

# Moving from large to small city: Big difference?

Yevgen Kryukov

August 2, 2020

## 1. Introduction

### 1.1 Background

In the United States it is common to switch jobs and relocate from city to city, however relocation often poses a significant discomfort, as new place might be quite different from the place person one currently likes and enjoys a lot. One of the drastic changes could be if a person considers moving from a large to a small city or vice versa, in this project we will explore the problem of moving from large to small city and try to figure out if there are similar neighbourhoods with similar venues and lifestyle.

### 1.2 Interested audience

The target audience for such a project is anyone who is considering moving and would like a quick comparison between current and potential new location. While in this project we only looking at venues, it could be significantly expanded to compare additional resources such as schools, crime, housing prices etc.

## 2. Data

### 2.1 Data Sources

We are going to choose two random cities one large (Washington, DC) and one small (Kalamazoo, MI) and locate postal codes of the neighborhoods in this these cities from the following source:

<https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/?q=kalamazoo&location=12,38.83316,-77.1678&basemap=jawg.streets>

<https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/?q=washington&refine.state=DC&location=12,38.83316,-77.1678&basemap=jawg.streets>

Once data is preprocessed, Foursquare API will be used to extract venues for both cities. The data includes:

- Place
- Name
- Latitude and Longitude of the place
- Neighborhood
- Category

- Other venue details.

### 3. Methodology and Exploratory Data Analysis

First, Geopy tool was used to perform visual analysis to make sure neighborhoods coordinated shown on correctly on the map. Longitude and latitudes of both Alexandria and Kalamazoo were calculated.

```
[7]: address = 'Alexandria, VA'

geolocator = Nominatim(user_agent="ny_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Alexandria are {}, {}'.format(latitude, longitude))
```

The geograpical coordinate of Alexandria are 38.8949855, -77.0365708.

```
[8]: address = 'Kalamazoo, MI'

geolocator = Nominatim(user_agent="ny_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Kalamazoo are {}, {}'.format(latitude, longitude))
```

The geograpical coordinate of Kalamazoo are 42.291707, -85.5872286.

Next, using Foursquare API, names and coordinates of venues were extracted and

(1919, 7)

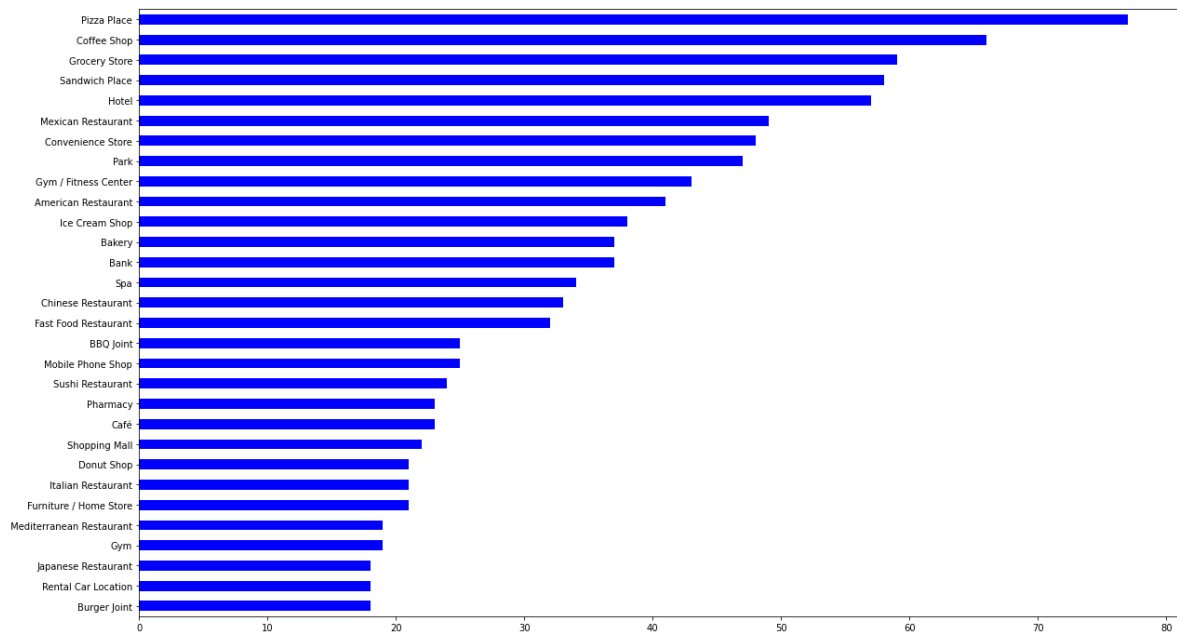
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	22320	38.804413	-77.046699	Old Town Alexandria	38.805065	-77.047792	Neighborhood
1	22320	38.804413	-77.046699	Nando's Peri-Peri	38.805002	-77.047422	Portuguese Restaurant
2	22320	38.804413	-77.046699	Sugarhouse Day Spa & Salon	38.805927	-77.048893	Spa
3	22320	38.804413	-77.046699	The Spice & Tea Exchange of Alexandria	38.804555	-77.043206	Herbs & Spices Store
4	22320	38.804413	-77.046699	The Alexandrian, Autograph Collection	38.804445	-77.044550	Hotel

```
[88]: #Kalamazoo Venues
print(kz_venues.shape)
kz_venues.head()
```

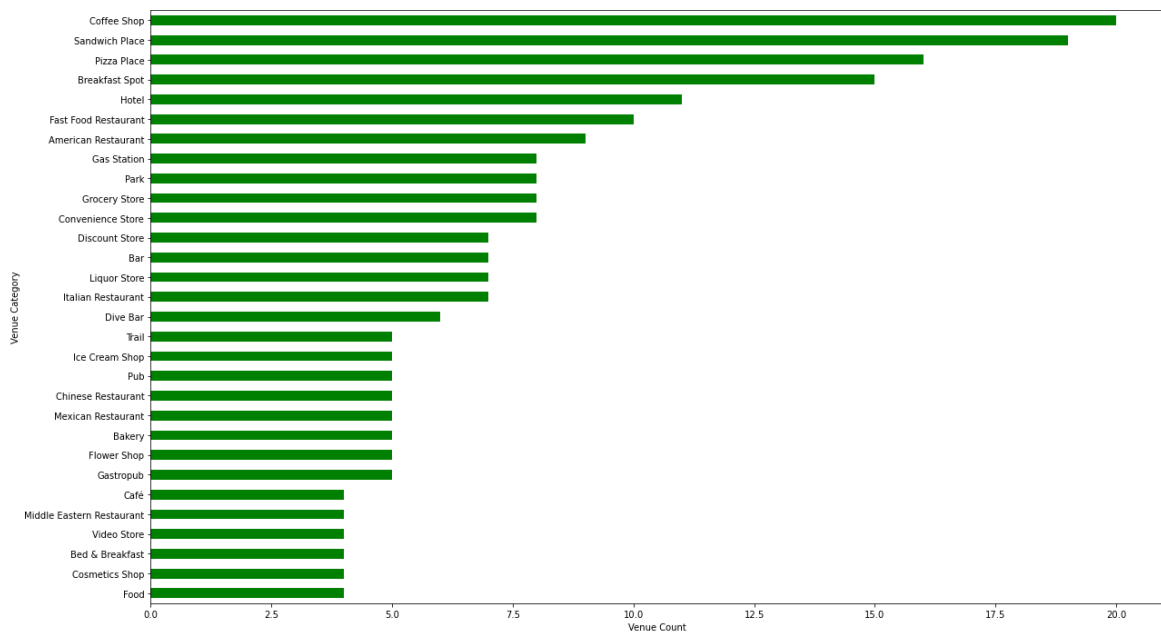
(381, 7)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	49005	42.323045	-85.493232	BIGGBY COFFEE	42.328002	-85.513721	Coffee Shop
1	49005	42.323045	-85.493232	Qdoba Mexican Grill	42.327867	-85.514227	Mexican Restaurant
2	49005	42.323045	-85.493232	Culver's	42.329434	-85.510709	Fast Food Restaurant
3	49005	42.323045	-85.493232	Penn Station East Coast Subs	42.327758	-85.513881	Sandwich Place
4	49005	42.323045	-85.493232	SUBWAY	42.331761	-85.510393	Sandwich Place

Frequency of each type of venue was calculated and subsequently graphed as a bar chart



*Figure 1: Top 30 Venues Count For Alexandria, VA*



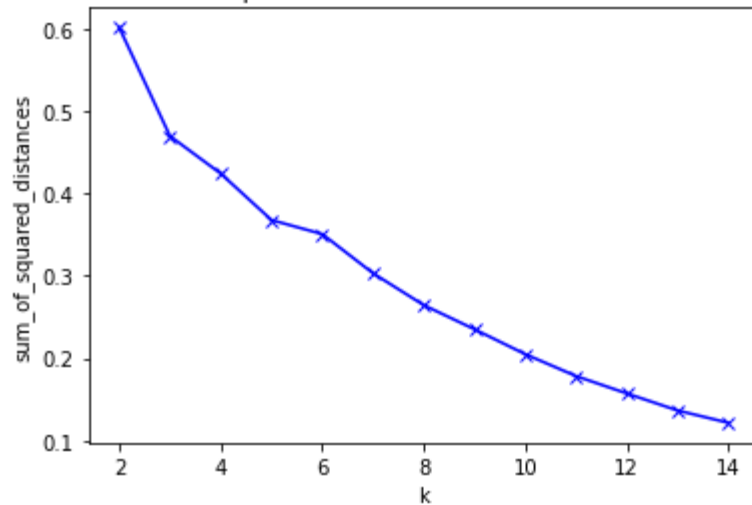
*Figure 2: Top 30 Venues Count For Kalamazoo, MI*

From the graph above, we can see that Alexandria is much bigger city with many more venues, however the types of top venues are similar for both cities. This is good news for someone looking to relocate.

Next step is to cluster neighborhoods using unsupervised k-means method will be based on frequency and types of venues. An elbow method was used in order to decide the number of optimal clusters. The optimal number corresponds to the bend in the graph similar to how elbow bends (see picture below)

234567891011121314

Elbow method for Optimal K for Alexandria and Kalamazoo Clustering



Though this graph doesn't show clear bending area, it looks like there is a slight bend around cluster size of 5, which was decided to use for clustering in this project.

After clustering was complete corresponding labels with clustered neighborhoods were placed on the map using Geopy tool.

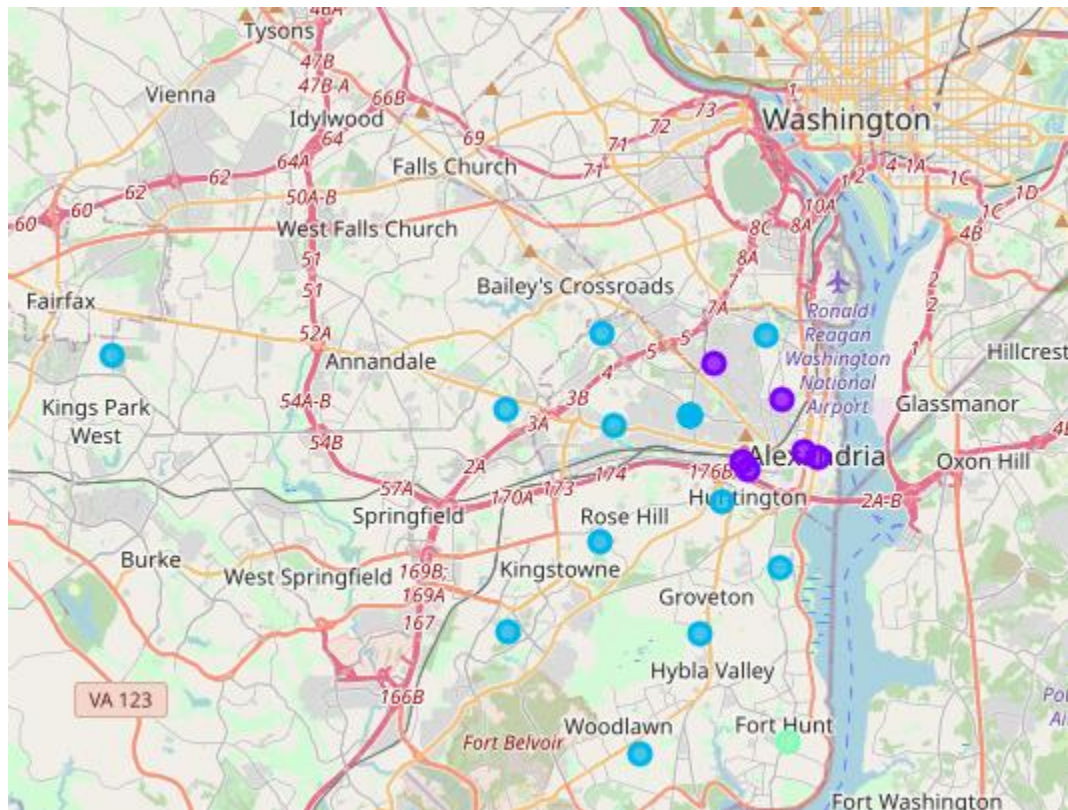


Figure 4: Clustered neighborhoods in Alexandria

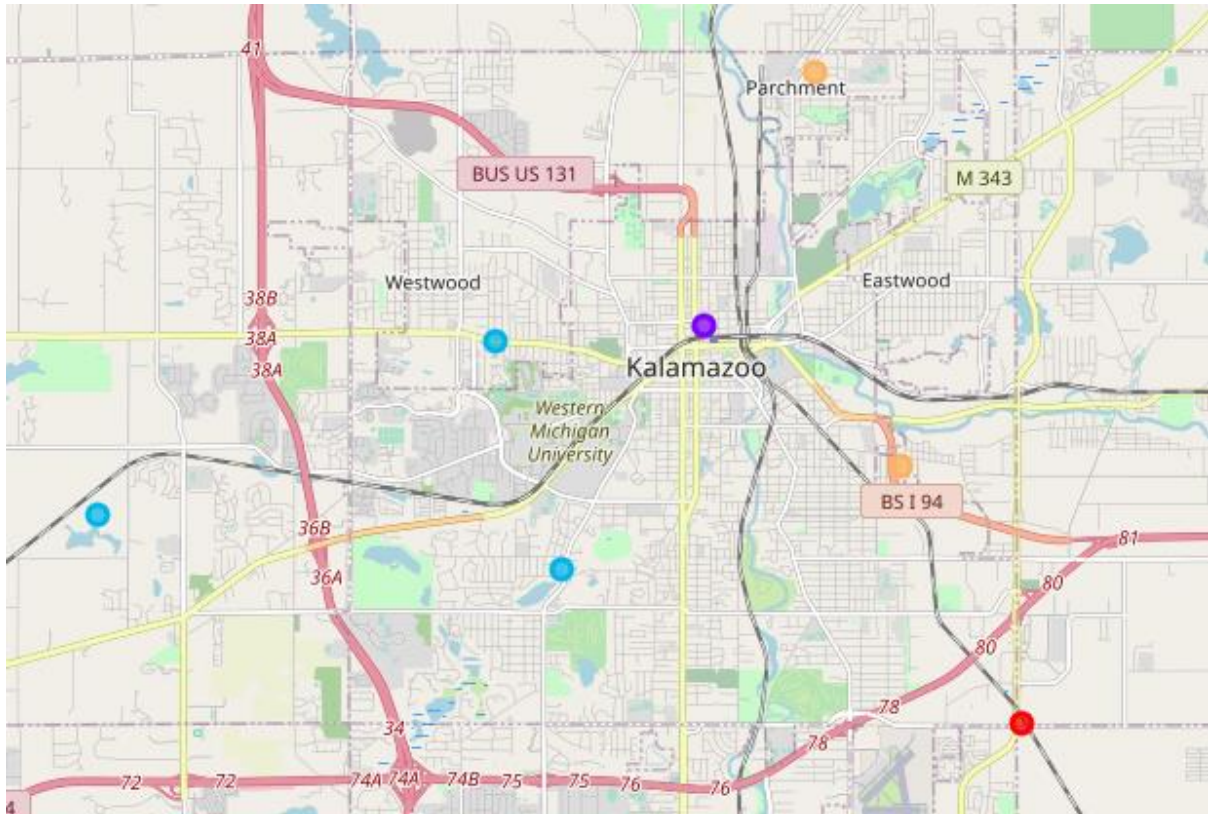


Figure 5: Clustered neighborhoods in Kalamazoo.

## 4. Results and Discussion

After exploring details of each cluster it was observed that there are only two clusters that present in both cities. Purple cluster is 1 and it corresponds to downtown area densely populated with cafes, restaurants and hotels. Second cluster is light blue and it corresponds to residential area populated with casual eateries such as sandwich shops, pizzas and grocery stores.

## 5. Conclusion

If one decides to relocate from Alexandria to Kalamazoo, based on the observation the person should be able to find similar neighborhood either in residential or downtown area. However, considering that Alexandria has significantly larger number of venues this person should explore downtown area first as it would offer more venues opportunities to reflect Alexandria lifestyle. Though it was not part of this project, it could be very interesting to do additional comparison of the cities based on other metrics besides venues. If one has family than schools parks, and safety of neighborhoods could be a priority. Additionally, it would be interesting to compare the ratings of the venues in both cities: are they higher rated in one on them or on par?