IN SEARCH OF LOWER COST OF LIVING

JONAS ANDERSSON

IBM COURSERA DATA SCIENCE

2.24.21

INTRODUCTION

Purpose

Everybody has a dream place where they want to live. But as housing prices have skyrocketed during the past years, the area where one ideally would want to live, **may not be in range** when it comes to affordability. In this project, I try to mitigate some of that issue by providing a framework where a person who has seen their **dream area** but think it's too expensive, can see what **other areas that are similar, but cheaper**.

SUMMARY

MAIN CONCLUSIONS

The main conclusion is that there indeed seem to be **alternatives to a perfect area**. The script would ideally only provide the person thinking about moving with viable options. In this specific case, **Seattle was taken as a dream area** to live in, and we found a **possible proxy in Eureka California**.

DATA ACQUISITION

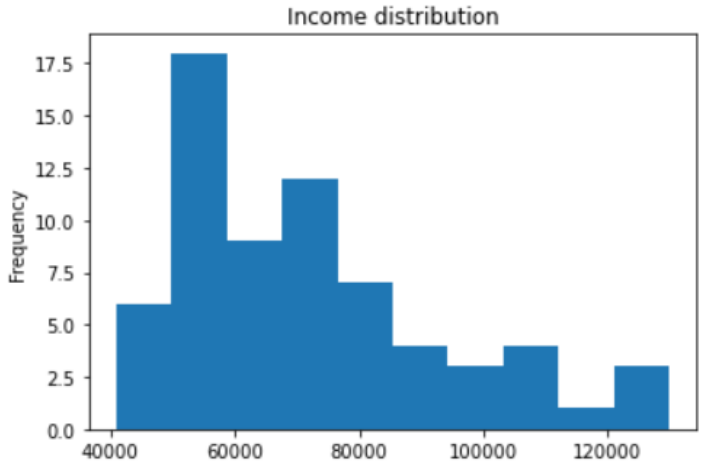
DATA FROM

ST. LOUIS FED

FOURSQUARE

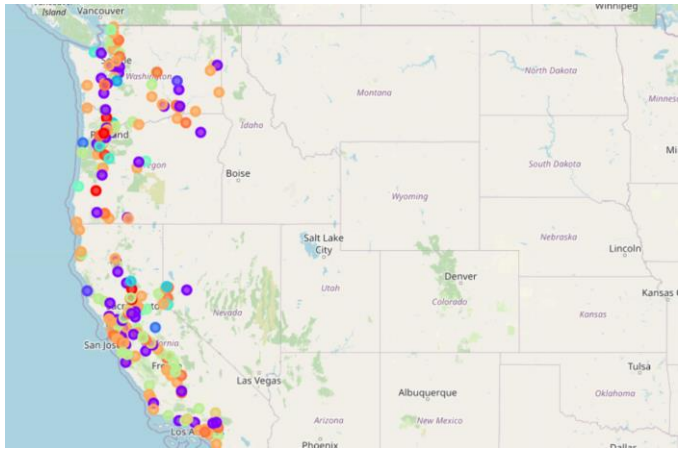
SIMPLEMAPS

SELECT DATA



Seattle cluster income distribution

- Most places are cheaper
- Possible to find something similar for less



Nice dispersion of different clusters

- The clustering algorithm did a decent job in the clustering.
- Only selected west coast cities

	city	Cluster Labels	MedianIncome	state_name	population
202	Pullman	11.0	40858	Washington	34506
31	Crescent City	11.0	43919	California	16849
30	Brookings	11.0	46747	Oregon	11162
49	McKinleyville	11.0	47446	California	17208
47	Eureka	11.0	47446	California	44236
70	Altamont	11.0	49412	Oregon	19341
57	Delano	11.0	51116	California	54917
192	The Dalles	11.0	52575	Oregon	20442
205	Sunnyside	11.0	52764	Washington	18352
203	Yakima	11.0	52764	Washington	133191

Final frame of top 10 cheap alternatives

- Eureka, California is right on the water with a median income of half of Seattle

As there are other metrics than just venues to a city, more data like closeness to water, or mountains should ideally be included to further refine the clustering

DISCUSSION