INTRODUCTION

The Titanic dataset is one of the most famous datasets used in data science and machine learning. It contains real data about the **passengers on the RMS Titanic**, which sank in the North Atlantic Ocean on **April 15, 1912**, after hitting an iceberg.

This dataset helps us analyze **what factors affected the survival of passengers**. It includes details like:

- **Name** and **Passenger ID**
- **Gender**
- **Age**
- **Ticket class** (1st, 2nd, or 3rd)
- **Fare** (price paid for the ticket)
- **Family members onboard** (siblings/spouse, parents/children)
- **Embarked** (where they boarded: Southampton, Cherbourg, or Queenstown)
- **Survived** (1 = Yes, 0 = No)

RELATIONSHIPS AND TRENDS

1. Pclass vs Survived
• **Visual Reference:** Heatmap (corr = -0.34) + Pairplot
• **Trend:** 1st class passengers were more likely to survive.
• **Reason:** Priority access to lifeboats and crew.

2. Fare vs Survived
• **Visual Reference:** Pairplot (Fare higher in Survived = 1)
• **Trend:** Higher ticket fares → Higher survival probability.
• **Reason:** Wealthier passengers likely traveled in 1st class with better rescue chances.

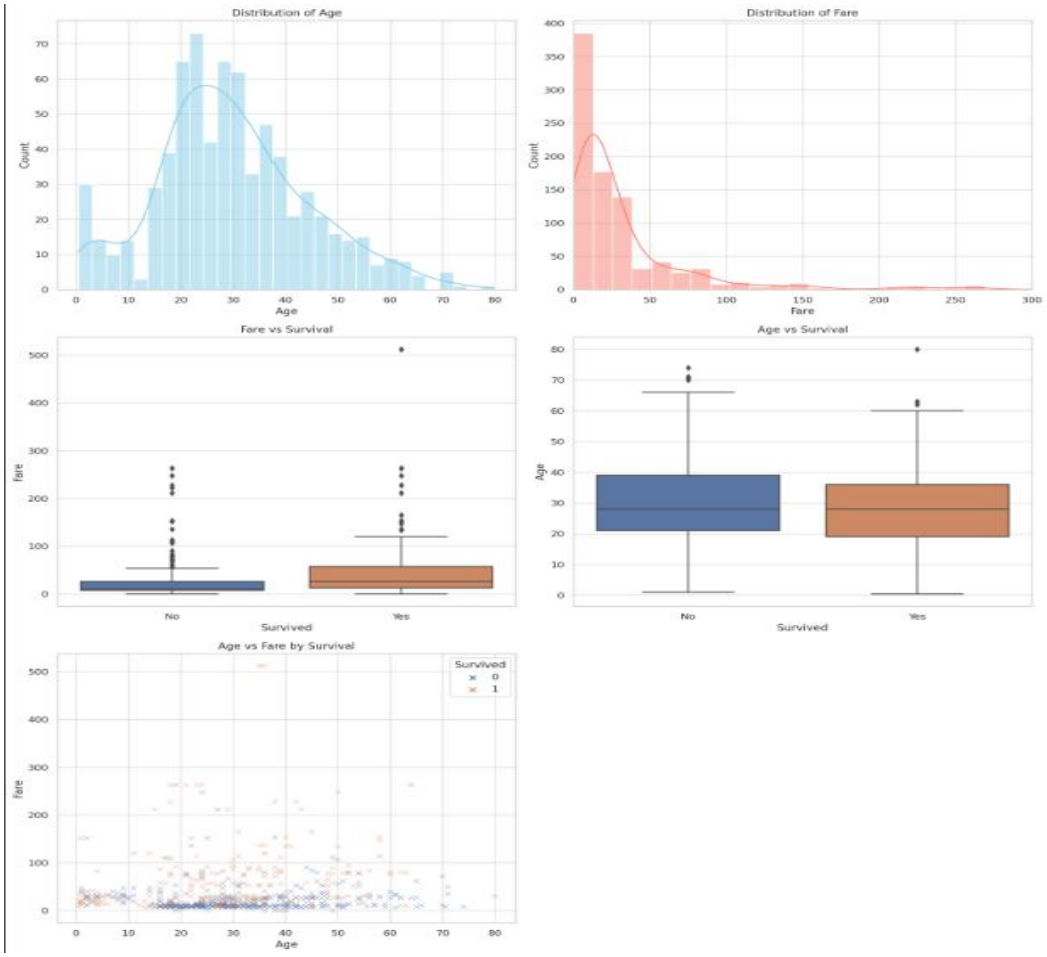
3. Age vs Survived
• **Visual Reference:** Pairplot & Histograms
• **Trend:** Children (age < 10) had better survival; older adults had poorer outcomes.
• **Reason:** "Women and children first" evacuation policy favored younger passengers.

4. SibSp and Parch vs Survived
• **Visual Reference:** Pairplot
• **Trend:**
  o Solo travelers (0 siblings/spouse or parents/children) had lower survival.
  o Those with 1–2 family members had higher survival.
  o Large family groups fared poorly.
• **Reason:** Smaller groups were easier to evacuate and support each other.

5. Fare vs Pclass
• **Visual Reference:** Heatmap & Scatterplot (if used)
• **Trend:** Strong inverse relationship.
• **Interpretation:** Ticket pricing scaled clearly with class level (1st class paid most).

ANALYSIS OF EACH PLOT

1. **Histogram – Age**:
    - Distribution is roughly normal with a slight right skew.
    - Large group between ages 20–40, some children.
2. **Histogram – Fare**:
    - Highly right-skewed; most passengers paid under $100.
    - Very few high-paying passengers.
3. **Boxplot – Fare vs Survival**:
    - Survivors paid higher fares on average.
    - Significant outliers among survivors (likely 1st class).
4. **Boxplot – Age vs Survival**:
    - Survivors are slightly younger.
    - Broad age range in both groups.
5. **Scatterplot – Age vs Fare by Survival**:
    - Survivors cluster at higher fares.
    - No strong visible trend with age alone.



SUMMARY

1. Survival Distribution
About 38% survived, 62% did not.
There is a significant class-based and gender-based disparity in survival rates.

2. Passenger Class (Pclass)
Higher-class passengers had higher survival rates.
Most first-class passengers survived, while the majority of third-class passengers did not.

3. Gender Impact
Female passengers had significantly higher survival (around 74%) than males (around 19%).
Gender was one of the most influential factors.

4. Age Distribution
Most passengers were between 20–40 years old.
Children had better survival outcomes — supporting "women and children first".

5. Fare Patterns
Right-skewed distribution: most paid under $100.
Survivors had higher median fares, indicating a correlation between fare/class and survival.
Outliers exist, especially among survivors with very high fares (likely VIPs or 1st class).

6. Family Size – SibSp & Parch
Passengers with 1–2 relatives onboard had better survival odds.
Too many or no family members resulted in lower survival rates.

7. Correlation Summary
Strongest correlations with Survived:
• Sex (encoded)
• Fare
• Pclass
Age, SibSp, Parch have weaker but noticeable effects.