

1. Obtain data (and explanation of the data) can be obtained from:
<https://www.kaggle.com/datasets/shitaljagtap/titanic-dataset>

2. Save the csv file as .xls named as titanic_data.xls

3. We have completed the assignment in Ubuntu 20.4. to set up first we needed to select the folder where data will be stored.

```
cd ~/python_project
```

4. Create a virtual environment for the project

```
virtualenv python_project_env
```

5. Activate the bin for python

```
source python_project_env/bin/activate
```

6. Finally run the jupyter notebook to create the new note book or open an existing one.

```
jupyter notebook
```

7. Below needed to be installed for regression analysis

```
pip3 install -U scikit-learn scipy matplotlib
```

Data Description

(from <https://www.kaggle.com/c/titanic>)

- **survival:** Survival (0 = No; 1 = Yes)
- **pclass:** Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)
- **name:** Name
- **sex:** Sex
- **age:** Age
- **sibsp:** Number of Siblings/Spouses Aboard
- **parch:** Number of Parents/Children Aboard
- **ticket:** Ticket Number
- **fare:** Passenger Fare
- **cabin:** Cabin
- **embarked:** Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)

Special Notes:

- Pclass is a proxy for socio-economic status (SES) 1st ~ Upper; 2nd ~ Middle; 3rd ~ Lower
- Age is in Years; Fractional if Age less than One (1) If the Age is Estimated, it is in the form xx.5

With respect to the family relation variables (i.e. sibsp and parch) some relations were ignored. The following are the definitions used for sibsp and parch.

- **Sibling:** Brother, Sister, Stepbrother, or Stepsister of Passenger Aboard Titanic
- **Spouse:** Husband or Wife of Passenger Aboard Titanic (Mistresses and Fiances Ignored)
- **Parent:** Mother or Father of Passenger Aboard Titanic
- **Child:** Son, Daughter, Stepson, or Stepdaughter of Passenger Aboard Titanic

Other family relatives excluded from this study include cousins, nephews/nieces, aunts/uncles, and in-laws. Some children travelled only with a nanny, therefore parch=0 for them. As well, some travelled with very close friends or neighbors in a village, however, the definitions do not support such relations.

1. what are the ages that survived most?

1. was there any class identification to survive?
2. which gender survived most?
3. gender and age combination survival rate..
4. class,gender and age another combination