Exploratory Data Analysis (EDA) on Titanic Dataset

1. Introduction

The Titanic dataset is a classic dataset used in data science and machine learning. It contains passenger details such as age, sex, class, fare, cabin, and survival status. **Objective:** Perform Exploratory Data Analysis (EDA) to identify patterns, trends, and anomalies that influenced passenger survival. **Tools Used:** Pandas, Matplotlib, Seaborn, Jupyter Notebook.

2. Dataset Overview

- Dataset contains passenger information such as age, gender, class, fare, and survival. - Missing values are found in Age, Cabin, and Embarked columns. - Target variable is 'Survived' (0 = No, 1 = Yes). - Fare distribution is skewed with some outliers.

3. Univariate Analysis

- Most passengers were between 20–40 years old. - Fare distribution is skewed with a few very high values. - About 38% passengers survived. - Females had a higher survival rate than males. - 1st class passengers survived more than 3rd class.

4. Bivariate & Multivariate Analysis

- Strong negative correlation between Passenger Class (Pclass) and Fare. - Fare is positively correlated with survival — higher fare passengers had better chances. - Weak correlation between Age and Survival. - Pairplots reveal survivors generally paid higher fares and were in higher classes.

5. Key Insights & Findings

1. Women and children had higher survival rates. 2. Passenger class strongly influenced survival chances (1st > 2nd > 3rd). 3. Fare amount was positively related to survival. 4. Age played a minor role — children had slightly better survival chances. 5. Missing Cabin data indicates many 3rd class passengers lacked assigned cabins.

6. Conclusion

The Titanic dataset demonstrates how socio-economic status, gender, and age significantly influenced survival rates. EDA shows that females, children, and upper-class passengers had better chances of survival. This analysis provides a foundation for predictive modeling of survival outcomes.