Capstone Project - The Battle of Neighbourhoods: Should I move from Lausanne to Bern?

Christopher Tse, Dec 2019

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.

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.

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.

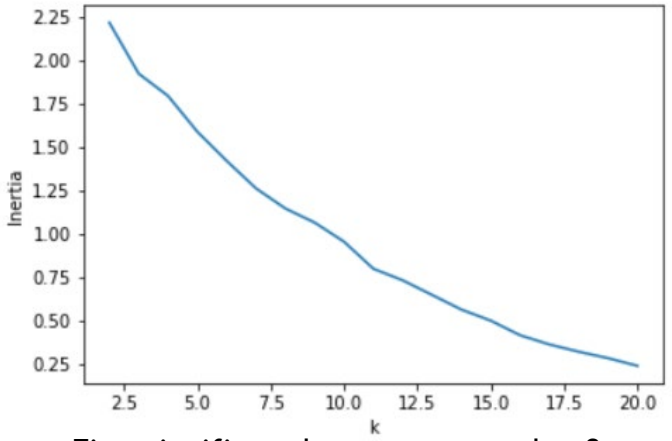
Methodology - Data acquisition and cleaning

- **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.

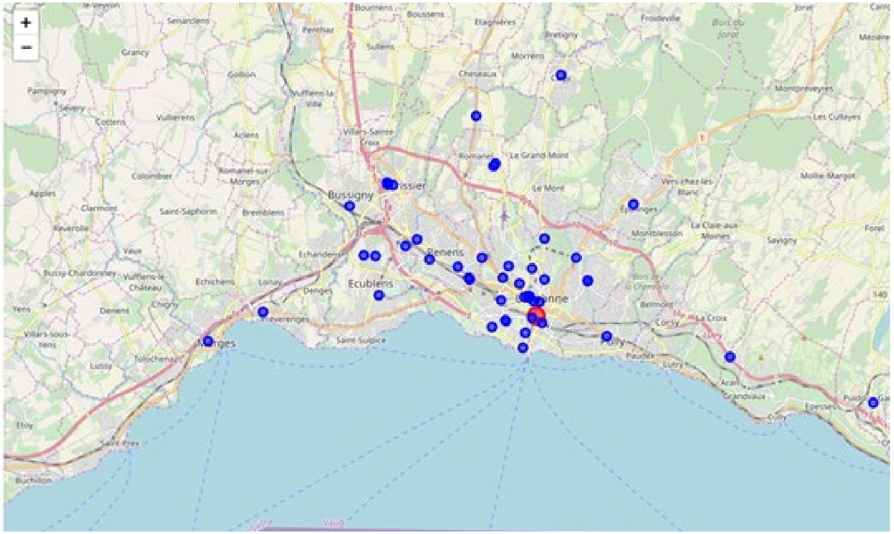


First significant bump occurs at k = 8

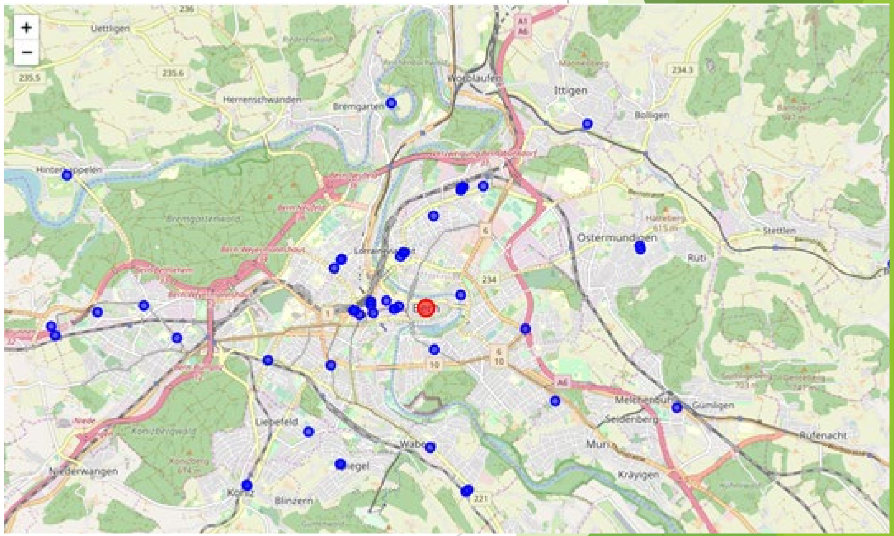
• Results and Discussion

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

- Lausanne -



Bern -



Most common venues in each cluster

- Lausanne -

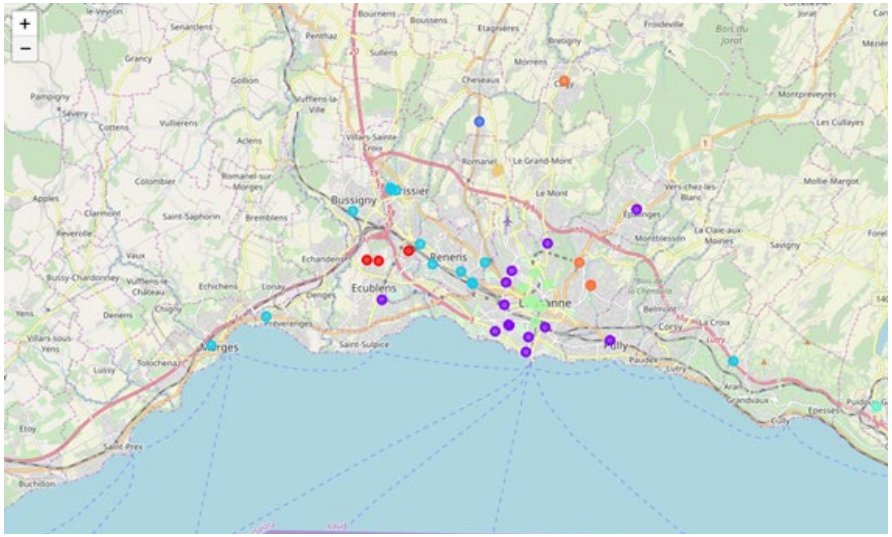
Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Shopping Mall	Supermarket	Department Store	Dance Studio	Bar	Fast 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 -

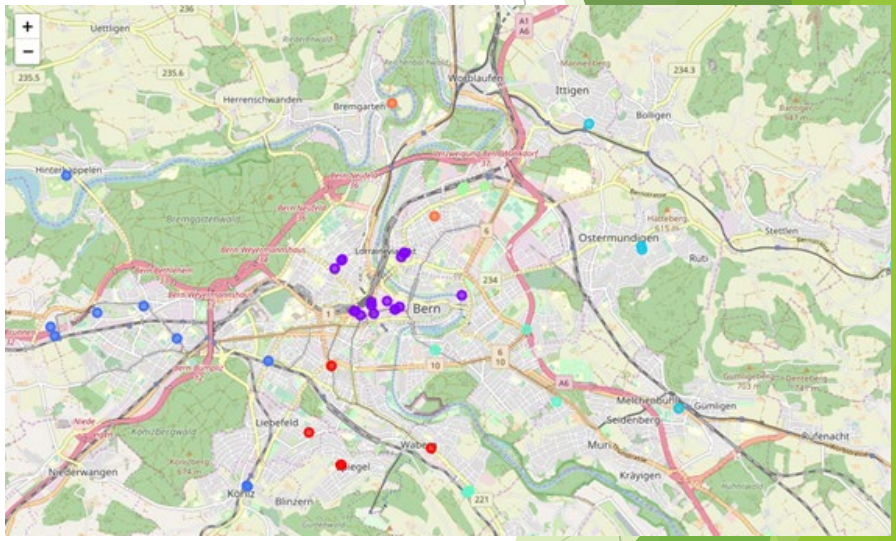
Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Plaza	Italian Restaurant	Café	Park	Restaurant	Monument / Landmark	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

Folium map to show each cluster

- Lausanne -



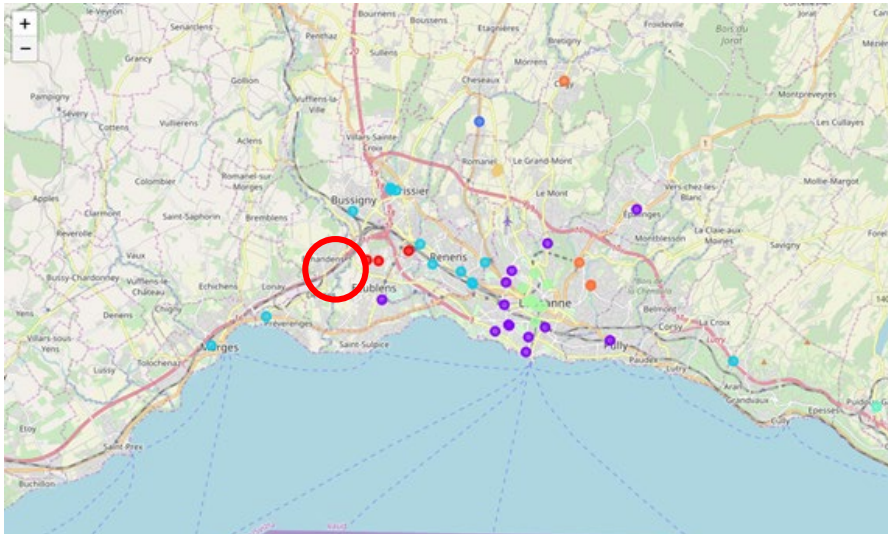
Bern -



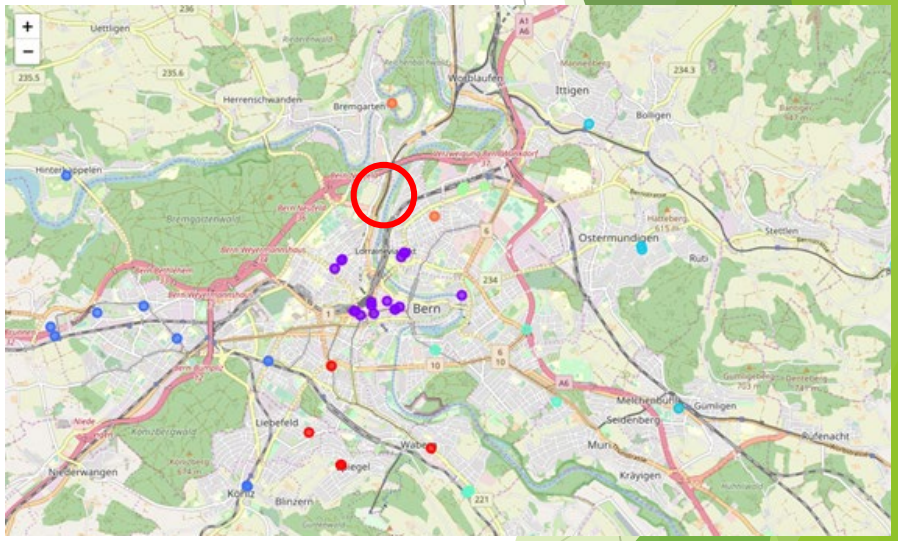
colour-coded to each clister; Red = 0, Purple = 1, Navy Blue = 2, Faded Blue = 3, Turquoise = 4, Lime Green = 5, Mustard Yellow = 6, Orange = 7

Best area to live for me

- Lausanne -



Bern -



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

Conclusion

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.
2. Similar venues i.e. restaurants can be merged together so more unique venues are shown in the top 10 list.