

Exploratory Data Analysis Report

Dataset: Titanic Dataset

Tools Used: Python, Pandas, Seaborn, Matplotlib

Author: [Your Name]

Date: [Insert Date]

1. Objective

The objective of this EDA is to perform a thorough statistical and visual investigation of the Titanic dataset to understand the data structure, identify key patterns, detect anomalies, and derive meaningful insights, especially focusing on the factors that influenced passenger survival.

2. Dataset Overview

The Titanic dataset contains information on passengers aboard the Titanic, including demographic details, ticket fare, travel class, and survival status.

Key Features:

- **Categorical:** sex, embarked, class, who, deck
- **Numerical:** age, fare, sibsp, parch
- **Target Variable:** survived (0 = No, 1 = Yes)

Basic Info:

- Total Rows: 891
 - Missing Data: Found in age, embarked, and deck
-

3. Data Summary

► Descriptive Statistics

- **Average Age:** 29.7 years
- **Average Fare:** \$32.20
- **Survival Rate:** ~38.4%

► Missing Values

Column	Missing Values
--------	----------------

age	~19%
-----	------

embarked	<1%
----------	-----

deck	~77%
------	------

4. Univariate Analysis

► Gender Distribution

- Males: 577
- Females: 314
- Survival was higher among females (~74%).

► Age Distribution

- Skewed slightly right; majority between 20–40 years
- Children had better survival rates than adults

► Fare Distribution

- Most passengers paid under \$50
- Some outliers with extremely high fares (> \$500)

Outlier and Anomaly Detection

- **Fare:** Several high-ticket prices (VIPs or family groups?)
 - **Age:** Some infants and elderly over 70 (rare cases)
 - Dealt with by visualizing using boxplots and deciding on imputation (age median fill)
-

8. Key Findings

- **Gender and class** were the strongest indicators of survival.
 - **Females and 1st class passengers** had the highest survival rates.
 - **Young children** were given priority in rescue.
 - **Fare price** had a positive correlation with survival.
 - **Deck** column has too many missing values for meaningful analysis.
-

9. Conclusion

The EDA successfully provided insights into which factors most influenced survival on the Titanic. These findings not only offer a data-driven understanding of the historical event but also lay the groundwork for predictive modeling.