

Task #6.1 Sourcing Open Data

Q1. If you haven't done so already, download your Achievement 6 project brief (.pdf).

A1: Completed, downloaded. Created files with 01, 02, 03, 04, 05 Sent to Client – same format as Python project.

Q2. The data you use for your project will need to meet certain criteria as defined in the brief. Read through the data requirements now to be sure the data you choose is appropriate.

A2: Criteria are met with this data set. Geographic component is the various property addresses all within New York State , 3+ Continuous variables are close price, property square feet, bedrooms, bathroom counts. 3 categorical variables are brokerage, address, sub localities. Data has 17 columns & 4802 rows. Titles do currently have acronyms but will be formatted to be as clear as possible. Data is recent – uploaded on January 2024, after doing further research these are all active listings – not yet sold homes.

Q3. Source your data. Use the requirements (and your own interests) to source an open data set from the web. We introduced you to several sources in this Exercise but feel free to look elsewhere, as well.

A3: <https://www.kaggle.com/datasets/nelgiriyeewithana/new-york-housing-market/data>

NYS active listings from Kaggle. I currently work in Real Estate so Real Estate data is a bit more familiar to me. Also this has extreme relevancy in terms of potential investment opportunities & the data is “raw” enough to where I can see potential changes to improve this dataset.

Q4. Create a new document to detail your project information.

A4. Completed, created a report document to track changes.

Q5. Create a “Data Source” section in your project document and provide the following information: A summary of your data source. We recommend you revisit [Exercise 1.4: Sourcing the Right Data](#) for a recap on what to include in your summary. An explanation for why you’ve chosen this data set.

A5. Created a data source section. Data was selected as this is relevant housing data – something that everyone wants to know about especially if they are an aspiring home owner, a current home owner, or investor. It’s relevant to the state I reside in. Also looking at housing data is already something I do in my day job.

Q6. Clean your data. Conduct some basic data cleaning and consistency checks in Jupyter to ensure your data is ready for further analysis.

A6. Data fully cleaned on Jupyter, with Final touches on Excel. (Attached). All processes added to Housing data report file.

Q7. Understand your data. Develop a basic understanding of your data set by reviewing the variables and performing basic descriptive statistical analysis. You might want to make a data profile similar to what you did in Achievement 1.

A7. Variables and Data Types:

- Address (A): qualitative, time-invariant, nominal
- Price (B): quantitative, time-variant, continuous
- Bedrooms (C): quantitative, time-invariant, nominal
- Bathrooms (D): quantitative, time-invariant, nominal
- Square Feet (E): quantitative, time-invariant, nominal
- Borough (F): qualitative, time-invariant,, nominal
- Neighborhood (G): qualitative, time-invariant, nominal
- County (H): qualitative, time-invariant, nominal
- Cleaned Address (I): qualitative, time-invariant, nominal
- Zip code (J): qualitative, time-invariant, nominal
- Property Type (K): qualitative, time-invariant, nominal
- Brokerage (L) qualitative, time-variant, nominal

Q8. Consider limitations and ethics. Outline any limitations and ethical considerations presented by the content of your data, its source, and/or how it was collected. Include the results of steps 6 to 8 in a second section of your project document. This second section can be titled something like “Data Profile.”

Q8. Put everything under “Housing data report” for ease of reviewing.

Q9. Define questions to explore. In a third section of your project document, define a list of questions to explore with your analysis. As mentioned in the Exercise, you may want to revisit Exercise 1.2: Starting with Requirements for a recap on writing good questions.

Q9. Questions under housing data report. Example questions buyers would ask are....

- A. What is the average property price within each borough?
- B. Average Property Price of a certain type within each borough? "How much do co-ops in Brooklyn cost? How about condos in Manhattan?" etc.
- C. What are the different price ranges? Can you categorize them into sections? How would you filter them?
- D. Categorize each neighborhood based on safety / crime rate.
- E. Relationship between square footage and size? How much price per square footage to look for?
- F. Square footage and size? How much price per square footage to look for?
- G. Is this property appropriately priced relative to similar properties?
- H. Can you give me an estimate on monthly mortgage?
- I. What is an appropriate bathroom to bedrooms ratio? Is it every 2 rooms one bathroom? Every 3?
- J. How much are closing costs for different brokerage? Approximately.