Titanic Data Analysis Report

Introduction

This report presents a complete exploratory data analysis (EDA) of the Titanic dataset. The goal is to clean the data, handle missing values, perform statistical analysis, and visualize patterns.

Dataset Overview

The Titanic dataset contains information about passengers aboard the Titanic, including age, gender, ticket class, fare, and survival status. The dataset is widely used for machine learning and data analysis practice.

Data Cleaning

• Missing Age values were filled using the median.
• Missing Embarked values were filled using the mode.
• Cabin column was dropped due to excessive missing values.
These steps help prepare the dataset for accurate analysis.

Statistical Analysis

Key summary statistics were computed to understand distributions such as Age, Fare, and Passenger Class. Additionally, group-wise survival statistics were calculated for gender, age, and class.

Visualizations

Various visualizations were created using Matplotlib and Seaborn:
• Age distribution histogram
• Survival by passenger class
• Survival by gender
• Correlation heatmap
These visualizations reveal clear survival patterns, such as higher survival rates among females and higher-class passengers.

Key Findings

• Females had a significantly higher survival rate compared to males.
• Passengers in 1st class were more likely to survive.
• Younger passengers had slightly better survival odds.

• Fare and class are strongly correlated.

Conclusion

The Titanic dataset provides insightful patterns about survival based on age, class, and gender. The EDA approach helps understand the structure of the data and prepares it for further machine learning modeling.