Task 5: Exploratory Data Analysis (EDA)
Dataset: Titanic (train.csv)
Tools: Python (Pandas, Matplotlib, Seaborn)
Your Name: Naki Qureshi
Date: 15th Aug 2025

1. Introduction

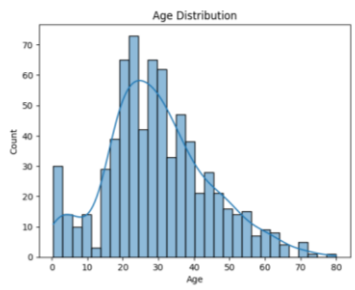
- The dataset used is the Titanic dataset (891 passengers).
- Objective: Explore data using statistical summaries and visualizations to identify patterns related to survival.

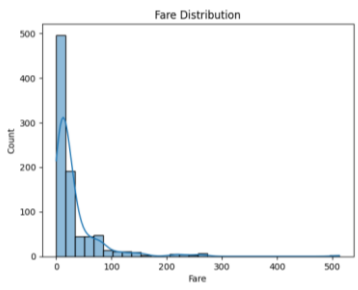
2. Dataset Overview

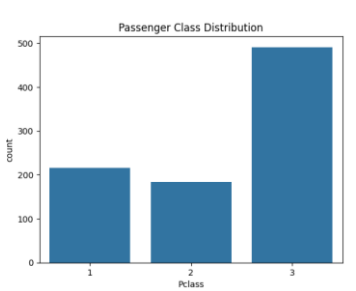
- 891 rows
- 12 columns
- PassengerId 0
- Survived 0
- Pclass 0
- Name 0
- Sex 0
- Age 177
- SibSp 0
- Parch 0
- Ticket 0
- Fare 0
- Cabin 687
- Embarked 2

3. Univariate Analysis

- Survival Rate: ~38% survived
- Passenger Class: Most passengers were in 3rd class
- Gender Distribution: More males (~65%) than females (~35%).
- Age Distribution: Most passengers aged between 20–40.
- Fare: Right-skewed — a few very high fares.

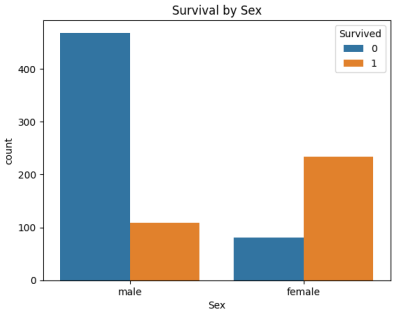


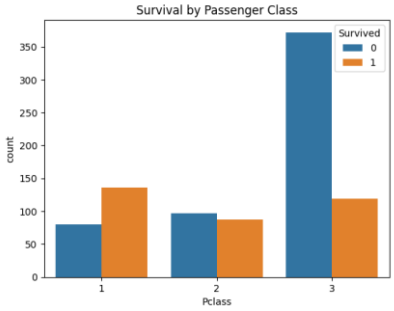


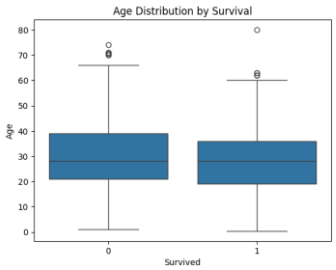


4. Bivariate Analysis

- Survival by Gender:
  - Females had much higher survival rate than males.
- Survival by Class:
  - 1st class had highest survival, 3rd class lowest.
- Survival by Age:
  - Children (<10 years) had higher chances of survival.
- Fare vs Survival:
  - Higher fares linked to higher survival (wealthier passengers in 1st/2nd class).







5. Multivariate Analysis

- Pclass + Sex + Survival:
  - Female 1st class had highest survival (~95%).
  - Male 3rd class had lowest survival (~15%).
- Family (SibSp + Parch):
  - Passengers with small families (1–2 members) had better survival chances.

6. Heatmap (Correlation)

- Positive correlation between Fare and Survived.
- Negative correlation between Pclass and Survived.
- Weak correlation of Age with survival.