EXPLORATORY DATA ANALYSIS (EDA) REPORT
TITANIC DATASET

OBJECTIVE:

To extract meaningful insights from the Titanic dataset using statistical summaries and visualizations. This includes identifying patterns, relationships, trends, and anomalies that help understand the survival of passengers.

TOOLS USED:

- Python
- Pandas
- Matplotlib
- Seaborn
- Jupyter Notebook

DATA OVERVIEW:

- **Total entries**: 891 passengers
- **Missing values**:
    - Age: ~19.9% missing
    - Cabin: >75% missing
    - Embarked: 2 missing

KEY FEATURES:

- Survived: Target variable (0 = No, 1 = Yes)
- Pclass: Passenger class (1st, 2nd, 3rd)
- Sex: Gender
- Age: Age of the passenger
- SibSp, Parch: Family relations
- Fare: Ticket fare
- Embarked: Port of embarkation

VISUAL ANALYSIS:

1. Survival Distribution

- **Countplot** showed more passengers **did not survive**.
- **Survival Rate**: ~38% survived, ~62% did not.

2. Survival vs Gender

- **Females had significantly higher survival rates** (~74%) compared to males (~19%).
- Gender was a strong predictor of survival.

3. Survival vs Passenger Class

- 1st Class passengers had the **highest survival rate**, followed by 2nd and 3rd.
- **Higher class = Higher survival**

4. Age Distribution

- Wide range of ages.
- Children had **better survival chances**.
- Many missing values in Age.

5. Fare vs Class

- Higher fare correlates with higher passenger class and survival.
- Boxplots showed **outliers** in fare values, especially for 1st class.

6. Correlation Heatmap

- Fare and Pclass were moderately correlated.
- Sex (after encoding) had a strong negative correlation with Survived.

7. Embarkation Point

- Most passengers boarded from **Southampton (S)**.
- Slightly higher survival rate from **Cherbourg (C)**.

8. Pairplot Observations

- Multivariate relationships visible:
    - Age and Fare had class-wise separation.
    - Sex and Pclass influenced survival significantly.

STATISTICAL SUMMARY:

Used .describe() and .info() to:

- Understand central tendencies (mean, median)
- Detect missing/null values
- Analyze standard deviation, min/max

KEY FINDINGS:

Insight	Description
Gender	Women were prioritized during evacuation – most women survived.
Class	Rich passengers (1st Class) had better survival chances.
Children	Young passengers had slightly better survival odds.
Fare	Higher fare indicated better class and hence, higher survival.
Embarked	Port of embarkation had a minor effect on survival.
Alone vs Family	Passengers with small families had slightly better survival. Solo travellers had lower survival rates.

CONCLUSION:

This EDA successfully revealed key survival patterns and relationships in the Titanic dataset. The findings highlight the influence of gender, class, and age on survival. Visualizations helped in understanding complex inter-variable interactions and set a foundation for building predictive models.