Exploratory Data Analysis Report

Dataset: Titanic (Modified version)
Total Records: 1,309 passengers
Total Columns: 28

---

1. Dataset Overview

The dataset represents passenger data from the Titanic. Important features include:

- Survived – Target variable (0 = No, 1 = Yes)

- Pclass – Passenger class (1 = First, 2 = Second, 3 = Third)

- Sex – Encoded gender (0 = Male, 1 = Female)

- Age – Passenger's age in years

- Fare – Ticket fare

- Embarked – Port of embarkation (0, 1, 2)

Additional features like SibSp (siblings/spouses aboard) and Parch (parents/children aboard) are present, as well as several unnamed or placeholder columns (e.g., zero.1, zero.2, etc.), possibly dummies or engineered features.

---

2. Summary Statistics

- **Average Age**: ~29.5 years

- **Fare Range**: 0 to 512; **Average Fare**: ~33.28

- **Survival Count**:

  - Survived: 342 (26.1%)

  - Not Survived: 967 (73.9%)

- **Sex Distribution**:

  - Male (0): 843

  - Female (1): 466

- **Passenger Class Distribution**:

  - 1st Class: 277

  - 2nd Class: 323

- o 3rd Class: 709
- **Port of Embarkation**:
  - o 0 (Cherbourg): 270
  - o 1 (Queenstown): 123
  - o 2 (Southampton): 914

---

3. Visual Insights & Observations

Histograms:

- **Age**: Most passengers were between 20–40 years.
- **Fare**: Skewed distribution; most fares were low, with a few very expensive tickets.

Count Plots:

- **Survival**: Majority did not survive.
- **Sex**: More male passengers than female.
- **Pclass**: 3rd class had the highest number of passengers.

Boxplots:

- **Age vs Survival**: Younger passengers had a higher chance of survival.
- **Fare vs Pclass**: 1st class passengers paid significantly higher fares.

Heatmap:

- **Fare** had a slight positive correlation with Survived.
- **Pclass** negatively correlated with Survived (higher class = better survival).

Pairplot:

- Clear separation in fare and class between survivors and non-survivors.

---

4. Key Insights

- **Gender and survival**: Females had a significantly higher survival rate.
- **Class matters**: 1st class passengers had better survival outcomes.
- **Younger passengers** were more likely to survive.

- **Port of embarkation** had some association with survival trends (passengers from Cherbourg had better survival).