

Predicting House Prices in CA Bay Area

Target: Predict house prices using multiple Machine Learning Algorithms considering different sets of house and zip code features:

(1) Main Dataset: Source www.redfin.com:

- Obtained sold houses from Dec 2019-Dec 2020.
- The sold houses located in North California distributed between 49 cities and 94 Zip Codes.
- Data has 8,790 Observations and 27 Variables.
- Data Wrangling is needed: Dataset has NaNs, Wrong values, insignificant columns and Duplicates.

***** Adding more Datasets to enhance our prediction *****

(2) Adding Median Income per Zip code from:

Source: <http://www.usa.com/rank/california-state--median-household-income--zip-code-rank.htm> and since the data from 2010-2014, and according to <https://www.statista.com/statistics> --> the median Income in California has grown by 6.36% we'll use this % to adjust this Dataset YoY.

(3) Adding Public School per zip code:

Source: https://hifld-geoplatform.opendata.arcgis.com/datasets/87376bdb0cb3490cbda39935626f6604_0

(4) Adding Hotness score (0-100) to reflect demand and supply per zip code:

Source: <https://www.realtor.com/research/data/>