**Data Work for ML Project**

**Missing Values:**

* Full analysis of all data for missing values

**Imputation:**

* Determine method for imputation of missing values
  + Consider using random forest for imputation
    - Based on lecture, it is OK to use both train set and test set together (removing the target variable of sale price) to impute missing values using random forest.
    - I believe this is the most sophisticated way to impute missing values and I want to gain practice at doing this.
    - Builds on my passion for data quality… everything hinges on the quality of the data you put through the model.

**Feature Engineering (for consideration):**

* Age of house (year sold – year built)
* Years since remodel (year sold – year of remodel)