Applied Data Science Capstone Project: Total population and the most popular venues in Zurich city, Switzerland

Authors note : The analysis and the graphics were generated on December 20, 2019. By this day, the situation may have slightly changed and the code would not necessarily generate the exact same graphics and clusters as in the present report.

Introduction / Business problem

In this project we will analyse the most popular venues in the neighborhoods of Zurich city, Switzerland. The final goal is to generate a map and a summary table showing the total population and the most abundant venues in each neighborhood. The purpose of this project is to explore the current situation in the city and to observe the general trends, for example: Which are globally the most popular venues in the neighborhoods? Are the most popular venues the same in each neighborhood? Are there differences between more and less populated neighborhoods?

This project could potentially be interesting for different stakeholders trying to find an optimal location for their business, such as a restaurant, shop, tea room, coffee shop, pharmacy, etc. The stakeholders would probably look for neighborhoods with a high population and which are not already crowded with the same type of venue as the one the stakeholder wants to implement. This information will be provided through this project.

Data

Based on the definition of the problem, we will need the following information :
- the names and exact locations of the neighborhoods in Zurich city
- total population of each neighborhood in Zurich city
- the type (and eventually the number) of the most popular venues in each neighborhood in Zurich city

Following data sources will be used to extract/generate the required information :
- https://www.geonames.org for the information about the latitude and longitude of each neighborhood
- https://wikipedia.org for the information about the boroughs and neighborhoods of Zurich city and their population and population density
This information will be inserted into a .csv file containing the following columns : Borough, Neighborhood, Latitude, Longitude, Population and Population density.
- https://opendata.swiss for the .json file with the neighborhoods of Zurich city (file uploaded by Zurich Stadt)
- Foursquare API for the type, location (and number) of the most popular venues in each neighborhood in Zurich city

Methodology

Dataframe
The data about boroughs, neighborhoods, latitude, longitude, total population and population density/km² was prepared and compiled together in a .csv file. This .csv file was loaded into the Jupyter

Notebook with the python *pandas* library. The head of the resulting dataframe of 34 rows and 6 columns can be seen on Figure 1.

	Borough	Neighborhood	Latitude	Longitude	Population (2018)	Population density / km²
0	Kreis 1 – Altstadt	Rathaus	47.37161	8.54501	3267	3194
1	Kreis 1 – Altstadt	Hochschulen	47.36601	8.54594	664	3194
2	Kreis 1 – Altstadt	Lindenhof	47.37188	8.54036	990	3194
3	Kreis 1 – Altstadt	City	47.37269	8.53576	829	3194
4	Kreis 2	Wollishofen	47.34010	8.53134	18923	3151

Figure 1 : Dataframe used for the analysis.

Choropleth map

The python *Nominatim* and *folium* libraries were used to generate a choropleth map of Zurich city with its 34 neighborhoods (« statistische Quartiers ») with an additional superimposed layer showing the population density/km² or the total population in every neighborhood. As the geo_data for the choropleth map, a previously downloaded .json file of Zurich city (see section **Data**) was used. Therefore, the python *json* library was also imported to handle .json files.

Neighborhoods exploration

Foursquare API was used to explore the most popular venues in every neighborhood in Zurich city. For each neighborhood, a limit of 100 venues and a radius of 500 meters were set from their given latitude and longitude. For this part, the python *requests* library was used. A new dataframe containing every neighborhood with its 10 most common venues was created.

Clustering neighborhoods

To cluster neighborhoods based on their most common venues, the python *metrics*, *cdist*, *matlplotlib.pyplot*, *KMeans* libraries were used. The « elbow method » was used to determine the optimal value of *k* for the *k*-means model (Figure 2). The *k*-means test was then performed using the calculated optimal value of *k*. A new dataframe was created incorporating the original dataframe, as well as the cluster labels and the 10 most common venues in every neighborhood. The clusters were added to the previously created choropleth map using the python *cm* and *colors* libraries.

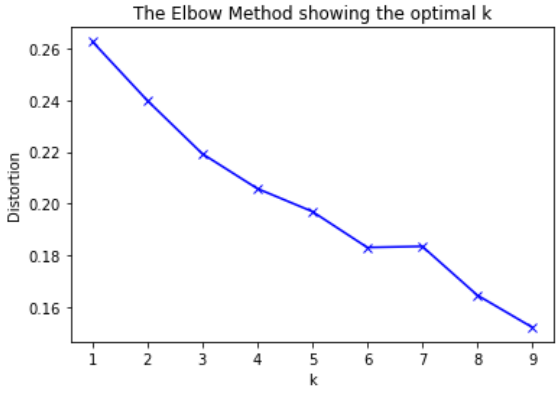

Figure 2 : Plot showing the result of the elbow method to find the optimal k for the k-means test.

Results

The choropleth maps showing the population density/km² in each neighborhood or the total population in each neighborhood, with the belonging of each neighborhood to one of 6 clusters based on the most popular venues in these neighborhoods can be seen on Figure 3 and Figure 4.

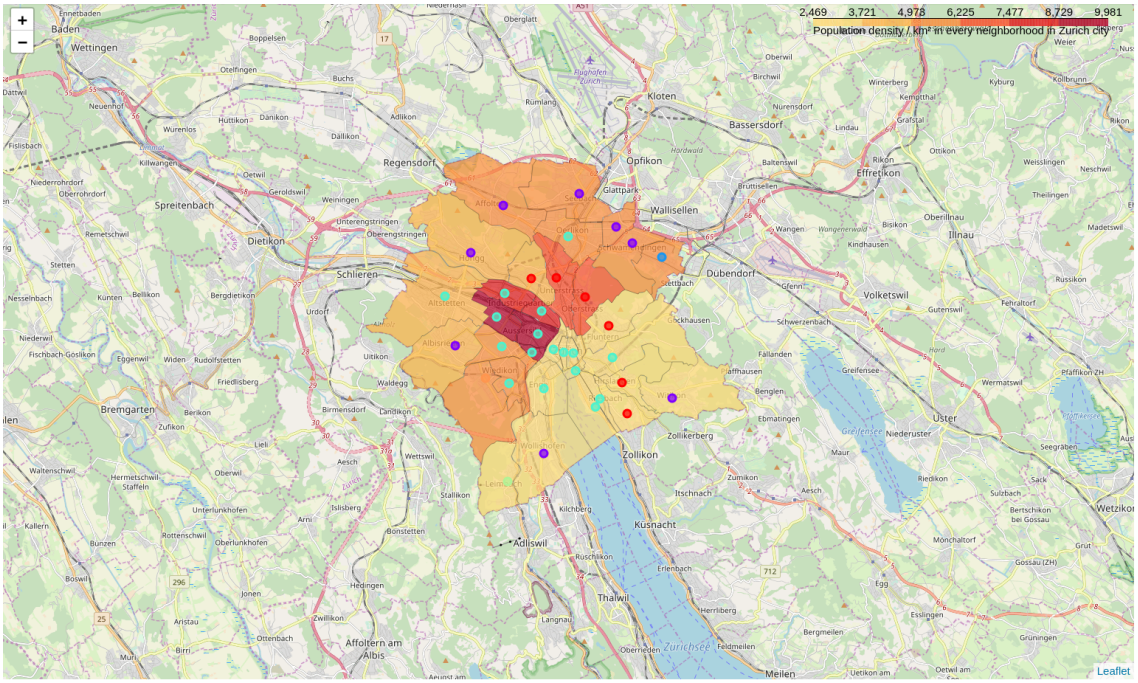

Figure 3 : Map showing the population density/km² in each neighborhood and the belonging of each neighborhood to one of 6 clusters based on the most popular venues in these neighborhoods.

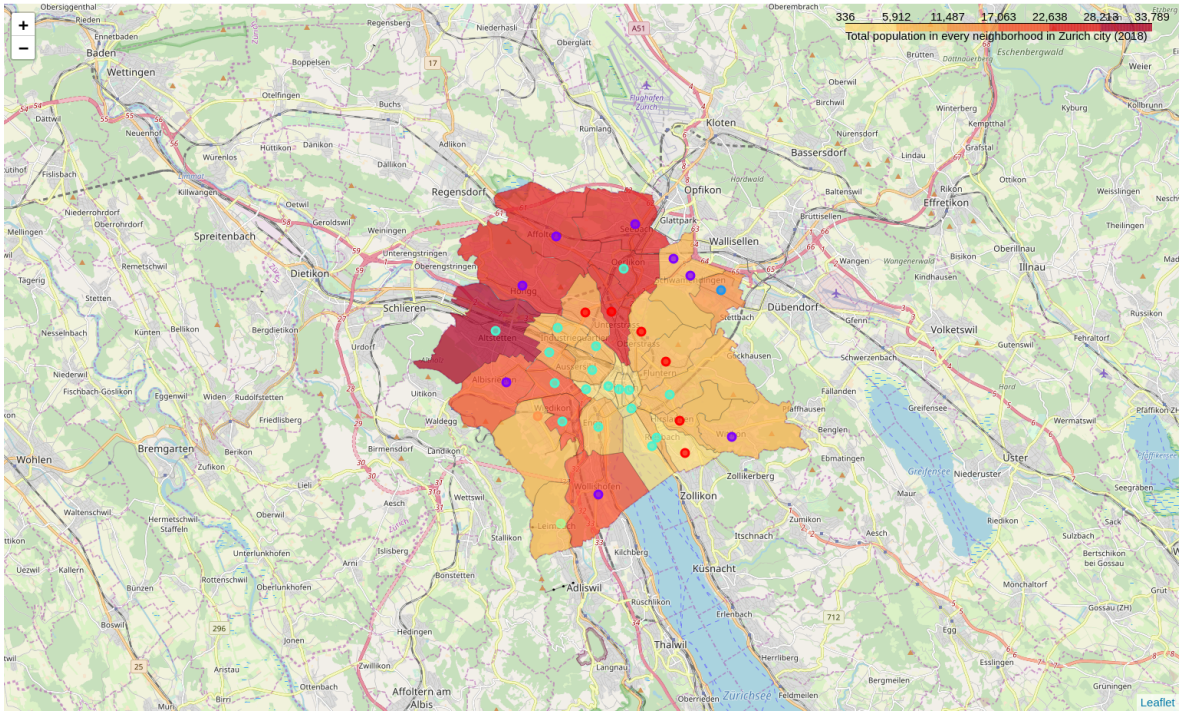

Figure 4 : Map showing the total population (year 2018) in each neighborhood and the belonging of each neighborhood to one of 6 clusters based on the most popular venues in these neighborhoods.

The Cluster 1 (red dots on the Figure 3 and Figure 4) contains 6 neighborhoods which are quite central, but are not located directly around the lake and are not the most heavily populated. The summary of the neighborhoods and the 10 most common venues of this cluster are shown on Figure 5. Among the most popular venues of this cluster, we observe tram stations, bakeries, restaurants (Asian, Italian, Oriental), various stores, hotel and parks.

	Neighborhood	Population density / km²	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
15	Unterstrass	6730	0	Tram Station	Pizza Place	Hotel	Bakery	Café	Doner Restaurant	Kids Store	Sporting Goods Shop	Flower Shop	Convenience Store
16	Oberstrass	6730	0	Tram Station	Asian Restaurant	Bakery	Supermarket	Theater	Middle Eastern Restaurant	Athletics & Sports	Medical Supply Store	Bus Station	Park
17	Fluntern	2543	0	Tram Station	Bakery	Gastropub	Pizza Place	Hotel	Grocery Store	Bus Station	Supermarket	Diner	Factory
19	Hirslanden	2543	0	Tram Station	River	Italian Restaurant	Hotel	Park	Swiss Restaurant	French Restaurant	Mediterranean Restaurant	Plaza	Discount Store
23	Weinegg	3490	0	Bakery	Medical Center	Tram Station	Modern European Restaurant	Café	Restaurant	Deli / Bodega	Cultural Center	Falafel Restaurant	Factory
27	Wipkingen	4438	0	Grocery Store	Bakery	Pizza Place	Bus Station	Business Service	Café	Supermarket	Swiss Restaurant	Bar	Design Studio

Figure 5 : Neighborhoods and the 10 most common venues of Cluster 1.

The Cluster 2 (purple dots on the Figure 3 and Figure 4) contains 8 neighborhoods which are mostly peripheral regions and having a high total population (but not necessarily the highest population density). The summary of the neighborhoods and the 10 most common venues of this cluster are shown on Figure 6. Among the most popular venues of this cluster, we observe bus stations, restaurants (Asian, Italian, Oriental), various stores, but also pools, bars, and yoga studios.

	Neighborhood	Population density / km²	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
4	Wollishofen	3151	1	Bus Station	Tram Station	Plaza	Bar	Supermarket	Thai Restaurant	Restaurant	Gas Station	Pool	Cheese Shop
20	Witikon	2543	1	Bus Station	Indian Restaurant	Optical Shop	Department Store	Tram Station	Church	Supermarket	Bakery	Yoga Studio	Doner Restaurant
24	Albisrieden	4620	1	Bus Station	Supermarket	Swiss Restaurant	Pizza Place	Café	Grocery Store	Trattoria/Osteria	Bar	Factory	Event Space
26	Höngg	4438	1	Grocery Store	Tram Station	Italian Restaurant	Steakhouse	Food & Drink Shop	Mexican Restaurant	Gas Station	Plaza	Supermarket	Fast Food Restaurant
28	Affoltern	5614	1	Bus Station	Supermarket	Diner	Light Rail Station	Department Store	Athletics & Sports	Italian Restaurant	Train Station	Restaurant	Doner Restaurant
30	Seebach	5614	1	Hookah Bar	Pizza Place	Korean Restaurant	Bakery	Supermarket	Eastern European Restaurant	Pool	Tram Station	Laser Tag	Bus Station
31	Saatlen	5441	1	Gastropub	Falafel Restaurant	Construction & Landscaping	Pedestrian Plaza	Arts & Crafts Store	Bus Station	Kebab Restaurant	Yoga Studio	Factory	Event Space
32	Schwamendingen-Mitte	5441	1	Bus Station	Thai Restaurant	Italian Restaurant	Fast Food Restaurant	Supermarket	Light Rail Station	Restaurant	Asian Restaurant	Swiss Restaurant	Tram Station

Figure 6 : Neighborhoods and the 10 most common venues of Cluster 2.

The Cluster 3 (dark blue dot on the Figure 3 and Figure 4) contains only 1 peripheral and moderately populated neighborhood located close to Zurich airport. The summary of the neighborhood and the 10 most common venues of this cluster are shown on Figure 7. The most popular venues are similar to those of the cluster 2, but with no bars. We can also observe new types of venues, such as a steakhouse, a farmers market, and a factory.

	Neighborhood	Population density / km²	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
33	Hirzenbach	5441	2	Tram Station	Steakhouse	Yoga Studio	Discount Store	Farmers Market	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant	Electronics Store

Figure 7 : Neighborhood and the 10 most common venues of Cluster 3.

The Cluster 4 (light blue dot on the Figure 3 and Figure 4) contains 28 neighborhoods which are mostly located in the center of the city or around the lake. This cluster contains some of the most densely

populated neighborhoods of the city. The summary of the neighborhoods and the 10 most common venues of this cluster are shown on Figure 8. Among the most popular venues of this cluster, we observe bars, cafés, restaurants, supermarkets, hotels (Swiss, Italian, Asian), cocktail bars and other (gyms, museums, theaters, music venues, coffee shops, etc.).

	Neighborhood	Population density / km²	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	Rathaus	3194	3	Swiss Restaurant	Bar	Café	Cocktail Bar	Restaurant	Spanish Restaurant	Plaza	French Restaurant	Italian Restaurant	Hotel
1	Hochschulen	3194	3	Swiss Restaurant	Bar	Hotel	Coffee Shop	Café	Plaza	French Restaurant	Bakery	Lounge	Italian Restaurant
2	Lindenhof	3194	3	Café	Swiss Restaurant	French Restaurant	Boutique	Bar	Restaurant	Hotel	Lounge	Cocktail Bar	Italian Restaurant
3	City	3194	3	Bar	Hotel	Restaurant	Cocktail Bar	Italian Restaurant	Swiss Restaurant	Café	Vegetarian / Vegan Restaurant	Department Store	Plaza
6	Enge	3151	3	Bar	Italian Restaurant	Park	Restaurant	History Museum	Tram Station	Supermarket	Swiss Restaurant	Burger Joint	Café
7	Alt-Wiedikon	5846	3	Italian Restaurant	Restaurant	Supermarket	Diner	Lounge	Shopping Mall	Sporting Goods Shop	Beer Garden	Burrito Place	Thai Restaurant
9	Sihlfeld	5846	3	Café	Italian Restaurant	Bar	Swiss Restaurant	Thai Restaurant	Ethiopian Restaurant	Plaza	Supermarket	Pizza Place	Coffee Shop
10	Werd	9907	3	Bar	Italian Restaurant	Restaurant	Indian Restaurant	Pizza Place	Japanese Restaurant	Music Venue	Café	Sushi Restaurant	Coffee Shop
11	Langstrasse	9907	3	Bar	Italian Restaurant	Restaurant	Swiss Restaurant	Café	Japanese Restaurant	Mediterranean Restaurant	Bakery	Vegetarian / Vegan Restaurant	Chinese Restaurant
12	Hard	9907	3	Plaza	Hotel	Mediterranean Restaurant	French Restaurant	Swiss Restaurant	Kebab Restaurant	Soccer Stadium	Gym / Fitness Center	Hotel Bar	Falafel Restaurant
13	Gewerbeschule	7829	3	Bar	Swiss Restaurant	Café	Asian Restaurant	Bakery	Vietnamese Restaurant	Food Truck	Italian Restaurant	Thai Restaurant	Yoga Studio
14	Escher Wyss	7829	3	Café	Hotel	Restaurant	Gym / Fitness Center	Dance Studio	Bar	Italian Restaurant	Art Museum	Tram Station	Hotel Bar
18	Hottingen	2543	3	Swiss Restaurant	Hotel	Gym Pool	Coffee Shop	Tram Station	Italian Restaurant	Modern European Restaurant	Supermarket	Bakery	Doner Restaurant
21	Seefeld	3490	3	Swiss Restaurant	Supermarket	Restaurant	Movie Theater	Art Museum	Park	Bakery	Wine Bar	Sushi Restaurant	Café
22	Mühlebach	3490	3	Supermarket	Swiss Restaurant	Restaurant	Italian Restaurant	Bakery	Tram Station	Café	Bar	Sushi Restaurant	French Restaurant
25	Altstetten	4620	3	Supermarket	Tram Station	Hotel	Mediterranean Restaurant	Italian Restaurant	Gym	Gas Station	Asian Restaurant	Doner Restaurant	Restaurant
29	Oerlikon	5614	3	Supermarket	Italian Restaurant	Hotel	Restaurant	Kebab Restaurant	Swiss Restaurant	Beer Bar	Pub	Steakhouse	Indonesian Restaurant

Figure 8 : Neighborhoods and the 10 most common venues of Cluster 4.

The Cluster 5 (light green dot on the Figure 3 and Figure 4) contains only 1 peripheral and feebly populated located in the South-West of the city. The summary of the neighborhood and the 10 most common venues of this cluster are shown on Figure 9. The most popular venues are similar to those of the cluster 3.

	Neighborhood	Population density / km²	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
5	Leimbach	3151	4	Bus Station	Light Rail Station	Grocery Store	Trail	Gas Station	Yoga Studio	Doner Restaurant	Falafel Restaurant	Factory	Event Space

Figure 9 : Neighborhood and the 10 most common venues of Cluster 5.

The Cluster 6 (orange dot on the Figure 3 and Figure 4) contains only 1 peripheral and moderately populated neighborhood located in the South-West of the city. The summary of the neighborhood and the 10 most common venues of this cluster are shown on Figure 10. The most popular venues are similar to those of the cluster 3.

	Neighborhood	Population density / km²	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
8	Friesenberg	5846	5	Bus Station	Supermarket	Garden	Yoga Studio	Fast Food Restaurant	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant	Electronics Store

Figure 10 : Neighborhood and the 10 most common venues of Cluster 6.

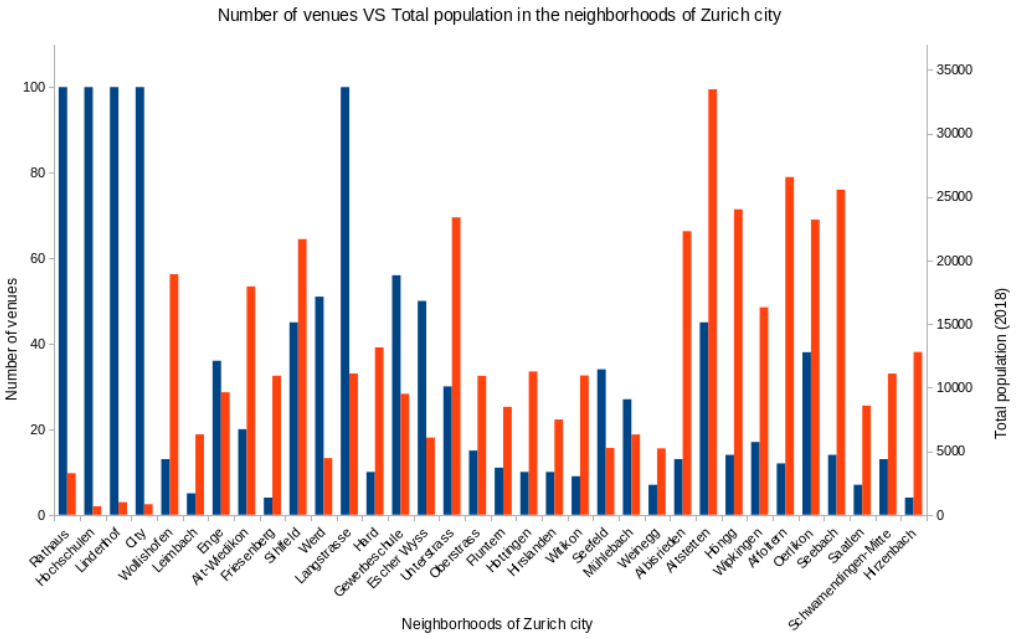

Figure 11 : Number of venues (in blue) VS Total population (in red) in the neighborhoods of Zurich city.

Discussion

First observation we can make, is that the neighborhoods with the biggest number of venues (100 and more) are the ones with the lowest number of total population living in these neighborhoods (below 12000). On the contrary, neighborhoods with the largest population usually have less than 50 venues in their neighborhoods, sometimes less than 20 (Figure 11). These differences show which neighborhoods are mostly residential (with houses and apartments) and which are the active neighborhoods, concentrating most of the commercial activities of the city. These differences can also be observed in the types of the most popular venues in these neighborhoods.

Clusters 2, 3, 5 and 6, can be considered as residential neighborhoods. Their main venues are bus/tram stations, restaurants, supermarkets/stores, yoga studios, event spaces and these neighborhoods have little to none nightlife activities. These neighborhoods are peripherical, but are well connected to the rest of the city. People probably chose these neighborhoods for their quiet and a their higer number of available and more abordable apartments.
Cluster 2 globally encompasses more populated and more developed neighborhoods than clusters 3, 5 and 6. The clusters 3, 5 and 6 (each containing only one neighborhood) have the lowest numbers of venues among all the neighborhoods and their population is also not very high. This could indicate that the neighborhoods from cluster 2 are older, so they had more time to develop than those of clusters 3, 5 and 6. In that case, the neighborhoods from the clusters 3, 5 and 6 have a great potential for developpment, which the city should exlpoit if it wanted to attract more people to these neighborhoods. Especially the neighborhood Leimbach from cluster 5 could be more developed, as the population density is lower than in the clusters 3 and 6. The venues suitable for these neighborhoods should be aimed for daylife activities and be family friendly.
Let’s name cluster 2 the Old residential neighborhood and clusters 3, 5 and 6 the New residential neighborhoods.

Cluster 4 on the contrary has a lot of commercial activities and a developed nightlife (restaurants, bars, pubs). Let's name cluster 4 the City Center. The neighborhoods of this cluster are the most suitable for shopping and going-out in general. The existing venues are already numerous and various, so the competition would be also very high for a new built venue, especially for Swiss and Italian restaurants. Asian/Oriental/African restaurants would maybe stand more chances, as well as those specialised in vegetarian cuisine. From this cluster, particular attention should be adressed to Altstetten and Oerlikon neighborhoods, as they have the highest total population and still a moderate number of venues.

Cluster 1 can be considered as a buffer zone between the two previous categories. The total population is higher than in the center, but lower than in the residential neighborhoods ; however the number of venues is globally higher than in the purely residential neighborhoods. There are some hotels, various restaurants and stores. The nightlife activities are not very present in this cluster. These neighborhoods seem to be lively and to present a lot of activities, but are more destined / suitable for people who appreciate more calmer areas than the city center. These neighborhoods seem also to be popular for wealthy people, which is shown by the presence of gastropubs and bodegas in the neighborhoods. Let's name cluster 1 the Wealthy Residential-Center. From this cluster, the neighborhoods which seem to have the greatest potential of growing are Unterstrass, Oberstrass and Wipkingen, which have a relatively high population, but are not overcrowded with venues.

Conclusion

The purpose of this project was to analyse the most popular venues in the neighborhoods of Zurich city, Switzerland and thus to explore the current situation in the city and to observe the general trends. By using Foursquare API data we could observe differences between the neighborhoods and make clusters representing whether we were dealing with residential or commercial neighborhoods. This analysis gave us information on the development potential of the neighborhoods, which can be very useful to different stakeholders. This project demonstrated again the importance and the interest of data and how this data can bring us answers, and even sometimes to questions that we didn't ask.