Titanic Dataset Analysis Report

1. Data Cleaning Insights

- Missing values in Age were replaced with the median value to avoid distortion from extreme values.
- Fare was also filled with the median value for consistency.
- Embarked was filled with the mode value, representing the most common point of embarkation.
- Categorical values for Sex and Embarked were encoded into numeric form for consistency and analysis.

2. Univariate Analysis Insights

- The distribution of Age shows a slight right skew, indicating that more passengers were younger.
- The majority of passengers belonged to 3rd class, followed by 1st and 2nd class.
- Fare distribution showed a right-skewed pattern, with higher fares concentrated among first-class passengers.

3. Bivariate Analysis Insights

- Weak correlation between Age and Fare (correlation ~ 0.1) indicates that age and fare are largely independent.
- Strong negative correlation between Pclass and Fare (correlation ~ -0.55) suggests that higher-class tickets cost more.
- Higher survival rate for 1st-class passengers compared to 2nd and 3rd classes.
- Women had a higher survival rate than men in all passenger classes, consistent with the 'women and children first' policy.

4. Multivariate Analysis Insights

- Pair plots showed survival patterns influenced by both Pclass and Fare.
- Passengers in higher classes with higher fares had a greater likelihood of survival.
- Heatmaps indicated low correlation between most numerical features, showing that the dataset has mostly independent variables.
- Grouped bar plots showed that 1st-class passengers, especially women, had the highest survival rates.

5. Overall Inferences

- Higher-class passengers and women had a higher survival rate.
- Low correlation between Age and Fare indicates that age was not a major determinant of ticket price.
- Strong correlation between Pclass and Fare shows that higher-class tickets were more expensive.
- The survival rate differences between sexes and classes suggest that social factors played a major role in survival outcomes.

Conclusion

The analysis confirms that survival on the Titanic was not random - it was influenced by factors like passenger class, gender, and fare. Higher-class passengers and women had a significant survival advantage, highlighting the societal norms and emergency protocols followed during the disaster.