Titanic Dataset - Exploratory Data Analysis (EDA) Report

Objective:

The objective of this analysis is to extract insights from the Titanic dataset using visual and statistical exploration techniques.

Tools Used:

- Python, Pandas, Matplotlib, Seaborn

Dataset Overview:

- The dataset consists of information on passengers aboard the Titanic, including demographic data, class, fare, survival status, and more.
- Key columns include: PassengerId, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked, and Survived.

1. Data Understanding and Preprocessing:

- Data loaded and previewed using .head() method.
- Dataset information and data types identified using .info().
- Descriptive statistics generated using .describe().
- Missing values identified and visualized using a heatmap.

2. Data Visualization and Analysis:

- Distribution Analysis:
  - Age distribution analyzed using histograms.
  - Significant right-skew in Age distribution.

- Correlation Analysis:
  - Correlation heatmap generated to identify relationships between features.
  - Notable correlations observed between Fare and Pclass, and Pclass and Survived.
- Visualizations:
  - Pairplots to understand interactions between Age, Fare, Pclass, and Survival.
  - Boxplots to observe Age distribution across different Pclass values.

3. Key Findings:

- The dataset has missing values in the Age and Cabin columns.
- Age distribution is right-skewed, with most passengers aged between 20-40 years.
- Pclass significantly impacts survival, with higher classes showing higher survival rates.
- Fare is positively correlated with survival, indicating wealthier passengers had higher survival rates.

4. Conclusion and Recommendations:

- Address missing values for Age through imputation based on Pclass or other related features.
- Further analysis on Cabin and Embarked columns may provide additional insights.
- Consider analyzing SibSp and Parch for familial relationships and survival patterns.

Appendix:

- Screenshots of key visualizations (histograms, heatmaps, pairplots, and boxplots).

End of Report.