

## Titanic Dataset – Introduction

**Brief Summary:** A fairly famous dataset that contains info about every passenger on the Titanic and whether or not they survived. The target variable is who lived on the ship, and who died. This is a useful dataset for learning different data analysis techniques because it contains a variety of data types as well as interesting opportunities for feature engineering and visual exploration.

- **Survived:** A 0 or 1, passenger gets a 1 if they survived, 0 if they didn't
- **PassengerId:** Numerical value that acts as a row index for each passenger listed
- **Pclass:** Passenger class, corresponds to what level of traveler they were. Takes values 1, 2, 3
- **Name:** Name printed on the ticket
- **Sex:** Gender of the passenger
- **Age:** Age of the passenger
- **SibSp:** Number of siblings on the board with passenger
- **Parch:** Number of parents or children on board with passenger
- **Ticket:** Unique identifier for every ticket bought per room
- **Fare:** Price paid for each ticket
- **Cabin:** Cabin number for every passenger
- **Embarked:** Abbreviation for port that people used to get on the ship. Takes on the values 'S', 'C', and 'Q'.