Titanic Data Exploration Report

A. Basic Statistics and Structure

- df.info() shows 891 entries and 12 columns. Age, Cabin, and Embarked have missing values.
- df.describe() reveals:
  * Mean age: ~29
  * Mean fare: ~32
  * Survival rate: ~38%
- df['Survived'].value_counts():
  * 0 (did not survive): 549
  * 1 (survived): 342

B. Visualizing Correlations

- sns.pairplot(): Survivors tend to be younger, in higher classes, and paid higher fares.
- sns.heatmap(): Strongest negative correlation between Pclass and Survived.
  * Fare positively correlates with survival.
  * Age has weak correlation.

C. Identifying Relationships and Trends

- Gender vs Survival: Females survived more.
- Class vs Survival: 1st class passengers had better survival.
- Age vs Survival: Children (<10) had better chances.
- Fare vs Survival: Higher fares linked to higher survival.

D. Distribution and Comparative Visuals

- Histogram of Age: Most passengers were aged 20-40.
- Boxplot (Fare vs Survived): Survivors paid more.
- Scatterplot (Fare vs Age): High fare and young passengers had higher survival.

E. Observations from Visuals

- Survival rate is ~38%.
- Most females and 1st class passengers survived.
- Survivors were generally younger and paid more.
- Fare and Pclass had the highest correlations with survival.

F. Summary of Findings

1. Gender and Class are major predictors of survival.
2. Fare and Age have noticeable but weaker effects.

3. Children and wealthy passengers had higher survival rates.

4. Data confirms the historical account of rescue priorities.