# Selecting a neighborhood to open a vegan restaurant in Berlin, Germany

Kristin Predeck

July 3rd 2020

## 1. Introduction

### 1.1. Business Problem and background

Background
Berlin is Germany's biggest city, with a population of 3.769 million people and a population density of 3944 inhabitants per km². The city as 12 boroughs, all very diverse and different. With around 80.000 vegans, the need for new restaurants to satisfy the high demand is paramount. With 12 boroughs and an area size of 891.1 km2, it might feel overwhelming to try and answer the questions as to where the best location to open a vegan restaurant would be and which cuisine it should have, as Berlin's food scene is just as diverse as its inhabitants.

Problem
We need to find out, which location would be the best to open a new vegan restaurant and which cuisine it should be. One way to achieve this goal is to use a map and information about where vegan restaurants are already located in Berlin and which venues are the most popular amongst Berliners. Another insightful information is where vegan restaurants cluster in Berlin. Based on this information, we can see which cuisine is the most popular and where the competition for vegan restaurants is the smallest.

Interest
New business owners who want to open vegan restaurants would be interested in the location data to have competitive advantages over their competitors and build a successful business.

## 2. Data

### 2.1 Getting geo data on Berlin

To obtain the necessary data, I web-scraped the borough information for Berlin off of Wikipedia. The data can be found here:
https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin
This information contained a few non-relevant columns, that I dropped and put the new information with borough names into a dataframe. Using the geocoder, I added latitude and longitude values for the boroughs, so that I could access venue information from Foursquare. The dataframe obtained looked like this:

Out[68]:

| | Borough | Latitude | Longitude |
|---|---|---|---|
| 0 | Charlottenburg-Wilmersdorf | 52.499620 | 13.323160 |
| 1 | Friedrichshain-Kreuzberg | 52.500000 | 13.450000 |
| 2 | Lichtenberg | 50.937420 | 6.947560 |
| 3 | Marzahn-Hellersdorf | 51.013100 | 6.900460 |
| 4 | Mitte | 50.856095 | 7.111685 |

## 3. Methodology

### 3.1 Using the foursquare API to obtain venue information

Using the Foursquare API, I first looked at the top 100 venues for all boroughs of Berlin. To get rid of duplicates, I printed out how many unique venue values the data had. There were 151 unique venue categories.

Out[90]:

(521, 7)

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Charlottenburg-Wilmersdorf | 52.49962 | 13.32316 | Mövenpick Weinkeller | 52.499556 | 13.320365 | Wine Shop |
| 1 | Charlottenburg-Wilmersdorf | 52.49962 | 13.32316 | Sweet2go | 52.497776 | 13.322638 | Dessert Shop |
| 2 | Charlottenburg-Wilmersdorf | 52.49962 | 13.32316 | BERLINRODEO interior concepts GmbH | 52.501163 | 13.325814 | Furniture / Home Store |
| 3 | Charlottenburg-Wilmersdorf | 52.49962 | 13.32316 | Piccola Taormina | 52.501231 | 13.324884 | Italian Restaurant |
| 4 | Charlottenburg-Wilmersdorf | 52.49962 | 13.32316 | EDEKA | 52.500454 | 13.322082 | Supermarket |

**Okay, how many unique categories are there?**

After printing out a list of those and looking at the different types, I checked the top 5 categories per borough to see whether vegan restaurants would pop up and to find out which venues were the most popular in general.

```
----Charlottenburg-Wilmersdorf----
                venue  freq
0  Italian Restaurant  0.12
1               Hotel  0.08
2               Café  0.07
3        Dessert Shop  0.04
4                 Bar  0.04


----Friedrichshain-Kreuzberg----
            venue  freq
0  Ice Cream Shop  0.09
1       Nightclub  0.06
2       Rock Club  0.06
3            Café  0.06
4      Distillery  0.03


----Lichtenberg----
                venue  freq
0  Italian Restaurant  0.07
1               Café  0.05
2          Restaurant  0.04
3               Hotel  0.04
4         Salad Place  0.03
```

### 3.2 Narrowing the dataframe down to our target variable

While Italian was in almost all of the top 5 venues for the boroughs, vegan was not one of the top categories. So, I further checked the boroughs, which did have vegan restaurants. The results showed that only 5 of the 12 boroughs had vegan restaurants, namely Charlottenburg-Wilmersdorf, Lichtenberg, Neukölln, Pankow and Spandau.

| | Neighborhood | Vegetarian / Vegan Restaurant |
|---|---|---|
| 0 | Charlottenburg-Wilmersdorf | 0.010000 |
| 1 | Friedrichshain-Kreuzberg | 0.000000 |
| 2 | Lichtenberg | 0.013333 |
| 3 | Marzahn-Hellersdorf | 0.000000 |
| 4 | Mitte | 0.000000 |
| 5 | Neukölln | 0.028571 |
| 6 | Pankow | 0.010000 |
| 7 | Reinickendorf | 0.000000 |
| 8 | Spandau | 0.010000 |
| 9 | Steglitz-Zehlendorf | 0.000000 |
| 10 | Tempelhof-Schöneberg | 0.000000 |
| 11 | Treptow-Köpenick | 0.000000 |

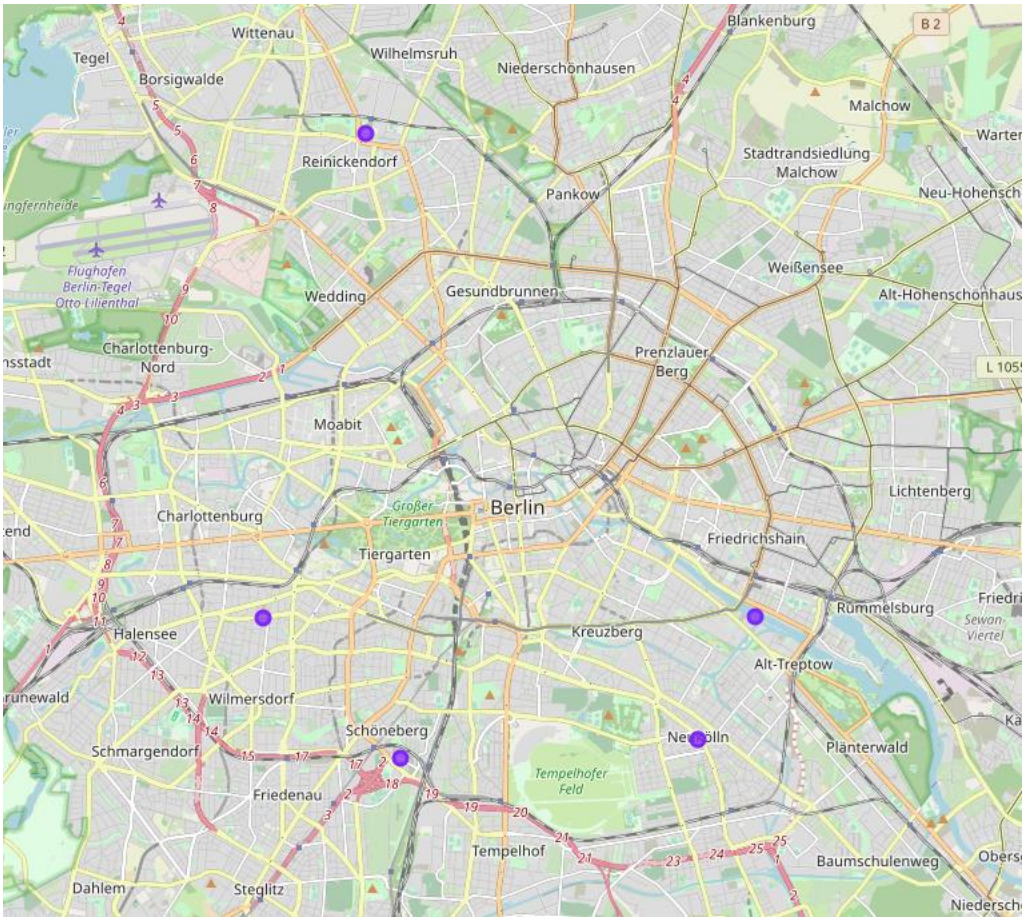
This returned valuable information, as all the other boroughs don't have vegan restaurants. The next thing I did, was to look at how many total venues per borough there are in Berlin.

| Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|
| Charlottenburg-Wilmersdorf | 100 | 100 | 100 | 100 | 100 | 100 |
| Friedrichshain-Kreuzberg | 35 | 35 | 35 | 35 | 35 | 35 |
| Lichtenberg | 75 | 75 | 75 | 75 | 75 | 75 |
| Marzahn-Hellersdorf | 5 | 5 | 5 | 5 | 5 | 5 |
| Mitte | 4 | 4 | 4 | 4 | 4 | 4 |
| Neukölln | 70 | 70 | 70 | 70 | 70 | 70 |
| Pankow | 100 | 100 | 100 | 100 | 100 | 100 |
| Reinickendorf | 6 | 6 | 6 | 6 | 6 | 6 |
| Spandau | 100 | 100 | 100 | 100 | 100 | 100 |
| Steglitz-Zehlendorf | 5 | 5 | 5 | 5 | 5 | 5 |
| Tempelhof-Schöneberg | 19 | 19 | 19 | 19 | 19 | 19 |
| Treptow-Köpenick | 2 | 2 | 2 | 2 | 2 | 2 |

Out of our boroughs that have vegan restaurants, Charlottenburg-Wilmersdorf has 100 total venues, Pankow has 100, Spandau has 100 as well, Lichtenberg has 75, and Neukölln has 70.

### 3.3 Clustering

In this step I grouped boroughs with similar features using the k-means algorithm. When we cluster our data by similarities, we get the following map:




Upon further investigation of the clusters, we can see that only 2 of these contain our target variable vegan, cluster 1 and 2. These include boroughs 9 and 3, so Marzahn-Hellersdorf and Steglitz-Zehlendorf.

## 4. Results and discussion

During the analysis we defined 4 clusters. From our analysis we can see that there are not a lot of vegan or vegetarian restaurants in Berlin - that is surprising! We only found 5 and they are pretty spread out. We also saw from our general observations, that the most popular cuisine seems to be

Italian. Our observations tell us that areas of great opportunity are clusters 3 and 4, as there is zero competition there. Cluster 1 should be avoided with 2/5 vegan restaurants in these neighborhoods, as well as cluster 2 with 3/5. Furthermore, if aiming for the best revenue, the vegan restaurant should be Italian themed, as this seems to be the most popular cuisine throughout all boroughs. We also found that out of the boroughs which have vegan restaurants, Lichtenberg and Neukölln have fewer total venues than the other 3 boroughs, so this would further narrow down our search for a good neighborhood. Further analysis could include features such as distance to public transportation, income of the residents, proximity to universities etc. This could help us to furhter refine the best possible location.

Keep in mind though that Foursquare does not represent the whole picture, because many venues are not listed. It is highly unlikely that there are only 5 vegan restaurants in all of Berlin. Other APIs such as Google or Yelp will likely deliver better results.

## 5. Conclusion

A basic exploratory data analysis has been performed, leveraging the Foursquare API venue data to identify the best borough for a vegan restaurant in Berlin. We identified the top venues per borough and also the top cuisine in Berlin. We also identified boroughs that already have vegan restaurants and which other venues there are in these boroughs using k-means clustering. Based on this we identified Marzahn-Hellersdorf and Steglitz-Zehlendorf as locations who do not yet have vegan restaurants and would therefore be competition free boroughs.