Customer Churn Analysis & Key Insights

Introduction: This project aims to analyze customer churn patterns, identify key factors influencing churn, and provide actionable insights to reduce churn rates.

Objectives

The main goal of this project is to predict customer churn using Machine Learning and identify the key factors that contribute to customer attrition.

Key Questions We Aim to Answer:

1.What are the most common characteristics of customers who churn?

2.How do factors like tenure, monthly charges, customer service interactions, and subscription plans affect churn?

3.Can we build an accurate machine learning model to predict which customers are likely to churn?

4.What strategies can businesses use to reduce churn and improve retention?

Data understanding.

- The dataset used for this analysis was sourced from SyriaTel (kaggle). It contains customer records with various attributes related to usage, service plans, and churn status.
- Key features of this data set include :
- **CustomerID** – Unique identifier for each customer.
- **Tenure** – How long the customer has been with the company.
- **Monthly Charges** – The amount paid by the customer each month.
- **Total Charges** – Total amount spent by the customer.
- **International Plan** – Whether the customer has an international call plan (Yes/No).
- **Service Calls** – The number of customer service calls made.
- **Churn** – The target variable (Yes = Churned, No = Retained)

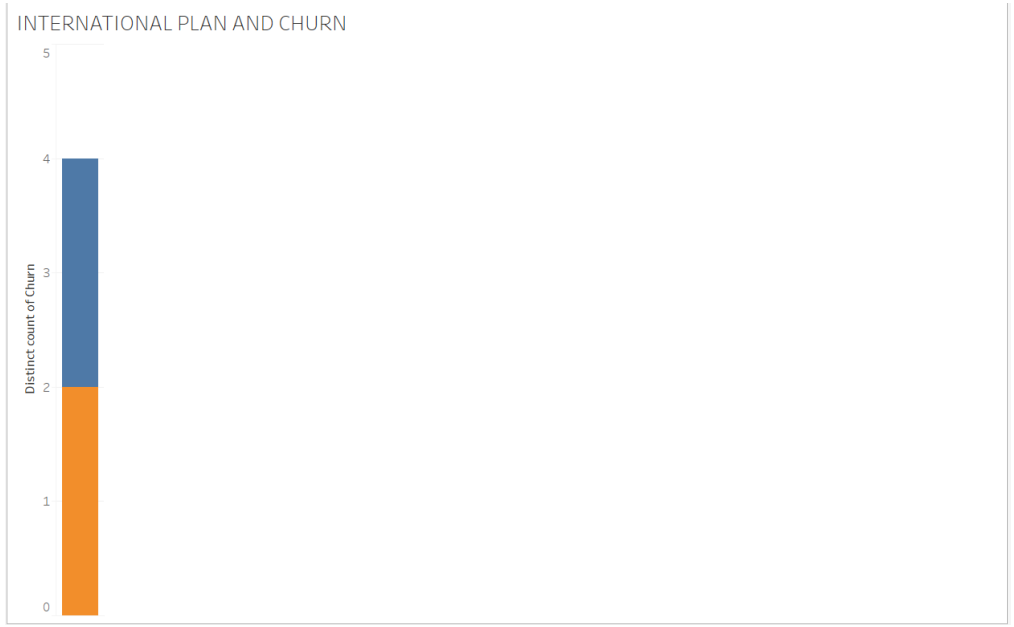
Exploratory Data Analysis (EDA)

Univariate Analysis (Single Variable)

- From the chart, we observe that:

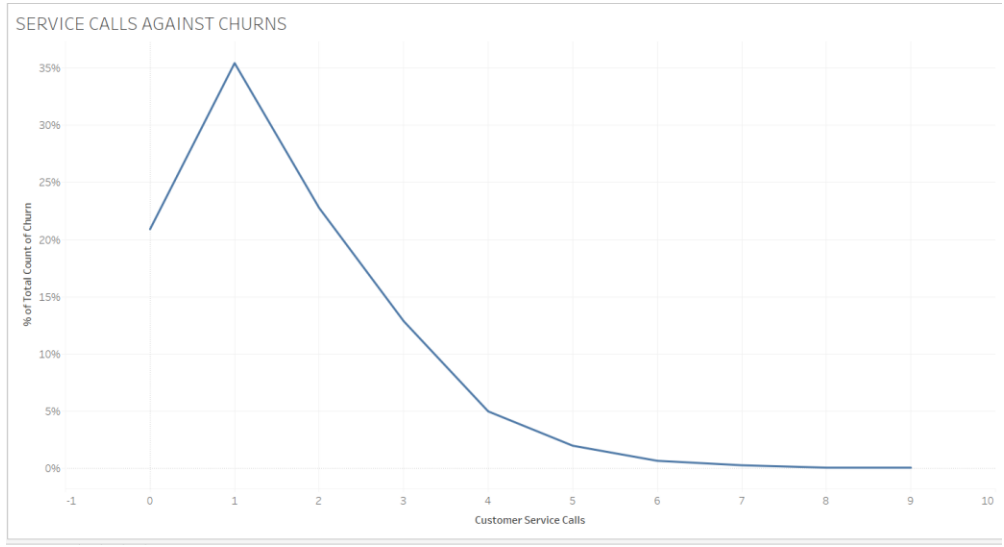
A significant number of customers with an International Plan have churned.

This suggests that customers who subscribe to the International Plan may be more likely to leave the service compared to those without it.



Bivariate Analysis (Comparing Two Variables)

- This line chart shows the relationship between customer service calls and the percentage of total churn.
- Customers who only make one service call may have had a negative experience, leading them to leave the service.
- Customers who make multiple service calls may have their issues resolved, increasing satisfaction and reducing churn.
- A low number of service calls (0 or 1) could indicate dissatisfaction, as customers do not bother reaching out before leaving.



Modeling Approach

We used a Logistic Regression model because it is effective for binary classification problems like churn prediction and other additional models was the decision tree.

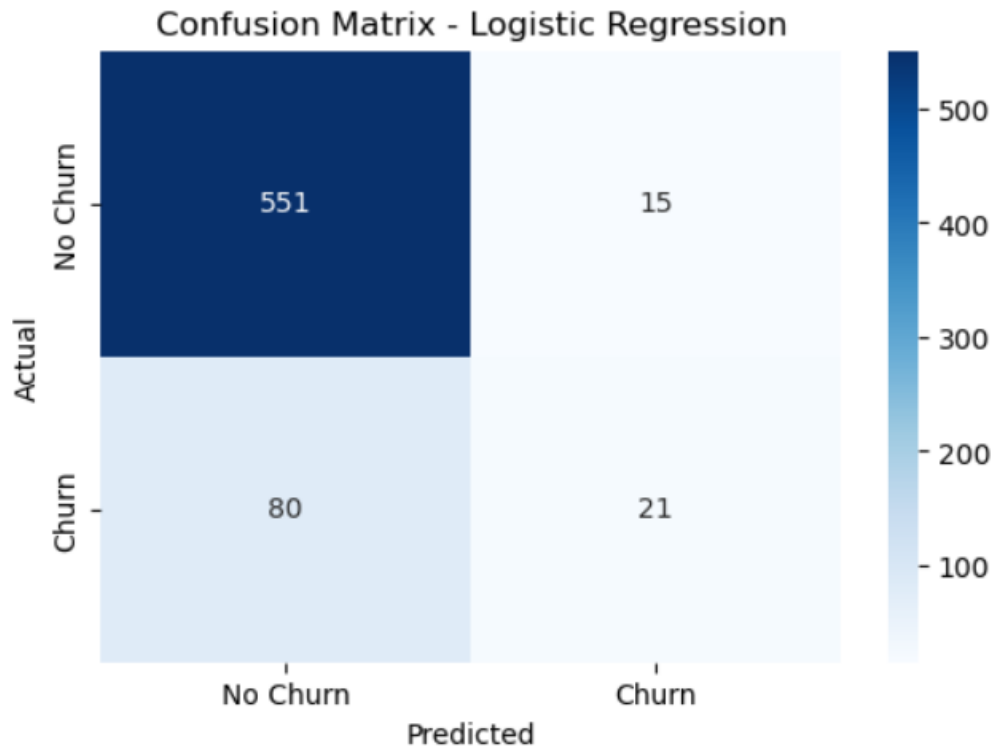
We first tackled the missing values using appropriate techniques then we encoded categorical variables.

The dataset is then divided into training (80%) and testing (20%) sets to evaluate model performance properly.

```
Logistic Regression Accuracy: 0.8575712143928036
              precision    recall  f1-score   support

           0       0.87      0.97      0.92       566
           1       0.58      0.21      0.31       101

    accuracy                           0.86       667
   macro avg       0.73      0.59      0.61       667
weighted avg       0.83      0.86      0.83       667
```

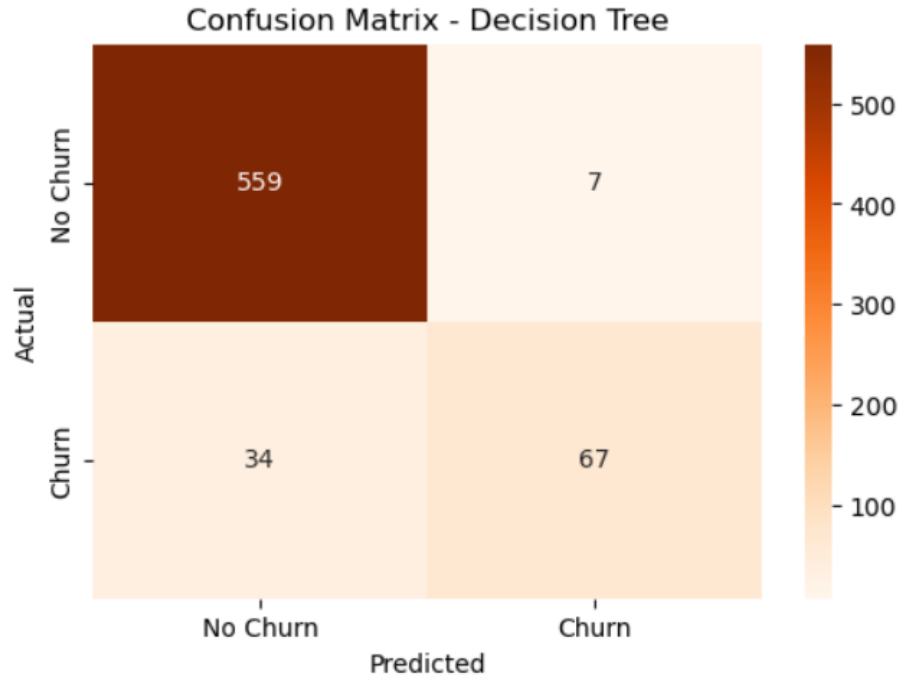


- This image shows the evaluation metrics and confusion matrix for a Logistic Regression model used to predict customer churn.
- The model correctly predicts churn/non-churn for 85.77% of customers.
- The model is good at predicting non-churners but struggles to identify actual churners.
- A high recall for No Churn (0.97) but a low recall for Churn (0.21) means the model misses a lot of actual churners.
- The macro average F1-score is 0.61, meaning the model is not balanced across both classes.

- This image presents the evaluation metrics and confusion matrix for a Decision Tree model used to predict customer churn.
- The model performs better than Logistic Regression in identifying churners (recall increased from 21% to 66%).
- The macro average F1-score is 0.87, showing a more balanced performance across both classes.

Decision Tree Accuracy: 0.9385307346326837

	precision	recall	f1-score	support
0	0.94	0.99	0.96	566
1	0.91	0.66	0.77	101
accuracy			0.94	667
macro avg	0.92	0.83	0.87	667
weighted avg	0.94	0.94	0.93	667



Recommendations

- Based on the analyzed data presented here are some recommendations:

1.Improve Customer Support-we have observed that customers with an International Plan have a higher churn rate so we should Improve customer service response time and resolution effectiveness to prevent frustration.

2.Personalized Retention Offers- Decision Tree model identified key churn patterns therefore we should use predictive analytics to offer targeted retention incentives (discounts, personalized plans) for customers predicted to churn.

3.**Focus on High-Risk Customers-**Customers with an International Plan have a higher churn rate therefore we would recommend that we offer discounts or loyalty programs for customers on international plans to retain them.

Conclusion

- In this data set we were able to successfully analyzed customer churn using various techniques, including exploratory data analysis, bivariate analysis, and predictive modeling. This helped in identifying patterns and making data-driven decisions to enhance business strategy.