CS 663 - Machine Learning Spring, 2020

Lab 02 - Titanic

The RMS Titanic is an infamous passenger ship which was built in 1912 and sank on its maiden voyage. A number of passengers survived its sinking; however, the probability of survival depended on the fare paid by the passenger.

In this lab, your task is to show the relationship between the fare paid and survival from the ship's sinking. This lab tests your ability to import and manipulate data and to display it using a graphing tool.

Data

The data for this lab comes from titanic.csv. This is available from several locations, including:

- Micheleen Harris' GitHub (https://gist.github.com/michhar/2dfd2de0d4f8727f873422c5d959fff5)
- Kaggle (https://www.kaggle.com/c/titanic/data)

Process

Your implementation must:

- Browse to https://classroom.github.com/a/hNPOH7UW and start a repo from there. (In the future, there may be "starter code" for you.) Choose your name and accept the assignment. If you do not see your name, contact the instructor / TA.
- Use a Jupyter notebook or Python script to complete your implementation.
- Read the data, which you may assume is in the same directory as the notebook / script.
- Show that you have inspected the data to determine fields.
- Show the difference in survival rates based on fare paid using a graph.

Grading

Grades for this assignment will be determined by the grader as follows:

- 100% = Code functions, is well-documented and clearly shows the relationship between fares and survival.
- 75% = Code functions but is not well-documented or does not clearly show the relationship between fares and survival.
- 50% = Code functions but is not well-documented -AND- does not clearly show the relationship between fares and survival.
- 0% = No submission / code does not function.

Submission

Submit the link to your GitHub repository on Canvas.