Capstone Project - The Battle of the Neighborhoods (Week 2)

Table of contents

Introduction: Business Problem

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an **Italian restaurant** in **Berlin**, Germany.

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 Italian restaurants in vicinity**. We would also prefer locations **as close to city center as possible**, assuming that first two conditions are met.

We will use our data science powers to generate a few most promissing 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 decission are:

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

We decided to use regularly spaced grid of locations, centered around city center, to define our 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 well known Berlin location (Alexanderplatz)

Neighborhood Candidates

Let's create latitude & longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is aprox. 12x12 killometers centered around Berlin city center.

Let's first find the latitude & longitude of Berlin city center, using specific, well known address and Google Maps geocoding API.

```
Coordinate of Alexanderplatz, Berlin, Germany: [52.5219184, 13.4132147]
```

Now let's create a grid of area candidates, equaly spaced, centered around city center and within ~6km from Alexanderplatz. Our neighborhoods will be defined as circular areas with a radius of 300 meters, so our neighborhood centers will be 600 meters apart.

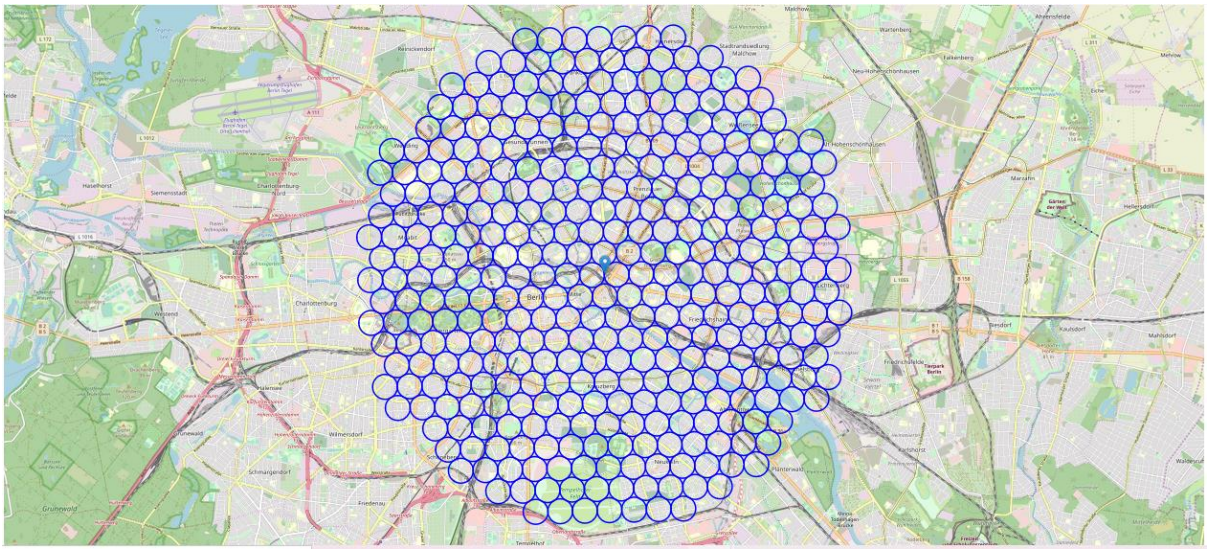
To accurately calculate distances we need to create our grid of locations in Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we'll project those coordinates back to latitude/longitude degrees to be shown on Folium map. So let's create functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

```
Coordinate transformation check
-------------------------------
Berlin center longitude=13.4132147, latitude=52.5219184
Berlin center UTM X=392341.28017572395, Y=5820273.243274779
Berlin center longitude=13.413214700000001, latitude=52.52191839999997
```

Let's create a **hexagonal grid of cells**: we offset every other row, and adjust vertical row spacing so that **every cell center is equally distant from all it's neighbors**.

```
364 candidate neighborhood centers generated.
```

Let's visualize the data we have so far: city center location and candidate neighborhood centers:



OK, 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.

Let's now use Google Maps API to get approximate addresses of those locations.

```
Reverse geocoding check
-----------------------
Address of [52.5219184, 13.4132147] is: Alexanderpl. 5, 10178 Berlin,
Germany
['Frankfurter Allee 147-149, 10365 Berlin',
 'Magdalenenstraße 12, 10365 Berlin',
 'Siegfriedstraße 207, 10365 Berlin',
 'Englische Str. 3, 10587 Berlin',
 'Händelallee 51, 10557 Berlin',
 'Spreeweg, 10557 Berlin',
 'John-Foster-Dulles-Allee 10, 10557 Berlin',
 'B96, 10557 Berlin',
 'Pariser Platz 6A, 10117 Berlin',
 'Unter den Linden 38, 10117 Berlin',
 'Unter den Linden 5, 10117 Berlin',
 'Spreeufer 6, 10178 Berlin',
 'Parochialstraße, 10179 Berlin',
 'Neue Blumenstraße 1, 10179 Berlin',
 'Blumenstraße 41, 10243 Berlin',
 'B5 85, 10243 Berlin',
 'Weidenweg 27, 10249 Berlin',
 'Rigaer Str. 96, 10247 Berlin',
 'Bänschstraße 58, 10247 Berlin',
 'Parkaue 30, 10367 Berlin']
```

	Address	Latitude	Longitude	X	Y	Distance from center
0	Ringbahnstraße 50, 12099 Berlin	52.470194	13.388575	390541.280176	5.814557e+06	5992.495307
1	09R/27L, 12101 Berlin	52.470314	13.397404	391141.280176	5.814557e+06	5840.376700
2	09R/27L, 12049 Berlin	52.470434	13.406234	391741.280176	5.814557e+06	5747.173218
3	Oderstraße 174, 12049 Berlin	52.470552	13.415063	392341.280176	5.814557e+06	5715.767665
4	Warthestraße 23A, 12051 Berlin	52.470670	13.423893	392941.280176	5.814557e+06	5747.173218
5	Altenbraker Str. 15, 12053 Berlin	52.470788	13.432722	393541.280176	5.814557e+06	5840.376700
6	Karl-Marx-Straße 212, 12055 Berlin	52.470904	13.441552	394141.280176	5.814557e+06	5992.495307
7	Hessenring 34, 12101 Berlin	52.474683	13.375159	389641.280176	5.815077e+06	5855.766389
8	Kleineweg 125, 12101 Berlin	52.474804	13.383989	390241.280176	5.815077e+06	5604.462508
9	09L/27R, 12101 Berlin	52.474924	13.392820	390841.280176	5.815077e+06	5408.326913

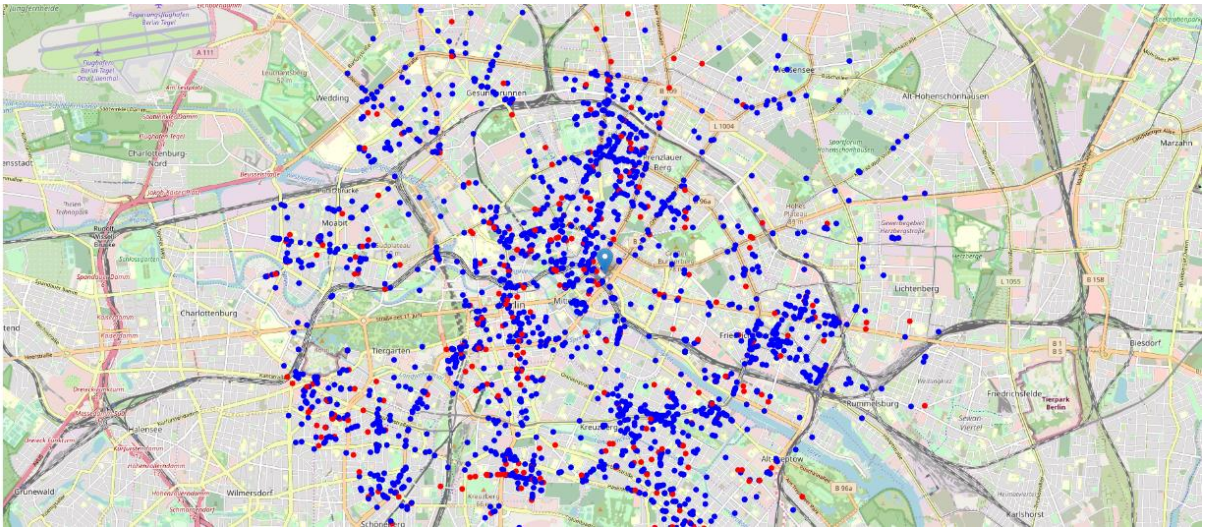
Foursquare

Now that we have our location candidates, let's 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 'Italian restaurant' category, as we need info on Italian restaurants in the neighborhood.

```
Total number of restaurants: 2122
Total number of Italian restaurants: 310
Percentage of Italian restaurants: 14.61%
Average number of restaurants in neighborhood: 5.1401098901098905
```

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



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

This concludes the data gathering phase - we're now ready to use this data for analysis to produce the report on optimal locations for a new Italian restaurant!

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 Italian restaurants. We will limit our analysis to area ~6km around city center.

In 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 Italian 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 center with low number of restaurants in general (*and* no Italian restaurants in vicinity) and focus our attention on those areas.

In third and final step we will focus on 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 radius of 250 meters**, and we want locations **without Italian restaurants in radius of 400 meters**. We will present 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

Let's perform some basic explanatory data analysis and derive some additional info from our raw data. First let's count the **number of restaurants in every area candidate**:

```
Average number of restaurants in every area with radius=300m: 5.1401098901098905
```

]:

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area
0	Ringbahnstraße 50, 12099 Berlin	52.470194	13.388575	390541.280176	5.814557e+06	5992.495307	5
1	09R/27L, 12101 Berlin	52.470314	13.397404	391141.280176	5.814557e+06	5840.376700	0
2	09R/27L, 12049 Berlin	52.470434	13.406234	391741.280176	5.814557e+06	5747.173218	0
3	Oderstraße 174, 12049 Berlin	52.470552	13.415063	392341.280176	5.814557e+06	5715.767665	0
4	Warthestraße 23A, 12051 Berlin	52.470670	13.423893	392941.280176	5.814557e+06	5747.173218	1
5	Altenbraker Str. 15, 12053 Berlin	52.470788	13.432722	393541.280176	5.814557e+06	5840.376700	7
6	Karl-Marx-Straße 212, 12055 Berlin	52.470904	13.441552	394141.280176	5.814557e+06	5992.495307	6
7	Hessenring 34, 12101 Berlin	52.474683	13.375159	389641.280176	5.815077e+06	5855.766389	0
8	Kleineweg 125, 12101 Berlin	52.474804	13.383989	390241.280176	5.815077e+06	5604.462508	0
9	09L/27R, 12101 Berlin	52.474924	13.392820	390841.280176	5.815077e+06	5408.326913	0

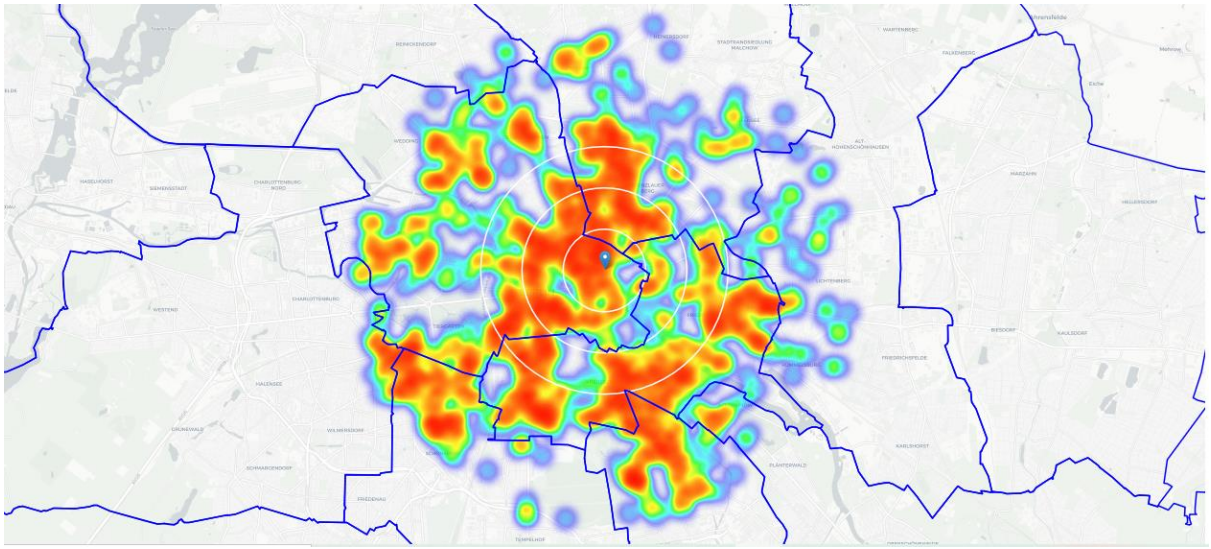
OK, now let's calculate the **distance to nearest Italian restaurant from every area candidate center** (not only those within 300m - we want distance to closest one, regardless of how distant it is).

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area	Distance to Italian restaurant
0	Ringbahnstraße 50, 12099 Berlin	52.470194	13.388575	390541.280176	5.814557e+06	5992.495307	5	264.408532
1	09R/27L, 12101 Berlin	52.470314	13.397404	391141.280176	5.814557e+06	5840.376700	0	830.999331
2	09R/27L, 12049 Berlin	52.470434	13.406234	391741.280176	5.814557e+06	5747.173218	0	1269.038823
3	Oderstraße 174, 12049 Berlin	52.470552	13.415063	392341.280176	5.814557e+06	5715.767665	0	829.067436
4	Warthestraße 23A, 12051 Berlin	52.470670	13.423893	392941.280176	5.814557e+06	5747.173218	1	575.681166
5	Altenbraker Str. 15, 12053 Berlin	52.470788	13.432722	393541.280176	5.814557e+06	5840.376700	7	293.966217
6	Karl-Marx-Straße 212, 12055 Berlin	52.470904	13.441552	394141.280176	5.814557e+06	5992.495307	6	317.390305
7	Hessenring 34, 12101 Berlin	52.474683	13.375159	389641.280176	5.815077e+06	5855.766389	0	776.047531
8	Kleineweg 125, 12101 Berlin	52.474804	13.383989	390241.280176	5.815077e+06	5604.462508	0	378.018237
9	09L/27R, 12101 Berlin	52.474924	13.392820	390841.280176	5.815077e+06	5408.326913	0	635.252552

```
Average distance to closest Italian restaurant from each area center:
491.66885234457226
```

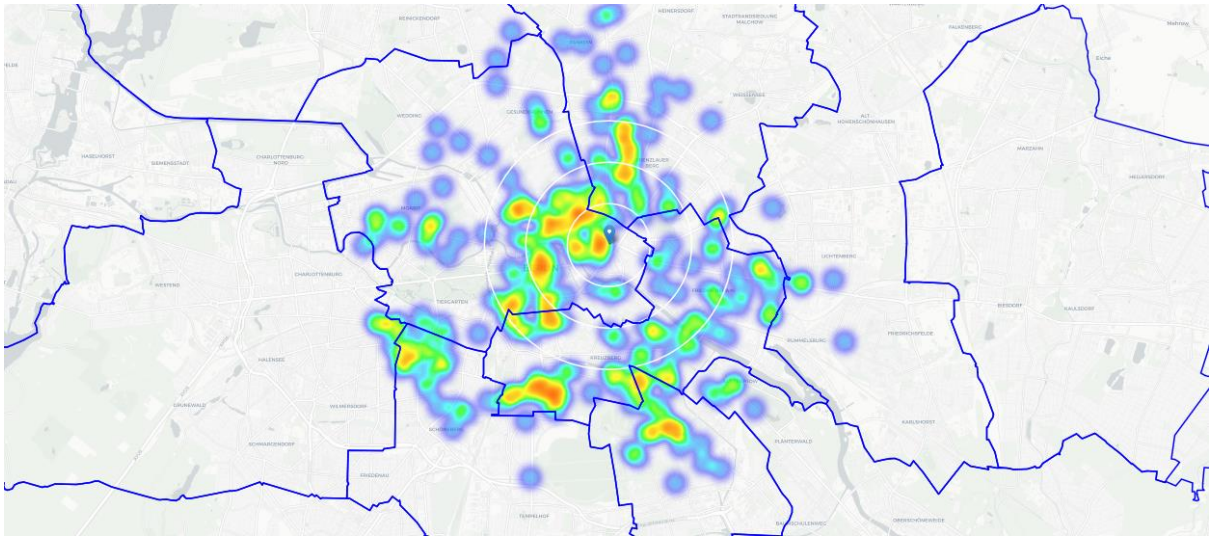
OK, so **on average Italian restaurant can be found within ~500m** from every area center candidate. That's fairly close, so we need to filter our areas carefully!

Let's crete a map showing **heatmap / density of restaurants** and try to extract some meaningfull info from that. Also, let's show **borders of Berlin boroughs** on our map and a few circles indicating distance of 1km, 2km and 3km from Alexanderplatz.



Looks like a few pockets of low restaurant density closest to city center can be found **south, south-east and east from Alexanderplatz**.

Let's create another heatmap map showing **heatmap/density of Italian restaurants** only.



This map is not so 'hot' (Italian restaurants represent a subset of ~15% of all restaurants in Berlin) but it also indicates higher density of existing Italian restaurants directly north and west from Alexanderplatz, with closest pockets of **low Italian restaurant density positioned east, south-east and south from 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 it's 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 residental and less popular with tourists).

Kreuzberg and Friedrichshain

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

"Bold and brazen, Kreuzberg's creative people, places, and spaces might challenge your paradigm." Tags: Nightlife, Artsy, Dining, Trendy, Loved by Berliners, Great Transit (airbnb.com)

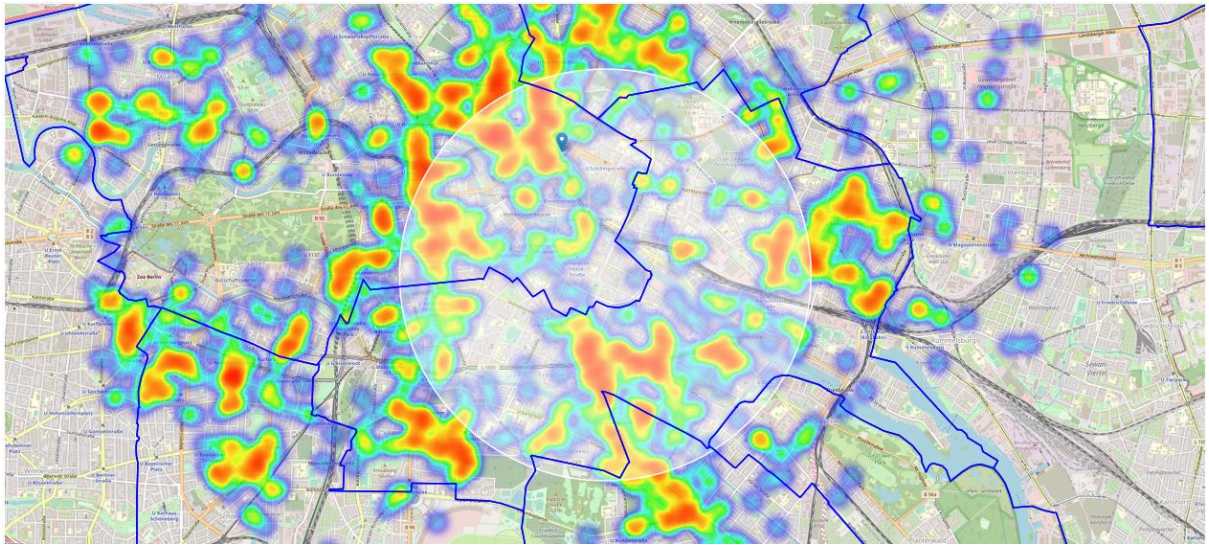
"Kreuzberg has long been revered for its diverse cultural life and as a part of Berlin where alternative lifestyles have flourished. Envisioning the glamorous yet gritty nature of Berlin often conjures up scenes from this neighbourhood, where cultures, movements and artistic flare adorn the walls of building and fills the air. Brimming with nightclubs, street food, and art galleries, Kreuzberg is the place to be for Berlin's young and trendy." (theculturetrip.com)

"Imagine an art gallery turned inside out and you'll begin to envision Friedrichshain. Single walls aren't canvases for creative works, entire buildings are canvases. This zealously expressive east Berlin neighborhood forgoes social norms" Tags: Artsy, Nightlife, Trendy, Dining, Touristy, Shopping, Great Transit, Loved by Berliners (airbnb.com)

"As anyone from Kreuzberg will tell you, this district is not just the coolest in Berlin, but the hippest location in the entire universe. Kreuzberg has long been famed for its diverse cultural life, its experimental alternative lifestyles and the powerful spell it exercises on young people from across Germany. In 2001, Kreuzberg and Friedrichshain were merged to form one administrative borough. When it comes to club culture, Friedrichshain is now out in front – with southern Friedrichshain particularly ranked as home to the highest density of clubs in the city." (visitberlin.de)

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

Let's define new, more narrow region of interest, which will include low-restaurant-count parts of Kreuzberg and Friedrichshain closest to Alexanderplatz.



Not bad - this nicely covers all the pockets of low restaurant density in Kreuzberg and Friedrichshain closest to Berlin center.

Let's also create new, more dense grid of location candidates restricted to our new region of interest (let's make our location candidates 100m appart).

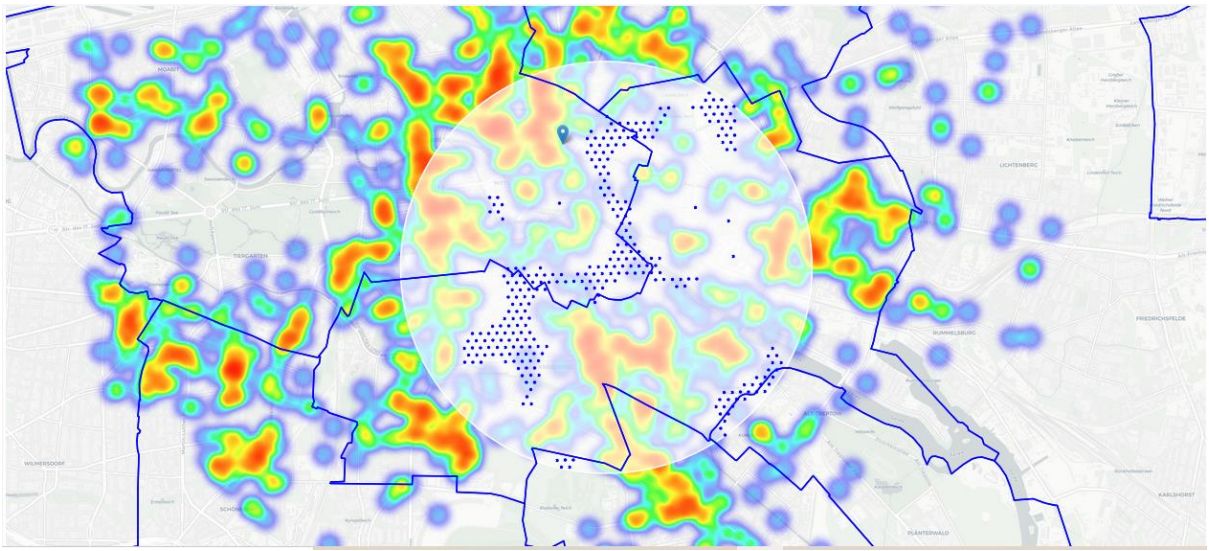
OK. Now let's calculate two most important things for each location candidate: **number of restaurants in vicinity** (we'll use radius of **250 meters**) and **distance to closest Italian restaurant**.

	Latitude	Longitude	X	Y	Restaurants nearby	Distance to Italian restaurant
0	52.486060	13.421133	392791.280176	5.816273e+06	6	158.565914
1	52.486080	13.422605	392891.280176	5.816273e+06	7	180.522171
2	52.486730	13.413009	392241.280176	5.816360e+06	0	523.935806
3	52.486750	13.414481	392341.280176	5.816360e+06	0	468.257436
4	52.486769	13.415953	392441.280176	5.816360e+06	0	369.743331
5	52.486789	13.417425	392541.280176	5.816360e+06	2	272.314591
6	52.486809	13.418897	392641.280176	5.816360e+06	4	177.764848
7	52.486829	13.420369	392741.280176	5.816360e+06	6	95.107551
8	52.486848	13.421841	392841.280176	5.816360e+06	6	80.563958
9	52.486868	13.423314	392941.280176	5.816360e+06	9	154.711526

OK. Let us now **filter** those locations: we're interested only in **locations with no more than two restaurants in radius of 250 meters**, and **no Italian restaurants in radius of 400 meters**.

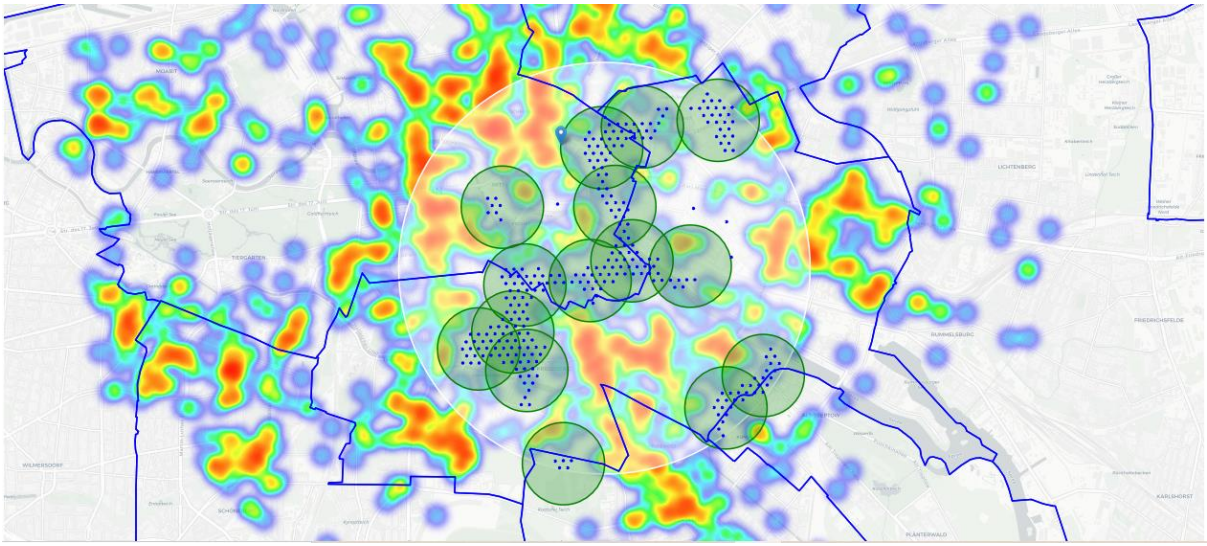
```
Locations with no more than two restaurants nearby: 773
Locations with no Italian restaurants within 400m: 362
Locations with both conditions met: 292
```

Let's see how this looks on a map.



Looking good. 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 Italian restaurant closer than 400m. Any of those locations is a potential candidate for a new Italian restaurant, at least based on nearby competition.

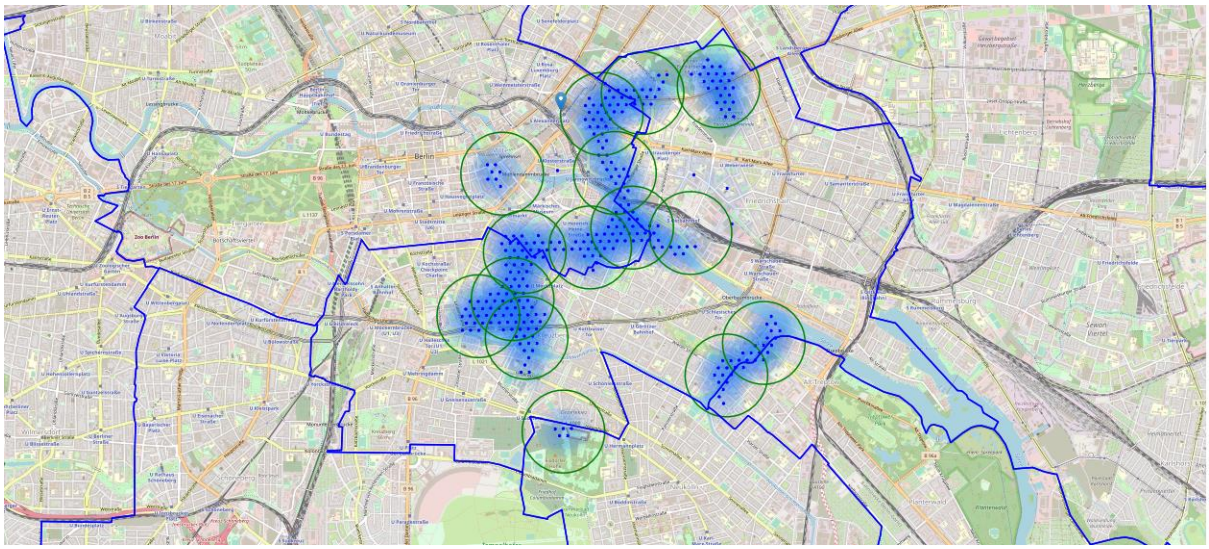
Let's now show those good locations in a form of heatmap:



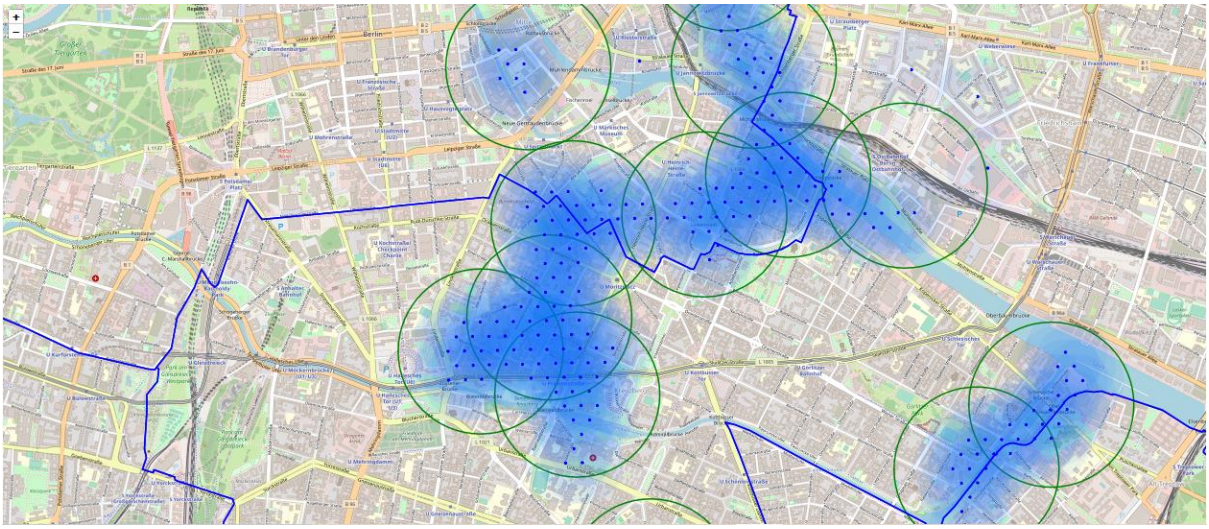
Not bad - our clusters represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates.

Addresses of those cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics.

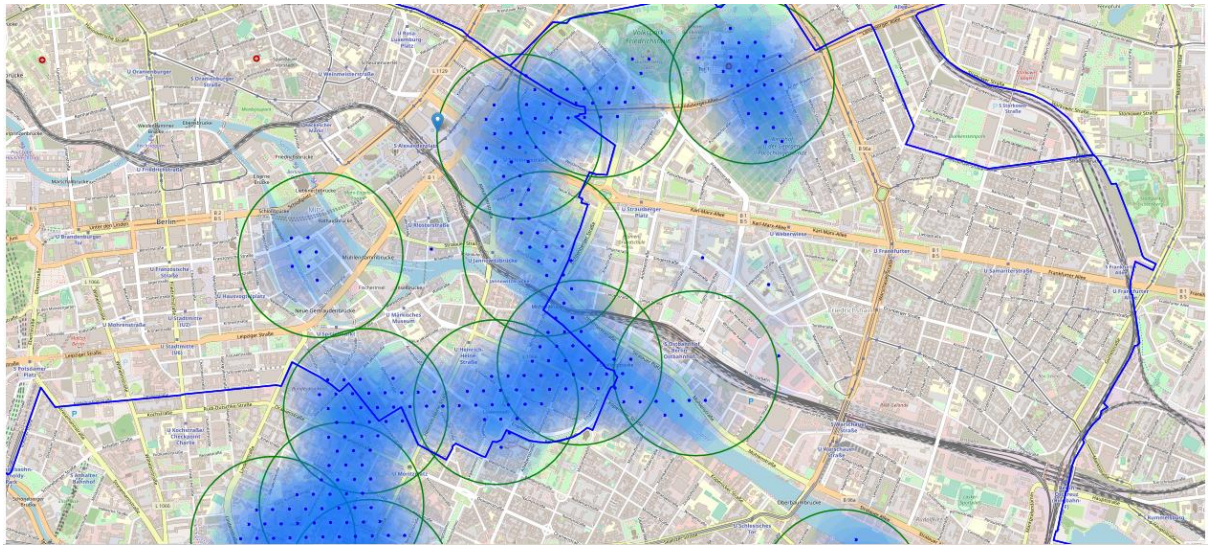
Let's see those zones on a city map without heatmap, using shaded areas to indicate our clusters:



Let's zoom in on candidate areas in **Kreuzberg**:



..and candidate areas in **Friedrichshain**:

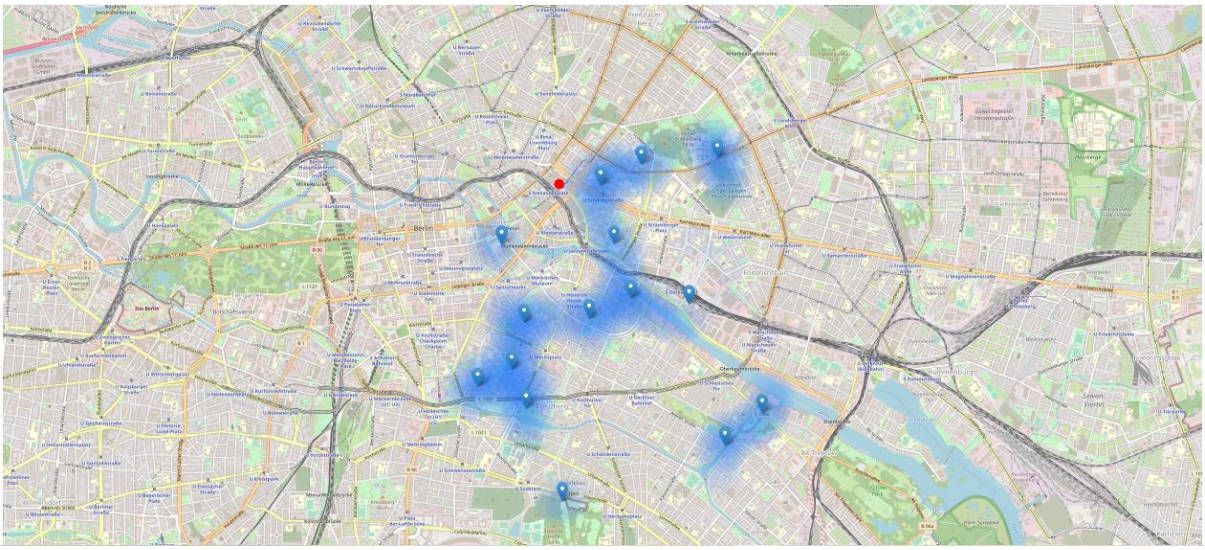


Finaly, let's **reverse geocode those candidate area centers to get the addresses** which can be presented to stakeholders.

```
============================================================
Addresses of centers of areas recommended for further analysis
============================================================

Vor dem Schlesischen Tor 2, 10997 Berlin          => 3.7km from Alexanderplatz
Stallschreiberstraße 46, 10969 Berlin             => 1.7km from Alexanderplatz
Platz der Vereinten Nationen 25, 10249 Berlin     => 1.0km from Alexanderplatz
Köpenicker Str. 40, 10179 Berlin                  => 1.6km from Alexanderplatz
Dog Park, Hasenheide 82, 10967 Berlin             => 3.9km from Alexanderplatz
Prinzenstraße 1, 10969 Berlin                     => 2.8km from Alexanderplatz
Breite Str. 1, 10178 Berlin                       => 1.0km from Alexanderplatz
Landsberger Allee 37, 10249 Berlin                => 1.9km from Alexanderplatz
Holzmarktstraße 55, 10179 Berlin                  => 1.0km from Alexanderplatz
Berolinastraße 14, 10178 Berlin                   => 0.5km from Alexanderplatz
Neuenburger Str. 16, 10969 Berlin                 => 2.7km from Alexanderplatz
Michaelkirchpl. 23, 10179 Berlin                  => 1.7km from Alexanderplatz
Lobeckstraße 62, 10969 Berlin                     => 2.3km from Alexanderplatz
Wiener Straße 58, 10999 Berlin                    => 3.8km from Alexanderplatz
Str. der Pariser Kommune 8, 10243 Berlin          => 2.2km from Alexanderplatz
```

This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of restaurants and no Italian restaurants nearby, all zones being fairly close to city center (all less than 4km from Alexanderplazt, and about half of those less than 2km from Alexanderplatz). Although zones are shown on map with a radius of ~500 meters (green circles), their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential restaurant locations. Most of the zones are located in Kreuzberg and Friedrichshain boroughs, which we have identified as interesting due to being popular with tourists, fairly close to city center and well connected by public transport.



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 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 radius of 250m and those with an Italian restaurant closer than 400m were removed.

Those location candidates were then clustered to create zones of interest which contain 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 largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Italian restaurants particularly. 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 Italian) - 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

Purpose of this project was to identify Berlin areas close to center with low number of restaurants (particularly Italian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Italian 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.

Final decission 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.