**Titanic Data Analysis Report**

**EDA Report**

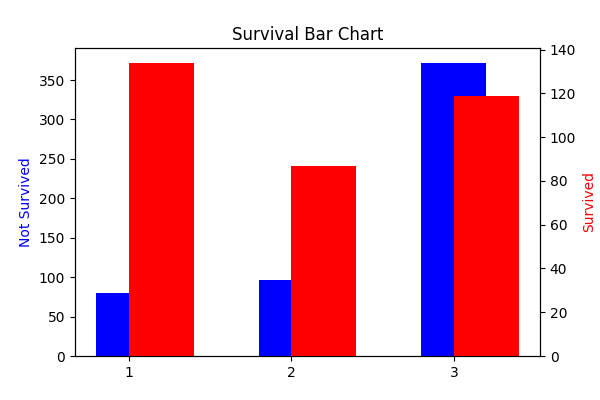
Introduction & Dataset Overview :

* **Title:** Titanic Survival Analysis Report
* **Objective:** Extract insights using visual and statistical exploration.
* **Dataset Summary:**
  + Number of passengers
  + Features included (Age, Sex, Pclass, Embarked, Fare, etc.)
  + Mention missing data issues (Age, Cabin, Embarked)
* **Analysis Steps:**
  + Data cleaning & imputation
  + Feature engineering
  + Exploratory data analysis (EDA)
  + Visualization of key trends and insights

Data Cleaning & Feature Engineering:

* **Missing Data Treatment:**
  + Age: Imputed missing values using median (or mean)
  + Cabin: Dropped due to excessive missing values
  + Embarked: Dropped
* **Data Type Conversion:**
  + Converted Sex to categorical variables
* **New Features Created:**
  + FamilySize = SibSp + Parch (size of family aboard)

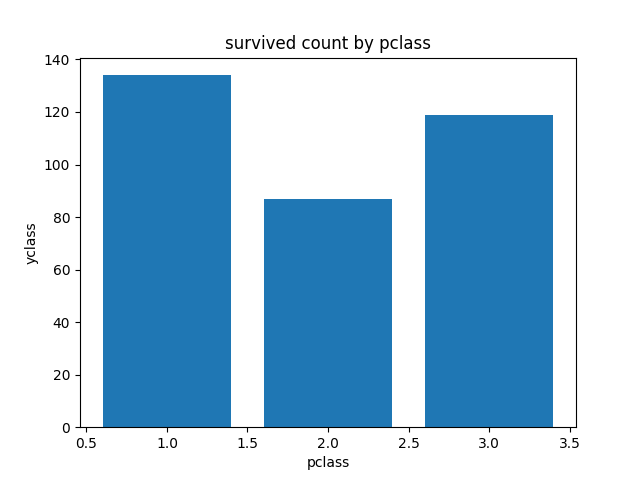
**📊 Important Visualizations (Maps/Charts)**

**1.** Survival Count

**Plot:** Bar chart of survival counts

**Insight:** Imbalanced classes — fewer passengers survived than perished Baseline for classification models

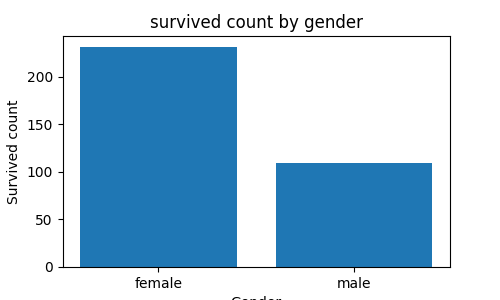
✔️ Shows class imbalance (how many survived vs died)

**2.** Survival by Passenger Class

**Plot**: Countplot showing survival across passenger classes (1st, 2nd, 3rd)

**Insight**: First-class passengers had the highest survival rates .Survival decreased with lower passenger classes, indicating socio-economic influence.

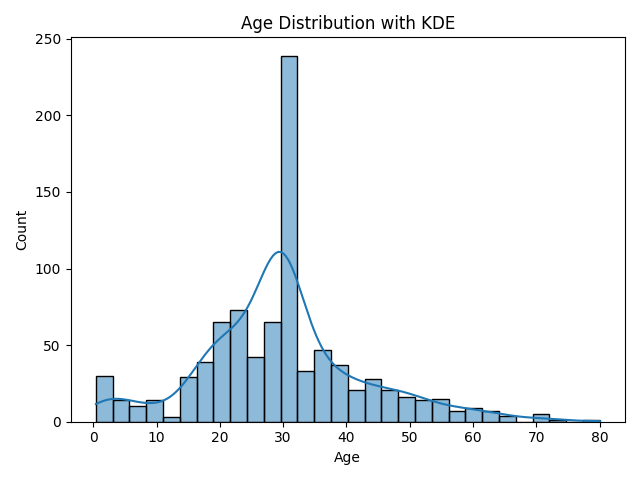
✔️ 1st class passengers survived more

**3.** Survival by Gender

**Plot**: Countplot of Survival by Sex (male/female)

**Insight**: Females had a significantly higher survival rate compared to males .Consistent with historical “women and children first” policy.

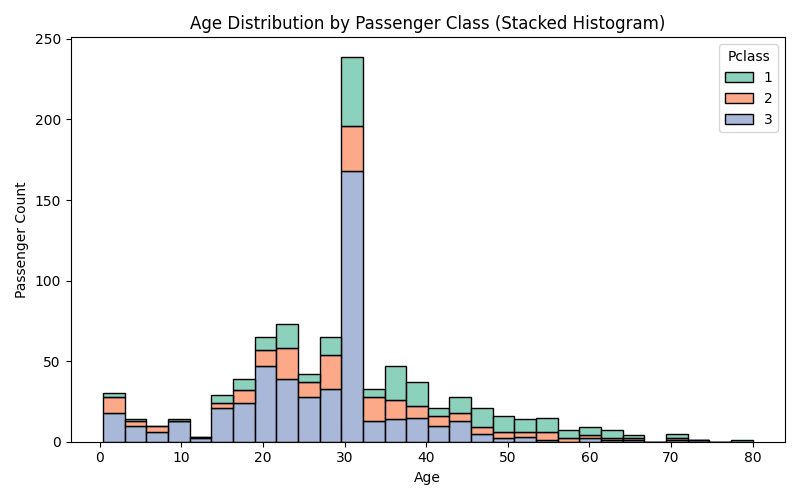
✔️ Females had higher survival rate

**4.** Age Distribution

**Plot**: Histogram with KDE showing age distribution of passengers

**Insight**: Age ranged widely, with concentration in young adults

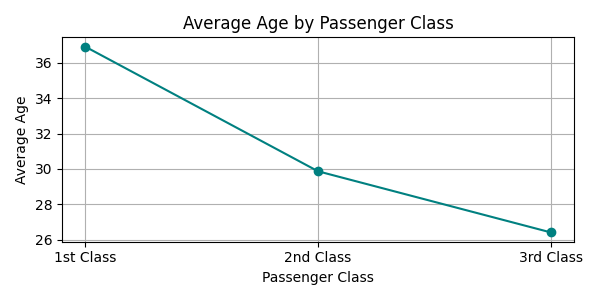
✔️ Age spread, possible age groups

**5.** Age vs Pclass

**Plot**: Age across different Pclass groups

**Insight**: Younger passengers were more frequent in lower classes (3rd class)

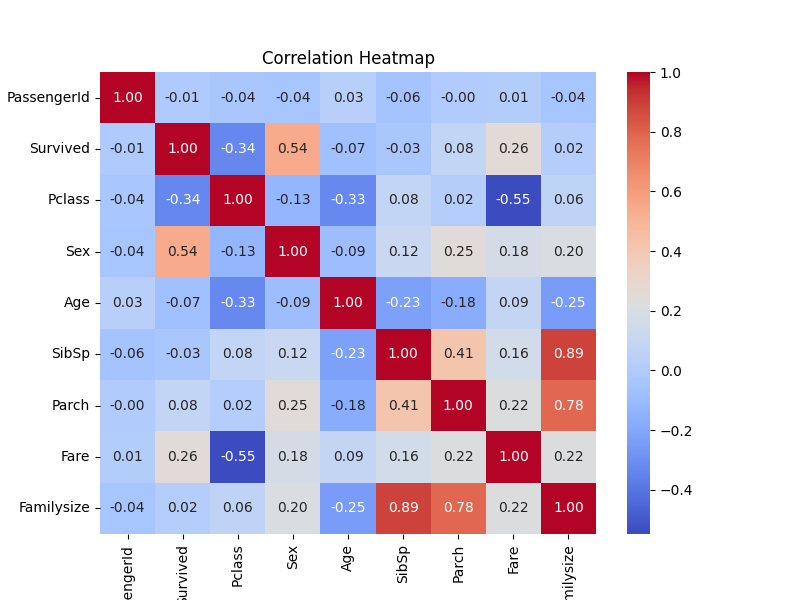
✔️ Younger people in lower class

**6.** Average Age by Passenger Class

**Plot**: Average age in every each class

**Insight**: We can find that what is the average age in class

✔️ Young age people average in 3rd class

**7.** Correlation Heatmap

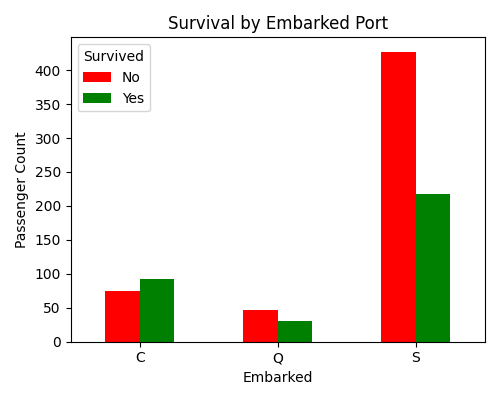
**Plot**: Heatmap showing correlation among numerical features (Age, Fare, Pclass, etc.)

**Insight**: Fare positively correlated with Pclass (higher class, higher fare).

Survival negatively correlated with Pclass (lower class, lower survival) .

Age weakly correlated with other features.

✔️ Correlation between Numerical features

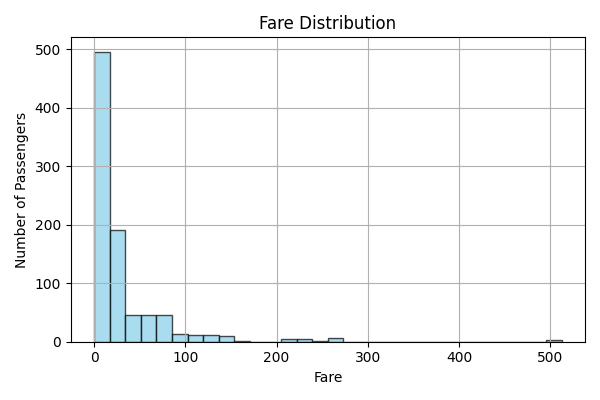
**8.** Survival by Embarked Port

**Plot**: Countplot of survival by port of embarkation (C, Q, S)

**Insight**: Passengers embarked at certain ports had different survival rates.

Possible influence of port socio-economic factors or demographics.

✔️ Port-wise survival trends

**9.** Fare Distribution

**Plot**: Histogram + KDE of Fare values

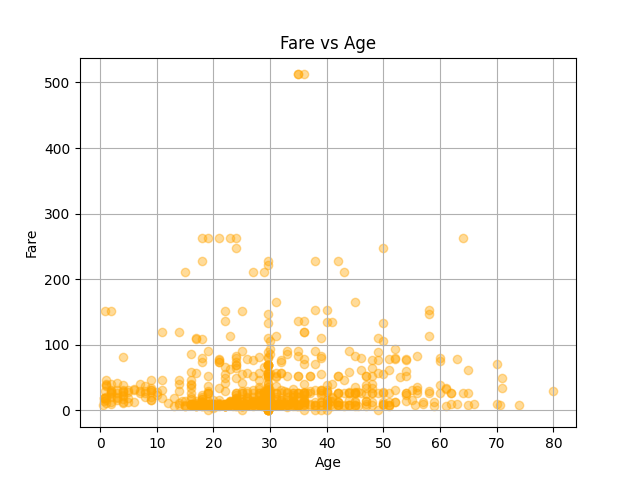
**Insight**: Fare values highly skewed with some outliers paying very high fares.

Higher fares generally associated with better survival (related to Pclass).

Outliers could affect modeling, may require treatment.

✔️ Useful in detecting Outliners

**10.** Fare vs Age



**Plot**: Scatter for age and Fare