

Capstone 3 project proposal
Real estate housing price prediction

Context

Real estate has been known to be one of the few investments people consider safe and low risk given that the housing market typically trends upward. This makes it seem as though investing in real estate will always be a safe place to put your money; however, it may not always be the case. There are areas that are experiencing more growth than others which would be the most ideal location to purchase property. I would like to build a model that predicts the future value of a given area based on the area's historical data. This should yield valuable information for the typical real estate investor or an individual who plans to purchase a new residence and is unsure where to buy.

Criteria for success

In order to have this project be successful, the categorization of areas needs to be well defined, in this case, I will be focusing on area codes and their projected value in the next 5 AND 10 years. Secondly, obtaining enough historical data from each postal code is also key. There may be more data from booming cities such as Austin Texas in comparison to a smaller town such as Eugene Oregon. I will most likely need to filter out the zip codes with less sales and focus on the top 10 to 15 states with the most houses sold.

Scope of solution space

I plan to use a regression model in order to predict the price in the near future such as linear regression, logistic regression or multiple regression, depending on the model performance. If the model outcomes are not as strong as i intend them to be then i will apply the ensemble method gradient boosting regression

Constraints

The main constraints of this project is the data. Obtaining the correct and sufficient amount of data per given area is key in this case to make valid and accurate predictions. A second constraint is also the manner in which the areas of interest are classified.i initially plan to group the areas by zip code but may need to change pending better grouping methods.

Stakeholders

The individuals who may be interested in the outcome of this project range from seasoned real estate investors to first time home buyers. If anyone is in the market to purchase or sell a home or property then the predictions may be valuable.

Data sources

There is a data set that was posted on kaggle which was obtained from realtor.com

https://www.kaggle.com/maedemaftouni/real-estate-market-trends?select=RD_C_Inventory_Core_Metrics_Zip_History.csv

This data contains 3 separate files each of which contain information from states, zip codes and counties in the United states. I ultimately would like to use an API to obtain data straight from the source but have yet to find a viable API to use. At this point I will be moving forward with this data set unless an API is found to take its place.