Titanic Survival Prediction

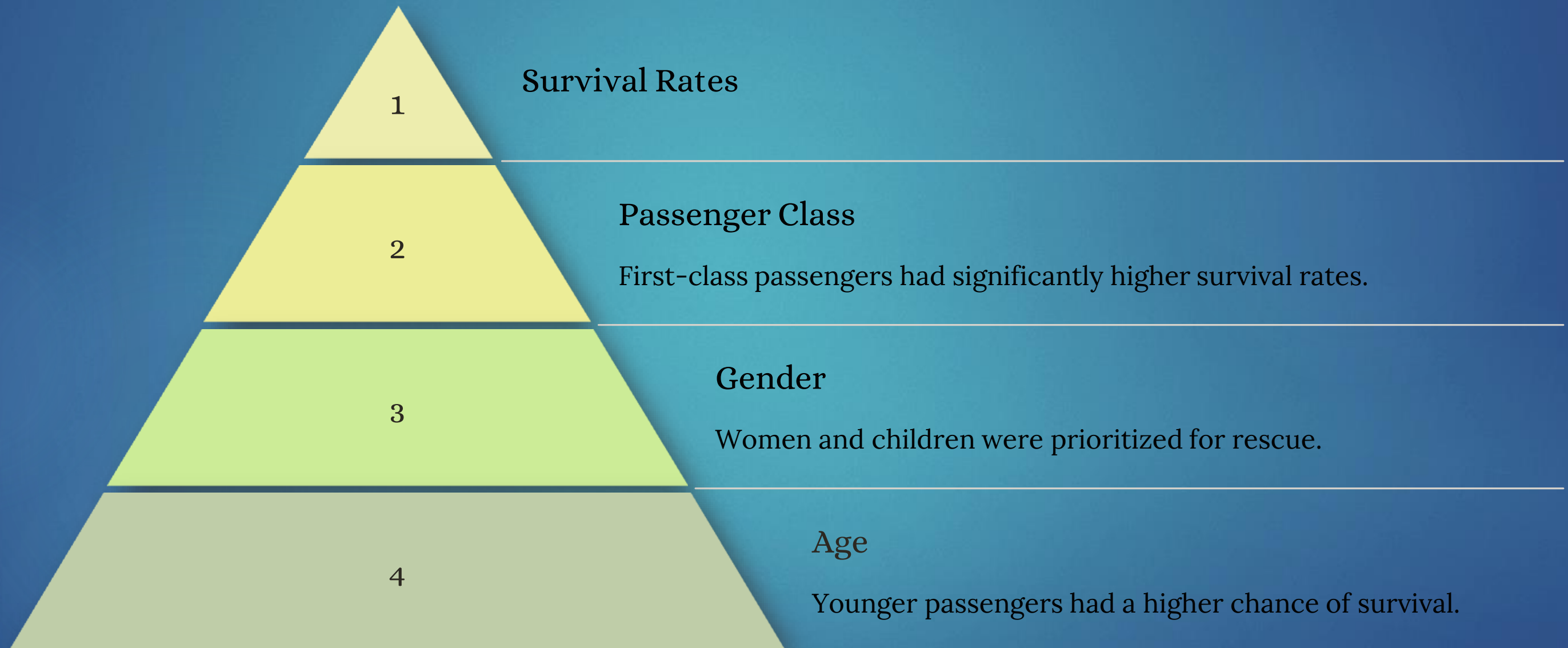PREPARED BY ANNAPURNA KALAL

Introduction to the Titanic Disaster

- The Titanic survival prediction aims to determine if a passenger survived or not based on various attributes such as age, sex, class, and more.

- The dataset contains features like gender, age, class, fare, and port of embarkation, which can be used to predict survival outcomes.

- Gender, class, and age were significant factors; for example, women and children had higher survival rates, and first-class passengers were more likely to survive.
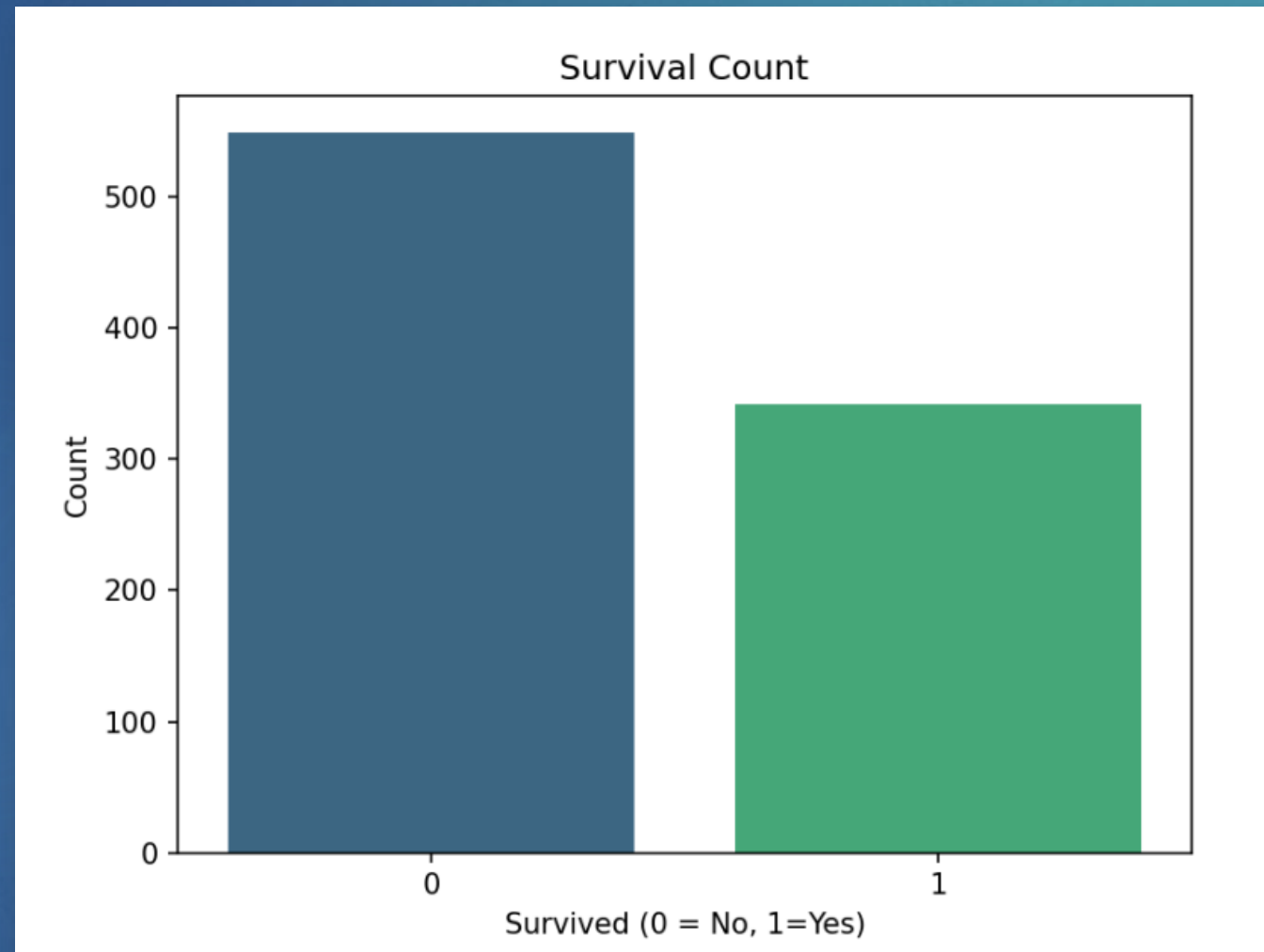
Key Features in the Dataset

1. **Passenger Class (Pclass):** 1st, 2nd, and 3rd class; higher class associated with higher survival rates.

2. **Sex:** Females had higher survival rates compared to males.

3. **Age:** Children had higher survival rates.

4. **Fare:** Higher ticket fares often correlated with higher survival rates.

5. **Embarked:** Port of embarkation; slight variations in survival rates.
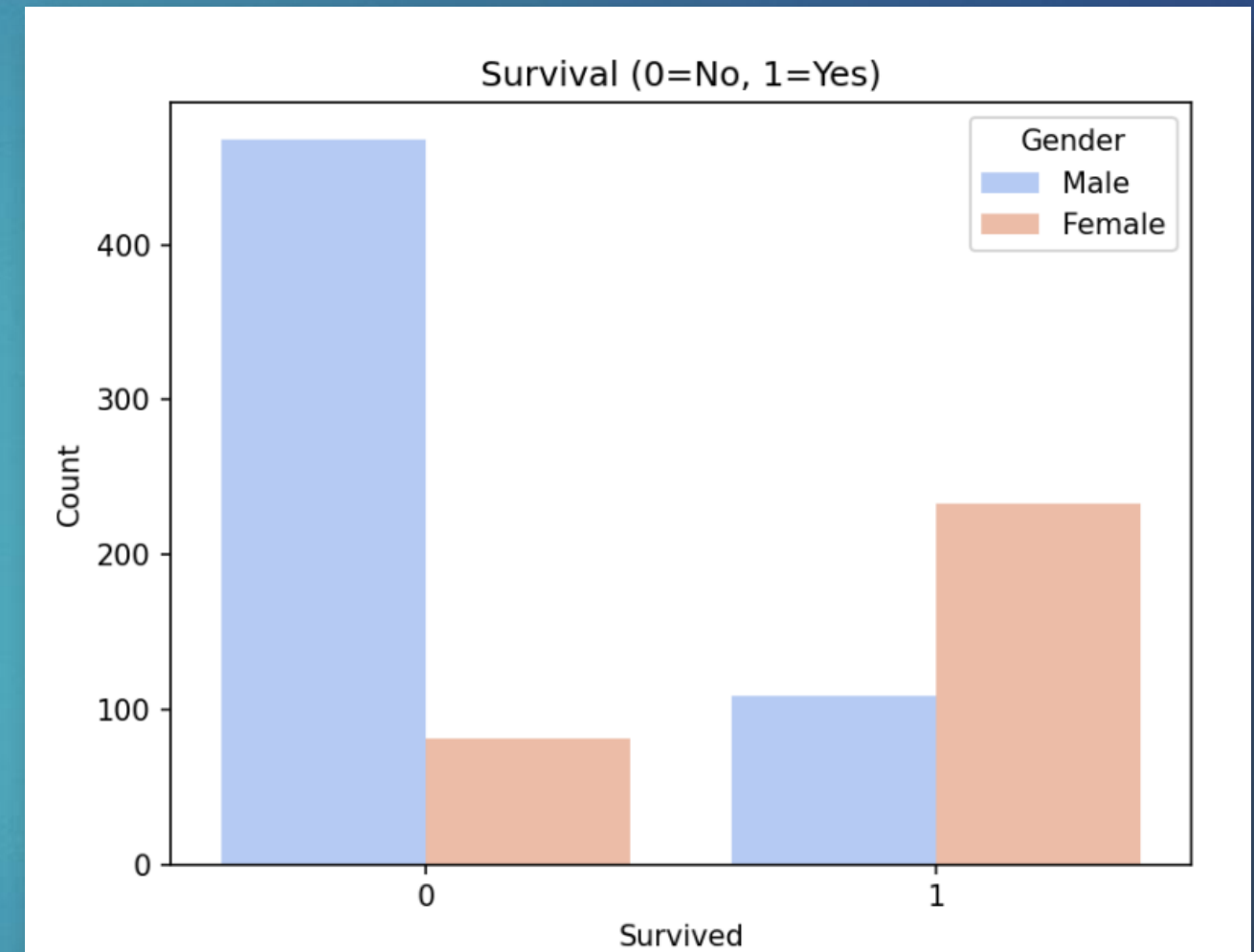
# Exploratory Data Analysis

1 — Survival Rates

2 — Passenger Class

First-class passengers had significantly higher survival rates.

3 — Gender

Women and children were prioritized for rescue.

4 — Age

Younger passengers had a higher chance of survival.

1. Survival Distribution

2. Survival by Gender

Survival Distribution shows the ratio of survivors v/s non-survivors.

Survival by Gender highlights that females had higher survival rates.

3. Age Distribution by Survival

4. Survival by Passenger Class
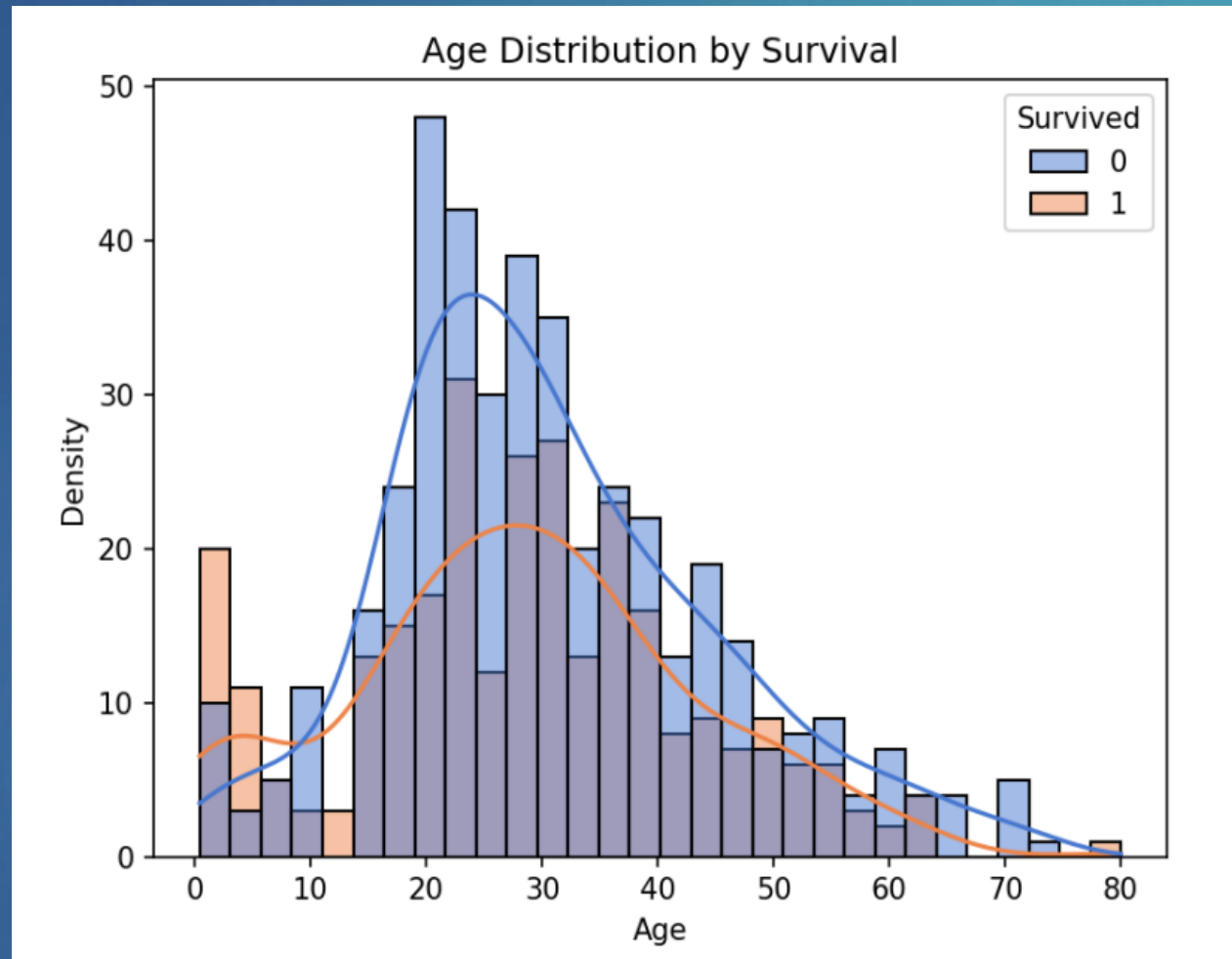
Age Distribution by Survival illustrates survival trends across different age groups.
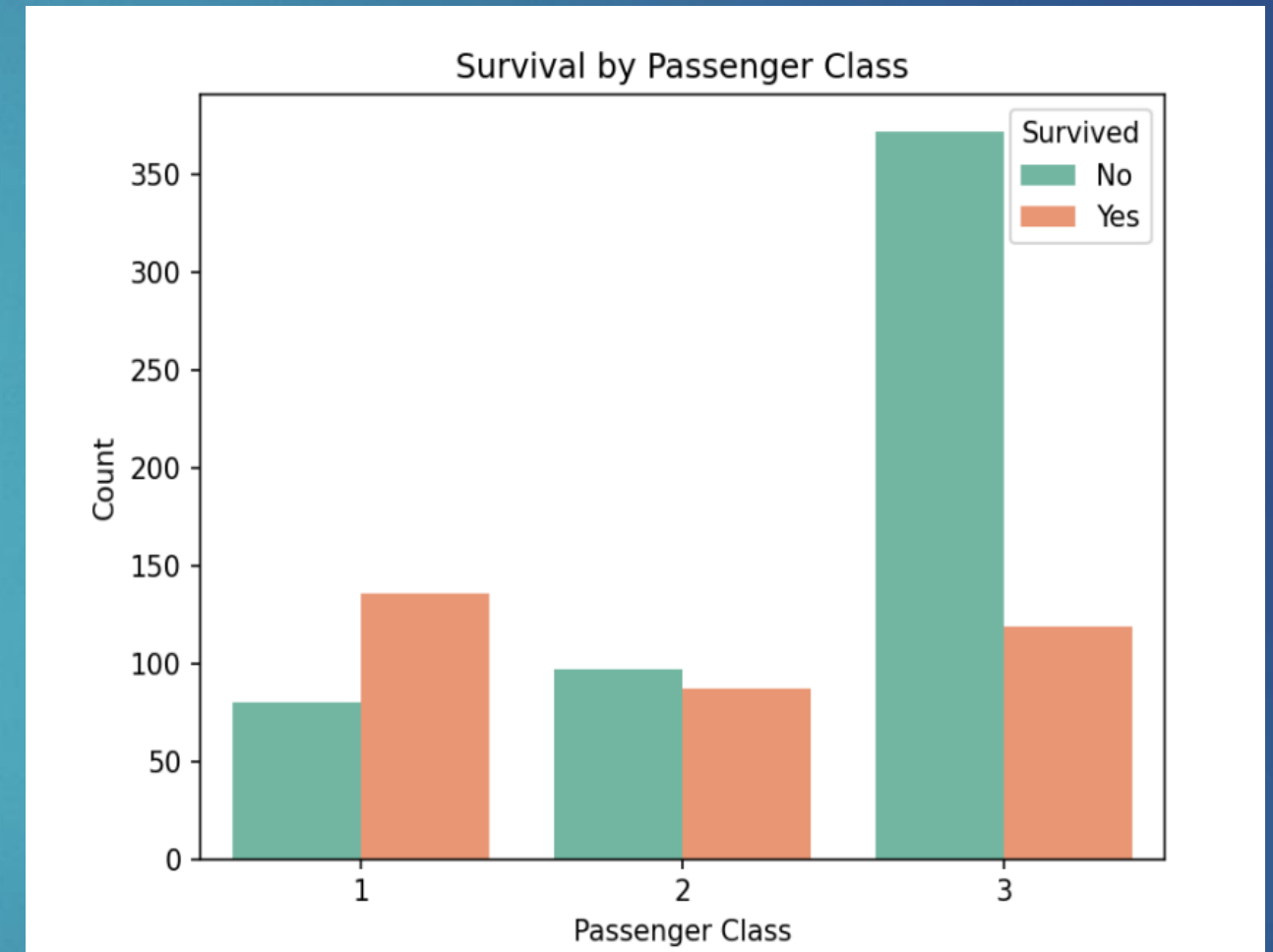
Survival by Passenger Class revels that first-class passengers had better chances of survival.

Feature Engineering

Family Size

Creating a new feature based on the number of family members on board.

A new **feature** (FamilySize) is created to indicate family connections

Machine Learning Model Selection

Logistic Regression

A simple and interpretable model for binary classification.

Decision Tree

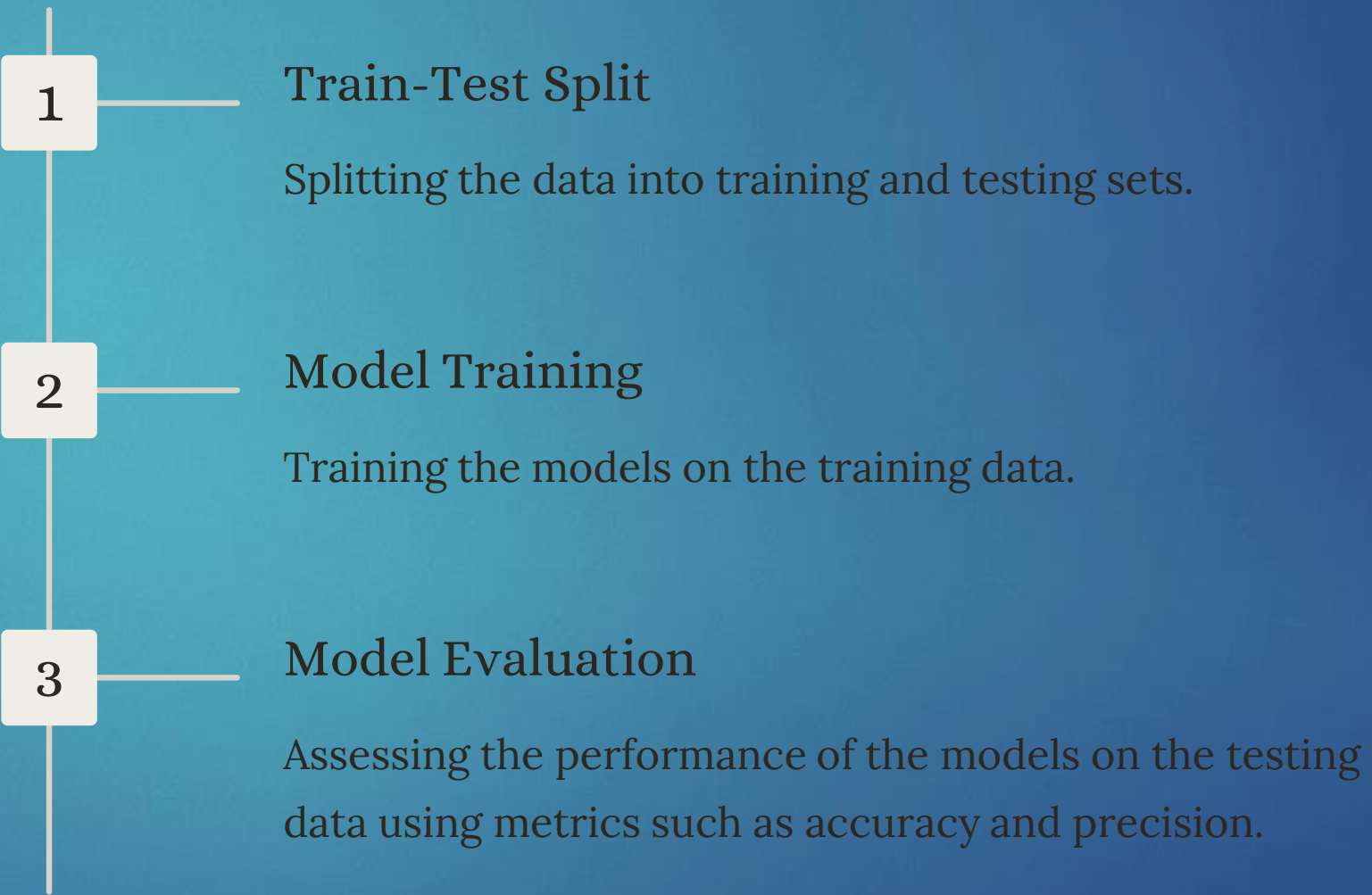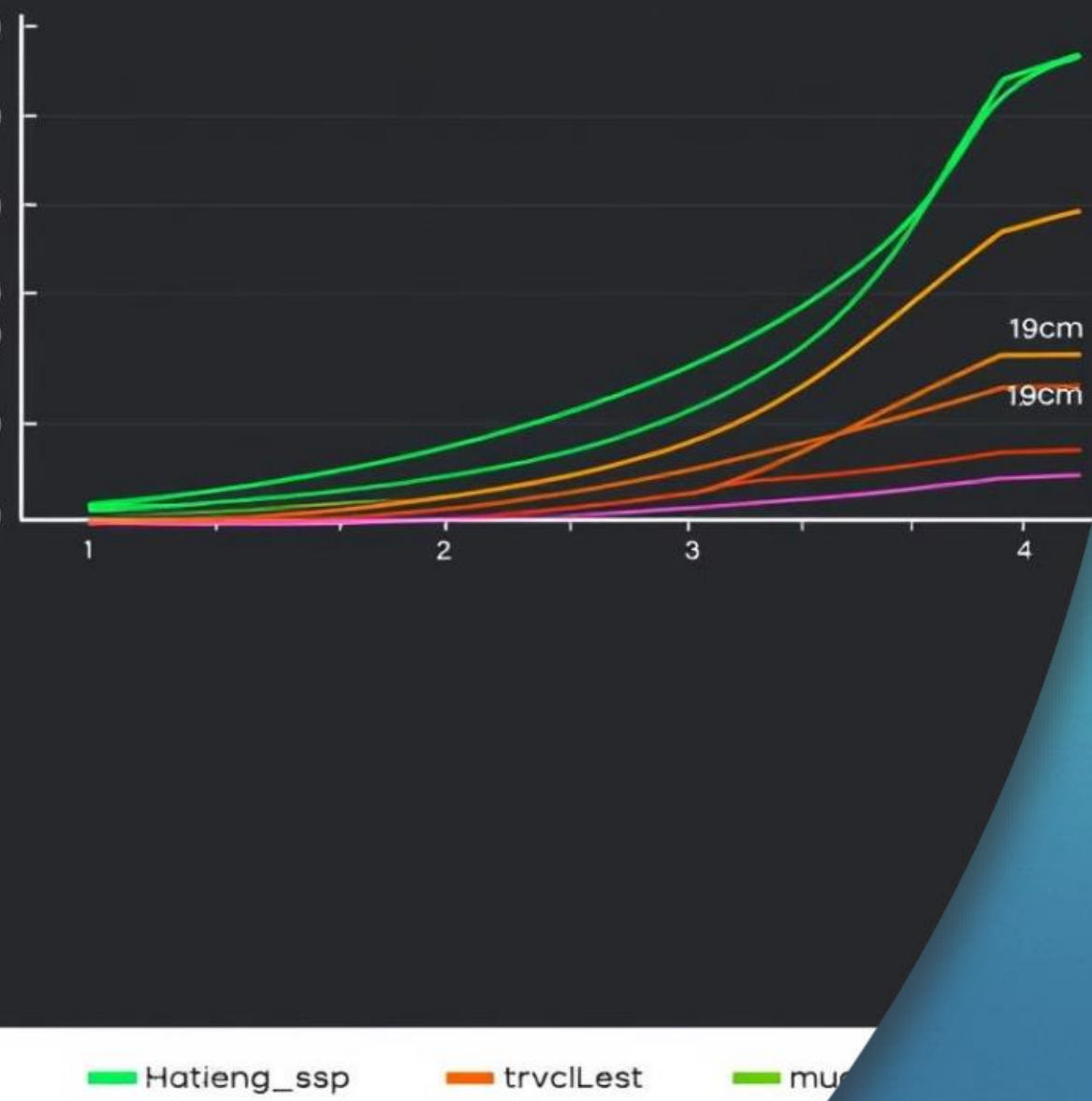A tree-based model that can handle both numerical and categorical features.

Random Forest

An ensemble method that combines multiple decision trees for improved accuracy.

Model Training and Evaluation

1 Train-Test Split

Splitting the data into training and testing sets.

2 Model Training

Training the models on the training data.

3 Model Evaluation

Assessing the performance of the models on the testing data using metrics such as accuracy and precision.

Conclusion and Key Takeaways

The analysis reveals that passenger class, gender, and age played significant roles in survival. Machine learning techniques can be applied to historical events to gain insights and predict outcomes. This project offers a glimpse into the complexities of the Titanic disaster, using data to understand its human impact.