Summary of Findings – Titanic EDA

1. **Survival Rate Insights:**

    o   The survival rate was approximately **38%**.

    o   **Females** had a significantly higher survival rate compared to males.

    o   **Passengers in 1st class** had higher chances of survival than those in 2nd and 3rd class.

2. **Demographic Trends:**

    o   The average passenger age was around **29.7 years**.

    o   Most passengers were adults, but a notable number of children were also onboard.

    o   The **age group 20–40** was the most common among passengers.

3. **Fare and Class Correlation:**

    o   Higher ticket **fares were associated with 1st class passengers**.

    o   Passengers in 3rd class generally paid lower fares and had the **lowest survival rate**.

4. **Missing Data:**

    o   **Age** has 177 missing values.

    o   **Cabin** has 687 missing entries – a significant portion of the dataset.

    o   **Embarked** has 2 missing values.

5. **Statistical Observations:**

    o   Strong positive correlation between **Fare** and **Pclass**.

- Weak correlation between **Age** and **Survival**, but notable trends visible in visual analysis.

- **Heatmap and pairplot** suggest that certain groups (e.g., younger passengers in higher classes) had better survival chances.

6. **Data Quality Notes:**

   - Many features like **Cabin** may not be useful without imputation or categorization due to high missing values.

   - Some categorical columns like **Sex** and **Embarked** can be encoded for machine learning purposes.