Titanic Dataset Exploratory Data Analysis (EDA) Report

1. Introduction

The Titanic dataset provides information about passengers aboard the infamous Titanic shipwreck. Using Exploratory Data Analysis (EDA), we aim to uncover hidden patterns, important factors influencing survival, and initial insights into the dataset.

2. Dataset Overview

- **Rows:** 891

- **Columns:** 12

- Key features include passenger details such as **Age**, **Sex**, **Fare**, **Pclass**, **SibSp**, **Parch**, and survival outcome **(Survived)**.

---

3. Exploratory Data Analysis (EDA)

3.1 Basic Data Checks

- Used .info() and .describe() to understand the data structure and statistical properties.

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

- Cabin has a large number of missing entries, suggesting it might not be very reliable for analysis.

3.2 Univariate Analysis

- **Histograms** revealed that:
    - **Age** is slightly right-skewed (more young passengers).
    - **Fare** has several outliers with extremely high values.

- **Countplots** showed:
    - More males than females traveled.
    - Most passengers belonged to the 3rd class.

- o Most passengers embarked from location 'S' (Southampton).

3.3 Bivariate Analysis

- **Sex vs Survived:**

  - o Females had a much higher survival rate compared to males.

- **Pclass vs Survived:**

  - o 1st class passengers were more likely to survive.

- **Age vs Survived:**

  - o Younger passengers tended to survive more than older passengers.

3.4 Correlation Analysis

- **Heatmap** revealed:

  - o Fare is negatively correlated with Pclass (higher class, higher fare).

  - o Sex (after encoding) strongly correlates with Survived.

- **Pairplot** further visualized relationships between numerical features.

---

4. Key Observations

- Being female significantly increased the chance of survival.

- First-class passengers had better survival rates.

- Younger passengers generally survived more than older ones.

- The **Cabin** feature has too many missing values.

- Fare values show outliers; some passengers paid very high fares.

---

5. Conclusion

The EDA provided important insights into passenger survival factors on the Titanic. Sex, class, and age were critical indicators of survival chances. Handling missing values (especially Cabin) and further feature engineering could enhance predictive models in future analysis.