

# Exploratory Data Analysis (EDA) Report

This report presents an Exploratory Data Analysis (EDA) on the Titanic dataset. The objective is to extract insights, identify trends, and understand relationships between features and passenger survival.

## Data Overview:

- Total Rows: 891
- Total Columns: 12
- Target Variable: Survived (0 = No, 1 = Yes)
- Missing Values: Age (177), Cabin (~687), Embarked (2)

## Univariate Analysis:

- Age: Majority between 20–40 years.
- Fare: Right-skewed; few passengers paid very high fares.
- Sex: More males than females.
- Pclass: Most passengers in 3rd class.
- Survival: About 38% survived.

## Bivariate Analysis:

- Females had significantly higher survival rates.
- 1st class passengers had higher survival than 3rd class.
- Younger children had higher chances of survival.
- Higher fares were associated with higher survival.

## Multivariate Analysis:

- Combination of Pclass and Sex shows strong survival differences.
- Titles like Mrs/Miss show higher survival compared to Mr.
- Large family size reduces survival chances; small families fare better.

## Handling Missing Values (suggestions):

- Age: Impute using median or by Title.
- Cabin: Too many missing, may drop or extract deck info.
- Embarked: Impute with mode (most common port).

## Summary of Insights:

1. Females had higher survival probability.
2. Wealthier passengers (higher class, higher fare) survived more.
3. Age distribution shows children had better chances.
4. Family size played a role in survival.
5. Missing data handling is essential before modeling.

*Prepared as part of Data Analyst Internship — Task 5.*