

## Titanic EDA – PDF Report Structure

Title: Exploratory Data Analysis - Titanic Dataset

Author: **Ekta Sinha**

Date: 11-08-2025

Objective: The objective of this analysis is to explore the Titanic dataset using statistical summaries and visualizations to identify trends, relationships, and key factors affecting survival.

## 1. Dataset Overview Dataset: Titanic (train.csv)

Source: Kaggle Titanic: Machine Learning from Disaster

Number of Rows: (after loading & cleaning)

Number of Columns: (after loading & cleaning)

Data Dictionary (Key Columns): Survived – Survival (0 = No, 1 = Yes)

Pclass – Passenger Class (1 = 1st, 2 = 2nd, 3 = 3rd)

Sex – Gender Age – Age in years SibSp – Number of siblings/spouses aboard

Parch – Number of parents/children aboard Fare – Ticket price Embarked – Port of Embarkation.

## 2. Data Inspection df.info() – Data types and null values.

df.describe() – Statistical summary.

value\_counts() for categorical columns.

Observations: Age column contains missing values. Cabin has many missing values (dropped/ignored in analysis). Most passengers are in 3rd class.

### 3. Univariate Analysis

#### 4.1 Survival Count Graph: Countplot of Survived

Observation: About 38% of passengers survived, while 62% did not.

#### 4.2 Passenger Class Distribution Graph: Countplot of Pclass

Observation: Majority of passengers were in 3rd class.

4.3 Gender Distribution Graph: Countplot of Sex  
Observation: There were more male passengers than female passengers.

### 4. Bivariate Analysis

5.1 Survival by Gender Graph: Barplot of Survived vs. Sex  
Observation: Female passengers had a much higher survival rate than males.

5.2 Survival by Passenger Class Graph: Barplot of Survived vs. Pclass  
Observation: 1st class passengers had the highest survival rate; 3rd class the lowest.

5.3 Age Distribution by Survival Graph: Boxplot of Age vs. Survived  
Observation: Younger passengers had a slightly better chance of survival.

## **5. Multivariate Analysis**

6.1 Pairplot (Survived, Age, Fare) Observation: Survivors tended to have paid higher fares; Age vs. Fare had no strong correlation.

6.2 Correlation Heatmap Observation: Survival negatively correlated with Pclass and positively correlated with Fare.

## **6. Summary of Key Findings :**

Survival rate was higher among females than males.

1st class passengers had the highest survival rate, followed by 2nd class, then 3rd class.

Higher fare amounts are associated with higher survival probability.

Most passengers were young adults (20 to 40 years old).

Age had a weaker correlation with survival compared to passenger class and fare.

## **7. Multivariate Analysis**

6.1 Pairplot (Survived, Age, Fare) Observation: Survivors tended to have paid higher fares; Age vs. Fare had no strong correlation.

6.2 Correlation Heatmap Observation: Survival negatively correlated with Pclass and positively correlated with Fare.

## **8. Summary of Key Findings :**

Survival rate was higher among females than males.

1st class passengers had the highest survival rate, followed by 2nd class, then 3rd class.

Higher fare amounts are associated with higher survival probability.

Most passengers were young adults (20 to 40 years old).

Age had a weaker correlation with survival compared to passenger class and fare.



## **Conclusion**

This EDA reveals that passenger class, gender, and fare were strong determinants of survival. Such insights can guide predictive modeling for the Titanic dataset.