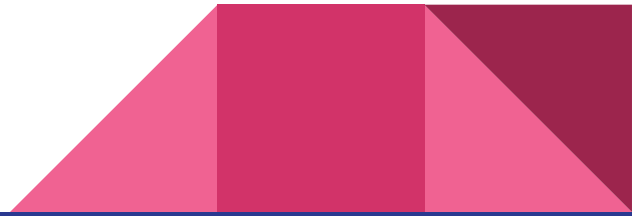
Telco Customer Churn Analysis

By Dhwani Zala

Outline

- Introduction
- Data Sources
- Documentation Of Cleaning & Manipulation
- Summary Of Data Analysis
- Key Visualization & FIndings
- Preprocessing & Feature Engineering
- Modeling & Predictions
- Recommendations

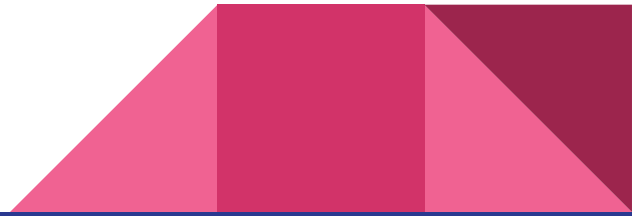


Introduction

This Exploratory Data Analysis (EDA) focuses on understanding customer churn in the telecom industry. The objective is to identify key factors influencing customer attrition and provide insights to help improve retention strategies.

The dataset includes various customer attributes such as tenure, monthly charges, contract type, and service usage. Through this analysis, we aim to uncover patterns and trends that differentiate churned and non-churned customers.

By leveraging data-driven insights, businesses can make informed decisions to reduce churn, enhance customer satisfaction, and optimize retention strategies.



Data Sources & Overview

Data Sources

The dataset used in this analysis comes from a telecom company’s customer records. It includes information about customer demographics, service subscriptions, billing details, and churn status. The data is structured and extracted from the company’s customer management system.

Data Sources & Overview

Overview

- Total Records: 7,043 customers
- Total Features: 33 columns

Key Categories:

- Customer Demographics: Gender, Senior Citizen status, Partner, and Dependents
- Service Subscriptions: Internet service, phone service, online security, and other add-ons
- Billing Information: Monthly charges, total charges, and contract type
- Churn Indicator: Whether the customer has left the service (Yes/No)

Documentation Of Cleaning & Manipulation

Used tools: Python, Pandas, seabron, matplotlib, statsmodels libraries for analysis

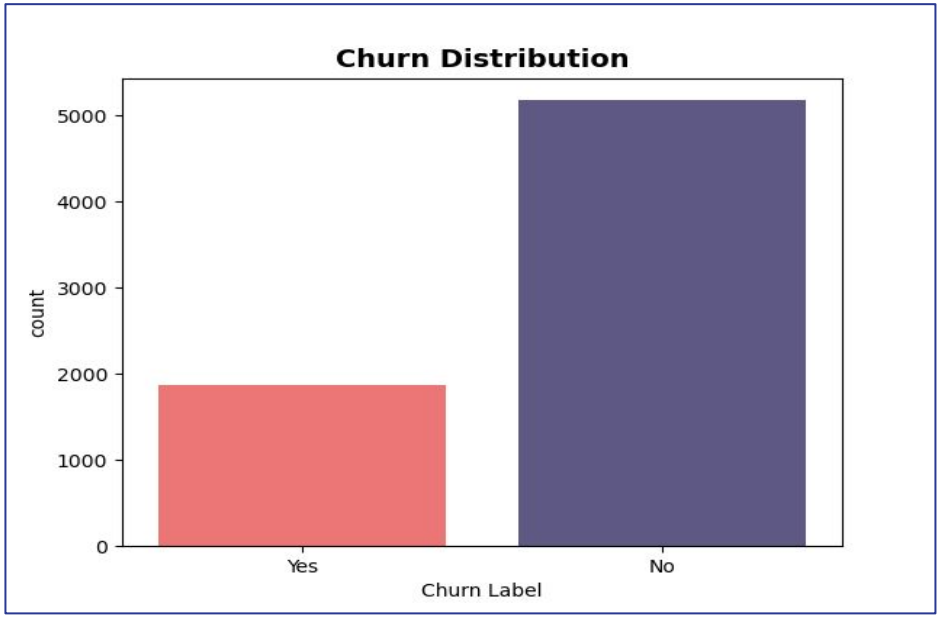
- Handled Missing Values → Filled missing "TotalCharges" with median & missing "Churn Reason" with "Not Mentioned Reason".
- Data Type Corrections → Converted "SeniorCitizen" to categorical & "TotalCharges" to numeric.

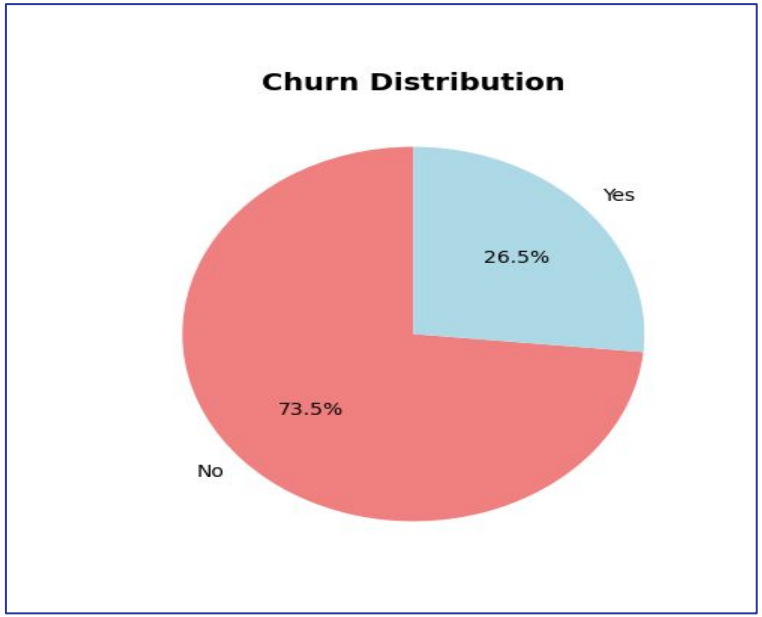
Summary Of Data Analysis

- Churn Rate → 📉 26.5% of customers have churned, while 73.5% remain.
- Tenure Impact → 🕒 Shorter tenure customers are more likely to churn.
- Monthly Charges → 💰 Higher charges lead to higher churn rates.
- Contract Type → 📜 Month-to-month contracts have the highest churn, while long-term contracts improve retention.
- Service Impact → Customers without tech support & security services churn more.
- Demographics → Senior citizens & customers without dependents show higher churn rates.

Key Visualization & FIndings

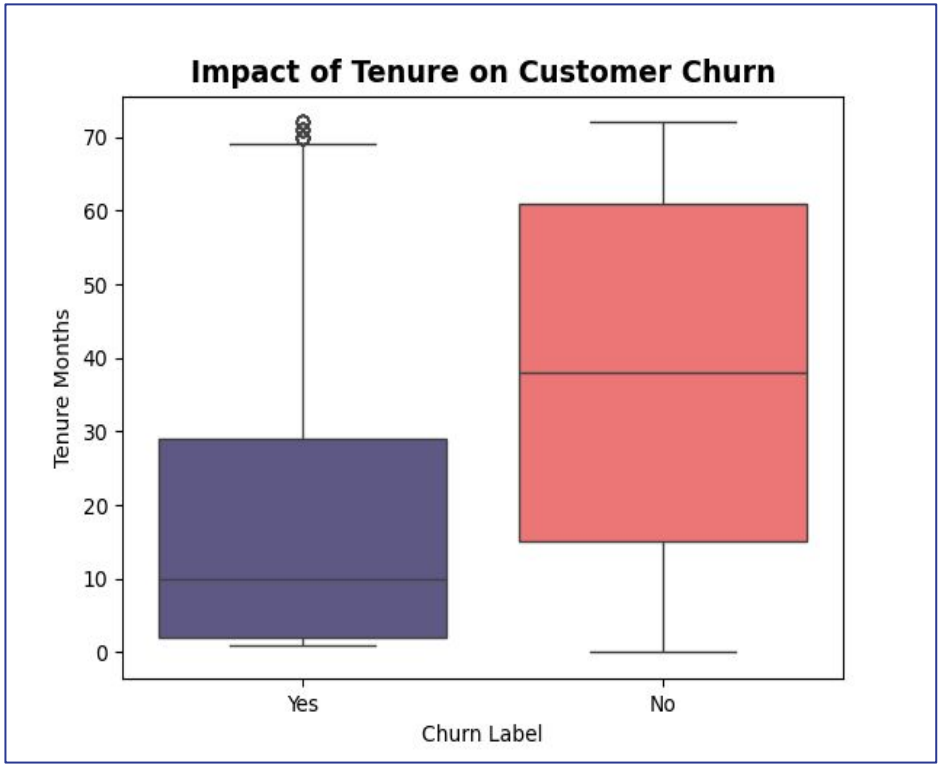
❑ Churn Distribution 📉





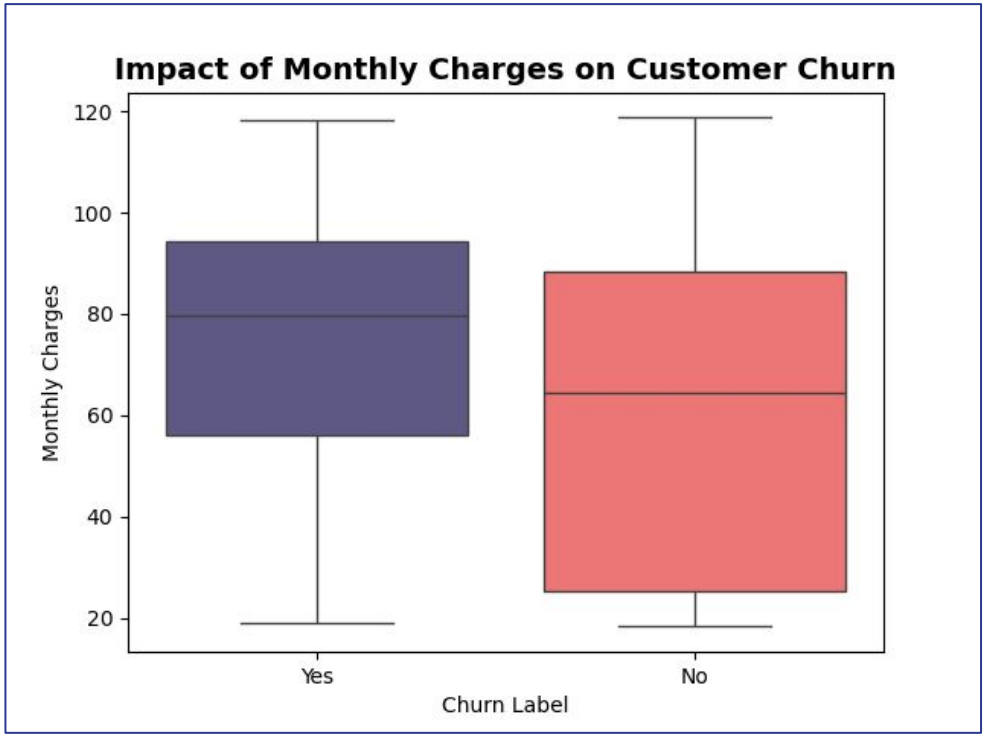
- 26.5% of customers churned, highlighting a retention challenge.
- **Finding**: A significant portion of customers leave, requiring better engagement strategies.

❑ Tenure vs. Churn 🕑



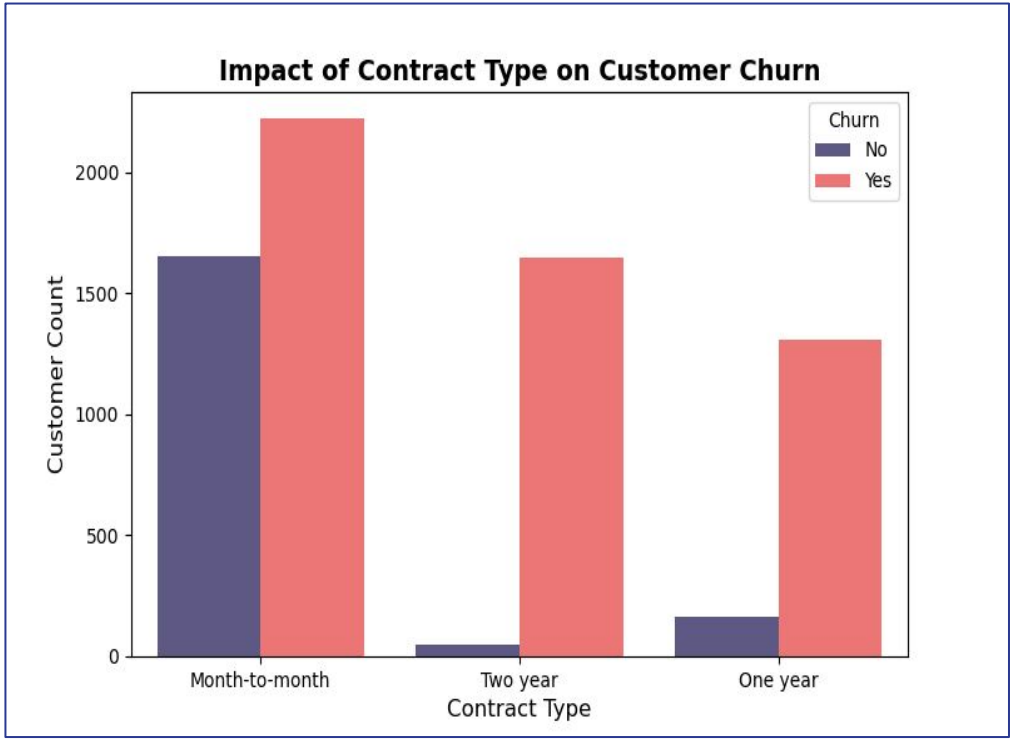
- Churned Customers (Yes) → Have a lower median tenure (mostly below 10 months).
- Non-Churned Customers (No) → Have a higher median tenure (around 40 months).
- There are outliers in churned customers, meaning a few long-tenure customers also churned.
- **Customers with shorter tenure are more likely to churn.**
- **Retaining customers in the first few months is critical.**

❑ Monthly Charges & Churn 💰



- Churned Customers (Yes) → Have a higher median monthly charge (~rs.80).
- Non-Churned Customers (No) → Have a lower median monthly charge (~rs.50).
- The overall distribution suggests that customers paying higher charges are more likely to churn.
- **Higher monthly charges correlate with higher churn rates.**
- **Customers with lower charges are more likely to stay.**

❑ Contract Type Impact 📜



- Month-to-month contracts have the highest churn, while long-term contracts retain customers better.

❑ Service Impact on Churn 🛠️

1.Phone Service:

- Most customers have phone service.
- Churn is slightly higher among customers with phone service than those without.

2. Multiple Lines:

- Customers with multiple lines tend to churn more than those with a single line.
- "No phone service" category has a low churn count, likely due to fewer customers in that category.

3. Internet Service:

- Fiber optic users churn more compared to DSL or those without internet service.
- This suggests fiber-optic users might face pricing or service-related issues.

4.Online Security & Online Backup:

- Customers without security or backup services churn at a higher rate.
- Providing these services might improve retention.

5. Device Protection:

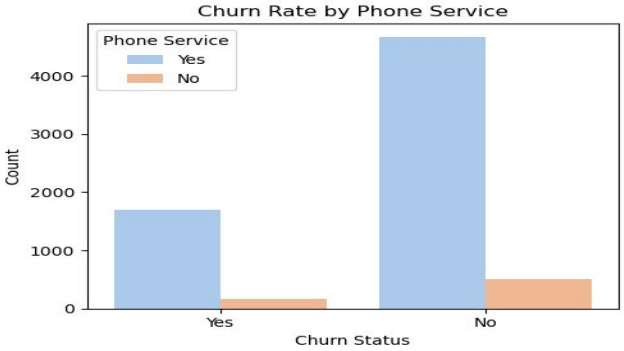
- No device protection → Higher churn.
- Customers might feel less secure or face additional costs when devices fail.

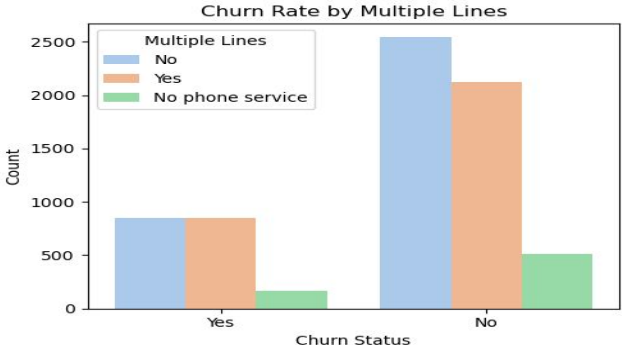
6. Tech Support:

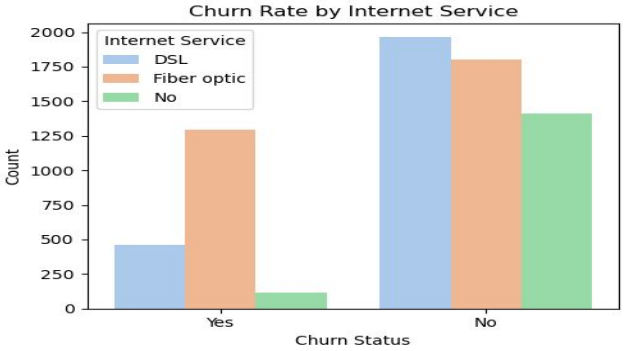
- No tech support leads to higher churn.
- Customers without support might struggle with service issues and leave.

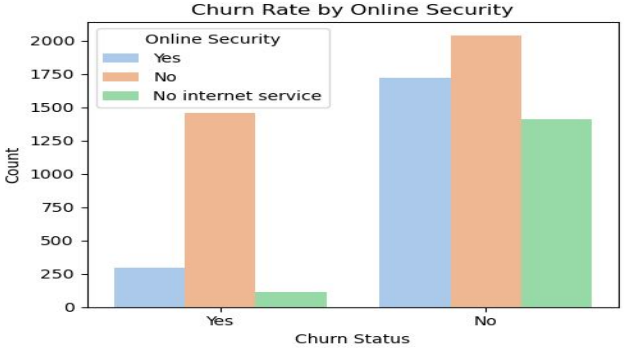
7. Streaming TV & Streaming Movies:

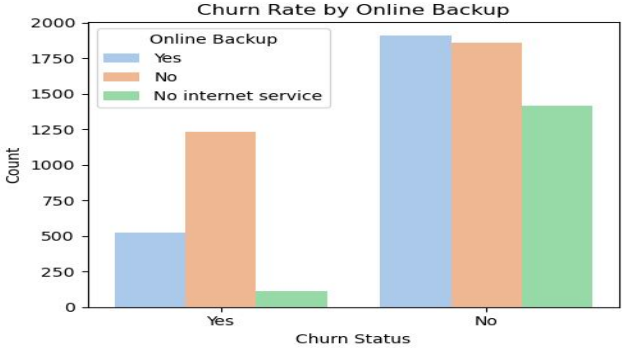
- Customers without streaming services churn more.
- These services could act as incentives for retention.

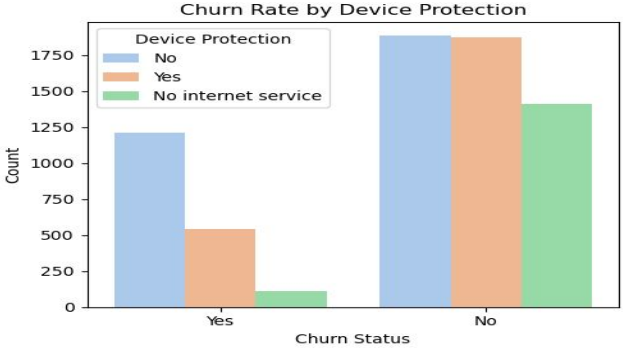

Churn Rate by Phone Service
Phone Service
Yes
No
Count
Churn Status
Yes
No

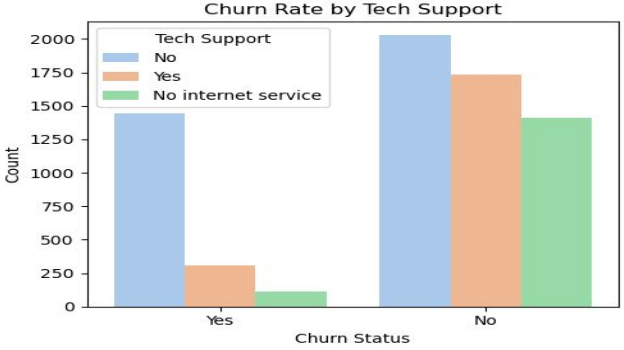

Churn Rate by Multiple Lines
Multiple Lines
No
Yes
No phone service
Count
Churn Status
Yes
No

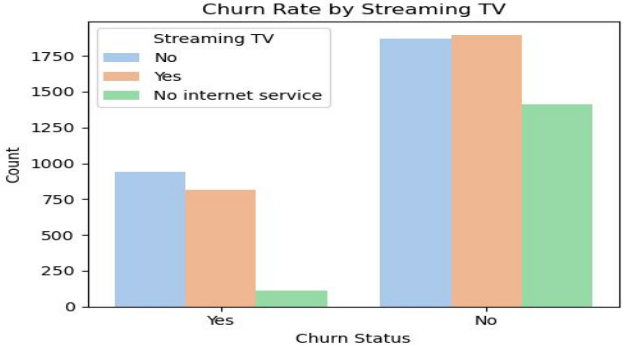

Churn Rate by Internet Service
Internet Service
DSL
Fiber optic
No
Count
Churn Status
Yes
No

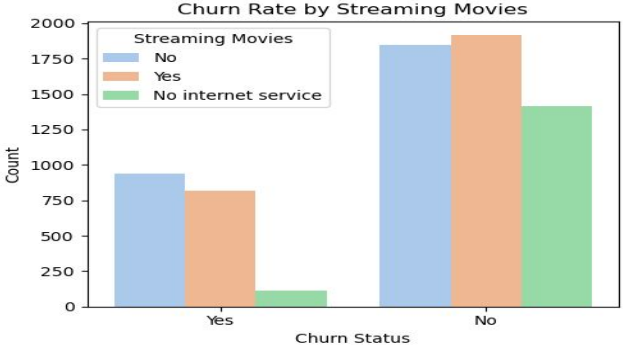

Churn Rate by Online Security
Online Security
Yes
No
No internet service
Count
Churn Status
Yes
No


Churn Rate by Online Backup
Online Backup
Yes
No
No internet service
Count
Churn Status
Yes
No

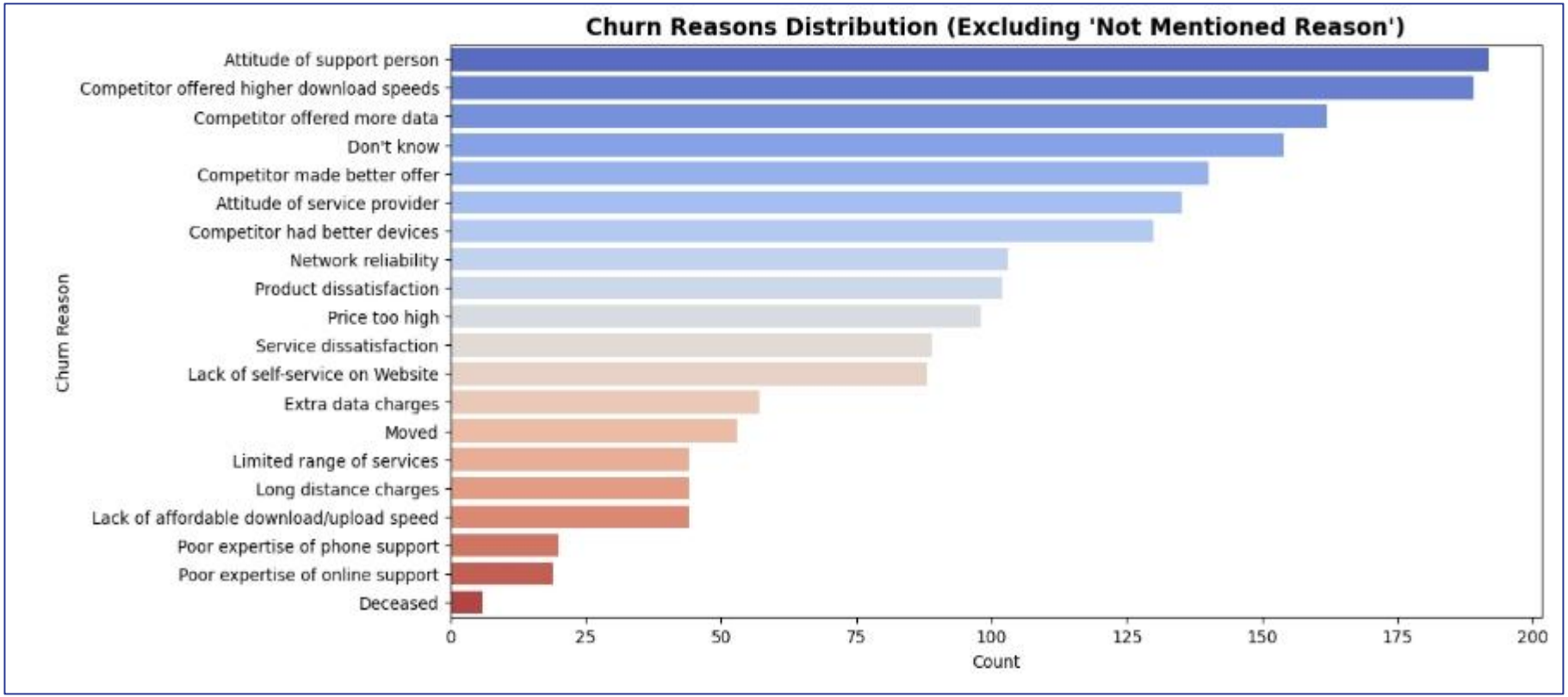

Churn Rate by Device Protection
Device Protection
No
Yes
No internet service
Count
Churn Status
Yes
No


Churn Rate by Tech Support
Tech Support
No
Yes
No internet service
Count
Churn Status
Yes
No


Churn Rate by Streaming TV
Streaming TV
No
Yes
No internet service
Count
Churn Status
Yes
No


Churn Rate by Streaming Movies
Streaming Movies
No
Yes
No internet service
Count
Churn Status
Yes
No

❏ 🔍 Uncovering Churn Reasons



❑ Top 3 Reasons for Churn:

- Attitude of support person: Poor customer support is a leading cause of churn.
- Slower download speeds: Speed issues drive dissatisfaction, likely affecting internet users.
- Competitor offered more data: Customers leave for better deals, indicating a pricing or package issue.

Preprocessing & Feature Engineering

Objectives

Prepare data for model training by cleaning, encoding, and scaling.

Dropped Irrelevant or Risky Columns:

- Removed columns like CustomerID, Lat Long, and Senior Citizen (null or redundant).

Excluded post-churn or derived columns such as:

- Churn Label, Churn Value, Churn Score, CLTV, and Churn Reason — to prevent data leakage.
- Target Variable: Encoded Churn Label as binary (0 = No churn, 1 = Yes churn).
- Train-Test Split: Stratified split (80/20) for robust model evaluation.

Preprocessing Pipelines

Modular Pipelines for Numerical & Categorical Features:

Numerical Features:

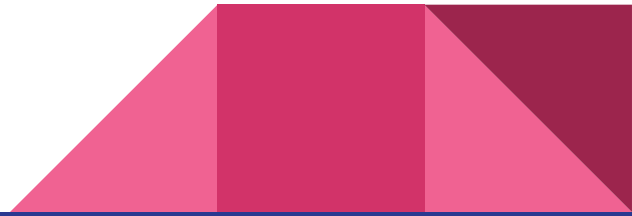
- Missing values handled with mean imputation.
- Features standardized using StandardScaler.

Categorical Features:

- Missing values imputed using most frequent category.
- Encoded using One-Hot Encoding with handle_unknown='ignore' for robustness.
- Combined using ColumnTransformer to apply specific transformations to corresponding feature types.

Benefits:

- Ensures consistency, reproducibility, and clean separation of preprocessing logic.
- Scalable approach for real-world data pipelines and deployment.



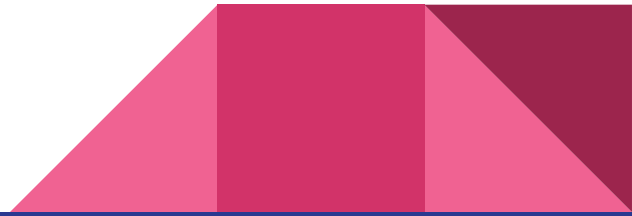
Modeling & Predictions

We compared two models:

- Logistic Regression (simple)
- Random Forest (advanced)

Why Random Forest is better:

- It understands complex patterns in the data.
- More accurate predictions with fewer mistakes.
- Great at capturing hidden relationships in customer behavior.



What We Measured	What It Means	Logistic Regression	Random Forest	Better Model
Average Error (MAE)	On average, how far off the model's predictions are	0.2982	0.1959	Random Forest
Prediction Mistakes (MSE)	Big mistakes count more — we want this to be lower	0.1594	0.0623	Random Forest
Overall Fit (R^2 Score)	How well the model explains why customers leave (0 to 1)	0.2158	0.6208	Random Forest

- Random Forest predicts customer churn more accurately. It makes fewer mistakes and does a much better job of understanding customer behavior patterns.

Churn Prediction Demo

Objective

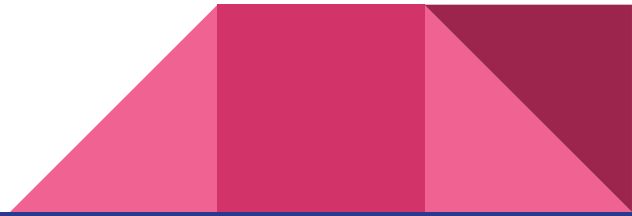
Show how the trained model can be used to predict if a customer is likely to leave the telecom service.

Steps Taken:

- Loaded the trained Random Forest pipeline using pickle.
- Created a sample customer profile with attributes like:
- Location, Gender, Tenure, Internet Service, Monthly Charges, etc.

Used the model to predict:

- Churn Class (1 = Yes, the customer will likely churn)
- Churn Probability (67% confidence in this case)

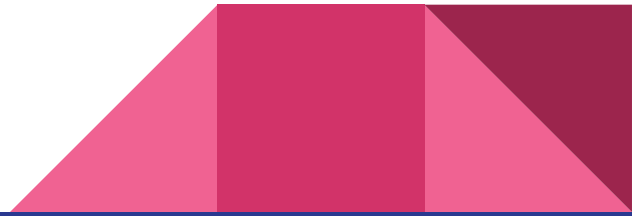


Output Example:

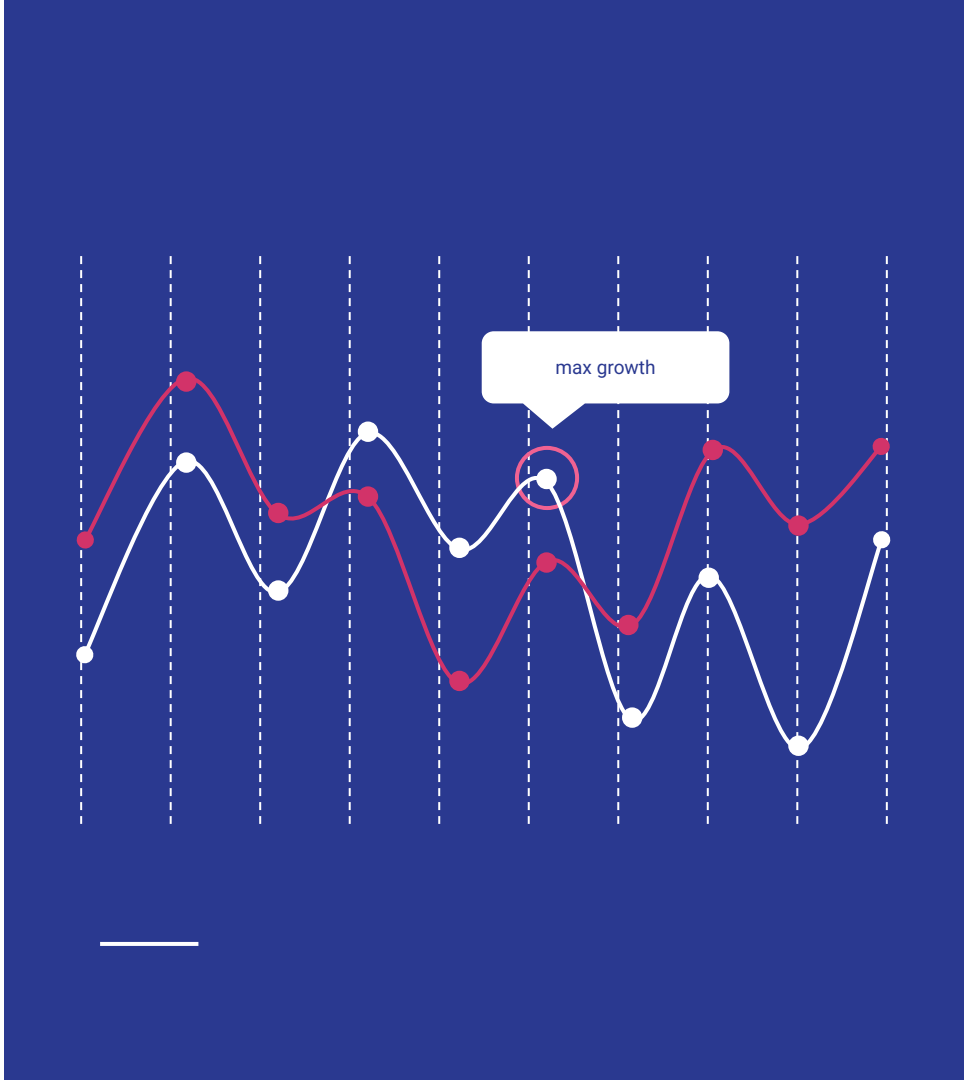
- Prediction: Customer is likely to churn
- Confidence: 67% probability

Business Impact

- This allows telecom companies to:
- Take proactive action (e.g., send offers, support outreach)
- Focus retention efforts on high-risk customers
- Personalize communication based on churn risk



Recommendations



- **Promote Long-Term Contracts:** Offer discounts and perks to encourage annual or biennial plans.
- **Enhance Customer Support:** Train staff, improve response times, and implement chatbots for quick issue resolution.
- **Improve Service Bundling:** Upsell Online Security, Tech Support, and Device Protection with attractive packages.
- **Boost Network Reliability:** Invest in infrastructure, optimize speed, and provide proactive outage alerts.
- **Boost Network Reliability & Speed:** Slow speeds and network issues drive churn—invest in infrastructure, optimize routing, and provide proactive outage alerts.
- **Competitor Awareness & Retention Offers:** Many customers leave for better deals—track competitor pricing and offer targeted retention discounts.

- **Strengthen Fiber Optic Services:** High churn in fiber customers—focus on resolving service issues and improving perceived value.
- **Leverage Streaming & Entertainment Perks:** Customers with Streaming TV & Movies have lower churn—offer exclusive content deals or bundle with premium plans.
- **Simplify Online Self-Service:** Improve website self-service options for bill payments, upgrades, and troubleshooting to reduce dissatisfaction.
- **Optimize Pricing & Extra Charges:** High churn due to extra fees—introduce transparent pricing and loyalty rewards for long-term customers.
- **Address Specific Churn Reasons:**
  - "Better devices elsewhere" → Partner with leading brands for device upgrade deals.
  - "Moved" → Offer relocation benefits or flexible service transfers.
  - "Lack of online support" → Enhance digital support channels (live chat, self-help articles).

Thank You!