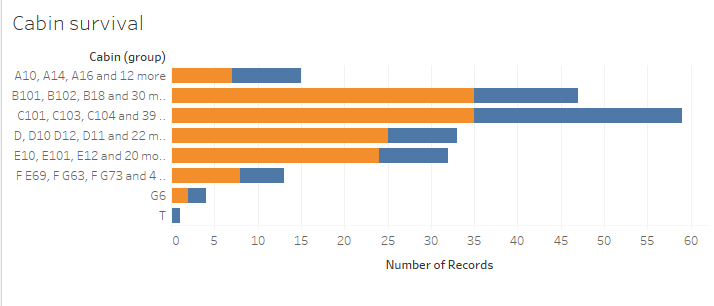
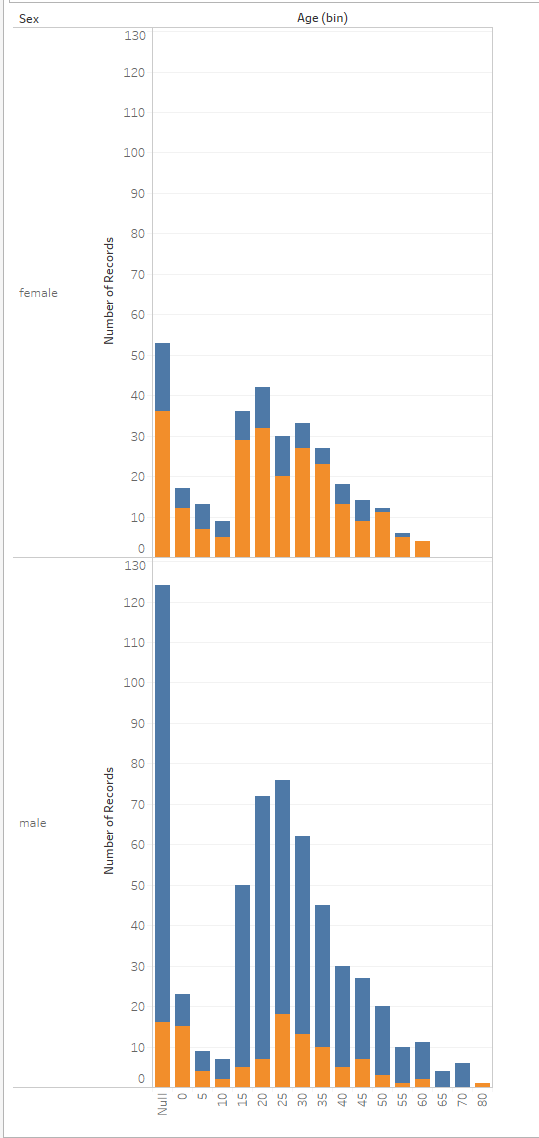
I’m new to this, but I’m excited to give it a go. Looking through the data here are some initial thoughts I have on how feature engineering could be beneficial.

First off looking at the Cabins, it seems that anyone who has a cabin listed has pretty much a better than average chance of surviving, although the cabins show some variance. I feel that grouping them by their letter may be a good way to start:



Age is an issue because there are so many missing values. What to do with them? My initial thought was to turn all the missing values into the average, but that seemed to make early models worse. My new thought is to make the feature into different bins. “Unknown” will be one. Others will be “0-4”, “5-9”, … “75-80, “80+”



Let’s see how it goes!