

Titanic Dataset Exploratory Data Analysis

Project Report

Student Name: Samir Ansari

Project Title: Titanic Dataset Exploratory Data Analysis

Domain: Python / Data Analysis / Data Science

1. Introduction

Exploratory Data Analysis (EDA) is a crucial step in data science that helps in understanding the structure, patterns, and insights hidden within data. This project focuses on performing EDA on the famous Titanic dataset, which contains information about passengers such as age, gender, passenger class, family size, embarkation port, and survival status. The goal of this project is to analyze the dataset and identify key factors that influenced passenger survival during the Titanic disaster.

2. Objective

- To understand the structure and features of the Titanic dataset
- To perform data cleaning and handle missing values
- To analyze relationships between different features and survival
- To visualize data patterns using graphs and plots
- To extract meaningful insights that explain survival trends

3. Tools and Technologies Used

- Python 3
- Pandas Library
- NumPy Library
- Matplotlib
- Seaborn
- Jupyter Notebook / Google Colab

4. Dataset Description

The Titanic dataset consists of passenger-related information including Passenger Class (Pclass), Sex, Age, Fare, Embarked Port, Family Size, and Survival Status. The target variable is Survived, which indicates whether a passenger survived (1) or did not survive (0).

5. Project Description

The dataset was loaded and explored to understand its structure and data types. Missing values were handled appropriately. Feature engineering such as Family Type was performed. Categorical variables were converted into numerical form using One-Hot Encoding. Visualizations like count plots, bar plots, and box plots were used to analyze survival patterns.

6. Key Features of the Project

- Data cleaning and preprocessing
- Handling missing values
- Feature engineering (Family Type)
- One-Hot Encoding of categorical variables
- Visual analysis of survival trends
- Statistical insights from data distribution

7. Results and Observations

- Female survival rate was higher than male survival rate
- Pclass 3 passengers had the highest mortality
- Passengers embarked from port C survived more
- Age group 20–40 had higher mortality
- Small families had higher survival chances

8. Learning Outcomes

- Understanding of EDA concepts
- Hands-on experience with Pandas and NumPy
- Improved visualization skills
- Knowledge of data preprocessing techniques

9. Conclusion

This project demonstrates how EDA helps uncover meaningful insights from data. It strengthened the understanding of data analysis and built a strong foundation for machine learning and predictive modeling.