

**Report
Applied Data Science Capstone
Capstone Project - The Battle of Neighborhoods**

Exploring Restaurants in Berlin

Author: André Schlinke

Date: Wednesday 6th March, 2019

Abstract

This report explores the distribution of different types of restaurants in the center of Berlin. It specifically tries to answer the question where restaurants are located which offer German cuisine. Therefore the concentration of German restaurants are examined for the different boroughs of Berlin. Also, as a last objective, the clustering algorithm k-means is being used to detect culinary hot-spots for opening potential new German restaurants.

Contents

1	Introduction	1
1.1	Background	1
1.2	Problem	1
1.3	Interest	2
2	Data	3
2.1	Data Sources	3
2.2	Data Cleaning	3
3	Exploratory Data	5
4	Clustering	10
5	Conclusion	13
6	Outlook	14

1 Introduction

1.1 Background

As a enthusiast of good, tasty food, I like to go out with my friends to new restaurants and try new dishes. Luckily this is pretty easy as a local of Berlin, the German capital. The city has a rich culinary scene to offer. Here one can find culinary styles from all over the world. And eating delicious, and exotic meals aren't solely for the ones with a bigger pocket money. The competition is fierce for the hungry bellies walking by.

I live in Wedding, a neighborhood which is part of the borough called Mitte (center). As a neighborhood for the working class, a lot of Turkish and Arabic families live here. It therefore makes sense that mostly doner and falafel restaurants are placed here. In general, Berlin has also a strong Vietnamese community. Immigrants fleeing from the war back then. That is why you can find very good Vietnamese restaurants all over Berlin. It is also the Vietnamese restaurant called Monsieur Vuong which is my favorite restaurant of all time. I even bought their cookbook. Whenever I have someone visiting me from out of town I use the opportunity to go to Monsieur Vuong and order some delicious meals.

1.2 Problem

Berlin has lots of tourists visiting the city every year. Last year alone Berlin had over 10 million guests. It happens very quickly to bump into one of these tourists and have a little chat about the city. In these occasions I like to recommend my favorite spots of Berlin. However, whenever I get asked if I know a place where one could eat traditional German food, I shake my head and shrug my shoulders. To be honest, I have no idea. I barely see any German restaurants and I start to ask myself - Where are all those German restaurants in the German capital?

To highlight the problem even further - I love to watch the documentary series at Netflix called 'Chef's Table' portraying different chefs from all over the world. In one episode the chef Tim Raue is being portrayed and even he, a local of Berlin, is famous for making extraordinary Asian fusion cuisine.

Therefore, the question I want to investigate in this notebook is if there really is a lack of German restaurants in Berlin? And if not, where are all those restaurants located?

1.3 Interest

This topic is not only of interest for me or my friends but also for the several million guests visiting the city every year. As tourist from outside of Germany probably want to try German cuisine, this study could be a good advisor of where one would probably find a possible restaurant for lunch or dinner. Also future restaurateurs will get three suggestions at the end where to open a potential new restaurant.

2 Data

To tackle the above depicted problem we need a list of restaurants, their type of cuisine as well as their latitude and longitude. These informations can be gathered by the **Foursquare API**. However, to explore the different venues of a specific area, a radius as well as latitude and longitude values need to be provided by the user. Therefore we take the latitude and longitude of the different postal codes which we get by the **pgeocode** library. A hand-picked selection of postal codes, their corresponding neighborhoods, as well as their boroughs are provided by a **CSV-file** which contains data gained by this wikipedia article [here](#).

2.1 Data Sources

In this study a hand-picked selection of the core neighborhoods of Berlin are being analyzed and a listing can be found in the Table 2.1. The data of the different neighborhoods and boroughs were taken from a wikipedia article which you can find [here](#).

The postal codes of the neighborhoods, [Mitte](#), [Moabit](#), [Hansaviertel](#), [Tiergarten](#), [Wedding](#), [Gesundbrunnen](#), [Friedrichshain](#), [Kreuzberg](#), [Prenzlauer Berg](#), [Charlottenburg](#), [Wilmersdorf](#), [Westend](#), [Schöneberg](#), and [Neukölln](#) were gained by their according wikipedia article. The data are provided by a CSV-file which contain the different postal codes, neighborhoods and boroughs.

In a next step we used the library **pgeocode** to extract the **latitude** and **longitude** per each postal code. This new data was then combined with the data of the postal codes, neighborhoods and boroughs. With this new data it now possible to explore the different venues nearby the different postal code areas with the **Foursquare API**.

2.2 Data Cleaning

In a first step the postal codes which are part of multiple neighborhoods were summarized to one postal code sample with the several neighborhoods being separated by comma. Therefore we minimize the requests for the **Foursquare API**.

The exploration of the venues within a given radius returns a number of different categories of venues. As we are only interested in restaurants for our problem, we only consider categories which contain the string 'Restaurant' in their category. It is clear,

2.2 Data Cleaning

Table 2.1: Neighborhoods and boroughs examined in this study

Neighborhood	Borough
Mitte	Mitte
Moabit	Mitte
Hansaviertel	Mitte
Tiergarten	Mitte
Wedding	Mitte
Gesundbrunnen	Mitte
Friedrichshain	Friedrichshain-Kreuzberg
Kreuzberg	Friedrichshain-Kreuzberg
Prenzlauer Berg	Pankow
Charlottenburg	Charlottenburg-Wilmersdorf
Wilmersdorf	Charlottenburg-Wilmersdorf
Westend	Charlottenburg-Wilmersdorf
Schöneberg	Tempelhof-Schöneberg
Neukölln	Neukölln

that we thereby neglect related venues which belong to categories such as 'Fast Food' or 'Food & Drink'. However, as these types do not provide further details on the origin of the respective cuisine style, we consider this neglect as reasonable. Therefore we only focus on restaurants as they generally provide also the local origin of their cuisine within their category name.

Also, in order to not consider restaurants multiple times, we drop duplicate listings of restaurants.

3 Exploratory Data

In this chapter we explore the data we gained so far. In a first step we take a look at the location of the postal codes as well as of all the restaurants found. The locations of the postal codes are depicted in Figure 3.1 and the locations of all the restaurants in Figure 3.2.

Figure 3.2 depicts a very chaotic distribution of the restaurants. To gain a better sense for the area density Figure 3.3 shows the distribution of the standardized latitude and longitude values for the whole dataset. Here you can see that there are more restaurants in the east of the city than in the west in general. Also you can see that there are generally more restaurants in the south of the city and fewer in the north. Seems like Wedding, the area where I live, is a pretty bad spot to go out and eat.

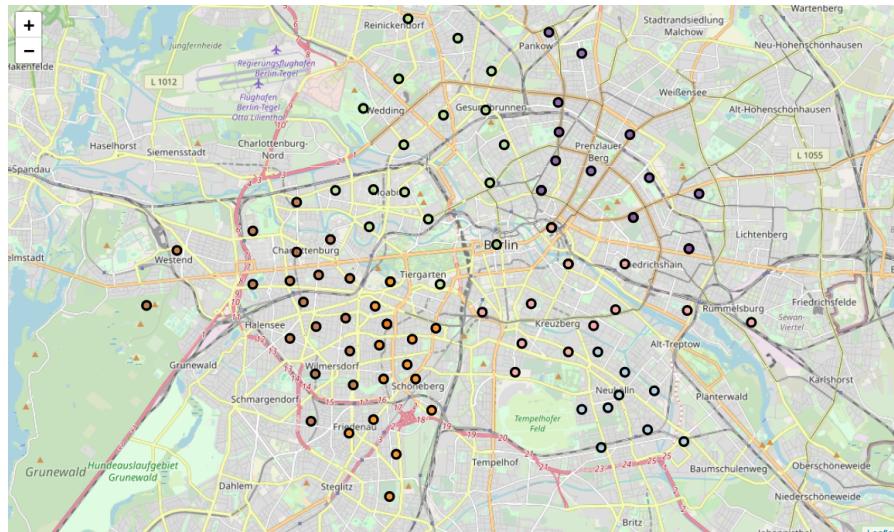


Figure 3.1: Location of all postal codes examined

Now we take a look at the total number of restaurants per category for the whole city. As it turns out, **German restaurants** are on the **third place** of number of restaurants in Berlin with a total of 88 restaurants which you can see in Figure 3.4. To be honest, I did not expect this. There are **only more Vietnamese and Italian restaurants** with Italian being the most type of restaurants in Berlin with a total number of 140 restaurants found. This confirms more or less the impression I got from the restaurants

2.2 Data Cleaning

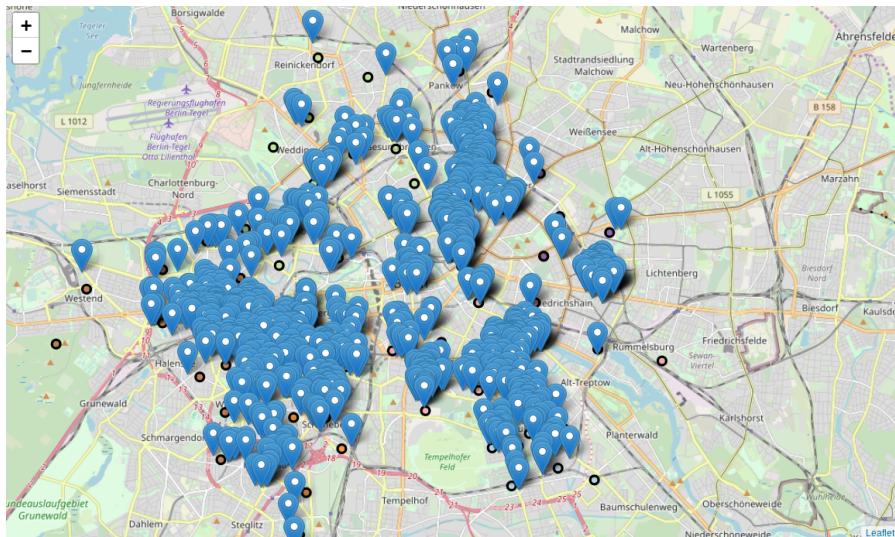


Figure 3.2: Locations of all restaurants found in Berlin

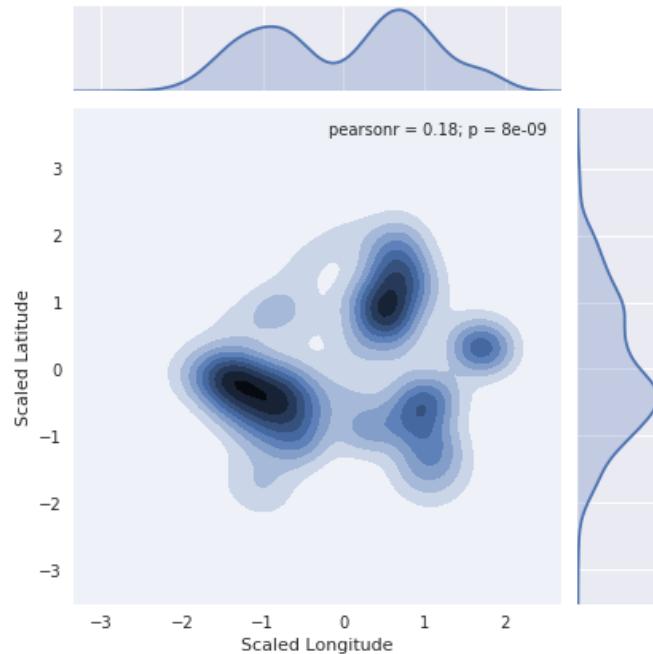


Figure 3.3: Scaled distribution of all restaurants found in Berlin

in Berlin. It is a common fact that Germany is a pasta loving country.

2.2 Data Cleaning

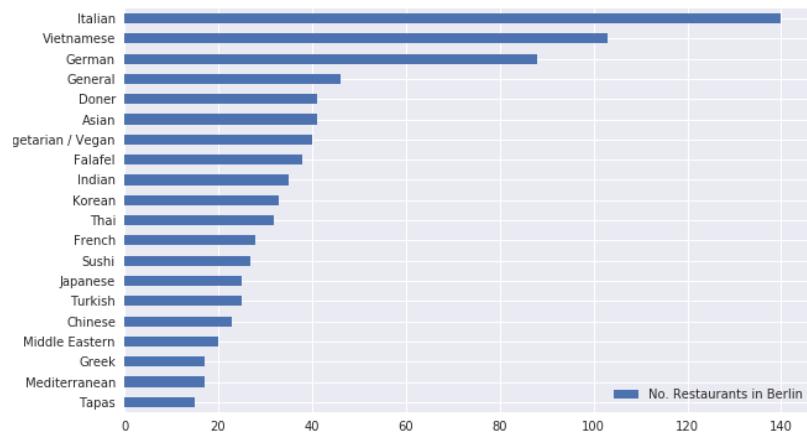


Figure 3.4: Excerpt of total no. of restaurants per category (top 20)

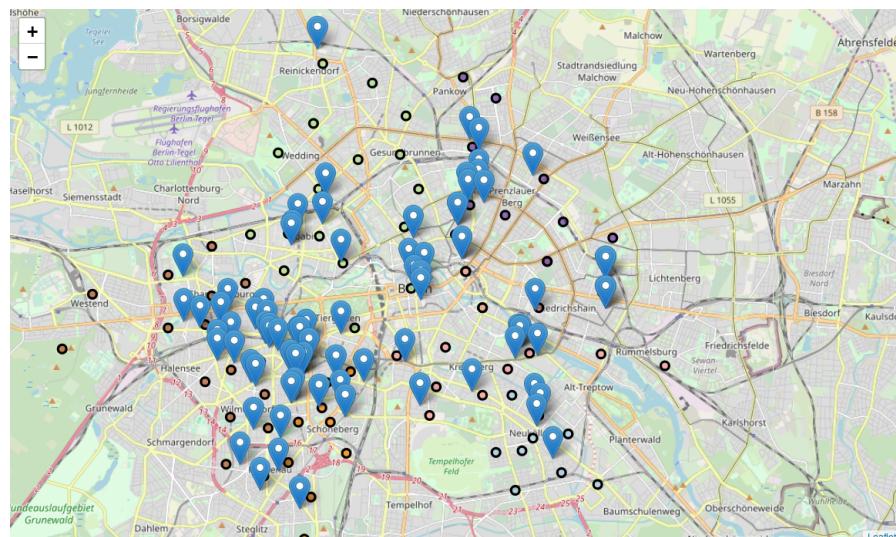


Figure 3.5: Locations of all German restaurants found in Berlin

To get an overview of the general locations of only German restaurants in Berlin, we take a look at Figure 3.5. It seems like there are a lot of German restaurants in the borough Charlottenburg-Wilmersdorf. To confirm this, we take a look at the Figure 3.6.

The Figure 3.6 clearly indicates a concentration of German restaurants in the south west of Berlin. We analyze this further by listing the number of German restaurants per

2.2 Data Cleaning

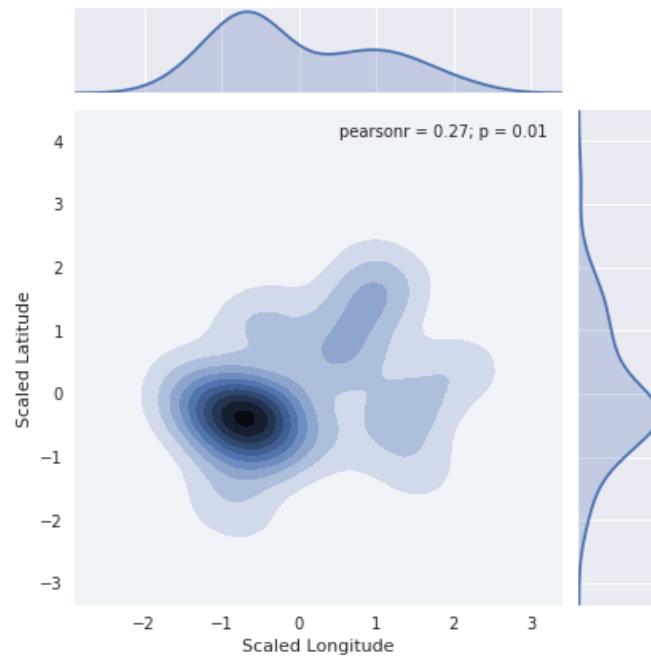


Figure 3.6: Scaled distribution of all German restaurants found in Berlin

borough.

Table 3.1: Percentage of German restaurants per borough

Borough	No. of Restaurants	No. of German Restaurants	Percentage of German Restaurants per Borough
Tempelhof-Schöneberg	149	18	12.1%
Charlottenburg-Wilmersdorf	225	26	11.6%
Mitte	170	17	10.0%
Friedrichshain-Kreuzberg	189	12	6.3%
Pankow	188	11	5.9%
Neukölln	88	4	4.5%

As expected, Charlottenburg-Wilmersdorf is the borough with the most German restaurants. Interestingly, Neukölln, a borough similar to wedding, is a borough for the working class with also lots of Arabic and Turkish families. This borough is very famous for its rich nightlife as well as for its exotic culinary scene, whereas Charlottenburg-Wilmersdorf is more inhabited by the middle and upper class. It is interesting that this fact shows off so clearly by the type of cuisine located in each borough. To back this up we investigate the percentage of German restaurant for the different boroughs.

Neukölln has the lowest rate of German restaurants with only 5% whereas Tempelhof-

2.2 Data Cleaning

Schöneberg has the highest concentration with almost 13% followed by Charlottenburg-Wilmersdorf with 11%.

However, Prenzlauer Berg in Pankow is a neighborhood known to have a lot of German inhabitants from the south of Germany. I would consider it to be a neighborhood for the solid middle class. Nevertheless it has the second least concentration of German restaurants. But Prenzlauer Berg is also known to have a lot of vegetarian and vegan restaurants. It would be interesting to see which restaurant category is the most common in Prenzlauer Berg. But this is out of scope for this report.

4 Clustering

In this chapter we use the clustering algorithm k-means to determine 3 hot-spots of non-German restaurants, so that motivated newcomer restaurateurs of the German cuisine can have a recommendation of where to open their new restaurant. We, therefore, only need the latitude and longitude values of the non-German restaurants to train the algorithm. The out coming three cluster centers are then considered to be our recommendations.

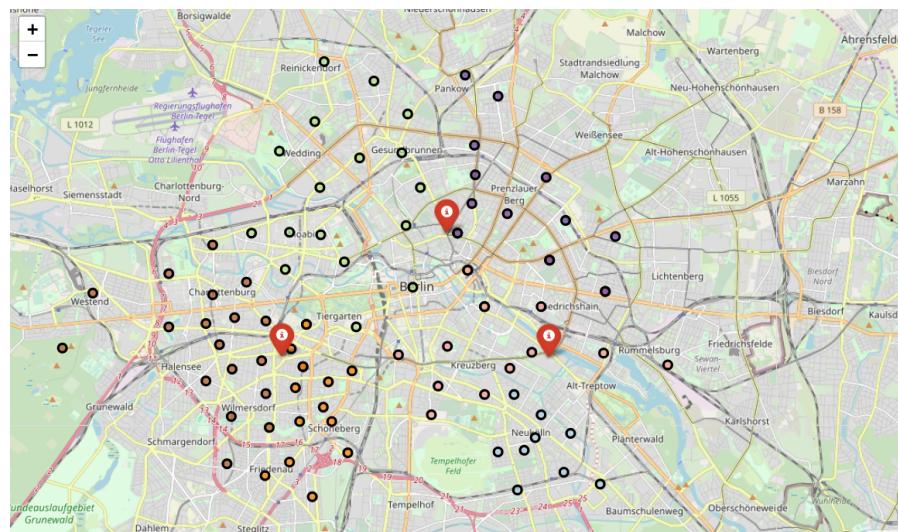


Figure 4.1: Location of the three recommendations

In Figure 4.1 you can see the results of clustering method. I am pretty impressed by the locations of the cluster centers, as two of them are my favorite spots to go out to eat delicious food.

- The first recommendation is located at the metro station **U Schlesisches Tor**, see Figure 4.2. It's in Kreuzberg and one of the go to areas if you want to party. There are a lot of nightclubs playing techno music and also some very good spots to get something to eat. My recommendation there would be the **Burgermeister** which is located in an old public toilet. It sounds disgusting but it is delicious. Check it out.
- The second cluster center and therefore our recommendation would be the **Rosen-**

2.2 Data Cleaning

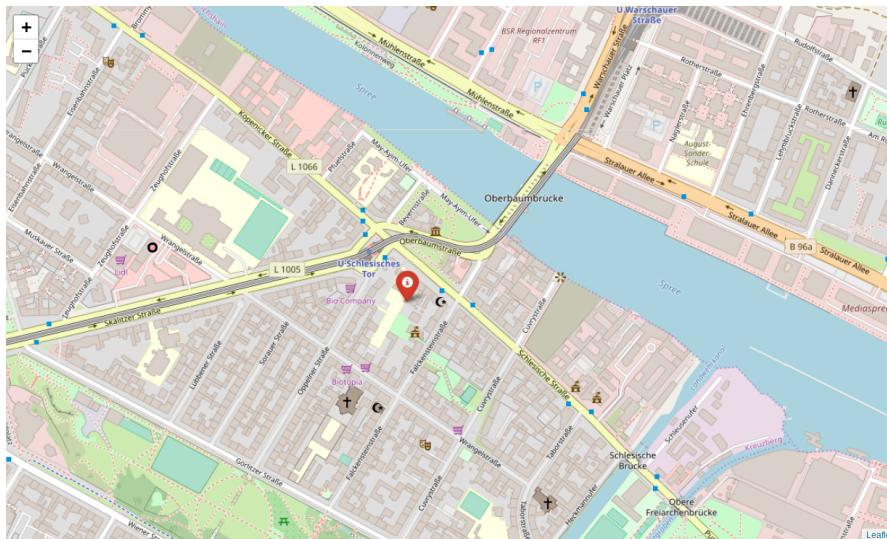


Figure 4.2: First recommendation

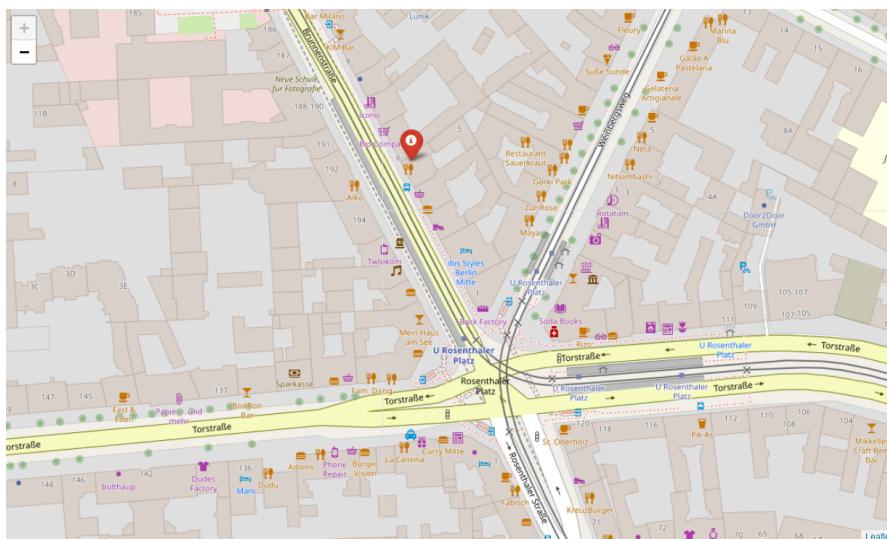


Figure 4.3: Second recommendation

thaler Platz in Prenzlauer Berg, see Figure 4.2. This makes total sense since I would consider this the number one spot of delicious food in Berlin. Here a number of very good Asian food restaurants. This is definitely a culinary hot-spot and it is funny that the clustering algorithm is so precisely in detecting it. Just to give you a view recommendations: [Transit](#), [District Mot](#), or [Shiso Burger](#) are definitely

2.2 Data Cleaning

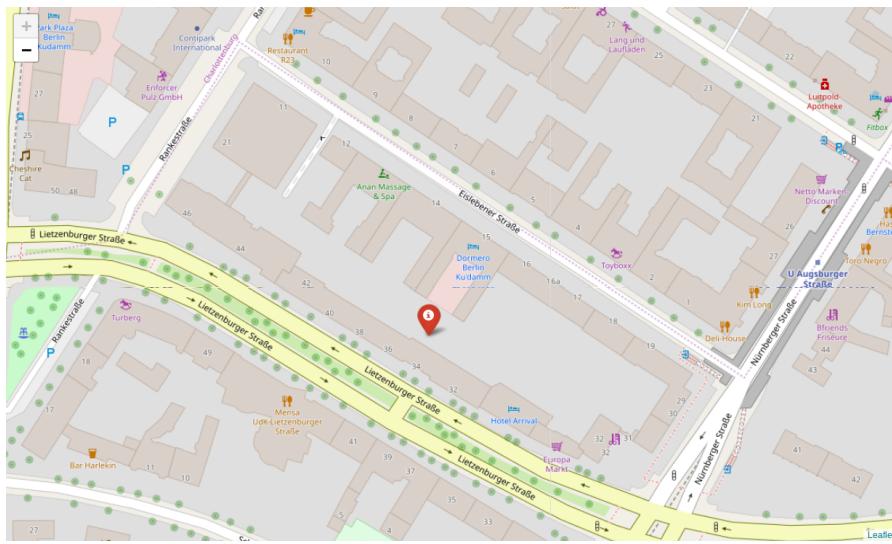


Figure 4.4: Third recommendation

places one should check out there. But what is missing there is definitely a German restaurant. So this would be a very good spot.

- The last recommendation would be in the **Lietzenburger Straße** at the border of Schöneberg and Wilmersdorf, see Figure 4.2. To be honest, I barely go out to eat in this area of the city so I have no recommendation of places to eat nearby. However, I should definitely check it out. As we discussed earlier, the **competition** in this area for German restaurants is higher, nevertheless German restaurants seem to be pretty successful there. So this spot makes also total sense.

5 Conclusion

We have successfully analyzed the amount of German restaurants in the center of Berlin, as well as their exact locations. We have detected a concentration of German restaurants in the boroughs Charlottenburg-Wilmersdorf and Tempelhof-Schöneberg. We also have determined the favorite types of restaurants in Berlin with Italian being the most common type followed by Vietnamese and then German restaurants. The finding of German restaurants being at the third place of amount of restaurants in Berlin does not support our assumptions at the beginning that there would be a lack of German restaurants in Berlin. As it turns out, this type of cuisine is centered specially in the south west of Berlin. The borough with the least amount and concentration of German restaurants is Neukölln.

As a last point we have used the location of the non-German restaurants to create three recommendations of culinary hot-spots. These would probably be very interesting for opening a new German restaurant or a new restaurant in general using the k-means clustering algorithm.

6 Outlook

In this study the focus was on German restaurants. But the insights gained in this research could be extended by other type of restaurants. For example, which borough has the most Italian or Vietnamese restaurants. Also which borough has the most uncommon types of restaurants?

Also we examined only a small fraction of neighborhoods of Berlin. Mostly the party located in the heart of Berlin. It would be also possible to extend this selection to include all neighborhoods, boroughs, and postal codes belonging to Berlin.

These are questions which could be further investigated.