Problem Identification Capstone_Two

Problem Statement Worksheet (Hypothesis Formation)

Which features can significantly affect housing prices?

Build a model in one month to reflect on the relationships between price of houses and the features to improve the accuracy of house price predictions to 60%.

1. Context

House price predictions are very important for investors to make decisions of buying houses. There are many features which play roles in house price decisions. Based on historical data of house sales in different regions in the U.S., we need to predict house price based on features.

2. Criteria for success

Build a house price prediction model to improve the accuracy of price prediction to 60%.

3. Scope of solution space

It is a regression problem. Based on features of the houses including location, we can build a prediction model of house price with machine learning prices model will indicate how housing price and location attributes are correlated, and whether house attributes (bedroom, bathroom count) strongly correlate with the price.

We are going to use linear regression and random forest algorithm, if necessary, try Neural Networks Model. We will choose the best model by comparing their prediction accuracy.

4. Constraints within solution space

The features about the houses are not enough, because we don’t have time when the houses were built. It is an important feature for the valuation of houses.

We have the data of house locations, but we don’t have economic and demographic data about the city the house was located at.

5. Stakeholders to provide key insight

- Manager of Real estate sales department
- Real estate investor
- Real estate valuation information provider

6. Key data sources

- Kaggle.com/USA Real Estate Dataset

  This dataset contains Real Estate listings in the US broken by State and zip code.