Ramsey Basma Progress Report:

Project scope update

As of 11/13/25, I have created a GitHub repository for my project and connected it to a local directory via Visual Studio Code. I have formatted the directory as required and kept out unnecessary files via .gitignore. Within the project, I have completed the data collection portion successfully pulling data from all three sources via API. I've structured the code to allow testing either directly through the data_collection.py file or via the tests.py file. Additionally, I've structured the tests.py file to be run through the command terminal as explained in the readme file.

Data sources

The code pulls from three data sources – Kaggle housing data sets, Federal Reserve Economic Data, and Google Trends.

Kaggle housing data is provided specifically through "usa real estate dataset" (a CSV file extracted from a zip folder) and provides several columns of data with the most important being date, region, and home price. This data is that of actual individual homes sold and, in its raw form, is too much to be displayed on the Excel grid. This data is pulled via Kaggle's API with a personal account.

Federal Reserve Economic data (FRED) is provided specifically through the Federal Reserve Bank of St. Louis and provides 30 year fixed mortgage rates each month dating back to 1971 in CSV file format. This data is pulled via FRED's API with a personal account.

Google Trends uses Google Search Interest to provide data for keywords. This project specifically uses search interest data of "homes for sale" for the past 10 years. Date and Google Search interest index are provided. This index is a relative measure of search popularity at that given time. This data is pulled via Pytrends API with no need for an account but a limit of pull requests hence the limit to one phrase search of "homes for sale."

Issues / difficulties

Throughout this project, I've faced the most difficulty in finding, developing, and using new code for the first time. This project has incorporated language not yet covered in the course or homework which causes more time to be spent learning the code as well as various programs including GitHub and VS Code rather than providing meaningful results in regards to the actual data analysis, but I am aware this may be the intent. Additionally, I had to watch a few online tutorial videos regarding using VS code with GitHub for reasons such as difficulty with commits and specific tasks such as deleting previous commits.