

## **Problem Identification**

- **Problem Statement Formation:**

Real Estate companies need help in accurately predicting the price of a home based upon the appropriate variables such as square feet. This can be useful to their clients when listing or buying a home. They need to know if a house is over or undervalued. It can help a company assist their clients better and give them an advantage over other real estate brokers by providing accurate price predictions.

- **Context:**

Potential clients of real estate agents need to know if a home is worth the price based upon different traits of the house. Real estate companies need to help clients understand which variables have the biggest impact on the price of a property. This can be accomplished by assessing historical data of homes using a regression model.

- **Criteria for Success:**

Finding which elements have the highest correlation with the price of a home. Also to accurately predict the price of a property. This will help a real estate company to have an advantage over other firms and provide value to their clients.

- **Scope of Solution Space:**

We will use the 18 columns from our dataset to predict the price of a home. The information we have will paint a picture of what particular elements play a major role in the price of a house.

- **Constraints:**

There might be other elements that are not in our data that have an effect on the evaluation of a house. If the house is next to a busy intersection or if there was a serious crime committed at the property are two variables that we cannot account for.

- **Stakeholders:**

Real Estate Companies, Realtors, Buyers, Sellers, Potential Clients of Real Estate Companies, NAR, Zillow, Redfin, Investors, Housing Market Companies, HOAs, Mortgage Companies, Banks and Appraisers.

- **Data Sources:**

The columns for the data are: Date, Price, Bedrooms, Bathrooms, sqft living, sqft lot, floors, waterfront, view, condition, sqft above, sqft basement, year built, year renovated, street address, city, state zipcode, country. I will add a few other columns as well.

<https://www.kaggle.com/datasets/shree1992/housedata?select=data.csv>

I will create this analysis by wrangling and cleaning the data first. Then I will use graphs, scatter plots, heat maps, etc. to find the correlation between the independent variables and the price of the house. I will formulate a predictive model in python. Here I will use a regression analysis such as a random forest model or lasso regression to predict the numeric home value. The deliverables will be python code to clean, analyze and to create a model on the data. I will deliver the predictive model via a slidedeck, project report in github.