1. Business Understanding
   1. Objective

Predict the survival on the Titanic

* 1. Description

The sinking of the Titanic is one of the most infamous shipwrecks in history. On Apr 15,1912, during her maiden voyage, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. This sensational tragedy shocked the international community and led to better safety regulations for ships.

One of the reasons that the shipwreck led to such loss of life was that there were not enough lifeboats for the passengers and crew. Although there was some element of luck involved in surviving, some groups of people were more likely to survive than others, such as women, children, and the upper-class.

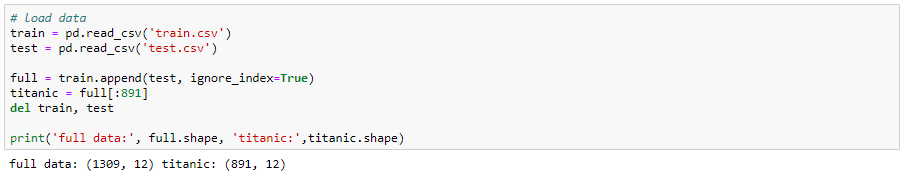
In this project, I will complete the analysis of what sorts of people were more likely to survive. In particular, I will apply machine learning to predict which passengers survived the tragedy.

1. Data Understanding
   1. Import library

First of all, I need import python libraries containing the necessary functionality will need.

* 1. Load data

Next, load train.csv and test.csv dataset, append train and test dataset as full dataset. Have a peek at the dataset size.



* 1. Statistical summaries and visualization

To understand the data, I am going to consider some key facts about various variables including their relationship with the target variables, i.e. survived.

Here is the variable descriptions:

Age: passenger’s age

Cabin: cabin

Embarked: port of embarkation

Fare: fare

Name: passenger’s name

Parch: number of parents/children aboard

Pclass: passenger’s class

Sex: passenger’s sex

SibSp: number of sibling/spouses aboard

Survived: survived (0) or died (1)

Ticket: ticket number

