



EDA of Titanic Survival Dataset

Title Page

Project : Exploratory Data Analysis (EDA)

Dataset : Titanic Survival Dataset

Author : Kevin Lazarus

Tools Used : Python (Pandas, Seaborn, Matplotlib)

Date : July 1, 2025

Objective

The objective of this analysis is to explore the Titanic passenger dataset using statistical and visual techniques. The goal is to identify key factors that influenced passenger survival and to uncover meaningful patterns or trends that could inform future predictions.

Dataset Description

The dataset contains data on **891** passengers who were aboard the Titanic. Each record includes details such as:

PassengerId : Unique ID

Survived : 0 = No, 1 = Yes

Pclass : Passenger class (1 = First, 2 = Second, 3 = Third)

Name, Sex, Age

SibSp/Parch : Number of siblings/spouses and parents/children aboard

Ticket, Fare

Cabin : Cabin number (many missing)

Embarked : Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton)

Missing Values

* Age : 177 missing

* Cabin : 687 missing (dropped)

* Embarked : 2 missing

✂ Methods and Tools Used

Python Libraries:

- * `'pandas'` for data manipulation
- * `'seaborn'` and `'matplotlib'` for visualizations

Techniques Used:

- * Summary statistics (`'describe()'`, `'value_counts()'`)
- * Univariate & Bivariate visual analysis
- * Correlation heatmaps
- * Missing value treatment

Visualizations with Observations

1. Survival Distribution

- * About 38% of the passengers survived.

2. Passenger Class and Survival

1st class had the highest survival rate.

3rd class had the highest mortality.

3. Gender and Survival

Females had significantly higher survival rates than males.

4. Age Distribution

- * Age is right-skewed; most passengers were between **20-40 years old**.
- * Children (<10) had relatively higher chances of survival.

5. Fare Distribution

- * Highly right-skewed.
- * Higher fare generally corresponded with higher survival rates.

6. Correlation Heatmap

- * `'Fare'` positively correlated with `'Survived'`.
- * `'Pclass'` negatively correlated with `'Survived'`.

7. Embarked Port

- * Most passengers boarded from **Southampton (S)**.
- * Highest survival rate among those who embarked from **Cherbourg (C)**.

✅ Key Findings Summary

Insight	Observation
Gender	Female passengers more likely to survive
Class	First-class passengers had better survival rates
Age	Children had higher survival chances
Fare	Higher-paying passengers were more likely to survive
Cabin	Dropped due to >75% missing values

📌 Conclusion / Next Steps

This EDA reveals that **gender**, **passenger class**, and **fare** are significant predictors of survival. Based on the trends, a model can now be trained to predict survival based on features like `Sex`, `Pclass`, `Fare`, and `Age`.