# Titanic Dataset — Detailed Report of Findings

## 1. Dataset Overview

The Titanic dataset contains passenger-level data such as age. gender. class. embarkation point, fare. and survival status. It is Widely used Ior predictive modeling and exploratory data analysis due to its balanced mix ot numerical and categorical variables.

## 2. Data Quality Summary

The dataset has missing values in columns like 'Age', 'Cabin', and 'Embarked'. 'Cabin' has the highest percentage 01 missing entries. Numerical fields such as •Age' and •Fare' contain outliers. indicating variance in passenger demographics and ticket pricing.

## 3. Key Insights from EDA

- Women had significantly higher survival rates compared to men.
- First-class passengers survived more frequently than second- and third-class passengers.
- Younger passengers, especially children, showed better survival chances.
- Higher tares were with greater Survival odds, reflecting advantages.

## 4. Distribution Observations

The Age distribution skews toward younger adults between ages 2040. Fare distribution shows many low-value entries with few extremely high fares. Categorical distrü)utions in&cate more male passengers and more third-class travelers.

## 5. Missing Values Handling Strategy

Age can be imputed using median ages or group medians based on Sex and Pclass, Embarked values are typically tilled with the mode Cabin data can either dropped or simplified

## 6. Manually Written-Looking Observations

When reviewing the Titanic dataset. it's easy to notice that survival wasn't random. Women and tirst-class passengers clearly had outcomes. Many cabin entries are missing. Likely because passengers in lower classes didn't have assigned cabins. Fare amounts range Widely, showing differences in economic status. Overall, the dataset highlights strong inequalities that influenced survival patterns during the tragedy.