

Titanic Dataset Exploratory Data Analysis Report

What is EDA?

Exploratory Data Analysis (EDA) is a process of examining datasets to summarize their main characteristics, often using statistical graphics and other data visualization techniques. EDA helps uncover patterns, detect anomalies, and test assumptions using descriptive statistics.

Steps Performed:

1. Data Import & Cleaning

- Dataset shape: 891 rows, 15 columns
- Duplicates removed

2. Data Type Handling

- Converted 'age' to numeric for consistency

3. Univariate Analysis

- Age and Fare distributions analyzed
- Age distribution is right-skewed

4. Bivariate Analysis

- Count plot of Survival by Passenger Class
- Fare distribution by Class using Boxplot
- Survival rate by Age Group

5. Categorical Feature Analysis

- Analyzed embarkation point (S, C, Q)

6. Correlation Analysis

- Computed correlation between age, fare, survival, and class

Key Insights:

- Most passengers were in 3rd class.
- 1st class passengers had higher survival rates.
- Children had the highest survival rate among age groups.
- Seniors had the lowest survival rate.
- Fare is positively correlated with survival.
- Class is negatively correlated with survival.
- Most passengers boarded from Southampton.

Next Steps:

- Handle missing values (age, deck, embarked).
- Explore interaction between sex and survival.
- Analyze family size vs. survival.
- Use more advanced visualizations like heatmaps or pairplots.
- Build predictive models using logistic regression or decision trees.