TITANIC EDA REPORT

1. Dataset Info

- Total entries: 891
- Total columns: 12
- Data types:
  - Numerical: int64, float64
  - Categorical: object
- Missing values detected in:
  - Age (177 missing)
  - Cabin (687 missing)
  - Embarked (2 missing)

2. Summary Statistics

Numerical Features:
- Age: Mean = 29.7, Min = 0.42, Max = 80
- Fare: Mean = 32.2, Max = 512.33
- SibSp (Siblings/Spouses Aboard): Max = 8
- Parch (Parents/Children Aboard): Max = 6

Categorical Features:
- Sex: Male (577), Female (314)
- Embarked: Southampton (644), Cherbourg (168), Queenstown (77)
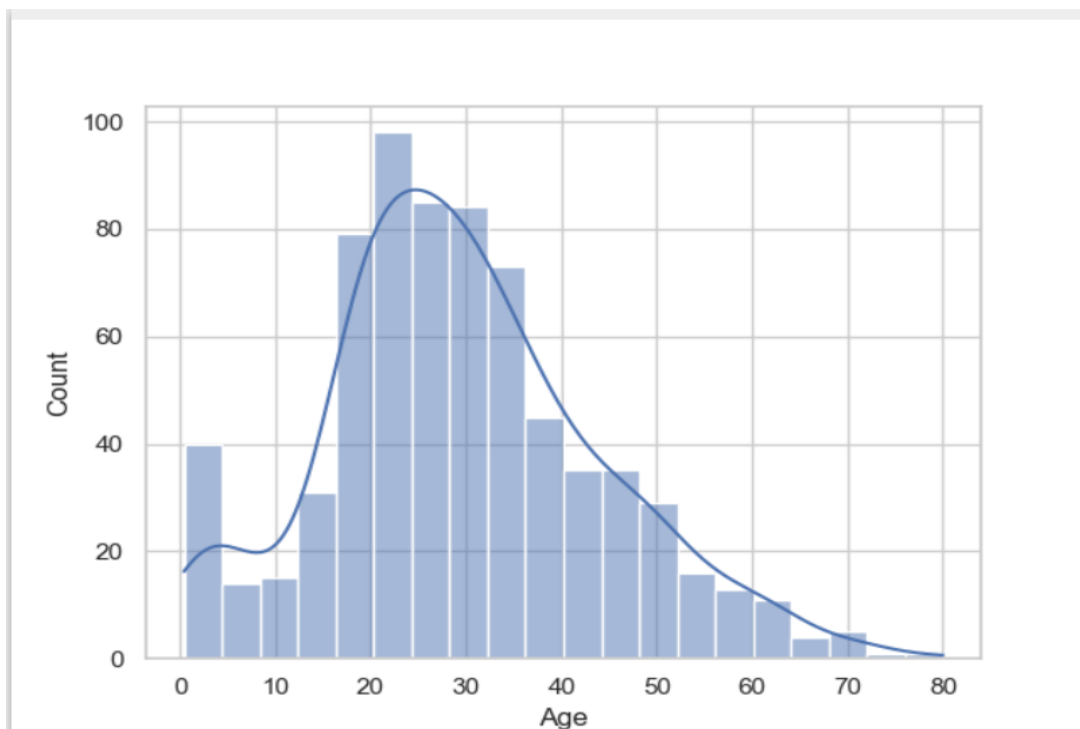- Pclass: Class 3 most common

3. Missing Values

Column	Missing Count
Age	- 177
Cabin	- 687
Embarked	- 2

4. Key Observations & Visual Insights

- Survival Rate:
  - Overall survival rate: ~38%
  - Higher survival for females and passengers in 1st class
- Age Distribution:
  - Young children and adults aged 20–40 form the majority
- Fare vs. Pclass:
  - Higher class → higher average fare
- Embarked Port:
  - Most passengers boarded from Southampton ('S')
- Missing Data:
  - Cabin has a large number of missing values, possibly due to non-assignment
  - Embarked and Age can be imputed using common strategies

5. Summary of Findings

- Gender Impact: Females had a much higher survival rate than males.
- Class Matters: 1st class passengers had better survival rates, likely due to proximity to lifeboats and privilege.
- Missing Data:
  - Age: Could affect model accuracy; can impute using median by Pclass & Sex
  - Cabin: Too many missing → best to drop or simplify (e.g., deck letter)
- Fare and Class: Positive correlation between fare and survival.

-Embarked Port: Could reflect socioeconomic status affecting survival.

Survival Count

Correlation Heatmap