Exploratory Data Analysis (EDA) – Titanic Dataset

Objective: Extract insights using visual and statistical exploration.

Tools Used: Python (Pandas, Matplotlib, Seaborn)

Dataset Overview:

• 418 rows × 12 columns
• Target variable: **Survived** (0 = No, 1 = Yes)
• Features: Passenger details such as Pclass, Sex, Age, SibSp, Parch, Fare, Cabin, Embarked.

Missing Values:

• Age: 86 missing
• Fare: 1 missing
• Cabin: 327 missing (mostly empty, can be dropped or feature engineered)

Key Findings:

1. Majority of passengers were in 3rd class.
2. Women had higher survival rates compared to men.
3. Younger passengers (children) had better chances of survival.
4. Higher fares were positively correlated with survival (wealthier passengers had better survival rates).
5. Embarked port 'S' had the most passengers, but survival was higher for 'C'.

Visual Exploration Summary:

• Histograms revealed skewness in Fare distribution.
• Boxplots showed outliers in Fare and Age.
• Heatmap showed correlations: Pclass, Fare, and Sex were strongly related to survival.
• Bar plots highlighted clear differences in survival by gender and class.

Conclusion:

Survival on the Titanic was strongly influenced by socio-economic factors such as passenger class, gender, and fare paid. Women, children, and wealthier passengers had significantly higher chances of survival. This dataset is well-suited for classification modeling (e.g., Logistic Regression, Decision Trees, Random Forest).