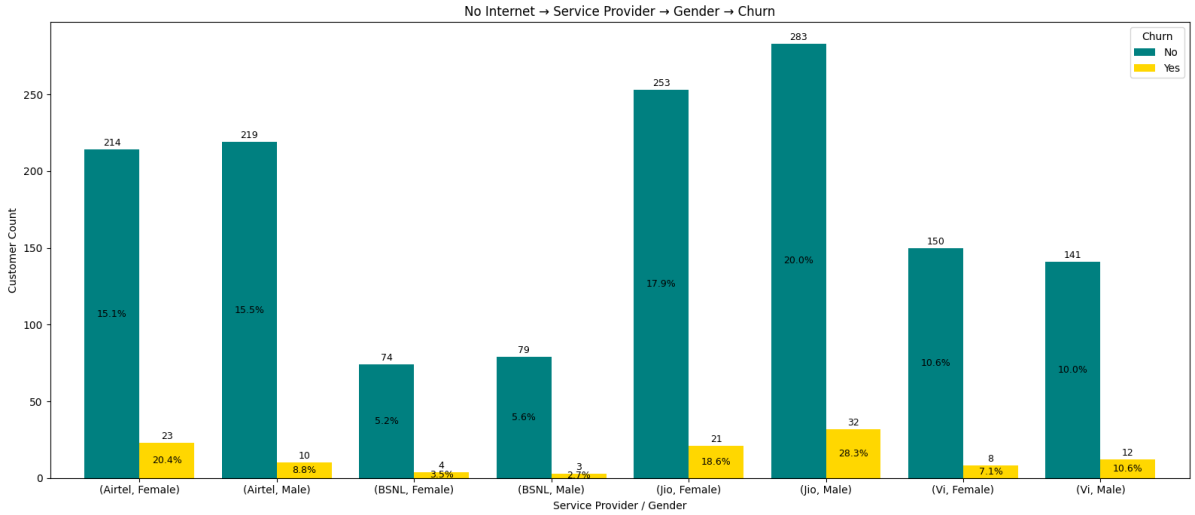
FOR DSL:



FOR FIBER OPTIC:



FOR NO SERVICE:

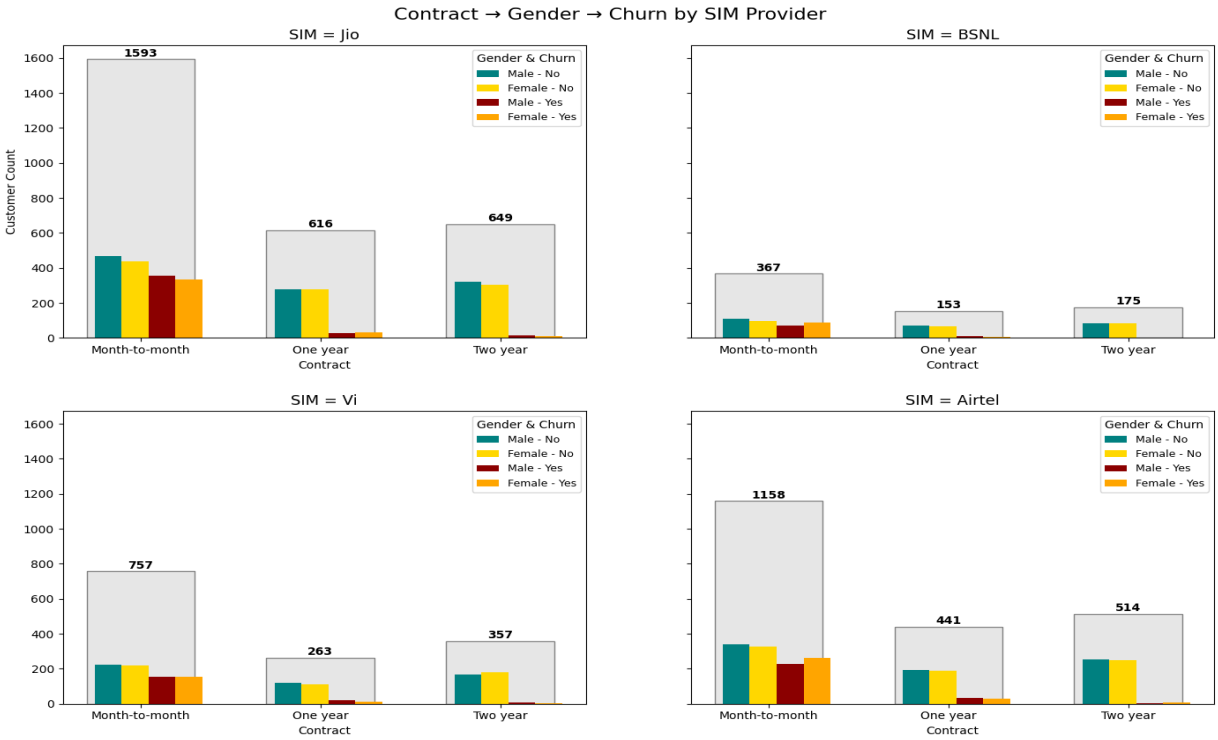


Summary & Insight:

- **Fiber Optic is High Risk:** Customers with Fiber Optic connections consistently show the **highest churn rates** (percentage of "Yes") compared to other internet types. This is a critical area for improvement.
- **DSL is More Stable:** DSL customers have moderate churn rates, significantly lower than Fiber users.
- **Non-Internet Users are Loyal:** Customers who do not subscribe to internet services (No Internet) have the **lowest churn rates**. They are the most stable and loyal customer segment.
- **Gender Consistency:** Across all three internet types and all service providers, male and female customers exhibit nearly identical behavior.

3.8. Contract Duration Analysis by Service Provider & Gender

Description: This figure displays four charts, one for each Service Provider (Jio, Airtel, Vi, BSNL). For every contract type (Month-to-month, One year, Two year), a **gray background bar** shows the *total* number of customers. The smaller colored bars in front break that total down by Gender and Churn status (e.g., Male-No, Female-Yes).

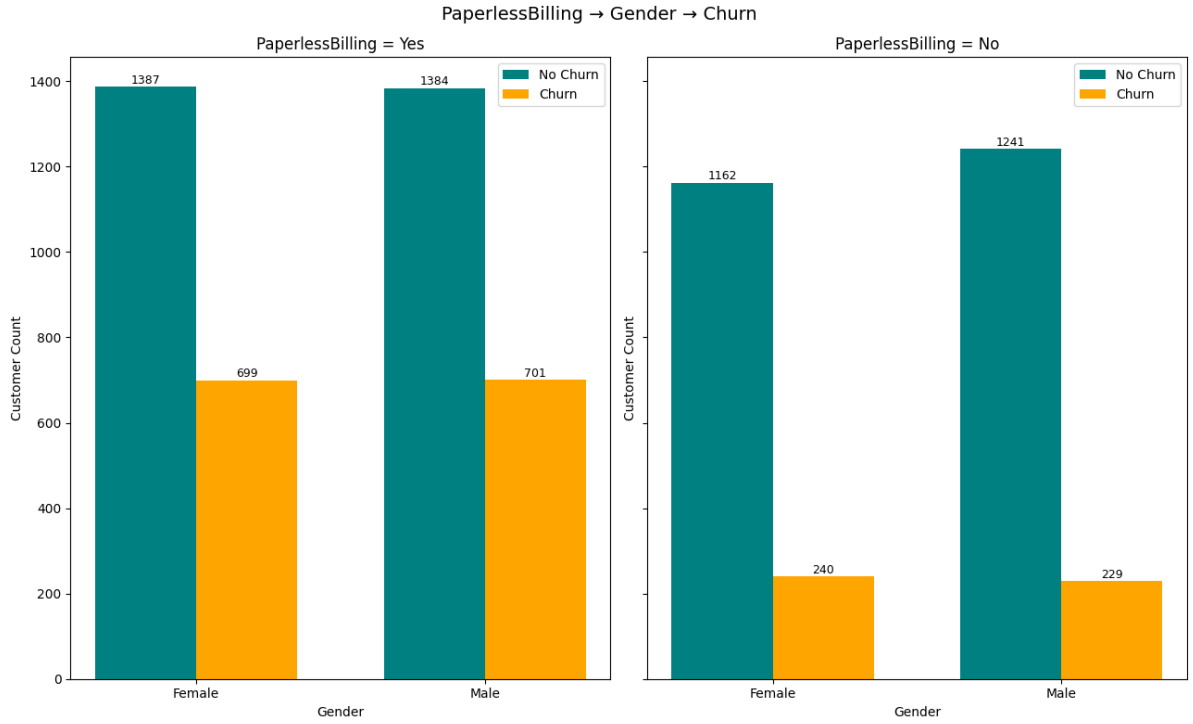


Summary & Insight:

- **High Churn in Short-Term:** "Month-to-month" contracts are the riskiest. They have the highest number of customers leaving (Red/Orange bars).
- **Loyalty in Long-Term:** Customers with "Two year" contracts almost never leave. Their churn bars are tiny or non-existent.
- **Volume Insight:** The gray bars show us that a large portion of customers prefer flexible (Month-to-month) contracts, despite the higher churn risk.
- **Gender Neutral:** Men and women show identical behavior across all contract types.

3.9. Paperless Billing & Churn by Gender

Description: This visualization splits the customers into two groups: those who use **Paperless Billing** (Yes) and those who receive traditional paper bills (No). For each group, grouped bar charts compare the number of customers who stayed (Teal) versus those who left (Orange), separated by gender.

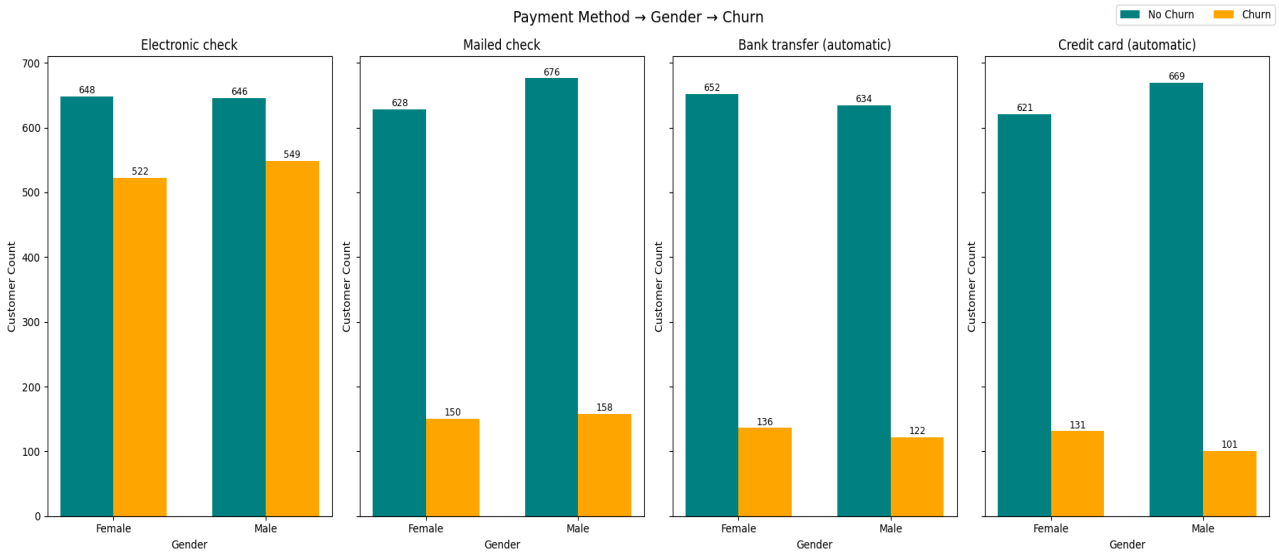


Summary & Insight:

- **Higher Churn with Paperless:** Customers who opt for Paperless Billing tend to have a **higher churn rate** (the orange bars are taller relative to the total) compared to those using paper bills.
- **Sticky Traditional Users:** Customers receiving paper bills appear more loyal and less likely to switch providers.
- **Gender Consistency:** As seen in previous plots, men and women show identical behavior. Gender does not influence whether a paperless user decides to churn.

3.10. Churn by Payment Method & Gender

Description: This visualization displays four separate charts, one for each payment method (e.g., Electronic Check, Mailed Check, Bank Transfer, Credit Card). Inside each chart, the bars compare the number of customers who stayed vs. those who left, separated by gender.

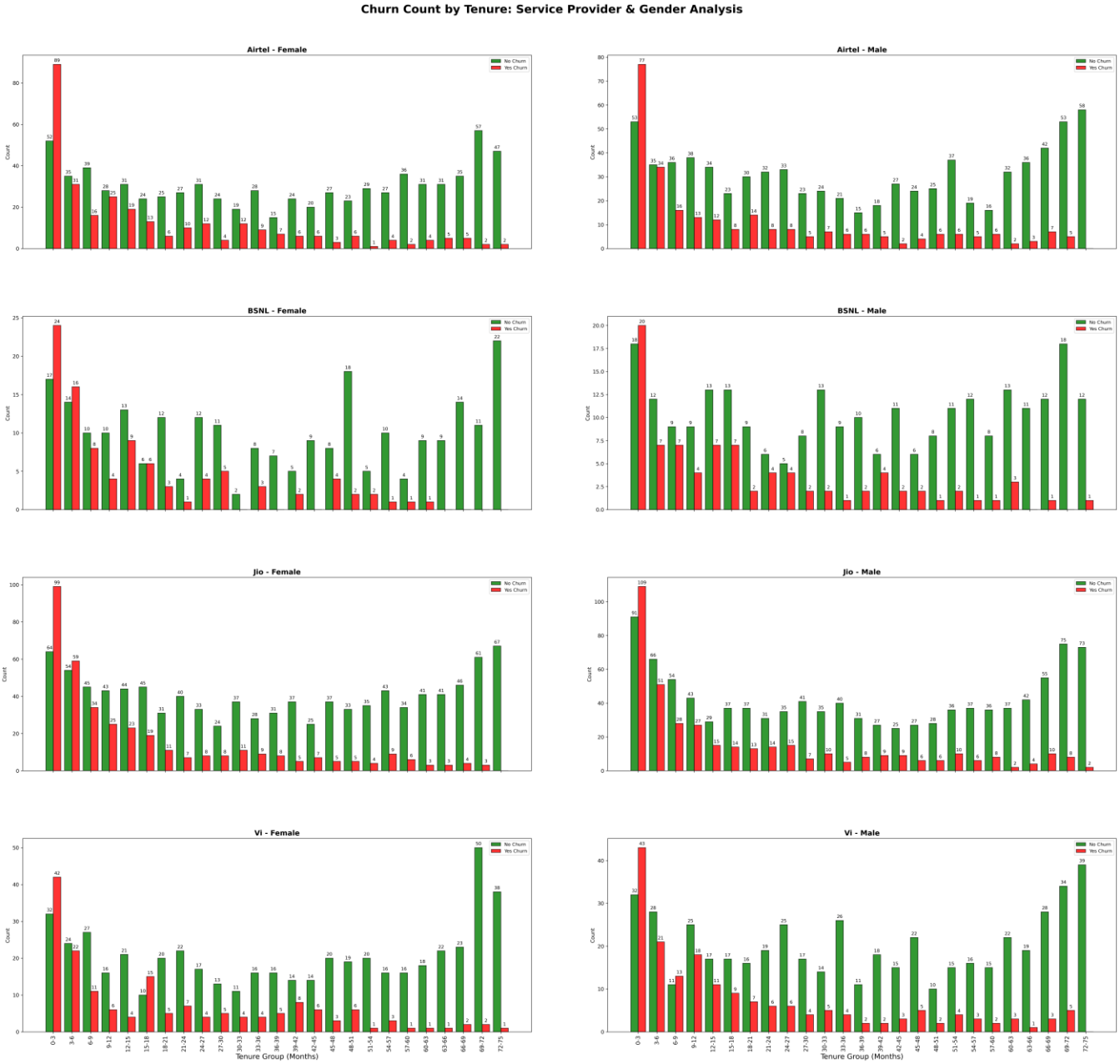


Summary & Insight:

- **High Risk with Electronic Checks:** Customers paying via "Electronic Check" have the **highest churn rate** by far. This is the most volatile segment.
- **Stability with Automatic Payments:** Customers using automatic methods like "Bank Transfer" or "Credit Card" are much more stable and loyal (very low churn).
- **Gender Neutral:** Once again, men and women show the exact same behavior across all payment methods. The type of payment influences churn, not the gender.

3.11. Tenure-Based Churn Analysis by Service Provider & Gender

Description: This comprehensive grid visualization analyzes *when* customers churn, broken down by **Service Provider** (Rows) and **Gender** (Columns). The X-axis represents customer tenure in **3-month intervals** (e.g., 0-3 months, 3-6 months). The side-by-side bars display the count of retained customers (Green) versus those who left (Red) for each specific time period.



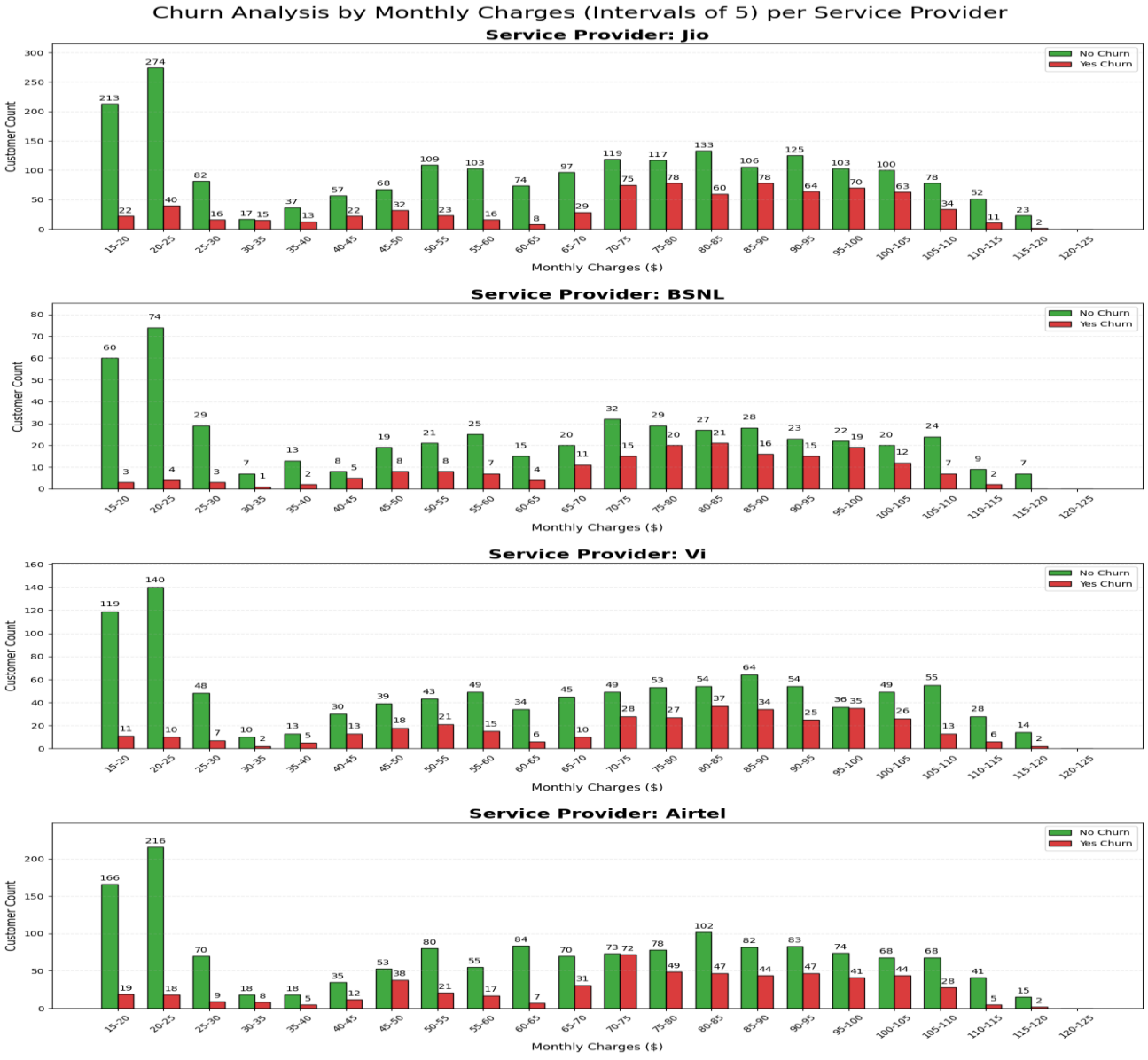
Summary & Insight:

- **The "Danger Zone" (0-6 Months):** Across all service providers, the **Red bars (Churn)** are tallest in the very first few bins (0-3 and 3-6 months). This indicates that new customers are the most likely to leave, making the onboarding experience critical.

- **Loyalty:** As tenure increases (moving right on the graph), the Red bars shrink and the Green bars grow. Customers who stay past the first year are significantly less likely to churn.
- **Provider Patterns:** This view allows for a direct comparison of provider health. If one provider (e.g., Vi or BSNL) has taller Red bars in the mid-tenure range compared to others, it indicates a specific retention issue for that brand.
- **Gender Neutrality:** Consistent with previous findings, the tenure-churn patterns are nearly identical for Males and Females.

3.12. Monthly Charges (Intervals of 5) & Churn per Service Provider

Description: This visualization analyzes customer retention based on spending habits by binning **Monthly Charges** into intervals of $5 (e.g., 15-20, 20-25). It generates a dedicated histogram for each Service Provider (Jio, BSNL, Vi, Airtel) to isolate price sensitivity trends. The dual-bar system (Green for No Churn, Red for Yes Churn) illustrates the volume of customers in each price bracket and reveals how attrition rates fluctuate as the cost of service increases.

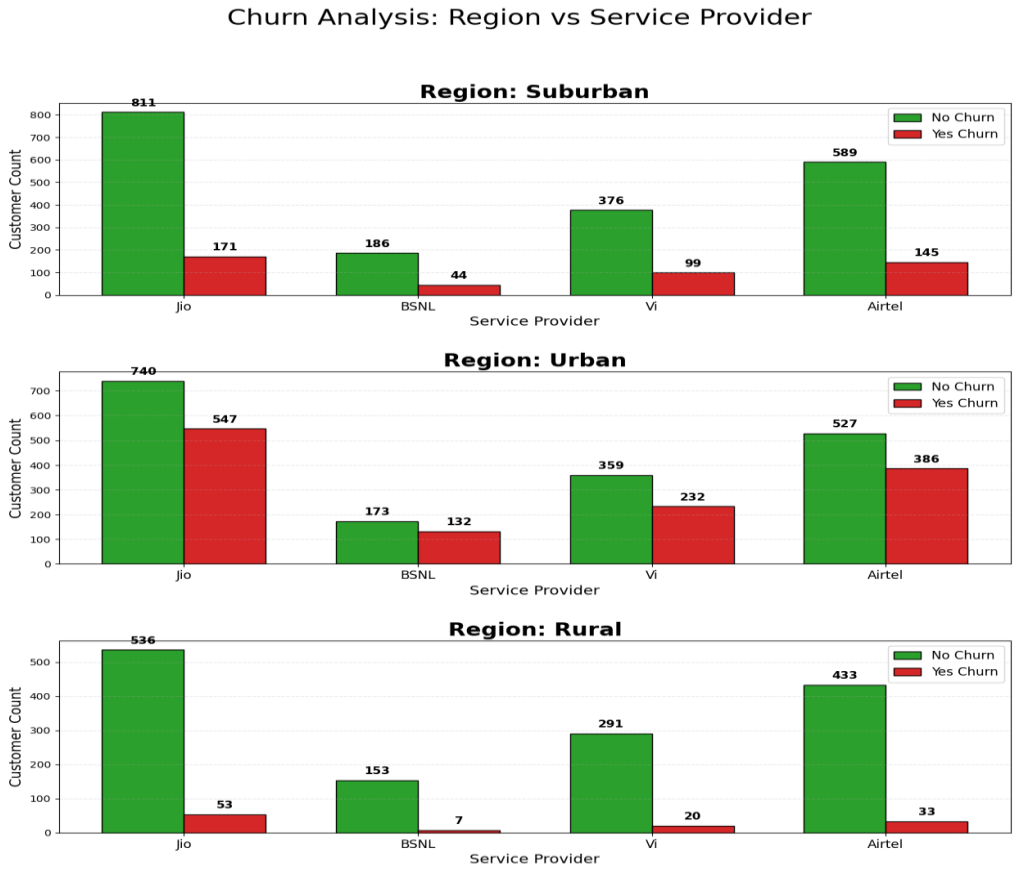


Summary & Insight:

- **High-Value Customer Risk:** A clear trend visible across all providers (particularly **Vi** and **Airtel**) is that churn rates increase significantly as monthly charges rise. For example, in the **$95-100** bracket for **Vi**, the number of churned customers (35) is nearly equal to retained customers (36), indicating a dangerous 50% churn rate among premium users.
- **Stickiness of Budget Segments:** The **$15-$25** price range consistently holds the largest volume of customers with the lowest proportional churn across all providers. This suggests that budget users are highly stable and less likely to switch, providing a reliable revenue baseline.
- **Provider-Specific Vulnerabilities:** While **Jio** has the highest overall customer volume, it sees a noticeable spike in absolute churn numbers in the mid-to-high range (**$70-$90**), suggesting that while they capture the mass market easily, they face stiffer competition when retaining mid-tier spenders. **BSNL** remains a niche player with lower overall volume but maintains very low churn relative to its size in the lowest price bands.

3.13. Region & Churn by Service Provider

Description: This visualization segments the analysis by geographic location (e.g., Urban, Suburban, Rural). It generates a dedicated chart for each region to compare the performance of all Service Providers side-by-side. The bars illustrate the volume of retained customers (Green) versus churned customers (Red), allowing for a direct comparison of market share and attrition rates within each specific area.



Summary & Insight:

- **Regional Market Share:** The chart highlights which providers dominate specific geographies. For instance, one provider might have a massive user base in Urban areas but a smaller footprint in Rural regions.
- **Localized Churn Issues:** It helps identify if a provider's churn is a general issue or a regional one. If a provider has high churn *only* in the "Rural" chart, it suggests coverage or infrastructure issues specific to that terrain.
- **Competitive Landscape:** By comparing the ratio of Red to Green bars within a single plot, we can see which provider is best at retaining customers in that specific region compared to its local competitors.

3.14. Region Vs Gender Vs Senior Citizen Vs Service Provider -> Churn

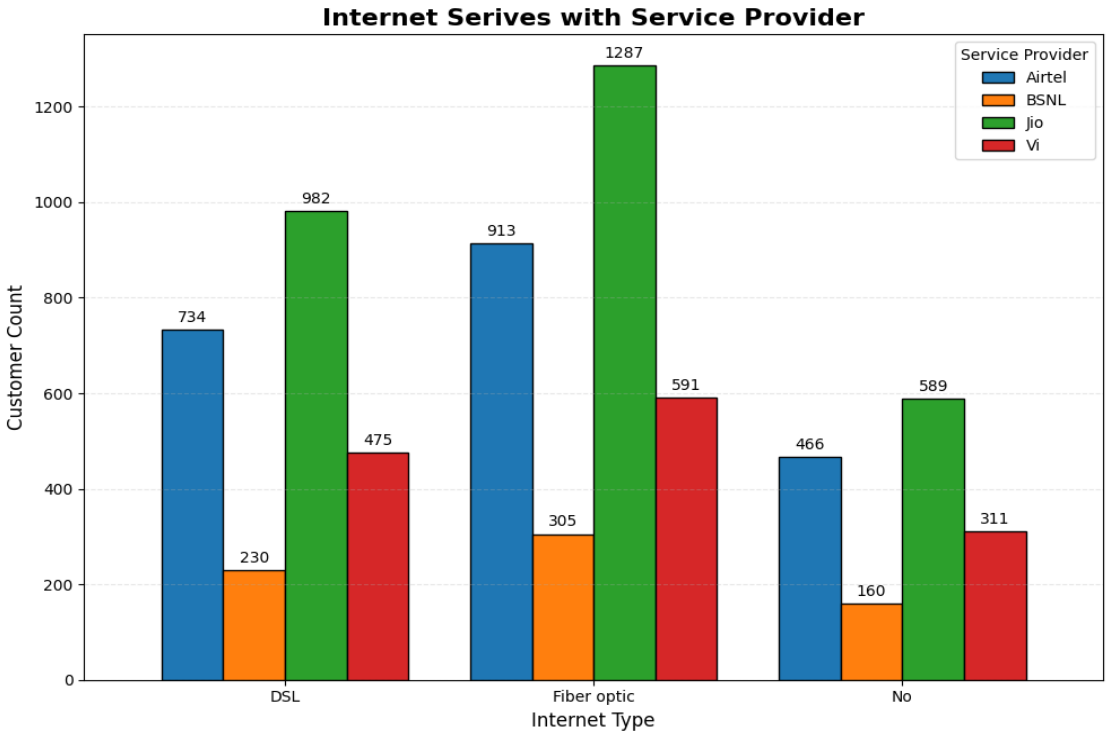
Description: This multi-dimensional matrix visualization integrates all key categorical variables into a single high-level view. The grid is structured with **Regions** (Urban, Suburban, Rural) as rows and **Service Providers** (Jio, Airtel, Vi, BSNL) as columns. Within each specific subplot, the customer base is segmented by **Gender** and **Senior Citizen Status** (Sr: No vs. Sr: Yes). The green and red bars allow for a simultaneous comparison of volume (market share) and attrition (churn rate) across every possible segment of the customer population.

Summary & Insight:

- **Hyper-Local Performance:** This view exposes specific regional weaknesses that broad averages might hide. For instance, a Service Provider might have excellent retention with Urban Seniors but face a massive churn spike among Rural Seniors, indicating a coverage gap specific to that demographic's location.
- **Age-Based Vulnerability Check:** By comparing the "Sr: Yes" bars against the "Sr: No" bars within the same provider's column, we can instantly validate if the "High Senior Churn" trend is consistent across all providers or if one specific provider (e.g., BSNL) manages to retain older customers better than competitors.
- **Strategic Segmentation:** This chart is the ultimate tool for targeted marketing. It identifies exact "problem pockets" (e.g., "Male Seniors using Vi in Rural areas") rather than general issues, allowing for surgical retention strategies rather than blanket campaigns.

3.15. Internet Services with Service Provider

Description: This grouped bar chart segments the total customer base by **Internet Service Type** (DSL, Fiber optic, and No Internet) and **Service Provider** (Airtel, BSNL, Jio, Vi). It visualizes the market share distribution to identify which providers dominate the high-speed infrastructure (Fiber) versus legacy technology (DSL) before analyzing their respective churn rates.



Summary & Insight:

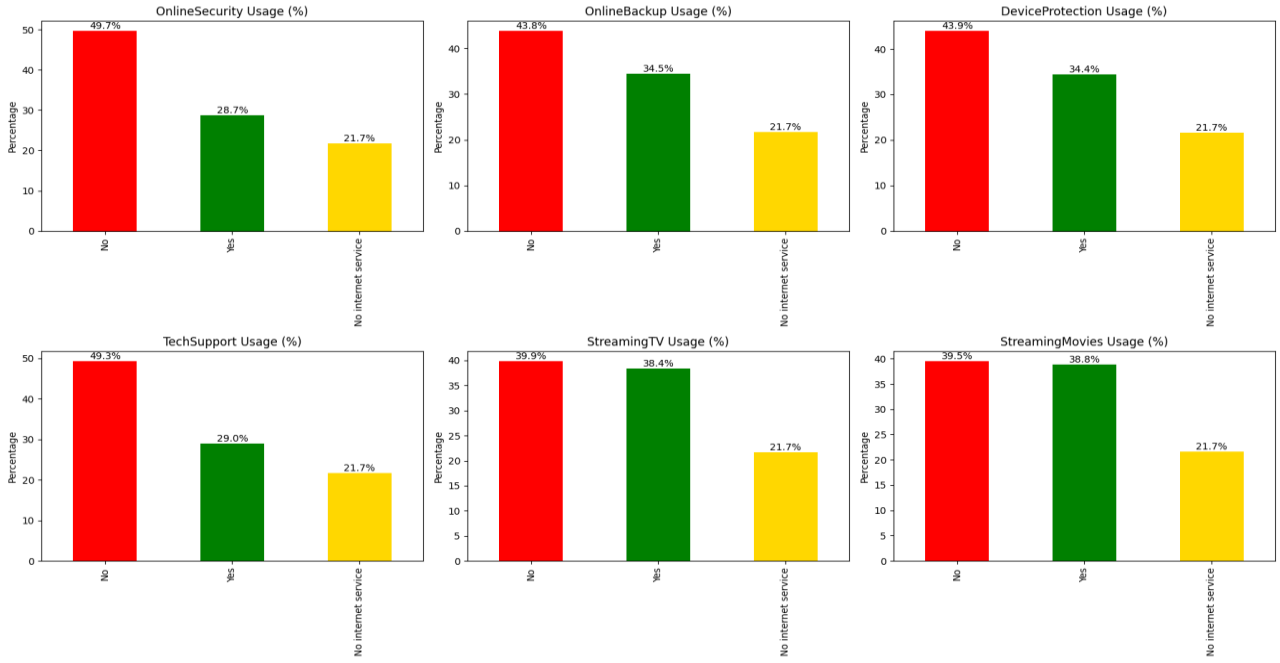
- **Fiber Optic Dominance:** The visualization clearly identifies a duopoly in the high-speed market. **Jio (1,287 customers)** and **Airtel (913 customers)** are the overwhelming leaders in the Fiber Optic segment. This suggests that the "Fiber

Churn" risks identified later in the report will disproportionately impact these two providers.

- **Lagging Competitors: Vi** (591) and **BSNL** (305) have significantly smaller footprints in the Fiber sector. Their inability to capture the high-volume Fiber market suggests they may be struggling to compete on infrastructure speed.
- **Jio's Market Lead: Jio** (represented by the Green bars) holds the highest customer count across all three categories—DSL, Fiber, and even non-internet users—indicating it is the most aggressive player in acquiring subscribers across the board.

3.16. Additional Services Vs Churn

Description: This grid of six bar charts provides a snapshot of customer adoption for various value-added services: **Online Security, Online Backup, Device Protection, Tech Support, Streaming TV,** and **Streaming Movies**. Each chart displays the percentage of customers who have subscribed ("Yes"), those who haven't ("No"), and those without internet service ("No internet service").



Summary & Insight:

- **Low Security Adoption:** Services like **Online Security** and **Tech Support** have relatively low adoption rates (high "No" bars). This represents a significant upselling opportunity, as customers with these services are typically "stickier."
- **Streaming Popularity:** Entertainment services like **Streaming TV** and **Streaming Movies** generally show higher adoption rates compared to technical support services.
- **Correlation with Churn:** (Contextual Insight) Typically, customers who subscribe to "sticky" services like Tech Support and Online Security exhibit lower churn rates. Identifying the low adoption here highlights a pathway to reduce churn: bundling these services more effectively.

4.OVERALL SUMMARY

Key Data Statistics

- **Churn Rate:** 26.5% (High risk) vs. 73.5% (Retained).
- **Gender:** Balanced (50.5% Male, 49.5% Female).
- **Service Providers:** Jio (Largest), Airtel, Vi, BSNL.

Based on the analysis of all 13 visualizations, here are the primary reasons customers are leaving:

A. The "New Customer" Cliff

The most critical finding is the **Tenure** analysis. The vast majority of churn happens in the **first 0–6 months**. If a customer stays past the first year, they are highly unlikely to leave. The onboarding experience is currently failing to lock customers in.

B. Contract & Payment Friction

- **Contract Type:** Customers on **"Month-to-Month"** contracts are extremely volatile and leave frequently. In contrast, those on **1-Year or 2-Year contracts** are nearly 100% loyal.
- **Payment Method:** Users paying via **"Electronic Check"** have the highest churn rate. Users on automatic payments (Credit Card/Bank Transfer) rarely churn. This suggests that the manual effort of paying bills monthly leads to cancellation.

C. Service-Specific Issues

- **Fiber Optic Dissatisfaction:** Customers with **Fiber Optic** internet churn at much higher rates than DSL or No-Internet users. This points to potential pricing issues or technical instability with the Fiber product.
- **Missed "Stickiness":** Services that usually make customers stay—like **Online Security** and **Tech Support**—have very low adoption rates.

D. Demographics

- **Senior Citizens:** This group churns at a significantly higher rate than younger users.
- **Gender:** Gender is **not** a factor. Men and women behave identically.