

Titanic DataSet – Determining the Best Odds of Survival and Analyzing the Given Factors

I investigated the titanic data provided by Kraggle to explore which factors affected the survival rate of those on board. The following three questions inspired my data investigation:

- 1. Which socio-economic status (pclass) had the highest survival rate?**
 - a. Is there a correlation between those who survived and their social class?**

As one might would assume, the upper class had the highest survival rate as they generally would be first in queue to board rescue boats. The breakdown for the Pclass survival rate is as follows:

| | Pclass |
|---|----------|
| 1 | 0.629630 |
| 2 | 0.472826 |
| 3 | 0.242363 |

- 2. Which sex had the highest survival rate?**

Females had the highest survival rate to males. The means for their survival rate are below:

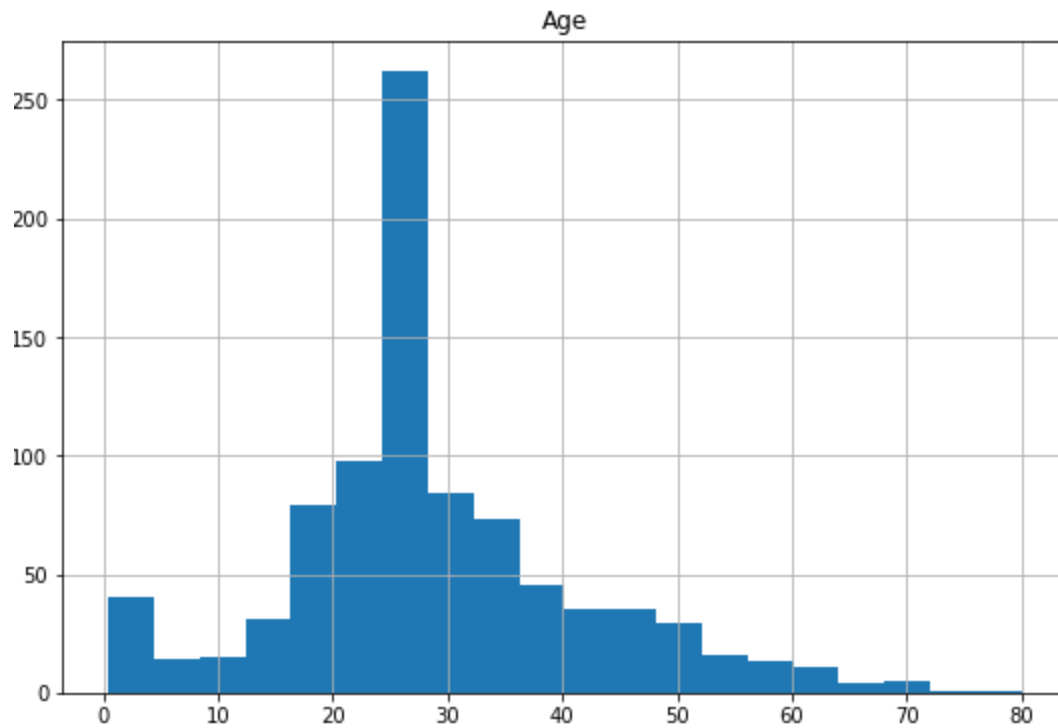
| | Sex |
|--------|----------|
| female | 0.742038 |
| male | 0.188908 |

- 3. Did a person's family size affect survival rate?**

Determined family size by condensing the SibSp and Parch variables into one. The family size of the total population on the boat ranged from 0 to 10 family members. Of those, 7+ persons in a family resulted in a 0% survival rate. It can be argued that the greater the family size, the more chaotic rounding them up would have been when the titanic crashed. Those with family sizes of 3 total had a very high survival rate of +70%.

- 4. How does age affect the likelihood of survival?**
 - a. Are young or old people more likely to survive?**

The age variable through a bit of a curveball in the investigation because the data had a lot of missing ages for the people on board, 177, to be exact. In order to continue with the investigation of this data, a simple replacing of the null values with the median age (28 years old) filled in the gaps.



At a glance, it seems that infants (< 1 years old) had a 100% survival rate. Whereas, seniors (ages 60 and above) did not fare so well.