Since there are so many missing values in *deck*, remove the *deck* feature. Define a lambda function to impute age using the median of the passenger class you computed

earlier.

Drop the remaining records containing null values, and show there are no remaining null values.

Several of the remaining features are either duplicates of another feature, or engineered features from other features. For example, *pclass* and *class* are identical, as well as *alive* and *survived*. Therefore, *class, who, adult_male, alive,* and *embarked* will be removed from the dataset. Before removing these features, examine *who* and *adult_male*, and identify the logic/relationship from the other features. The feature names should provide clues as to which features were used. Document which features were used to construct each of these two features and the logic employed. For example, "*who* is engineered from *X* and *Y*, such that if *X* is less than, greater than, or equal to BLAH, and *Y* is less than, greater than, or equal to BLEH, then *who* is *child*."

Convert categorical variables to numeric dummies using pandas' *get_dummies()* method, and add these to your training dataframe. If *get_dummies()* is called correctly, the original categorical features will be replaced with the dummified version.

Create the feature set by dropping s*urvived*.
The resulting feature set should include *pclass*, *age*, *sibsp*, *parch*, *fare*, and the categorical dummy columns you created earlier.

Implement a target dataframe by copying the contents of the *survived* column of your training set to a new dataframe.

Split your clean data into a training and test set.