Summary of Findings - Titanic Dataset

1. Importing Required Libraries

All necessary libraries for data analysis and visualization were successfully imported, including pandas, numpy, matplotlib, seaborn, etc.

2. Loading Dataset

The Titanic dataset was loaded, allowing for exploration of passenger demographics, travel classes, survival outcomes, and more.

3. Data Cleaning

Initial cleaning steps were taken, including handling missing values and checking data types to ensure correct processing.

4. Finding Duplicates

Duplicate entries were checked and managed appropriately to maintain data integrity.

5. Checking Data Types

Data types for each column were verified and converted where necessary, enabling proper numerical and categorical operations.

6. Checking Missing Values

Missing values were identified in columns such as 'Age', 'Cabin', and 'Embarked'. Suitable strategies like imputation or omission were considered.

7. Visualizations

- Pairplot: Helped visualize relationships between numerical features.
- Heatmap: Showed that 'Fare' and 'Pclass' had strong negative correlation, while 'SibSp' and 'Parch' showed

positive correlation.

- Countplot: Visualized categorical distributions like survival count by gender.

- Barplot: Compared survival rates across different classes and genders.

- Histplot: Displayed age and fare distributions.

- Boxplot: Revealed outliers in fare and age data.

- Scatterplot: Explored relationships between two continuous variables.

8. Summary of Key Findings

- Higher-class passengers (1st class) had a significantly higher survival rate.

- Women and children survived at much higher rates compared to men.

- Passengers embarking from Cherbourg had slightly better survival rates.

- Ticket fare and family size influenced survival chances.

- Extensive missing data in 'Cabin' column requires careful handling.