**Final Assignment**

**Problem 1**

Problem Description

Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. But this competition's dataset proves that much more influences price negotiations than the number of bedrooms or a white-picket fence.

With 79 explanatory variables describing (almost) every aspect of residential homes, this project challenges you to predict the final price of each home.

Data

This assignment file has training set as train.csv, testing set as test.csv, description of the variables as data\_desc.

Tasks:

Hypothesis Building

Understand the problem and make hypothesis about what could potentially have good impact on the outcome . Hypothesis, by definition is an unproven truth. It could be a hunch which needs validation through data.

State your hypotheses in separate word file.

Data Exploration

This process is exercised to extract inferences from the dataset, validation of hypothesis on the given dataset.

Data Cleaning

Treating Missing Values and Outliers

Model Building

1. .Make a baseline Model (Model which require no predictive modelling.Just an informal guess )  eg .Just filling the Response Variable for Test dataset with mean or median or mode of Response Variable of Train .

The Response Variable in the given dataset is “Sale Price”.

1. Fit linear Regression Model. Check the assumptions of regression models and rectify it when it gets violated.
2. Plot Residual Vs Fitted plot  and develop inferences from it. Make necessary transformation if required.
3. Bring the concept of polynomial regression if you find non linear relationship between predictor and response variable.
4. Cross Validate the dataset to reduce overfitting.

Final Prediction on test dataset

Predict the values of Response Variable for test dataset with the model having lowest RMSE.

Submission – Follow the same procedure as before.

\*\*NOTE: Make sure you add comments to explain your thought process and other necessary things

Problem 2 will be updated soon…….