

# Exploratory Data Analysis (EDA) Report

Dataset: Titanic Survival Dataset ( [Kaggle Link](#) )

## 1. Introduction & Objective

The Titanic dataset is one of the most popular datasets used in data science and machine learning. It contains passenger information such as age, sex, class, and survival status.

### **Objective:**

The purpose of this analysis is to explore the dataset, identify patterns, trends, and anomalies, and gain insights into the factors that influenced passenger survival during the Titanic disaster.

## 2. Dataset Overview

**Source:** Titanic train dataset

**Target variable:** Survived (0 = Did not survive, 1 = Survived)

### **Features include:**

- a. PassengerId – Unique ID of passenger
- b. Survived – Target variable
- c. Pclass – Ticket class (1 = 1st, 2 = 2nd, 3 = 3rd)
- d. Name, Sex, Age
- e. SibSp, Parch – Number of family members aboard
- f. Ticket, Fare
- g. Cabin, Embarked – Port of embarkation

### **Initial Checks:**

- a. `.info()` shows column data types and missing values.
- b. `.describe()` provides summary statistics.
- c. `.value_counts()` highlights categorical distributions.

### **Key Findings from Overview:**

- a. Missing values in Age, Cabin, and some in Embarked.
- b. Categorical variables: Sex, Pclass, Embarked.
- c. Continuous variables: Age, Fare.

## 3. Descriptive Statistics

### **`.describe()` results (summary):**

- a. **Age:** Mean  $\approx$  29.7 years, Range 0.4 – 80.
- b. **Fare:** Mean  $\approx$  32, Range 0 – 512, skewed distribution.

- c. **Pclass:** More passengers in 3rd class.

**.value\_counts() examples:**

- a. **Survived:** About 38% survived, 62% did not.
- b. **Sex:** Male passengers were more common than females.
- c. **Embarked:** Most passengers boarded at Southampton (S).

## 4. Visual Analysis & Insights

### 4.1 Histograms

- **Age distribution:** Most passengers between 20–40 years.
- **Fare distribution:** Right-skewed; few passengers paid very high fares.

*Observation:* Younger adults formed the majority. A small group of high-paying passengers existed, likely first-class.

### 4.2 Boxplots

- **Age vs Sex:** Males and females had similar age ranges.
- **Fare vs Pclass:** 1st class passengers paid significantly higher fares.

*Observation:* Clear socio-economic divide; wealthier passengers more likely to afford survival advantages.

---

### 4.3 Countplots

- **Survival Counts:** More deaths (0) than survivals (1).
- **Survival by Sex:** Higher survival rate for females compared to males.
- **Survival by Pclass:** 1st class had highest survival, 3rd class the lowest.

*Observation:* Gender and ticket class strongly influenced survival chances.

---

### 4.4 Correlation Heatmap

- Positive correlation: Fare and Pclass (higher class, higher fare).
- Strong negative correlation: Pclass and Survived.
- Moderate correlation: Sex (encoded as numeric) and Survived.

*Observation:* Socio-economic status and gender were significant predictors of survival.

---

### 4.5 Pairplot (Age, Fare, Pclass vs Survival)

- Survivors tended to be **younger** and **paid higher fares**.
- Clear separation between 1st and 3rd class survival rates.

*Observation:* Wealthier and younger individuals had better survival chances.

## 5. Key Findings & Summary

- a. Survival Rate:** Only ~38% of passengers survived.
- b. Gender Factor:** Women had significantly higher survival rates than men.
- c. Class Factor:** 1st class passengers survived at higher rates compared to 2nd and 3rd class.
- d. Age Factor:** Children and younger passengers had better survival chances.
- e. Fare Factor:** Higher fares (indicator of wealth/class) correlated with survival.
- f. Embarkation:** Majority boarded from Southampton; survival distribution varied across ports.
- g. Missing Data:** Age and Cabin have missing values that may need imputation.

## 6. Conclusion

The EDA reveals that **gender, class, and fare** were critical determinants of survival on the Titanic. Females and passengers from higher classes had a distinct advantage. This aligns with the "women and children first" evacuation protocol as well as socio-economic advantages.