

Capstone Project Report (Revised)

Finding Alternative City in the Washington State

Fadhel Hariz Dzulfikar

May 2020

1. Introduction

1.1 Background

Washington, formally the State of Washington, is a state in the Pacific Northwest region of the United States. Washington is the eighteenth largest state, with an area of 71,362 square miles (184,827 km²), and the thirteenth most populous state, with more than 7.6 million people. Approximately 60 percent of Washington's residents live in the Seattle metropolitan area, the center of transportation, business, and industry along Puget Sound, an inlet of the Pacific Ocean consisting of numerous islands, deep fjords, and bays carved out by glaciers. The Pacific Northwest, often alluded to as Cascadia, is quite possibly the prettiest region of the United States.[1]

1.2 Problem

Seattle is the largest city in both Washington state and North America's Pacific Northwest region. Under the U.S. Census data released in 2019, the population of the Seattle metropolitan area is 3.98 million and ranks 15th-largest in the United States. It was the fastest-growing major city in the United States in July 2013 and remained in the top five in May 2015 with an annual growth rate of 2.1%. In July 2016, Seattle was again the fastest-growing major U.S. city with an annual growth rate of 3.1%[2]. Some people might think to live in another city (leave Seattle) where the city is not too densely populated, but similar to Seattle, in order to solve population distribution problem. They need some alternatives city to think of where to buy their new house. So, are there nearby city around Washington state that similar to Seattle to live in?

1.3 Target Audience

Target audience of this study are people who might be interested in buying properties, or finding good city to live around the state of Washington.

2. Data

To get a list of cities in Washington State, I scraped Wikipedia page. Then I used the Foursquare venue recommendation API to obtain a list of the most popular venues for each city and find location data (latitude/longitude) using the Mapquest Geocoding API. I used the Folium mapping library to visualize the cities and how they cluster using kmeans cluster method.

2.1 Preprocessing Data

First, I searched google to find Wikipedia page about Washington State. On that page, there was a list of cities in the Washington State. Using pandas library, it was easy to convert those list into a dataframe. I found 212 cities in the Washington State.

	City	State	Latitude	Longitude
0	Aberdeen	Washington		
1	Airway Heights	Washington		
2	Algona	Washington		
3	Anacortes	Washington		
4	Arlington	Washington		
5	Asotin	Washington		
6	Auburn	Washington		
7	Bainbridge Island	Washington		
8	Battle Ground	Washington		
9	Bellevue	Washington		
10	Bellingham	Washington		
11	Benton City	Washington		
12	Bingen	Washington		

Figure 1. Sample of cities listed on df_wa

Next, I used MapQuest API to get location of each city in Washington state. The location data are in the form of latitude and longitude coordinates. You need to sign up on MapQuest website to get credential information that will be used to request data from the MapQuest API. Here is the sample of data that I got.

	City	State	Latitude	Longitude
0	Aberdeen	Washington	46.9755	-123.816
1	Airway Heights	Washington	47.643	-117.593
2	Algona	Washington	47.2791	-122.25
3	Anacortes	Washington	48.5054	-122.632
4	Arlington	Washington	48.1913	-122.126
5	Asotin	Washington	46.3406	-117.049
6	Auburn	Washington	47.3075	-122.226
7	Bainbridge Island	Washington	47.6431	-122.527
8	Battle Ground	Washington	45.7807	-122.548
9	Bellevue	Washington	47.6137	-122.191
10	Bellingham	Washington	48.7549	-122.478
11	Benton City	Washington	46.2649	-119.488

Figure 2. Sample of cities listed on df_wa after requesting data on MapQuest API

Using above information, I could render map of Washington State to visualize location of each city on the list. In order to do that, I used Folium mapping library to visualize the data. Here is the result of rendering the map.

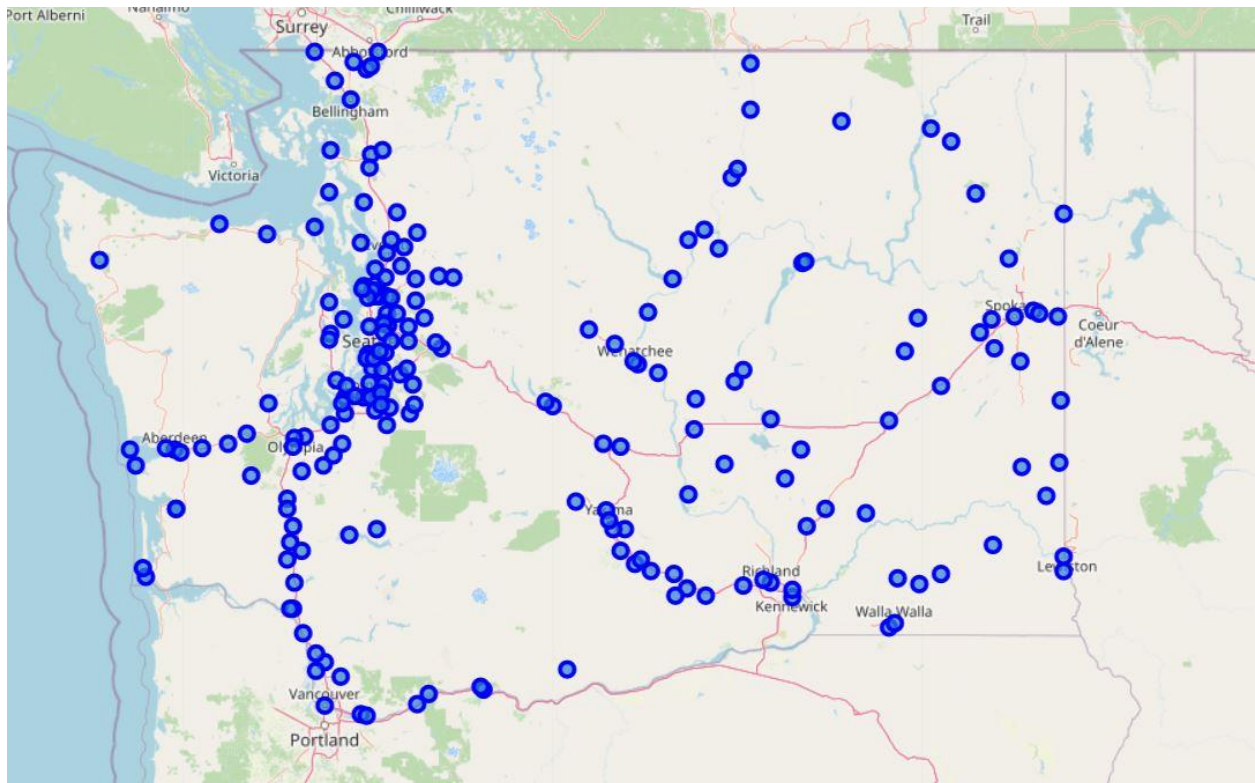


Figure 3. Map of Washington States

Next, I used the Foursquare API to query the top 5 venues in each city based on user rating of Foursquare apps. On this step, I got 2740 different venues spread in the Washington State. In order to make the clustering easier, I filtered out some cities that have less than 10 avenues. After this filter step, I got 82 cities with more than 10 avenues each and 301 unique categories.

3. Methodology

To analyze the data of the project, I used KMeans clustering to group the cities in the Washington State that have similarities with each other. First, I needed to calculate frequency of the top 5 venues in each city. Using pandas dataframe library, I could convert each venue category into Boolean using one hot encoding method. By using those data I could find the most common venues for each city and listed the top 5.

Then, I used KMeans clustering algorithm to start grouping the city. The output of KMeans clustering is a column of cluster of each city. After that, I could visualize each cluster using Folium mapping library. Here is the result.

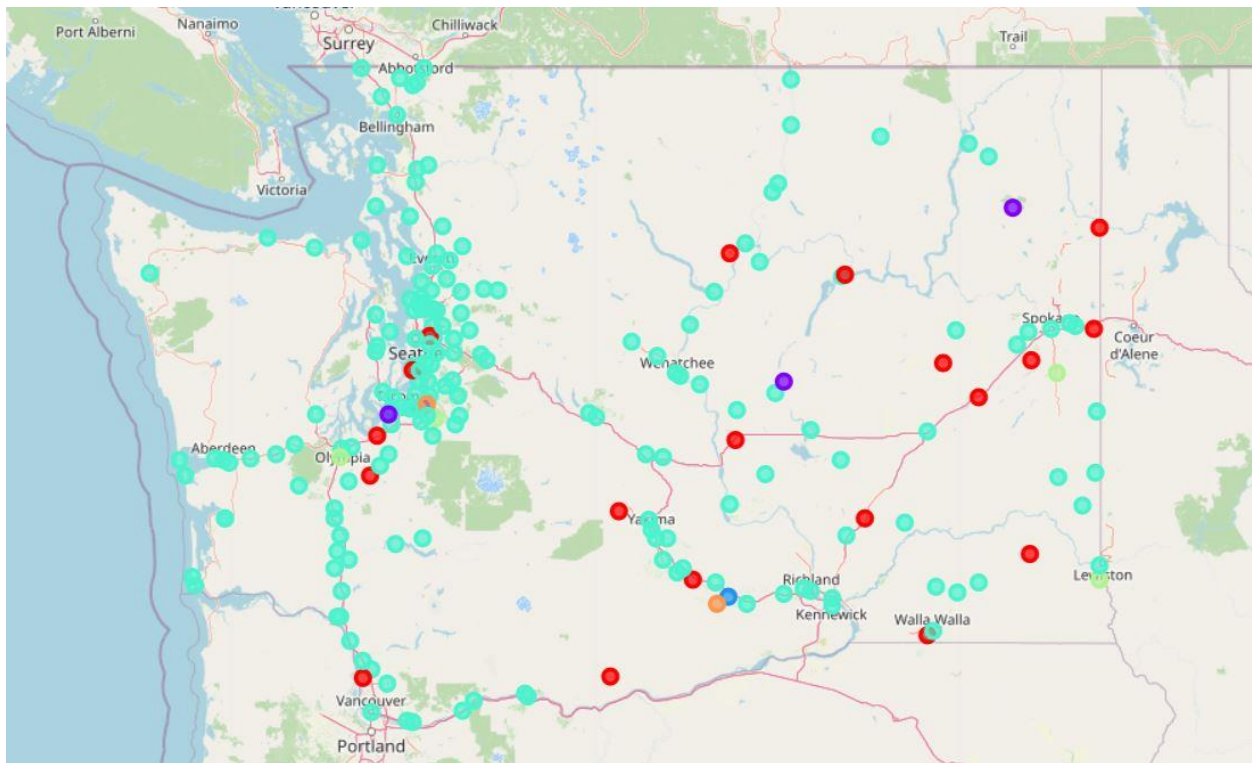


Figure 4. Clusters of Washington States

4. Analysis

- Cluster 0: It has 19 cities with mainly parks and farms avenue. So I think these cities have many wide open areas and suitable for people who love nature.
- Cluster 1: There are only four cities listed in cluster 1. Therefore, this cluster cannot be used as reference to find your city.
- Cluster 3: This cluster has many cities on the list. The most striking feature is that Seattle belongs to this cluster. So, I think these cities on the list have many similarities to Seattle and might be people's alternative to find home where the city are less populated than Seattle but have Seattle vibes. Moreover, the cities on this cluster have many entertainment avenues as the top five avenues such bar, restaurant, cafe, coffee shop, and many more which is good for business.
- Cluster 4: It has only four cities listed on this cluster. The top avenues here are fishing spot and farmers market. I think this cluster has similar vibes with cluster 0.
- Cluster 5: This cluster has unique avenues on the top 5 avenues such as candy store and falafel restaurant making them unique in this dataset, but unreliable.

5. Discussion

To get those information, we used MapQuest API. After that, we used FourSquare API to get FourSquare query of venues, there were 2740 venues in the Washington State including 301 unique category venues. To solve the problem, we used KMeans clustering in order to group the cities into 6 clusters. However, there are only four data cluster that are reliable. This project needs to be improved whether using more reliable dataset or clustering method.

6. Conclusion

Based on this project, we sure can answer the main problem stated in the beginning of this project that is "are there nearby city around Washington state that similar to Seattle to live in?". And the answer are listed on cluster 3. To get those information, we used MapQuest API. After that, we used FourSquare API to get FourSquare query of venues, there were 2740 venues in the Washington State including 301 unique category venues. To solve the problem, we used KMeans clustering in order to group the cities into 6 clusters. However there are only four data that are reliable. This project needs to be improved whether using more reliable dataset or clustering method. Based on this project, we sure can answer the main problem stated in the beginning of this project that is "are there nearby city around Washington state that similar to Seattle to live in?". And the answer are listed on cluster 3.

References

- [1] https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_Washington
- [2] <https://developer.foursquare.com/docs/api/venues/explore>
- [3] <https://developer.mapquest.com/documentation/geocoding-api/>
- [4] <https://python-visualization.github.io/folium/>