

# Titanic Dataset - Exploratory Data Analysis (EDA)

---

## Project Overview

This project involves performing Exploratory Data Analysis (EDA) on the famous Titanic dataset to uncover patterns, clean the data, and visualize important trends that influence survival.

---

## 1. Initial Data Exploration

- **Dataset Shape:** 891 rows × 12 columns
- **Key Variables:**
  - Numerical: Age, Fare, SibSp, Parch
  - Categorical: Pclass, Sex, Embarked
- **Missing Values:**
  - Age: 177 missing
  - Cabin: 687 missing
  - Embarked: 2 missing

### Actions Taken

- Imputed missing **Age** values with median based on **Pclass**.
  - Created **Has\_Cabin** feature (binary flag for Cabin availability).
  - Filled missing **Embarked** values with the mode.
  - Extracted **Title** (e.g., Mr, Mrs) from the **Name** field.
  - Created **Family\_Size** feature by adding **SibSp** and **Parch**.
- 

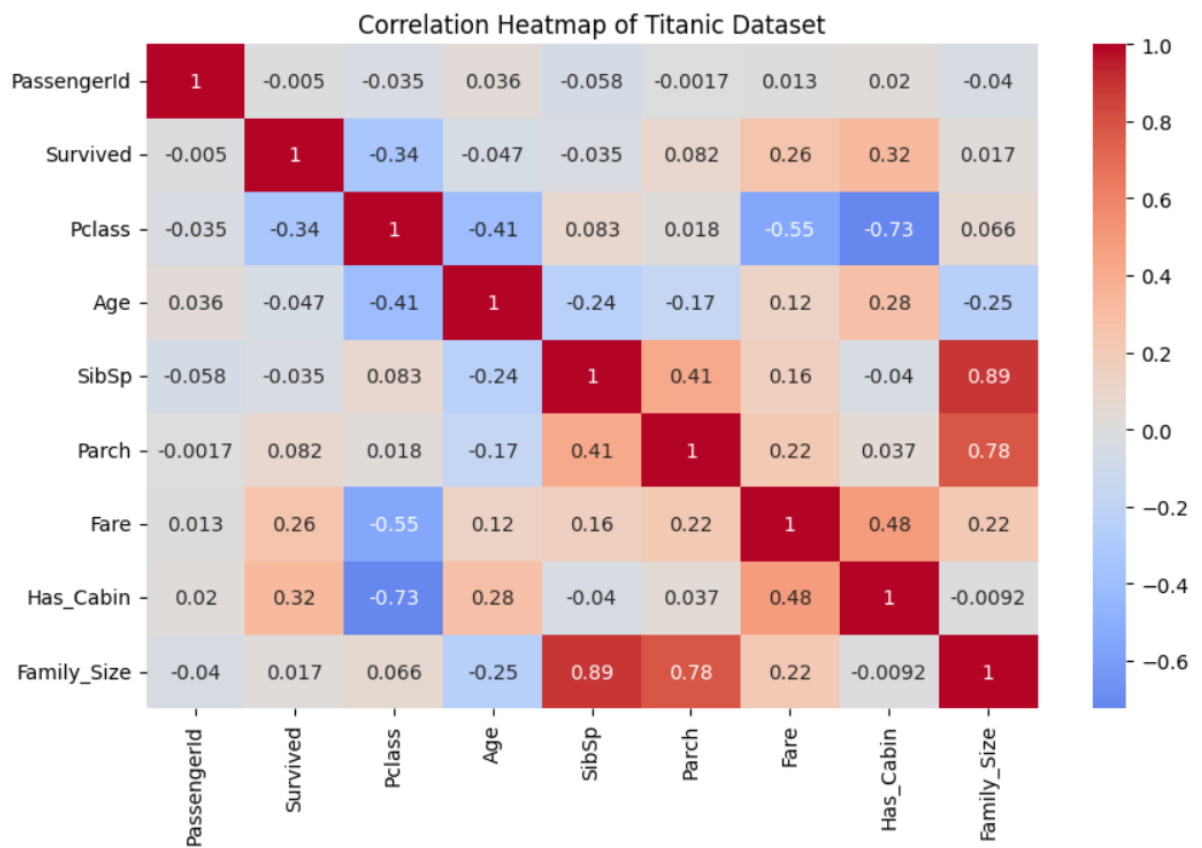
## 2. Data Visualization

### a. Pairplot for Numerical Variables

- Relationships between variables like Age, Fare, SibSp, Parch were visualized with Survived as the hue.

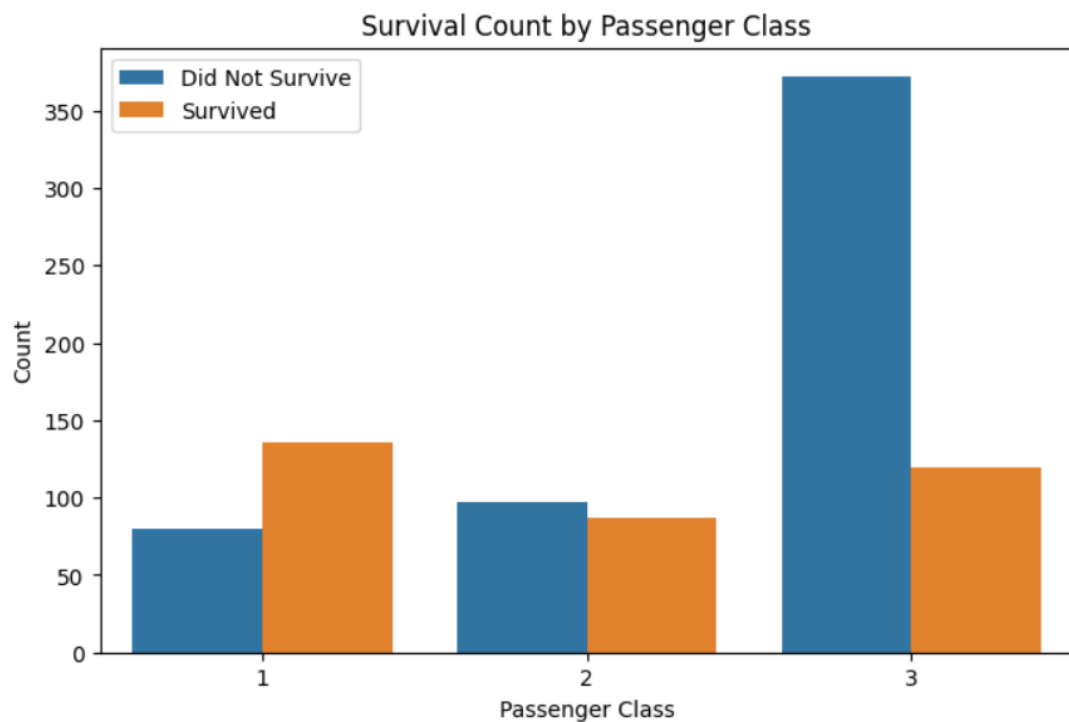
### b. Correlation Heatmap

- A heatmap was created to show correlations among numerical features.



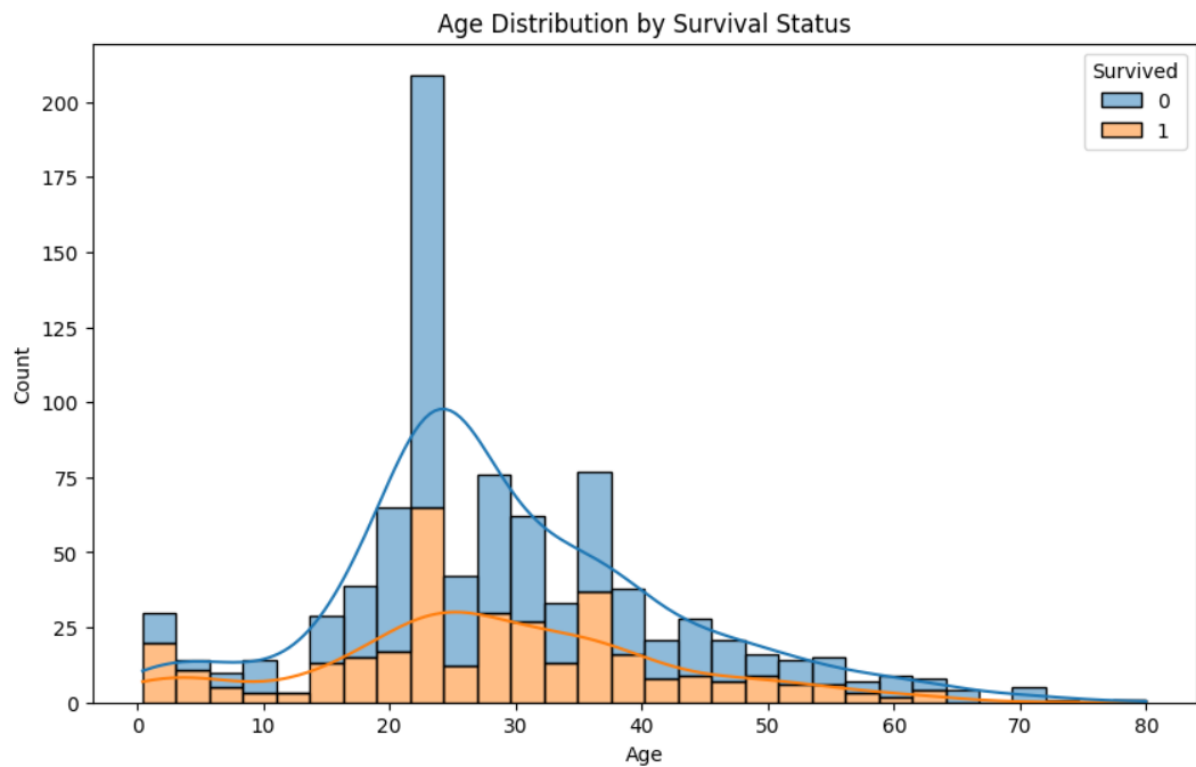
### c. Survival by Passenger Class

- Survival counts compared across different passenger classes (Pclass).



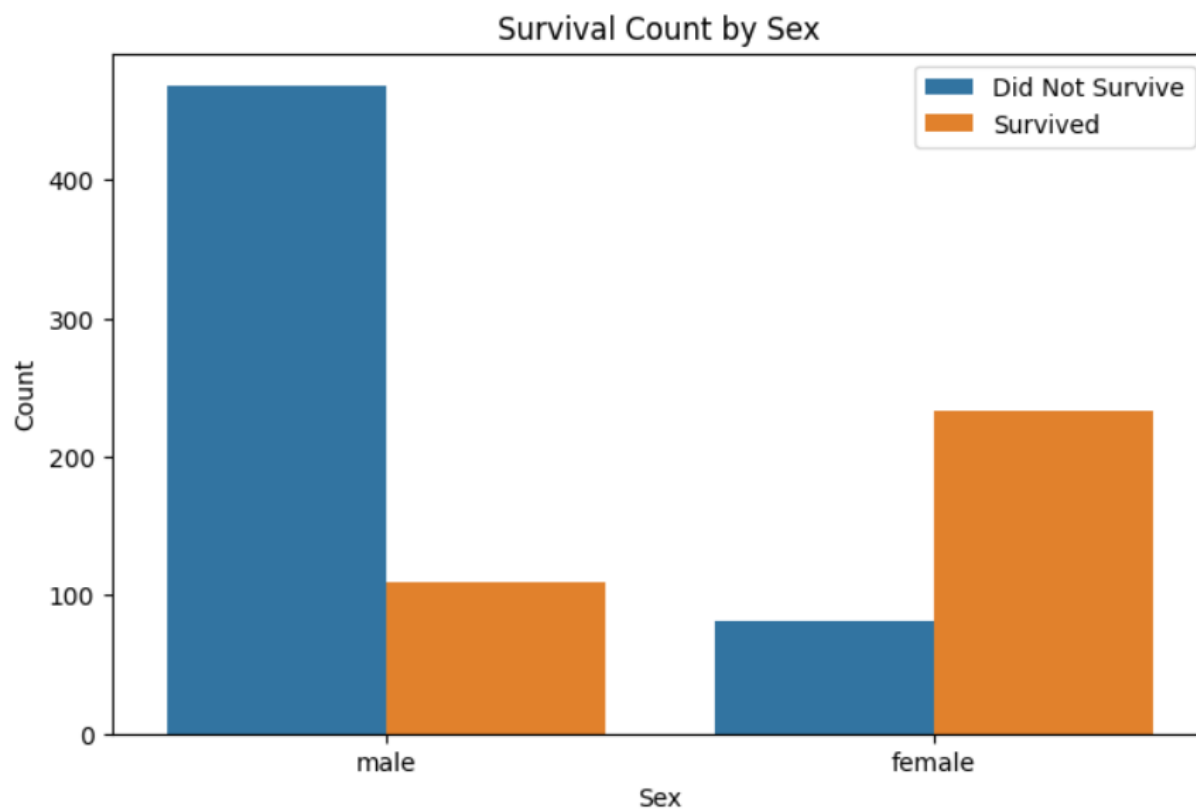
### d. Age Distribution by Survival

- Compared age distribution between survived and non-survived passengers.



#### e. Survival by Sex

- Survival counts compared across genders (male and female).



- pandas
- numpy
- seaborn
- matplotlib