

WinnyC3 / seriatel-project

<> Code Issues Pull requests Actions Projects Wiki Security Insights Settings

0 stars 0 forks 1 watching 1 Branch 0 Tags Activity

Public repository

1 Branch 0 Tags

Go to file

t

Go to file

+

Add file

Code

WinnyC3 updates26d52ac · 3 minutes ago

Images	updates	15 hours ago
.gitignore	checking data shape	4 days ago
README.md	readme further updates	6 hours ago
bigml_59c28831336c6604c8...	new commit	4 days ago
non-technical presentation.pdf	presentation	16 minutes ago
seriatel.ipynb	updates	3 minutes ago

README

CUSTOMER CHURN PREDICTION OF SYRIATEL COMPANY



Overview

SyriaTel is addressing customer churn in the competitive telecommunication industry by developing machine learning models to predict churn based on customer data such as demographics and usage patterns. The objectives include building and evaluating these models to identify the most accurate predictor, analyzing key features impacting churn rates, and providing actionable recommendations to enhance customer retention. Additionally, the company aims to refine its retention strategies and continuously adapt its predictive models to stay aligned with changing customer behaviors and market conditions, ultimately reducing churn and improving profitability.

Business Understanding

In the competitive telecom industry, predicting customer churn is vital for SyriaTel to maintain its edge. Retaining existing customers is more cost-effective than acquiring new ones, making it crucial to prevent churn and improve retention. By analyzing customer demographics, usage patterns, and other data points, SyriaTel can build predictive models to forecast churn and take proactive retention measures. To stay ahead, SyriaTel must continually update these models to reflect changing customer behaviors and market dynamics, ensuring effective retention strategies and sustained competitive advantage.

Objectives

1. Develop Predictive Models: Create machine learning models to predict customer churn using data on customer features, such as demographics and usage patterns.
2. Evaluate Model Accuracy: Compare the performance of different machine learning models to determine the most accurate one for predicting customer churn.
3. Feature Impact Analysis: Identify key features that significantly impact customer churn rates to understand the factors driving customer retention and attrition.

4. Provide Actionable Recommendations: Offer strategic recommendations based on the analysis to help SyriaTel implement effective retention strategies and reduce churn rates.
5. Optimize Retention Strategies: Utilize the predictive insights to refine and optimize customer retention strategies, ensuring they align with identified churn predictors and changing customer behaviors.

Data Understanding

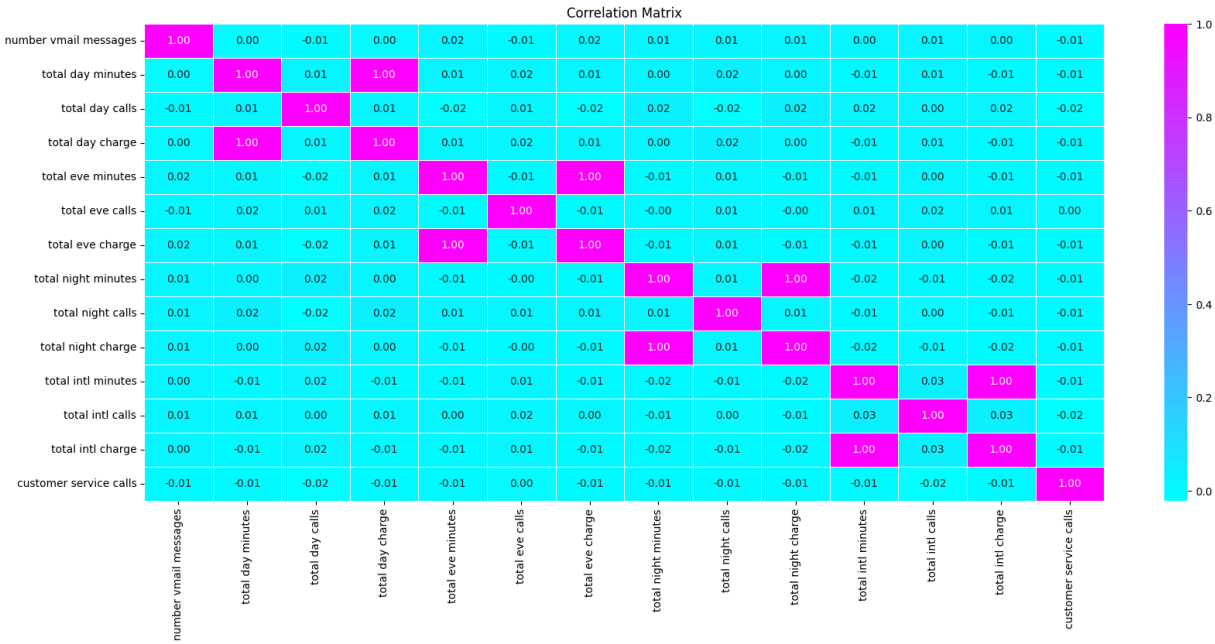
The project utilizes historical customer data, including demographic information and transactional data of Seriatel telecom company. This data is used to build a predictive model that can classify customers as churned or non-churned. The data has 3333 rows and 21 columns and the company is based in California USA. The column titles are as follows:

- State: The state where the customer resides.
- Area code: The area code associated with the customer's phone number.
- International plan: A binary variable indicating whether the customer has an international calling plan (1) or not (0).
- Voice mail plan: A binary variable indicating whether the customer has a voicemail plan (1) or not (0).
- Number vmail messages: The number of voicemail messages the customer has.
- Total day minutes: The total number of minutes the customer used during the daytime.
- Total day calls: The total number of calls the customer made or received during the daytime.
- Total day charge: The total charges incurred by the customer for daytime usage.
- Total eve minutes: The total number of minutes the customer used during the evening.
- Total eve calls: The total number of calls the customer made or received during the evening.
- Total eve charge: The total charges incurred by the customer for evening usage.
- Total night minutes: The total number of minutes the customer used during the night.
- Total night calls: The total number of calls the customer made or received during the night.
- Total night charge: The total charges incurred by the customer for night usage.
- Total intl minutes: The total number of minutes the customer used for international calls.
- Total intl calls**: The total number of international calls the customer made.
- Total intl charge: The total charges incurred by the customer for international calls.
- Customer service calls: The number of customer service calls made by the customer.
- Churn: A binary variable indicating whether the customer has churned (1) or not (0).
- Total_calls: The total number of calls made or received by the customer across all periods (day, evening, and night).
- Total_charge: The total charges incurred by the customer across all usage periods.

Data Processing

- We checked for missing values, duplicates, categorical and numerical values to ensure that data is clean and in correct format for modelling.
- We further removed the class imbalance from the data and standardized to make data consistent and uniformity of modelling.
- This was done to ensure that the data is valid, accurate and complete for modelling.

- The chart shows heatmap for feature selection and relationship understanding.

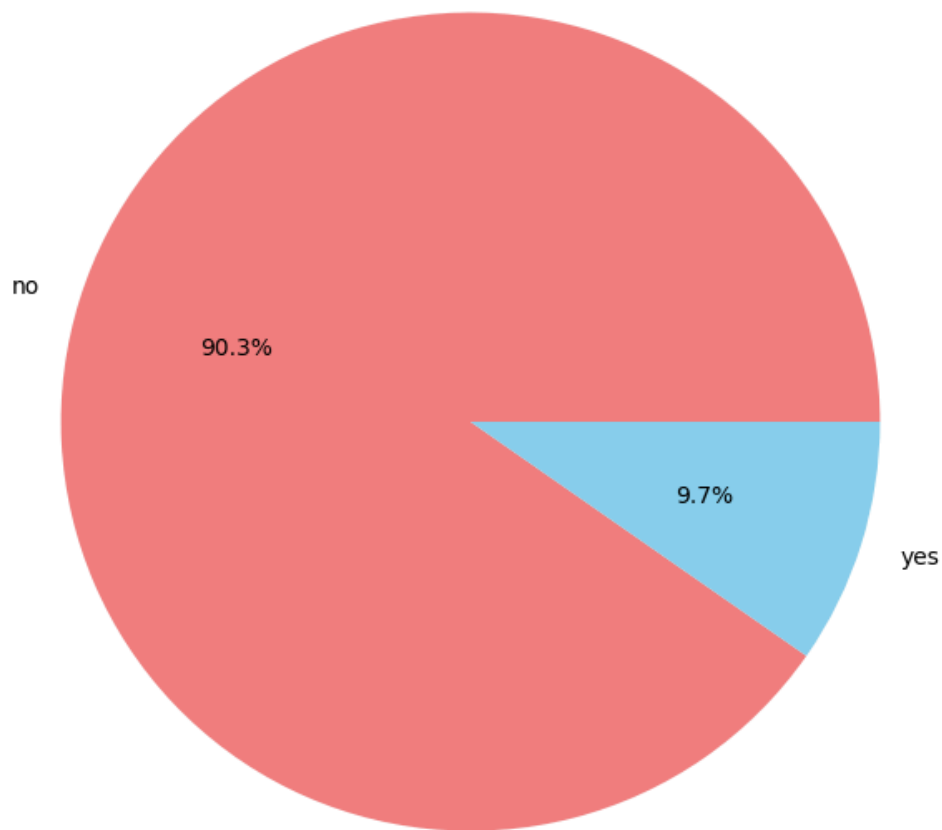


Exploratory Data Understanding

We visualized several columns on churn rates, voice mail and internation plan relationship with churn and calls made in United States to get more insights about our data.

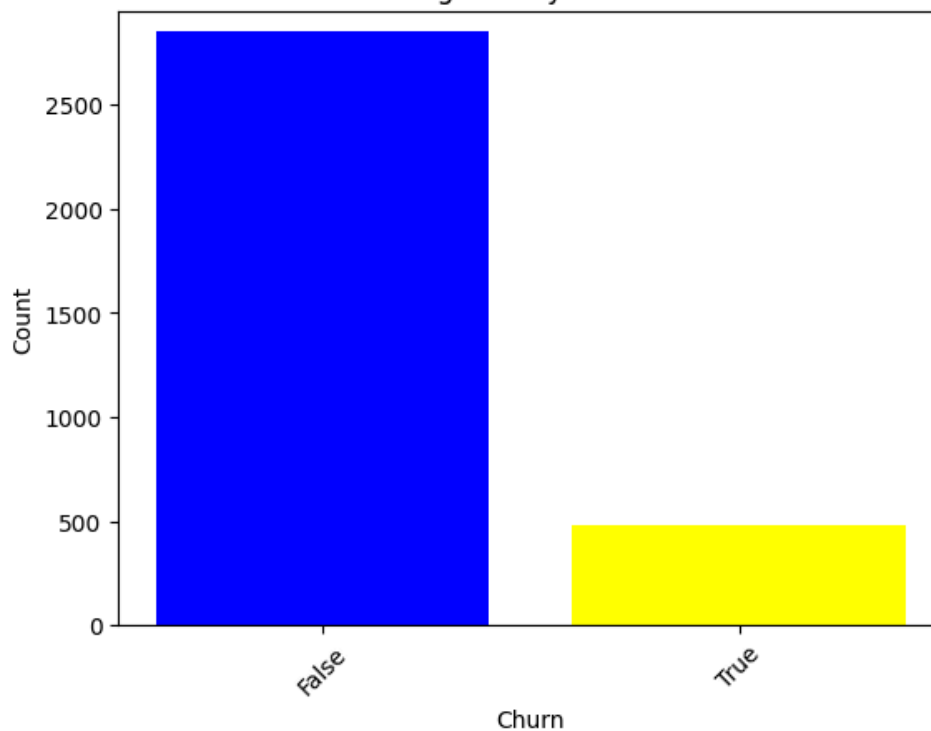
Univariate Analysis

Distribution of International Plan



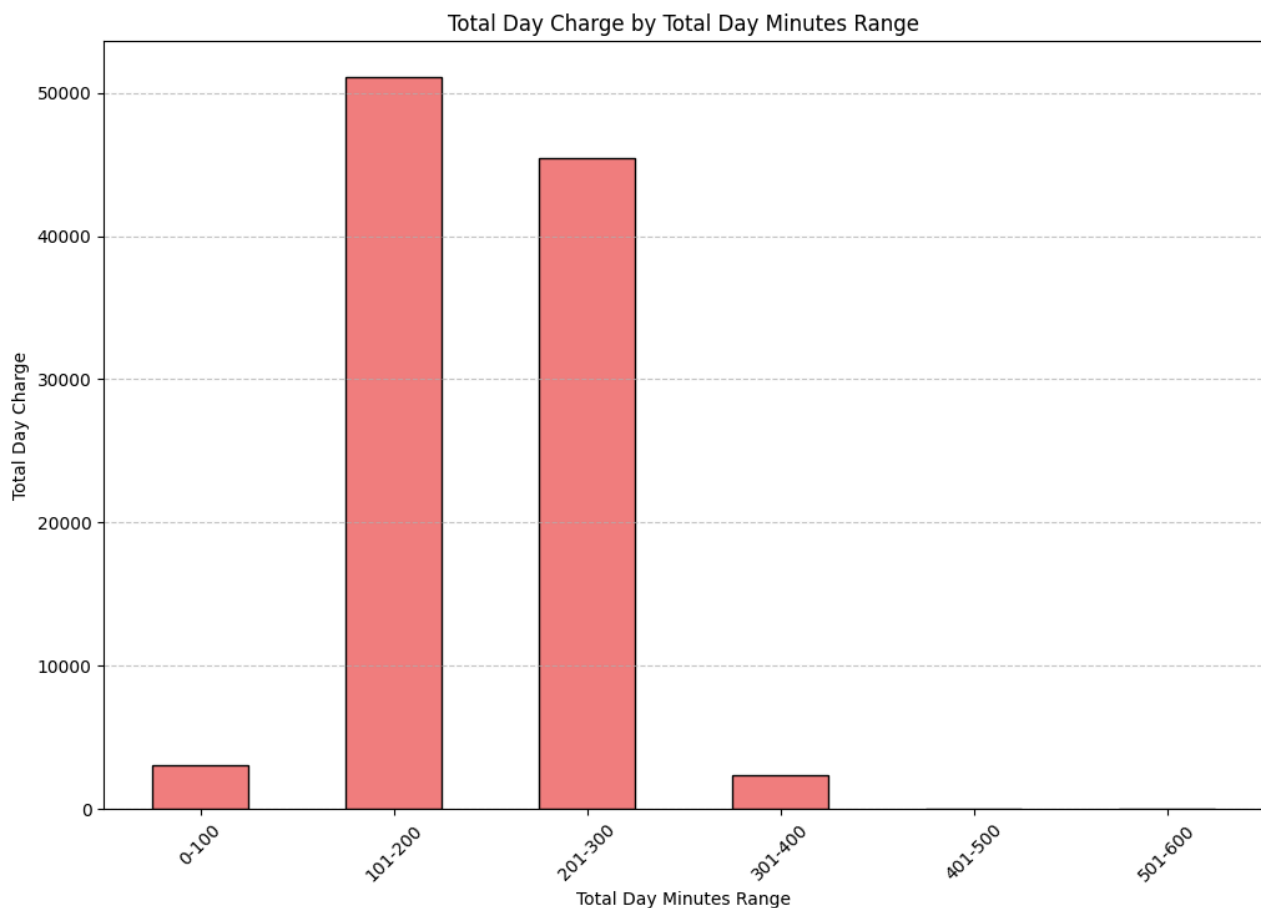
- The pie chart reveals that 90.3% of customers do not have an international plan, indicating its low popularity among SyriaTel's customer base.

Percentage of Loyal Customers

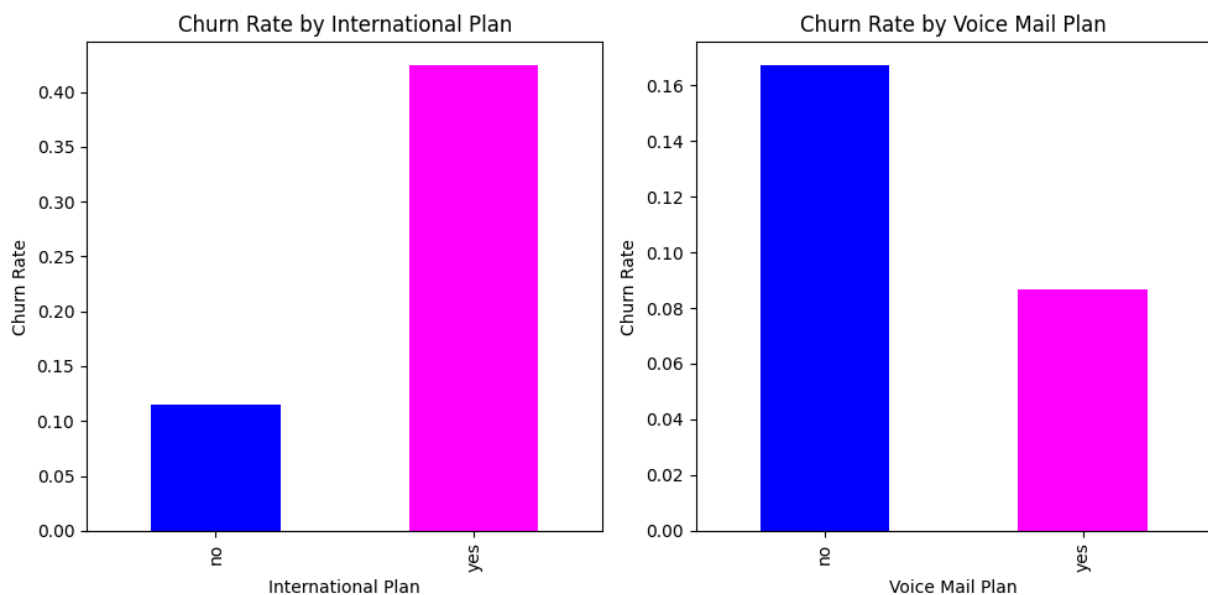


- There is higher percentage of customers not to churn, indicating good customer retention.

Bivariate Analysis



- Highest Charges : The 101-200 minute range has the highest total day charges, indicating it's the most common usage category.
- Lower Charges : The 0-100 minute range has the lowest total day charges, while charges for usage above 401 minutes decrease, likely due to pricing strategies or caps.



- Voice Mail Plan : Lower churn rate indicates higher customer loyalty.
- International Plan : Higher churn rate suggests dissatisfaction and a greater likelihood of switching.
- Customer Preference : The Voice Mail plan is preferred over the International Plan due to lower churn and higher satisfaction.

Modelling and Evaluation



Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

● Jupyter Notebook 100.0%