Exploratory Data Analysis (EDA) Report

Dataset: Titanic Dataset
Prepared by: Angel Suri
Internship Task: Task 5 - Data Analyst Internship

Objective:

To perform an exploratory data analysis on the Titanic dataset in order to understand the data, identify key patterns, and generate insights related to passenger survival based on various factors such as gender, class, age, and fare.

Dataset Summary:

- **Total Entries:** 891 passengers

- **Columns:** 12 (PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked)

- **Target Variable:** Survived (1 = survived, 0 = did not survive)

Data Cleaning:

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

- Converted categorical columns like Sex and Embarked for analysis.

- Removed irrelevant columns such as Name and Ticket for simplicity in visualizations.

Key Visualizations:

- **Countplot:** Showed survival counts by gender and class.

- **Boxplot:** Showed distribution of Age across Passenger Class.

- **Histograms:** Used for Age and Fare distributions.

Insights:

1. **Gender Impact:** Females had a significantly higher survival rate than males.

2. **Passenger Class:** First-class passengers had the highest survival rate.

3. **Age Factor:** Children and younger individuals had higher survival rates.

4. **Fare Influence:** Higher fare passengers were more likely to survive.

5. **Embarked Port:** Passengers who embarked from port 'C' had higher survival rates compared to those from 'S' and 'Q'.

6. **Family Size:** Passengers traveling with family (moderate SibSp + Parch) had a better chance of survival than those traveling alone.

Summary:

The analysis confirmed that survival rates were strongly influenced by socio-economic status, gender, age, and fare. First-class female passengers had the highest chances of survival, while third-class males had the lowest.

This exploratory study provides a foundation for further predictive modeling using machine learning techniques. The insights gained highlight the importance of social structures in emergency situations and can inform future safety protocol planning.

Tools Used:

- Python

- Pandas

- Matplotlib

- Seaborn

- Jupyter Notebook