

## Titanic Dataset Analysis

### Introduction

Titanic was a famous luxury ship sailing from Southampton, England to New York city, however, it never made it to its destination. It sank on April 15, 1912, when the ship collided with an iceberg in the North Atlantic Ocean, taking with it more than 1,500 lives, including passengers and crew.

### Project information

The goal of this project is to identify the patterns of the survival passengers and understanding the relationship with their age, gender, and cabin class.

### About the data

Data was downloaded directly from [Kaggle](https://www.kaggle.com/c/titanic). It included 2 datasets gender\_submission.csv and train.csv, which are combined into one file name titanic.csv and we will use this combined file for our analysis.

### Visualization and Insights

Figure 1: Survival rate by Passenger Class

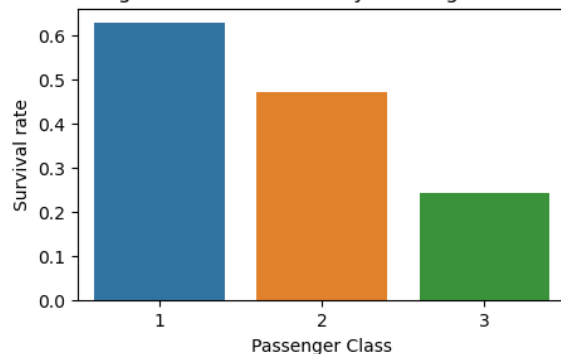


Figure 2: Survival Rate by Gender

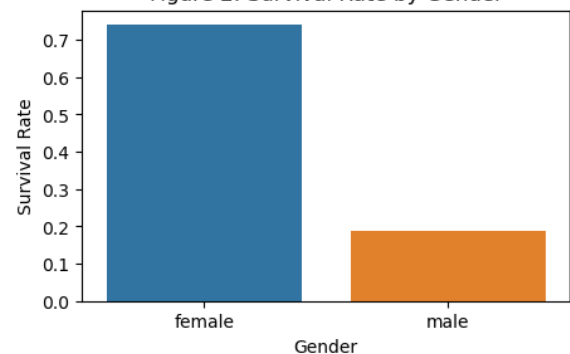


Figure 3: Age Distribution for Survivors and Non-Survivors

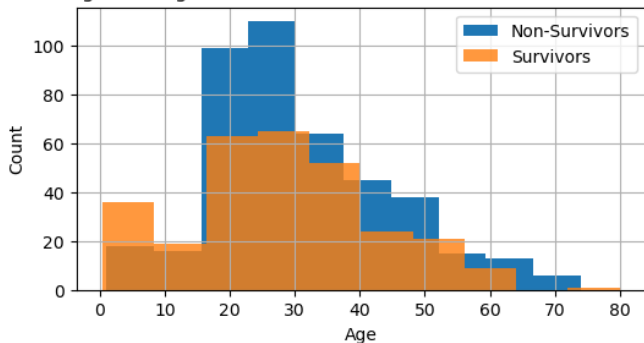
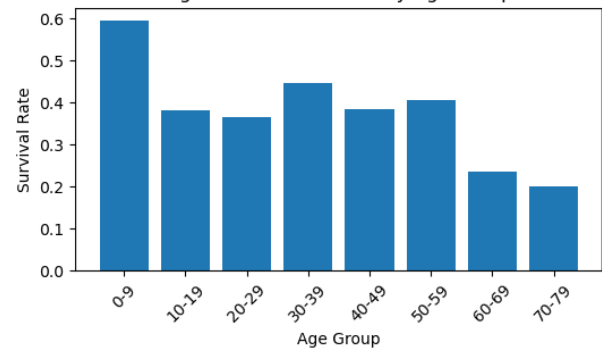


Figure 4: Survival Rate by Age Group



Among 891 passengers, the survival rate is the highest when they are from first class cabin at 63%, and that ratio gets lower to the second (47%) and third class (24%). Female passengers also have a significantly higher survival rate than male – 74% and 19% respectively.

The ratio of surviving infants and kids under 10 years old are the highest at 60%, possibly due to their priority to be rescued first. In contrast, senior passengers (above 60 years old) get the lowest survivor rate among all passengers. Young Adults (from 20 – 29 years old) were the majority age group of the passengers but only 36% of them survived, which is also the second lowest ratio.