Customer Churn Analysis Project (Telco Dataset)

Project Title :

Customer Churn Analysis using Exploratory Data Analysis (EDA) and Feature Engineering

---

Problem Statement

Customer churn refers to customers who discontinue using a company's services. High churn rates directly impact revenue and long-term business growth.

The objective of this project is to analyze customer behavior using exploratory data analysis (EDA) and identify key factors that influence customer churn. These insights can help businesses design effective customer retention strategies and serve as a foundation for predictive machine learning models.

Dataset Description

- **Dataset Name:** Telco Customer Churn
- **Source:** IBM Sample Dataset
- **Number of Rows:** ~7,000
- **Number of Columns:** 21

Key Features

- **Demographic Features:** gender, SeniorCitizen
- **Service-related Features:** InternetService, Contract, PaymentMethod
- **Account Information:** tenure, MonthlyCharges, TotalCharges
- **Target Variable:** Churn (Yes / No)

Tools & Libraries Used

- **Python**
- **Pandas:** Data manipulation and preprocessing
- **Matplotlib:** Data visualization

Initial Data Checks

The following checks were performed to understand the dataset structure and quality: -

df.head() – Preview of the dataset

df.info() – Data types and missing values

df.describe() – Statistical summary of numerical features

Data Cleaning

Steps Performed

- Checked for duplicate records and removed them
- Handled missing values
- Converted TotalCharges from object to numeric
- Ensured consistency in data types

Purpose

✓ Ensure clean, consistent, and reliable data for analysis and modeling

---

Dataset Enhancement: SIM Type Feature

To enhance the depth of analysis, an additional categorical feature called SIM Type was manually introduced into the dataset. This feature represents the customer's mobile network provider such as Airtel, Jio, Vodafone, or Idea.

Note: The original IBM Telco Customer Churn dataset does not contain SIM or network provider information. This column was added purely for analytical and learning purposes to simulate real-world telecom datasets.

Data Preparation for SIM Type

- The SIM Type column was added after data cleaning.
- Values were assigned consistently across records.
- The feature was treated as a categorical variable during EDA.

Purpose:

✓ Analyze churn behavior across network providers
✓ Improve business-oriented insights

Exploratory Data Analysis – Visualization Insights

1. Churn Distribution

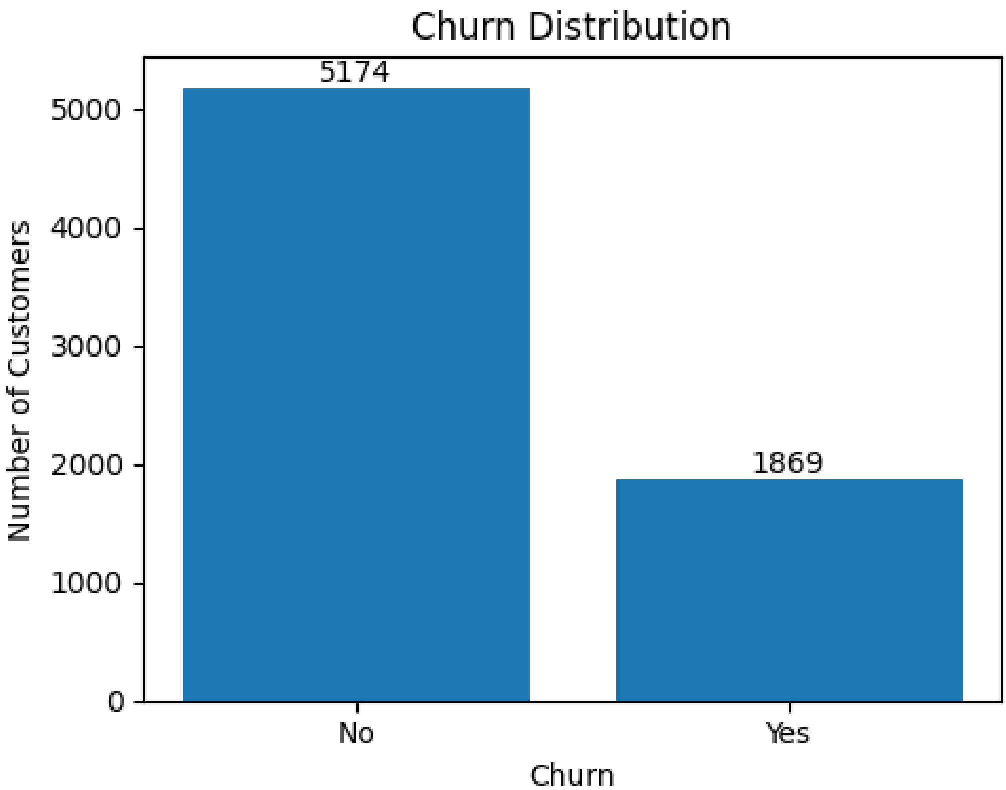


Figure: Overall Churn Distribution

Counts:
- Non-Churn Customers: ~5174
- Churned Customers: ~1869

Insight:
- Majority of customers did not churn.
- A significant minority of customers churned, indicating a class imbalance.

Business Interpretation:
✓ Retention strategies should focus on the smaller but impactful churn segment.
✓ Imbalanced data should be handled carefully during model building.

2. Contract Type vs Churn

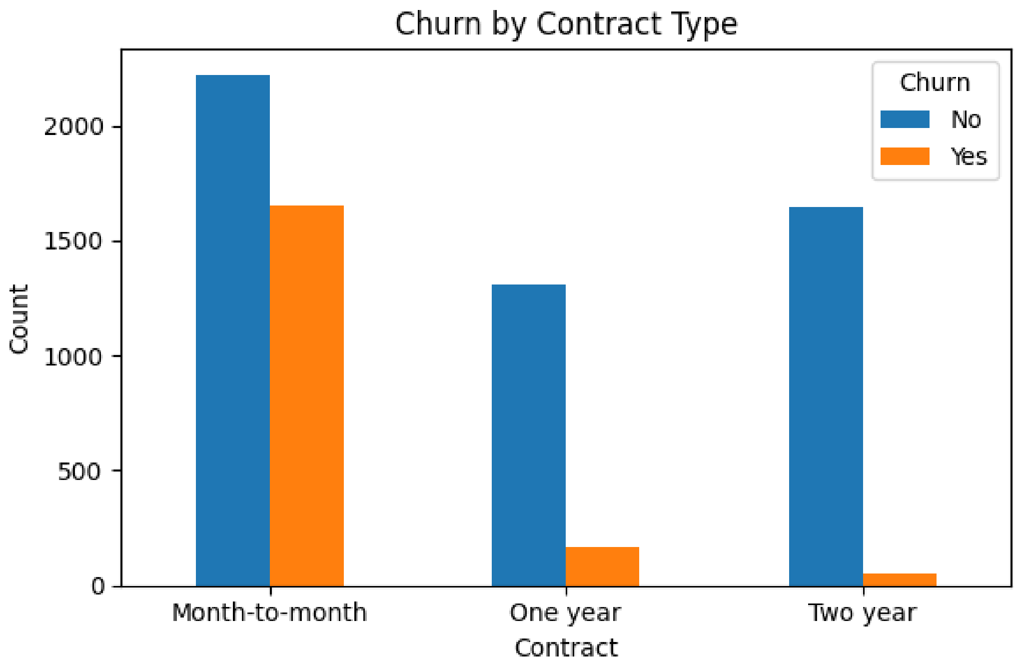


Figure: Churn by Contract Type

Insight:
- Month-to-month customers have the highest churn count.
- One-year contract customers show much lower churn.
- Two-year contract customers have minimal churn.

Business Interpretation:
✓ Longer contract durations improve customer retention.
✓ Incentivizing long-term contracts can significantly reduce churn.

✓ Contract type is the strongest churn driver.

3. Gender vs churn

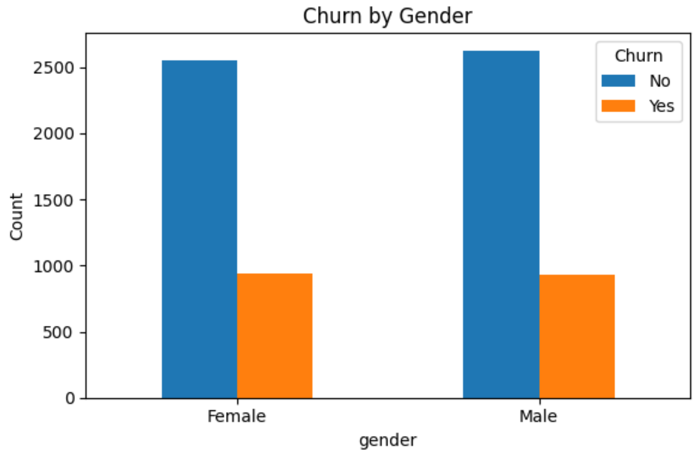


Figure: Churn by Gender

Insight:
Male and female customers show nearly identical churn patterns.
Gender has minimal impact on churn.
Churn distribution is nearly equal across genders.

Business Interpretation:
✓ Gender is not a strong predictor of churn.
✓ Retention strategies should not be gender-specific.

4. Customer Tenure Distribution

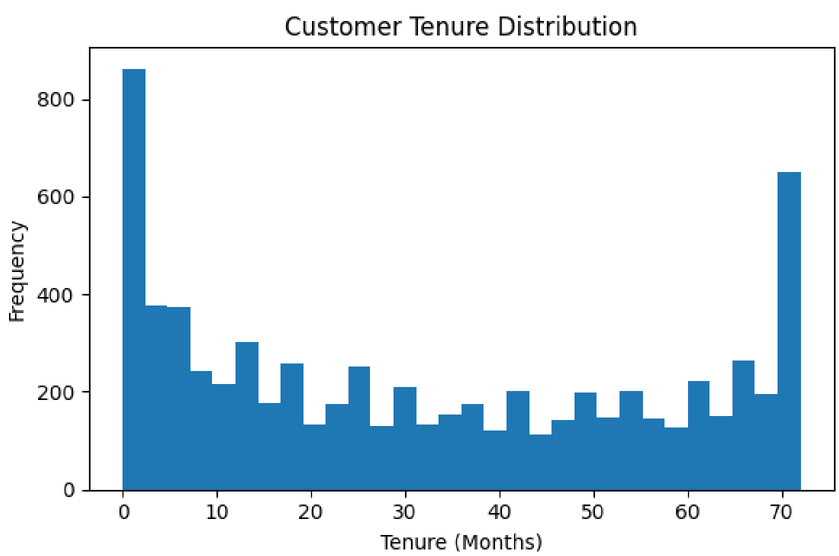


Figure: Customer Tenure Distribution

Insight:
- Many customers churn within the early months of service.
- Customers with longer tenure are more stable.

Business Interpretation:
✓ Early engagement and onboarding are critical.
✓ Long-term customers demonstrate higher loyalty.

Conclusion:
✓ Low tenure is a major churn risk.

5. Gender Distribution by Senior Citizen Status

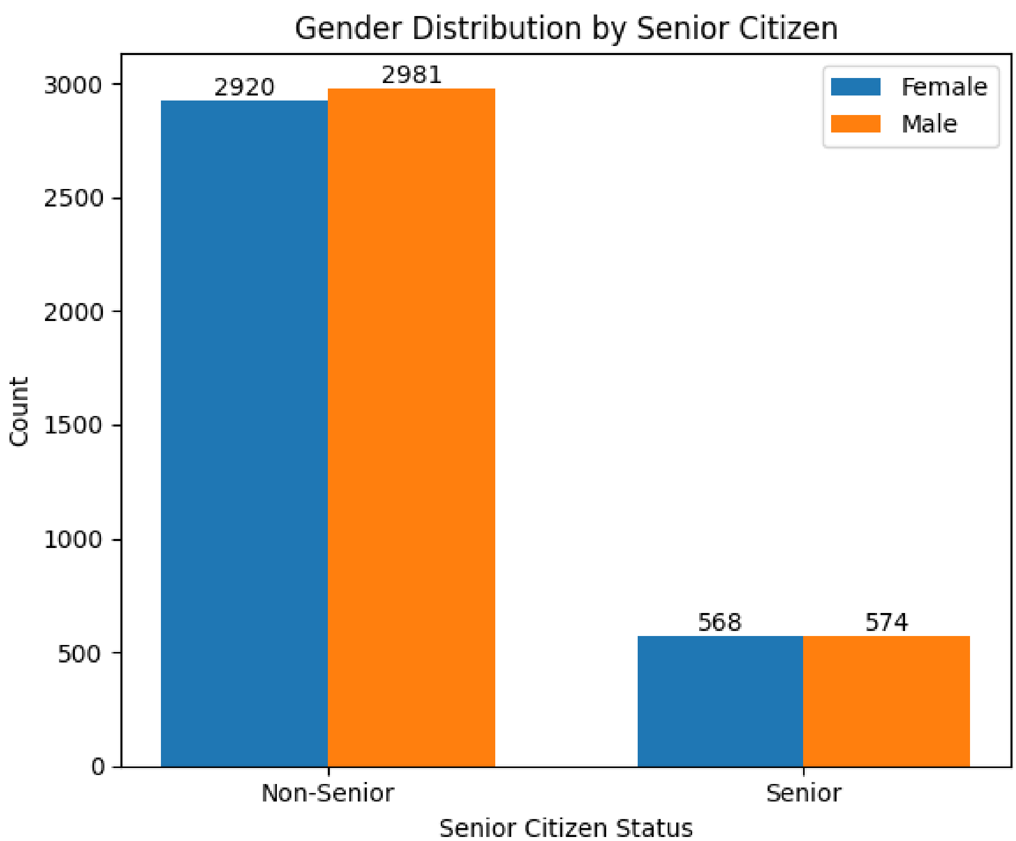


Figure: Gender Distribution by Senior Citizen

Insight:
- Senior citizens form a smaller group but show higher churn tendency.
- Senior citizen status is more relevant than gender.

- Gender distribution among senior and non-senior customers is balanced.

Business Interpretation:
✓ Senior citizen status may influence churn more than gender.
✓ Specialized plans for senior customers may improve retention

6. Overall Gender Distribution

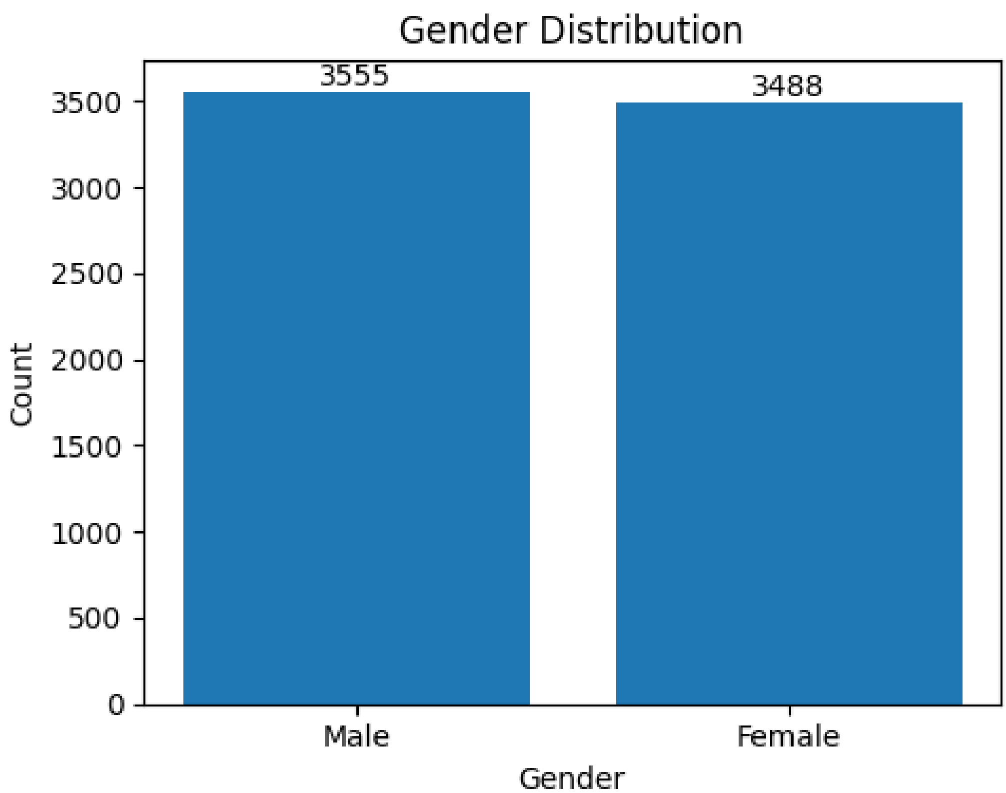


Figure: Overall Gender Distribution

Insight:
- Dataset is gender-balanced, reducing bias.

7. Gender vs Internet Service Type

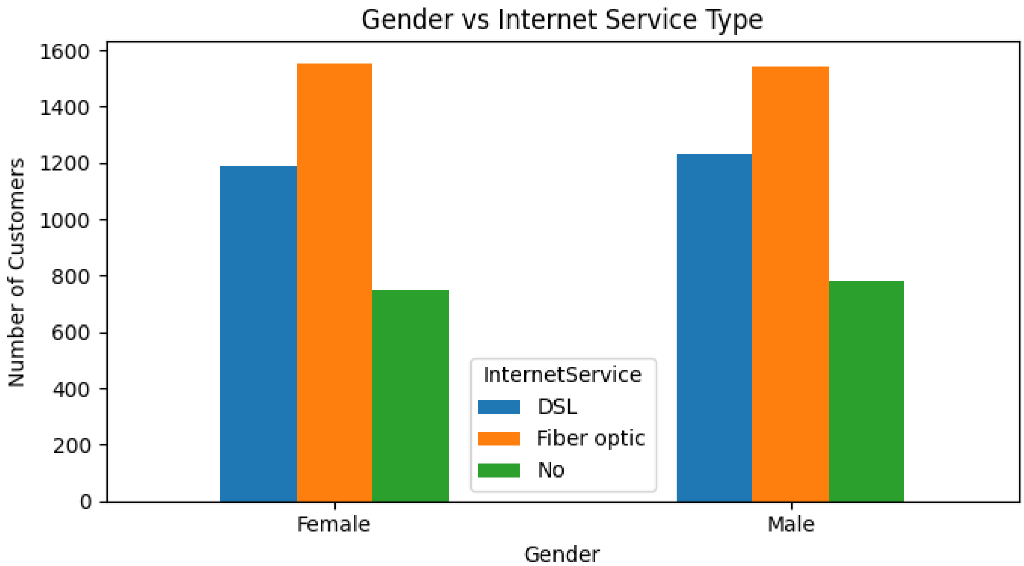
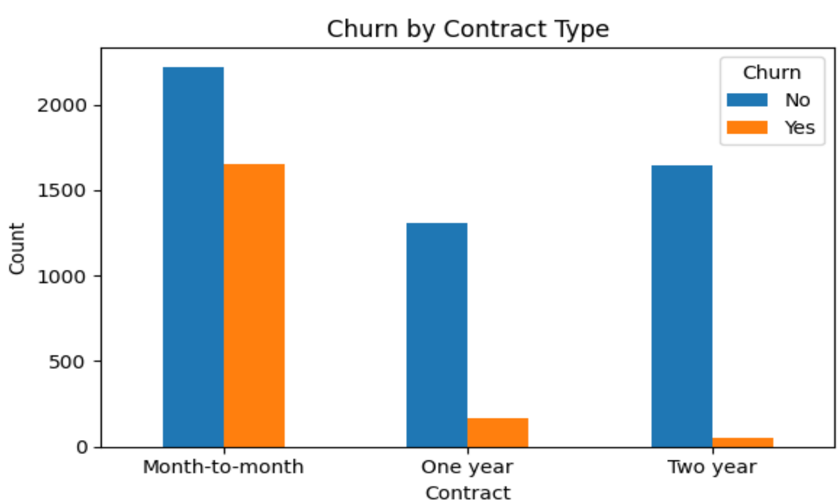


Figure: Gender vs Internet Service Type

Insight:
- Fiber optic users dominate and also show higher churn in previous analysis.
- Fiber optic is the most used internet service across genders.
- DSL and non-internet users are fewer in comparison.

Business Interpretation:
✓ High fiber optic adoption aligns with higher churn observed earlier.
✓ Service quality and pricing of fiber optic plans should be reviewed.

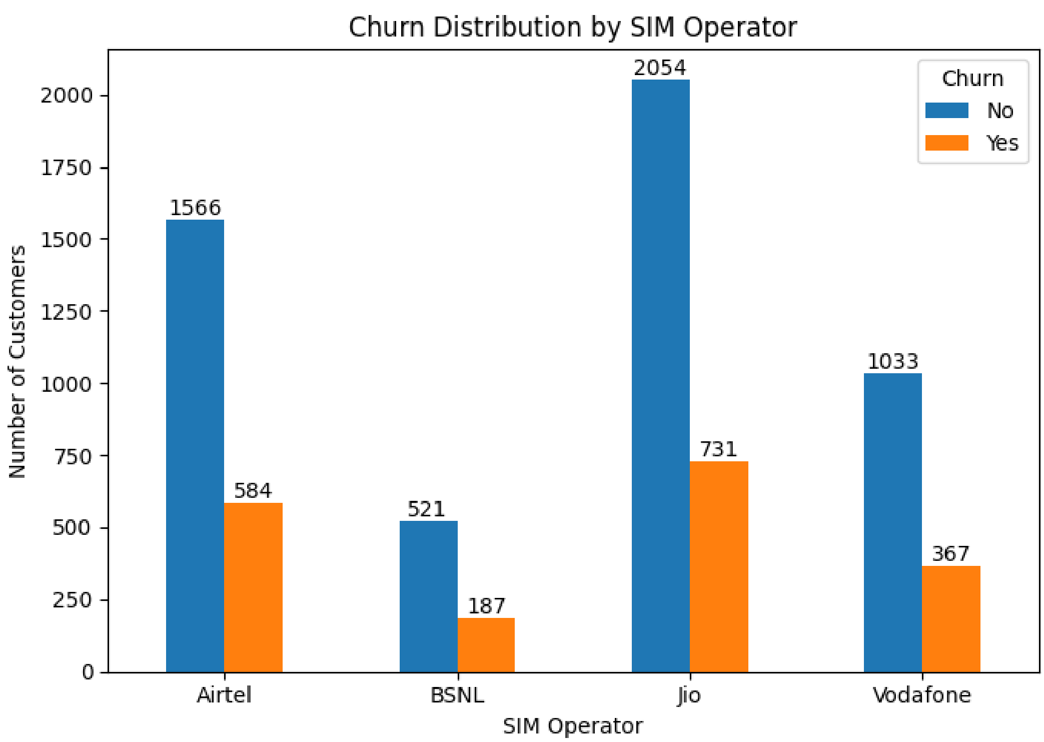
8. Monthly & Quarterly Charges vs Churn



Insight:
- Customers with higher monthly charges churn more frequently.
- Quarterly aggregation shows churn spikes in high-billing periods.

Conclusion:
✓ Pricing sensitivity is a key churn factor.

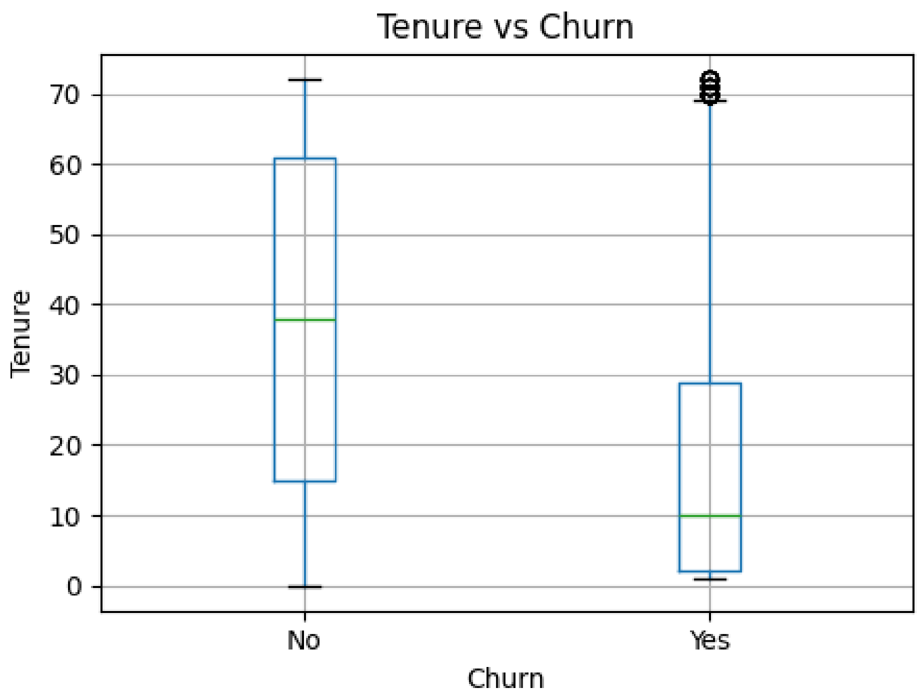
9. SIM Type vs Churn



Insights:
- Certain SIM providers show higher churn than others.
- Indicates impact of network quality, pricing, or coverage.

✓ SIM Type adds domain realism and improves churn understanding.

10. Tenure vs Churn



Insight:
- Churn probability decreases as tenure increases.
- Early-stage customers require proactive engagement.

Overall Performance & Churn Drivers

Major contributors to churn are
✓ Month-to-month contracts
✓ Low tenure customers
✓ High monthly charges
✓ Fiber optic internet users
✓ Certain SIM types

These factors should be prioritized in churn reduction strategies and predictive modeling.