**Working Notes for ML Project**

General Workflow:

* Data Work
  + EDA
  + Create New Features
  + Impute missing values using random forest
  + Create as automated pipeline to ensure exact same code is run later on test data
* Modeling Work
  + MLR
  + Random Forest
  + Boosting
  + Testing for underfit vs overfit
  + What measure to use to determine best model… make this clear from start of presentation
* ?

Questions:

* Should I use R at all… maybe for some of the EDA?