The Location Value of Pharmacies at Zurich

A Coursera Capstone Project

for

IBM Data Science

by

Pascal Keller
Grundackerweg 47
8965 Berikon, Switzerland

10th of August 2020
Coursera.org

Content

1 Introduction

1.1 Background - high prices, competition and monopoles

Pharmacies are an important contributor to the health system of Switzerland. While Europe has 31 pharmacies per 100'000 habitants, Switzerland has 21. Pharmacies play in a strongly regulated and tariff dependent market for the selling of drugs by prescription und medicine products without prescription, but they also sell beauty and health products like soaps, hand creams, shampoo and perfumes. These products have a high quality and price, even if such products can also be bought in supermarkets with much lower prices. In short, pharmacies have monopoles for some products and competition for other products. Products sold by Swiss pharmacies are associated with highest quality and pricing. All products sold by pharmacies are for the well-being of the human body.

Location is the number one business driver. Two questions arise: What is the location value for existing pharmacies in the centre of Zurich and where is the best location for opening new pharmacies?

Success for pharmacies depends much on their location. The location is the most important factor for the creation of revenue. Here I try to quantify the location value of pharmacies in the City of Zurich. The location value, in addition to other factors, is determined by the following:

The locations of competitors like other pharmacies and stores that also sell perfumes and cosmetics.

The locations of enablers like drugstores, and cosmetic and luxury stores. Foursquare can help analysing location data for Zurich, because using it not only provides coordinates of locations, it also helps to visualize the situation on a map. The report and the tool help to analyze and to understand the current situation for pharmacies in the City of Zurich. But this is also useful for future developments. Where should I open a pharmacy? Where should I close a pharmacy? Are there marketing instruments to make a situation better? Should we - as a chain - cooperate with competitors or even buy them? Should we sell stores or focus on other products? And so on. One important basis for decision makers are facts. We can identify them with the help of Foursquare.

1.2 A: the Location Value for existing Pharmacies

There are several pharmacies in the city of Zurich. What is the location value of these pharmacies? If we can put a real value on them, we can compare these pharmacies and visualize the result table with a barchart.

1.3 B: finding the best Places for new Pharmacies

Can we identify areas in the city, where new pharmacies are needed? How can we do that and what influences the value of pharmacies. What is the role of contributors and competitors? We could calculate every spot in a map and visualize the result as in a heatmap.

1.4 Audience

The audience is everyone who is interested in the pharmacy business. The results of this research are interesting for pharmacy marketing leaders, pharmacy consultants and pharmacist. But it might also be interesting for the urban administration for developing planning the city's future and caring about the attractiveness of the city for residents and tourists.

2 Data acquisition and cleaning

2.1 Data sources and selections

2.1.1 Location data from Foursquare

The source for the location data is Foursquare. Foursquare is a social network for location data and was very popular from 2010 till 2014. But Foursquare is probably still one of the best data providers for location data. Foursquare has an API interface where developers can get some data for free after a registration. My experience is that Foursquare limits the answers with the free sandbox account to 50 objects. With the sandbox account I can get:

- 950 Regular Calls/Day
- 50 Premium Calls/Day
- 1 Photo per Venue
- 1 Tip per Venue

But even with the commercial enterprise account, the answer is limited to 50 objects:

Enterprise

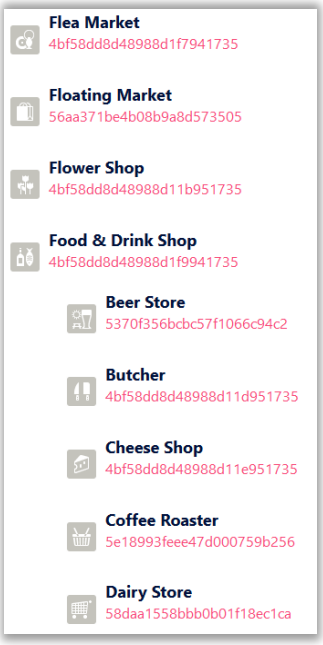
	Venue Search, Trending, Similar, Next, Photos, Tips	Venue Explore	Venue Details
# of Venues per Call	Up to 50 results	Up to 50 results	1 result

Selecting the right research area

This means, that developers have to narrow the area they want to explore or to send several request with different types of information requests. I suggest to define a small research area. In my case i defined a rectangle of research, calculated the middle and used the middle, corner distance as a radius for the Foursquare circle, because Foursquare does the research within a circle.

Selecting the right categories

There are about 200 categories defined in Foursquare. The categories can be found at: https://developer.foursquare.com/docs/build-with-foursquare/categories/ and have a hierarchical structure. Selecting the useful categories is a job that should be done with marketing experts.



	categoryId	categoryname
0	4bf58dd8d48988d10f951735	Pharmacy
1	52f2ab2ebcbc57f1066b8b23	Perfume Shop
2	58daa1558bbb0b01f18ec206	Medical Supply Store
3	4bf58dd8d48988d177941735	Doctor s Office
4	54541900498ea6ccd0202697	Health Beauty Service
5	5745c2e4498e11e7bccabdbd	Drugstore
6	4bf58dd8d48988d10c951735	Cosmetics Shop

2.1.2 Map Data from Openstreetmap

For visualizing the location of stores, shops and some offices I used Openstreetmap (osm) from openstreetmap.org. The data can be used for free by a fair use policy. The zoomable map can display houses at the street.

2.2 Data cleaning

The limit of 50 results makes programming a little bit more difficult. In my case I sent a own a selection for every category ID. I did not send 1 request including all categories, but I sent 7 requests and accumulated the result in a dataframe (table). Every request had less than 50 results.

Probably caused by the hierarchical structure, I received a few duplicates. I eliminated these duplicates from my table. A further data cleaning was not necessary

3 Methodology: Exploratory Data Analysis

Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why

3.1 Definition of the research area

Working with map data means working with the coordinates longitude and latitude. The research rectangle was defined by coordinates:

```
min_lat    = 47.355   # Below, lower border
max_lat    = 47.385   # Top, upper border
min_lng    =  8.515   # Left border
max_lng    =  8.560   # Right border
```

Bringing coordinates and meters together was quite tricky and the results depend on the location:

```
The Zurich Rectangle is 3336 Meters wide and 5004 Meters high.
1 Latitude has a Length of             : 111195 Meters
1 Longitude has a Length of            : 75308 Meters
Latitude per Longitude Meter Proportion: 1.476536357359112
Grid in Meters Latitude                : 22.239
Grid in Meters Longitude               : 22.239
Grid Difference Lat Lng in Millimeters : 0.021
step stp_lat                           : 0.0002
Step stp_lng                           : 0.000295
```

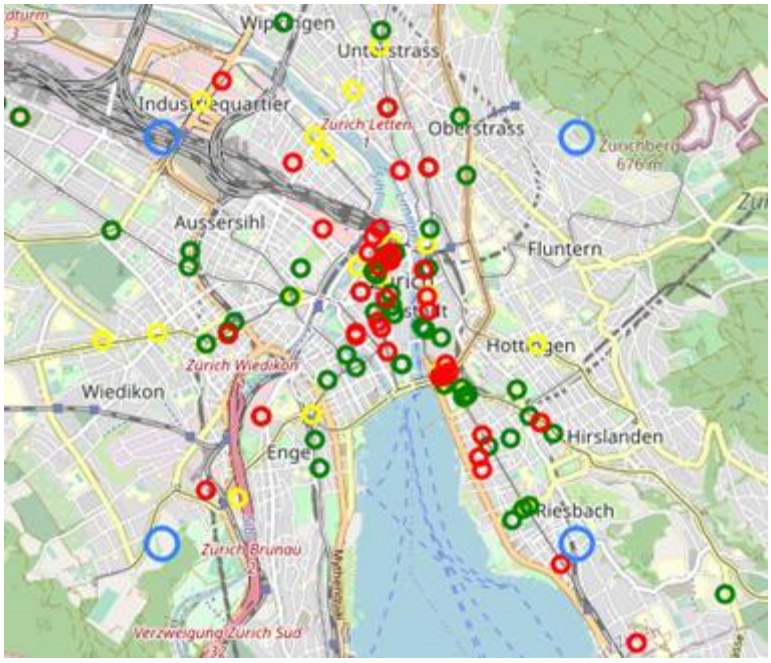
The research area is between the blue circles:

3.2 Finding and marking objects

With the help of Foursquare and without any manual manipulation I could identify pharmacies (yellow circles) contributors (green circles) and competitors (red circles).

The color rules are in this table:

categoryname	dotcolor
Pharmacy	yellow
Perfume Shop	red
Medical Supply Store	red
Doctor s Office	green
Health Beauty Service	green
Drugstore	green
Cosmetics Shop	green

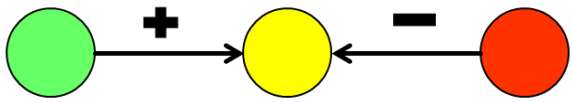


High object density at the street "Bahnhofstrasse".

3.3 Calculating Value for A

3.3.1 Rule 1

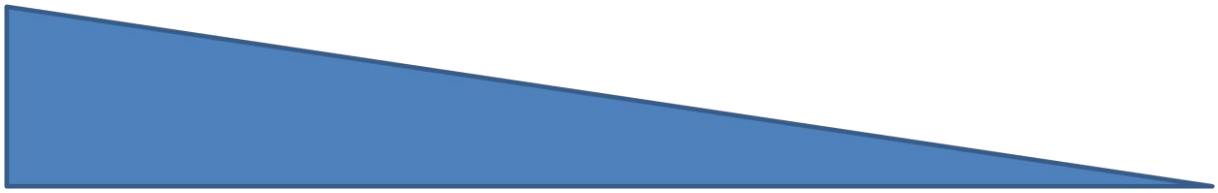
Pharmacy value is influenced by surrounding stores. These stores have a positive or negative effect on the value of pharmacies:



3.3.2 Rule 2

The closer the object is, the higher is the influence. Objects far away have no influence on the pharmacy.

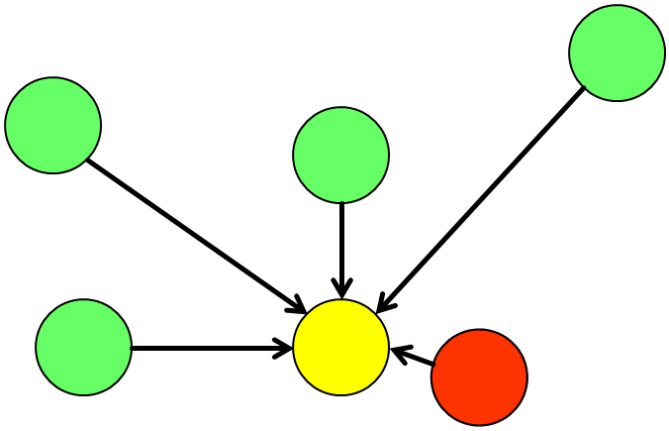
I decided to use a linear model to calculate influences.



Objects with a distance of 0 (zero) have an influence of 1 (positive) or -1 (negative case). Objects with a distance of 400 meters and more have no, zero influence. I did no further differentiation of the objects like give different weights to different types.

3.3.3 Rule 3

The final location value of the Pharmacy is the sum of all single values.



3.4 Calculation for B

Now we have seen how the location value for an existing pharmacy can be calculated. But how we can identify best locations for new pharmacies?

The solution is that we make a grid into the rectangle and calculate every cell of the grid as done before.

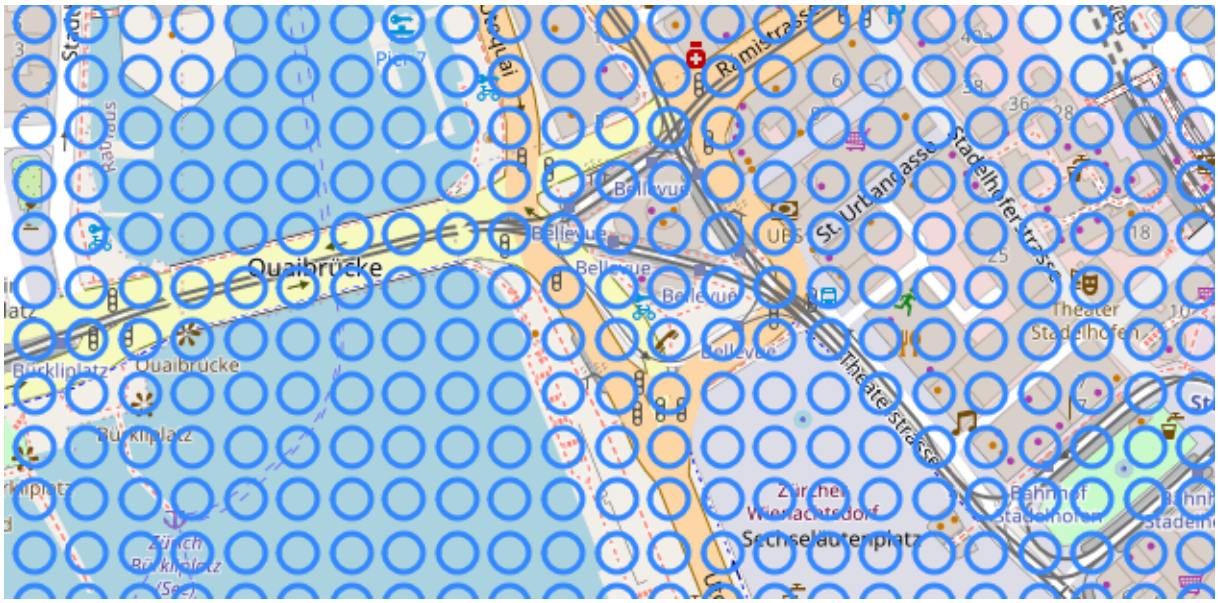
I decided to define the grid by fractions (steps) of one Latitude.

```
stp_lat     =  0.0002
```

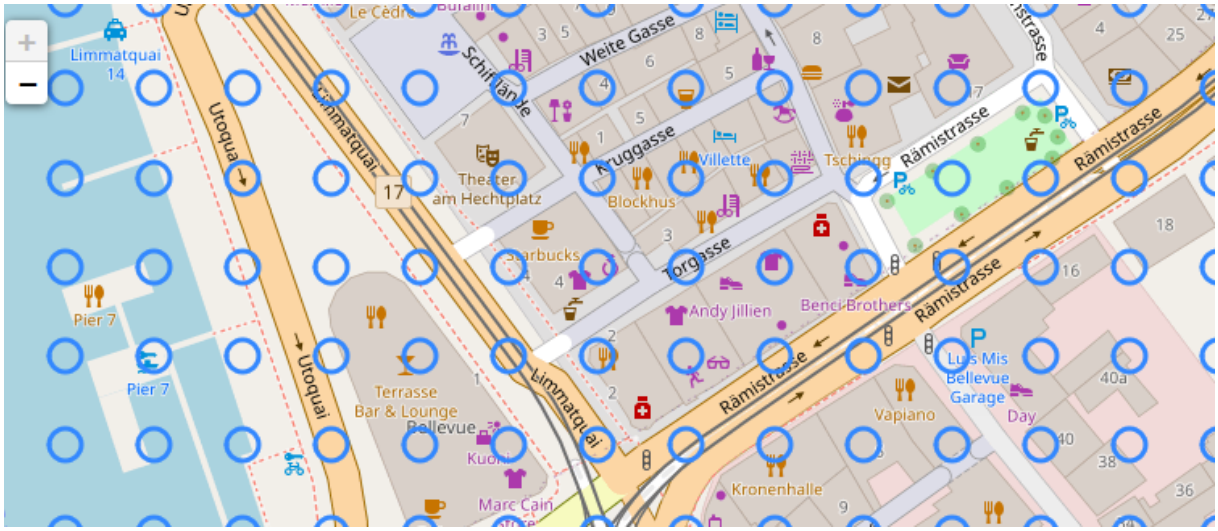
This number creates a cell distance of 22 meters. The smaller the grid the higher is the calculation time. With this grid my PC calculated for 6 and a half minutes.

How does the grid look like? This is the rectangle for the grid.

Zooming in:



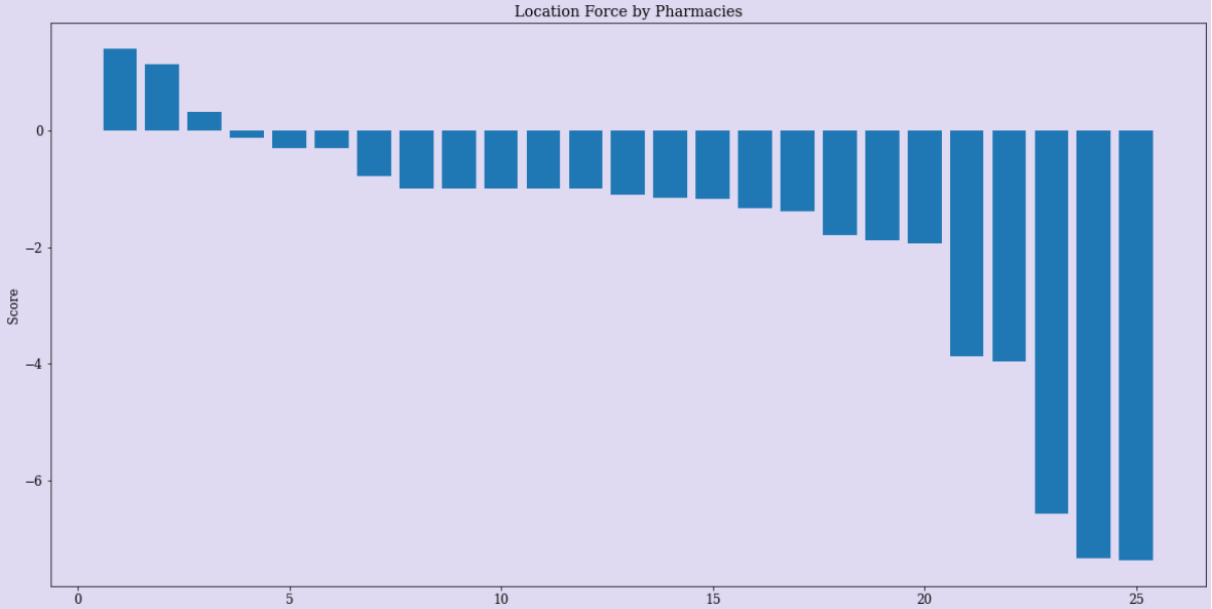
Closer:



The distance from circle to circle is 22 meters.

4 Results

4.1 Results for A



The chart shows the total location value for each pharmacy. The numbers on the x-axis are replacements for pharmacy names.

This is the list of the 25 pharmacies:

1	TopPharm Apotheke & Drogerie Höschgasse
2	Dr. Andres Apotheke Stadelhofen
3	Berg-Apotheke
4	Anrig Drogerie Naturathek
5	Topwell Apotheke-Drogerie
6	Apotheke Schaffhauserplatz
7	Vision hair
8	DROPA Apotheke & Post Hottingen
9	Drogama Apotheke Drogerie
10	Wehntal Apotheke
11	Bären-Apotheke
12	Neumarkt Apotheke Drogerie
13	Bellevue Apotheke
14	TopPharm Leonhards-Apotheke
15	Rosen Apotheke
16	Coop Vitality
17	apodoc
18	Odeon Apotheke
19	DROPA Drogerie Apotheke Limmatplatz
20	TopPharm Limmatplatz Apotheke
21	Sun Store Apotheke
22	Victoria Apotheke
23	Bahnhof Apotheke
24	Coop Vitality Zürich Bahnhofstrasse
25	Amavita Apotheke

4.2 Results for B

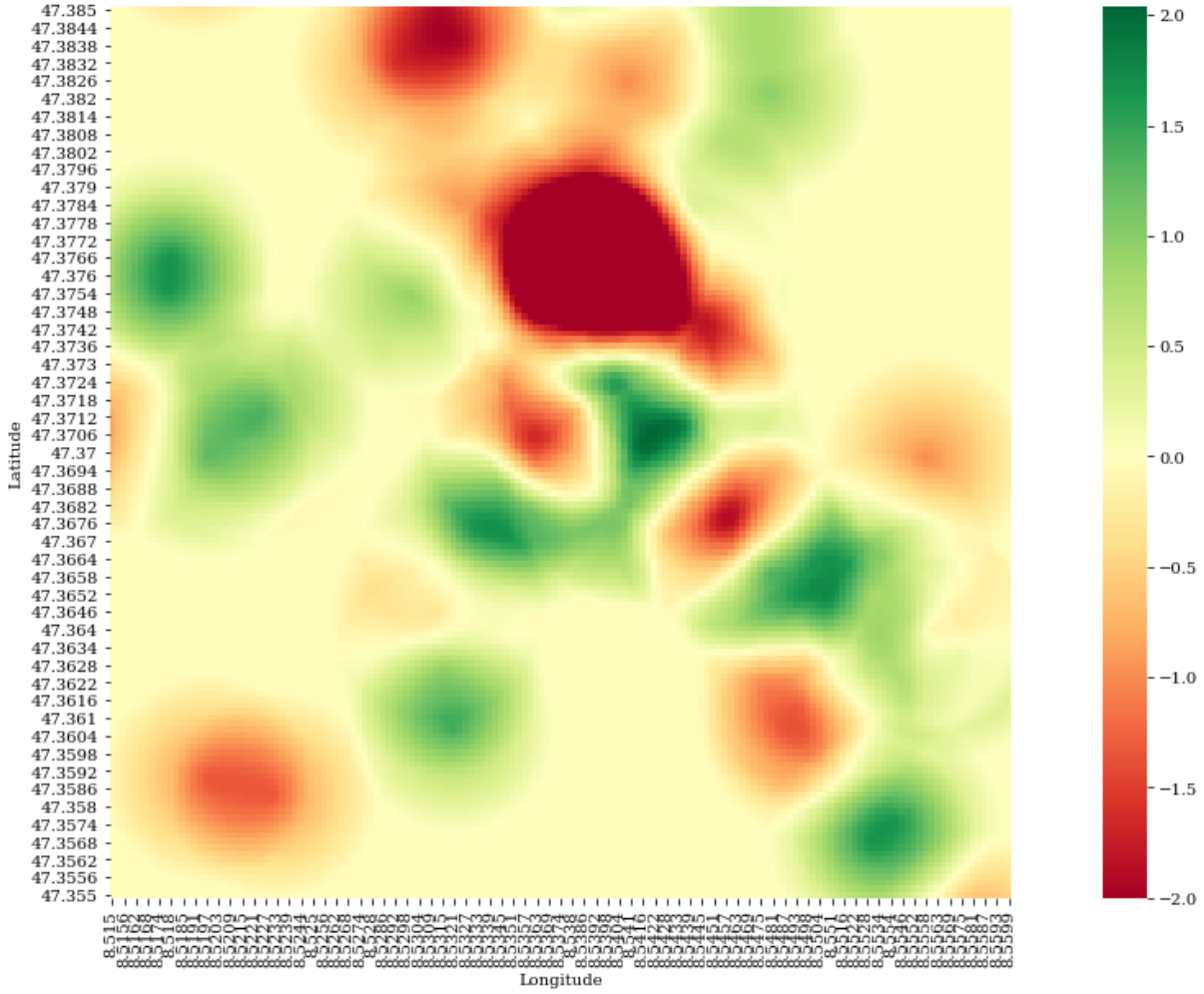
4.2.1 Table

Calculating each cell generates a table. Using pivot I can create the following table with 151 rows and 153 columns. Here we see the force by cell:

lng / lat	8.515000	8.515295	8.515591	8.515886	8.516181	8.516477	8.516772
47.3850	0.0	0.0	-0.003501	-0.040363	-0.075352	-0.108248	-0.138809
47.3848	0.0	0.0	0.000000	0.000000	-0.030759	-0.062089	-0.091097
47.3846	0.0	0.0	0.000000	0.000000	0.000000	-0.014955	-0.042534
47.3844	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.000000
47.3842	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.000000

4.2.2 Heatmap

The visualizing in a heatmap is easy to understand:

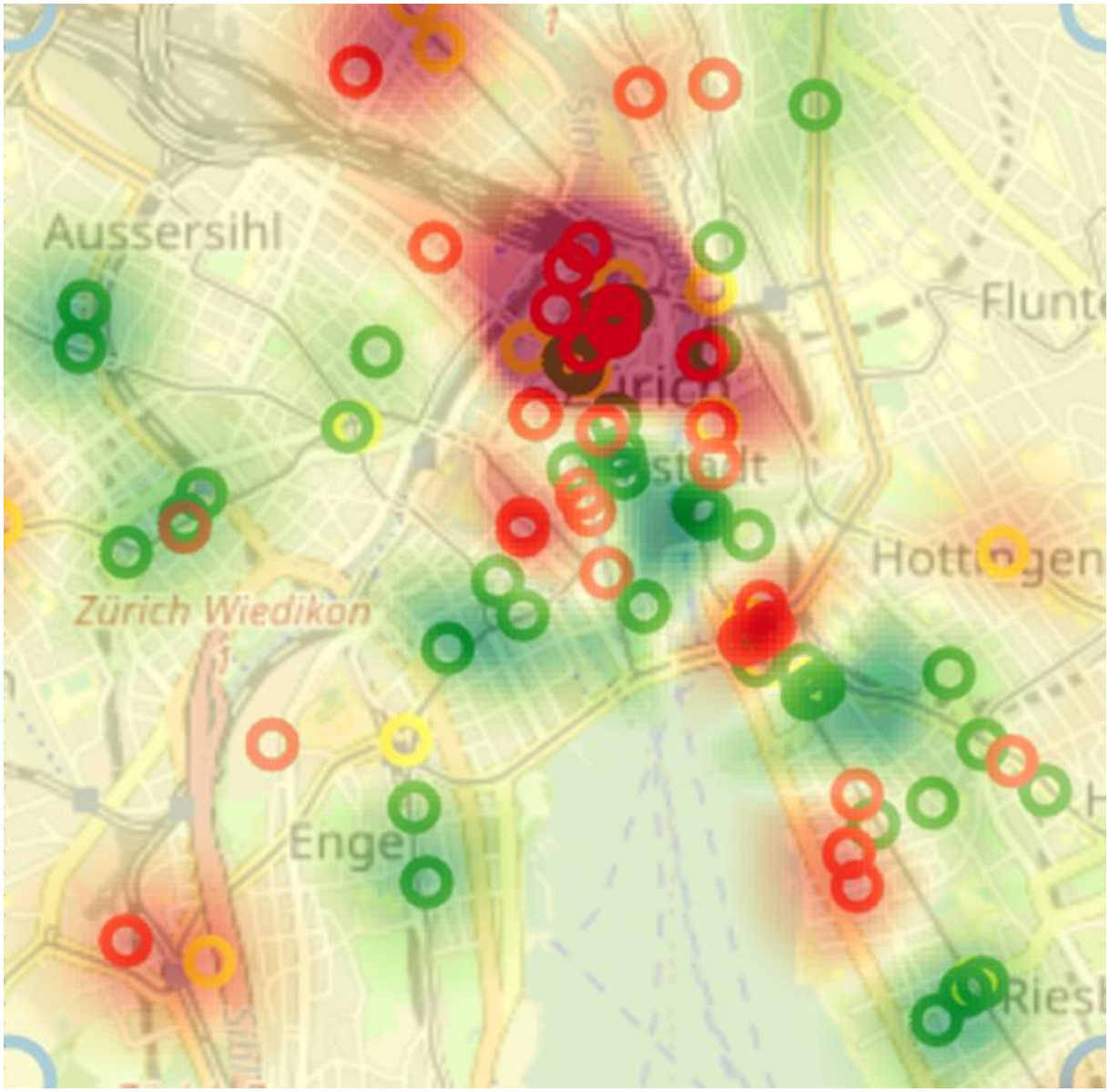


4.2.3 The Heatmap in Details

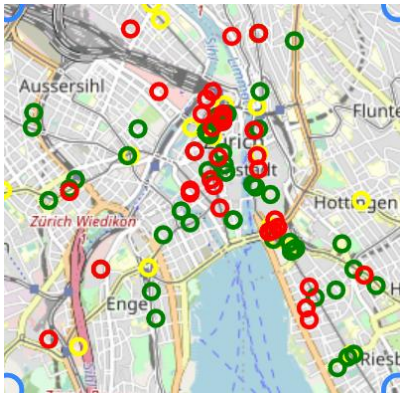
Here is a detailed extract of the heatmap. Note that each pixel represents a grid cell, which is a real square with a length of 22 meters:

4.2.4 Combination of Heatmap and Streetmap

If I bring the city map and the heatmap together, we can clearly see and understand the good places for new pharmacies.



Even with location information, without the heatmap it is difficult to identify the good places.



5 Conclusion, Discussion and Future Development

I think the basic work is done. But there are many parameters that can be changed. For such changes we need the expertise from professionals with a deep understanding of the business rules. Questions for the future development are:

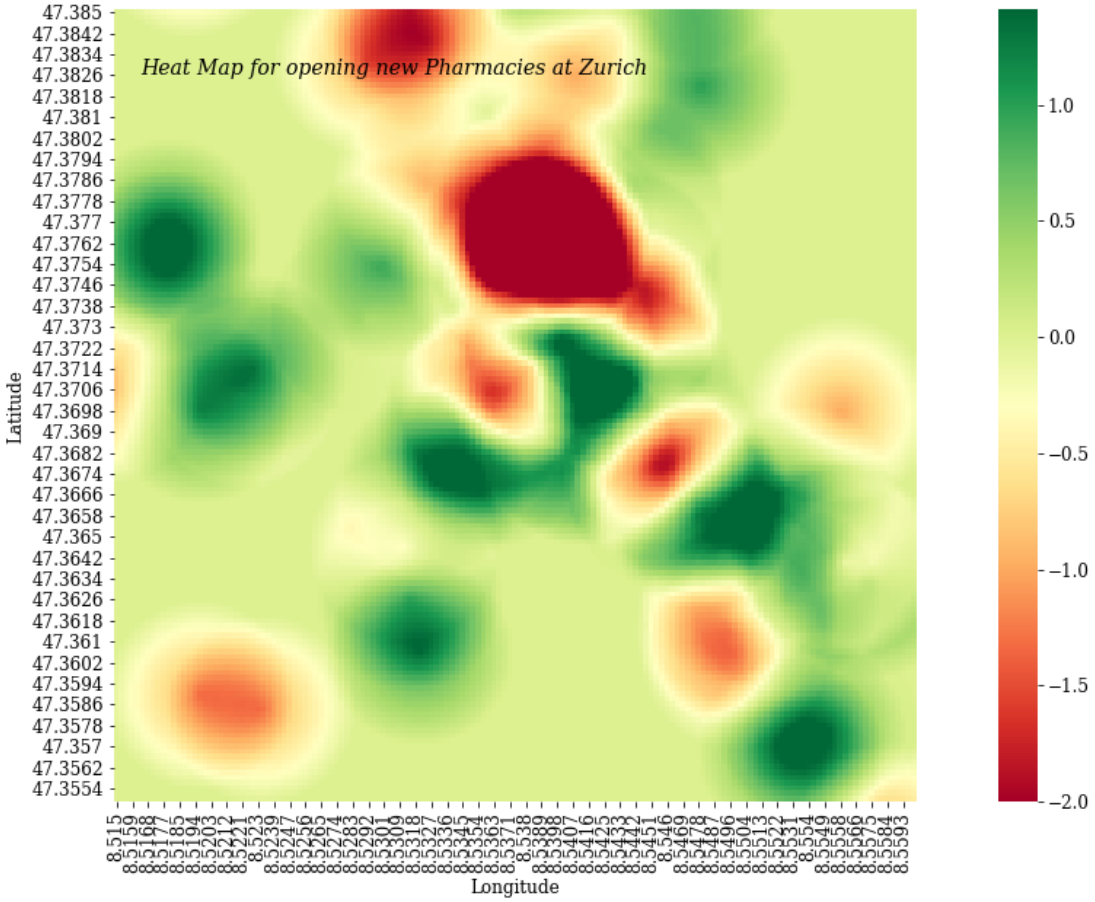
- We assume a linear model for the study. Do better models already exist?

- We ignored other important factors like the pedestrian traffic per day because this data is not available for me. Are there additional public sources? Should I make my own estimates or should I even count pedestrians?

- We did not weigh the categories. We only differentiate by a positive or negative influence. Should we weigh categories or treat them all the same? Did we select the right categories or do we ignore important ones?

We assume that after the radial distance of 400 meters the influence is zero. This is however, just a guess. Should we interview people find and identify a better parameters?

The implementation can be adopted and adapted for other locations and fields. The power of this software is the full automation of the process of gathering data, calculating and visualizing Results by creating intuitive and powerful graphics.



6 Appendix

6.1 What is Coursera?

Coursera.org is an online platform where you can take courses in many fields. The courses are provided by the best professors from the best universities worldwide.

Coursera offers:

- Courses
- Guided Projects
- Specializations
- Professional Certificates
- MasterTrack™ Certificates
- Online Degrees

"Coursera provides universal access to the world's best education, partnering with top universities and organizations to offer courses online."

(https://about.coursera.org)

6.2 What is the IBM Data Science Specialization program?

The IBM Data Science Specialization is a intensive learning program running on coursera. The Program contains 9 courses and is produced by IBM. The program exists for several years but is adapted to the newest technologies and methods. The current teachers of the program are:

Alex Aklson
Ph.D., Data Scientist
IBM Developer Skills Network
355,513 Kursteilnehmer
7 Kurse

Polong Lin
Data Scientist
IBM Developer Skills Network
322,424 Kursteilnehmer
2 Kurse

Romeo Kienzler
Chief Data Scientist, Course Lead
IBM Watson IoT
227,640 Kursteilnehmer
6 Kurse

Svetlana Levitan
Senior Developer Advocate with IBM Center for Open Data and AI Technologies
142,385 Kursteilnehmer
1 Kurse

Joseph Santarcangelo
Ph.D., Data Scientist at IBM
IBM Developer Skills Network
337,313 Kursteilnehmer
5 Kurse

Rav Ahuja
AI and Data Science Program Director
IBM
213,346 Kursteilnehmer
6 Kurse

SAEED AGHABOZORGI
Ph.D., Sr. Data Scientist
IBM Developer Skills Network
151,441 Kursteilnehmer
1 Kurse

Source: https://www.coursera.org/professional-certificates/ibm-data-science

The Course Lead and Chief Data Scientist Remo Kienzler is a former Student of the ETH Zurich. The 9 Courses are:

1. What is Data Science?
2. Tools for Data Science
3. Data Science Methodology
4. Python for Data Science and AI
5. Databases and SQL for Data Science
6. Data Analysis with Python
7. Data Visualization with Python
8. Maschine Learning with Python
9. Applied Data Science Capstone

This paper is my final work to get the IBM Data Science Certificate.

Pascal Keller 10.8.2020

* * *