vkeya /
Phase3_Project

<> **Code** | ⊙ Issues | ⅂⅃ Pull requests | ▷ Actions | ⊞ Projects | ▱ Wiki | ⊙ Security | ✓ Insights | ⚙ Settings

☆ **0** stars | ⅂⅃ **0** forks | ⊙ **1** watching | ⅂⅃ **2** Branches | ⬙ **0** Tags | ⌁ Activity

⊕ **Public repository**

⅂⅃ **master** ▾ | ⅂⅃ **2 Branches** ⬙ **0 Tags** | ⅂⅃ | ⬙ | 🔍 Go to file   t | **Go to file** | + | **Add file** ▾ | **Code** | ⋯

This branch is 18 commits ahead of, 1 commit behind `main` .   ⅂⅃ **Contribute** ▾

👤 **vkeya** Update README.md		1 hour ago ⋯ ⟳	
📁 .ipynb_checkpoints	Finished lab	3 hours ago	
📁 Phase3_images	Finished lab	1 hour ago	
📄 Phase3_Notebook.ipynb	Finished lab	3 hours ago	
📄 Phase3_Notebookv1.pdf	Finished lab	3 hours ago	
📄 Phase3_Presentation.pptx	Finished lab	3 hours ago	
📄 Presentation.pdf	Finished lab	3 hours ago	
📄 README.md	Update README.md	1 hour ago	
📄 SyriaTel_Dataset.csv	Finished lab	3 hours ago	

📖 **README**   ✎ ☰

Customer Churn Analysis Using SyriaTel Dataset

Overview

This project aims to analyze customer churn behavior using the SyriaTel dataset from the telecommunications industry. The analysis involves understanding customer attributes, identifying patterns, and predicting churn using various machine learning models.

Business and Data Understanding

Business Understanding

Customer churn is a critical issue in the telecommunications industry, directly impacting revenue and growth. By understanding the factors that contribute to churn, SyriaTel company can devise strategies to retain customers and enhance their services.

Data Understanding

The SyriaTel dataset contains detailed information about customer usage patterns, plans, and service interactions. Key attributes include account length, international plan, voice mail plan, total call minutes across different times of the day, customer service calls, and whether the customer churned or not.

Dataset Choice

The SyriaTel dataset is chosen for its comprehensive coverage of customer behavior and service usage. It includes attributes such as: Account length: The duration (in days) that the customer has had an account with the company. International plan: Binary indicator of whether the customer has an international calling plan (1 = Yes, 0 = No). Voice mail plan: Binary indicator of whether the customer has a voice mail plan (1 = Yes, 0 = No). Number vmail messages: Number of voice mail messages the customer has. Total day minutes: Total number of minutes the customer has used during the day. Total day calls: Total number of calls the customer has made during the day. Total day charge: Total charge for the customer's day usage. Total eve minutes: Total number of minutes the customer has used during the evening. Total eve calls: Total number of calls the customer has made during the evening. Total eve charge: Total charge for the customer's evening usage. Total intl calls: Total number of international calls made by the customer. Total intl charge: Total charge for the customer's international usage. Customer service calls: Number of calls the customer has made to customer service. Churn: Binary indicator of whether the customer has churned (1 = Yes, 0 = No). Total_calls: Total number of calls made by the customer. Total_minutes: Total number of minutes the customer has used. Total_charge: Total charge for the customer's usage. Day_calls_per_minute: Ratio of day calls to day minutes. Eve_calls_per_minute: Ratio of evening calls to evening minutes. Night_calls_per_minute: Ratio of night calls to night minutes.

Stakeholder Audience

The primary stakeholders for this analysis are the customer retention team, marketing strategists, and senior management within the telecommunications company. They are interested in identifying the key factors driving customer churn and developing strategies to reduce churn rates.

Methodology

The analysis of this dataset was conducted using a series of methodical steps to extract meaningful insights and develop a predictive model to identify potential churners. Below is a breakdown of the methodology employed:

Data Preprocessing:

Data Cleaning: Identifying and handling missing values, outliers, and any inconsistencies in the data.

Data Transformation: Converting categorical variables into numerical formats using techniques such as one-hot encoding.

Exploratory Data Analysis (EDA):

Descriptive Statistics: Calculating summary statistics for each column to understand the central tendency and dispersion of the data.

```
df.describe()
```

	account length	area code	number vmail messages	total day minutes	total day calls	total day charge	total eve minutes	total eve calls	total eve charge	total night minutes	total night calls	
count	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3333.000000	3
mean	101.064806	437.182418	8.099010	179.775098	100.435644	30.562307	200.980348	100.114311	17.083540	200.872037	100.107711	
std	39.822106	42.371290	13.688365	54.467389	20.069084	9.259435	50.713844	19.922625	4.310668	50.573847	19.568609	
min	1.000000	408.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	23.200000	33.000000	
25%	74.000000	408.000000	0.000000	143.700000	87.000000	24.430000	166.600000	87.000000	14.160000	167.000000	87.000000	
50%	101.000000	415.000000	0.000000	179.400000	101.000000	30.500000	201.400000	100.000000	17.120000	201.200000	100.000000	
75%	127.000000	510.000000	20.000000	216.400000	114.000000	36.790000	235.300000	114.000000	20.000000	235.300000	113.000000	
max	243.000000	510.000000	51.000000	350.800000	165.000000	59.640000	363.700000	170.000000	30.910000	395.000000	175.000000	

Visualization: Using various plots (histograms, box plots, scatter plots) to visually inspect the distribution of variables and relationships between them.
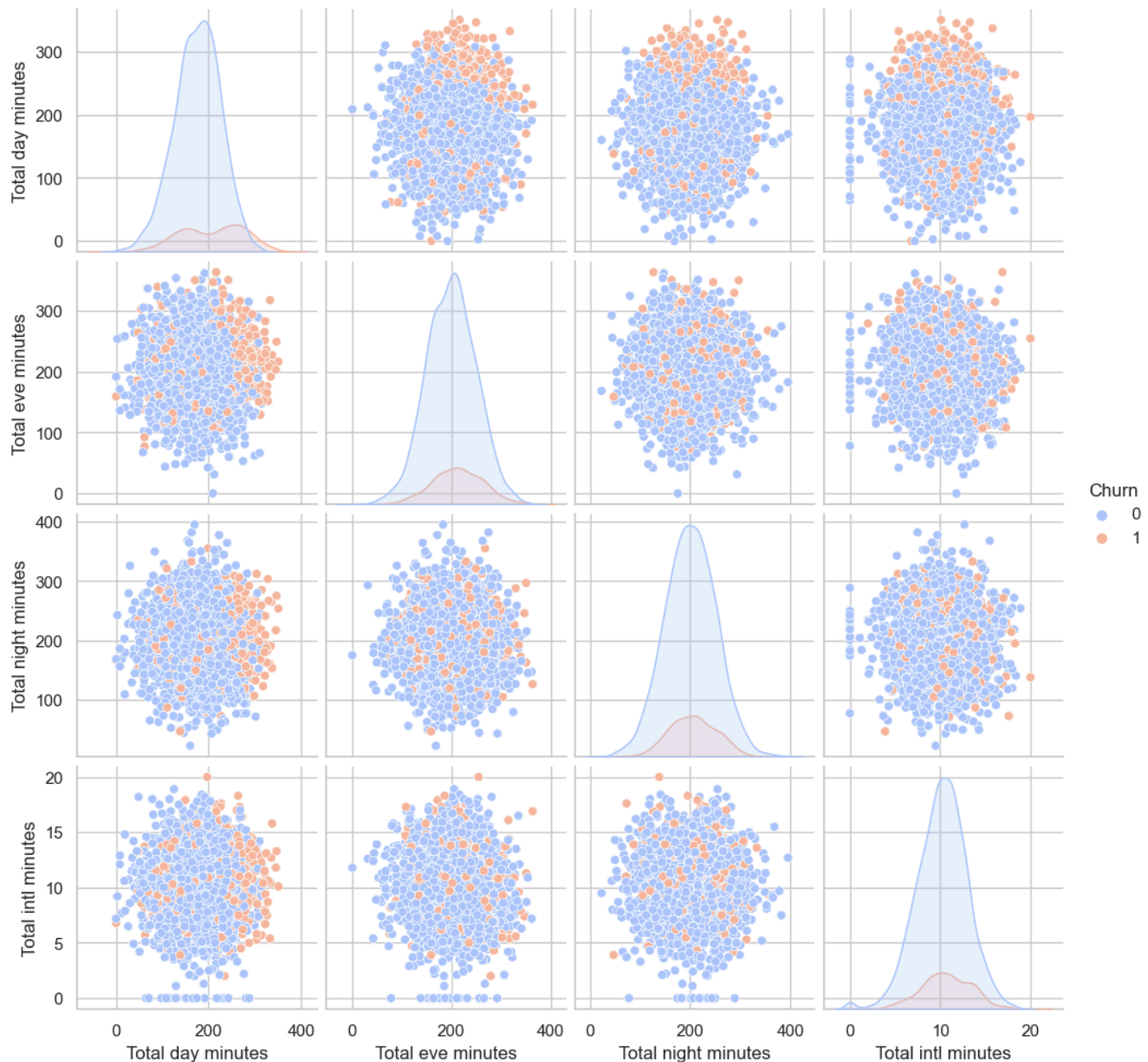
Univariate Analysis
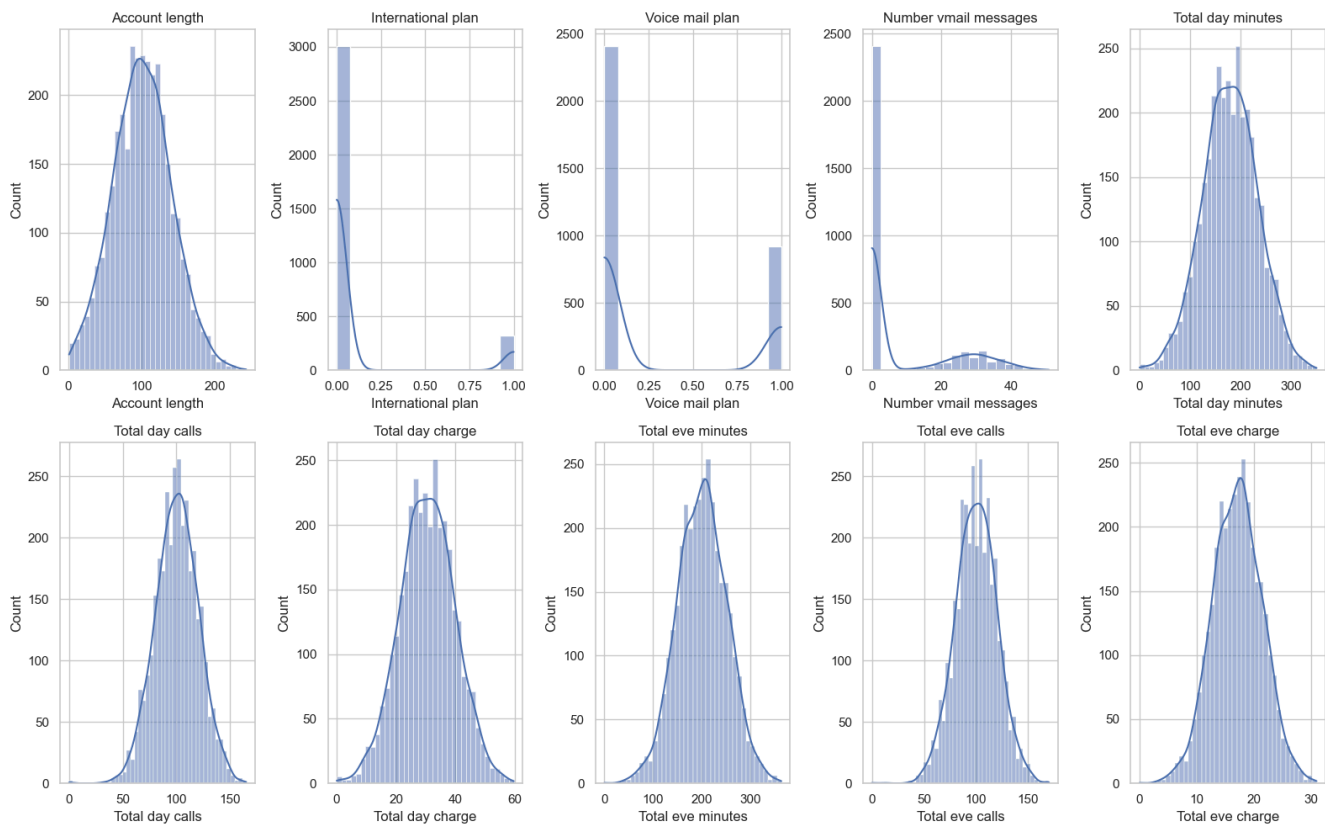
Bivariate Analysis

Correlation Heatmap
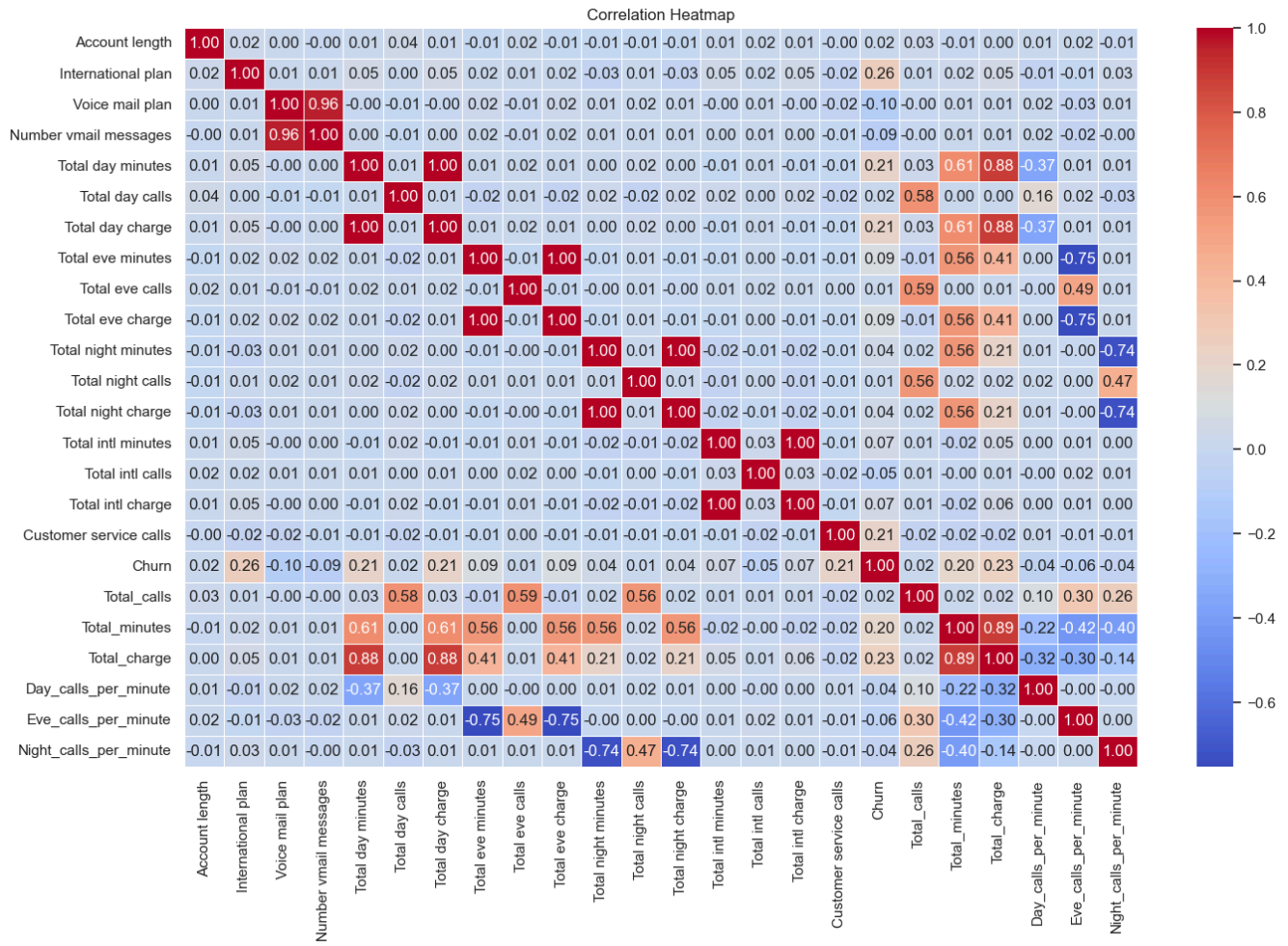
Multivariate Analysis

Pair Plot for Selected Features

Feature Engineering:

New Features: Creating new features such as Total_calls, Total_minutes, and Total_charge to provide additional insights. Ratios and Proportions: Calculating ratios such as Day_calls_per_minute, Eve_calls_per_minute, and Night_calls_per_minute to capture customer behavior more effectively.

Modeling

Model Selection: We Evaluated different machine learning models (i.e., logistic regression, decision trees, random forests) to predict churn. Model Training: Splitting the dataset into training and testing sets to train the selected model. Model Evaluation: Assessing the performance of the model using metrics such as accuracy, precision, recall, and F1-score.

Evaluation

The models are evaluated based on:

- **Accuracy**: Proportion of correctly predicted churn and non-churn cases.

- **Precision**: Proportion of true positive churn predictions out of all positive predictions.
- **Recall**: Proportion of true positive churn predictions out of all actual churn cases.
- **F1 Score**: Harmonic mean of precision and recall.

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package