

Task 5: Exploratory Data Analysis (EDA)

Internship: Data Analyst Internship

Tools Used: Python (Pandas, Matplotlib, Seaborn)

Dataset: Titanic Dataset (Kaggle)

■ Objective:

To perform Exploratory Data Analysis (EDA) on the Titanic dataset to identify patterns, trends, and relationships between features that affected passenger survival.

■■ Steps Followed:

1■■ **Import Libraries:** pandas, numpy, seaborn, matplotlib

2■■ **Load Dataset:** Read the CSV file and preview the data.

3■■ **Data Overview:** Used `.info()`, `.describe()`, and `.isnull()` to understand structure and missing values.

4■■ **Univariate Analysis:** Analyzed individual columns using histograms and boxplots.

5■■ **Bivariate Analysis:** Checked relationships (e.g., survival vs. gender/class).

6■■ **Correlation Analysis:** Used heatmap to study variable correlations.

7■■ **Data Cleaning:** Handled missing data and removed unnecessary columns.

■ Observations:

- Most passengers were aged 20–40 years.
- Fare distribution is skewed (few passengers paid very high fares).
- 38% passengers survived.
- Females had a higher survival rate.
- 1st-class passengers had better survival chances.
- Fare and survival are positively correlated.

■ Summary of Insights:

- Survival rate was higher among females and 1st-class travelers.
- Younger passengers had better survival chances.
- Fare and class strongly influenced survival probability.
- Data cleaning improved accuracy of insights.

■ Outcome: Gained hands-on experience in identifying patterns, trends, and correlations using Python visualization tools.