Exploratory Data Analysis (EDA) on Titanic Dataset

Debolina Das

Data Analyst Internship – Task 5

December 16, 2025

1 Introduction

Exploratory Data Analysis (EDA) is a crucial step in data analysis that helps in understanding the structure of the dataset and identifying patterns, trends, and anomalies using statistical and visual techniques. This report presents the EDA performed on the Titanic dataset.

2 Dataset Description

The Titanic dataset contains information about passengers including age, gender, passenger class, fare, and survival status. The dataset consists of 891 records with survival as the target variable.

3 Tools Used

- Python

- Pandas

- Matplotlib

- Seaborn

- Google Colab

4 Basic Data Exploration

Initial exploration was performed using functions such as `.info()`, `.describe()`, and `.value_counts()`. Missing values were observed in the Age and Embarked columns.

5 Univariate Analysis

Histogram analysis showed that most passengers were between 20 and 40 years of age. Box-plot analysis of fare revealed the presence of outliers and a right-skewed distribution.

6 Bivariate Analysis

Analysis of survival based on gender showed that females had a significantly higher survival rate. Passenger class also influenced survival, with first class passengers having higher survival chances.

7 Multivariate Analysis

Pairplot and correlation heatmap were used to analyze relationships between multiple variables. Fare showed a positive correlation with survival, while passenger class showed a negative correlation.

8 Handling Missing Values

Missing values in the Age column were filled using the median, and Embarked missing values were filled using the mode to maintain data consistency.

9 Summary of Findings

- Females had a higher survival rate than males

- First class passengers survived more than third class passengers

- Fare positively influenced survival

- Age had a moderate impact on survival

- Dataset contained outliers and skewed data

10 Conclusion

The EDA process provided meaningful insights into the Titanic dataset. Visualization and statistical analysis helped in understanding key factors influencing passenger survival and prepared the data for further modeling.