Titanic Dataset - Exploratory Data Analysis Summary

Dataset Overview:

The Titanic dataset is a classic machine learning dataset that contains information about passengers aboard the RMS Titanic. It includes features such as Age, Gender, Passenger Class, Fare, Embarkation Port, and whether the passenger survived the disaster. The goal of this exploratory analysis is to uncover meaningful patterns and trends that may have influenced survival.

Objective:

To extract meaningful insights from the Titanic dataset through statistical and visual exploratory data analysis using Python (Pandas, Matplotlib, and Seaborn).

Data Cleaning:

- Filled missing Age values with the median.
- Filled missing Embarked values with the mode.
- Dropped the Cabin column due to 77% missing values.- Converted cleaned data into a usable format for analysis.

Key Statistical Insights:

- Most passengers were aged between 20 and 40 years.
- Majority of passengers were from the 3rd class.
- Female passengers had a significantly higher survival rate.
- Passengers in 1st class had better survival chances than those in 2nd or 3rd class.

Key Visual Insights:

- Countplot showed survival rate was higher for females.
- Class-wise analysis revealed survival rate drops from 1st to 3rd class.
- Histogram of Age showed right-skewed distribution with a peak in young adults.
- Correlation heatmap showed a negative correlation between Fare and Pclass.
- Pairplot illustrated general trends but lacked strong patterns for Age vs Survival.

Final Inference:

From this analysis, we infer that survival on the Titanic was heavily influenced by socio-economic factors such as gender and class. Women and first-class passengers had significantly higher survival rates, reflecting real-world prioritization during rescue operations. EDA allowed us to detect missing data issues, clean the dataset effectively, and generate insights using visual trends and correlations. The dataset is now prepared for further steps such as modeling or prediction.