

Capstone Project

By Hannah Donnelly

Table of Contents:

1. [Executive Summary](#)
2. [Motivation](#)
3. [Data Question](#)
4. [MVP](#)
5. [Schedule](#)
6. [Data Documentation & Sources](#)
7. [Known Issues & Challenges](#)

Executive Summary:

The goal of this capstone project is to identify the most suitable city in the United States to relocate to, based on an analysis of infrastructure, diversity, and affordability. This exploration is motivated by a desire to find a place that supports both personal well-being and professional flexibility, while offering a vibrant, connected community.

To determine the five cities for comparison in my capstone project, I initially used my astrological chart, which suggested the following cities: Asheville, NC, Santa Fe, NM, Seattle, WA, Portland, OR, and Boston, MA. However, after further consideration, I made a few adjustments. I decided against Portland, OR, as I felt I wasn't weird enough to live there. Additionally, I ruled out Asheville, NC, due to concerns about the city's recovery following a recent disaster. While I love Asheville and have enjoyed visiting, I was unsure if the data would accurately reflect the current state of the city. I also wanted to explore cities I had not yet visited, so I replaced Boston, MA. Ended up excluding Santa Fe, NM due to the lack of raw data on their city data site. The final four cities selected for this capstone project are Raleigh, NC, Seattle, WA, Philadelphia, PA, and Austin, TX. These cities were chosen based on a combination of personal interest and alignment with my astrological birth chart.

Key assumptions include the availability of reliable internet, given that the four selected cities are all major metropolitan areas. It is also assumed that [greater community diversity correlates with a more diverse food scene and broader access to cultural events](#). Additionally, this project assumes relative stability in infrastructure and cost-of-living metrics over the next few years, along with access to accurate and up-to-date data sources for each criterion.

Challenges may include inconsistencies in data availability or definitions across cities, particularly in metrics related to public transportation and diversity, as well as potential gaps in affordability data at the city level. Additionally, assigning appropriate weights to each factor while balancing objective analysis with personal values may present difficulties.

Motivation:

As a remote worker with a dog, a fish, and a collection of (occasionally) thriving houseplants, my lifestyle prioritizes access to outdoor spaces, walkable neighborhoods, and a strong sense of community. While not all of these elements are directly measurable in this project, they inform my focus on public transportation, diversity, and affordability. I'm seeking a city where car-free living is an option, green space is accessible, and ['third spaces'](#) like parks, cafes, libraries, and local events support meaningful social interaction.

To help narrow down potential cities, I began with a set of recommendations based on my astrological birth chart. While unconventional, this method provided a personally resonant starting point. The final city selections are a combination of that initial guidance, my lifestyle values, and practical considerations such as recent events or familiarity with the area. These lifestyle-driven priorities directly influenced the cities chosen for comparison and the criteria used in this analysis.

Data Question:

Where should I live?:

Which U.S. city among my selected options - Raleigh, NC; Seattle, WA; Philadelphia, PA; and Austin, TX - offers the best balance of public transportation, diversity, and affordability for a remote worker seeking a vibrant, walkable, and connected lifestyle?

- Livability, Top 100 Best Places to Live in the US:
<https://livability.com/best-places/top-100-best-places-to-live-in-the-us/>
- Niche, Places to Live Rankings: <https://www.niche.com/places-to-live/rankings/>
- U.S. News, Best Places to Live in the U.S. in 2024-2025:
<https://realestate.usnews.com/places/rankings/best-places-to-live>

Minimum Viable Product (MVP):

The final capstone will consist of a Google Slides presentation and an interactive Power BI dashboard analyzing four U.S. cities — Raleigh, NC; Seattle, WA; Philadelphia, PA; and Austin, TX, across three key quality-of-life metrics:

- **Urban Mobility & Infrastructure**
 - Public transportation budget
 - Public transportation budget per capita
 - Transit service area population per sq mile
 - Average commute time
- **Diversity**
 - Racial diversity per city
 - Foreign-born population %
 - Number of languages spoken
 - Non-European Ancestry
- **Affordability**
 - Median monthly rent or avg cost of a home
 - Annual per capita personal consumption expenditures
 - Annual healthcare, housing, and dining costs
 - Average monthly utility bills

The intended audience is myself, a remote worker evaluating relocation options. This project is designed to support a personal decision based on lifestyle priorities: accessible public transportation, cultural diversity, and long-term affordability.

To reach a recommendation, I will evaluate and compare each city's performance across the selected factors, identifying which location best aligns with my personal and professional growth goals.

Schedule (through 7/10/2025):

1. Get the Data (5/17/2025)
2. Clean & Explore the Data (6/10/2025)
3. Create Presentation of your Analysis (6/20/2025) - Should be a presentation, but could include a Jupyter Notebook or dashboard in Excel, Tableau, or PowerBI
4. Internal demos (6/21/2025)
5. Demo Day!! (7/10/2025)

Data Documentation & Sources:

This project will evaluate four cities - Seattle, Philadelphia, Austin, and Raleigh based on three key categories: Urban Mobility & Infrastructure, Diversity, and Affordability.

City Candidates & Sources:

- **Raleigh, NC:**
<https://raleighnc.gov/apps-maps-and-open-data/services/gis-and-open-data>
<https://data-ral.opendata.arcgis.com/>
- **Seattle, WA:** <https://data.seattle.gov/>
- **Philadelphia, PA:** <https://www.phila.gov/programs/open-data-program/> and <https://opendataphilly.org/>
- **Austin, TX:** <https://data.austintexas.gov/>

Below is a breakdown of the data sources and metrics used in each category:

- **Urban Mobility & Infrastructure:** access to public transportation.
 - **Metrics:**
 - Availability of public transit systems
 - Average Commute time
 - Walkability Score
 - **Sources:**
 - **Seattle:** Public transport: King County Metro
(<https://kingcounty.gov/en/dept/metro>)
 - <https://www.soundtransit.org/help-contacts/business-information/open-transit-data-otd/otd-downloads>
 - <https://kingcounty.gov/en/dept/kcit/data-information-services/gis-center>
 - <https://www.seattle.gov/transportation/document-library/reports-and-studies>
 - **Philadelphia:** Public transport: <https://www.septa.org/>
 - **Austin:** Public transport: <https://www.capmetro.org/>
 - https://data.austintexas.gov/Transportation-and-Mobility/TRANSPORTATION_urban_trails_network/jdwm-wfps/about_data
 - <https://www.capmetro.org/facts/#>
 - **Raleigh:** Public transport: <https://goraleigh.org/>
 - Census S0801 Commuting Characteristics by Sex:
<https://data.census.gov/table?q=S0801>
 - Monthly Modal Time Series (Safety and Service):
<https://www.transit.dot.gov/ntd/data-product/monthly-modal-time-series-safety-and-service>

- **Diversity:** Racial, ethnic, and cultural diversity
 - **Metrics:**
 - Racial diversity per city
 - Foreign-born population %
 - Number of languages spoken
 - **Sources:**
 - U.S. Census Bureau – American Community Survey (ACS) (data.census.gov)
 - DP02 Selected Social Characteristics in the United States: <https://data.census.gov/table?q=DP02>
 - The American Community Survey (ACS) is an annual, nationwide survey conducted by the U.S. Census Bureau to provide communities with up-to-date data on demographic, social, economic, and housing characteristics. (<https://www.census.gov/programs-surveys/acs>)
 - Brookings Institution (<https://www.brookings.edu/>)
 - Wallethub (<https://wallethub.com/edu/most-diverse-cities/12690>)
- **Cost of Living & Housing:** affordability based on rent, housing prices, utilities, and other living expenses.
 - **Metrics:**
 - Median monthly rent
 - Median home price
 - Average monthly utility and internet costs
 - Grocery and transportation prices
 - Living wage per city
 - Cost of Living Index
 - **Sources:**
 - Zillow Research (rent and housing cost): <https://www.zillow.com/research/data/>
 - BLS (Bureau of Labor Statistics): Regional averages for housing, utilities, transportation, etc. (<https://www.bls.gov/>)
 - MIT Living Wage Calculator: <https://livingwage.mit.edu/>
 - Cost of Living Index by C2ER (Council for Community and Economic Research): (<https://www.coli.org/>)
 - Fair Market Rents (FMRs): https://www.huduser.gov/portal/datasets/fmr.html#data_2025
 - U.S. Census Bureau - American Community Survey (ACS): <https://data.census.gov/table?q=Housing+costs>
 - U.S. Bureau of Economic Analysis (BEA): <https://www.bea.gov/news/2024/real-personal-consumption-expenditures-state-and-real-personal-income-state-and>
 - RPPs show how much more or less expensive it is to live in a particular area relative to the national average, which is set to 100.
 - RPP > 100: The area's prices are above the national average.

- RPP < 100: The area's prices are below the national average.
- There are overall RPPs (all goods and services) and specific RPPs, like for housing rents.
- Open Energy Information (OpenEI) – Utility Rate Database:
https://openei.org/wiki/Utility_Rate_Database
- U.S. Energy Information Administration (EIA):
https://www.eia.gov/electricity/sales_revenue_price/
- World Population Review:
<https://worldpopulationreview.com/state-rankings/water-prices-by-state>

Technology Selection:

- ☐ Excel
- ☐ Power BI
- ☐ Python
- ☐ Google Sheets
- ☐ Google Slides

Known Issues and Challenges:

Some anticipated challenges with this project include inconsistencies in data availability and formatting across the four selected cities. Additionally, subjective qualities such as cultural richness or sense of community are difficult to quantify, so I will assume that factors like language diversity and foreign-born population percentage correlate with a more diverse food scene and broader access to cultural events. Another challenge involves ensuring data from different sources aligns in terms of time frame and geographic scope.

Data cleaning steps may include standardizing city names, handling missing values, and normalizing cost-related metrics for easier comparison. I will also ensure that all data is as current as possible and clearly document any assumptions or gaps.