Project Summary

Problem Statement

California real estate is once again at an all-time high. Houses can range from $500,000 to $50,000,000, what exactly contributes to this cost? Is there a quantifiable approach to accurately approximate this cost? If there is, what are the most and least important factors that affect the price? By how much and how many?

Solution

The problem focuses mainly on interpretation, which means a simple understanding model shall be used. In this project, the multi-linear regression will be a good choice, since it is easy to interpret, everything can be broken down into coefficients. This can be done using Sklearn on Jupyter Lab to give steps by steps approaches, both good for analysis and documentation for the readers.

Data Source

The ideal data source is from Zillow, Zillow has many housing data, working with multi-sources brokers listing. Another good source will be Redfin, Redfin has a good amount of data as well, but Redfin does not have a public API. However, web scraping like BeautifulSoup or Selenium is always possible. A non-business owns data actually found from Santa Barbara County, it provides all the recent 6 months housing sold reports, which is perfect for this project. The data will be download .xls and imported to Pandas. There is a total of 2350 rows in this data set.

Programing Environment

Window 10/11, Python 3.9, Jupyter Lab, Google Chrome

Libraries Used

Pandas, Sklearn, Matplotlib, Seaborn, Xlrd

Result

The most important finding in this project, can be summarized in this chart.

Ranks of Importance	Factors	Absolute Coefficient Percentage
#1	Locations	47.6%
#2	Bedrooms	31.9%

Project Summary

#3	Bathrooms	11.8%
#4	Property Types(House or Condo)	8.1%
#5	All Others	0.8%

This is the result produce by multilinear regression model training by 1886 data after clean (before cleaning: 2350). The R-Squared Error of this model is 0.64.

Project Link and Source Code

Main Finding and Command Line Interface Application in IPYNB:

https://github.com/GiggleSamurai/Multi-linear-Regression-Approximation-on-Santa-Barbra-s-Real-Estate/blob/0e1a837c020621e28e9e4a3438782b15e4838a49/Project_Summary.ipynb

Walkthrough Process in IPYNB:

https://github.com/GiggleSamurai/Multi-linear-Regression-Approximation-on-Santa-Barbra-s-Real-Estate/blob/0e1a837c020621e28e9e4a3438782b15e4838a49/Project_Summary.ipynb

Detail Documentation

APA Style PDF Document:

https://github.com/GiggleSamurai/Multi-linear-Regression-Approximation-on-Santa-Barbra-s-Real-Estate/blob/0e1a837c020621e28e9e4a3438782b15e4838a49/Real%20Estate%20ML.pdf