Milestone 1

June 18, 2023

Project Description

For this project, I will be exploring GDP Data in the United States, House Price Data, and City Population data. Using this data, an analysis can be conducted on the relationship between the three and can be looked at by city and state. This data can give further undestanding into how GDP, home prices, and population interact and if there is any correlation between the three.

3 Data Sources along with Description and Links

1. List of U.S. states and territories by GDP (Website). The link to this data can be found here: https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_GDP. This data will include GDP information for the years 2022 and 2021.
2. Zillow House Price Data (CSV/Flat File). This data contains average prices of home sold by city and by month dating back to 2007. The link for this data can be found here. https://www.kaggle.com/datasets/paultimothymooney/zillow-house-price-data/discussion
3. GEO API (API). GEO API contains City Details such as Latitude, Longtitude, and Population. The link for this api can be found here https://rapidapi.com/bixby-bixby-default/api/geoapi13

Relationship between Data

These sources are all related by State/City and the data will be merged by those fields. Using the information from the sources, I will be able to view GDP by state, as well as average price of homes sold in a particular city, and GEO API will provide population data for each city. These three sources will be merged into one to analyze the data found within the three.

Project Plan

The plan for this study is analyze the trends between GDP, Population, and Home prices. This can mean seeing if higher populated states have a higher GDP, or if there are other relationships between home prices and population. The project plan will involve a series of steps. This The first step will be to import all the data appropriately. There will be a challenge when calling the API due to the limits set by the api owner. After import, the data will need to be cleaned and prepped for merging. This can involve renaming columns or selecting relevant columns that will be used for this study. Once the data sets have been cleaned, they can be joined to the most granular dataset to have all the data condensed. Lastly, visualizations can be made to understand any trends and get a better idea of the data. A few examples of visualizations that can be made are states with the highest GDP and their respective population, and the opposite as well. Another visualization can be average home price by city. These are a few potential possibilities of visualizations that can be made.