

Introduction

In the modern era with a globally connected economy, young professionals are increasingly willing to relocate for the right opportunity. Often these opportunities come from among many active applications for the applicant, with 30 days being a typical timeframe for their relocation. 30 days is a crunch for any individual to assess a living situation and accurately anticipate the actual annualized cost of living that relies on more than simply the monthly rental fee associated with living. How can a young professional quickly and easily determine what areas are within their budget and what their commute will be like? This capstone will aim to give a young professional summary information about areas surrounding a potential job location so that they can make an informed subjective decision with a more complete objective dataset than would otherwise likely be achievable in such a short timeframe for relocation.

Target Audience:

This idea is focused on the state of Massachusetts with a job based in the center of Boston. Selfishly, the idea is derived from my own struggles to find an appropriate living situation when a job opportunity presented itself in a similar way in another city. Knowing the stress of the scramble for living accommodations, this tool could be of great use to anyone seeking deeper knowledge about cost of living on a tight timeline.

Data

For the given problem, the number of cost factors is quite significant and so the scope will be limited to giving a holistic view of recurring monetary costs associated with accommodations and transportation, namely median rent in all surrounding areas and the commuting distance to the job location to be extrapolated to cost. For this the following data is required:

- A list of cities/towns and counties in a state
- Geolocation data for cities/towns and counties for map display
- Home pricing data from the active housing market
- Reported apartment rental prices

For this data to be manageable and appropriately complex for this project, the scope is limited by the following assumptions:

- The target audience is searching for a job based in the heart of Boston, MA
- The target audience has excellent credit
- The target audience is a potential first time home buyer
- The target audience has a compact car for commuting purposes
- Property inflation is not a predictable factor and will not be accounted for
- Any prices for the previous year (2020) are identical in this year so that the dataset will accurately represent a whole year of price fluctuation
- Annual averages will be used as seasonal price fluctuation is highly volatile

To complete the analysis of this data and provide the target audience with a more holistic assessment of living options, several unique formulas will be used to add depth to the gathered data:

- The Haversine formula, used to calculate the great-circle distance between two coordinates, will be used to approximate commuting costs using the job location coordinates and all towns/cities. Once calculated as a distance in miles, an approximate monthly commuting cost can be approximated based on average mpg and an average gas price
- A monthly mortgage payment calculation. Using the principal cost of the home and assuming a 3% down payment (first time home buyer option) and 3% interest rate for a 30 year-fixed mortgage, a monthly cost can be approximated and compared to rental costs in a less intimidating way than the total home price. Buying is often not as expensive in some areas especially when compared to renting within large cities, this contextualized number will allow the target audience to make a more informed decision

Sources of data and methods to extract them:

Mapsofworld.com (<https://www.mapsofworld.com/usa/states/massachusetts/lat-long.html>) provides a list of all cities and towns in with their corresponding latitudes and longitudes in a neat table. This data can be scraped into a data frame through the html processing contained in pandas.

Wikipedia (https://en.wikipedia.org/wiki/List_of_villages_in_Massachusetts) contains a table of towns and their counties in Massachusetts in a formatted table. Again using pandas html

read functionality, this data can be scraped into a data frame and the geolocation data joined with the county data through a city/town match case.

Geolocation data for the coordinates of the job location can be obtained through the provided geolocation package and all other points of consideration for the user will be achieved through rigorous data wrangling, calculation, and organized visualization.

Although machine learning might suggest what could be considered ideal through methods such as k-nearest neighbor, the weights of these points will vary wildly between potential users within the target audience. To avoid over-processing and allowing the user to apply weight as they see fit, machine learning will not be used to provide a recommendation.