

## Task 5: Exploratory Data Analysis – Titanic (Synthetic)

### Objective

Perform exploratory analysis to extract insights from a Titanic-like dataset using pandas, matplotlib, and seaborn.

### Dataset

Synthetic Titanic training-like data (800 rows). Columns include Survived, Pclass, Sex, Age, SibSp, Parch, Fare, Cabin, Embarked, etc.

### Key Steps

- Data overview with `.info()`/`.describe()`
- Missing values inspection and median imputation for Age
- Univariate plots (hist, boxplot)
- Bivariate analysis (group survival rates)
- Correlation heatmap and pairplot for multivariate glimpse

### Highlights

- Females show higher survival rates than males
- Survival is higher in 1st class than 2nd, and 2nd higher than 3rd
- Higher fare has a mild positive association with survival
- Fare contains a few high-value outliers; Age requires imputation

### Next Steps

- Try logistic regression / tree models for predictive insights
- Engineer features (e.g., FamilySize, Title extracted from Name)
- Use stratified train/validation splits for robust checks