

# Kaggle Project Update

---

14 April, 2017

# Initial Steps

1. Explored data
2. Fit simple models on unaltered data
3. Removed *NA* in data set where *NA* had meaning as given in data document
4. Fit various models on “clean” data with no new features

## Model Results

Model	CV Error	Test Error
GBM	0.1307	0.13066
Ridge	0.1437	
PCR	0.1412	
PLS	0.1255	
Ensemble	0.116	0.121

# Next Steps

1. Data dimension reduction
  - Tried PCA
  - Clustering for grouping high correlation predictors
  - Possibly remove variables with near zero variance
2. Still unsure what to do with truly missing/unknown values
3. Feature engineering
  - Combine variables with basement and garage features
  - Neighborhoods