

# Titanic Survival Prediction - Exploratory Data Analysis (EDA)

## Objective

Analyze Titanic dataset to understand data structure, handle missing values, explore survival patterns, and identify factors influencing survival.

## Data Overview

- Source: Titanic Kaggle competition
- Rows & Columns: As per train.csv
- Key variables: Survived, Pclass, Sex, Age, Fare, Embarked
- Missing Values: 'Cabin' (large), 'Age' (some), 'Embarked' (few)

## Key Findings

1. **Survival Rate**: ~38% survived, 62% did not.
2. **Gender**: Females had much higher survival rate.
3. **Class**: 1st class survival > 2nd > 3rd.
4. **Age**: Survivors slightly younger; children had better odds.
5. **Embarked**: Port C had highest survival, Port S had most passengers but lower survival.
6. **Fare**: Higher fares correlated with higher survival and better classes.

## Data Cleaning

- Filled missing Age with median.
- Filled missing Embarked with mode.
- Dropped Cabin, PassengerId, Name, Ticket.

## **Observations**

- Fare distribution is highly skewed with extreme outliers.
- Higher class and fare, female gender, and younger age improved survival odds.
- Outliers in Fare likely represent luxury passengers.

## **Recommendations**

- Consider feature engineering (e.g., family size, title extraction from names).
- Handle skewed distributions for modeling.