

## Capstone Project 1 Proposal

House Prices: Advanced Regression Techniques from Kaggle competitions is chosen to be my first Capstone Project.

Typically, a home buyer would not describe their dream house beginning with the height of the basement ceiling or the proximity to an east-west railroad. But this playground competition's dataset proves that much more influences price negotiations than the number of bedrooms or a white-picket fence. The project is to explore 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa to predict the final price of each home.

Real estate sales agents, sellers and buyers will be interesting in predicting the sales price as it will help them to set their listing price, offering price and final sales price.

The dataset from Kaggle is modified from the Ames Housing dataset which was compiled by Dean De Cock for use in data science education. It's an incredible alternative for data scientists looking for a modernized and expanded version of the often-cited Boston Housing dataset. The data set is provided in csv files.

The work would include data wrangling, exploratory data analysis and machine learning.

Creative feature engineering and advanced regression techniques like random forest and gradient boosting skills are needed to complete this project.

The result will be submitted to Kaggle. It will be evaluated on Root-Mean-Squared-Error (RMSE) between the logarithm of the predicted value and the logarithm of the observed sales price.

The deliverables will include code and slides.