Customer Churn Analysis Report

1. Executive Summary

This project focused on analysing customer churn patterns at a telecom company and building predictive models to identify customers at risk of leaving. Using SQL queries and Exploratory Data Analysis (EDA), we examined customer demographics, service usage, pricing factors, and contract details to uncover key drivers of churn. Machine learning models — Logistic Regression, Random Forest, and XGBoost — were developed to predict churn risk, achieving up to 79% accuracy.

Key findings revealed an overall churn rate of approximately 26%, with higher churn rates among month-to-month contract customers, Fiber optic internet users, and customers with high monthly charges. Shorter customer tenure and use of Electronic check payment methods were also strong churn indicators.

To reduce churn, it is recommended that the company offer loyalty rewards to high-risk segments, enhance onboarding programs for new customers, and promote longer-term contracts with improved service offerings. Implementing these strategies could significantly improve customer retention and positively impact revenue growth.

2. Project Introduction

Project Objective

The objective of this project is to analyse customer data from a telecom company to uncover patterns and drivers of customer churn. Using insights from Exploratory Data Analysis (EDA) and SQL queries, and applying predictive modeling techniques, the goal is to identify key risk factors and build a model capable of predicting customers most likely to churn.

Importance of the Problem

Customer churn has a direct and significant impact on revenue growth, customer lifetime value, and business sustainability. Acquiring new customers is often far more expensive than retaining existing ones. By better understanding and proactively addressing churn, companies can strengthen customer loyalty, increase profitability, and improve long-term operational stability.

3. Data Overview

The dataset used for this analysis originates from a telecom company and contains detailed information on customer demographics, account information, services subscribed to, and whether or not the customer churned.

Key Details:

Customer Churn Analysis Report

- Number of rows: 7,044 customers
- Number of columns/features: 21 features
- Target variable: `Churn` (Yes/No)

Major Categories of Features:

- Customer Demographics:
  - `gender`
  - `SeniorCitizen` (binary: 0 = No, 1 = Yes)
  - `Partner` (Yes/No)
  - `Dependents` (Yes/No)
- Account Information:
  - `tenure` (months with the company)
  - `Contract` (Month-to-month, One year, Two year)
  - `PaperlessBilling` (Yes/No)
  - `PaymentMethod` (e.g., Electronic check, Mailed check, etc.)
  - `MonthlyCharges`
  - `TotalCharges`
- Services Signed Up:
  - `PhoneService` (Yes/No)
  - `MultipleLines` (Yes/No/No phone service)
  - `InternetService` (DSL, Fiber optic, No)
  - `OnlineSecurity`, `OnlineBackup`, `DeviceProtection`, `TechSupport`, `StreamingTV`, `StreamingMovies` (all Yes/No/No internet service)
- Churn Outcome:

- `Churn` (target: Yes = churned, No = stayed)

Notes on Data Preparation:

- Missing Values: A few missing or problematic entries were found in `TotalCharges`, likely due to customers with very short tenure. These were handled appropriately during preprocessing.
- Categorical Features: Several features are categorical and required encoding for machine learning models.
- Numeric Features: `tenure`, `MonthlyCharges`, and `TotalCharges` are continuous variables that provide important insights into customer behavior.

4. Exploratory Data Analysis (EDA) & SQL Insights

Methodology

The exploratory data analysis was performed using SQL queries executed in Google Colab via Pandas' SQL extension.

Data aggregation and segmentation were conducted through SQL, and Python libraries (Matplotlib and Seaborn) were used for creating visualisations to support the findings.

1. What is the overall churn rate in our customer base?

Findings:

- Total Customers: 7,044
- Churned Customers: ~1,869
- Overall Churn Rate: 26.5%

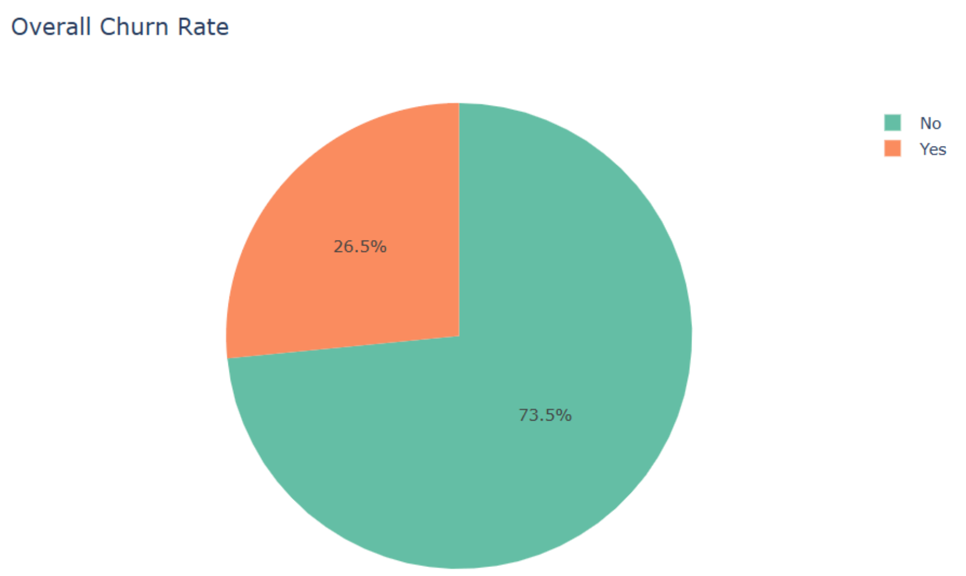

Figure 1: Pie Chart showing Churn vs. No Churn distribution

Insight:

A churn rate of 26.5% indicates a serious churn issue — more than one-fourth of the customer base is leaving.

2. Which customer demographics are most at risk of churning?

SQL Queries and Segments Analysed:

Segmented churn rates by gender, Senior Citizen status, Partner status, and Dependents.

Findings:

- Senior Citizens:
  - Churn Rate: 41%
  - Non-Senior Citizens: 24%
- Partner Status:
  - Customers without a partner churned at 32%, compared to 17% for those with partners.
- Dependents:
  - Customers without dependents churned at 31%, compared to 16% with dependents.
- Gender:
  - Churn rate was relatively similar between males and females (~26%), indicating gender is not a major churn driver.

Insight:

Senior Citizens, customers without partners, and customers without dependents are at higher risk of leaving.

<u>Customer Churn Analysis Report</u>

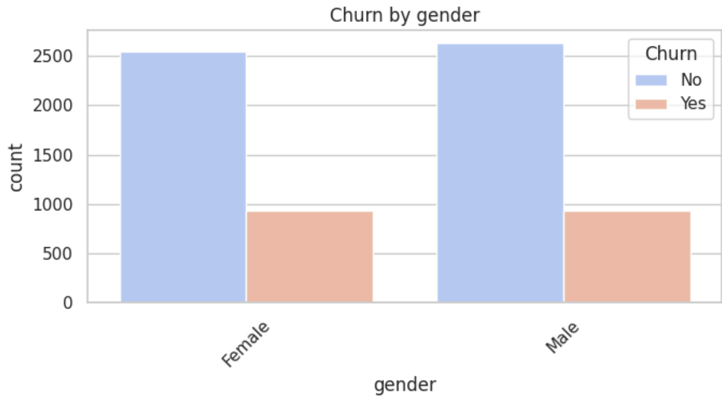

Figure 2: Grouped bar charts for churn rate by gender

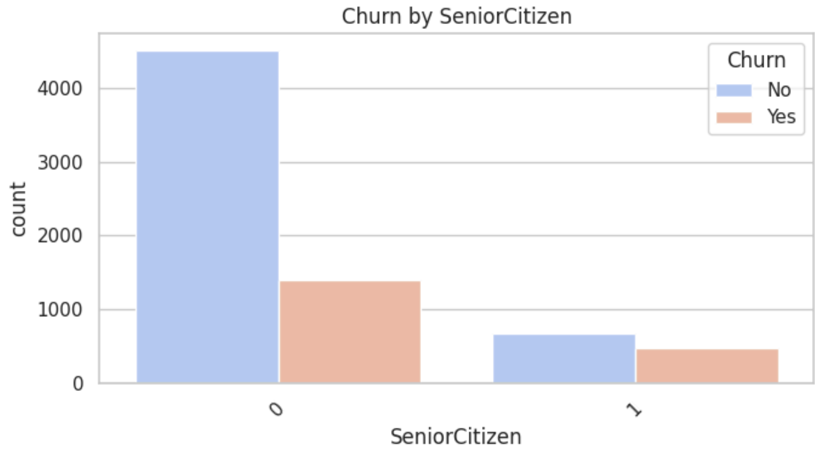

Figure 3: Grouped bar charts for churn rate by senior citizens

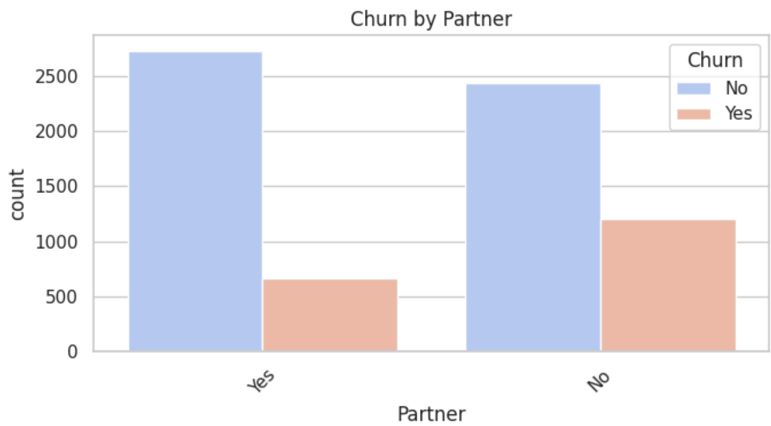

Figure 4: Grouped bar charts for churn rate by partner status

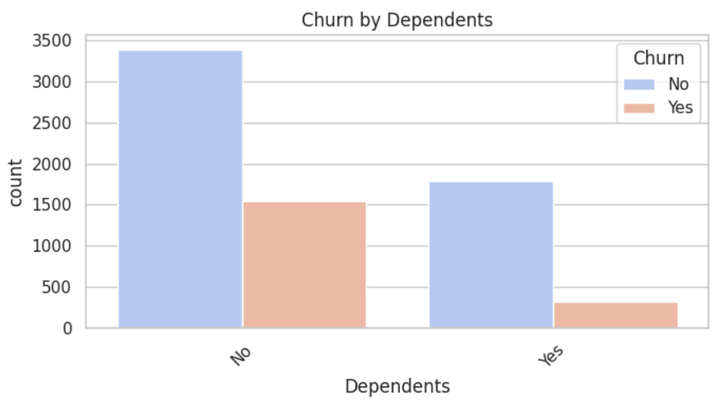

Figure 5: Grouped bar charts for churn rate by departments

3. Are there specific service features that correlate with higher churn?

SQL Queries Analysed:

Segmented churn rates by InternetService and Contract Type

Key Findings:

- Month-to-month contracts: Highest churn rate (~43%).
- Fiber optic Internet service users: Higher churn compared to DSL users.

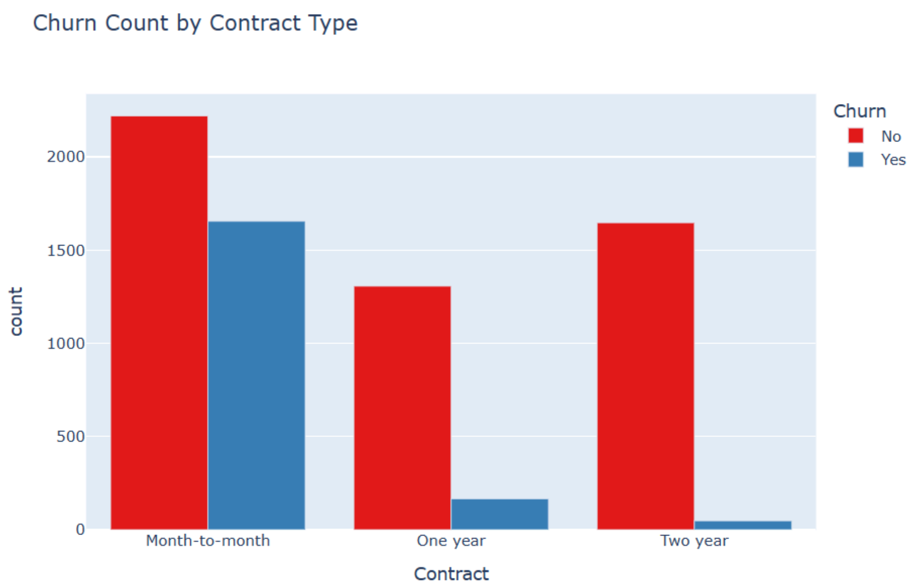

Figure 6: Bar chart showing churn rate by contract type

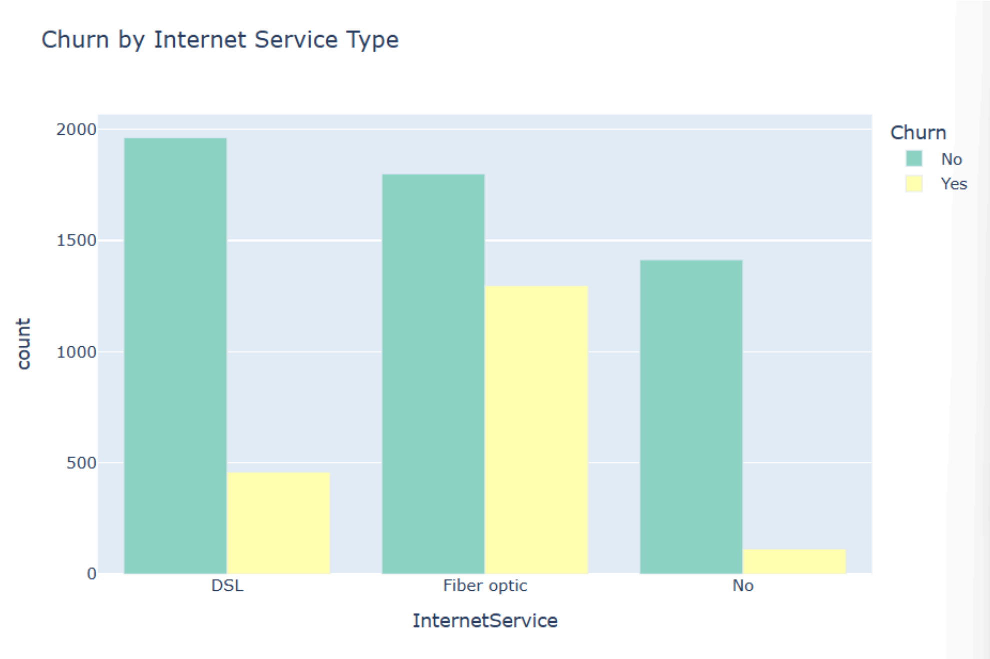

Figure 7: Bar chart showing churn rate by service type

Insight:
Customers with Fiber internet, month-to-month contracts, and electronic check payments are more likely to churn — suggesting operational levers for intervention.

4. How do pricing factors like Monthly Charges impact churn?

SQL Queries:

Compared average MonthlyCharges between churned and loyal customers.

Findings:

- Average Monthly Charges:
  - Churned customers: $74.44
  - Loyal customers: $61.27

Insight:

Churned customers are paying significantly higher monthly charges on average, suggesting that pricing sensitivity could be driving churn.

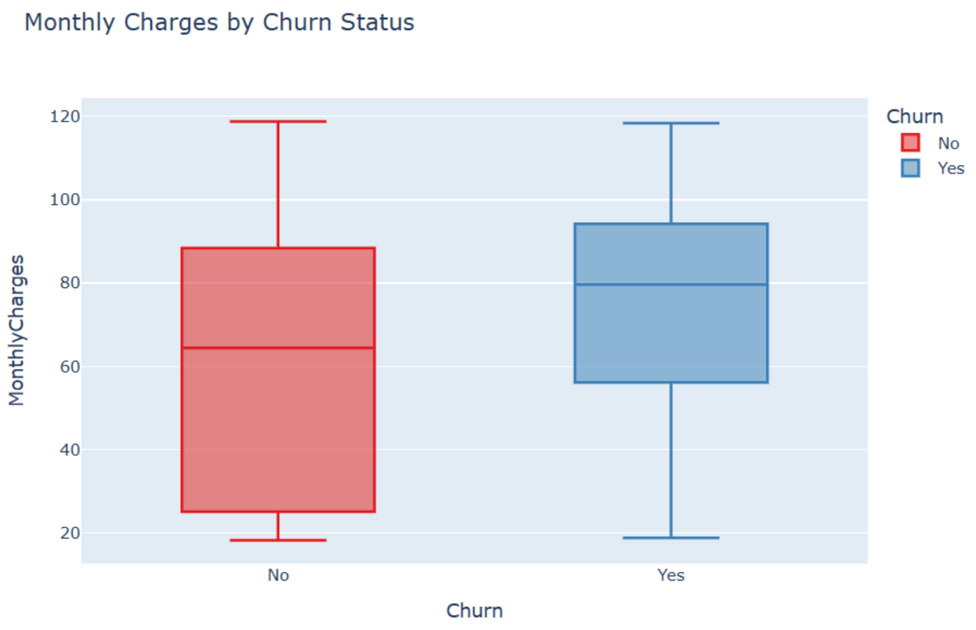

Figure 9: Boxplot of monthly charges by churn status

5. How does customer tenure relate to churn?

SQL Queries:
Grouped customers into tenure buckets (0–12 months, 13–24 months, etc.) and analysed churn rates within each group.

Findings:

- Customers with tenure <12 months had a churn rate of 45%.
- Customers with tenure >48 months had a churn rate of only 11%.

Insight:

Newer customers are much more likely to leave, highlighting the importance of improving early retention strategies and customer onboarding.

Customer Churn Analysis Report

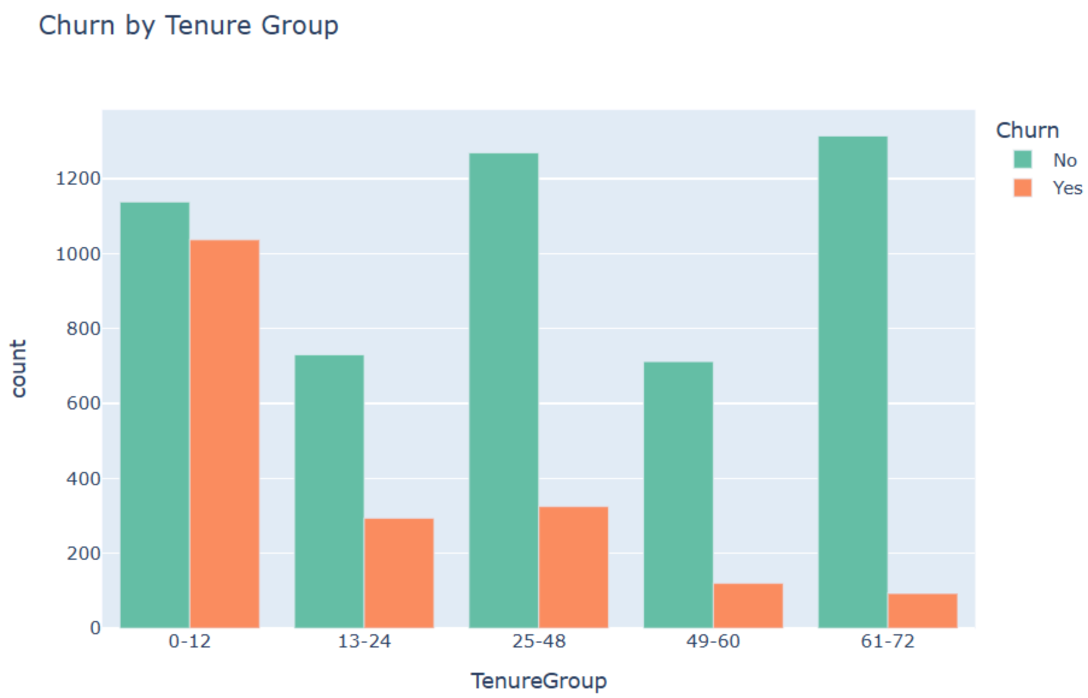

Figure 10: Grouped bar chart of churn rate by tenure bucket

(Tenure grouped into: `0-1 year`*,* `1-2 years`*,* `2-4 years`*,* `4+ years`*)*

Key Takeaways:

- Churn is concentrated among Senior Citizens, customers without dependents/partners, and those with month-to-month contracts.
- High monthly charges and Fiber Internet service are risk factors for churn.
- New customers (especially within their first 12 months) are most vulnerable to leaving.
- Operational interventions (e.g., incentives for longer contracts, auto-payment discounts, tailored support for new customers) could significantly reduce churn.

These insights lay the groundwork for building a predictive model to proactively identify and retain at-risk customers.

5. Predictive Modelling

To better understand the drivers of customer churn and predict customers at risk of leaving, I built and evaluated three supervised machine learning models: **Logistic Regression**, **Random Forest**, and **XGBoost**.

5.1 Model Performance

The models were evaluated using precision, recall, F1-score, and overall accuracy on the test dataset.
Here is a summary of the results:

Model	Accuracy	Precision (Churn)	Recall (Churn)	F1-Score (Churn)
Logistic Regression	79%	62%	52%	56%
Random Forest	79%	63%	48%	54%
XGBoost	76%	57%	47%	51%

- Logistic Regression and Random Forest both achieved an overall accuracy of 79%.
- XGBoost performed slightly worse with 76% accuracy.
- All models struggled more with correctly identifying churned customers (lower recall for the churn class), which is typical in imbalanced churn datasets.

Between the models, Logistic Regression and Random Forest provided similar performance, with Logistic Regression having slightly better recall on churned customers.

5.2 Feature Importance Analysis

Understanding the features that contribute most to churn predictions is crucial for actionable business insights.

Using the Random Forest model, the top 10 most important features were identified:

Feature	Importance Score
TotalCharges	0.195
MonthlyCharges	0.135
tenure	0.125
InternetService_Fiber optic	0.050
PaymentMethod_Electronic check	0.040
OnlineSecurity_Yes	0.035
Contract_Two year	0.030
gender_Male	0.025
TechSupport_Yes	0.025
PaperlessBilling_Yes	0.020

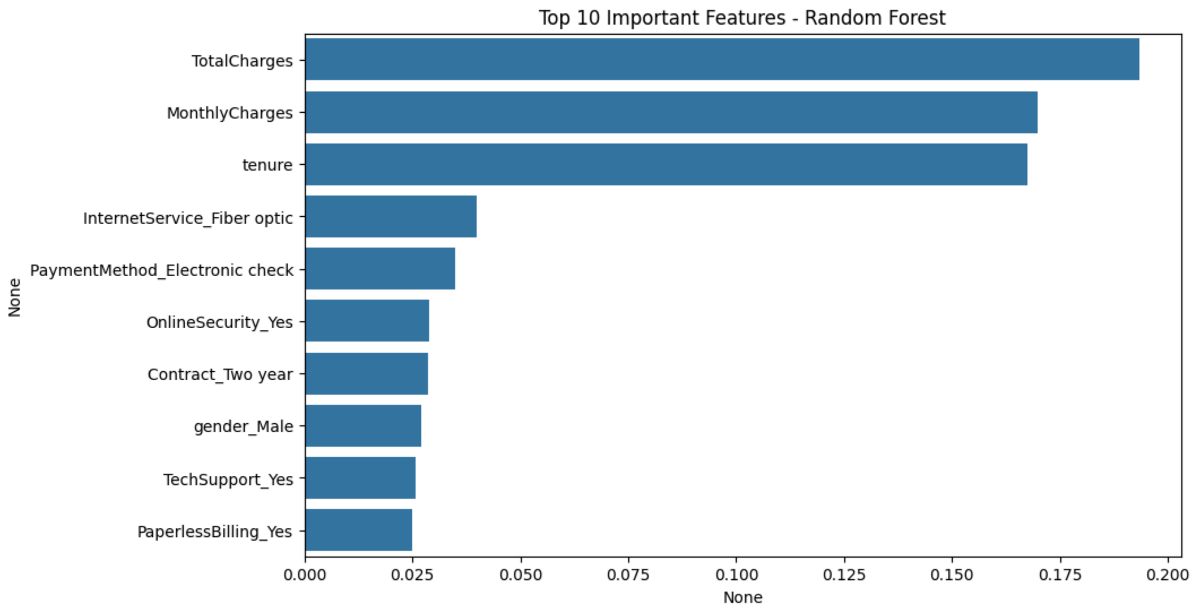

Figure 11: horizontal bar chart was plotted to visualise the feature importances

Key Observations:

- Financial variables such as TotalCharges and MonthlyCharges were the strongest predictors of churn.
- Tenure (the length of time a customer has stayed) was highly influential — shorter tenure typically increases churn risk.
- Customers with Fiber optic internet service, Electronic check payment methods, and Paperless billing were more likely to churn.
- Availability of support services like Online Security and Tech Support reduced churn risk.
- Longer Contract durations (e.g., Two year contracts) were associated with lower churn.

5.3 Interpretation

The predictive modeling phase confirmed that price sensitivity, length of customer relationship, and service quality (including support options) are critical drivers of churn.
By targeting customers with higher monthly charges, shorter tenure, and certain service profiles (e.g., fiber optic users without support services), the company can develop strategies to reduce churn.

6. Key Findings

Executive Summary:
The analysis revealed a churn rate of approximately 27%, driven by factors such as month-to-month contracts, high monthly charges, Fiber optic internet service, and short customer tenure. Demographic factors like senior citizen status and lack of household stability (no partner or dependents) also increased churn risk. Strengthening early customer engagement, offering value-added services, and promoting longer-term contracts are key strategies to reduce churn.

After completing the exploratory data analysis (EDA) and predictive modelling phases, several clear patterns and risk factors for customer churn emerged:

6.1 Main Takeaways

- Churn Rate: Approximately 27% of the customers in the dataset had churned, indicating a significant business challenge that requires focused retention strategies.
- Demographic Risks:
  - Senior citizens were notably more likely to churn compared to non-senior customers.
  - Customers without a partner or without dependents also exhibited higher churn rates, suggesting that household stability may be linked to customer loyalty.
- Service and Contract Factors:
  - Customers with month-to-month contracts had the highest churn rates, reinforcing the importance of locking in customers with longer-term agreements.
  - Customers using Fiber optic internet showed greater churn risk compared to DSL or customers without internet service.
  - The Electronic check payment method was associated with higher churn compared to automatic payments.
- Pricing and Financial Factors:
  - Customers with higher monthly charges were more likely to churn.
  - Lower tenure (newer customers) correlated strongly with higher churn, emphasising that the early stages of the customer lifecycle are particularly critical.

    - Customers with lower total charges also churned more, aligning with the trend of early lifecycle exits.

- Support Services:
    - Availability of services such as Online Security and Tech Support were linked with lower churn rates. Customers who subscribed to these add-ons were more likely to remain loyal.

- Predictive Modeling Insights:
    - Financial metrics (TotalCharges, MonthlyCharges) and tenure were the most influential predictors of churn.
    - Service-related factors, particularly internet type, payment method, and support services, were also strong drivers of churn risk.

6.2 Top Churn Risk Factors

Based on both data exploration and feature importance analysis, the top risk factors contributing to customer churn are:

Risk Factor	Description
Month-to-Month Contracts	Customers on flexible, short-term contracts churned at significantly higher rates.
High Monthly Charges	Customers with elevated monthly bills were more prone to leave.
Fiber Optic Internet Service	Fiber users were more likely to churn compared to DSL or no internet users.
Electronic Check Payment	Customers paying by electronic check churned more than those using credit card or auto-payment methods.
Short Customer Tenure	Newer customers (shorter tenure) were much more likely to churn, suggesting onboarding issues.
Lack of Online Security and Tech Support	Customers without value-added support services were more vulnerable to churn.
Senior Citizen Status	Older customers had a higher likelihood of leaving, indicating the need for tailored retention strategies.

7. Business Recommendations

Based on the insights gained through exploratory analysis and predictive modeling, the following targeted actions are recommended to help reduce customer churn:

7.1 Offer Incentives to High-Risk Segments

- Target customers with month-to-month contracts, high monthly charges, and Fiber optic service with personalised offers, such as loyalty discounts, bundled service packages, or rewards for contract extensions.

- Senior citizens and customers without partners or dependents could be offered tailored loyalty programs to build stronger relationships and brand affinity.

7.2 Strengthen Early Customer Engagement

- New customers with low tenure are at significantly higher risk of early churn.

- Implement a robust onboarding program in the first 3–6 months, such as:

    - Proactive customer support check-ins.

    - Educational content on service benefits and troubleshooting.

    - Welcome offers or first-year loyalty rewards.

7.3 Promote Longer-Term Contracts and Service Quality

- Month-to-month customers should be encouraged to transition to annual or two-year contracts through incentives (e.g., discounted rates, added features like premium support).

- Focus on improving service reliability and support quality, particularly for Fiber optic customers, who show higher churn risk when support services (e.g., Tech Support, Online Security) are missing.

- Promote the adoption of auto-pay methods to reduce churn associated with electronic check users.

7.4 Potential Impact of Churn Reduction

- With an estimated churn rate of ~27% across 7,044 customers, approximately 1,900 customers are lost annually.

- Reducing churn by even 5 percentage points (from 27% to 22%) could retain approximately 350 customers per year.

- Assuming an average TotalCharges value of ~$2,000 per customer (based on analysis), this translates into potential revenue retention of ~$700,000 annually.

8. Limitations & Future Work

8.1 Limitations

- Limited Feature Scope:
  The dataset primarily includes customer demographics, billing information, and service details. It lacks richer behavioral data such as customer support interactions, service usage patterns, or customer satisfaction scores, which are often critical drivers of churn.

- Static Snapshot:
  The analysis is based on a single snapshot of customer data, rather than tracking behavior over time. Without a time-series perspective, it is difficult to capture dynamic changes leading up to churn.

- Class Imbalance:
  Although not extreme, the dataset has an inherent churn imbalance (~27% churners vs. 73% non-churners), which made it challenging for models to achieve high recall when predicting churn.

- Assumptions in Feature Engineering:
  Some transformations (e.g., tenure bucketing) and one-hot encodings simplify complex relationships, but may overlook more nuanced interactions between variables.

8.2 Future Work

- Integrate Customer Satisfaction and Support Data:
  Incorporating metrics like Net Promoter Score (NPS), customer complaint rates, and support ticket frequency could significantly improve churn prediction and customer segmentation.

- Adopt Time-Series Modeling:
  Tracking customer behavior longitudinally (e.g., monthly usage trends, billing changes, support events) would allow for the development of early warning churn models and better capture churn triggers.

- Advanced Machine Learning Techniques:
  Explore ensemble methods (e.g., stacking models) or deep learning approaches to potentially boost model performance, particularly for minority churn classes.

- Segmented Churn Strategies:
  Build separate models for high-value vs. low-value customers to tailor retention strategies more effectively based on customer lifetime value (CLV).

- A/B Testing of Retention Strategies:
  Following model deployment, test the effectiveness of interventions (like loyalty rewards or improved onboarding) through controlled experiments to quantify real-world impact on churn reduction.