House Price Prediction

1. What is the problem you want to solve?

- Real estate business involves committing a large sum of money, for both, the Developers/ Builders and the Buyers.
- The Builders would generally have been in the business for a while and know their way around building/ selling a house and often take the customers (inexperienced) for a ride by selling them substandard houses with premium prices.
- As a client, one would want to know, the actual worth of a house i.e. for a house with the specifics known (e.g. Area, number of rooms, etc.), the approx selling price of the house could be predicted.

2. 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 that they wouldn't have done otherwise?

- Our clients are the house buyers, who wouldn't want to shell out money for the unfair pricing set by the builders.

- After this analysis, the clients can reason if the price quoted to them by the Builders is justifiable or not and then negotiate with a proper understanding of why the house price should be brought down.

3. What data are you using? How will you acquire the data?

- We are using the Ames Housing dataset, a dataset of 79 explanatory variables which describes in detail about the various aspects of the residential homes in Ames, Iowa.

- The data is made available as part of on-going Kaggle competition.

4. Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution. This might include:

a. Is this a supervised or unsupervised problem?

- This would be a supervised problem. Here, we have structured and labeled data to predict made available with us.

b. If supervised is it a classification or regression problem?

- This is a regression problem. Here, the given data set would be used to predict a numerical value (Selling Price)

c. What variable is it you are trying to predict?

- We are trying to predict the Selling price(expected) of the house.

d. What variables will you use as predictors?

- We have a set of 80 independent variables that would be used as a predictor here. These range from the basement condition, its quality to that of fireplace availability.

e. What will be your training data?

- The training data is a set of 1461 rows containing the 81 columns i.e 80 predictors and the resulting Selling price of the house.

5. What are your deliverables? Typically, this includes code, a paper, or a slide deck.

- The project deliverable is a google sheet, containing the consolidated report of the project. PPTs for Story Telling, Python Notebook Files(.ipynb)