Executive Summary: Titanic Dataset Analysis

Objective

The analysis aims to uncover patterns in the survival of Titanic passengers using data cleaning, visualization, and statistical analysis techniques. The ultimate goal is to provide actionable insights and improve decision-making for similar scenarios in the future.

Key Findings

1. **Overall Survival Rate:**
    - Only 38.38% of passengers survived the Titanic disaster, emphasizing the severity of the event.
2. **Gender-Based Survival Trends:**
    - Survival rates were significantly higher for females compared to males. This highlights the prioritization of women during the evacuation process.
3. **Class-Based Survival Trends:**
    - Passengers in higher classes (e.g., First Class) had better survival rates than those in lower classes, reflecting social and resource-based advantages during the crisis.

Data Cleaning Process

- **Outlier Detection and Handling:**
    - Utilized the Interquartile Range (IQR) method to identify and manage outliers in numerical columns.
    - Boxplots were employed to visualize and address extreme values by capping them within acceptable bounds.
- **Missing Values:**
    - Addressed missing entries in critical columns like `Age`, `Cabin`, and `Embarked`. Methods included imputing missing values with median or mode based on the context.
- **Data Transformation:**
    - Converted categorical variables such as `Sex` and `Embarked` into a numerical format for analysis.

Visualizations and Insights

- **Survival by Gender:**
    - A bar chart demonstrated the stark contrast in survival rates, with females significantly outnumbering males among survivors.
- **Boxplots for Outliers:**
    - Boxplots illustrated the presence of outliers in variables such as `Fare` and `Age`. These outliers were effectively managed to enhance the reliability of subsequent analyses.
- **Class and Survival Correlation:**
    - Graphical analysis revealed that First-Class passengers were more likely to survive than those in Second or Third-class.

Conclusion

This analysis underscores critical factors influencing survival, such as gender and socioeconomic status. The robust data-cleaning process and effective visualisation techniques ensured accurate and actionable insights. These findings can guide future studies and preparedness measures for large-scale emergencies.