# TITANIC

## Introduction

The Titanic disaster of 1912 is one of the most well-known maritime tragedies in history. Analyzing the data related to the passengers can provide valuable insights into survival rates, passenger demographics, and ticket pricing trends. This project visualizes various aspects of the Titanic dataset using Python and Matplotlib.

## Objective

The main objectives of this project are:

- To analyze the distribution of passengers based on class.
- To examine the survival rate of passengers.
- To study the age distribution of passengers.
- To explore the relationship between age and fare paid.
- To analyze survival rates based on gender.

# Dataset Description

The dataset used for this project contains information about Titanic passengers, including:

- **Pclass**: Ticket class (1st, 2nd, 3rd)
- **Survived**: Survival status (0 = No, 1 = Yes)
- **Sex**: Gender of passengers
- **Age**: Age of passengers
- **Fare**: Price paid for the ticket
- **Embarked**: Port of embarkation
- **SibSp & Parch**: Number of family members onboard

# Data Quality Observations

- The Age column has 177 missing values.
- The Cabin column has a significant number of missing values (687 out of 891).
- The Embarked column has 2 missing values.

# Data Visualization and Analysis

1. Passenger Count by Class

- A bar chart is used to show the number of passengers in each class.
- It reveals that the majority of passengers belonged to the third class.

2. Survival Distribution

- A bar chart represents the number of passengers who survived versus those who did not.
- The chart shows that a significant number of passengers did not survive.

3. Age Distribution

- A histogram is used to display the distribution of passengers' ages.
- The majority of passengers were between 20 to 40 years old.

4. Survival Rate by Gender

- A bar chart shows survival counts for male and female passengers.
- Females had a significantly higher survival rate than males.

5. Relationship Between Age and Fare

- A scatter plot is created to analyze the relationship between a passenger's age and the fare they paid.
- The plot indicates that higher fares were paid mostly by younger and first-class passengers.

# Conclusion

The visualization of the Titanic dataset provides meaningful insights into passenger demographics and survival trends. The analysis highlights:

- The dominance of third-class passengers.
- A lower survival rate overall.
- A diverse age range among passengers.
- Females had a much higher chance of survival.
- First-class passengers paid higher fares, with some outliers paying exceptionally high amounts.

This project demonstrates how data visualization techniques can be effectively used to explore and understand historical datasets.

## Future Scope

- Further analysis can be performed on family size and its impact on survival.
- Machine learning models can be applied to predict survival chances.
- Additional features such as embarkation points and social status can be analyzed for deeper insights.