**Introduction:**

This Project predicts the Price of a house based on various internal and external characteristics. One of the big dreams of each and every individual is having their own house. Now-a-days the business of selling and buying houses has been rapidly increasing. So, a good housing price prediction would better prepare the customers for what to expect before they make one of the most important financial investments. Also, a recent report from popular housing database website indicates that house sellers and buyers are increasingly turning to online research in order to estimate house price before contacting real estate agents. Researching how much the house which the Customer is interested in is very difficult because the potential Price of a house is affected by various factors. This makes it more complicated for an individual to decide on whether they buy or sell a house for a reasonable price. So, they need an external help to fulfill their dream with more accurate information possible. Therefore, we are here to design a house prediction model that best guides a customer to reach their destination without any problem.

**Business Problem:**

*The problem is to build a model that will predict house prices with a high degree of predictive accuracy given the available data*

**About the data:**

The idea behind selecting this dataset is to display the skills that we had learned in the classes because this dataset includes reasonably large number of variables so that we would go beyond a simple algorithm, such as forward or stepwise selection, to construct a final model. This also provides us an opportunity to explore more on feature engineering aspect which is an important feature a data scientist should possess to construct a best Model.

**Data source:**

To build a Prediction model, the dataset was sourced from Kaggle, “With 80 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, is a challenge to predict the final price of each home.”

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

**Our Goal:**

1) Create an effective price prediction model

2) Validate the model’s prediction accuracy

3) Identify the important home price attributes which feed the model’s predictive power.

**Methodology:**

Exploratory Data Analysis and Feature Engineering

Collect the Raw data

Clean the data

Preprocess the data

Visualizations and Report Findings

Machine Learning Algorithms and Statistical Models

Model evaluation and deployment

**Strategy:**

1. **Understand the problem**: We'll look at each variable and do a philosophical analysis about their meaning and importance for this problem.
2. **Univariate study**: We'll just focus on the dependent variable ('Sale Price') and try to know a little bit more about it.
3. **Multivariate study**: We'll try to understand how the dependent variable and independent variables relate.
4. **Dimensionality reduction:** Using PCA algorithm.
5. **Basic cleaning**: We'll clean the dataset and handle the missing data, outliers and categorical variables.
6. **Test assumptions**: We'll check if our data meets the assumptions required by most multivariate techniques such as regression analysis.
7. **Fit a Model**: Planning to start with regression technique and try to extend our analysis using advanced regression techniques such as Random Forest and Gradient Boosting.
8. **Evaluate the Model**
9. **Predictions**: Use the best fit model to predict the prices of the houses.