House Prices: Advanced Regression Technique

Problem Identification

- **Problem statement**
  - Understanding what the house prices are
- **Context : Background**
  - When asking a new home owner to describe the house they would want to stay in, the dataset is to prove that there is a lot more to price negotiations.
- **Criteria for success**
  - " Submissions are evaluated on Root-Mean-Squared-Error (RMSE) between the logarithm of the predicted value and the logarithm of the observed sales price. (Taking logs means that errors in predicting expensive houses and cheap houses will affect the result equally.)" - Kaggle
- **Scope of solution space**
  - Advanced regression techniques like random forest and gradient boosting
- **Constraints**
  - 79 explanatory variables at first - would need to narrow it down for trend analysis
- **Stakeholders**
  - Home buyer
  - Landlord
  - Ames Iowa Village officials
  - Realtors
  - Surrounding residents
- **Data sources**
  - Kaggle- Ames Housing Dataset compiled by Dean Se Cock

- **PROJECT GOAL**
  - Predicting the sales price for each house.

- **What is the problem you want to solve?**
  - Would like to solve and predict the prices for each house particularly in Ames, Iowa

- **Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis?**

  Our clients here are the home buyers, who care about the problem to understand the lowest house prices to understand to see if it fits their budget. All of our stakeholders can use it for various use such as the following: