FINDING THE BEST LOCATION FOR A FUTURE FLOWER SHOP IN ZURICH

1. BACKGROUND AND PROBLEM

The problem we want to solve is to find the best possible location(s) to start a 'mom & pop' retail shop in the city of Zurich, Switzerland. More specifically, we are looking for the best location for a flower shop, as this will help a friend who is about to launch her flower business.

Location is a key success factor for both pillars of such a business, which are:
- Selling flowers to customers visiting the shop;
- Delivering flowers, ordered either by internet or over phone.

Finding the best location is thus very important for the future business owner. Also, it is important to be able to explain and demonstrate why a specific location is suitable, to convince potential investors and banks of the viability of the future business. So this data visualization exercise is interesting for the business owner, her investors / partners, as well as the bank(s) expected to extend a credit line to the future business.

More specifically, the ideal location should have the following characteristics:

a) In a densely populated district of the city (we assume here that purchasing power and propensity to buy flowers is similar in each district, so that only the population density matters).
b) Near high-traffic public transportation hubs. Zurich mostly relies on a tram system, so that means the shop should be located near the tram stops with the highest number of daily passengers.
c) In a quite central area of the city, so that most of the city is a quickly reachable for deliveries. A location at the periphery would severely restrict the delivery business' potential.
d) At least 1'000 meters away from the next flower shop. Being closer from established competition would make it difficult to establish a new business. On the other hand, finding a 'white space' away from any competitor may help a lot.
e) Ideally - but only if all other criteria are met - the shop should be in a pedestrian area, as (i) passerby are more relaxed and more inclined to shop flowers in pedestrian areas and (ii) this brings nicer working conditions for the shop's employees.

2. DATA AND HOW IT WILL BE USED TO SOLVE THE PROBLEM

To solve this problem, we will mostly use data from three sources:
- The folium library
- The Open Data catalogue of Zurich City (https://data.stadt-zuerich.ch, in German)
- Foursquare.com

This location problem should be easier to solve visually, rather than using formulas with quantitative outputs. In that context, the folium will help a lot, with the following features:

a) Background map of the city of Zurich
b) Various overlays, like the map of the districts, markers for the main tram stations, etc…

The Open Data catalogue of Zurich City will provide us with the following data:

a) Population of the city, with a split per city district;
b) Limits of the city districts (geoJSON file). This will allow us to display the districts on the map, and create a choropleth based on the population density;

c) Limits of the pedestrian areas;
d) Location and daily number traffic of the tram stations. We will focus on the 20 most busy stations and mark them on the map.

The Foursquare API will provide us with:

a) The list of existing flower shops in the city of Zurich;
b) Their location, so that we can display them on the map with a marker;
c) If needed, additional information on select competitors, like rating and tips…

3. METHODOLOGY

3.1 Population per district

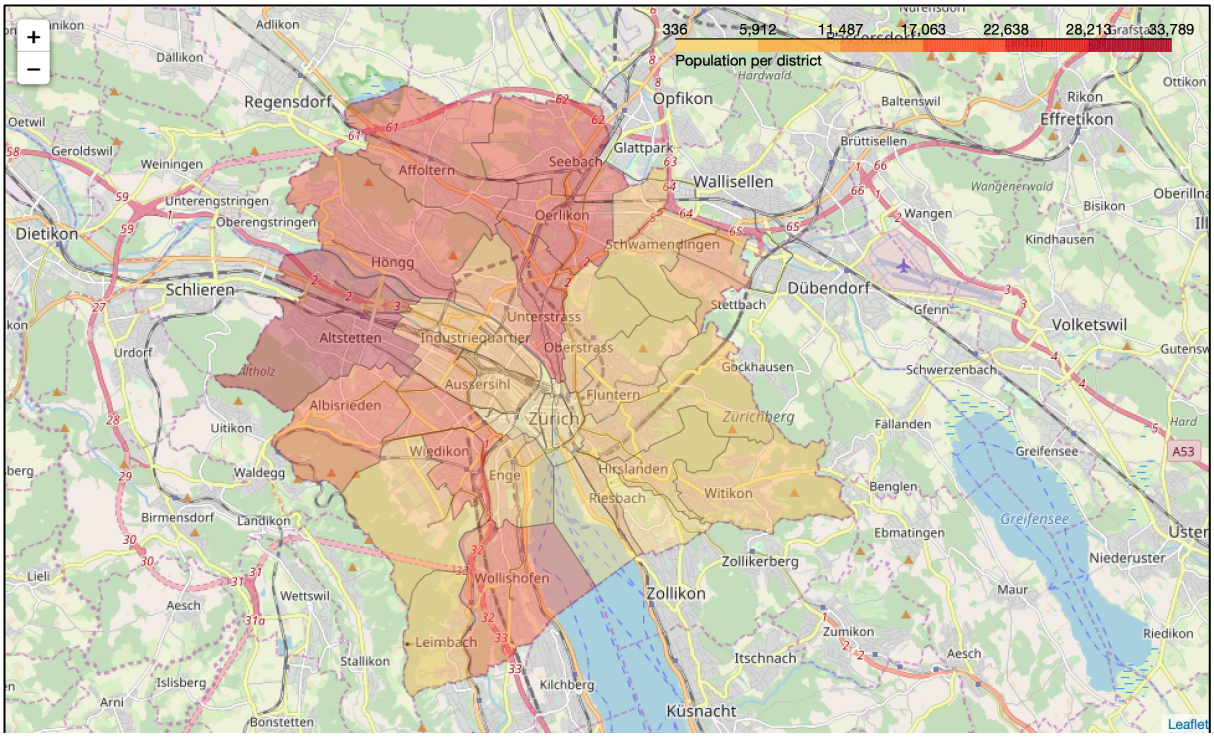
We first start by mapping the population per district, which is a key driver for the size of the potential client franchise, and thus the achievable business turnover.

To this end, we load a population dataset provided by the city of Zurich under their open data initiative.

	StichtagDatJahr	QuarSort	QuarCd	QuarLang	NationHistSort	NationHistLang	NationSort	NationLang	RegionSort	RegionLang	KontinentSort	KontinentLang	AnzBestWir
0	1993	11	11	Rathaus	8100	Schweiz	8100	Schweiz	13	Zentraleuropa	1	Europa	2487
1	1993	11	11	Rathaus	8205	Bulgarien	8205	Bulgarien	17	Südosteuropa	1	Europa	1
2	1993	11	11	Rathaus	8206	Dänemark	8206	Dänemark	11	Nordeuropa	1	Europa	2
3	1993	11	11	Rathaus	8207	Deutschland	8207	Deutschland	13	Zentraleuropa	1	Europa	96
4	1993	11	11	Rathaus	8211	Finnland	8211	Finnland	11	Nordeuropa	1	Europa	2
5	1993	11	11	Rathaus	8212	Frankreich	8212	Frankreich	14	Westeuropa	1	Europa	22
6	1993	11	11	Rathaus	8214	Griechenland	8214	Griechenland	16	Südeuropa	1	Europa	1
7	1993	11	11	Rathaus	8215	Grossbritannien	8215	Grossbritannien	14	Westeuropa	1	Europa	18
8	1993	11	11	Rathaus	8216	Irland	8216	Irland	14	Westeuropa	1	Europa	2
9	1993	11	11	Rathaus	8218	Italien	8218	Italien	16	Südeuropa	1	Europa	69

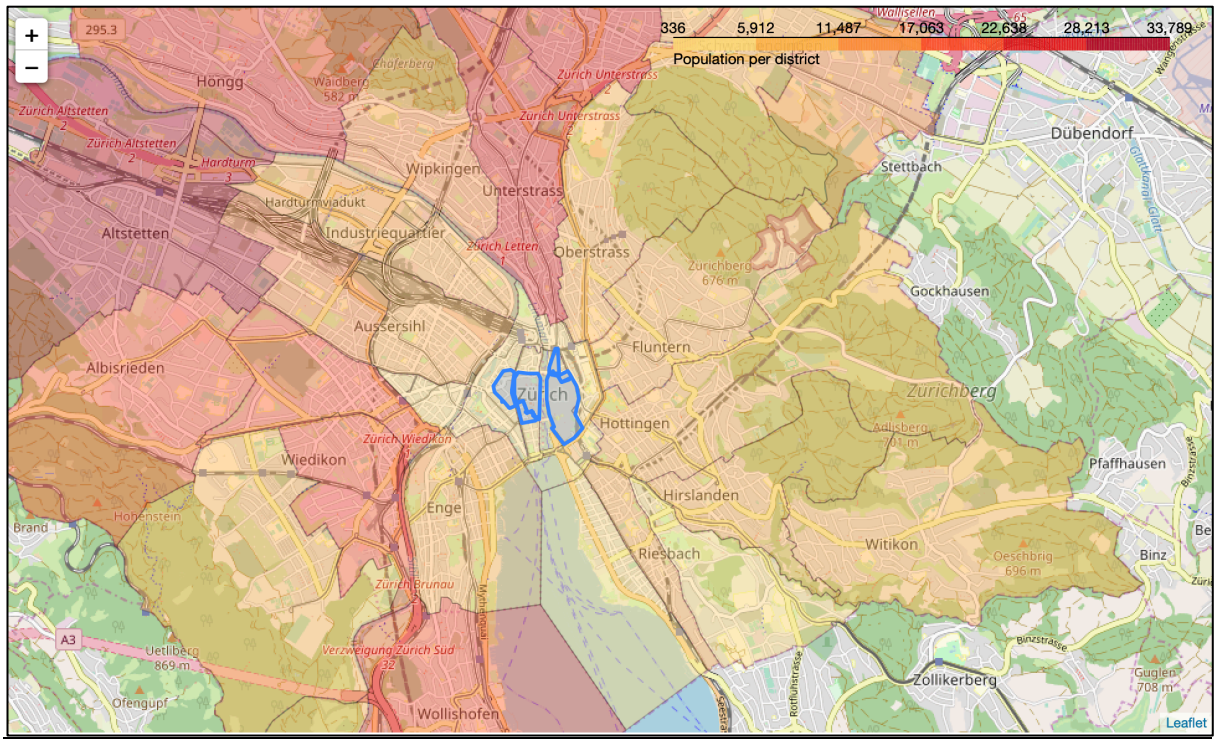
The dataset requires significant preprocessing, as it was not meant for our specific purpose, but rather as a census of populations from different national origins, with an indication of their residence. The data is also very granular (population is reported for very small areas of the city), extensive (all years since 1993, detail of all nationalities…) and sometimes incomplete (a number of NaN values). Therefore, significant data cleaning and aggregation is required.

This dataset allows us to derive a choropleth (below) of the city districts, where the color scale indicates the density of population (yellow = lower, red = higher). This makes clear that the densely populated districts are mostly close to the city periphery, while the historic center is less residential. This is not surprising, as the historic center is mostly dedicated to business offices and shops. People commute into the center for work, and back towards the periphery at the end of the day. Also, buildings in the center are lower rise, therefore hold fewer inhabitants.



3.2 Mapping of the pedestrian areas

The pedestrian areas are available in a GeoJSON provided by the city of Zurich (fussgaengerzone.json). Adding these to the map, it becomes evident that pedestrian areas are quite small, and concentrated in the heart of the historic center, where the density of population is lower. This makes it less likely that the location of the future shop will meet the last – subsidiary – criterion of being with a pedestrian area.



3.3 Mapping of the public transportation stations, ranked per daily traffic

Within the city of Zurich, public transportation carries many more passengers than individual cars. The city relies mostly on a very dense network of tramways, some buses are available as well.

We decide to map the 20 most busy public transportation stations. This seems a large enough number, because:
- The 20th station (Escher-Wiss Place) is already quite small (76'000 daily passengers), at roughly a third of the traffic of the top station (Bellevue, 221'000 daily passengers);
- Mapping more than 20 stations would make it visually difficult to make sense of the map.

To this end, we need to load, preprocess, and concatenate (left join) data frames built from different csv files available on the website of the city.

The first step is to obtain the list and details of the stations:

	Haltestellen_Id	Haltestellennummer	Haltestellenkurzname	Haltestellenlangname
0	272	6594	AHOR	Adliswil, Ahornweg
1	270	148	ABAD01	Adliswil, Badstrasse
2	683	28	BADL01	Adliswil, Bahnhof
3	273	6593	EWEG	Adliswil, Eichenweg
4	682	6961	AGRU	Adliswil, Grundstrasse
5	267	6437	GRUT	Adliswil, Grüt

NOTE:
- 'Haltestellen' means 'station', or 'stop'
- 'Haltestellennummer' is the station's number
- 'Haltestellenkurzname' is the station's short name
- 'Haltestellenlangname' is the station's long name

The second step is to obtain the daily passenger statistics per staion, from another csv file:

	Tagtyp_Id	Linien_Id	Linienname	Plan_Fahrt_Id	Richtung	Sequenz	Haltestellen_Id	Nach_Hst_Id	FZ_AB	Anzahl_Messungen	Einsteiger	Aussteiger	Besetzung
0	15	60	704	91534	2	1	576	575.0	05:38:00	11	4.63	0.00	4.63
1	15	60	704	91534	2	2	575	574.0	05:39:54	11	0.00	0.00	4.63
2	15	60	704	91534	2	3	574	451.0	05:41:12	11	0.09	1.63	3.09
3	15	60	704	91534	2	4	451	511.0	05:42:48	11	1.63	1.27	3.45
4	15	60	704	91534	2	5	511	512.0	05:44:00	11	0.09	0.00	3.54
5	15	60	704	91534	2	6	512	577.0	05:45:12	11	0.00	0.00	3.54

NOTE:
- 'Einsteiger' are passengers getting into the tram
- 'Austeiger' are passengers getting out of the tram
- The other columns are not required, and therefore dropped from the data frame

The third step is to compute the total number of daily passengers for each station, by adding up 'Einsteiger' and 'Aussteiger':

	Haltestellen_Id	Einsteiger	Aussteiger	Passengers
0	4	16482.38	15044.61	31526.99
1	5	139.51	151.20	290.71
2	6	507.74	472.18	979.92
3	7	964.77	861.10	1825.87
4	8	715.36	796.65	1512.01

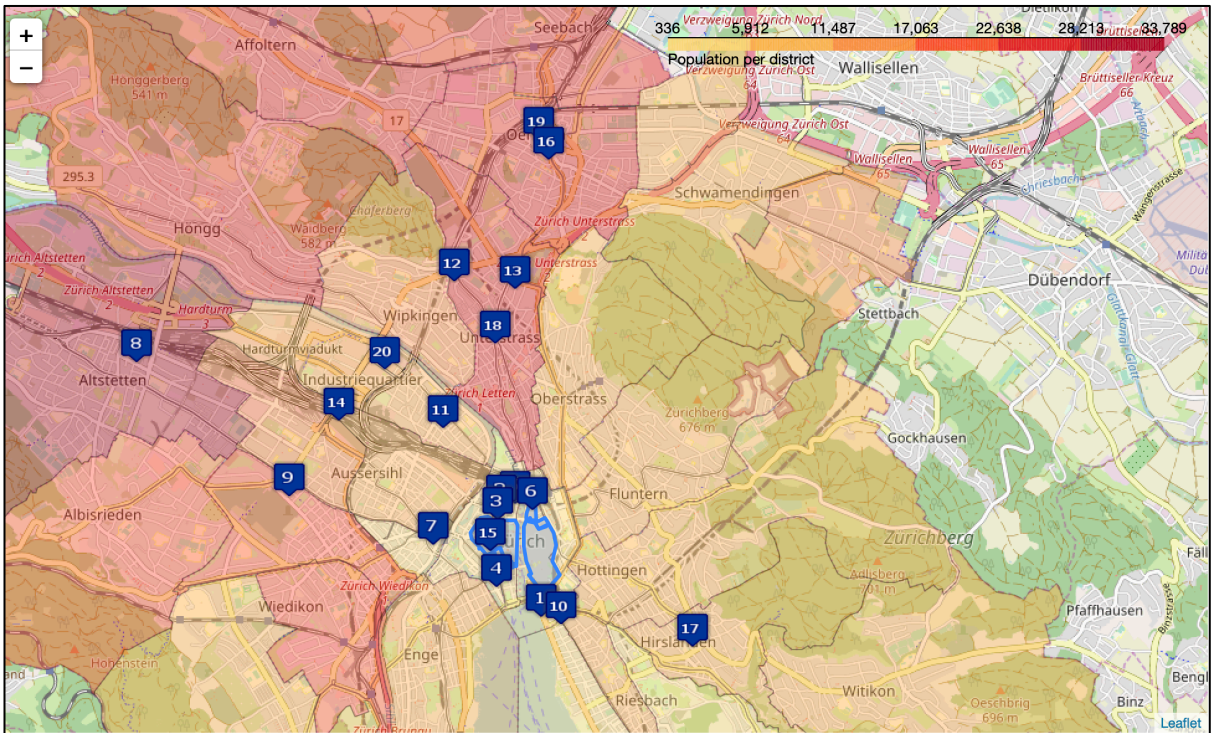
The fourth step is to combine the previous data frames with a left join, focusing on the most relevant columns:

	Haltestellen_Id	Haltestellennummer	Haltestellenkurzname	Haltestellenlangname	Passengers
0	272	6594	AHOR	Adliswil, Ahornweg	508.0
1	270	148	ABAD01	Adliswil, Badstrasse	590.0
2	683	28	BADL01	Adliswil, Bahnhof	2311.0
3	273	6593	EWEG	Adliswil, Eichenweg	455.0
4	682	6961	AGRU	Adliswil, Grundstrasse	98.0
5	267	6437	GRUT	Adliswil, Grüt	476.0

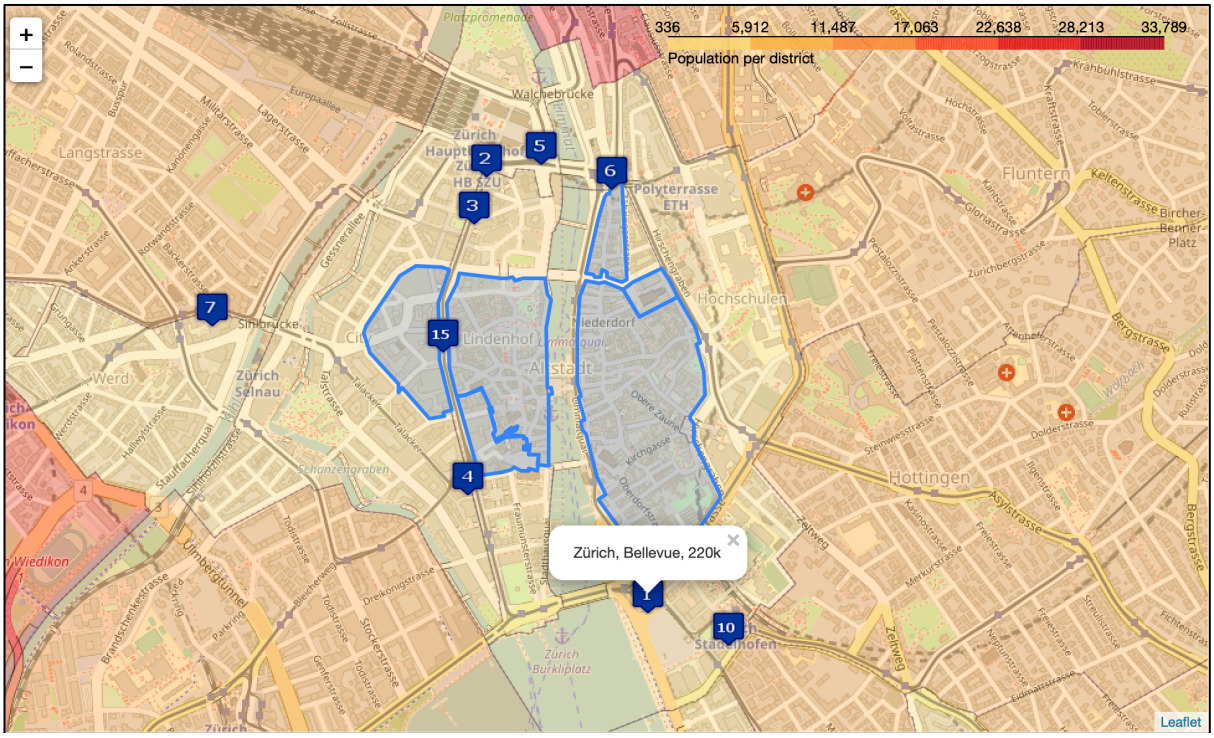
The fifth step is to add the geodata of the stations, from another file, keeping the stations sorted by descending order of traffic:

	Haltestellen_Id	Haltestellennummer	Haltestellenkurzname	Haltestellenlangname	Passengers	stop_id	stop_name	stop_lat	stop_lon
0	22	440	BELL	Zürich, Bellevue	220792.0	gen:23026:440:1:0	Zürich, Bellevue	47.366873	8.545332
1	244	412	BPLA	Zürich, Bahnhofplatz/HB	194430.0	gen:23026:412:1:0	Zürich, Bahnhofplatz/HB	47.377081	8.539771
2	114	416	BSTR	Zürich, Bahnhofstrasse/HB	185631.0	gen:23026:416:1:0	Zürich, Bahnhofstrasse/HB	47.375974	8.539340
3	116	1861	PARA	Zürich, Paradeplatz	177712.0	gen:23026:1861:1:0	Zürich, Paradeplatz	47.369629	8.539125

Finally, we can add the top 20 stations to the map, using numbered custom icons (clickable for name and passenger traffic):



It becomes clear most of the top 20 tram stations are (i) in the city center or (ii) on a Southwest/Northeast axis orthogonal to the train tracks.



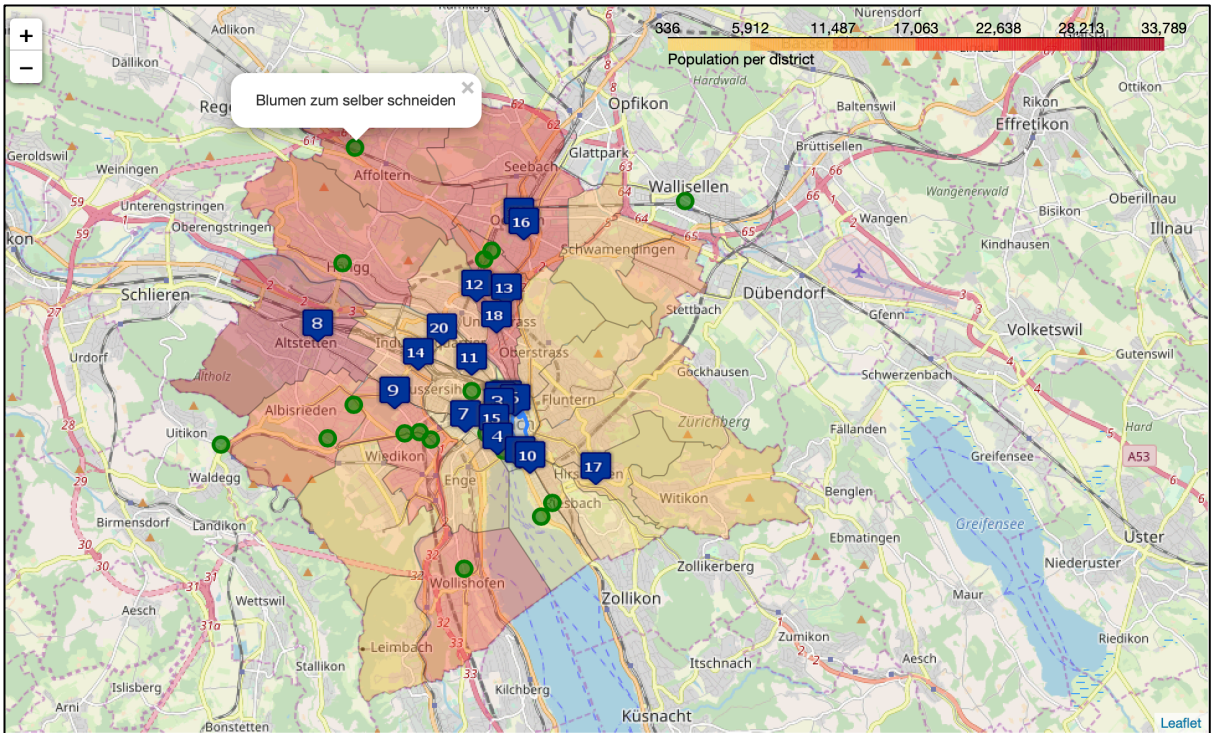
Moreover, 8 of the top 10 tram stations are in the city center. Bellevue is the busiest, with 220k passengers per day.

3.4 Mapping of existing flower shops – the competition

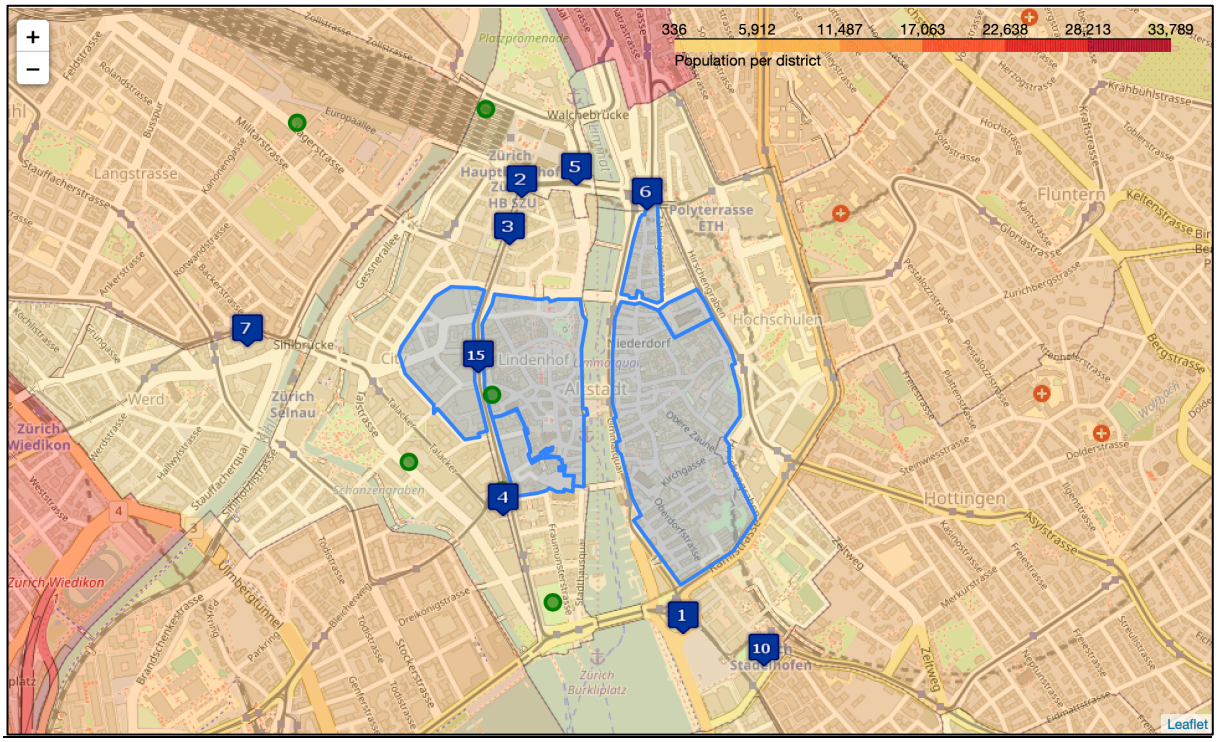
First, we obtain the full list of flowers shops via a Foursquare request.

	name	categories	lat	lng
0	Bürkliplatz Markt	Market	47.367228	8.540819
1	Chinagarten	Garden	47.354768	8.551098
2	Stadtgärtnerei	Garden	47.375887	8.499444
3	Blume 3000	Flower Shop	47.378753	8.538510
4	Veg And The City	Flower Shop	47.378449	8.531995
5	Bubbles	Café	47.372590	8.528307

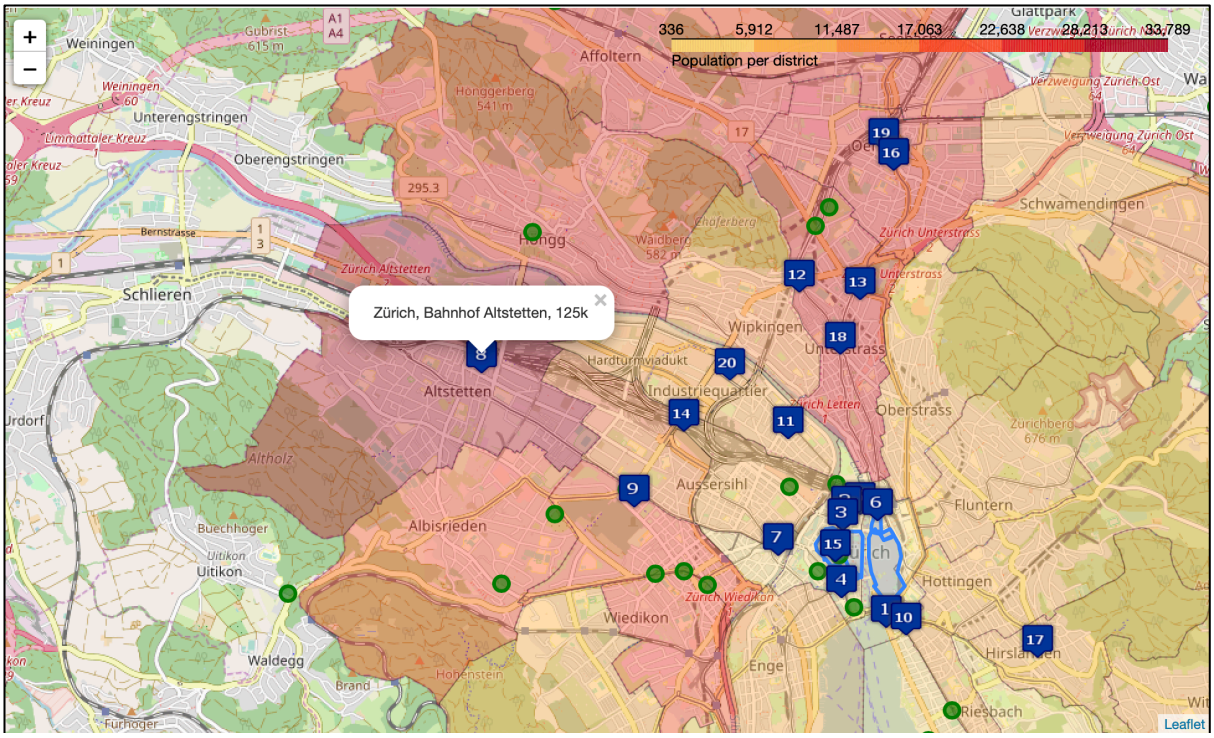
Then, we clean up the list, as a quick internet search reveals that two of the shops listed by Foursquare actually don't sell flowers but are cafés with a flowery atmosphere. We are then ready to add the existing flower shops to the map, marked with a green spot as below. This should help us find a white spot for the future business.



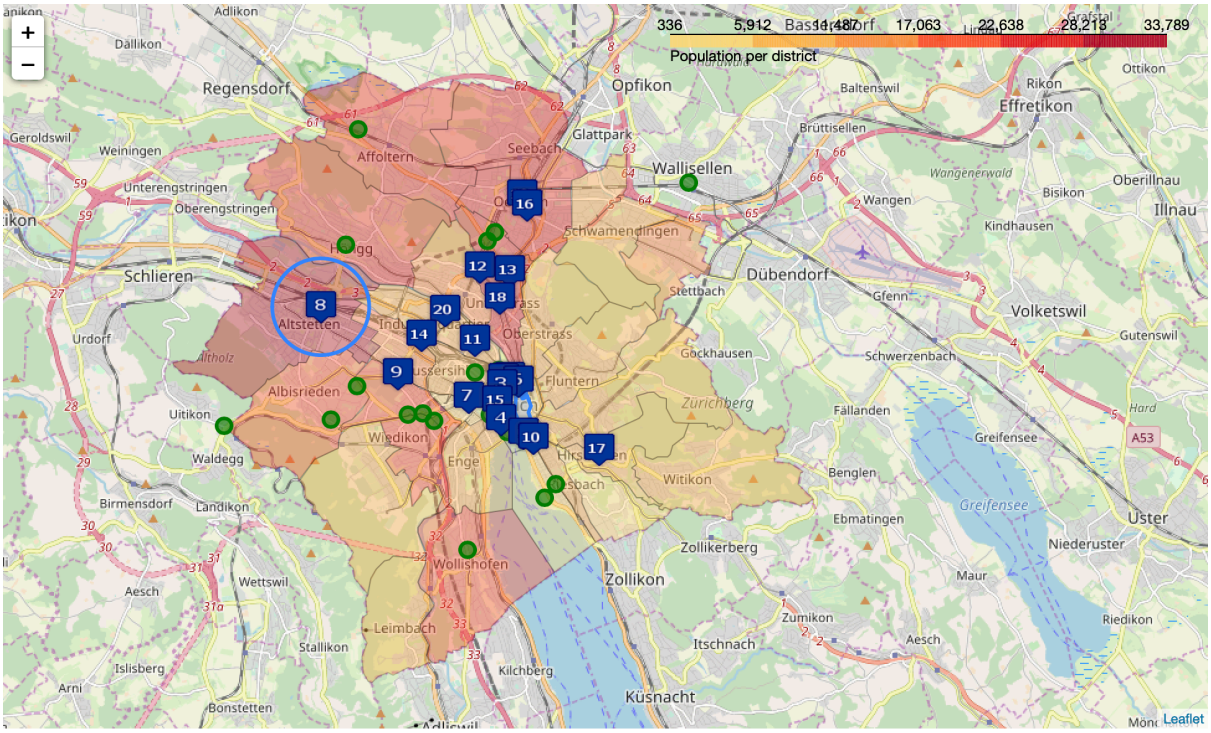
We then zoom-in into the city center (below). Unfortunately, there are already 5 well established flower shops in the city center, close to the top tram stations:



However, Alstetten station seems to offer the most promising white spot:



Adding a circular marker allows us the check whether some competitors are already established within 1'000 meters of the station:



4. RESULTS AND DISCUSSION

Going through the location criteria we defined at the beginning, we were able to zoom in on very promising white spot for our future flower shop.

Altstetten station meets all our criteria, but one:
- It is the 8th most busy station, 125'000 passengers per day
- It is right in the middle of one of the most populated districts of the city
- It is still quite central for deliveries
- It is more than 1'000 meters away from any existing competitor

Alstetten is not located in one of the pedestrian areas. However, this one only a less important, subsidiary, criterion. Also, the pedestrian areas are located in the city historic center, which is already saturated with incumbent flower shops. This area is therefore not a solid choice to establish a new flower business.

It makes sense however to keep an eye on these existing businesses, in case one of them would be up for sale in the near future:
- Busing an established and profitable business is most often a safer option than starting up from scratch;
- Existing businesses nearby the pedestrian area meet the criteria we defined earlier. In particular, 8 of the top 10 tram stations are located there, so that the potential customer franchise is very large.

5. CONCLUSION AND FUTURE DIRECTIONS

Going through the location criteria we defined at the beginning, we were able to zoom in on very promising white spot for the future flower shop. The approach was based on visual analysis, rather than on any optimization algorithm (which would have been quite an artificial approach in this specific case and was anyway not practical given limited data available from the city of Zurich website).

Moving forward, it would be possible to improve the analysis if additional data became available. For example:
- Turnover (and any additional financial data) from existing competitors could help us build the business plan for the new flower shop, using regression methods;
- Data on commercial real estate prices (rental, sale) would provide us with a very important criterion to choose the best possible location;
- Income per capita for the different city districts would also be a very meaningful metrics;
- Statistics on passerby flows for different areas would complement the statistics of public transportation passengers. As of today, these are available only for a few experimental areas, and therefore could not be included in this report.

That being said, it is important to stay pragmatic. For the problem at hand, data analytics and visualization provide a very powerful aid to decision making. However, they complement and cannot replace business sense and local market knowledge.