Project Steps

Downloaded Data From Kaggle

<https://www.kaggle.com/c/house-prices-advanced-regression-techniques>

Hi Ben, sounds good.  I think you will want to do some serious feature selection here.  Here are some ideas:

<https://machinelearningmastery.com/feature-selection-with-the-caret-r-package/>

<http://scikit-learn.org/stable/modules/feature_selection.html>

When you build a model, typically you can use varImp() in R to get feature importances (from almost any model trained with caret I think).  You can also use other metrics to judge feature importance, such as variance, or maybe even chi2: <https://nlp.stanford.edu/IR-book/html/htmledition/feature-selectionchi2-feature-selection-1.html>

I'd say go ahead and get started, and we'll see how far you can get.

6/3/18 -Trying to upload the new graphs to Github (rplot1) is the new one I need to keep updated.

6/7/18 – Still cleaning up some of my data, still cleaning up some of github. I might try to use XGBoost

What do I use for the missing values? Avg of non missing values is normal distribution, normal distribution for a continuous variable. KNN Neighbors to use those.   
  
Categorical fill it with constant value, if it is a pool value use it with a value -1. Fill in all missing values with -1.

Caret -

Linear model, random forest,

What I want to do is run some Variable Importance?

Atom –

What I need to do is keep messing with this, but I can’t figure out why my headers are not there

summary(as.factor(df$PoolQC))

## Ex Fa Gd NA's

## 4 2 4 2909

table(df$PoolArea > 0, df$PoolQC, useNA = "ifany")

train1=Using it RandomForest

What exactly does the Set Seed function do?

**Step 1:** suppose we have N observations, we weight each observation with 1/N weight (Wi). Di=Wi/sum(Wi). In other word, Di is the normalized weight.

Plot each quarters of the plot

Using caret package.

Set.Seed -Helpful for sharing results and making comparisons.

Figure out what it’s telling you.

So when I try to run my random forest I have factors

Doing Webcam:

Between 10-15 mintues, no more than 15-30 dates

-Go to Youtube.com/upload