

# Problem Statement

## About Data Set -

The data has been split into two groups:

Training Data and Testing data.

The training data set should be used to build your machine learning model.

For the training set data, we provide the outcome (also known as “ground truth”)

For each passenger.

Your model will be based on “features” like passenger, gender and class.

The test data set should be used to see how well your model performs on unseen data.

Even for the test data set, we are providing the ground truth for each passenger, so that you can compute the required deliverables yourself. But it is your job to predict these outcomes.

For each passenger in the test data use the model to predict whether or not they survived the sinking of the titanic.

## Variable Notes-

P-class: A proxy for socio-economic status (SES)

1<sup>st</sup> = Upper

2<sup>nd</sup> = Middle

3<sup>rd</sup> = Lower