Seth Robertson

Final Project

Identification of your research question

* People often talk about the cost of living versus the average income across the United States. I have heard things about the high cost of living in California and also the high income- so I questioned myself, is it worth the income if the cost of living is high as well? What cities in the United States have the highest income to lowest cost of living ratio?

Description of your data

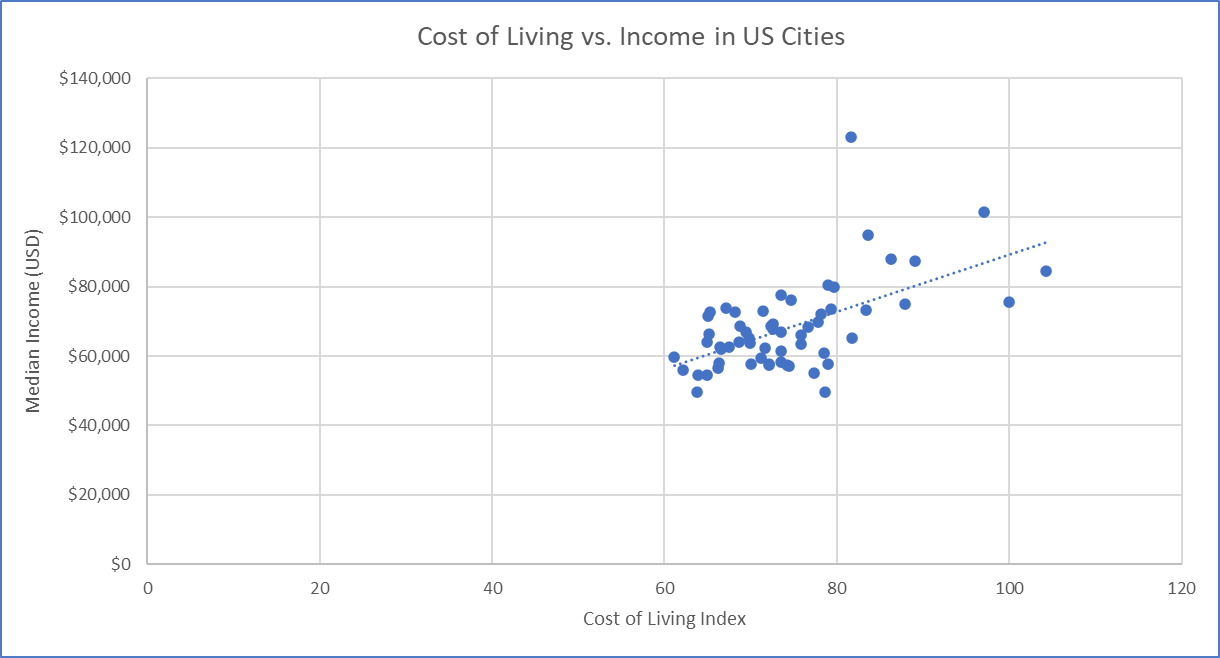
* To find my data for income, I used the [United States Census Bureau](https://data.census.gov/cedsci/table?q=median%20income&g=0100000US%2431000M1). I used median income so my data would not be as affected by outliers. I had to use the geographical filter of the Metropolitan Statistical Areas in order to see cities.To find the cost of living index, I used a website called [Numbeo](https://www.numbeo.com/cost-of-living/region_rankings_current.jsp?region=021). It had the major cities in all of Northern America.
* The census dataset was very large, so I had to target the data that was useful to me. It included information about families, races, marital status, and the margin of error. The only rows that I needed were median income, but they were bracketed into different age groups. I identified the four age groups, averaged them, and then created a new column of my overall average median income matched with the Metropolitan Statistical Area.
* For my cost of living index, I just had to pull the index and city from a smaller dataset that included unnecessary things such as the rent index, purchasing power, and cost of groceries index.
* Once I had my income data and index data in new sheets, I had to match the cost of living with the median incomes using vlookup. Since the census data was giving me the Metropolitan Statistical Area and the cost of living index was just cities, it gave me some issues. For example, the census data had titles such as “Atlanta-Sandy Springs-Alpharetta, GA Metro Area,” while the numbeo data simply had “Atlanta, GA, United States.” I had to find and replace certain pieces of the titles with ctrl+h so I could be positive that vlookup made the exact match.

Initial exploratory data analysis

* My initial exploratory data analysis involved examining the ratios of the median income and cost of living index. In my target data sheet, I created a new row with the ratios and filtered it from largest to smallest. With these ordered ratios, I could get a sense of how spread out or clustered the data was, and allowed me to see the best or worst ratios.

Model of analysis

* I noticed that my data was more clustered than I anticipated. As the cost of living or median income increases, you can see that there seems to be more variation from the trend line. Other than that, towards the beginning of the data, cities have relatively similar ratios- something that I was not expecting to see.

Data visualization

Unpacking your results

* With the work that I did, I will be able to compare cities when it comes to the ratio of income vs. cost of living. If someone is offered the same type of job in different parts of the country, they may want to know how far their income gets them. Also, people may want to look at jobs in cities that have an above average ratio vs cities with a below average ratio. For example, shown below, I highlighted the city of San Francisco, as it is the city with the highest ratio of income compared to its cost of living. On the other hand, the amount of income employees earn in New Orleans is not standard for their cost of living.
* My analysis raises other questions such as what economic factors impact this ratio of income to cost of living? How do the industries of these cities compare, such as San Francisco and New Orleans? Are there trends being shown that give reason to believe the ratios of certain cities are going up or down?
  + To answer these questions, there would have to be in-depth analysis of the economies of these cities, along with consideration of the projections of the industries within them.