Capstone Project – Indian Restaurant in Berlin

Applied Data Science Capstone by IBM/Coursera

Table of contents

Introduction:

Problem: - This report we will try to find optional location for restaurant business in Berlin. Since there are lots of restaurants in Berlin we will try to find :-

1. Locations that is less crowded with restaurants.
2. Locations with no Indian restaurants in vicinity
3. Locations close to city center (as possible), (if first two conditions are met).

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

Target Group – Targeted of this report will be stakeholders interested in opening an Indian restaurant in Berlin, Germany.

Data

To find a solution of above problem we would be looking for the following data:-

- Number of existing restaurants in the neighborhood
- Number of Indian restaurants in the neighborhood
- Distance of Indian restaurants in the neighborhood
- Distance of neighborhood from city center

To Define our neighborhood we will regularly spaced grid of locations centered around city center.

Author- Umakant

Data Source:-

1. Centers of candidate areas will be generated through algorithm
2. Google Maps API - Using Google Map API approximate addresses of generated centers will be obtained.
3. Coordinate of Berlin center will be obtained using Google Maps API geocoding of well known Berlin location (Alexanderplatz)
4. Foursquare API - Number of restaurants, type and location in every neighborhood will be obtained using Foursquare API
5. For Boundaries Of Berlin - https://raw.githubusercontent.com/m-hoerz/berlin-shapes/master/berliner-bezirke.geojson

Coordinates of City Center: Using Google Maps geocoding API Latitude & longitude of Berlin city center found with well known address "Alexanderplatz, Berlin, Germany"

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

Creating a Grid:
A grid of cells covering our area of interest has been created which is aprox. 12x12 kilometers around Berlin city center.

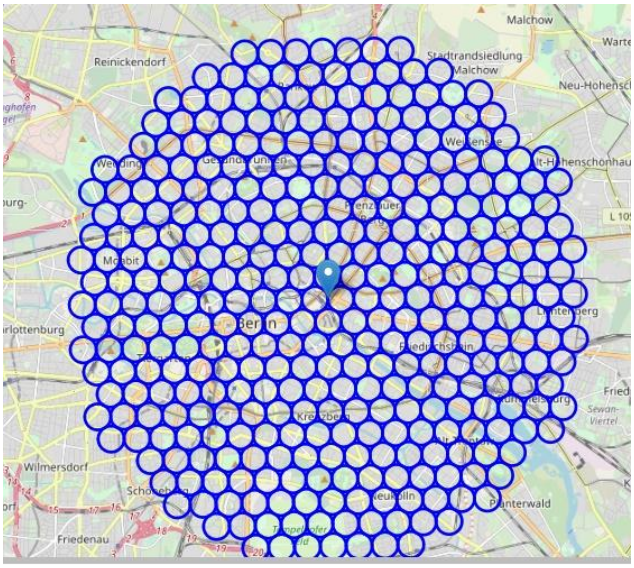
Candidate Area: A grid of candidate area, equally spaced, centered on city center and within ~6km from Alexanderplatz has been created .

Defining Neighborhood: our neighborhoods will be defined as circular areas with a radius of 300 meters, so our neighborhood centers will be 600 meters apart

Distance Calculation: 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).

H**exagonal 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**.**

Visualize City center location and candidate neighborhood centers:



Number of Candidate Neighborhood Center = 364

Approximate addresses neighborhood centers using latitude and longitude.

	Address	Latitude	Longitude	X	Y	Distance from center
0	A 100, 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 23, 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 213, 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:

Use of Foursquare API to get info on restaurants in each neighborhood

We are 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 have include in out list only venues that have 'restaurant' in category name, and we have make sure to detect and

include all the subcategories of specific 'Indian restaurant' category, as we need info on Indian restaurants in the neighborhood.

Food category: - '4d4b7105d754a06374d81259' # 'Root' category for all food-related venues

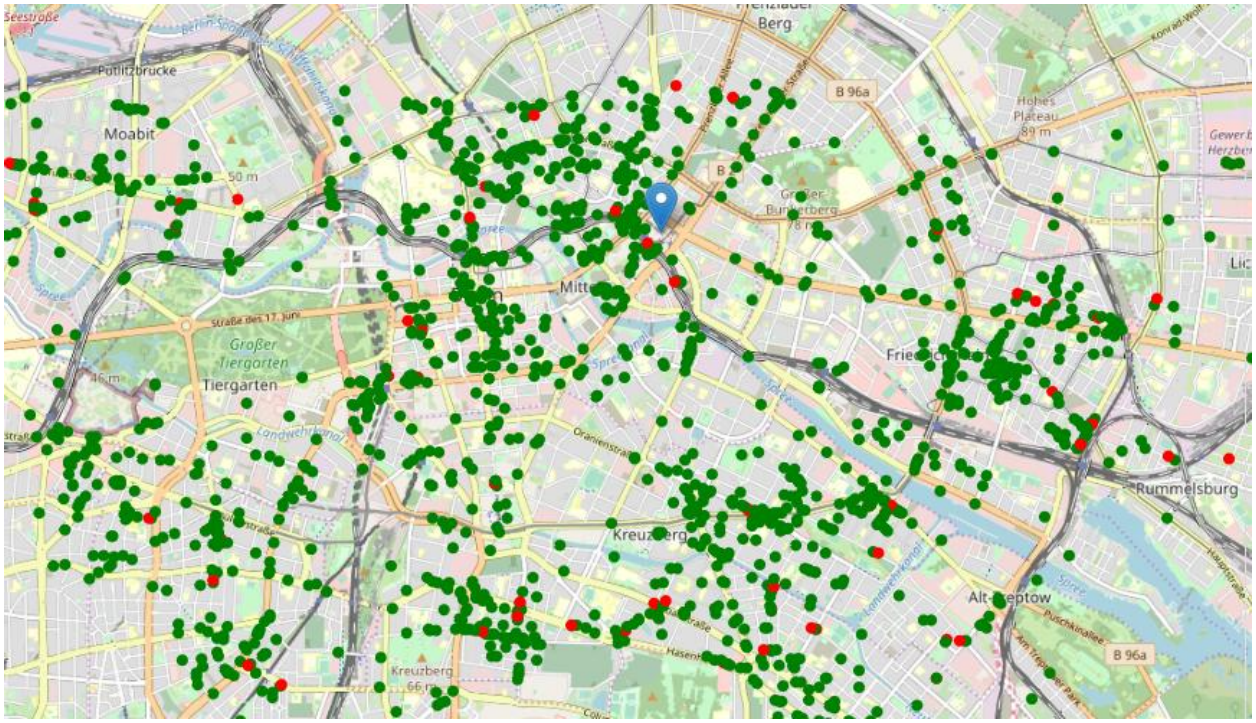
Indian restaurant categories :-

['4bf58dd8d48988d10f941735','54135bf5e4b08f3d2429dfe5','54135bf5e4b08f3d2429dff3',
'54135bf5e4b08f3d2429dff5','54135bf5e4b08f3d2429dfe2','54135bf5e4b08f3d2429dff2',
'54135bf5e4b08f3d2429dfe1','54135bf5e4b08f3d2429dfe3','54135bf5e4b08f3d2429dfe8',
'54135bf5e4b08f3d2429dfe9','54135bf5e4b08f3d2429dfe6','54135bf5e4b08f3d2429dfdf',
'54135bf5e4b08f3d2429dfe4','54135bf5e4b08f3d2429dfe7','54135bf5e4b08f3d2429dfea',
'54135bf5e4b08f3d2429dfeb','54135bf5e4b08f3d2429dfed','54135bf5e4b08f3d2429dfee',
'54135bf5e4b08f3d2429dff4','54135bf5e4b08f3d2429dfe0','54135bf5e4b08f3d2429dfdd',
'54135bf5e4b08f3d2429dff6','54135bf5e4b08f3d2429dfef','54135bf5e4b08f3d2429dff0',
'54135bf5e4b08f3d2429dff1','54135bf5e4b08f3d2429dfde','54135bf5e4b08f3d2429dfec']

Foursquare Exploration Result Summary:

Total number of restaurants	**:**	**1582**
Total number of Indian restaurants	**:**	**62**
Percentage of Indian restaurants	**:**	**3.92%**
Average no. of restaurants in neighborhood	**:**	**3.8076923076923075**

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



Author- Umakant

So now we have all the restaurants in area within few kilometers from Alexanderplatz, and we know which ones are Indian 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 Indian restaurant!

Methodology

In this project we will put our efforts on exploring the areas of Berlin that have

1. Low restaurant density,
2. Less number of Indian restaurants.
3. Our Analysis area Limit will be ~6km around city center.

Step 1: 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 Indian restaurants (according to Foursquare categorization).

Step 2: In second step our analysis will be calculation and exploration o**f** '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 Indian restaurants in vicinity) and focus our attention on those areas.

Step 3: 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 Indian 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

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

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area
0	A 100, 12099 Berlin	52.470194	13.388575	390541.280176	5.814557e+06	5992.495307	3
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 23, 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 213, 12055 Berlin	52.470904	13.441552	394141.280176	5.814557e+06	5992.495307	5
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

```
Average no of restaurants in every area with radius=300m: 3.8076923076923075
```

Calculate Distance : Calculate the distance to nearest Indian 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 Indian restaurant
0	A 100, 12099 Berlin	52.470194	13.388575	390541.280176	5.814557e+06	5992.495307	3	309.974553
1	09R/27L, 12101 Berlin	52.470314	13.397404	391141.280176	5.814557e+06	5840.376700	0	845.409396
2	09R/27L, 12049 Berlin	52.470434	13.406234	391741.280176	5.814557e+06	5747.173218	0	1432.951454
3	Oderstraße 174, 12049 Berlin	52.470552	13.415063	392341.280176	5.814557e+06	5715.767665	0	1909.911067
4	Warthestraße 23, 12051 Berlin	52.470670	13.423893	392941.280176	5.814557e+06	5747.173218	1	1516.978218
5	Altenbraker Str. 15, 12053 Berlin	52.470788	13.432722	393541.280176	5.814557e+06	5840.376700	7	1044.773416
6	Karl-Marx-Straße 213, 12055 Berlin	52.470904	13.441552	394141.280176	5.814557e+06	5992.495307	5	515.168177
7	Hessenring 34, 12101 Berlin	52.474683	13.375159	389641.280176	5.815077e+06	5855.766389	0	746.064537
8	Kleineweg 125, 12101 Berlin	52.474804	13.383989	390241.280176	5.815077e+06	5604.462508	0	308.618076
9	09L/27R, 12101 Berlin	52.474924	13.392820	390841.280176	5.815077e+06	5408.326913	0	594.876408

Average between closest Indian restaurant & area centers: 1409.08

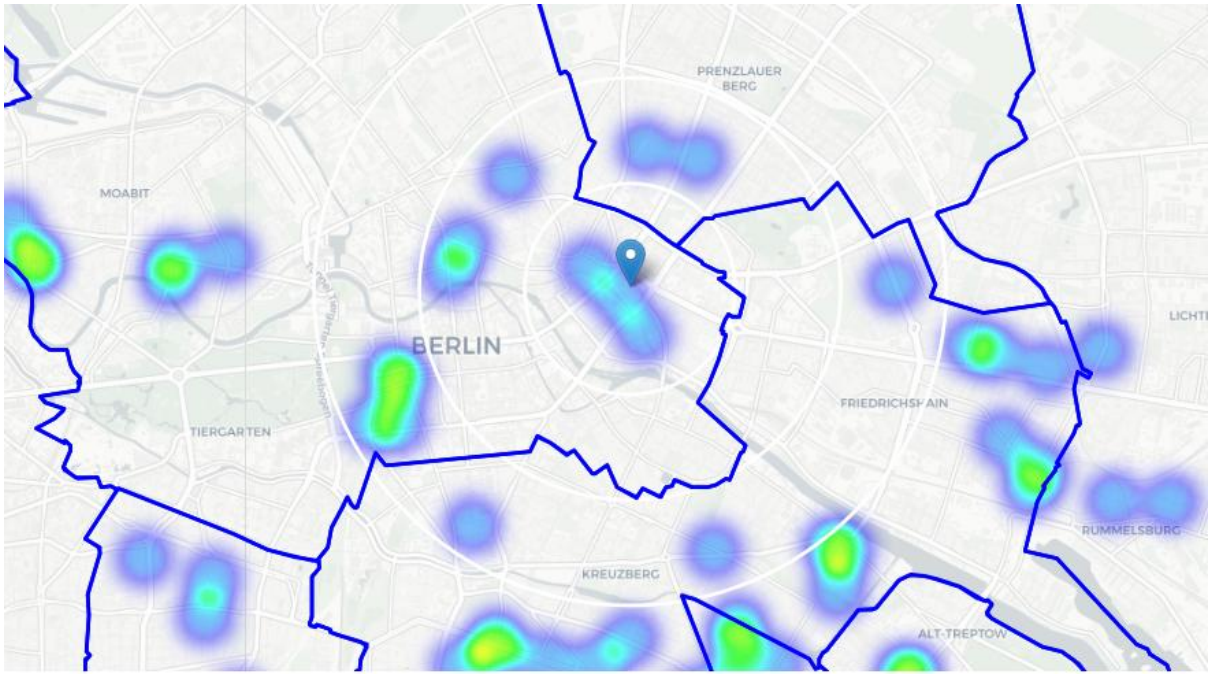
Average Indian restaurant can be found within ~1500m from every area center candidate. That's fairly close

Author- Umakant

Visualization of Restaurants- Heatmap / density of restaurants with Borders of Berlin boroughs on our map and a few circles indicating distance of 1km, 2km and 3km from Alexanderplatz.

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

Visualization of Indian Restaurants: Heatmap showing the density of Indian restaurants only.



above map is not so 'hot' (Indian restaurants represent a subset of ~4% of all restaurants in Berlin) but it also indicates higher density of existing Indian restaurants directly north and west from Alexanderplatz, with closest pockets of low Indian 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 residential and less popular with tourists.

Berlin's Popular Tourist Area - 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.

Kreuzberg - *It 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 neighborhood, where cultures, movements and artistic flare adorn the walls of building and fills the air. Brimming with nightclubs, street food, and art galleries,*

Friedrichshain*: "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)

Justification for Further Analysis: Selection of Kreuzberg and Friedrichshain for further analysis is based on three main aspects -

1. Popular with tourists,

2. Alternative and bohemian but booming and trendy,

3. Relatively close to city center 4. Well connected;

So these 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).

Candidate neighborhood centers generated= 2261

Now let's calculate two most important things for each location candidate:

1. Number of restaurants in vicinity (Radius of 250 meters)

2. Distance to closest Indian restaurant.

	Latitude	Longitude	X	Y	Restaurants nearby	Distance to Indian restaurant
0	52.486060	13.421133	392791.280176	5.816273e+06	8	136.415045
1	52.486080	13.422605	392891.280176	5.816273e+06	10	141.008943
2	52.486730	13.413009	392241.280176	5.816360e+06	0	455.582507
3	52.486750	13.414481	392341.280176	5.816360e+06	0	495.466614
4	52.486769	13.415953	392441.280176	5.816360e+06	0	395.931337
5	52.486789	13.417425	392541.280176	5.816360e+06	2	296.708747
6	52.486809	13.418897	392641.280176	5.816360e+06	6	198.268853
7	52.486829	13.420369	392741.280176	5.816360e+06	8	102.883408
8	52.486848	13.421841	392841.280176	5.816360e+06	9	43.121375
9	52.486868	13.423314	392941.280176	5.816360e+06	13	114.603274

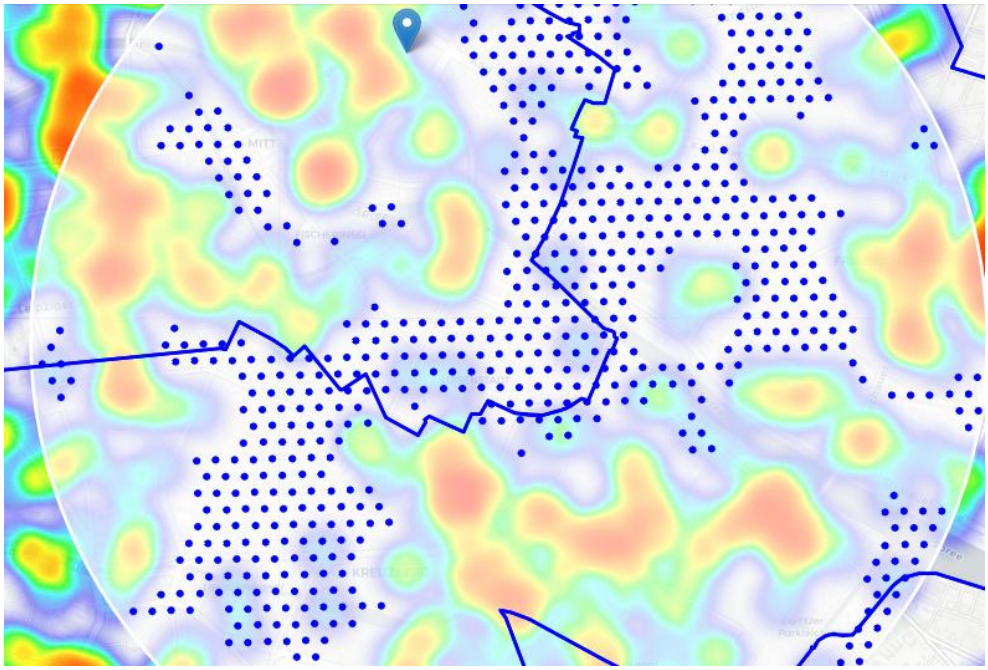
Filter those locations: we are interested only in locations with

1. No more than two restaurants in radius of 250 meters,
2. No Indian restaurants in radius of 400 meters.

Results of Filtration:

1. Locations with no more than two restaurants nearby : 802
2. Locations with no Indian restaurants within 400m : 1581
3. Locations with both conditions met : 707

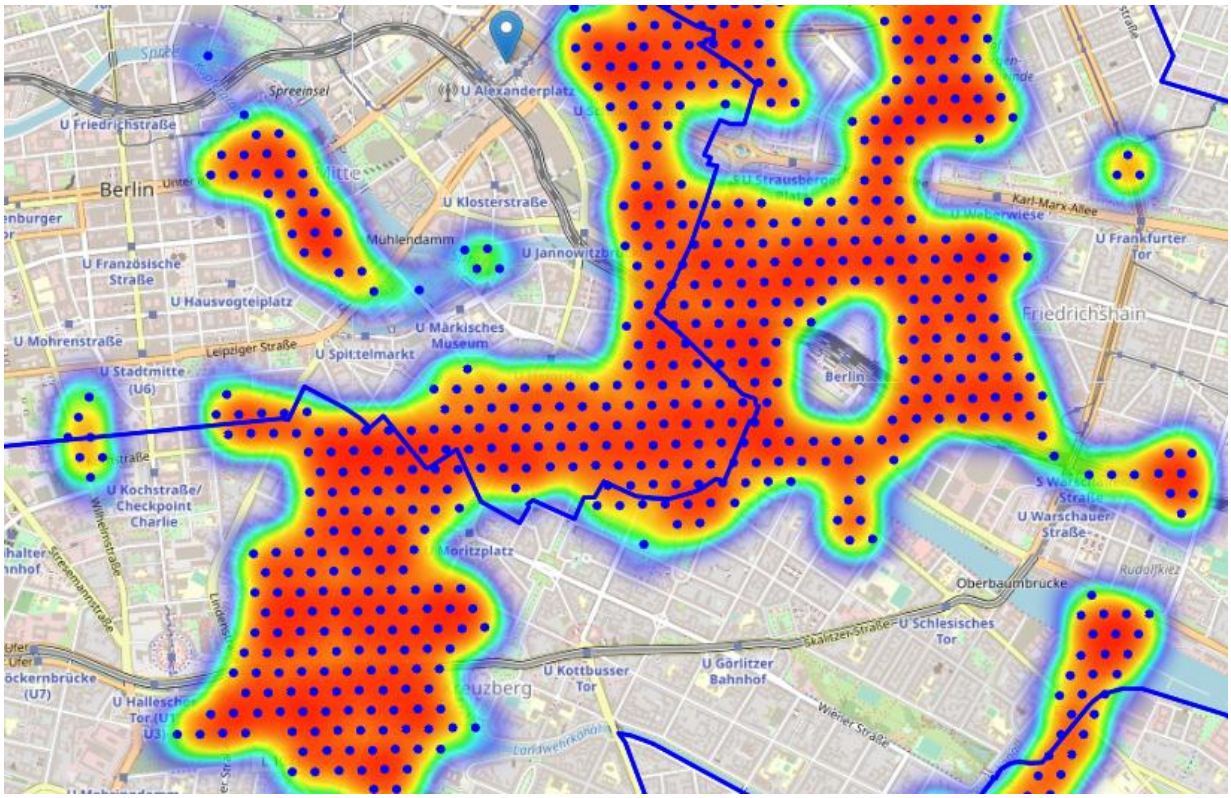
So Let's see how these results will looks on a map.



Author- Umakant

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

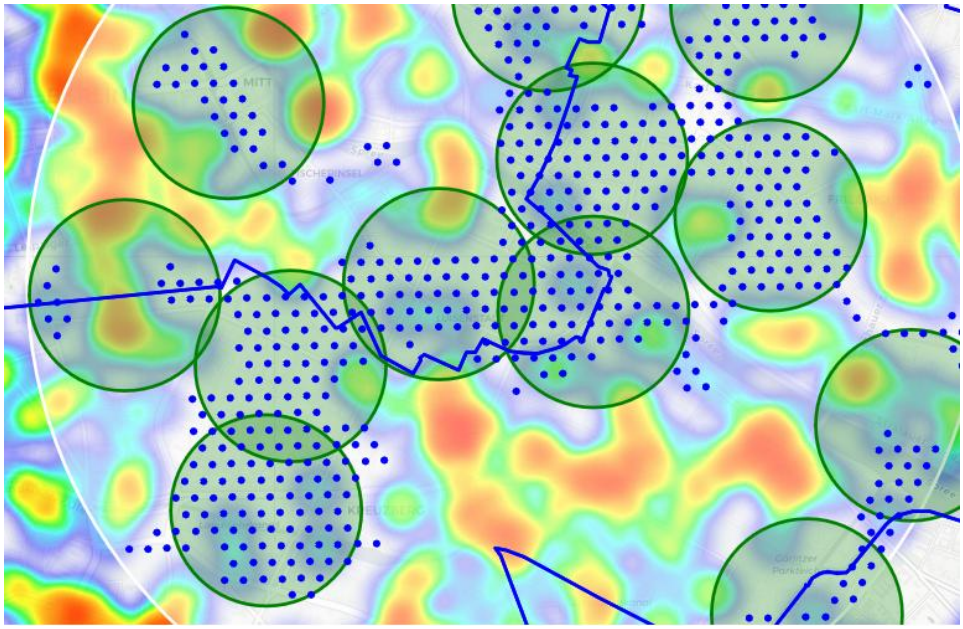
Visualization of good locations in a form of heatmap:



What we have now is a clear indication of zones with low number of restaurants in vicinity, and *no* Indian restaurants at all nearby.

Clustering: Cluster these locations to create **centers of zones** containing good locations. Those zones, their centers and addresses will be the final result of our analysis.

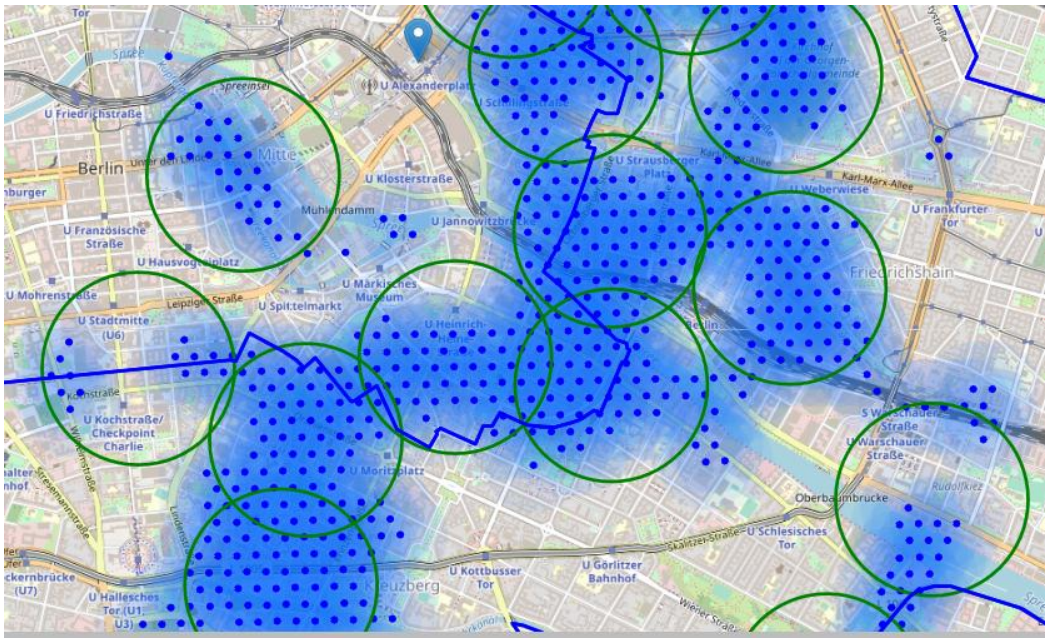
Author- Umakant



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.

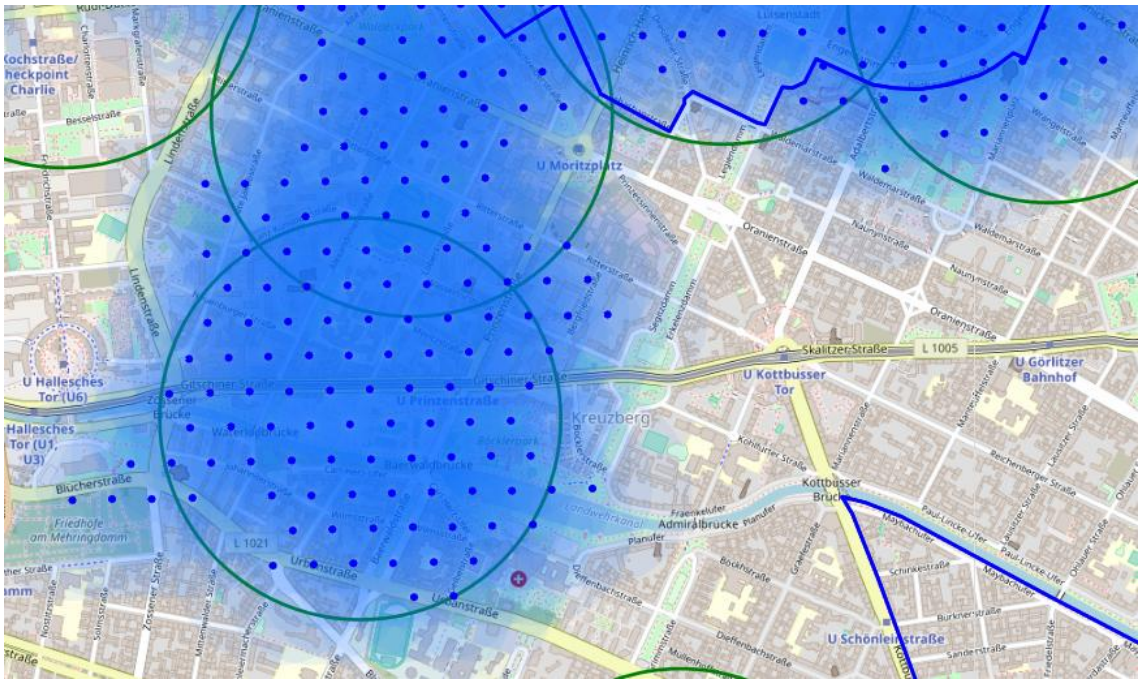
Addresses of those cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics.

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

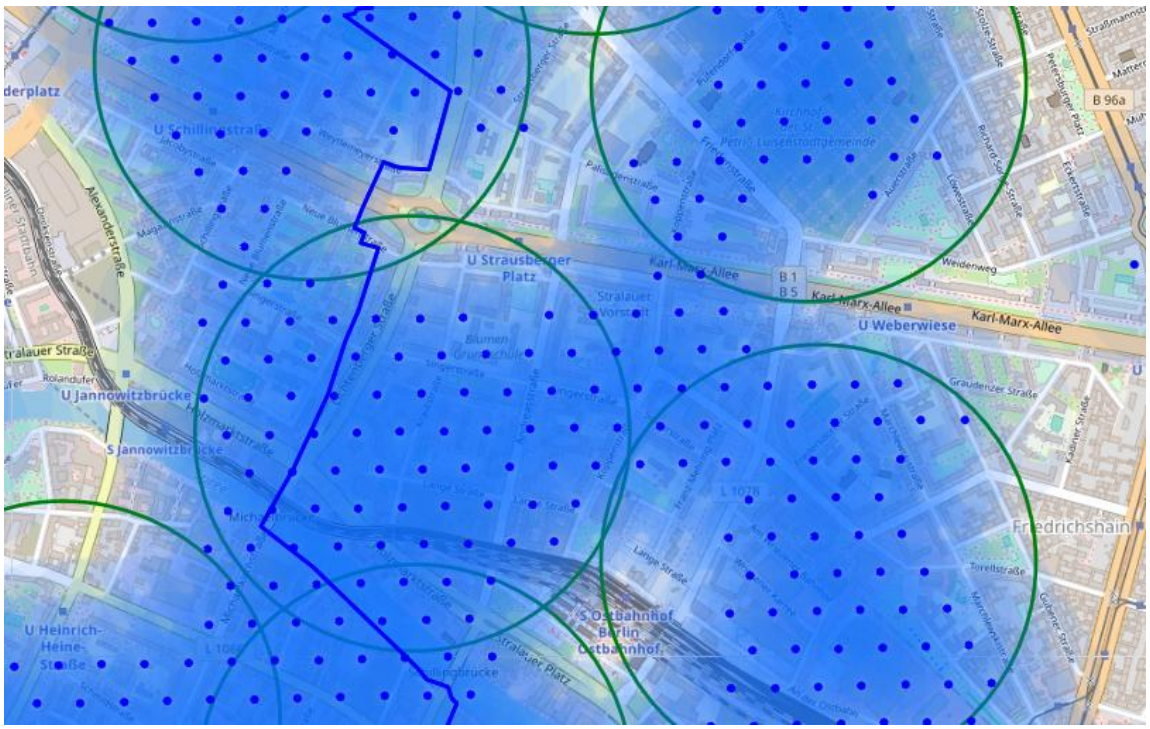


Author- Umakant

Kreuzberg - Coordinates(52.498972, 13.409591) Source - Google Maps API geocoding :
Let's zoom in on candidate areas in



Friedrichshain - Coordinates(52.516347, 13.428403) Source - Google Maps API :
Let's zoom in on candidate areas in



Author- Umakant

Finally: Use reverse geocode for those candidate area centers to get the addresses which can be presented to stakeholders.

Recommendations:

==========

Addresses of centers of areas recommended for further analysis

==========

Unnamed Road, 10249 Berlin	=> 1.5km from Alexanderplatz
Alexandrinenstraße 93, 10969 Berlin	=> 2.0km from Alexanderplatz
On Wriezener Station 1, 10243 Berlin	=> 2.3km from Alexanderplatz
Gitschiner Str. 21-22, 10969 Berlin	=> 2.8km from Alexanderplatz
Berolinastraße 7C, 10178 Berlin	=> 0.8km from Alexanderplatz
Karl-Kunger-Straße 1, 12435 Berlin	=> 3.9km from Alexanderplatz
Schloßpl. 3, 10178 Berlin	=> 1.1km from Alexanderplatz
Annenstraße 3, 10179 Berlin	=> 1.5km from Alexanderplatz
Stralauer Allee 35A, 10245 Berlin	=> 3.5km from Alexanderplatz
Friedrichstraße 45, 10969 Berlin	=> 2.1km from Alexanderplatz
Köpenicker Str. 143, 10997 Berlin	=> 1.9km from Alexanderplatz
Krautstraße 24, 10243 Berlin	=> 1.3km from Alexanderplatz
Diestelmeyerstraße 4A, 10249 Berlin	=> 1.9km from Alexanderplatz
Hasenheide 81, 10967 Berlin	=> 3.8km from Alexanderplatz
Mendelssohnstraße 29, 10405 Berlin	=> 0.7km from Alexanderplatz

This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of restaurants and no Indian 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.

Author- Umakant

Results and Discussion

Through our analysis we get a great number of restaurants in Berlin (~**1582** in our initial area of interest which was 12x12km around Alexanderplatz), there are also 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.

We found another interesting borough Prenzlauer Berg which is north-east from Alexanderplatz and it is largely a residential area.

So our attention was focused on Kreuzberg and Friedrichshain which offer a combination of -

1. Popularity among tourists,
2. Closeness to city center,
3. Strong socio-economic dynamics
4. 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 Indian 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 Indian 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 Indian) - 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

Objective of this project was to assist stakeholders in identifying areas close to center of Berlin with low number of restaurants (particularly Indian restaurants) for opening a new Indian restaurant.

1. Using Foursquare data we calculated density of restaurants to identify boroughs that justify the further analysis of two major tourist locations (Kreuzberg and Friedrichshain), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants.

2. 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.

3. Final decision on location of restaurant will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone.

4. Additional factors: Stakeholder can consider additional factors like

- Attractiveness of each location (proximity to park or water)
- Levels of noise / proximity to major roads
- Real estate availability
- Real estate Prices
- Social and economic dynamics of every neighborhood etc.

Author- Umakant