

Exploratory Data Analysis (EDA) Report

1. Objective

Understand the structure, relationships, and patterns in the Titanic dataset using statistical summaries and visualizations. The goal is to gain insights that can help in feature engineering and predictive modeling.

2. Dataset Used

Titanic Dataset

Widely used for binary classification problems (Survived or Not). It contains demographic and travel information for passengers aboard the RMS Titanic.

3. Steps Performed

1. Generated Summary Statistics

- Used `df.describe()` and `df.info()` to explore:
 - Central tendencies (mean, median)
 - Spread (standard deviation, min, max)
 - Missing values

Key Insights:

- Age: Mean approx. 29.7, Std Dev approx. 14.5, Missing approx. 177 entries
- Fare: High skew, wide range (0 to 512)
- Cabin: Mostly missing, dropped
- Sex, Pclass, and Embarked are key categorical features

2. Visualized Distributions with Histograms and Boxplots

- Histograms showed:
 - Fare is right-skewed

Exploratory Data Analysis (EDA) Report

- Age is somewhat normally distributed but with long tails
- Boxplots revealed:
 - Outliers in Fare ($> \$300$)
 - Some potential outliers in Age

3. Analyzed Feature Relationships

- Created a correlation matrix using Seaborn heatmap:
 - Pclass vs Fare: Negative correlation (~ -0.55)
 - Fare vs Survived: Weak positive (~ 0.26)
- Used pairplots for Age, Fare, Pclass, Survived:
 - Showed visually separable clusters based on survival

4. Identified Patterns, Trends, and Anomalies

- Survival by Gender: Females had a significantly higher survival rate
- Survival by Pclass: 1st class passengers more likely to survive
- Age vs Survival: Children had better survival rates
- Fare: Passengers who paid higher fares tended to survive more

5. Made Basic Feature-Level Inferences

Feature-Level Inferences:

- Sex: Strong survival predictor - most females survived
- Pclass: Higher class better rescue priority
- Age: Children prioritized, middle-aged passengers more vulnerable
- Fare: High fare correlated with class & survival
- Embarked: Port of embarkation slightly influenced survival, especially in 1st class