Report of Findings – Titanic Dataset EDA

Overview:

The objective of this analysis was to perform **Exploratory Data Analysis (EDA)** on the Titanic dataset to understand survival patterns and important features influencing passenger survival. This involved statistical summaries, visual exploration, and identifying correlations.

Dataset Used:

- **train.csv** – Contains 891 entries with features like Survived, Pclass, Sex, Age, Fare, etc.
- **test.csv** – Additional data for inference (not evaluated for survival).
- **gender_submission.csv** – Provided for comparison.

Key Insights:

1. Survival Distribution

- 38.4% of the passengers survived.
- Majority of non-survivors were male and belonged to lower classes.

2. Gender and Survival

- **Females** had a significantly higher survival rate (~74%).
- **Males** had a much lower survival rate (~19%).

3. Passenger Class (Pclass)

- Passengers in **1st class** had the highest survival rate.
- Survival rates decreased with class:
  1st > 2nd > 3rd

4. Age

- Younger passengers, especially **children**, had higher survival rates.
- Elderly passengers had lower chances of survival.

5. Fare

- There is a **positive correlation** between fare and survival.
- Passengers who paid higher fares were more likely to survive.

6. Correlation Matrix

- Fare and Pclass showed negative correlation (i.e., 1st class = higher fare).
- Sex, Pclass, and Fare showed meaningful influence on Survived.

Visualizations Used:

- **Bar Plot** – Survival counts
- **Pie Chart** – Gender-based survival
- **Boxplot** – Fare and class distribution
- **Scatterplot** – Age vs Fare vs Survival
- **Heatmap** – Correlation matrix
- **Pairplot** – Relationships among multiple features

Final Summary:

EDA reveals that survival on the Titanic was influenced heavily by **gender**, **passenger class**, and **age**. Women and children from higher classes were more likely to survive. These insights are foundational for any further modeling or predictions.