

Project:1 Report

Exploratory Data Analysis (EDA) on Titanic Dataset

Introduction

The Titanic dataset is one of the most popular datasets used in data science for practice. It provides passenger details such as age, gender, class, fare, and survival status. The objective of this project is to perform Exploratory Data Analysis (EDA) to identify survival patterns and factors that influenced survival on the Titanic.

Objectives

1. Clean and Preprocess the Titanic dataset.
2. Handle missing values appropriately.
3. Perform statistical and visual analysis to identify key insights.
4. Analyze survival rates based on demographics (age, sex, class, etc.).
5. Visualize data distributions and correlations.

Methodology

1: Data Collection:

Importe the Titanic dataset (Titanic Dataset.csv) into Pandas dataframe.

2: Data Cleaning:

Handled missing values in Age, Cabin, and Embarked columns.

Converted categorical variables (Sex, Embarked) into numeric/usable form.

Removed irrelevant columns such as PassengerId, Name, and Ticket for analysis.

3: Data Exploration:

Check dataset dimensions and summary statistics.

Analyzed the distribution of continuous variables (Age, Fare).

Studied categorical variables (Sex, Pclass, Embarked) in relation to survival.

4: Data Visualization:

1. Count Plots: Gender vs Survival, Class vs Survival.
2. Boxplots: Age distribution by survival status.
3. Histograms: Distribution of fares.
4. Heatmap: Correlation among numerical features.

Results

->Key insights from the Titanic dataset:

1. **Gender:** Females had a much higher survival rate compared to males.
2. **Passenger Class:** Passengers in 1st class had the highest survival chances, while 3rd class had the lowest.
3. **Age:** Children had a relatively higher survival rate than adults.
4. **Fare:** Higher fare passengers were more likely to survive (linked to class).

5. Embarked Port: Passengers who boarded from Cherbourg (C) had higher survival rates compared to Southampton (S) and Queenstown (Q).

Conclusion

The EDA of the Titanic dataset revealed that **gender, class, and fare** were the most important factors affecting survival. Women and children in higher classes had significantly better chances of survival. This analysis highlights how social and economic factors played a role in survival outcomes during the Titanic disaster.