📄 Exploratory Data Analysis (EDA) Report - Titanic Dataset

---

1. Introduction

This report presents an exploratory data analysis (EDA) of the Titanic dataset to identify patterns, trends, and anomalies that could help in predictive modelling or deeper insights into survival rates.

---

2. Dataset Overview

- **Rows:** 891
- **Columns:** 12
- **Target Variable:** Survived (0 = No, 1 = Yes)

Main features:

- Passenger Class (Pclass)
- Sex
- Age
- Fare
- Embarked
- Family-related features (Sib-Sp, Parch)

---

3. Initial Observations

- Some missing values in:
    - **Age** (~19%)
    - **Cabin** (many missing)
    - **Embarked** (2 missing)
- 'Sex', 'Pclass', and 'Fare' appear to have strong connections to survival.

---

4. Key Statistics

plaintext

Copy-edit

- Mean Age: ~29.7 years

- Mean Fare: ~32.2

- Survival Rate: ~38%

- Most passengers are from 3rd class.

- Males are more than females.

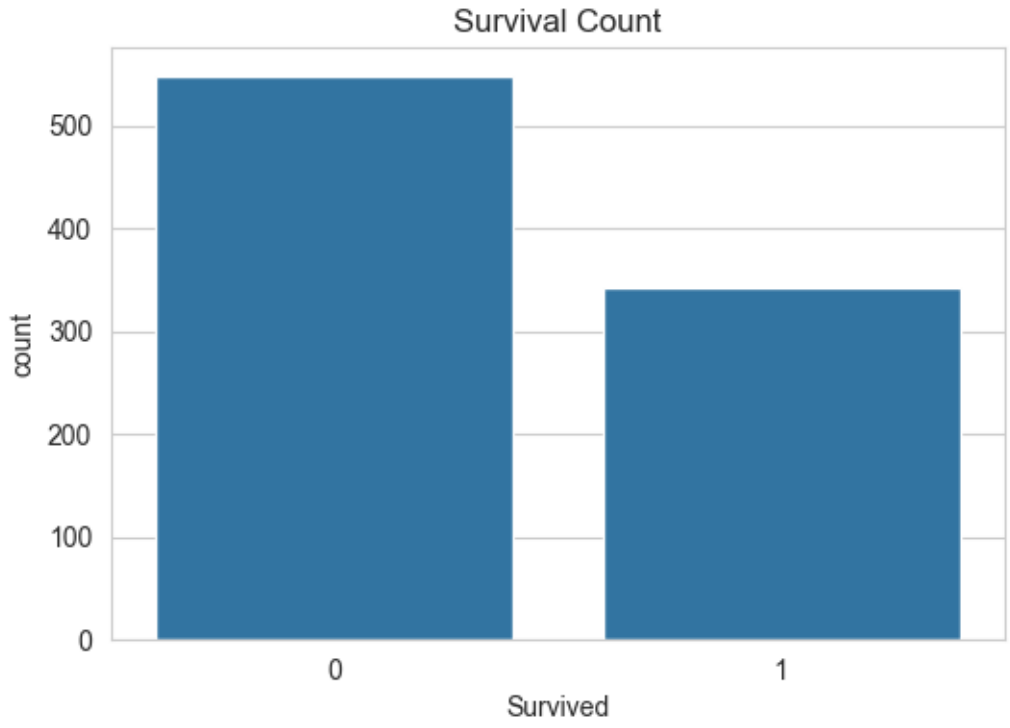
---

5. Visual Explorations

5.1 Survival Count

Observation:

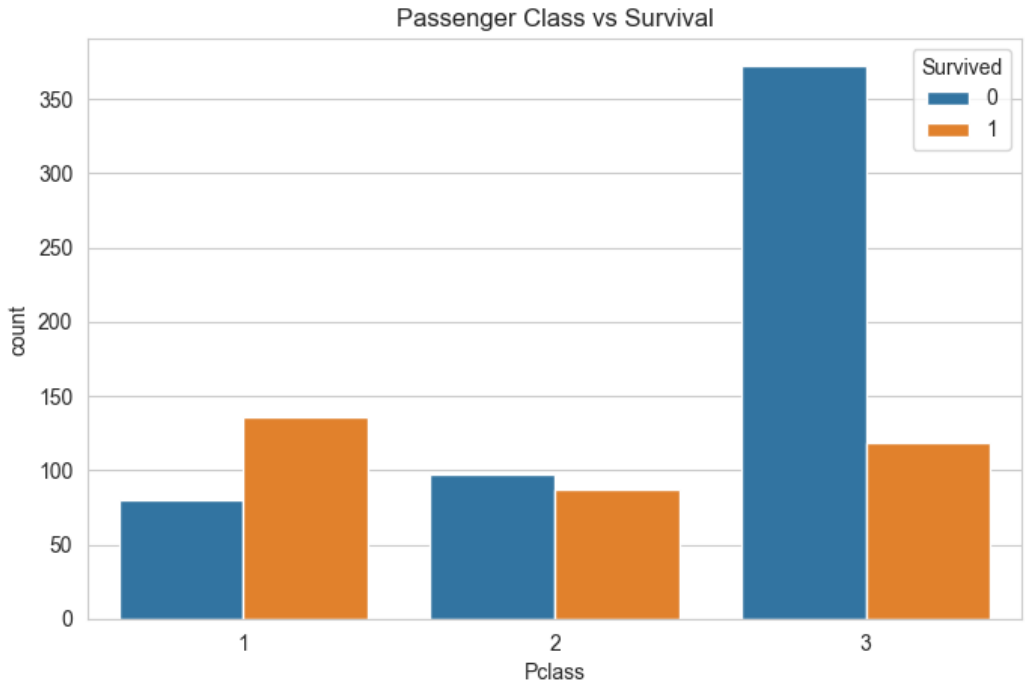
- Fewer passengers survived than died.



5.2 Passenger Class vs Survival

Observation:

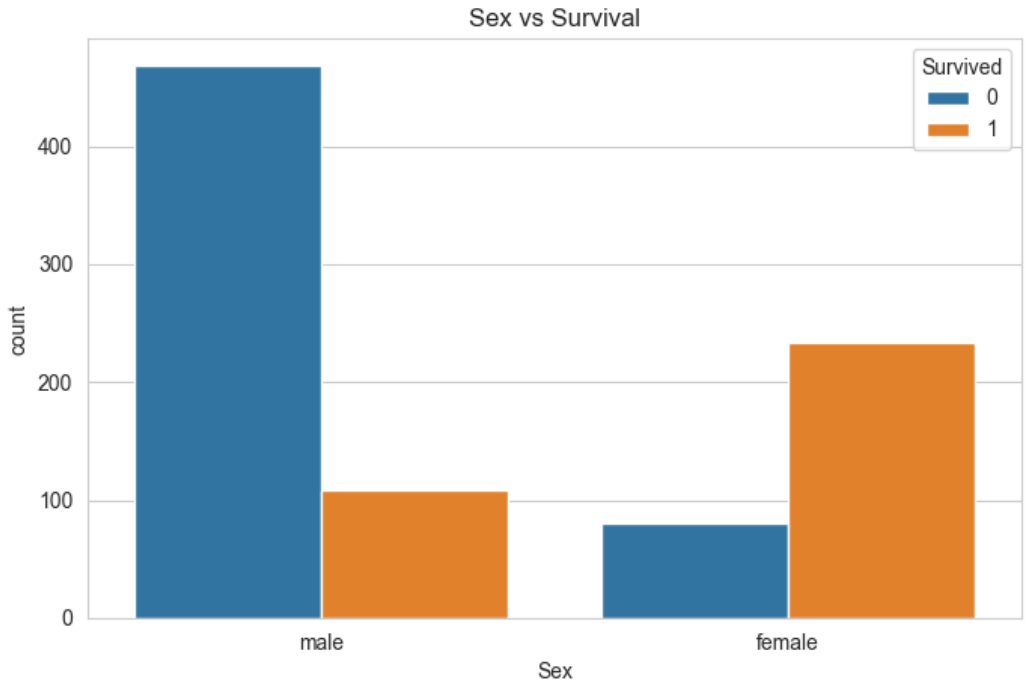
- 1st class passengers had the highest survival rates.



5.3 Sex vs Survival

Observation:

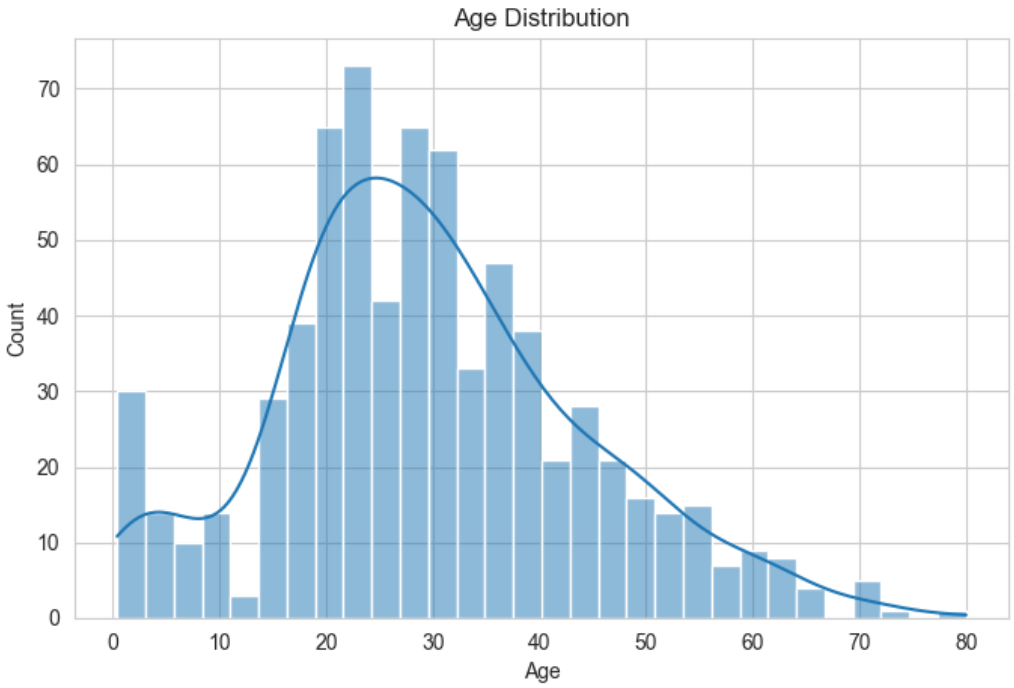
- Females had a much higher chance of survival compared to males.



5.4 Age Distribution

Observation:

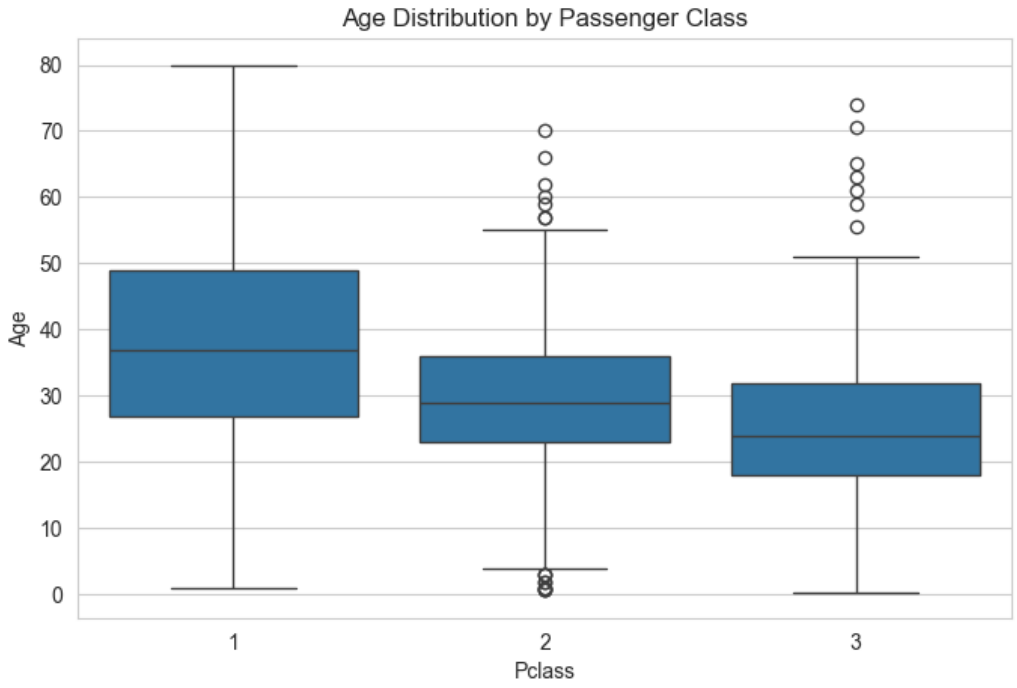
- Many passengers were young adults.
- Some children also onboarded.
- A slight positive survival chance for younger ages.



5.5 Age vs Passenger Class

Observation:

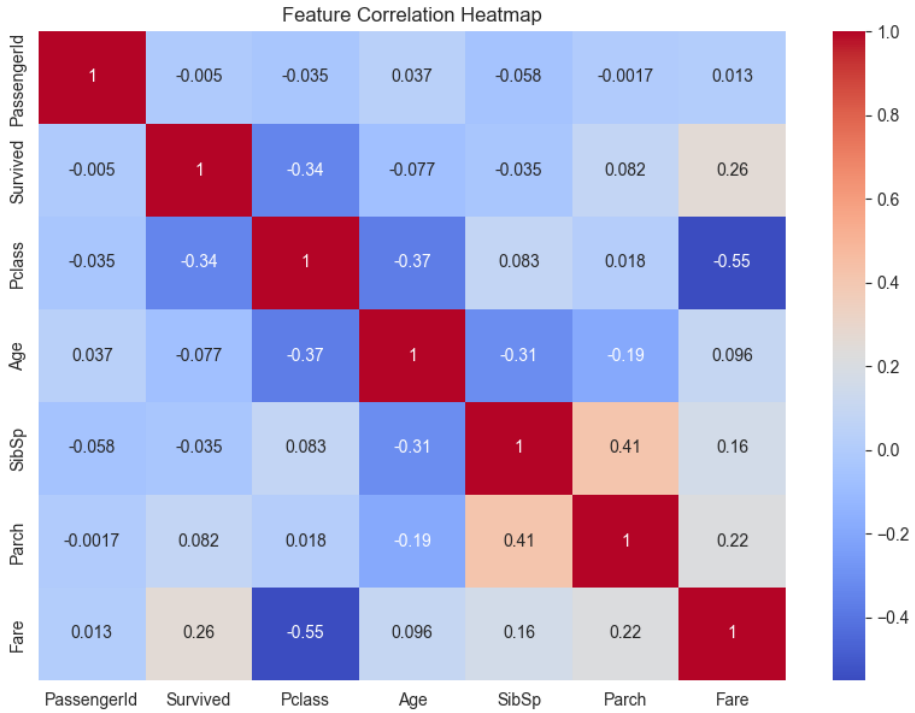
- Younger passengers were mostly in 3rd class.
- 1st class passengers are generally older.



5.6 Correlation Heatmap

Observation:

- Fare and Pclass are correlated.
- Sex is strongly correlated with survival.

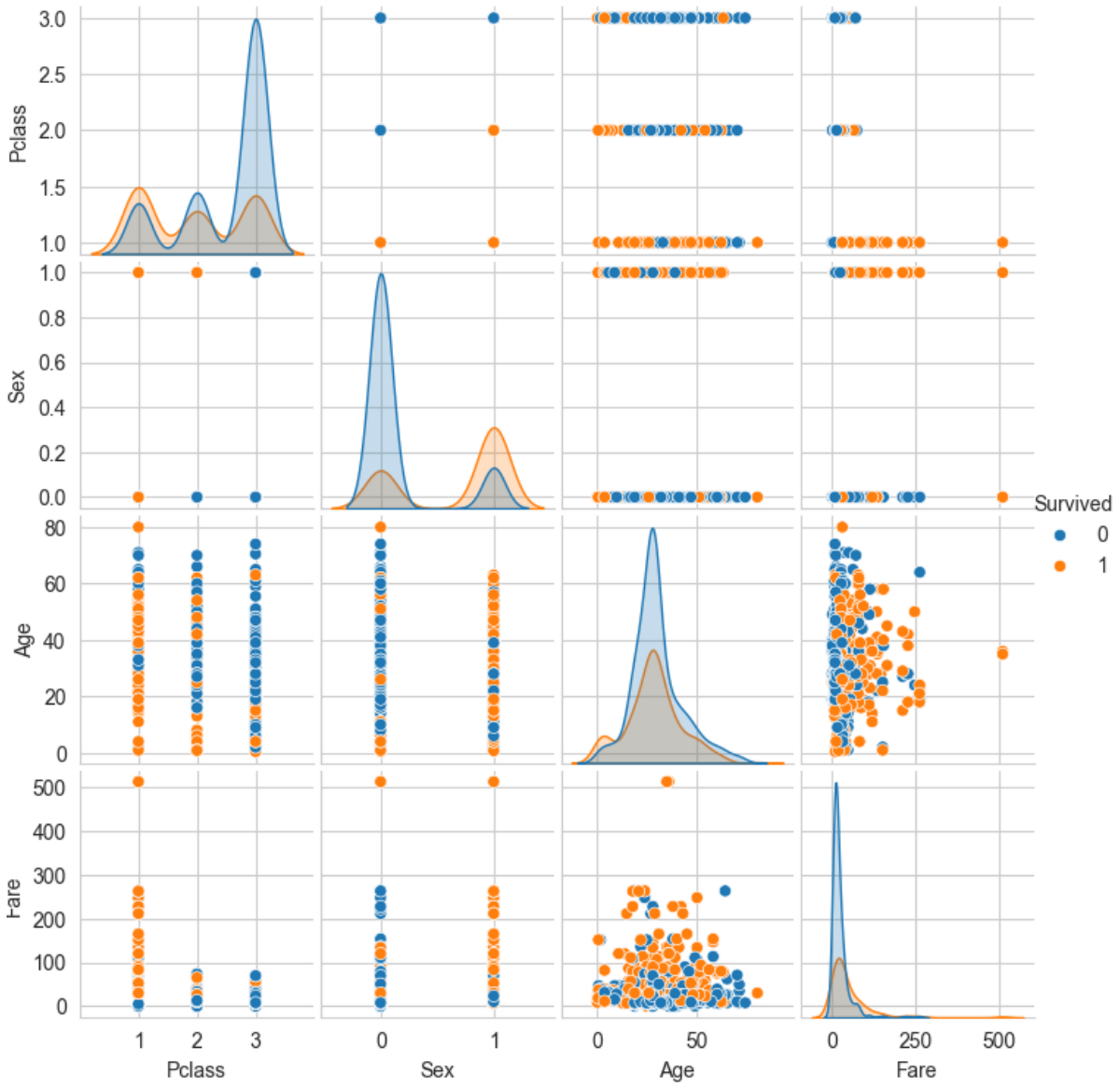


-

5.7 Pair plot (Selected Features)

Observation:

- Clear separation between survivors and non-survivors by Sex, Pclass, and Fare.



---

6. Conclusion and Summary of Findings

- Females and 1st class passengers had a significantly higher survival rate.
- Younger age showed slightly better survival chances.
- High Fare is slightly positively associated with survival.
- Missing data in Age and Cabin should be handled before modelling.