

Titanic project proposal

Datasets: This datasets contains 1309 passengers and 14 columns.

Link to dataset: <https://www.openml.org/d/40945>

Column Names:

- pclass: Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)
- survival: Survival (0 = No; 1 = Yes)
- 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)
- boat: Lifeboat - body: Body Identification Number - home.dest: Home/Destination

Question:

1. What percentage of the passengers are male and female?
2. What percentage of the passengers are in different age groups and classes?
3. What percentage of passengers was travelling alone?
4. What percentage of passengers survived?
5. Does gender appear to affect a passenger's chance of survival?
6. Does age have an effect?
7. Does class appear to have an effect?
8. What was the average fare?
9. Did the survival rate change based on the access to lifeboat ?

Potential Charts:

1. Stacked Bar Chart of Sex vs. Survival
2. Scatterplot of Age vs. Gender where each point represents an individual who survived
3. Frequency chart of fares to determine average fare
4. Regression Lines

Initial Analysis

