Sandra Carrasco

The battle of neighborhoods

Final Capstone Project

IBM / Coursera

4th May 2021

INTRODUCTION: BUSINESS PROBLEM

In this project we will try to find an optimal location for a restaurant. Specifically, given the ever increasing demand (and scarcity of options) for gluten-free options in restaurants, this report will be targeted to stakeholders interested in opening a Gluten-free restaurant in Berlin.

Since there are lots of restaurants in Berlin we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no Gluten-free restaurants in vicinity. We would also prefer locations as close to the city center as possible, assuming that the first two conditions are met.

We will use data science to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

DATA

Based on definition of our problem, factors that will influence our decision are:

- number of existing restaurants in the neighborhood (any type of restaurant)
- number of and distance to Gluten Free restaurants in the neighborhood, if any
- distance of neighborhood from city center

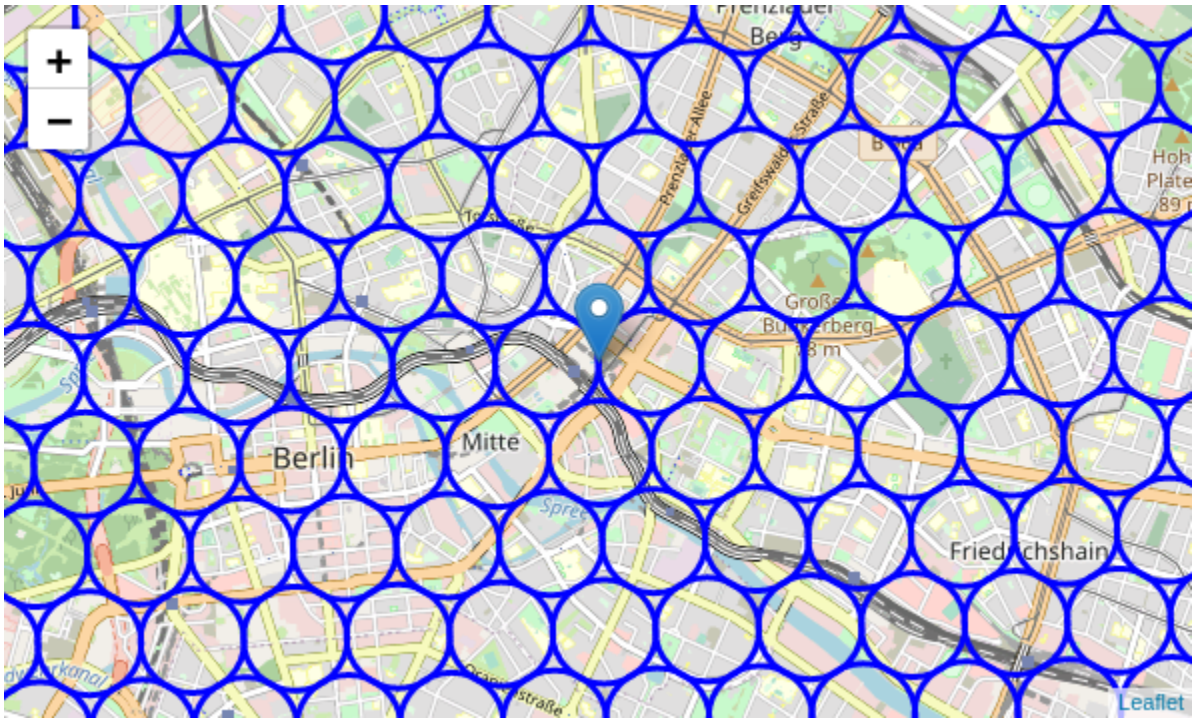
I decided to use a regularly spaced grid of locations, centered around the city center, to define the neighborhoods.

Following data sources will be needed to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding
- number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API
- coordinate of Berlin center will be obtained using Google Maps API geocoding of a well known Berlin location.

Neighborhood Candidates

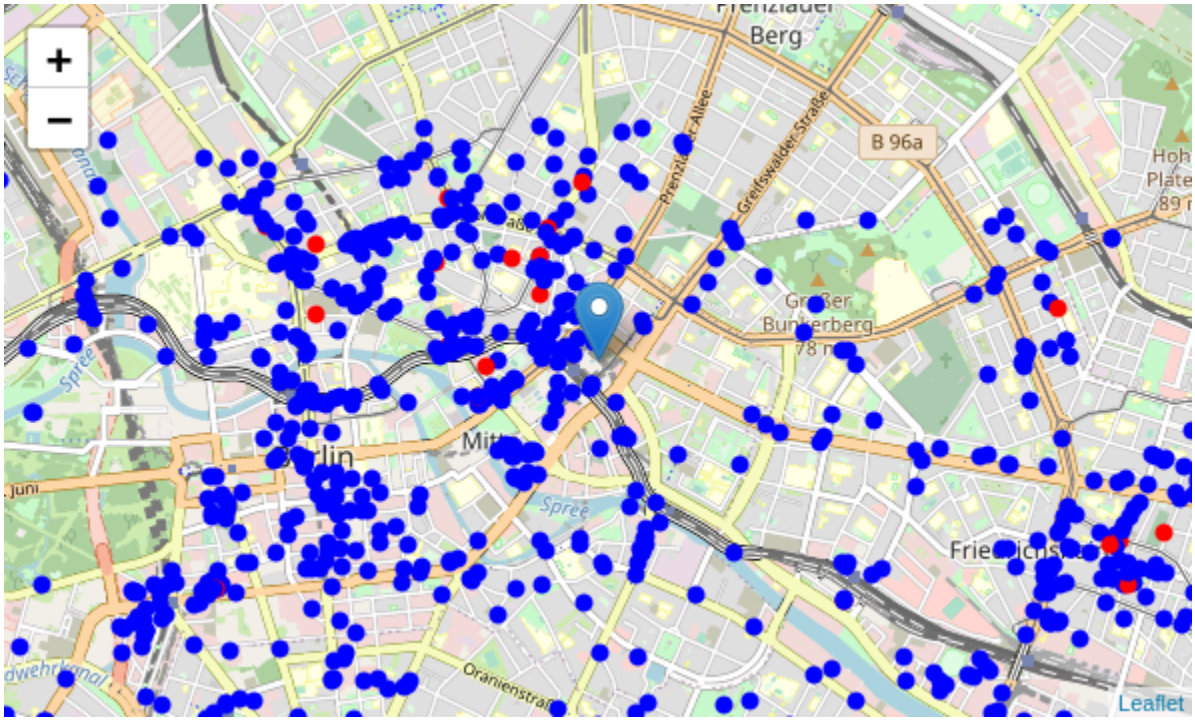
We will create a grid of cells covering our area of interest which is aprox. 12x12 kilometers centered around Berlin city center. Our neighborhoods will be defined as circular areas with a radius of 300 meters, so our neighborhood centers will be 600 meters apart. We now have the coordinates of centers of neighborhoods/areas to be evaluated, equally spaced (distance from every point to it's neighbors is exactly the same) and within ~6km from Alexanderplatz. Using Google Maps API we get approximate addresses of those locations.



Now that we have our location candidates, we use Foursquare API to get info on restaurants in each neighborhood. We're interested in venues in 'food' category, but only

those that are proper restaurants - coffe shops, pizza places, bakeries etc. are not direct competitors so we don't care about those. So we will include in out list only venues that have 'restaurant' in category name, and we'll make sure to detect and include all the subcategories of specific gluten-free restaurant' category.

Let's now see all the collected restaurants in our area of interest on map, and let's also show gluten-free restaurants in different color.



So now we have all the restaurants in an area within a few kilometers from Alexanderplatz, and we know which ones are Italian restaurants! We also know which restaurants exactly are in the vicinity of every neighborhood candidate center.

METHODOLOGY

In this project we will direct our efforts on detecting areas of Berlin that have low restaurant density, particularly those with low number of gluten-free restaurants. We will limit our analysis to areas ~6km around the city center.

In the first step we have collected the required data: location and type (category) of every restaurant within 6km from Berlin center (Alexanderplatz). We have also identified gluten-free restaurants (according to Foursquare categorization).

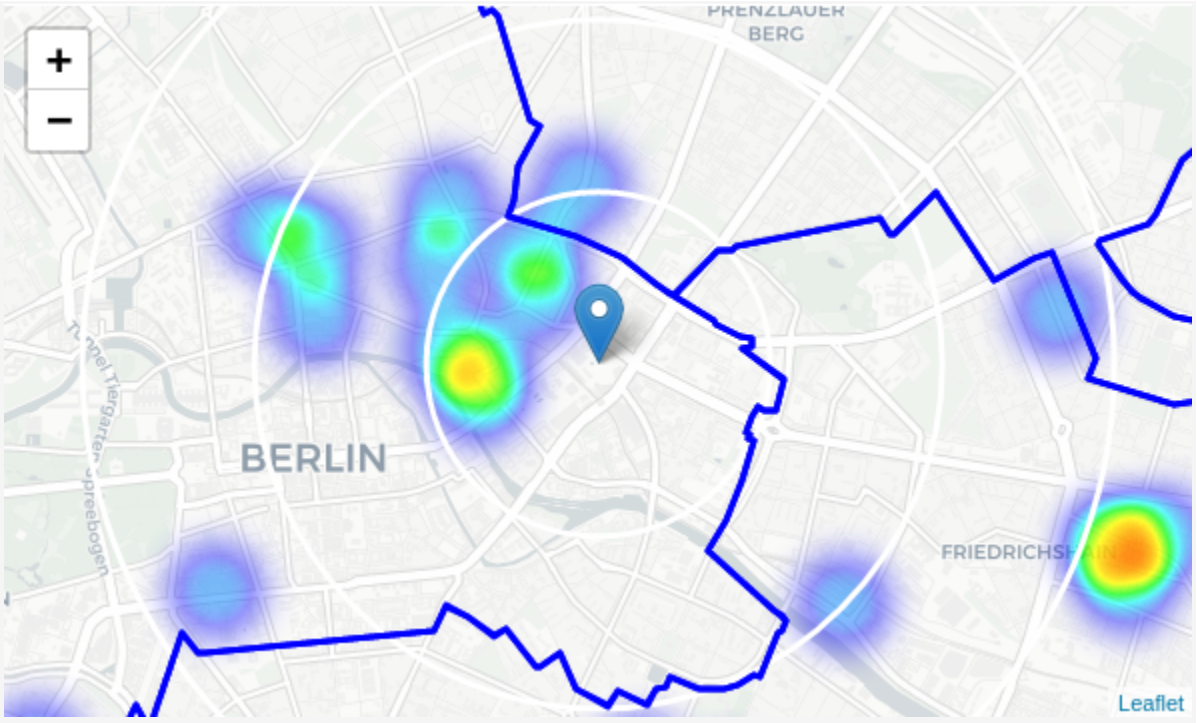
Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of Berlin - we will use heatmaps to identify a few promising areas close to the center with low number of restaurants in general (and no gluten-free restaurants in vicinity) and focus our attention on those areas.

In the third and final step we will focus on the most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two restaurants in a radius of 250 meters, and we want locations without gluten-free restaurants in radius of 400 meters. We will present a map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

ANALYSIS

We first perform some basic explanatory data analysis and derive some additional info from our raw data. For example, the average number of restaurants in every area with radius=300m: 4.912. Then we calculate the distance to the nearest gluten-free restaurant from every area candidate center. On average, gluten-free restaurant can be found within ~500m from every area center candidate. That's fairly close, so we need to filter our areas carefully! We create a map showing the density of restaurants in order to extract meaningful information.

Looks like a few pockets of low restaurant density closest to the city center can be found south, south-east and east from Alexanderplatz. The following heatmap shows the density of gluten-free restaurants only.



This map is not so 'hot' but it also indicates higher density of existing gluten-free restaurants directly north and west from Alexanderplatz, with closest pockets of low gluten-free restaurant density positioned east, south-east and south from the city center.

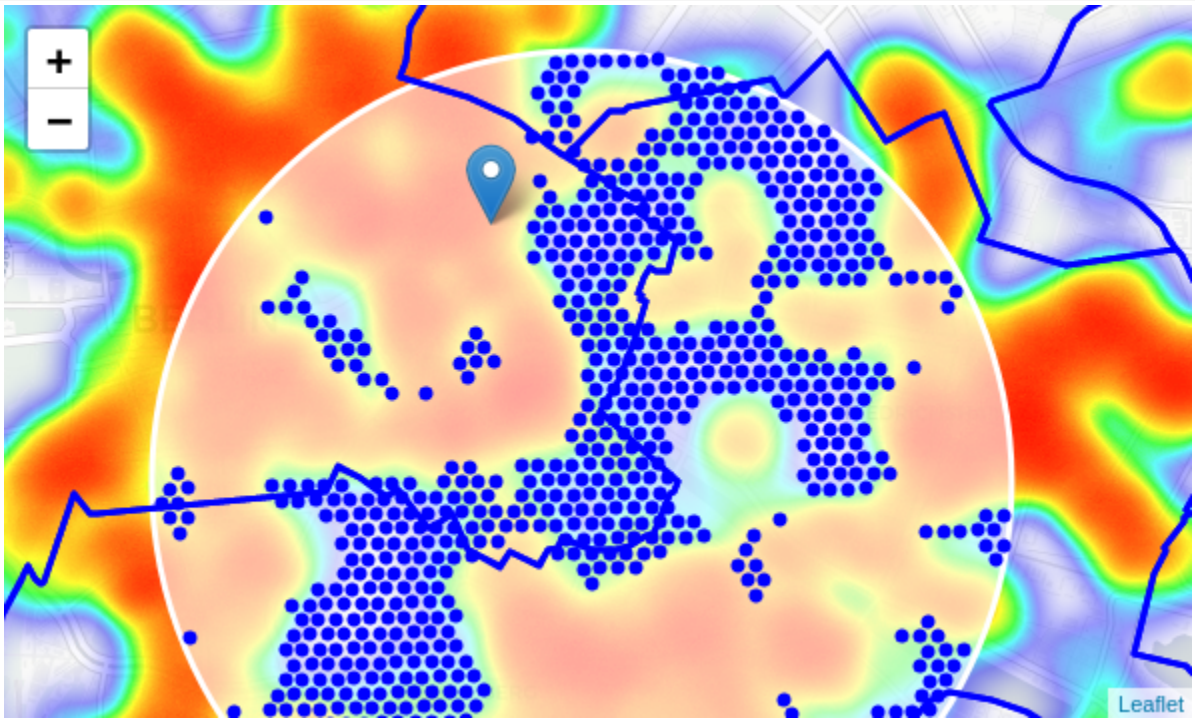
Based on this we will now focus our analysis on areas south-west, south, south-east and east from Berlin center - we will move the center of our area of interest and reduce its size to have a radius of 2.5km. This places our location candidates mostly in boroughs Kreuzberg and Friedrichshain (another potentially interesting borough is Prenzlauer Berg with large low restaurant density north-east from city center, however this borough is less interesting to stakeholders as it's mostly residential and less popular with tourists).

Kreuzberg and Friedrichshain

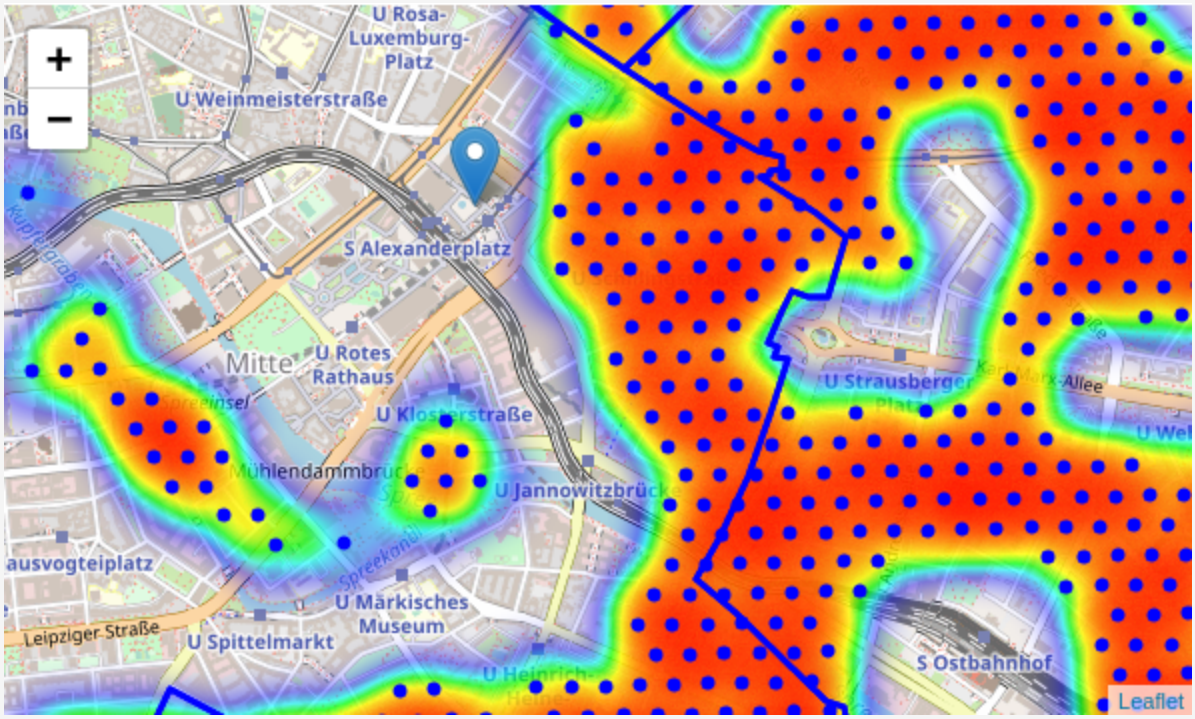
Analysis of popular travel guides and web sites often mention Kreuzberg and Friedrichshain as beautiful, interesting, rich with culture, 'hip' and 'cool' Berlin neighborhoods popular with tourists and loved by Berliners.

Popular with tourists, alternative and bohemian but booming and trendy, relatively close to the city center and well connected, those boroughs appear to justify further analysis.

Let's define a new, more narrow region of interest, which will include low-restaurant-count parts of Kreuzberg and Friedrichshain closest to Alexanderplatz. We also create new, more dense grid of location candidates restricted to our new region of interest, generating a total of 2261 candidate neighborhood centers. We calculate the number of restaurants in vicinity (we'll use radius of 250 meters) and distance to closest Italian restaurant in order to filter the locations with no more than two restaurants in radius of 250 meters, and no gluten-free restaurants in radius of 400 meters.

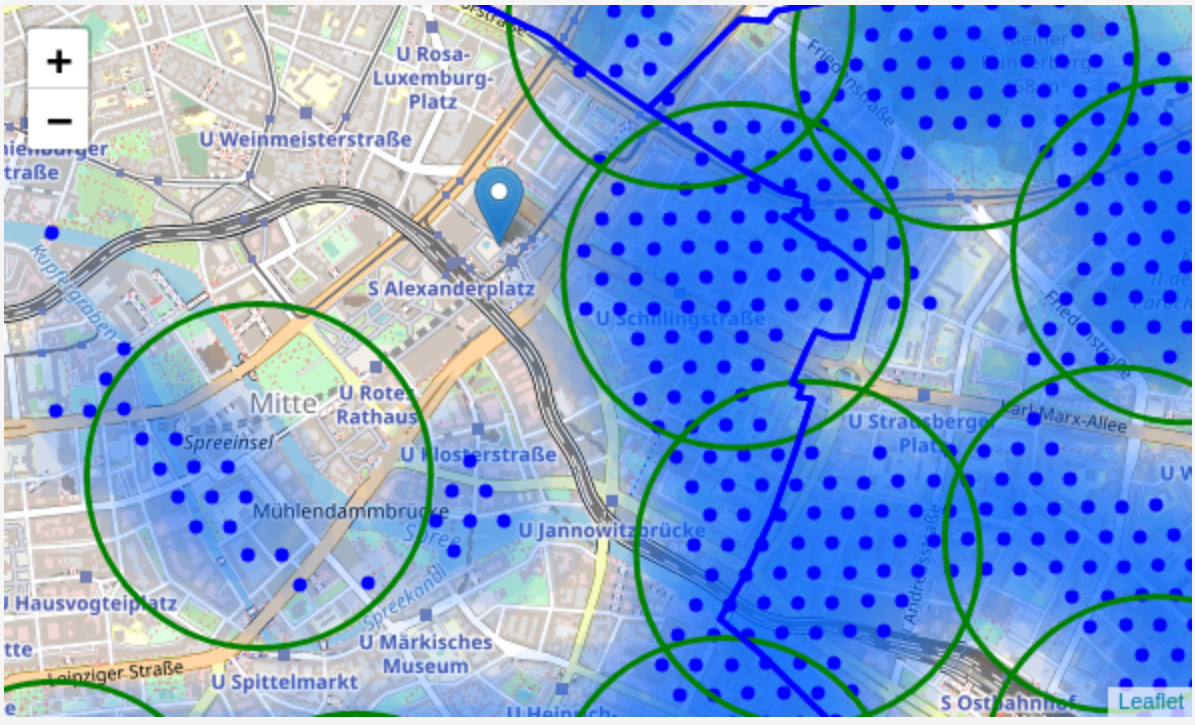


We now have a bunch of locations fairly close to Alexanderplatz (mostly in Kreuzberg, Friedrichshain and south-east corner of Mitte boroughs), and we know that each of those locations has no more than two restaurants in radius of 250m, and no gluten-free restaurant closer than 400m. Any of those locations is a potential candidate for a new gluten-free restaurant, at least based on nearby competition.

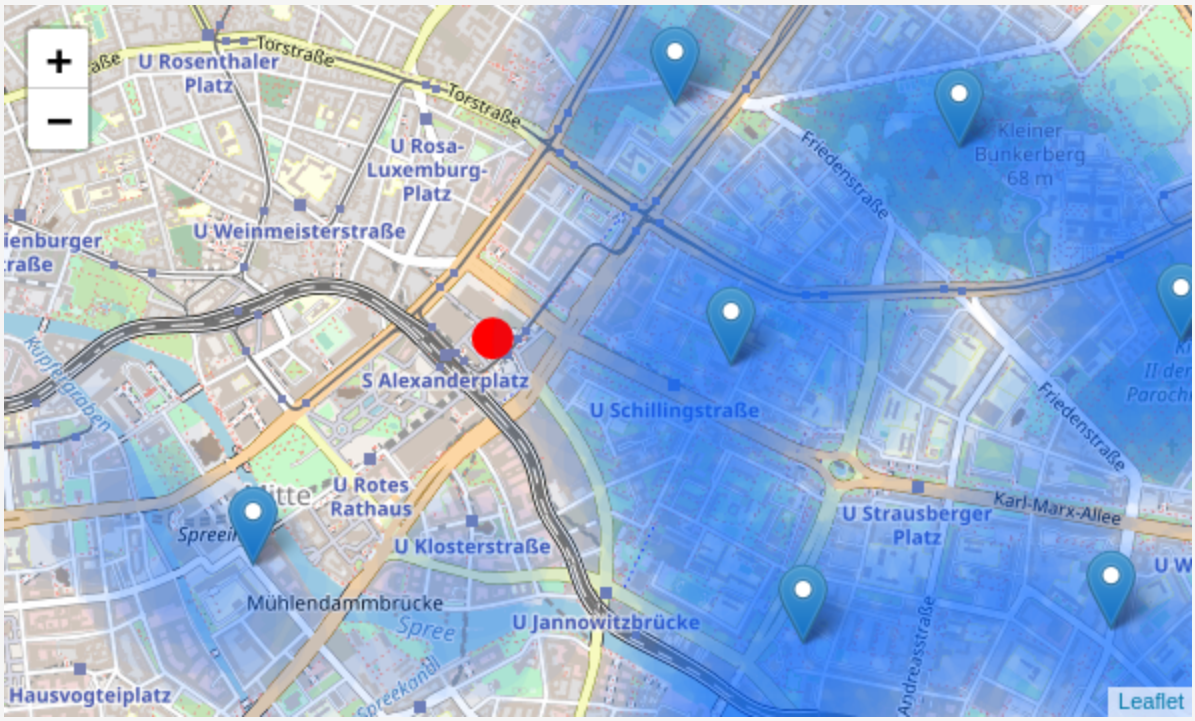


Now we have a clear indication of zones with low number of restaurants in vicinity, and no Italian restaurants at all nearby.

Finally, we cluster those locations to create centers of zones containing good locations. Those zones, their centers and addresses will be the final result of our analysis.



Our clusters represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates. Finally, we reverse geocode those candidate area centers to get the addresses which can be presented to stakeholders.



RESULTS AND DISCUSSION

Our analysis shows that although there is a great number of restaurants in Berlin (~2000 in our initial area of interest which was 12x12km around Alexanderplatz), there are pockets of low restaurant density fairly close to the city center. Highest concentration of restaurants was detected north and west from Alexanderplatz, so we focused our attention to areas south, south-east and east, corresponding to boroughs Kreuzberg, Friedrichshain and south-east corner of central Mitte borough. Another borough was identified as potentially interesting (Prenzlauer Berg, north-east from Alexanderplatz), but our attention was focused on Kreuzberg and Friedrichshain which offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics and a number of pockets of low restaurant density.

After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Alexanderplatz) we first created a dense grid of location

candidates (spaced 100m appart); those locations were then filtered so that those with more than two restaurants in a radius of 250m and those with an gluten-free restaurant closer than 400m were removed.

Those location candidates were then clustered to create zones of interest which contain the greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 15 zones containing the largest number of potential new restaurant locations based on number of and distance to existing venues. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to Berlin center but not crowded with existing restaurants (particularly gluten-free) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

CONCLUSION

The purpose of this project was to identify Berlin areas close to the center with a low number of restaurants (particularly gluten-free restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new gluten-free restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis (Kreuzberg and Friedrichshain), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

The final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location

(proximity to park or water), levels of noise, proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.