Titanic Dataset - Exploratory Data Analysis Report

1. Objective

The goal of this project is to perform Exploratory Data Analysis (EDA) on the Titanic dataset to extract meaningful patterns, trends, and anomalies using statistical and visual tools.

2. Tools Used

- Python
- Pandas
- Matplotlib
- Seaborn
- Jupyter Notebook

3. Dataset Overview

The Titanic dataset contains information about the passengers aboard the Titanic, including attributes like age, sex, ticket fare, class, and whether they survived.

4. Key Analysis Steps

- Checked data types and missing values using `.info()` and `.isnull().sum()`.
- Used `.describe()` and `.value_counts()` to understand distributions.
- Created visualizations including histograms, boxplots, scatterplots, pairplots, and heatmaps.
- Analyzed correlations between variables and survival outcomes.

5. Key Findings

- Females had a significantly higher survival rate than males.
- First-class passengers were more likely to survive compared to other classes.
- Younger passengers had slightly higher survival rates.

- Fare paid showed some correlation with survival probability.
- There were missing values in 'age' and 'embarked' columns.

6. Conclusion

EDA helped uncover important relationships and trends in the Titanic dataset. The insights can guide further predictive modeling or deeper domain analysis.