Exploratory Data Analysis (EDA)

Observations for Each Visual

1. Pairplot

The pairplot compares multiple variables (like Age, Fare, Pclass) and uses color to separate survivors from non-survivors.

- **What it shows:** Females have a visibly higher survival rate than males.
- Younger passengers, especially children, have more survivors.
- 1st and 2nd class passengers show better survival rates than 3rd class.
- People who paid higher fares tended to survive more often — suggesting that wealth (and therefore higher class) increased survival chances.

---

2. Correlation Heatmap

This heatmap shows how strongly each numeric variable is related to survival.

- **Fare:** Positive correlation (~0.26) — higher fare → higher survival rate.
- **Pclass:** Negative correlation (~-0.34) — higher class number (3rd class) → lower survival.
- **Age:** Slight negative correlation (~-0.08) — older passengers had a slightly lower survival rate.
- Also, Fare and Pclass are strongly negatively correlated — meaning higher-class passengers paid more.

---

3. Histogram – Age

The histogram shows how passenger ages are distributed.

- Most passengers were between **20 and 40 years old**.
- There's a smaller number of children (under 10) and older adults (over 60).
- This age distribution helps explain why many middle-aged passengers didn't survive — as survival advantages were higher for children and younger women.

---

4. Histogram – Fare

This histogram shows how fares were spread out.

- The distribution is **right-skewed** — most people paid less than $100, but a few paid much more (luxury tickets).
- These high-fare outliers are often linked to higher survival chances because they were usually in 1st class.

---

5. Boxplot – Age by Pclass

This boxplot compares the ages of passengers in each ticket class.

- 1st class passengers tend to be older on average.
- 3rd class passengers are generally younger — possibly families and working-class individuals.
- The variability (spread of ages) is largest in 1st class, meaning you had both younger and older wealthy passengers.

---

6. Boxplot – Fare by Survival

This boxplot compares fares for survivors and non-survivors.

- Survivors' fares are generally higher, suggesting they had better accommodations (closer to lifeboats, better crew attention).
- Non-survivors mostly paid lower fares, meaning they were more often in 3rd class.

---

7. Scatterplot – Age vs Fare (Survival)

This scatterplot plots passengers by age and fare, coloring by survival.

- Most passengers cluster at low fares and mid-range ages (20–40 years).
- Survivors are more common in **high-fare areas**, especially among younger passengers.
- There's also a small cluster of children with low fares who survived — likely rescued earlier due to the "women and children first" policy.

Final Summary

From these visuals, we can conclude:

- **Gender:** Women survived at much higher rates than men.
- **Class & Fare:** Wealthier (1st and 2nd class) passengers had far better survival odds.
- **Age:** Children had higher survival, elderly had lower.
- **Port of Embarkation:** Passengers from Cherbourg tended to survive more, likely due to a higher proportion of 1st class tickets.
- **Overall:** Social and economic status (class, fare), combined with gender and age, were the most influential factors in survival.