This is meant to be a comprehensive guide to all the steps to take when attempting to make a predictive model.

1. Clean Data
   1. Missing Data or Values that represent missing data
      1. Drop
      2. Replace
         1. Categorical
            1. Make dummy variables and include missing as an option
         2. Continuous
            1. Aggregate function like mean, median, mode.
   2. Convert Categorical Data
      1. Dummy Variables
      2. Continuous (only if logical)
   3. Cast Object Data
2. Drop Data
   1. Highly correlated (Multicollinearity)
   2. High P value
   3. Low R^2?
3. Initial Modeling - (What set of features work best?)
   1. Run all features in OLS, look at OLS summary.
      1. Per Feature
         1. P value should be less than .05
      2. AdjR^2 - the higher the better
      3. The difference between R^2 and AdjR^2 should be small
         1. If it is large, this indicates that some of your features are not adding to the overall predictability?
      4. P(F) : the likelihood your model is useless (again should be less than .05)
      5. JB
         1. Smaller the better. If >=6 this indicates Skew and Kurtosis
         2. P(JB) : if smaller than .05 the normality assumption should be rejected.
      6. Other Stats (Come back and fill this out)
   2. Histograms
4. Feature Engineering
   1. Logical Conversions
      1. Ex. I have date - but maybe the day of the week is more important
   2. Normalization (normalized features return better R^2)
      1. Log function
   3. Scaling (Makes features easier to compare)
      1. Min-Max scaling
      2. Mean normalization
5. Final Modeling and Presentation
   1. Now that all the data has been cleaned and engineered
      1. Create
         1. Models
         2. Presentation
   2. Model fit
      1. Run RMSE against you model.
      2. Run your model with test-split.
   3. See anything else interesting? Some other lead to chase?
      1. Go back to feature engineering to see if you can model your new target