House Price Predictor

UC Berkeley bootcamp

Project Proposal

*By: Amitava Samadaar, Prem Ganesh Elango & Suvrangshu Ghosh*

Using :

* Random Forest Regressor
* Gradient Boosting Regressor (if time permits)
* Python
* Pandas
* Matplotlib
* Numpy
* Ski-Kit
* Seaborn

Predicting the real estate home price using Kaggle dataset:

<https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data>

Finding a dream home, selling the house with best price!!!

Here we will try to solve the problem of predicting house prices for house buyers and house sellers.

There are many factors which plays a major role We are going to take advantage of all of the feature variables available to use in the dataset and use it to analyze and predict house prices.

We are going to break everything into logical steps that allow us to ensure the cleanest, most realistic data for our model to make accurate predictions from.

1. Package and Data load
2. Analyzing the Test Variable
3. Multivariable Analysis
4. Impute Missing Data and Clean Data
5. Modeling and Predictions

The data in Kaggle is from Ames, Iowa .

Here we will be using regression techniques with multi variable analysis using multiple models to see which gives the best results.

Understanding the data:

* Understand the missing values:

Many columns has 'NaN' value.

Closer look tells that they are not missing, instead it means it doesn’t exist in that property.

like for PoolQC - Nan means no swimming pool

Find total missing

Take decision depending on the importance of the features.

Understand the Skewness and Kurtosis