**Introduction**

This is my first competition submission, where i put together what i have learned about logit models to predict which passengers aboard the titanic survived the shipwreck or not. I came into the competition knowing the sex, age, and ticket class contributed to whether a person survived or not, so i jumped straight into processing the data to fit the model.

**Table of Content**

1. [The Challenge](http://localhost:8888/notebooks/Desktop/Data_Science%20projects/%5BKaggle%5D%20%20Titanic%20-%20Machine%20Learning%20from%20Disaster/Code.ipynb#cell1)
2. [Variables Overview](http://localhost:8888/notebooks/Desktop/Data_Science%20projects/%5BKaggle%5D%20%20Titanic%20-%20Machine%20Learning%20from%20Disaster/Code.ipynb#cell2)
3. [Importaing relevant libraries](http://localhost:8888/notebooks/Desktop/Data_Science%20projects/%5BKaggle%5D%20%20Titanic%20-%20Machine%20Learning%20from%20Disaster/Code.ipynb#cell3)
4. [importing dataset](http://localhost:8888/notebooks/Desktop/Data_Science%20projects/%5BKaggle%5D%20%20Titanic%20-%20Machine%20Learning%20from%20Disaster/Code.ipynb#cell4)
5. [Preprocessing dataset](http://localhost:8888/notebooks/Desktop/Data_Science%20projects/%5BKaggle%5D%20%20Titanic%20-%20Machine%20Learning%20from%20Disaster/Code.ipynb#cell5)
   * Dealing with missing values
   * Exploring The descriptive statistics
   * Creating binary columns
   * Creating dummies for categorical variables
6. [Logit Regression Model](http://localhost:8888/notebooks/Desktop/Data_Science%20projects/%5BKaggle%5D%20%20Titanic%20-%20Machine%20Learning%20from%20Disaster/Code.ipynb#cell6)
   * Declaring depedent and indepedent Variable
   * Train\_Test\_Split data
   * Fitting Logit Model
   * Logit Regression Results
7. [Accuracy Score of model](http://localhost:8888/notebooks/Desktop/Data_Science%20projects/%5BKaggle%5D%20%20Titanic%20-%20Machine%20Learning%20from%20Disaster/Code.ipynb#cell7)
8. [Testing](http://localhost:8888/notebooks/Desktop/Data_Science%20projects/%5BKaggle%5D%20%20Titanic%20-%20Machine%20Learning%20from%20Disaster/Code.ipynb#cell8)
   * Importing Test Data
   * Filter data to match inputed data
   * Creating Predicated data with test data

**The Challenge**

The sinking of the Titanic is one of the most infamous shipwrecks in history.

On April 15, 1912, during her maiden voyage, the widely considered “unsinkable” RMS Titanic sank after colliding with an iceberg. Unfortunately, there weren’t enough lifeboats for everyone onboard, resulting in the death of 1502 out of 2224 passengers and crew.

While there was some element of luck involved in surviving, it seems some groups of people were more likely to survive than others.

In this challenge, we ask you to build a predictive model that answers the question: “what sorts of people were more likely to survive?” using passenger data (ie name, age, gender, socio-economic class, etc).

**Variables**

**survival** - 0 = No, 1 = Yes

**pclass** - Ticket class 1 = 1st, 2 = 2nd, 3 = 3rd

**sex** - Sex M or F

**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

**pclass** - A proxy for socio-economic status (SES) 1st = Upper 2nd = Middle 3rd = Lower

**age** - Age is fractional if less than 1. If the age is estimated, is it in the form of xx.5

**sibsp:**- The dataset defines family relations in this way... **Sibling** = brother, sister, stepbrother, stepsister **Spouse** = husband, wife (mistresses and fiancés were ignored)

**parch** - The dataset defines family relations in this way... **Parent** = mother, father **Child** = daughter, son, stepdaughter, stepson Some children travelled only with a nanny, therefore parch=0 for them.pclass: A proxy for socio-economic status (SES) 1st = Upper 2nd = Middle 3rd = Lower