

Task 2: Exploratory Data Analysis (EDA) — Titanic Dataset

■ Objective

Perform detailed exploratory analysis on the Titanic dataset to uncover meaningful business insights using visualizations and statistical summaries.

■ Tools Used

Python: Pandas, Matplotlib, Seaborn

■ Steps Performed

1. Loaded dataset (`titanic.csv`) and checked structure using `info()`, `describe()`, `shape`.
2. Checked missing values using `isnull().sum()` and visualized them with a heatmap.
3. Univariate Analysis: Histograms for Age, Countplot for Gender, Boxplots for numeric columns.
4. Bivariate Analysis: Survival rate compared with Age, Pclass, and Gender using boxplots and barplots.
5. Correlation Heatmap: Showed numeric feature correlations.
6. Grouped analysis: Calculated average survival by Pclass and Gender.

■ Key Findings

- Females had significantly higher survival rates compared to males.
- Passengers in 1st Class had higher chances of survival compared to 2nd and 3rd Class.
- Younger passengers showed slightly better survival probability than older ones.
- Higher fare values correlated with better survival chances (proxy for class).

■ Business Summary

The Titanic EDA reveals clear demographic and socio-economic survival patterns. Women and children had priority during evacuation, and socio-economic class (Pclass) played a significant role in survival outcomes. These insights demonstrate how demographic and financial factors influenced survival, which can guide decision-making in risk analysis and safety planning for similar contexts.