Titanic Dataset Analysis

1. Overview of Dataset

There are data entries of **891** passengers present, which contain fields such as age, sex, class of ticket, price paid, and survived or not.

2. Data Preprocessing

- Dropped Duplicate Entries.
- Handled Missing Values:
    - Used median for missing Age values.
    - Assigned "Unknown." for missing Cabin values.
    - Dropped rows that were missing the value for Embarked.
    - Eliminated Outliers from numerical fields with the z score to gain precise analysis.
    - Categorized Data: Changed Sex and Embarked to numerical type.

3. Major Conclusions from the Analysis

Survival Rates

- **38%** of the passengers survived, and **62%** did not.
- This indicates that survival was not uniform and based on various factors.

Survival by Gender

- **Women had a considerably better survival rate than men.**
- This validates the "women and children first" policy adopted while evacuating.

Survival by Passenger Class

- First-class passengers had a superior survival rate in comparison to second and third class.
- Third-class passengers recorded the lowest survival rate.
- This implies that a **higher social standing gave more hope for survival**.

Age and Survival

- **Kids (particularly the infants) survived more often.**
- This corroborates the "women and children first" decree further.
- Elderly travelers were less likely to survive.

Fare and Survival

- Travellers paying higher fares survived more often.
- This implies that the richer travellers had greater lifeboat access.

3. Conclusion

- The Titanic tragedy was not haphazard—survival depended largely on gender, class of ticket, age, and cost of fare.
- Women, children, and first-class travelers had the best survival prospects.
- Men, third-class travelers, and older passengers had the worst survival prospects.
- Money mattered, with those who spent more on tickets having better access to lifeboats.
- This analysis demonstrates how economic and social realities influenced survival rates during the Titanic disaster.