Titanic EDA & Data Cleaning - Output Report

1. Project Summary

This report presents the findings and data cleaning process for the Titanic dataset.
The goal was to identify survival trends by performing exploratory data analysis (EDA) and cleaning the dataset using Python tools like Pandas, NumPy, Seaborn, and Matplotlib.

2. Data Cleaning Steps

- Handled missing values in 'Age' using median.
- Filled missing 'Embarked' with mode.
- Dropped 'Cabin' due to high missing rate.
- Capped 'Fare' at the 99th percentile to handle outliers.
- Converted 'Sex' to numeric (male: 0, female: 1).
- Applied one-hot encoding to 'Embarked'.

3. Exploratory Data Analysis (EDA)

Key visualizations were created using Seaborn:

- Survival by Sex: Showed higher survival rate among females.
- Survival by Pclass: Higher class passengers had better survival chances.
- Age Distribution: Revealed younger passengers were more likely to survive.
- Fare vs Survival: Higher fare passengers had higher survival probability.

4. Key Insights

- Females had a significantly higher survival rate than males.
- 1st class passengers were more likely to survive compared to 2nd and 3rd class.
- Younger passengers and children had better survival outcomes.
- Passengers who paid higher fares had a slightly better chance of survival.

5. Tools & Technologies

- Python 3.x
- Pandas
- NumPy
- Matplotlib
- Seaborn
- Jupyter Notebook