

David Bender - Springboard Capstone #1 Proposal

The Problem:

Real Estate investment and development has to optimise the value of the homes that they build, but estimating what types of homes sell for the highest price based on geography and property type is challenging. Predicting where to focus development geographically and what type of homes to build would improve the process and improve the investment decision process.

The Client:

Real estate developers who plan and construct the properties and institutional real estate investors (e.g., investment funds, private equity) who invest in the developments. The value for these two groups is a better initial planning process and better estimates for their financial models.

Data:

The UK Housing Prices Paid Database for England and Wales from 1995-today

- Via HM Land Registry Prices Paid data
- (<https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads>)

This dataset has over 22MM lines that includes the price paid, date of transfer, property type (Detached, Semi-detached, Apartment, etc.), new/old development, duration of leasehold, property postcode, and property address.

Initial Solution Proposed:

1. Exploratory analysis of data, cleaning any missing values and adding labels where needed
 - a. Cleaning any missing values
 - b. Adding labels where needed
 - c. Visualizing key summary statistics, charting total distributions, plotting geographically, etc.
2. Scikit-learn for series of supervised learning regression analysis to predict future home value
 - a. I.e., predicting sale price based on location and property type
3. Test the scikit-learn packages to determine best model for prediction
4. Visualize the distribution of home values

Deliverables:

Codebook that can do the following;

1. Estimate home value (based on price to be paid) of a property given geography (Post Code, City, County) and property information (Property Type, Old/New, Date Sold) inputs.
2. Auto-generate a report that charts and presents the key stats, factors, or distributions given inputs (i.e., price distribution of Detached homes in Devon over past 5 years)