EDA - Findings

DATASET : Titanic Survival Dataset

1. Import Libraries :

Import all required libraries such as pandas, numpy, seaborn, matplotlib.pyplot

2. Load Dataset :

Load dataset i.e csv file using pandas
pd.read_csv('Titanic_cleaned.csv')

3. Overview :

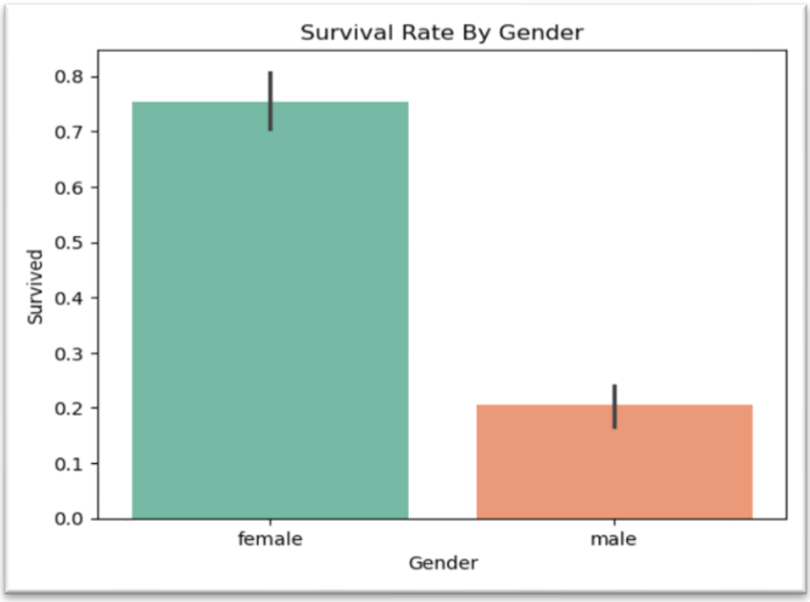
- **df.info() =>** DATATYPES AND NULL COUNT FOR EACH COLUMNS
- **df.describe() =>** STATISTICAL SUMMARY
- **df.columns =>** COLUMNS IN DATASET
- **df.shape =>** ROWS AND COLUMNS

4. Find Missing/Null values :

- **df.isnull() =>** Find null values in each columns (TRUE/FALSE)
- **df.isnull().sum() =>** Total null values in each columns

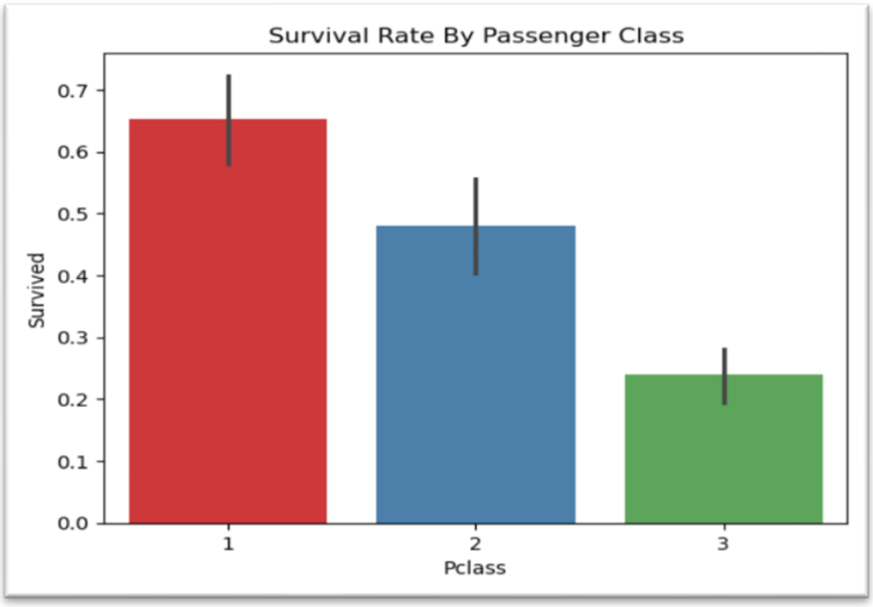
5. Data Visualization :

- **Bar Plot – Survival By Gender**



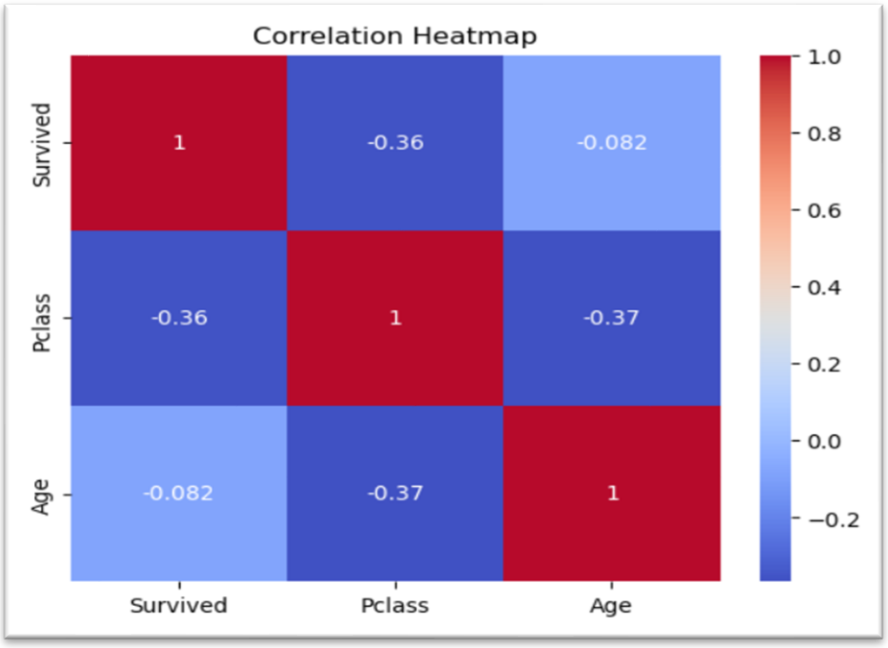
Females had a much higher survival rate (74%) compared to males (18%), supporting the "women and children first" evacuation policy.

- **Bar Plot – Survival By Passenger Class**



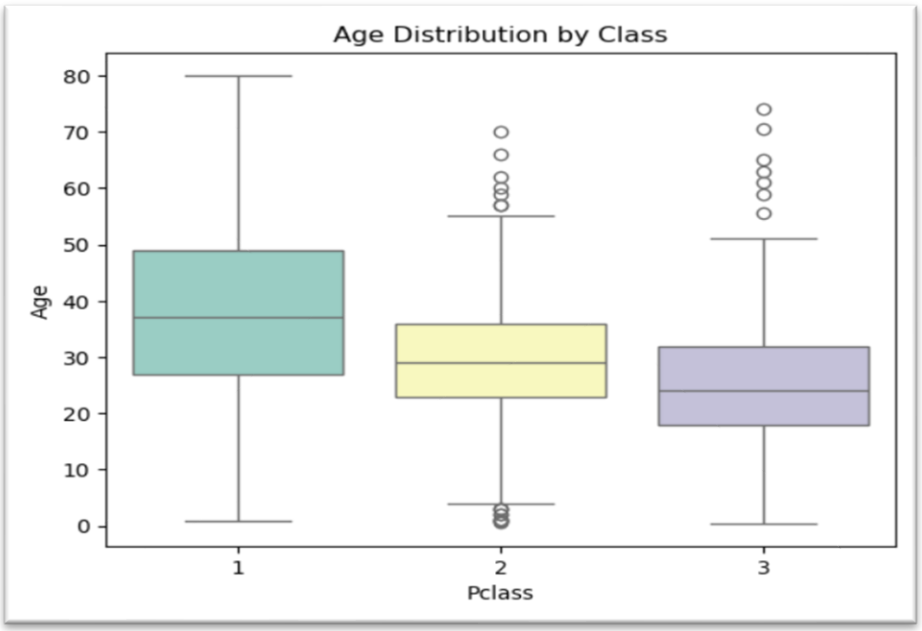
1st class passengers had the highest survival rate (63%), followed by 2nd (47%), and 3rd (24%).

- **Heatmap – Correlation Analysis**



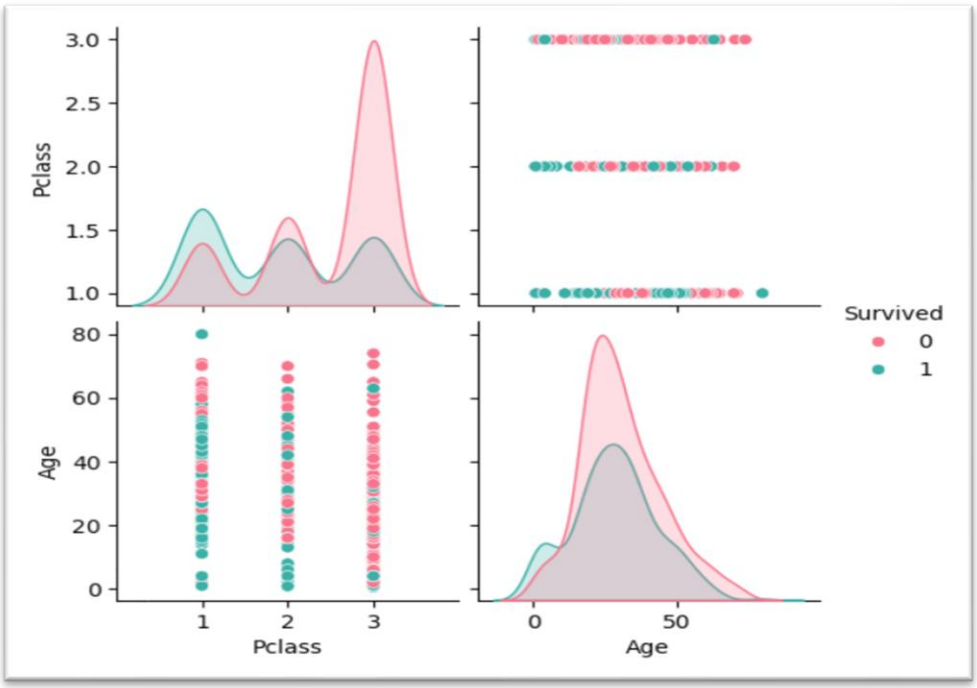
i. **Pclass has a moderate negative correlation with Survived (~-0.34), meaning higher classes had better survival chances.**
ii. **Age has a weak negative correlation with Survived.**

- **Box Plot – Age Distribution By Class**



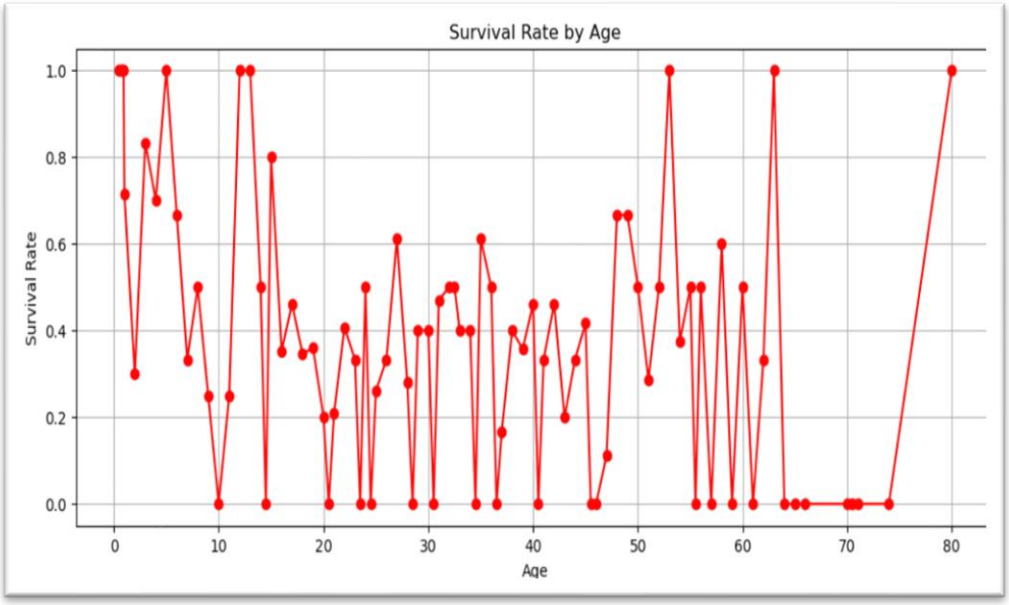
i. **Median age increases from 3rd class (23 years) to 1st class (37 years).**
ii. **Younger passengers were more common in lower classes.**

- **Pair Plot – Feature (Survived, Pclass, Age) Analysis**



i. **Survivors are concentrated in lower Pclass.**
ii. **Many younger passengers survived, especially in 1st and 2nd class.**

- **Line Chart – Survival Rate By Age**



Very young passengers (children under 10) had a higher survival rate, which decreases in middle age, then slightly fluctuates in older ages