

Title: Exploratory Data Analysis – Titanic Dataset

Author: *Anshu*

Date: *12-08-2025*

Objective

Perform Exploratory Data Analysis (EDA) on the Titanic dataset to uncover patterns, relationships, and insights that may influence survival rates.

Dataset Overview

- Source: Seaborn Titanic dataset (`sns.load_dataset('titanic')`)
 - Shape: ~891 rows × 15 columns
 - Features: Passenger demographics, ticket details, class, fare, survival, etc.
 - Missing Values: Age (~20%), Cabin (77%), Embarked (~0.2%)
-

Analysis Performed

Univariate Analysis

- Age distribution
- Survival count
- Gender distribution

Bivariate Analysis

- Survival by gender
- Age vs Passenger Class
- Survival vs Fare

Multivariate Analysis

- Correlation heatmap
- Pairplot (Age, Fare vs Survival)

Key Findings

- Majority Did Not Survive: ~62% of passengers perished.
 - Gender Impact: Females had a significantly higher survival rate.
 - Class Influence: First-class passengers more likely to survive.
 - Fare Relation: Higher fares correlated with higher survival rates and class.
 - Age Profile: Most passengers were between 20–40 years.
-

Recommendations / Insights

- Passenger class and gender were critical survival determinants.
- Fare is a weak but positive predictor for survival.
- Missing Cabin values may hold hidden patterns; imputation or feature engineering could help.

End of Report