

TELECOM CHURN PREDICTION

GROUP 2:

BHAVANA R C

AKANSHA

SAI RAKESH A.V

INTRODUCTION

Telecom churn refers to the rate at which customers stop using a telecom service, switching to competitors or canceling their subscriptions altogether. In the telecom industry, churn is a critical business metric because high churn rates can significantly impact revenue and market share. Factors influencing churn include customer satisfaction, service quality, pricing, and the availability of better alternatives.

Problem Statement: "The goal of this project is to predict customer churn in a telecom company, which helps in identifying customers who may leave the service."



PROJECT OBJECTIVE

- **Identify Key Drivers of Churn:**Analyze customer data to determine factors that contribute most to churn, such as pricing, service quality, or contract type.
- **Develop Predictive Models:**Create machine learning models to predict which customers are likely to churn, enabling proactive retention efforts.
- **Enhance Customer Retention:**Use insights to implement targeted retention strategies, like personalized discounts or loyalty programs.
- **Reduce Revenue Loss:** Minimize financial impact by retaining more customers and reducing new customer acquisition costs.

DATASET DISCRPTION

- dataset includes information on 5,000 telecom customers, with data points that help us understand customer behavior and usage patterns. Here's a breakdown:
- • **Basic Customer Info:** Where they're located (state and area.code) and how long they've been with the company (account.length).
- • **Service Plans:** Whether they have specific plans, like a voice plan or international plan (voice.plan, intl.plan).
- • **Usage Patterns:** Details on call and usage habits: how much time they spend on calls during different parts of the day (intl.mins, day.mins, eve.mins, night.mins), how many calls they make, and the associated charges.
- • **Customer Service Interaction:** The number of times they've called customer support (customer.calls).
- • **Churn Status:** A simple yes/no column (churn) showing if they've left The the service.

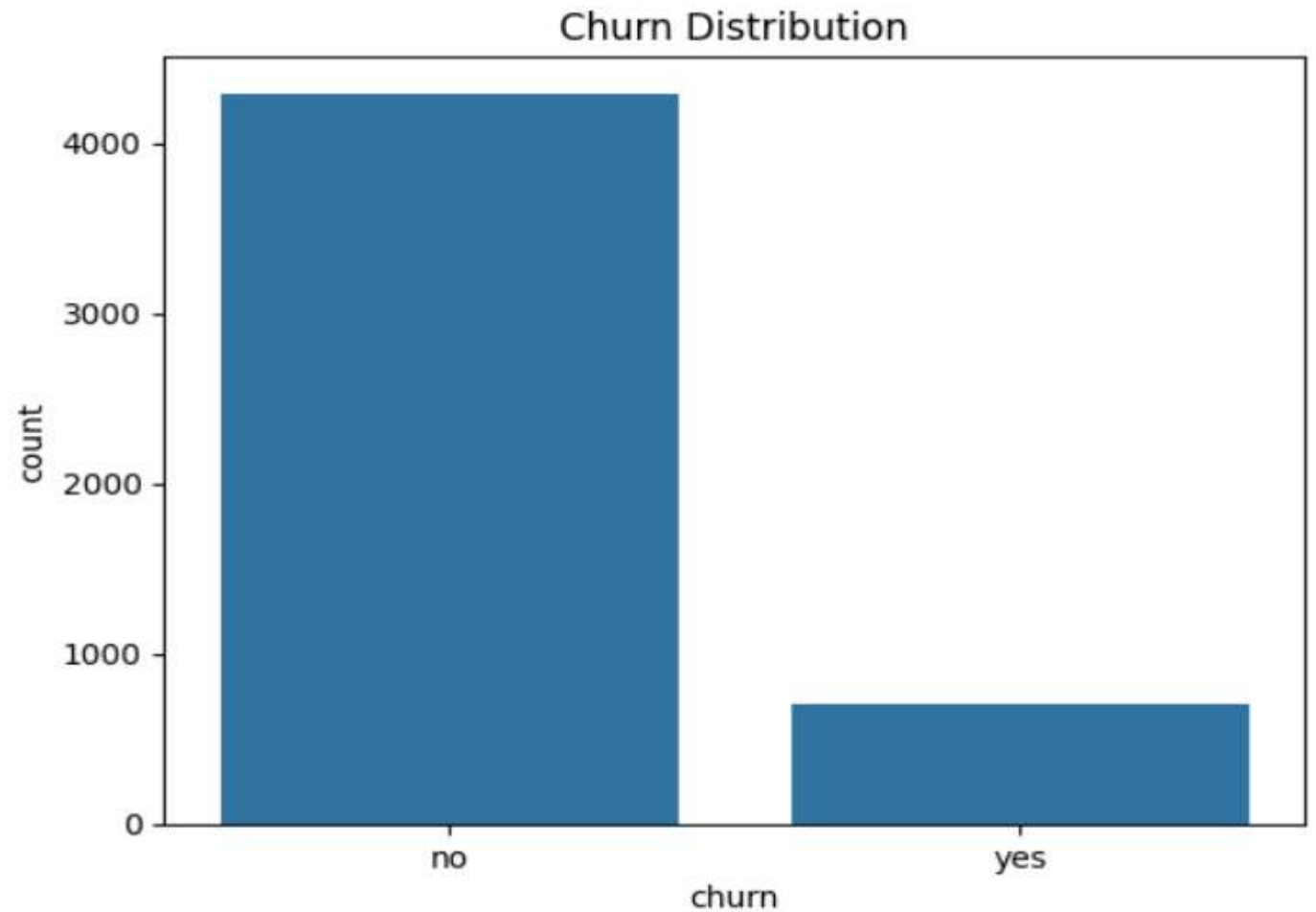
EXPLORATORY DATA ANALYSIS (EDA)

- **Data Structure and Summary:** Check data types, preview the dataset, and generate summary statistics to understand its general layout and distributions.
- **Missing Values and Duplicates:** Identify and handle missing values and remove duplicates to ensure data quality.
- **Univariate and Bivariate Analysis:** Examine individual features and relationships between pairs of features to detect patterns, distributions, and potential correlations.
- **Outlier Detection and Transformation:** Identify outliers and apply scaling, encoding, or other transformations as necessary for consistent data.
- **Insights and Feature Engineering:** Summarize findings and create new features based on observed patterns to enhance predictive power for modeling.

COUNT PLOT

```
churn_rate =  
data['churn'].value_counts(normalize=True) * 100  
print(churn_rate)  
  
sns.countplot(x='churn', data=data)  
plt.title('Churn Distribution')  
plt.show()
```

```
churn  
no      85.86  
yes     14.14  
Name: proportion, dtype: float64
```



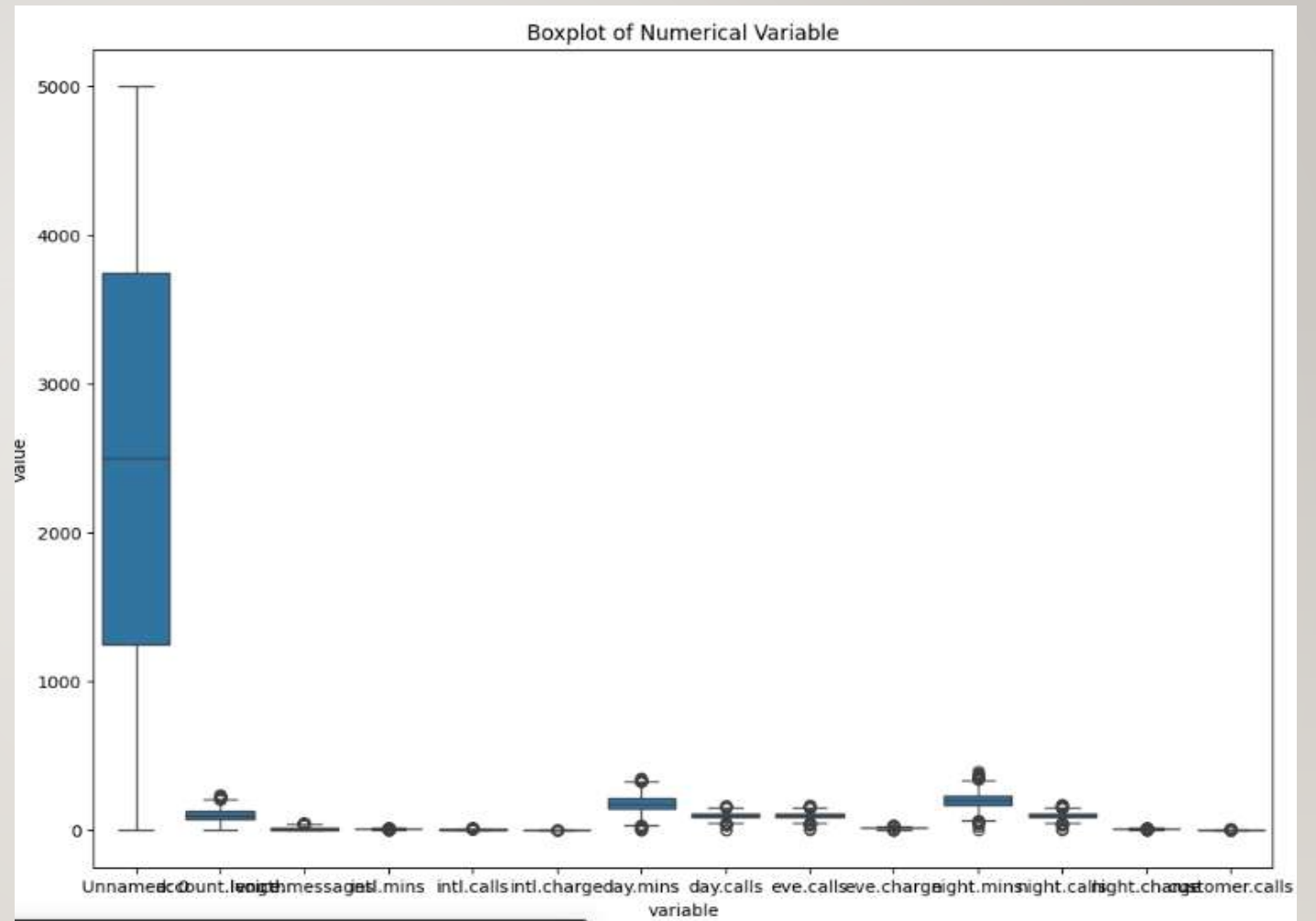
FEATURE DISTRIBUTIONS

```
plt.figure(figsize=(12,9))
```

```
sns.boxplot(x="variable",y="value",  
data=pd.melt(data[numerical_columns.c  
olumns]))
```

```
plt.title('Boxplot of Numerical Variable')
```

```
plt.show()
```



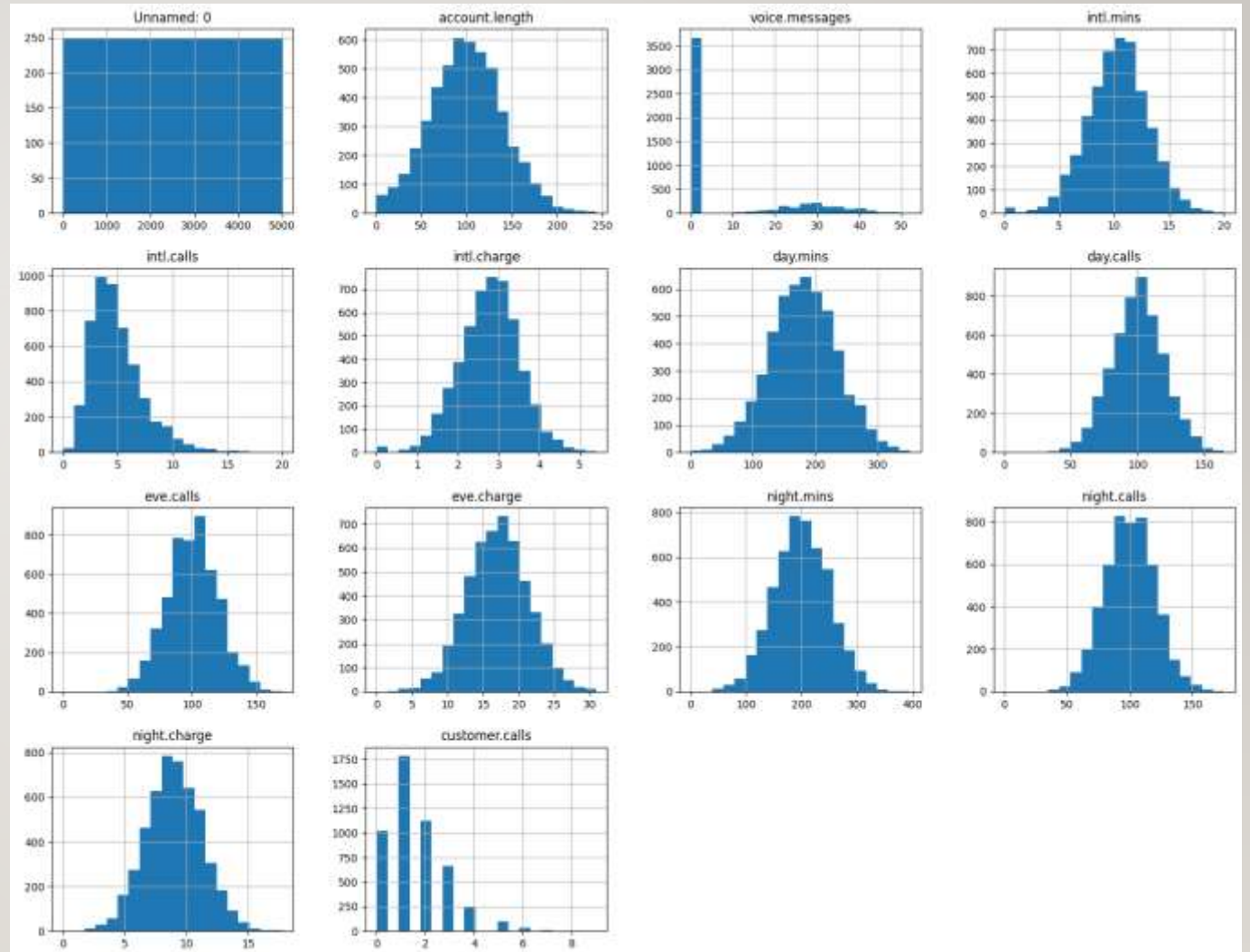
HISTOGRAM

```
plt.figure(figsize=(8, 5))
```

```
import matplotlib.pyplot as plt
```

```
data[numerical_columns].hist(bins=20,figsize=(20,15))
```

```
plt.show()
```



CORRELATION MATRIX

```
# Heatmap of numerical columns
```

```
numerical_columns =  
data.select_dtypes(include=['number'])
```

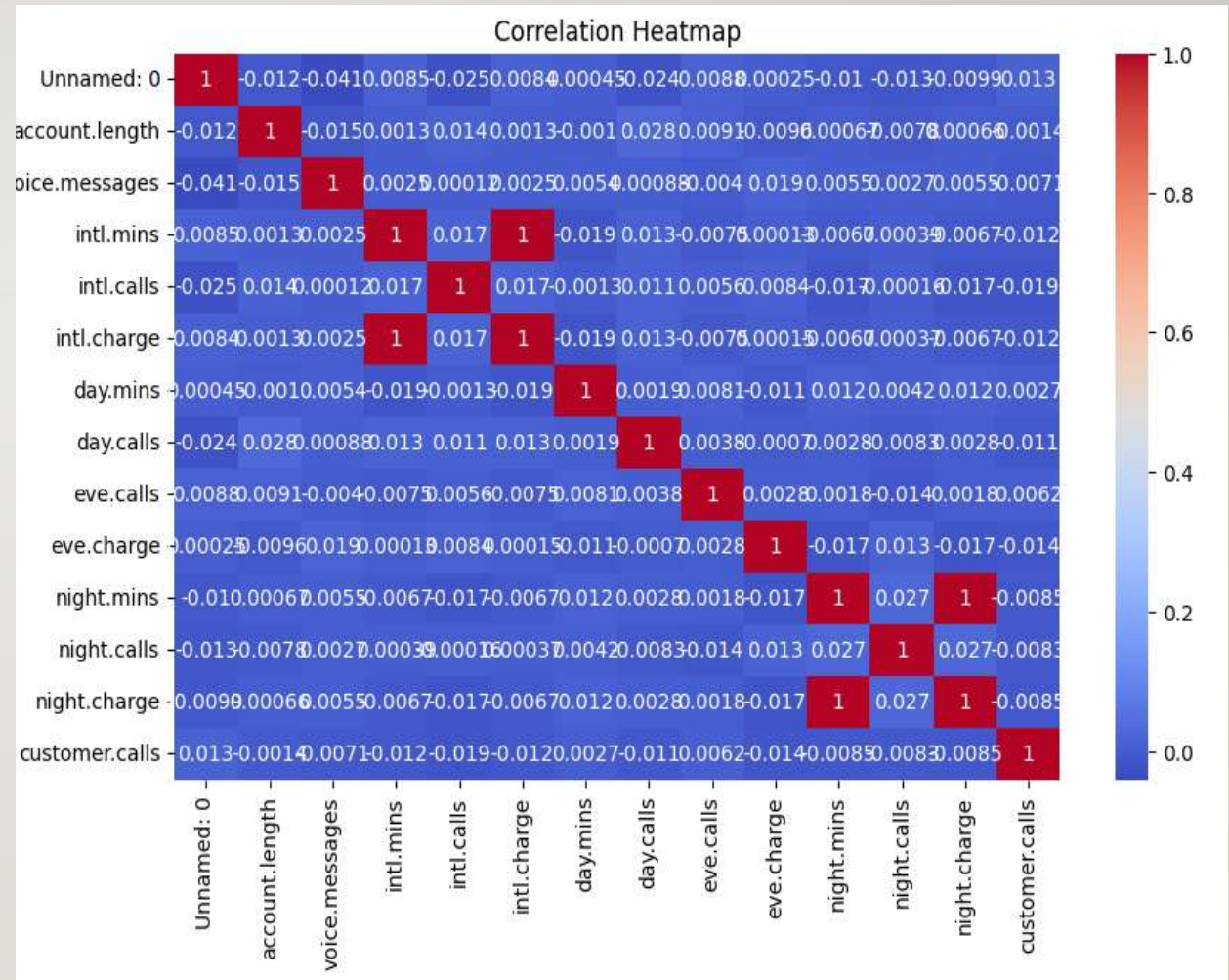
```
corr_matrix = numerical_columns.corr()
```

```
plt.figure(figsize=(10, 6))
```

```
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
```

```
plt.title('Correlation Heatmap')
```

```
plt.show()
```

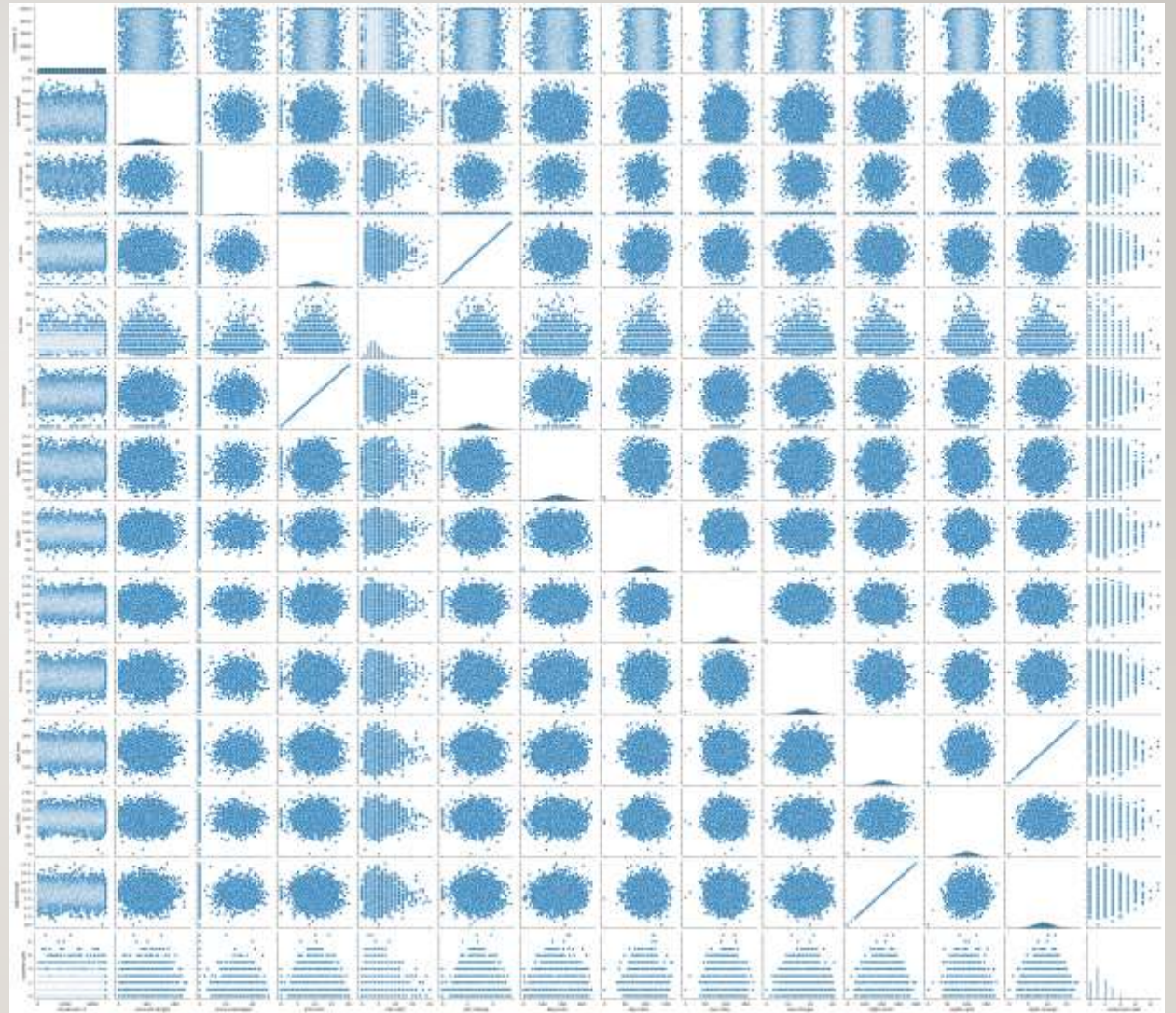


PAIR PLOT

```
plt.figure(figsize=(10, 15))
```

```
sns.pairplot(data)
```

```
plt.show()
```



MODEL SELECTION AND DEVELOPMENT

- Models Considered:

We trained several models, including Logistic Regression, KNN, Random Forest. But we have got the highest accuracy in Random Forest Classification.

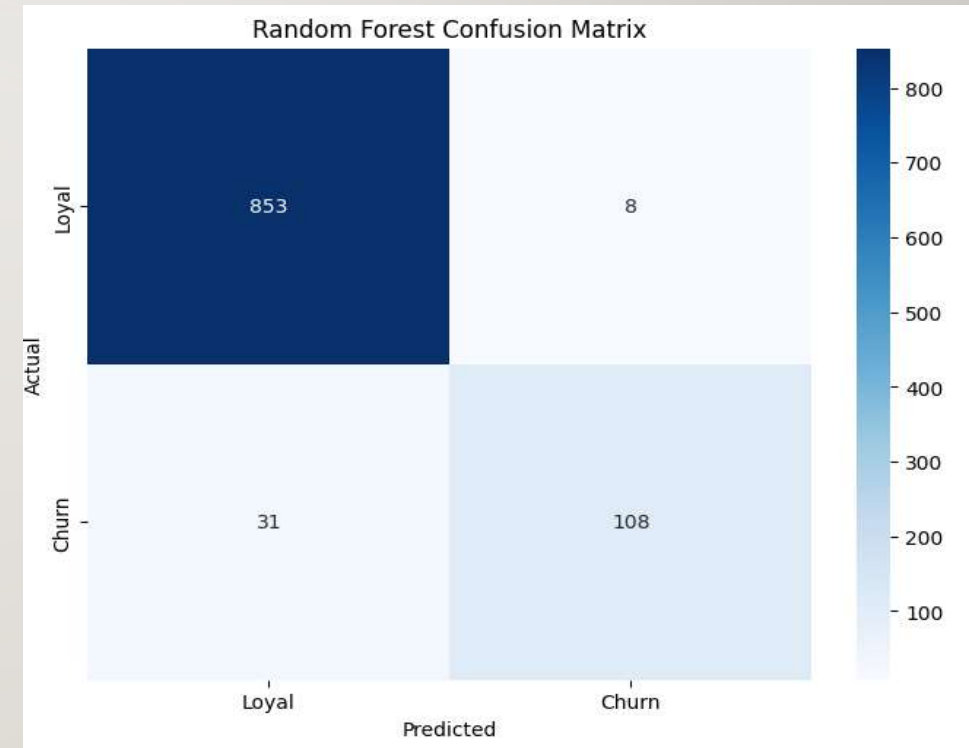
```
Random Forest Accuracy: 0.9610
Classification Report:
              precision    recall  f1-score   support

     no         0.96       0.99       0.98       861
     yes         0.93       0.78       0.85       139

 accuracy                   0.96       1000
 macro avg         0.95       0.88       0.91       1000
weighted avg         0.96       0.96       0.96       1000
```

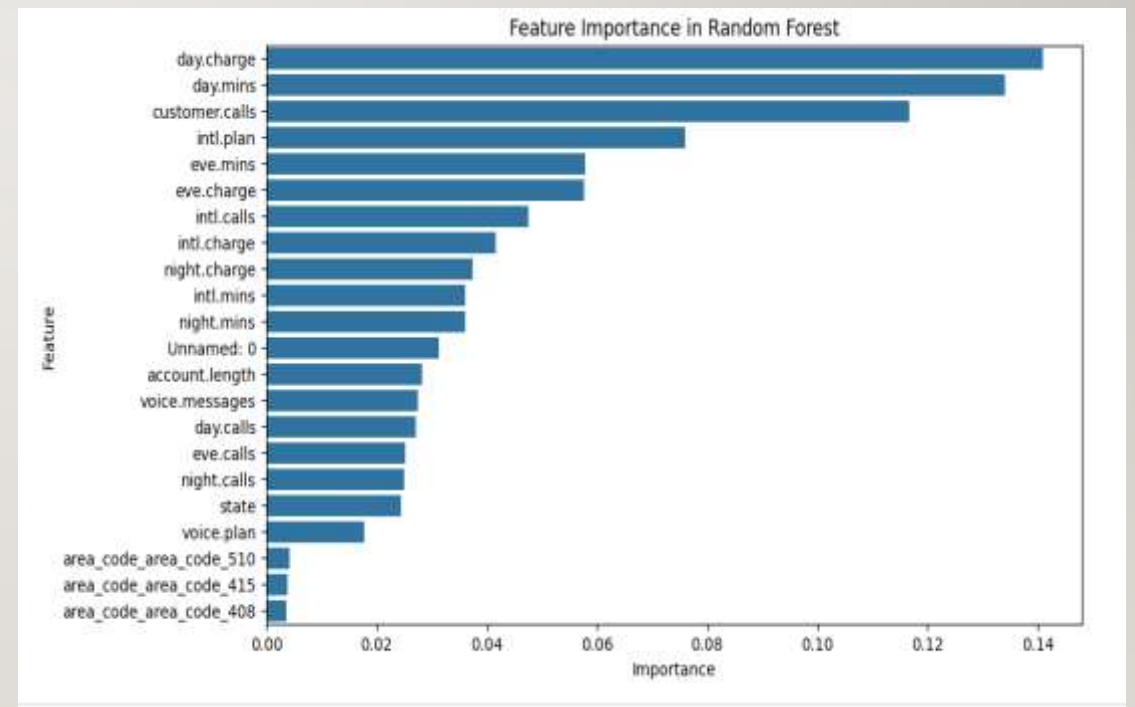
RANDOM FOREST CONFUSION MATRIX

- The confusion matrix help to visualize how many customers were correctly or incorrectly classified as churners or loyal



FEATURE IMPORTANCE IN RANDOM FOREST

Feature importance random forest shows that day.charge, days.mins and customer.calls are the most important for prediction and area.code features matter the least.



STREAMLIT

- Streamlit allows you to quickly create web apps that are interactive and easy to use. It turns your Python code into a fully functioning app in just a few lines. In this case, it's helping to make churn predictions accessible and understandable to anyone, without needing complex technical knowledge. This Streamlit telecom churn app lets you quickly predict which customers might leave, based on their data. By entering simple information like how long a customer has been with the company and how much they pay, the app tells you whether they are likely to churn or not. This can help your telecom business take action to keep customers happy and reduce churn

CONCLUSION:

The telecom churn analysis provides insights into customer behavior and the factors affecting their likelihood of leaving. By examining features like plan type, usage patterns, and customer service interactions, we can predict customers at risk of churn. This information allows telecom companies to proactively implement strategies like customized plans and special offers for high-risk customers. Such targeted actions help reduce churn, improve customer satisfaction, and stabilize revenue. This project emphasizes the value of data-driven decisions in helping telecom companies retain customers and support business growth.