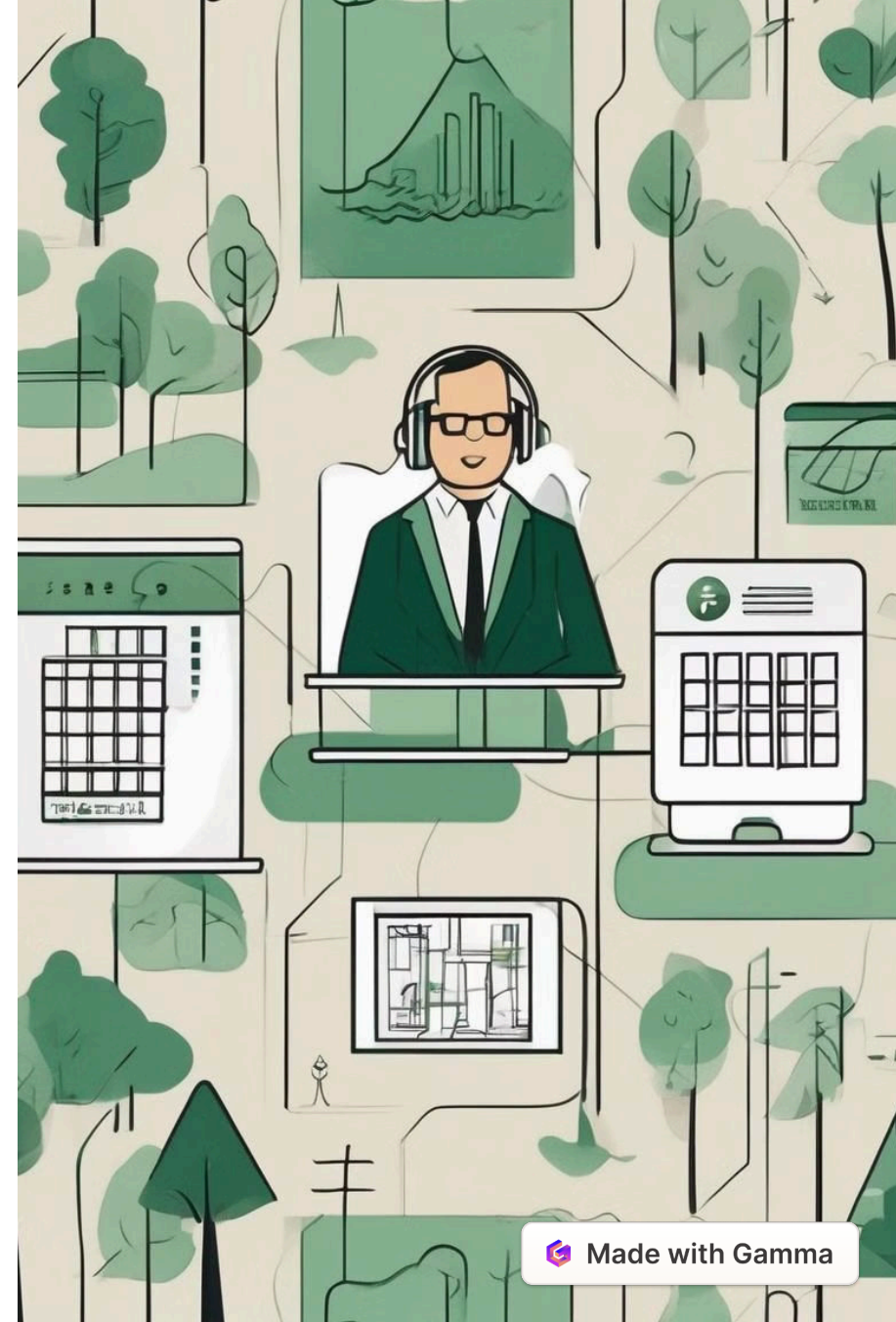# Customer Churn Analysis and Prediction

Chirag Raj (Cs21b1038) and BS Dhawal (Cs21b1005)

Supervisor - Dr Natesha B V

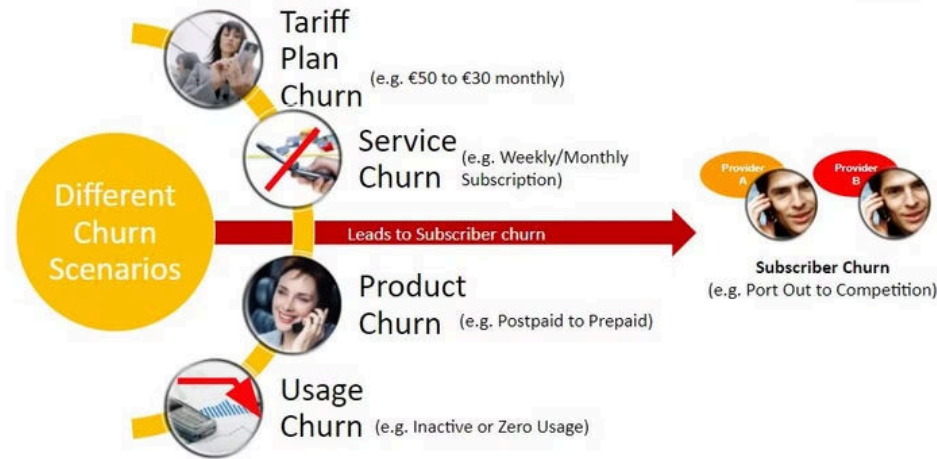# Introduction to Customer Churn Prediction

Churn - the phenomenon of customers discontinuing their relationship with a business with ending the use of thier product or services.

In the highly competitive telecom industry, understanding and predicting customer churn is crucial for business success. By identifying customers at risk of leaving, companies can proactively engage with them and improve customer retention.
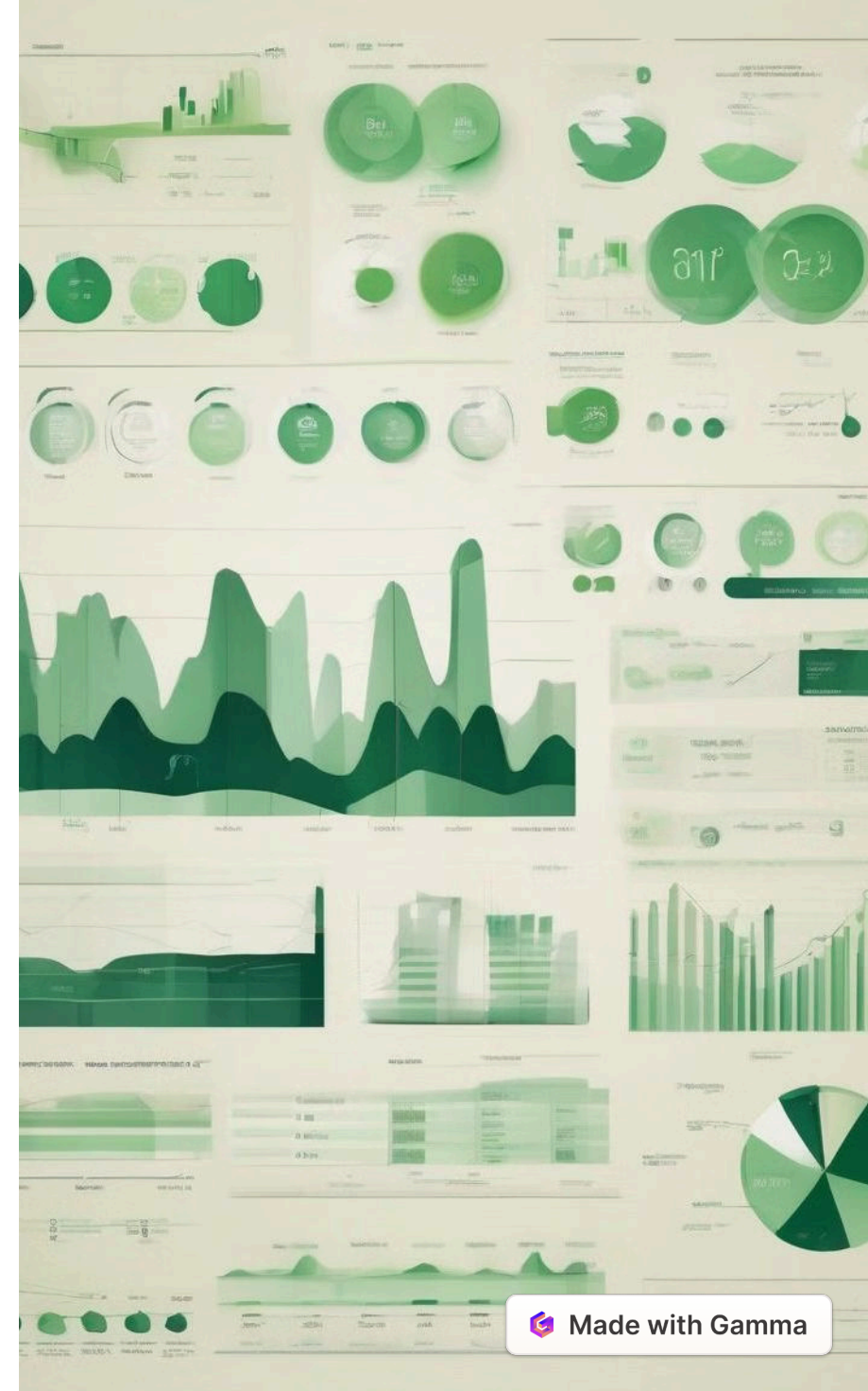
Different Types of Churn Scenarios

Subscriber Churn can be in different forms and not just exit from the base

Different Churn Scenarios

Tariff Plan Churn (e.g. €50 to €30 monthly)

Service Churn (e.g. Weekly/Monthly Subscription)

Product Churn (e.g. Postpaid to Prepaid)

Usage Churn (e.g. Inactive or Zero Usage)

Leads to Subscriber churn

Subscriber Churn (e.g. Port Out to Competition)

Provider A

Provider B

Problem Statement

- In today's competitive market, retaining customers is just as crucial as acquiring new ones.

- Businesses are increasingly facing the challenge of customer churn, which can lead to significant losses in revenue and increased operational costs.

- Predicting customer churn allows companies to proactively implement retention strategies to prevent loss of customers.

- The significant diversity in customer behaviors and profiles makes it difficult to identify and generalize the pre-churn indicators.

- This report addresses the problem by applying machine learning techniques to historical data to model and predict churn probability.

- The objective is to determine which factors are most influential in customer turnover and to assess the effectiveness of various predictive models in forecasting churn.

- This investigation will enable the deployment of more personalized and timely interventions aimed at retaining customers in various business sectors.

Methodology of Our Project

1

Step 1: Import the data

Retrieve and bring in the required 0data for analysis.

2

Step 2: Perform EDA

Explore and analyze the data to gain insights and understand its characteristics.

3

Step 3: Create data visualizations

Represent the data in visual formats to better understand patterns and trends.

4

Step 4: Build data models

Create mathematical models using the data to make predictions or draw conclusions.

5

Step 5: Import models

Bring in pre-built models or algorithms for further analysis.

6

Step 6: Select models with the best scores

Evaluate and choose the models with the highest performance based on defined criteria.

7

Step 7: Check accuracy

Assess the accuracy and reliability of the selected models.

8

Step 8: Verify results

Compare the predicted values with the actual values using a confusion matrix.

9

Step 9: Present the results

Communicate and showcase the findings and insights derived from the analysis.

# Data Preprocessing and Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a crucial step in the data analysis process. It involves examining and visualizing the data to gain insights, understand patterns, and identify relationships between variables. EDA helps in understanding the structure and characteristics of the data, detecting any anomalies or missing values, and selecting appropriate data preprocessing techniques. Here's the updated content with a brief explanation of Exploratory Data Analysis:
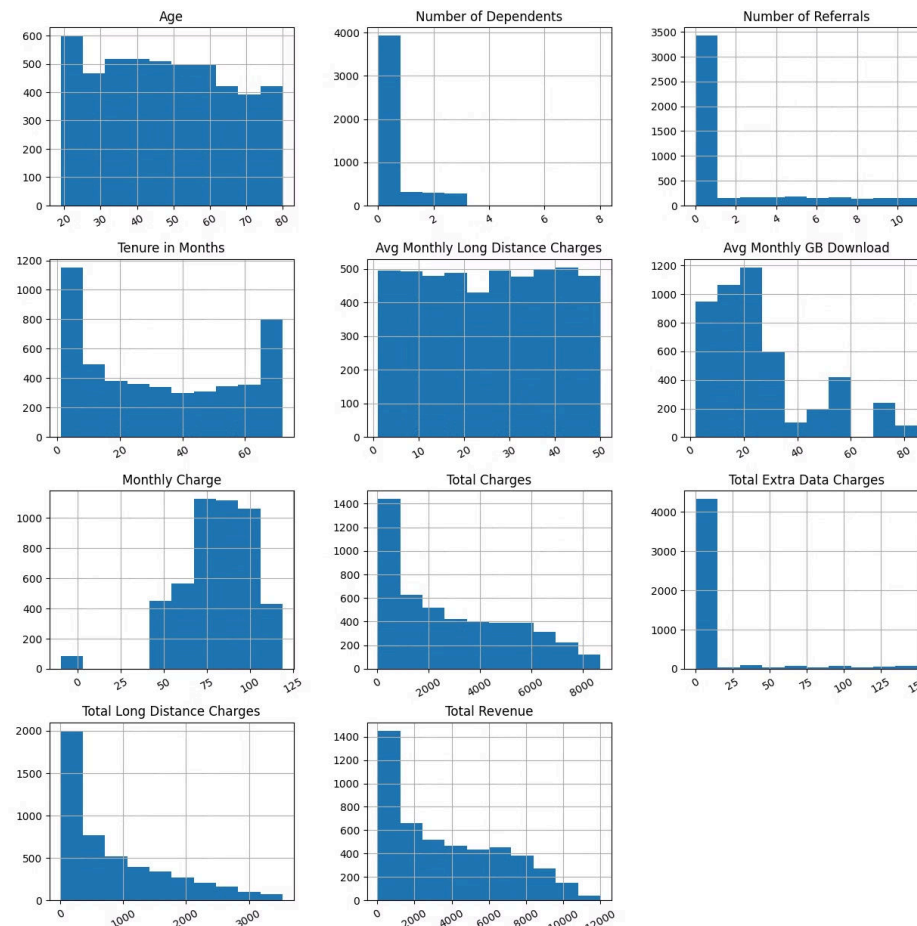
Defining Correlation between the columns in the dataset
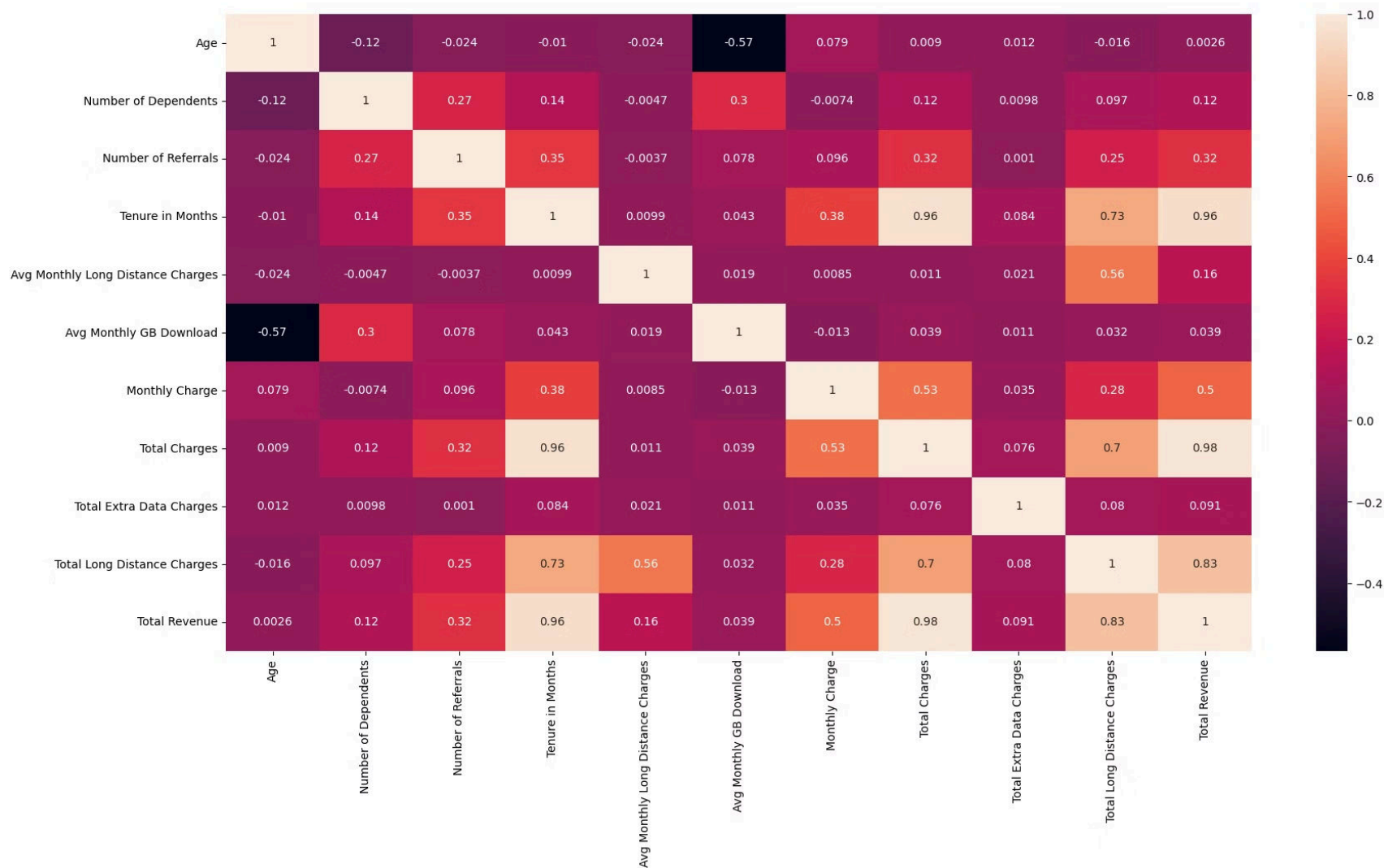
Customer Status
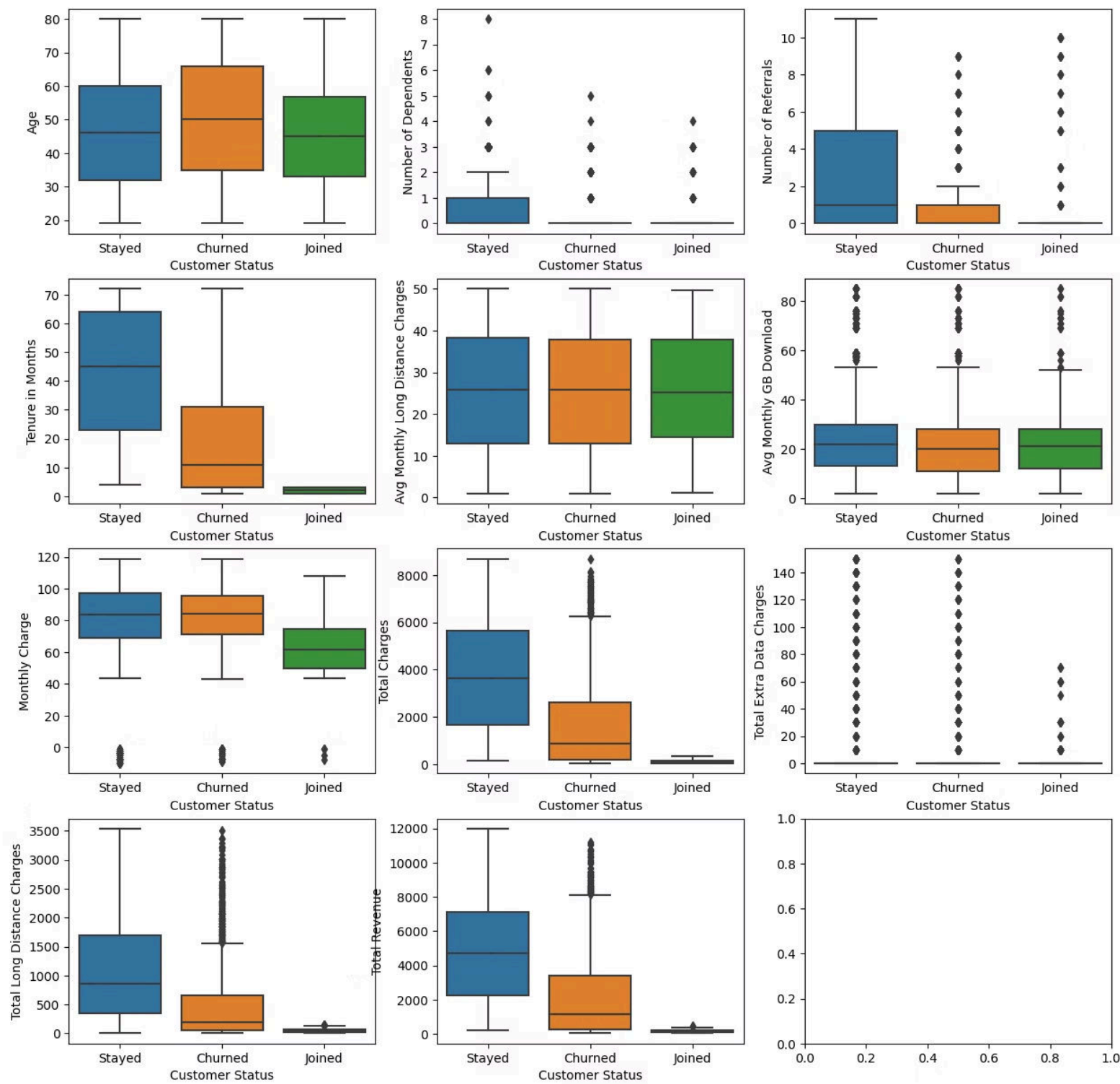
Model Evaluation and Comparison

Models used in this card:

- **RandomForestClassifier:** This is an ensemble learning method that constructs multiple decision trees and combines their predictions to make a final prediction. It is known for its high accuracy and ability to handle complex datasets.

- **LogisticRegression:** This is a statistical model used for binary classification. It calculates the probability of an instance belonging to a certain class using a logistic function. It is commonly used when the target variable is categorical.

- **GaussianNB:** This is a Naive Bayes classifier that assumes the features are independent and follows a Gaussian distribution. It is often used for text classification or problems with continuous input variables.

- **DecisionTreeClassifier:** This is a non-parametric supervised learning method that creates a tree-like model of decisions and their possible consequences. It is known for its interpretability and ability to handle both categorical and numerical data.

- **XGBClassifier:** This is an implementation of the gradient boosting algorithm that uses a combination of weak prediction models (decision trees) and combines their predictions to make a final prediction. It is known for its high performance and scalability.

After comparing the models, it was found that **XGBClassifier** has the highest accuracy score.

Comparison Graph of Models

To evaluate the performance of the various churn prediction models, we have created a comprehensive comparison table. This table showcases the key metrics, including accuracy, precision, recall, and F1-score, for each of the models under consideration.

Model	Best Score
RandomForestClassifier	0.785109
LogisticRegression	0.782834
GaussianNB	0.367735
DecisionTreeClassifier	0.772699
XGBClassifier	0.819235

As evident from the comparison, the XGBClassifier model emerges as the top performer, with the highest Best-score among the models evaluated.

Accuracy Score of Different Models
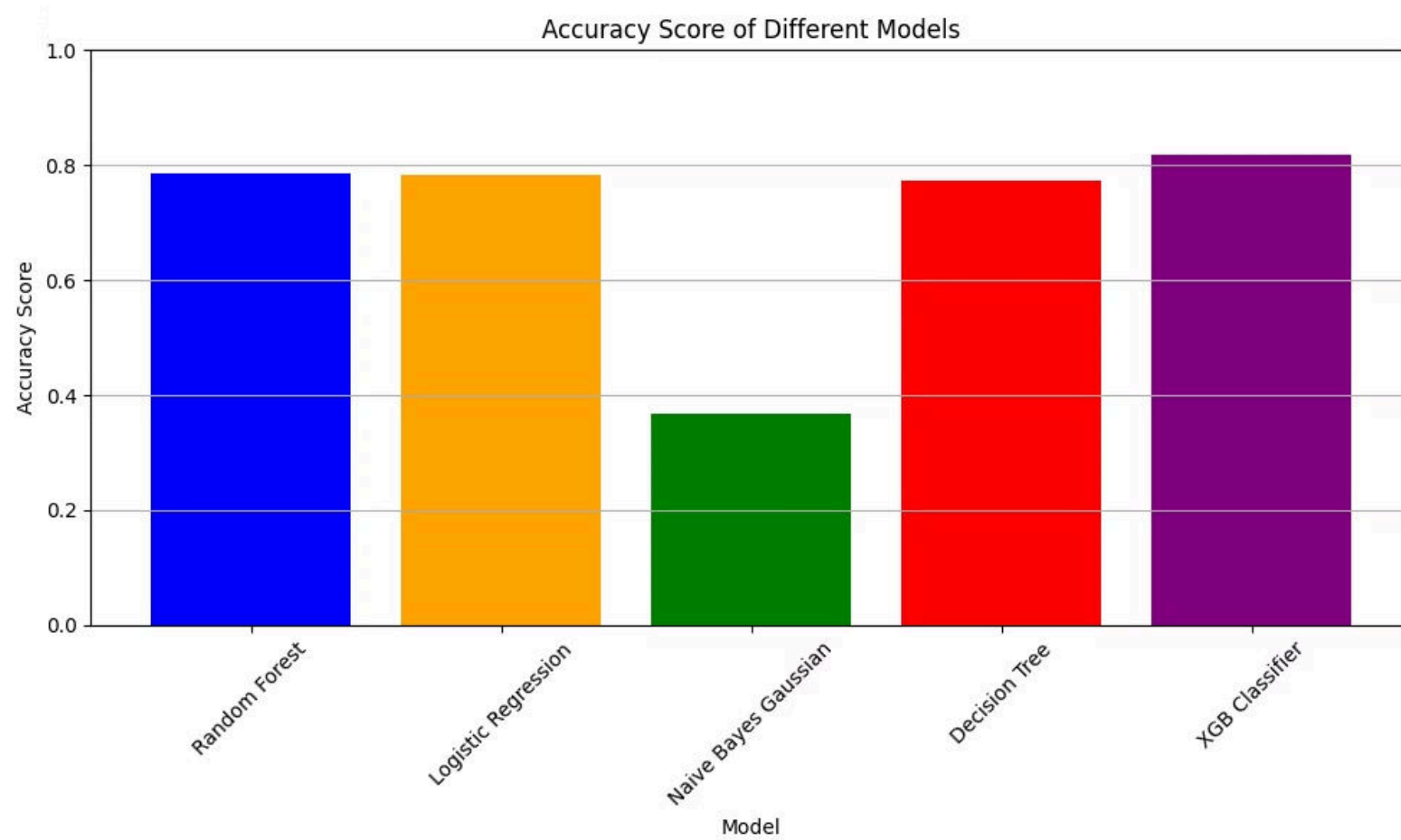
Verifying Actual data and predicted data through Confusion Matrix
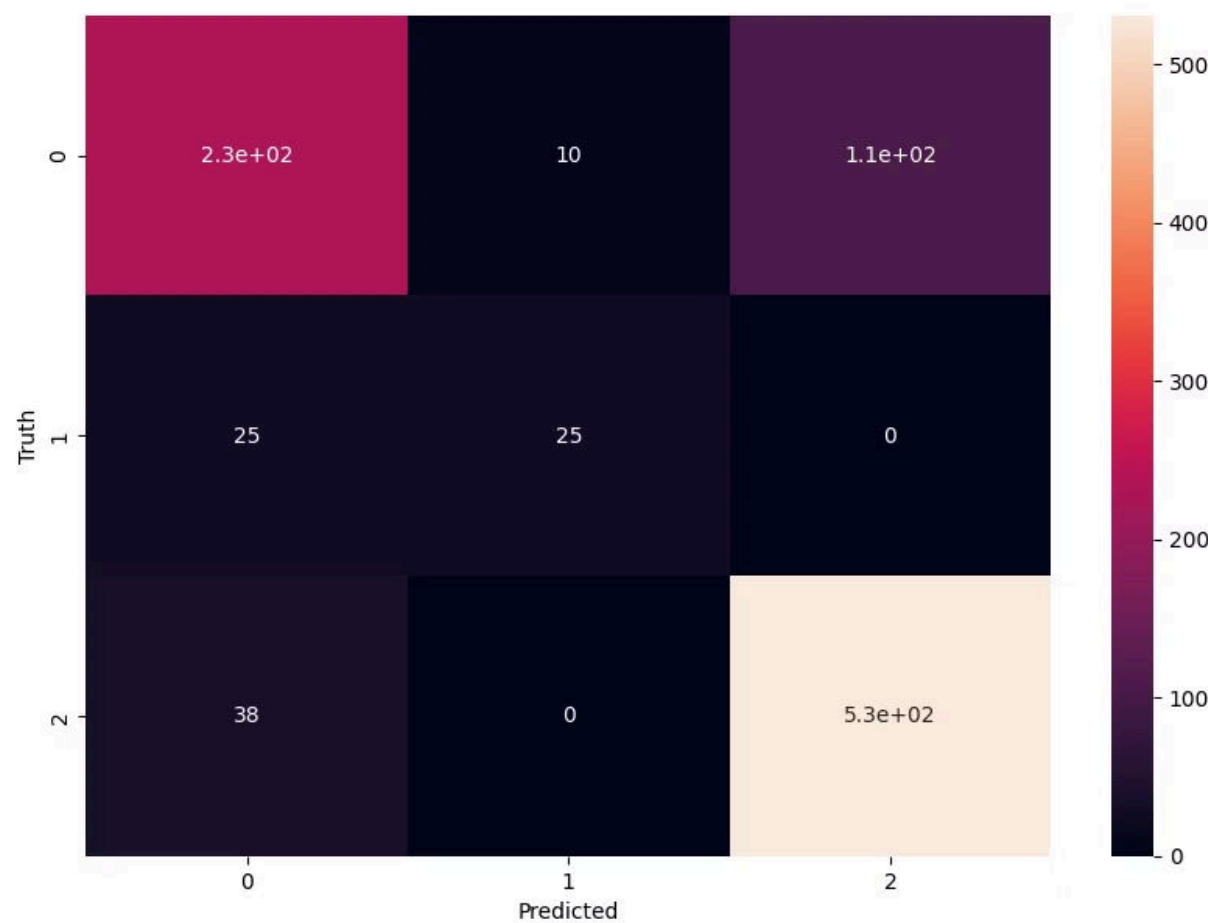
Here:-

0 → stayed

1→ churned

2→ joined

Result –

```
[ ] y_predicted[0]
```

    2

⌄ Stayed = 0

Chruned = 1

JOined = 2

Limitations of the Project

1. Limited **training data** - The dataset used for this project may not capture the full complexity of customer churn behavior, leading to potential biases or incomplete insights.

2. Difficulty in predicting long-term churn - The model's accuracy may decrease for predicting churn that occurs further in the future, as customer preferences and behaviors can change over time.

3. Inability to **account for external factors** - The model may not consider the impact of market conditions, industry trends, or other external variables that could influence customer churn.

Conclusion and Future Considerations

After conducting our customer churn analysis and implementing various predictive models, we have gained valuable insights into customer behavior and churn patterns. This information can guide strategic decision-making and help businesses take proactive measures to reduce customer churn.

In the future, we plan to further enhance our predictive models by incorporating additional data sources and exploring advanced machine learning techniques. By continuously monitoring customer churn and adapting our strategies, we aim to improve customer retention rates and drive long-term business growth.

we can proactively address issues and take appropriate actions to retain customers.

Q&A Session

?

Thank You