Capstone project 1
House price prediction

Problem description

Buying a house is one of the biggest investment in one's life. In U.S., for most families, house down payment and mortgage contribute to the biggest proportion of a family's expenses. Buying a house at the right moment and at the right price is therefore crucial to the financial wellbeing for lots of families.

Since property investment is often considered to be relatively safe and stable, lots of private and organization investors use property as an investment option. Purchasing property(s) at the right time can generate moderate return over a period of decades from two sources: 1) property rental income and 2) property value increase. Maintaining an up-to-date house price prediction is therefore crucial for their investment returns.

Potential clients

Both property sellers and buyers are potential clients of this projects. For sellers, selling their properties at the highest price within a given range of time is ideal to maximize their profits. For buyers, buying properties at the lowest price within a given range of time could save their tens of, even hundreds of, thousands of dollars.

Buyers	Newly-wed, Parents looking for better school district for kids, people upgrading/downgrading their house, property investors, etc.
Sellers	People upgrading/downgrading their house, parents looking for better school district, property investors, banks with foreclosure properties, property investors, etc.

Data source

For this project, the following dataset will be utilized:
https://www.kaggle.com/c/house-prices-advanced-regression-techniques
This data set is supposed to be clean (haven't explored the data yet), and designed for students with knowledge of data science who would like to practise their skillsets.

Alternatively, a backup dataset can be found here:
https://www.kaggle.com/zillow/zecon#DataDictionary.csv

This is zillow dataset with much more data entries.
The first dataset will be explored first. The backup dataset will only be used when the first dataset does not satisfy the project needs.

Methodology

The following procedure is proposed to predict the house price.
Step 1. Data loading
Step 2. Data wrangling
Step 3. Exploratory data manipulation and trend identification
Step 4. Identify key parameter affecting house price
Step 5. Divide data into training set and test set, and train model(s)