Capstone Project - The Battle of Neighbourhoods: Should I move from Lausanne to Bern?
Christopher Tse, Dec 2019

1. Introduction

I'm currently living in a city called Lausanne, on the French-speaking side of Switzerland, but there may be more job and life opportunities in a German-speaking city called Bern in Switzerland. I want to use data science to help me decide which city would suit me best to live.
I want to compare the livability of the two cities through using similar methods that have been used in the assignments from this course, and help me decide which area would be best for me - a young adult.

1.1 Problem

I have not been to Bern very often and don't know the area well. I don't know of a method to survey areas of a city to quickly assess the types of venues that are nearby. It would be immensely valuable for me to be able to see on a map, the kind of neighbourhood a location is to quickly give me an impression If it's a liveable place for me.

1.2 Interest

This can be used by anyone, who wants to assess the local area to help them decide if it's a suitable place for them to live, or open new business opportunities by identifying missing wanted venues.

2. Methodology - Data acquisition and cleaning

Using Foursquare, I first start at the train station from each city. From the train station, I select all Migros supermarkets in the region. From each Migros supermarket, I list out all nearby venues within 1 km, compile into a list and use kcluster to match similar neighbourhoods together. These will be displayed onto an interactive map with folium and help me decide which areas I would like to live. This will be done with k-means to segment and cluster related neighbourhoods to allow me to estimate the suitable demographic to live there. I will be interested in areas which are near a supermarket for convenience, and survey the types of venues in the area. The use of folium interactive maps would be very useful here also so I can visually see attractive features like rivers and parks in the respective areas to strengthen the validity of my recommendations.
There is a limit to how many venues each query can retrieve (up to 50), therefore I will need to break up my query into several queries to cover as much of each city as possible. A local governmnet-run supermarket chain called Migros are well distributed across all cities to provide the sale of groceries to the local populations. I will use these as local "neighbourhoods" to capture as many venues as possible.

2.1 Data sources

In this project, I used Foursquare's venue data to compare the most interesting regions of Lausanne and Bern.

2.2 Data cleaning

When retrieving the data from Foursqure, the data is collected in a JSON file, which needs to be converted into a pandas dataframe for further manipulation to only have the necessary data still present. Such unnecessary labels include "hasPerk", "location.cc", "location.country" etc.
What was kept was parameters which were useful such as id, lat and lng (identity, latitude, longitude respectively).
Such data types allowed me to create interactive maps to show the areas of interest.

Results and Discussion

Using the train station as a starting point, all Migros supermarkets are identified with the Foursquare API in the city (Fig. 1). They can be seen to be dotted around the cities, with an aggregated cluster in their respective town centres.

- Lausanne -

- - -

Bern -

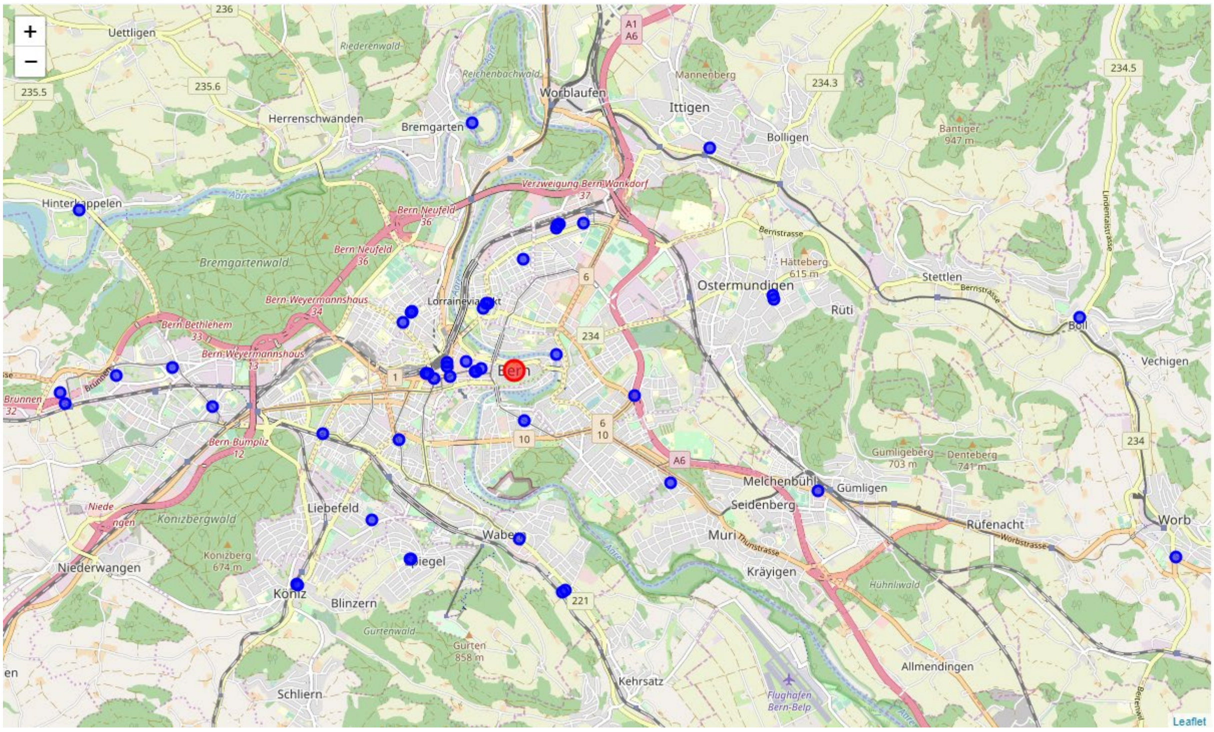


Fig. 1 showing screenshot of folium map of Migros in each city (blue), centred around the train station (red).

Using the supermarket at a reference point provides good coverage of each city. Having identified the supermarkets, their latitude and longitude positions are recorded and the “getNearbyVenues” function was used to list all local venues around each Migros supermarket. The closest venues are When writing the script, I used each venue’s unique id (“id”) as their name so each entry was unique because many of the supermarkets are simply called “Migros” and manipulation of the data would have aggregated results in later pieces of code.

A total of 1612 venues were collected from 131 unique categories, ranging from Art galleries to Zoos (Fig. 2)

```
In [38]: print(lausanne_areas.shape)
         lausanne_areas.head()

(1612, 7)
```

Out[38]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	4b644e06f964a5208ba92ae3	46.948132	7.444728	Bundesplatz	46.947414	7.443817	Plaza
1	4b644e06f964a5208ba92ae3	46.948132	7.444728	Adriano's Bar & Café	46.947914	7.447428	Café
2	4b644e06f964a5208ba92ae3	46.948132	7.444728	Zytglogge	46.948084	7.447555	Monument / Landmark
3	4b644e06f964a5208ba92ae3	46.948132	7.444728	Bellevue Palace Bern	46.946799	7.446866	Hotel
4	4b644e06f964a5208ba92ae3	46.948132	7.444728	Migros	46.948132	7.444728	Supermarket

```
In [39]: print('There are {} uniques categories.'.format(len(lausanne_areas['Venue Category'].unique())))

There are 131 uniques categories.
```

Fig. 2 showing a caption of the code that collected the total venues and categories all around Lausanne

To view each area more efficiently, the data was aggregated into a pandas dataframe with the top 10 venues shown (Fig. 3). By viewing the most common venues, I can get a grasp of the kind of environment each area has.

```
In [47]: neighborhoods_venues_sorted
```

Out[47]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	4b51ead6f964a520385b27e3	Supermarket	Bus Station	Pool	Park	Food Truck	Grocery Store	Shopping Mall	Asian Restaurant	Department Store	Bus Stop
1	4b58503cf964a520425228e3	Supermarket	Fast Food Restaurant	Grocery Store	Light Rail Station	Coffee Shop	Mexican Restaurant	Frozen Yogurt Shop	Shopping Mall	Sushi Restaurant	Swiss Restaurant
2	4b5f1cb1f964a520baa629e3	Bar	Restaurant	Café	Italian Restaurant	Creperie	Park	Electronics Store	Event Space	Pizza Place	Israeli Restaurant
3	4b644e06f964a5208ba92ae3	Café	Italian Restaurant	Plaza	Bar	Park	Swiss Restaurant	Hotel	Restaurant	Burger Joint	Food & Drink Shop
4	4b6f2023f964a520bdde2ce3	Plaza	Italian Restaurant	Café	Park	Restaurant	Monument / Landmark	Swiss Restaurant	Creperie	Hotel	Food & Drink Shop
5	4b75b135f964a520661d2ee3	Supermarket	Restaurant	Discount Store	Soccer Field	Swiss Restaurant	Bistro	Bus Station	Paper / Office Supplies Store	Office	Grocery Store
6	4b7a85a3f964a520bb2f2fe3	Plaza	Italian Restaurant	Café	Park	Restaurant	Monument / Landmark	Swiss Restaurant	Creperie	Hotel	Food & Drink Shop
7	4b7acda1f964a520923d2fe3	Café	Plaza	Creperie	Park	Restaurant	Vegetarian / Vegan Restaurant	Pizza Place	Bar	Hotel	Coffee Shop
8	4b7cee3cf964a5208ca92fe3	Tram Station	Supermarket	Hotel	Food & Drink Shop	Light Rail Station	Beach	Grocery Store	Restaurant	Bus Stop	Electronics Store
9	[illegible]	Italian Restaurant	Swiss Restaurant	Hotel	Bus Station	Restaurant	Stadium	Supermarket	Shopping Mall	Tram Station	Laundry Service

Fig. 3 showing a caption of the code that ranked the top 10 venues at each area in Lausanne

With 49 different areas to assess, and each having the 10 most common venues, I would need to inspect 490 fields of data for each city, which is too much for convenience. A better way to arrange and sort the venues was required and needed further data manipulation. Instead of reviewing each of the 49 areas, I used kcluster with a k-value of 8 to cluster similar areas together, reducing the analysis from 490 fields of data to 80 (Table 1). 8 was chosen as this was the first significant “elbow” in a range of k values (Fig. 4).

- Lausanne -

Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Shopping Mall	Supermarket	Department Store	Dance Studio	Bar	Fast Food Restaurant	Shoe Store	Stadium	Light Rail Station	Kebab Restaurant
1	Grocery Store	Café	Italian Restaurant	Supermarket	Sushi Restaurant	Train Station	Chinese Restaurant	Theater	Japanese Restaurant	Brewery
2	Pizza Place	Home Service	Train Station	Zoo Exhibit	Dessert Shop	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store	Discount Store
3	Supermarket	Gas Station	Hotel	Electronics Store	Shopping Mall	French Restaurant	Construction & Landscaping	Fast Food Restaurant	Restaurant	Massage Studio
4	Hotel	Train Station	Business Service	Furniture / Home Store	Tennis Stadium	Miscellaneous Shop	Health & Beauty Service	Fast Food Restaurant	Construction & Landscaping	Convenience Store
5	Bar	Burger Joint	Swiss Restaurant	Italian Restaurant	Brewery	Art Museum	Pizza Place	Chinese Restaurant	Plaza	French Restaurant
6	Train Station	Department Store	Clothing Store	Pizza Place	Furniture / Home Store	Pet Store	Shopping Mall	Outdoors & Recreation	Zoo Exhibit	Electronics Store
7	Supermarket	Swiss Restaurant	Farmers Market	Bakery	Diner	Food & Drink Shop	Fast Food Restaurant	Falafel Restaurant	Electronics Store	Discount Store

- Bern -

Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Plaza	Italian Restaurant	Café	Park	Restaurant	Monument / Landmark	Swiss Restaurant	Creperie	Hotel	Food & Drink Shop
1	Train Station	Mini Golf	Shopping Mall	Flower Shop	Grocery Store	Gym	Historic Site	Discount Store	Dessert Shop	French Restaurant
2	Gym Pool	Auto Garage	Grocery Store	Lake	Fast Food Restaurant	Forest	Food Truck	Food & Drink Shop	Food	Flower Shop
3	Supermarket	Swiss Restaurant	Tram Station	Restaurant	Bar	Hockey Arena	Park	Discount Store	Buffet	Light Rail Station
4	Mexican Restaurant	Grocery Store	Train Station	Zoo	Food	French Restaurant	Forest	Food Truck	Food & Drink Shop	Flower Shop
5	Train Station	Restaurant	Grocery Store	Spa	Fast Food Restaurant	Supermarket	Furniture / Home Store	Shopping Mall	Shoe Store	Sporting Goods Shop
6	Discount Store	Grocery Store	Bus Station	Sandwich Place	Gas Station	Bakery	Train Station	Supermarket	Food & Drink Shop	Swiss Restaurant
7	Supermarket	Bus Station	Pool	Park	Food Truck	Grocery Store	Shopping Mall	Asian Restaurant	Department Store	Bus Stop

Table 1. Comparative study of the top 10 venues in each city of Lausanne and Bern, by each cluster generated by kcluster with a k-value of 8

```
k_list = list(range(k_min,k_max+1))
plt.xlabel('k')
plt.ylabel('Inertia')
fig = plt.plot(np.array(k_list), np.array(inertia_list))
plt.savefig('inertia plot.png')
```

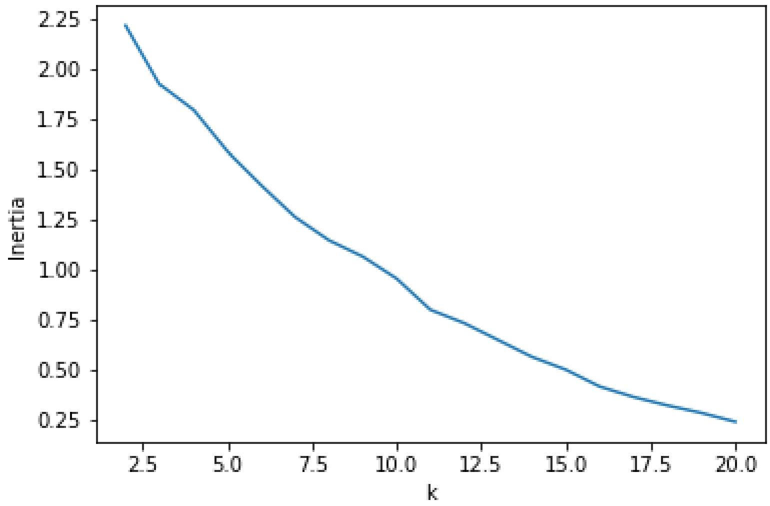


Fig. 4 showing k_list and the first significant bump occurs when k = 8.

Table 1 gives an excellent summary of each cluster that can help me narrow down my ideal area to live. However, it is still quite text heavy and some of the venues are of no interest to me. It would help to further refine the data so it's easier to discriminate against unwanted venues in each area. With the top venues now visible for each area, I can simplify each unique category to better decide which area is more attractive. I want to live near public transportation, food establishments, close to shopping venues and have leisure activities close by. All other things which are of interest to me can be considered unwanted i.e. dance studio, gas station, health and beauty service etc (Table 2). I created Table 2 through assessing each of the unique venues by hand. This could have been automated if I had assigned a label for each unique venue and automatically create a new table for me. This was considered too labour intensive at this time as I only wanted to generate two tables. If I use this same script presented in this project for several thousand different cities and areas of interest, it would be worthwhile to create this sort of table automatically through using pandas and dataframe manipulation.

Lausanne

Cluster	Public Transport	Food Shopping	Consumer shopping	Leisure	Restaurant	Unwanted
0	2	2	1	2	2	1
1	1	2		3	4	
2	1	1	3	1	4	
3		1	2		3	4
4	1	2	2	1		4
5			1	3	6	
6	1		4	2	1	2
7		4	1		4	1

Bern

Cluster	Public Transport	Food Shopping	Consumer shopping	Leisure	Restaurant	Unwanted
0		1	1	4	3	1
1	1	2	3	3	1	
2		3	1	3	2	1
3	2	1	1	3	3	
4	1	4	1	2	2	
5	1	2	3	1	2	1
6	2	5	1		1	1
7	2	3	2	2	1	

Table 2. Sorting all venues into relevant labels to better assess the attractiveness of each cluster.

With the data provided by Table 2, it is easier to pick the clusters which would be of least interest for me i.e. clusters with "unwanted" venues. Ideally there should be a lot of leisure venues, little consumer shopping venues and at least one public transport venue nearby. By setting these criteria, I get Table 3, which shows that cluster 1 in Lausanne, and cluster 3 and 7 in Bern would be ideal as they fulfil my criteria.

Lausanne

Cluster	Public Transport	Food Shopping	Consumer shopping	Leisure	Restaurant	Unwanted
1	1	2		3	4	

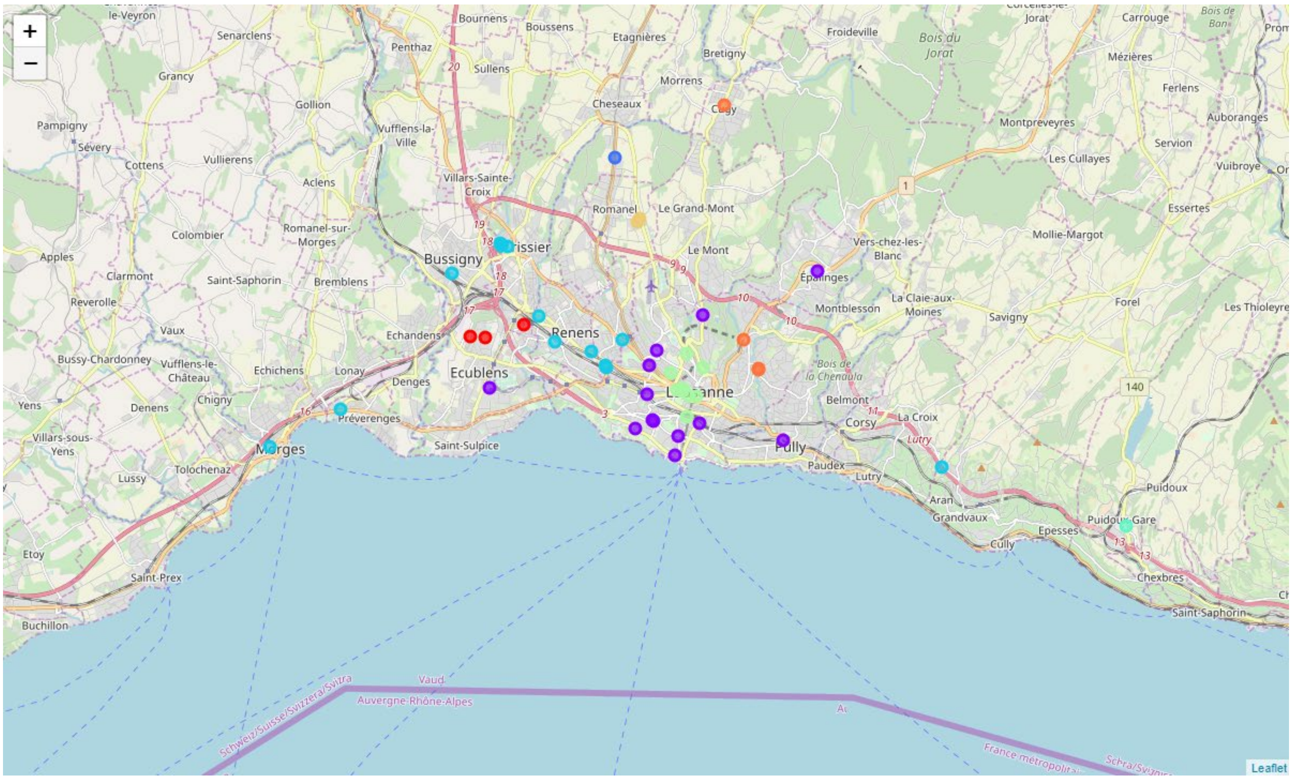
Bern

Cluster	Public Transport	Food Shopping	Consumer shopping	Leisure	Restaurant	Unwanted
3	2	1	1	3	3	
7	2	3	2	2	1	

Table 3.

To better visualise the clusters, Folium was used again and this time each Migros supermaket was colour-coded to the cluster they are associated with (Fig. 5). Cluster 1 in Lausanne is coloured purple, and cluster 3 and 7 in Bern are coloured faded blue and orange respectively.

- Lausanne -



- Bern -

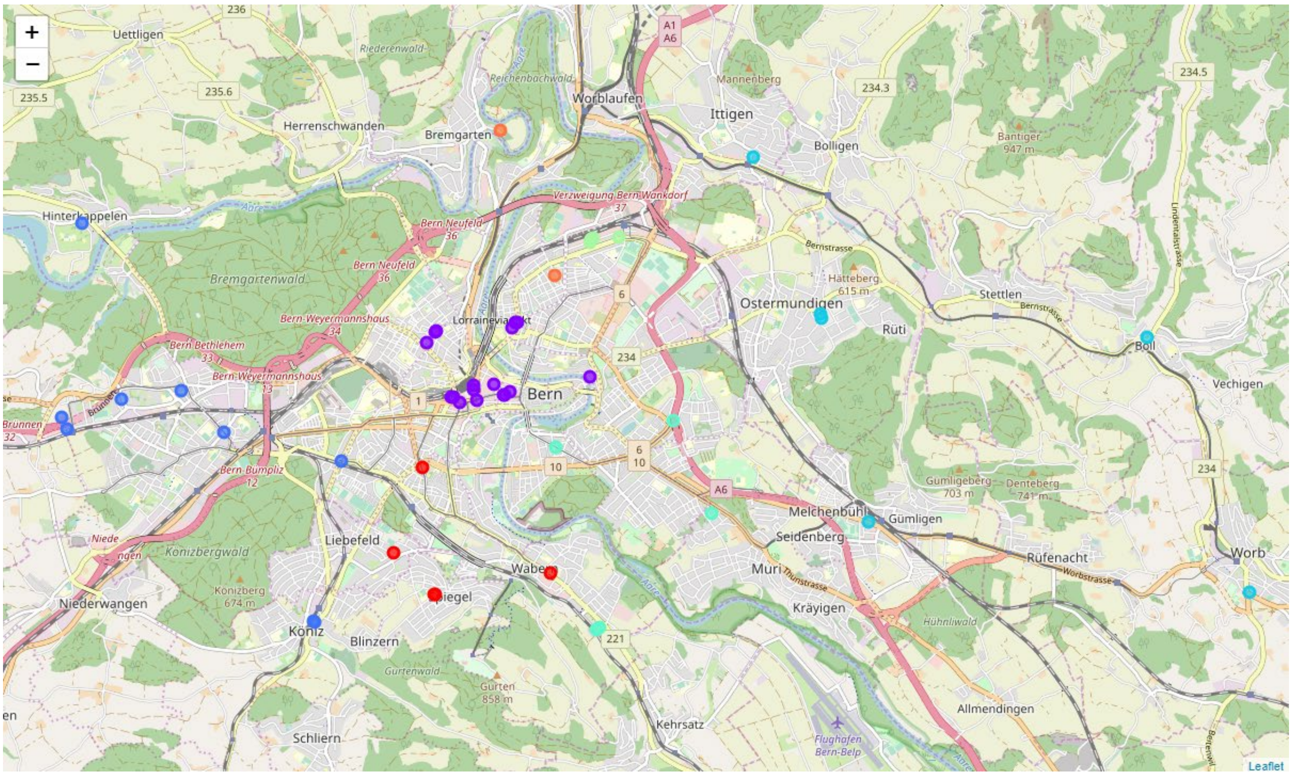


Fig. 5 showing screenshot of folium map of Migros in each city, colour-coded to each clister; Red = 0, Purple = 1, Navy Blue = 2, Faded Blue = 3, Turquoise = 4, Lime Green = 5, Mustard Yellow = 6, Orange = 7

Conclussion

The folium interactive map is very useful to represent the data and is very easy to survey the local area.

With Lausanne, through the use of data science, it was concluded that areas around cluster 1 (purple circles in the map of Fig. 4) are ideal. Seeing the locations, they are situated mostly in the city centre and there are three areas outside of the city centre; at Ecublens, Pully and Epalinges. Ecublens, with its close proximity to public transport and nice location, would be of most interest to me to live.

With Bern two areas identified with kcluster were ideal for me; 3 and 7 (faded blue and orange in Fig. 4). With this information, combined with the folium map, I conclude that cluster 7 (orange) would not be suitable as it looks too far away from the city and is too rural for me, therefore cluster 3 (faded blue) would be of most interest, specifically the orange area nearest to the centre of Bern.

I can conclude that if I were to move to Bern, I would want to move near the North of the city centre where there are the most attractive venues for me. If I should stay in Lausanne, then living somewhere near Ecublens would be ideal.

This script created for this project allows:

1. Contextual summary through combining Foursquare API and folium to assess attractiveness of each area.
2. Can be used for any city and any point of interest in the world making this translational for not just my needs, but for businesses and organisations to create contextual data of areas of interest.
3. Have up-to-date data as long as people still use the Foursquare API
4. Tailor the methodology easily i.e. radius, number of kclusters, vary the number of most common venues, change city and change areas to search can all be simply amended in the code.

However I did stumble upon some drawbacks which had caused inaccuracies and errors in the data;

1. Venues can have multiple unique venue ids, and human errors due to decentralised method of data collection by the general public. This is the strongest argument against using this technique. Less popular cities may not have up-to-date data. This is popular in the US, but less popular in Europe which can affect their accuracy. All further disadvantages to this method stems from this weakness.
2. There are many overlapping categories on the “restaurant” section and for future work, all restaurants could be consolidated into a single category so that other unique venues can be distinguished in the 10 top venues of an area.