Exploratory Data Analysis on Titanic Dataset

Internship Task: Data Analyst Internship

Submitted by: Sushree Subhadarsini Jena

Date: 09-06-2025

1. Introduction

Exploratory Data Analysis (EDA) is the process of analyzing datasets to summarize their main characteristics using visual and statistical methods. This report explores the Titanic dataset to uncover patterns, trends, and key relationships influencing survival outcomes.

2. Dataset Information

The dataset used is the Titanic dataset from Kaggle. It contains information on passengers such as Age, Gender, P class, Fare, and whether they survived. This data is commonly used for binary classification tasks.

3. Data Cleaning

- Checked for missing values using isnull().sum()
- Filled missing 'Age' values with the median
- Dropped 'Cabin' due to too many missing values
- Encoded categorical variables like 'Sex' and 'Embarked'

4. Visualizations and Observations

Age Distribution

Description of observation from the plot goes here. Replace this with actual insights.

Survival Count

Description of observation from the plot goes here. Replace this with actual insights.

Survival vs Gender

Description of observation from the plot goes here. Replace this with actual insights.

Survival vs Pclass

Description of observation from the plot goes here. Replace this with actual insights.

Scatterplot: Age vs Fare

Description of observation from the plot goes here. Replace this with actual insights.

Heatmap: Feature Correlation

Description of observation from the plot goes here. Replace this with actual insights.

5. Summary of Findings

- Females were more likely to survive.
- Higher class passengers (Pclass 1) had better survival rates.
- Fare and Pclass showed moderate correlation with survival.
- Majority of passengers were between 20-40 years old.
- Most passengers embarked from Southampton.

6. Conclusion

The EDA helped identify crucial insights in the Titanic dataset. These patterns are essential for building accurate predictive models and understanding data relationships.