Report On Exploratory Data Analysis (EDA)

Titanic Dataset - Exploratory Data Analysis (EDA)

Introduction

This report presents an Exploratory Data Analysis (EDA) on the Titanic dataset. The main objective is to uncover patterns, trends, and relationships between passenger features and survival outcomes using statistical summaries and visualizations.

Basic Data Overview

The dataset contains information like:

Passenger details (Name, Age, Sex, etc.)

Travel class (Pclass), Fare, Cabin

Target variable: Survived (0 = No, 1 = Yes)

Missing values were found in Age, Cabin, and Embarked columns.

Visual Insights

1. Survival Count

Observation: More passengers did not survive than those who did.

2. Age Distribution

Observation: Most passengers were between 20–40 years old.

3. Passenger Class vs Survival

Observation: 1st class passengers had the highest survival rate.

4. Gender vs Survival

Observation: Females had a significantly higher chance of survival compared to males.

5. Fare Distribution by Class

Observation: Passengers in 1st class paid much higher fares.

6. Correlation Heatmap

Observation: Fare and Pclass show notable correlation with Survived.

Key Stats

Survival Rate: ~38.38%

Total passengers: 891

Missing Age entries: 177

Summary of Findings

1st Class passengers had better chances of survival than those in 2nd or 3rd.

Gender was a major factor — females were more likely to survive.

Fare paid also influenced survival (higher fare = better survival chances).

Age played a minor role — young children had slightly better survival rates.