Exploratory Data Analysis (EDA) of the Titanic Dataset

Objective: Extract insights using visual and statistical exploration to identify relationships, trends, and anomalies.

1. Basic Data Exploration

1.1 .info()
Observation:
- The dataset contains 891 entries with 15 columns.
- Missing values detected in:
  - age (177 missing)
  - deck (688 missing)
  - embark_town (2 missing)
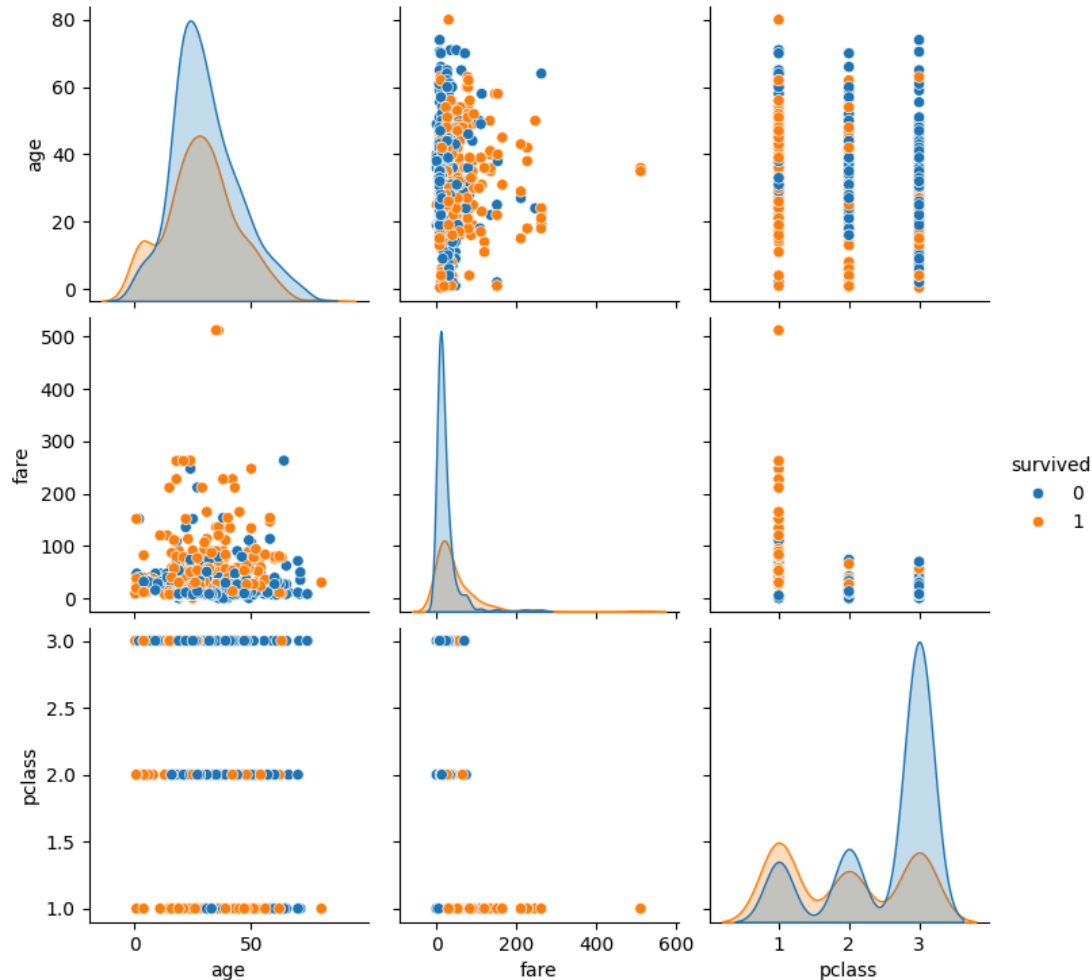- Data types: Mostly numeric (int, float) and categorical (object).

1.2 .value_counts()
Key Observations:
- Gender:
  - Male: 577
  - Female: 314
- Class:
  - 3rd Class: 491
  - 1st Class: 216
  - 2nd Class: 184
- Survival:
  - Died: 549
  - Survived: 342

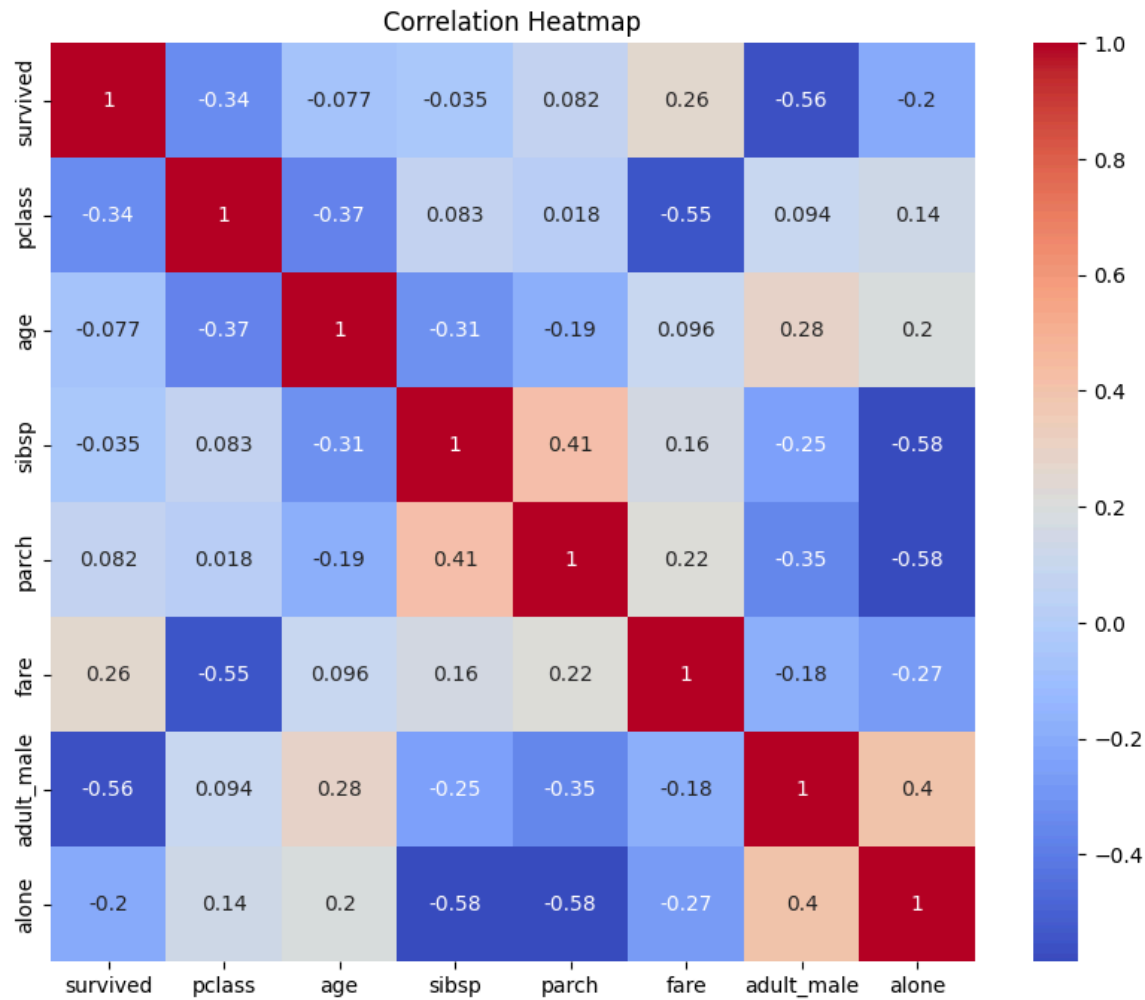## 2. Data Visualizations & Observations

2.1 Pairplot (sns.pairplot())

Observations:
- Age vs. Survival:
  - Higher survival rates for children (<10 yrs).
- Fare vs. Survival:
  - Passengers who paid higher fares (1st class) survived more.
- Pclass vs. Survival:
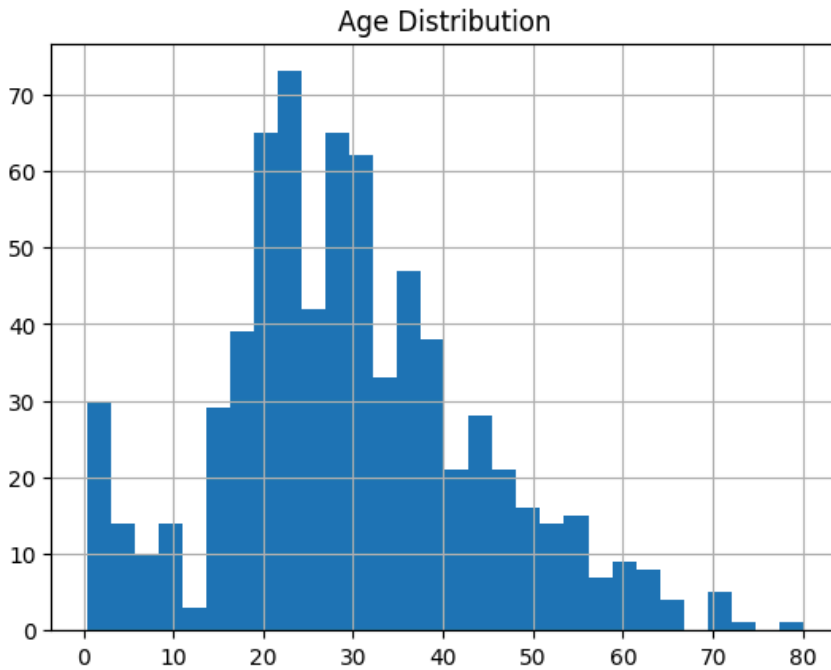 - 1st class had the highest survival rate

2.2 Heatmap (sns.heatmap())

Correlation Heatmap

Key Correlations:

| Feature Pair	Correlation	Interpretation
pclass & fare	-0.55	Higher class (1st) = Higher fare
fare & survived	0.26	Higher fare = Better survival
pclass & survived	-0.34	Lower class (3rd) = Lower survival

2.3 Histogram (Age Distribution)

Age Distribution

Observation:

- Most passengers were 20-40 years old.
- Few children (<10) and elderly (>60).

2.4 Boxplot (Age by Class)

Age Distribution by Passenger Class
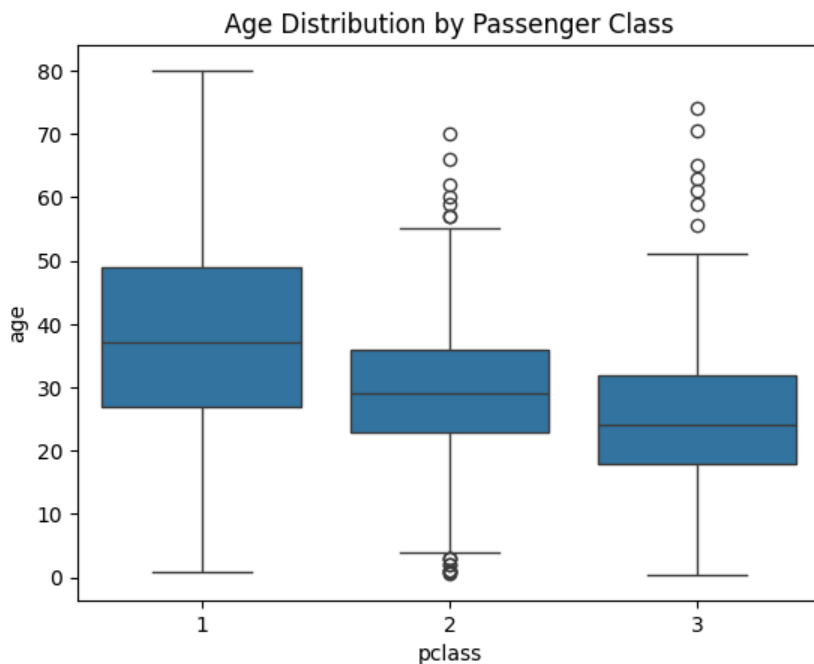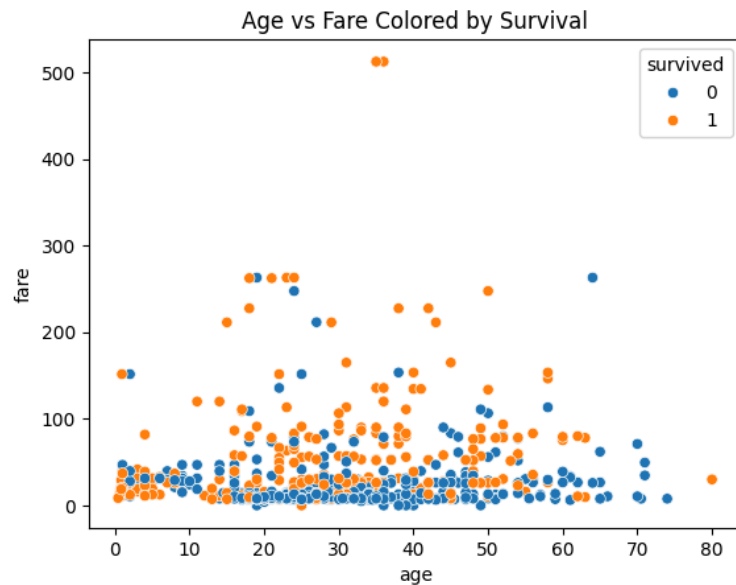
Observation:
- 1st class passengers were generally older (median ~37).
- 3rd class had more young passengers (median ~24).
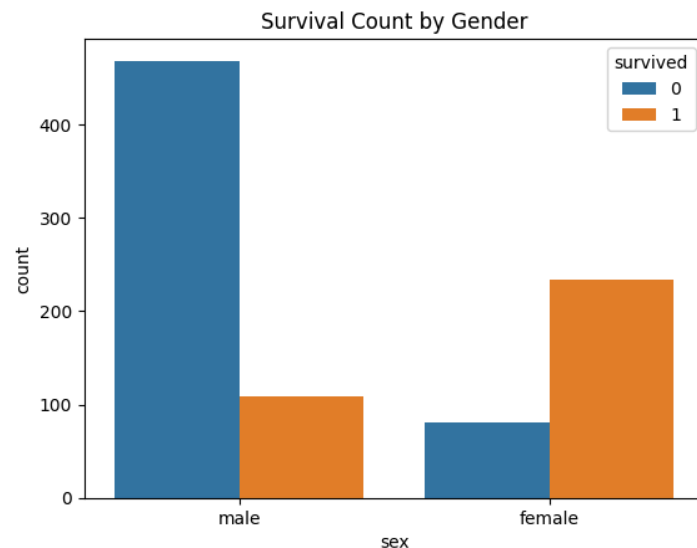- Outliers: A few very old passengers (>70) in all classes.

2.5 Scatterplot (Age vs. Fare, Colored by Survival)

Age vs Fare Colored by Survival

Observation:
- Higher fare = Higher survival (mostly 1st class).
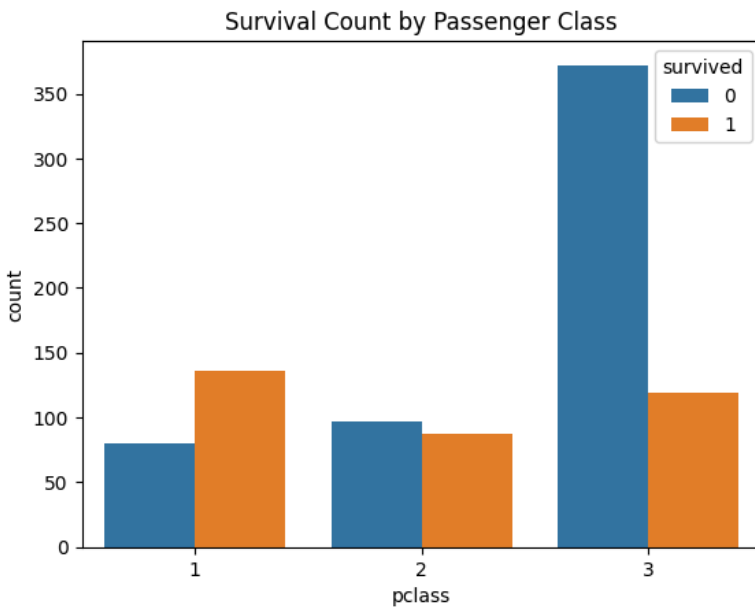- Children (<10) survived even with lower fares.

2.6 Countplot (Survival by Gender)

Survival Count by Gender

Observation:
- 74% of females survived vs. only 19% of males.
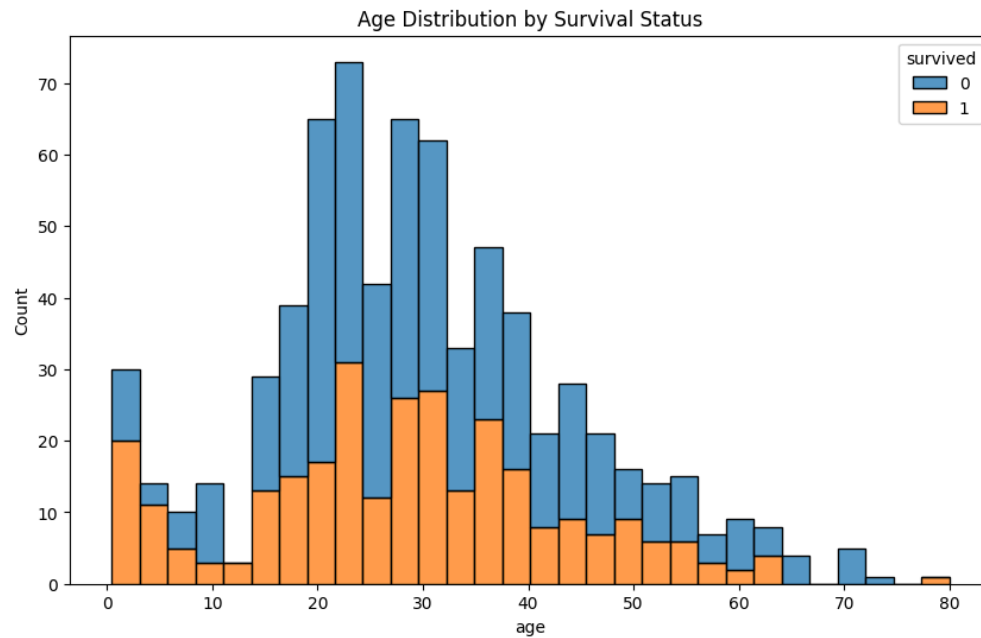- Strong "women and children first" policy effect.

2.7 Countplot (Survival by Class)

Survival Count by Passenger Class

Observation:
- 1st class: 63% survived
- 2nd class: 47% survived
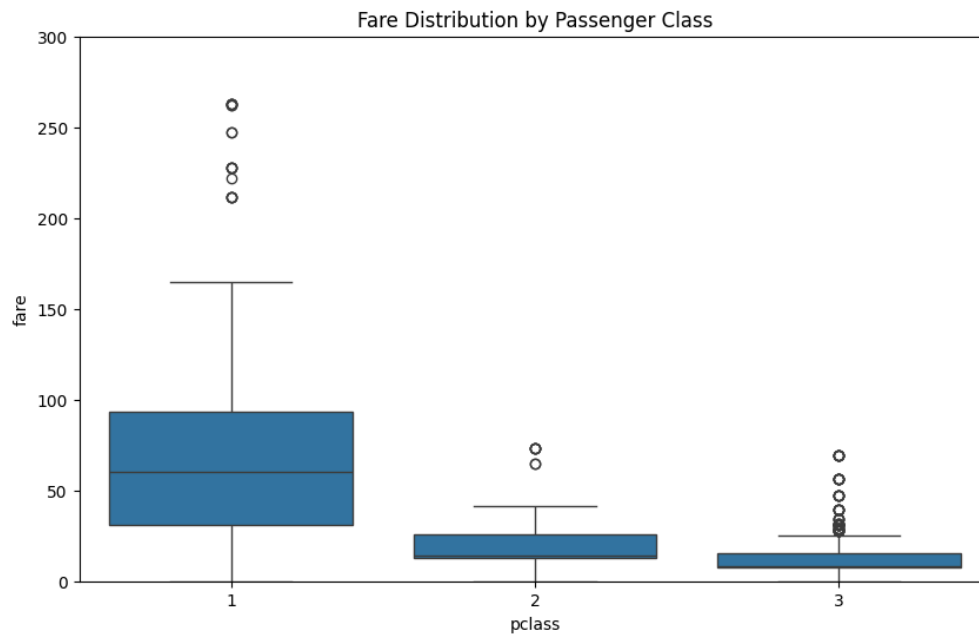- 3rd class: 24% survived
- Class was a major survival factor.

2.8 Stacked Histogram (Age by Survival)

Age Distribution by Survival Status

Observation:

- Children (<10) had the highest survival rate.
- Peak deaths in 20-40 age group (most passengers).

2.9 Boxplot (Fare by Class)

Fare Distribution by Passenger Class

Observation:
- 1st class fares were significantly higher.
- Outliers: Some 2nd/3rd class passengers paid very high fares.

3. Summary of Findings

Key Insights:
✅ Survival Rate: Only 38% survived, with major differences by:
- Gender: 74% females vs. 19% males
- Class: 63% in 1st class vs. 24% in 3rd class
- Age: Children (<10) had the highest survival rate

✅ Passenger Demographics:
- 65% male, 55% in 3rd class, avg. age ~30 yrs

✅ Fare Impact:
- Higher fare = Better survival (linked to class privilege)

✅ Anomalies:
- Some very old passengers (>70)
- A few 3rd-class passengers paid extremely high fares

Conclusion:
The Titanic disaster followed a "women and children first" policy, but class privilege (wealth) played a major role in survival. 1st-class passengers had the highest survival rates, while 3rd-class males had the worst odds.