1. Data Overview

- Total rows: 891
- Columns after cleaning: ~10 (dropped Cabin, handled missing Age and Embarked)
- Data types: Mix of categorical and numeric

2. Data Cleaning

- Age had 177 missing values → filled with median.
- Embarked had 2 missing values → filled with mode ('S').
- Cabin had 687 missing values → dropped.
- Fare had large range and high skewness → log-transformed.

3. Univariate Analysis

Feature	Key Findings
Survived	~38% survived, 62% died
Sex	More males than females
Pclass	Most were 3rd class passengers
Age	Many passengers were young adults
Fare	Some passengers paid very high fares (outliers)

4. Bivariate Analysis

Feature vs Survived	Observations
Sex	Women had a much higher survival rate
Pclass	1st class passengers survived more than 3rd class
Age	Children had better survival chances
Embarked	Passengers from Cherbourg (C) had better survival
Fare	Higher fares were linked to higher survival

5. Multivariate Analysis

- Heatmap: Strongest correlations were between Pclass, Fare, and Survived.

- Pairplot: Survivors cluster around high Fare, low Pclass, and are mostly female.
- Pivot Table:
  - Women in 1st class: highest survival (~97%)
  - Men in 3rd class: lowest survival (~13%)

6. Outliers & Skewness

- Fare: Highly skewed → fixed with log transformation.
- Age: Mild skew, no major treatment needed.
- Boxplots revealed outliers, especially in Fare.

Final Conclusions

- Sex, Pclass, and Fare are the most influential features for survival.
- Being female, young, in 1st class, and having a high fare increased survival chances.
- Titanic survival was not random — it was influenced by socioeconomic status and gender.