# Data Dictionary

* Survival- Survival 0 = No, 1 = Yes
* Pclass- Ticket class 1 = 1st, 2 = 2nd, 3 = 3rd
* sex
* Age- Age in years
* Sibsp- # of siblings / spouses aboard the Titanic
* Parch- # of parents / children aboard the Titanic
* Ticket- Ticket number
* Fare- Passenger fare
* Cabin- Cabin number
* Embarked- Port of Embarkation(C = Cherbourg, Q = Queenstown, S = Southampton)

# INSIGHTS

* All variables are categorical variables except for age.
* The death variable after converting date died to a categorical variable is highly skewed towards ‘no’ which is expected as the mortality rate of covid 19 is low.
* Sex variable is balanced with male and female being the two categories.
* Co – morbidities like pneumonia, asthma, copd, cardiovascular and renal chronic are skewed towards ‘no’.
* Tobacco consumption, diabetes, hypertension and obesity are skewed towards ‘no’ but have a substantial presence in ‘yes ’ as well.
* The distribution of ‘age’ is normal.
* DecisionTree, Random Forest, Extra trees and GaussianNB algorithms were used for prediction. All 4 models yielded an accuracy of more than 90 %