

Springboard Capstone Project 2: Project Proposal

Predict House Price

Lakpa Sherpa

The price of a house is dependent on various factors like size or area, how many bedrooms, location, the price of other houses, and many other factors. Real estate investors would like to find out the actual cost of the house in order to buy and sell real estate properties. They will lose money when they pay more than the current market cost of the house and when they sell for less than current market cost. The banks also want to find the current market price for a house, when they use someone's house as collateral for loans. Sometimes loan applicant overvalues their house to borrow the maximum loan from the bank. Banks and financial institutions also provide mortgage loan to home buyers. Local home buyers can also predict the price of the house to find out if a seller is asking for too much. A local seller can also predict their house price and find out how much is a fair market price.

The dataset for this project is downloaded as a csv file from the Kaggle website. This dataset contains house sale prices for King County, which includes Seattle. It includes homes sold between May 2014 and May 2015. The data set consists of 21613 observations and 19 features plus the house price and the id columns.

<https://www.kaggle.com/harlfoxem/housesalesprediction>

The columns are as follows:-

id: a notation for a house

date: Date house was sold

price: Price is prediction target

bedrooms: Number of Bedrooms

bathrooms: Number of Bathrooms

sqft_living: square footage of the home

sqft_lot: square footage of the lot

floors: total floors (levels) in house

waterfront: House which has a view of a waterfront

view: Has been viewed

condition: How good the condition is (Overall)

grade: overall grade given to the housing unit, based on King County grading system

sqft_above: square footage of house apart from basement

sqft_basement: square footage of the basement

yr_built: Built Year

yr_renovated: Year when the house was renovated

zipcode: zip

lat: Latitude coordinate

long: Longitude coordinate

sqft_living15: Living room area in 2015(implies -- some renovations) This might or might not have affected the lot size area

sqft_lot15: lot size area in 2015(implies--some renovations)

In this project, I will perform data wrangling to find out any missing data, bad or mistake data, and outliers. Then, I will perform exploratory data analysis and visualization using the dataset to get some insight. Then, I will perform inferential statistics to do hypothesis testing to find out more about the dataset. I would build different machine learning regression models to predict the price of a house, and I will compare them with each other to find out the best model to accurately predict house price.

The deliverables for this project would be a complete presentation of the project with all the documentation, visualizations and python codes on a notebook. All the model created to predict house price. The final submission will include the introduction, the steps, and methods involved in this project from the beginning to the end of the project. Finally, present how we evaluated the accuracy of different models created, and recommend a model.