Elevate labs

Data Analyst - Internship

Task - 5

Exploratory Data Analysis (EDA)

Sireesha Karagana

28 April 2025

Exploratory Data Analysis Report on Titanic Dataset

1. Introduction

This report presents the exploratory data analysis (EDA) performed on the Titanic dataset.
The main goal is to find patterns, relationships, and important insights using statistical summaries
and visualizations.

2. Tools Used

- Python
- Pandas
- Matplotlib
- Seaborn
- Jupyter Notebook

3. Dataset Information

After loading the dataset:

- Columns include: PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare,
- Cabin, Embarked
- Missing values found mainly in 'Age', 'Cabin', and 'Embarked'.

4. Data Exploration

- Used .info() to see data types and missing values.
- Used .describe() for statistical overview.
- Used .value_counts() for categorical columns like 'Survived' and 'Sex'.

5. Visualizations and Observations

5.1 Histograms

- Age Distribution: Most passengers are between 20 and 40 years old.5.2 Boxplots
- Survival vs Age: Younger passengers had a slightly higher chance of survival.

5.3 Scatterplots

Age vs Fare: Some older passengers paid higher fares.

5.4 Pairplot

Visualized relationships among features like Age, Fare, Pclass, and Survived.

5.5 Heatmap

Strong correlation found between Fare and Pclass (negative relation).

6. Key Insights

- Females had a higher survival rate compared to males.
- 1st class passengers had better survival chances.
- Younger passengers had slightly better survival chances.
- Fare amount and passenger class were closely related.
- Many missing values were present in 'Cabin' column.

7. Conclusion

EDA helped in understanding the Titanic dataset deeply.
The visualizations revealed important patterns about survival chances based on gender, class, age, and fare.
This task improved skills in data exploration and visualization.