Executive Summary of Titanic Passenger Survival Analysis

Objective

- To conduct an exploratory data analysis (EDA) on the Titanic dataset and identify key demographic, socioeconomic, and travel-related factors influencing passenger survival.
- To visualize survival disparities across passenger groups using statistical summaries and charts.

Key Findings

1. Overall Dataset Overview

- Total passengers analyzed: **~891**.
- Overall survival rate: **~38% survived**, **62% did not survive**.
- Significant missing data present in:
  - **Age** (≈20% missing)
  - **Cabin** (≈77% missing)

---

2. Survival by Gender

- Gender emerged as the **strongest predictor** of survival.
- **Female passengers:**
  - Approx. **74% survived**
- **Male passengers:**
  - Approx. **19% survived**
- In this charts clearly highlight a sharp divergence in survival outcomes between genders.

---

3. Survival by Passenger Class (Pclass)

- Survival rate significantly increased with socioeconomic standing:
  - **1st Class:** ~62% survived
  - **2nd Class:** ~47% survived
  - **3rd Class:** ~24% survived
- Visualizations show a disproportionate number of deaths among **third-class** passengers.

---

4. Age-Related Observations

- Children had higher survival probabilities:
  - **Passengers under age 10:** ~61% survived
- Age distribution charts show:
  - Younger passengers more likely to survive.
  - Majority of passengers were young adults (20–40), who had lower survival rates compared to children.

---

5. Fare Analysis

- Fare correlated positively with survival likelihood:
  - Survivor's median fare: **~£26**
  - Non-survivor's median fare: **~£10**
- Higher-paying passengers (upper socioeconomic status) had notably higher survival percentages.

---

6. Embarkation Port Influence (Embarked)

- Survival varied by port of departure:
  - **Cherbourg (C):** ~55% survived
  - **Southampton (S):** ~34% survived
  - **Queenstown (Q):** ~39% survived
- Charts imply that passengers boarding from **Cherbourg**, typically associated with wealthier travelers, had a distinct survival advantage.

---

7. Family Size (SibSp & Parch)

- Passengers traveling with **small families (1–3 members)** showed higher survival rates.
- Very large family groups had significantly lower survival percentages.
- Solo travelers had a moderate but lower survival rate compared to small-family groups.

---

8. Missing Data Assessment

- **Age** and **Cabin** fields contain substantial missing values.
- Absence of cabin data limits the ability to explore survival relationships tied to ship location.

Conclusions

- Gender and socioeconomic status (as captured by passenger class and fare) were the **most influential** factors in survival.
- The dataset clearly demonstrates:
    - "**Women and children first**" policy effectiveness.
    - Survivability advantage for **wealthier passengers**.
    - Marked survival inequality across class and social structures.
- This Analysis visualize (bar charts, pie plots, histograms, box plots and distribution plots) effectively highlight these disparities and reinforce the statistical findings.