Titanic project

These are the features that we will be using in the model  
  
pclass

sex

age (demographic age)

sibsp

spouse

parent

child

embarked

moniker

Embarked has two missing values that need to be filled in. I was going to fill it in with S, since this is the highest probability value. In looking at the data, both passengers are first class females and they make up a group. There is no other information. In looking at first class females, taking out the two missing values, 43 embarked from C, 1 embarked from Q and 48 embarked from S. Since the probability that they embarked from S is greater than 50%, I filled in the two missing values with S. It is unlikely that these two values will effect the accuracy of the model.

I will be constructing two csv data files with these nine features that we can import directly into our respective analysis programs. One file will include the training data and one will include the testing data. I will put the files in my folder.