

Applied Data Science Capstone Project: Finding the right spot for a Vegan Lifestyle Centre in Berlin

Danilo Grimm

02.December 2019

Table of contents

1. [Introduction](#)
2. [Business Problem](#)
3. [Data Aquisition, Selection & Wrangling](#)
4. [Methodology & Exploratory Data Analysis](#)
5. [Predictive Modeling](#)
6. [Results](#)
7. [Conclusion/Discussion/Outlook](#)

Introduction

As part of the Applied Data Science Capstone Project, I will try to help a vegan, health awareness, ecologically-friendly entrepreneur to find the best location to set up a vegan lifestyle centre in Berlin that includes a vegan food store, an animal-free butchery, coffee shop and a first class vegan restaurant as well as a vegan clothing shop and a yoga studio in one place. To achieve this, I'll scrape Berlin's boroughs and neighborhoods from the [Wikipedia Website](#) (https://de.wikipedia.org/wiki/Liste_der_Bezirke_und_Ortsteile_Berlins), to get it's location data to feed it into the Foursquare API to find the most common and frequent venue categories within the neighborhoods and use the k-means clustering algorithm to group the neighborhoods into clusters.

Business Problem

According to the [independent.co.uk](#) in 2016, "an estimated 80.000 vegan residents" ([source](https://www.independent.co.uk/travel/news-and-advice/top-10-vegan-cities-travellers-in-world-food-turin-berlin-melbourne-tel-aviv-london-helsinki-a7707031.html) (<https://www.independent.co.uk/travel/news-and-advice/top-10-vegan-cities-travellers-in-world-food-turin-berlin-melbourne-tel-aviv-london-helsinki-a7707031.html>)) live in Berlin. The Website HappyCow that specialized in finding vegan/vegetarian restaurants around the world with their location based service, crowned Berlin second most popular city for vegans ([source](https://www.happycow.net/vegtopics/travel/top-vegan-friendly-cities) (<https://www.happycow.net/vegtopics/travel/top-vegan-friendly-cities>)) with a vegan hotel, vegan bar, vegan food chain stores, animal-free butchers and more than 50 vegan restaurants ([source](https://www.livekindly.co/berlin-vegan-capital-of-world/) (<https://www.livekindly.co/berlin-vegan-capital-of-world/>)). Moreover, Germany has been dominating the market for launching new vegan products between July 2017 and June 2018 with a total share of "15% of vegan introductions worldwide" ([source](https://www.mintel.com/press-centre/food-and-drink/germany-continues-to-dominate-global-vegan-new-product-development) (<https://www.mintel.com/press-centre/food-and-drink/germany-continues-to-dominate-global-vegan-new-product-development>)) and over 14% of the products labeled vegan according to 4% in 2013 ([source](https://www.mintel.com/press-centre/food-and-drink/germany-continues-to-dominate-global-vegan-new-product-development) (<https://www.mintel.com/press-centre/food-and-drink/germany-continues-to-dominate-global-vegan-new-product-development>)). If we also take into consideration that there is a growth in popularity for veganism and vegetarism worldwide ([source](https://www.ipsos.com/sites/default/files/ct/news/documents/2018-09/an_exploration_into_dietsAroundTheWorld.pdf) (https://www.ipsos.com/sites/default/files/ct/news/documents/2018-09/an_exploration_into_dietsAroundTheWorld.pdf)) and the potential of the vegan market and peoples' health and ecological awareness ([source](https://www.vegansociety.com/news/media/statistics) (<https://www.vegansociety.com/news/media/statistics>)), tourists are very interested in a place that has it all, where no further travelling from one place to another within the city is involved, especially for the people that short stay in Berlin. Since Berlin has become such a popular city for tourists with around 14 million tourists ([source](https://en.wikipedia.org/wiki/Tourism_in_Germany) (https://en.wikipedia.org/wiki/Tourism_in_Germany)) that come visiting Berlin every year, the potential is high that they'll pay the centre a visit.

Therefore, finding the right spot for opening up a vegan lifestyle centre is vitally important for our business. Since we do not want to target just the tourists, but also the locals, it means we need to focus on boroughs and particular neighborhoods that are not only close to the city centre, but easy to access for locals as well. To avoid too much competition around, we should focus on finding a location that is surrounded by a couple of hotels or accomodations with only a few vegan/vegetarian places or restaurants nearby and not too many yoga studios in that spot. We might not be interested in neighborhoods that are too lively with a lots of activities, such as drinking and entertainment opportunities, since they might not come by to pay the "calm,relaxed,health-related"-oriented centre a visit, when going out for a drink etc.

Data Aquisition, Selection & Wrangling

To start with, I will collect all the available data that is necessary to tackle the problem. For this project, I will need the boroughs and neighborhoods of Berlin with their longitude and latitude values as well as the Foursquare API venue data for existing vegan/vegetarian restaurants, other type of restaurants, yoga studios, hotels/hostels, bars & tourist attractions.

Boroughs and Neighborhoods of Berlin from Wikipedia:

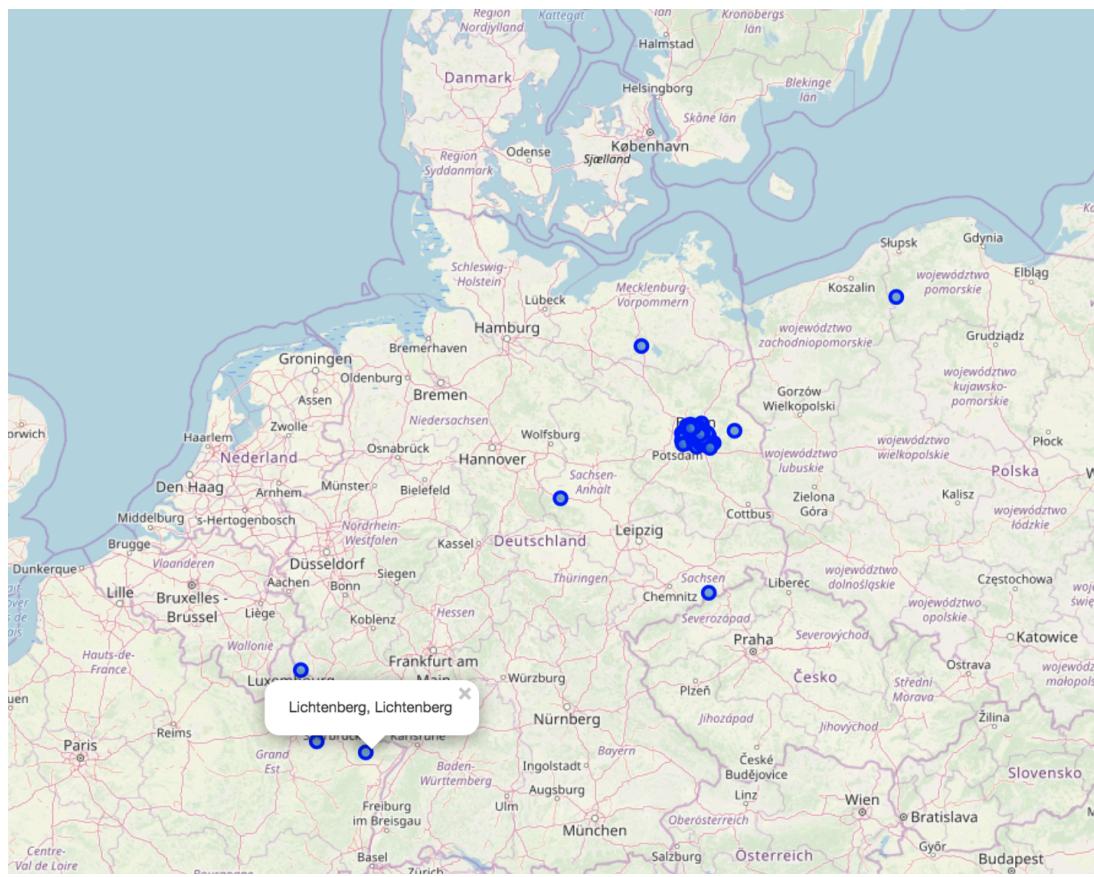
1. First of all, I used the table from the [Wikiedia Article](https://de.wikipedia.org/wiki/Bezirke_und_Ortsteile_Berlins) (https://de.wikipedia.org/wiki/Bezirke_und_Ortsteile_Berlins) that lists all 12 boroughs and 96 neighborhoods of Berlin ([github csv file](https://github.com/nllob/capstone_project/blob/master/boroughs_neighborhoods_berlin.csv) (https://github.com/nllob/capstone_project/blob/master/boroughs_neighborhoods_berlin.csv)). Since it was the second table in the article, I struggled to scrape the data from it using beautifulsoup. I also could not load the resources for [Pypi Wikipedia](https://pypi.org/project/wikipedia/) (<https://pypi.org/project/wikipedia/>). Therefore, I copied the data from the article and created a csv file out of it instead. One option to overcome this challenge would have been to loop through all the tables within the article and choose the one required. I noticed that when I imported the data from the csv file to pandas dataframe that there has been some hidden "\n" on the Neighborhood column, which needed to be replaced with an empty string to get a proper dataframe out of it. The table looked as follow:

	Boroughs	Neighborhood	City	Country
0	Mitte	Mitte	Berlin	Germany
1	Mitte	Moabit	Berlin	Germany
2	Mitte	Hansaviertel	Berlin	Germany
3	Mitte	Tiergarten	Berlin	Germany
4	Mitte	Wedding	Berlin	Germany

1. Moreover, I used geocoder to collect the latitude and longitude value for Berlin to map the neighborhoods on the Berlin area. The geographical coordinate of Berlin are 52.5170365, 13.3888599. The most important part of the table is the second column the neighborhoods. I will extend the neighborhoods by latitude and longitude values using Geocoder. To do so, I will reduce the dataframe to only having included Neighborhoods, City and Country to easier collect the longitude and latitude values for all neighborhoods at once and append them. I further checked if all neighborhoods have a latitude and longitude value or if there are any missing values.

	Boroughs	Neighborhood	City	Country	latitude	longitude
0	Mitte	Mitte	Berlin	Germany	52.517690	13.402376
1	Mitte	Moabit	Berlin	Germany	52.530102	13.342542
2	Mitte	Hansaviertel	Berlin	Germany	52.519123	13.341872
3	Mitte	Tiergarten	Berlin	Germany	52.509778	13.357260
4	Mitte	Wedding	Berlin	Germany	52.550123	13.341970

1. Then, I visualized the neighborhoods using the Folium library looking for outliers that have the wrong longitude or latitude values and are not in range for Berlin. From the map below, one can see that there have been some incorrect values for Berlin Neighborhoods placed around the world.



Here is a list of the values that have been wrong. One can clearly see that their latitude and longitude values are out of bounce for Berlin (52.5170365, 13.3888599)

	Boroughs	Neighborhood	City	Country	latitude	longitude
77	Lichtenberg	Lichtenberg	Berlin	Germany	48.921296	7.481227
78	Lichtenberg	Falkenberg	Berlin	Germany	49.048777	6.595821
79	Lichtenberg	Malchow	Berlin	Germany	53.477564	12.424517
84	Lichtenberg	Rummelsburg	Berlin	Germany	54.000148	16.977472
10	Pankow	Blankenburg	Berlin	Germany	51.820890	10.970734
90	Reinickendorf	Hermsdorf	Berlin	Germany	50.761263	13.631197
71	Marzahn-Hellersdorf	Biesdorf	Berlin	Germany	49.885462	6.305603
67	Treptow-Köpenick	Grünau	Berlin	Germany	-27.732656	18.378118

The reason for that is that there are some neighborhoods with the same name that exist in other locations. I put them in a separate dataframe, if needed later on, and remove the neighborhoods from the current dataframe. By checking the neighborhoods again using a different, more precise, address, I figured that all the neighborhoods that were located wrongly lay further away from the centre. See below the criteria I chose to pre-select the data. Having a closer look at the Berlin Neighborhoods map, I selected the neighborhoods of interest by zooming in on the map that meet our business objectives:

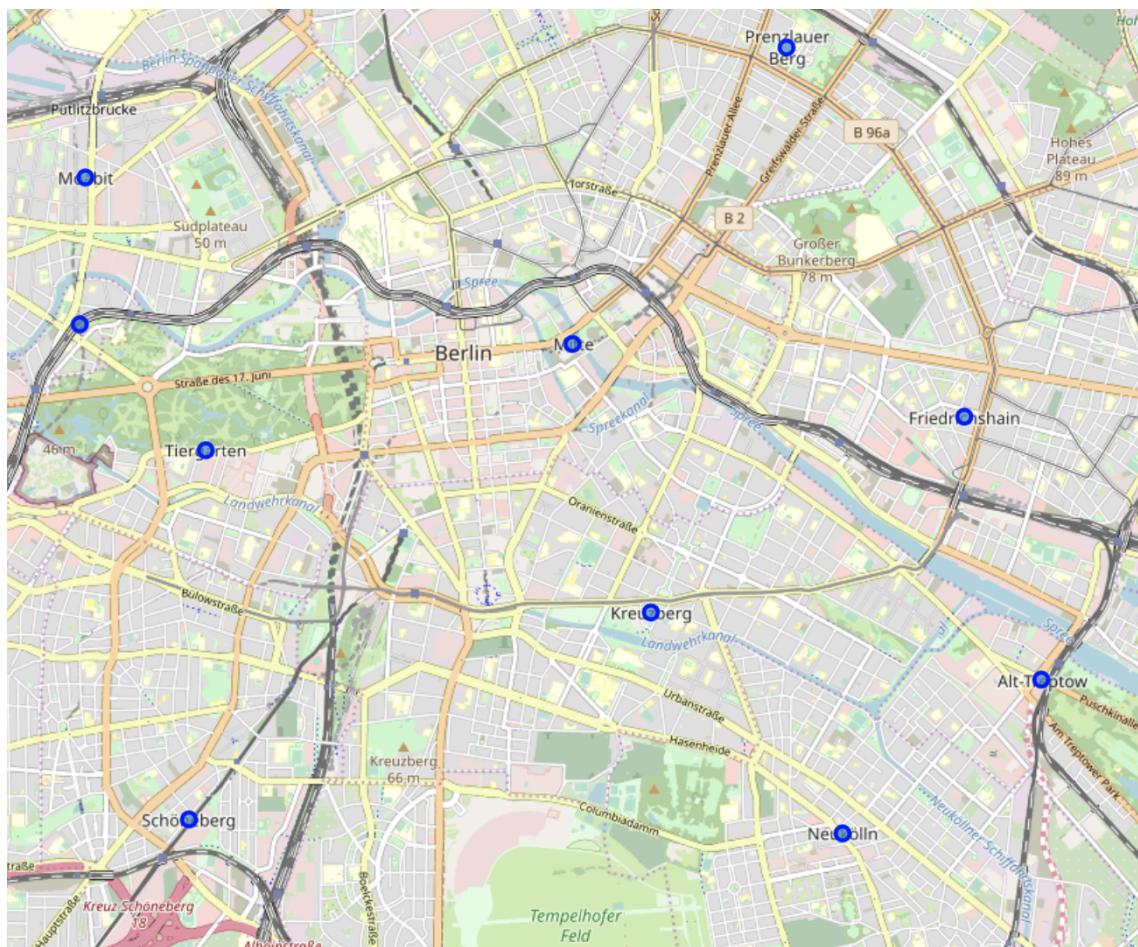
1. Neighborhoods that are near the city center
2. Neighborhoods that are centrally located
3. Neighborhoods that have easy access to public transport
4. Neighborhoods that are so called rising stars and tourist's know of

Since most of the tourist attraction are within Berlin Mitte, we consider this to be our starting point. From here on, I would suggest that the adjacent neighborhoods are important to discover and explore. Considering **Alexanderplatz Station** with the TV Tower & World Clock, **Friedrichstr. Station** with Museums-Insel, Stadtschloss, Dom and **Unter den Linden Station** with Reichstag and Brandenburg gate as one of the most busiest and most attractive places for tourist, which all are part of the borough Berlin Mitte, we can see on the public transport map that even regions like Moabit, Friedrichshain, Kreuzberg and Tiergarten can still be easily accessed by locals too S- und U-Bahn Netz (<https://sbahn.berlin/liniennetz/>)

Therefore, I came up with the following neighborhoods to be further discovered and selected for opening the vegan centre.

- Prenzlauer Berg
 - Mitte
 - Friedrichshain
 - Kreuzberg
 - Neukölln
 - Schöneberg
 - Moabit
 - Hansaviertel
 - Tiergarten
 - Alt-Treptow

Famous areas to go out or spend time at, match my criteria above pretty well ([Quelle](https://www.movinga.de/hub/expats/best-neighborhoods-in-berlin/)) (<https://www.movinga.de/hub/expats/best-neighborhoods-in-berlin/>)



This pre-selection might have a huge impact on the outcome of my analysis/project. Therefore, it might be possible that, once I have gone through the clustering, I have to go through the whole project re-adjusting the neighborhoods. This will be part of the discussion section to go more into detail about it.

1. After pre-selecting the neighborhoods, I'm interested in, I put them into a pandas dataframe to work with them throughout the project. Then, I used the Foursquare API to receive for each neighborhood the top venues available.

Using the Foursquare API to get the venue data:

To get an overview of the the most common venue categories in the interested neighborhoods, I will collect the top 200 venues within a 1500 meter radius to check if the categories of interest are among them (vegan/vegetarian restaurants, yoga studios, bio/organic supermarket, hotels, main attractions). The wider radius of 1500m is due to the fact that neighborhoods in Berlin can be rather big in comparison to New York's or Toronto's.

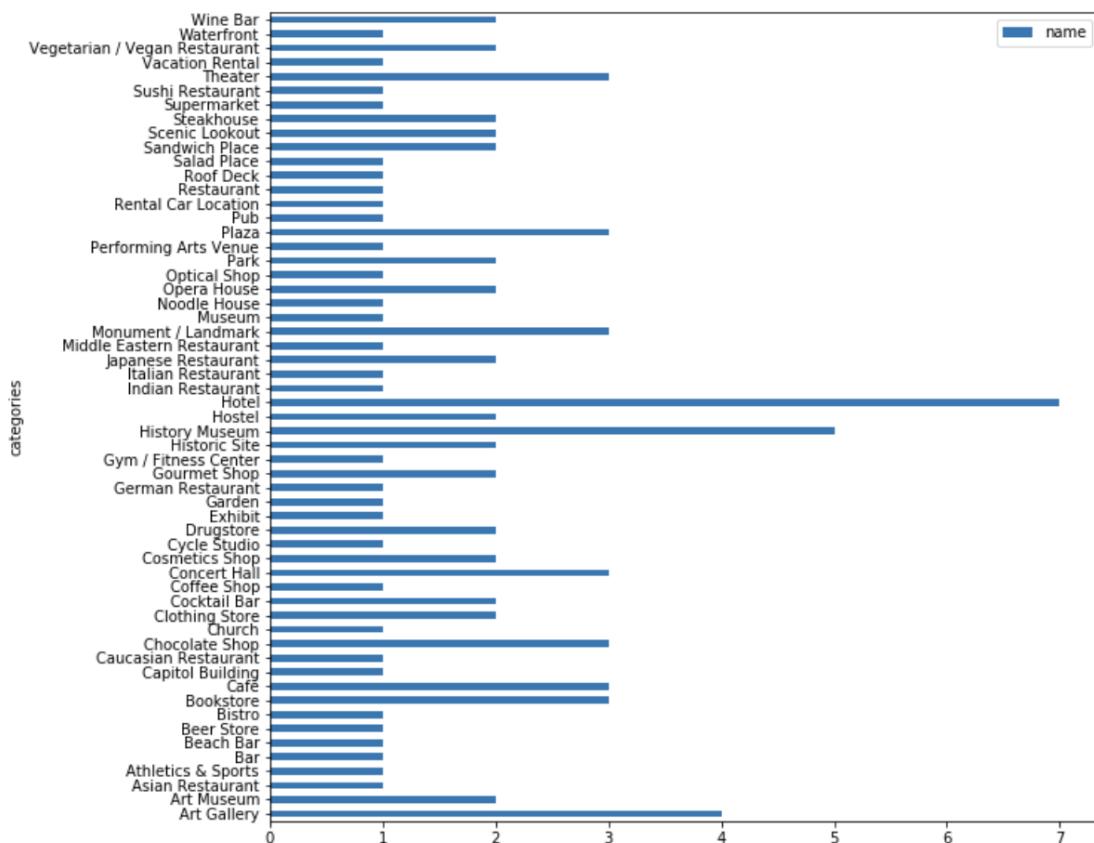
To receive all the required data, I create a function to loop through the Foursquare URL Requests for each neighborhood. Once, I've extracted the categories out of the JSON, returned from the Foursquare API, I'll beautify the data to fit it into a pandas dataframe and check for the desired neighborhood the amount of venues that were returned.

	name	categories	lat	lng
0	Dussmann English Bookshop	Bookstore	52.518223	13.389239
1	Dussmann das KulturKaufhaus	Bookstore	52.518398	13.388788
2	Cookies Cream	Vegetarian / Vegan Restaurant	52.516569	13.388008
3	Windhorst	Cocktail Bar	52.518553	13.386270
4	Freundschaft	Wine Bar	52.518294	13.390344

Methodology & Exploratory Data Analysis

Analysing the data from Foursquare API:

This allows us to find out how often a categorie eg. a vegan/vegetarian restaurant exists for each neighborhood on average. In this case for Berlin within a 1500m radius from the center. The results depicts that there are multiple categories among the top 100 venues and how often they exist.



I checked the amount of top venues available for each neighborhood in Berlin and narrowed it down to the neighborhoods that we were interested in. One can clearly see that the selected neighborhoods all have top100 venues within the region, while some of the more outside of Berlin neighborhoods only have around 20 to 40 venues within the radius of 1500m of its center. This could have to do with the size of the neighborhoods. The city centre is more dense and squeezed in, while the neighborhoods further away from the center are in general bigger and have fewer venues within a certain radius.

Neighborhood

Adlershof	38
Alt-Hohenschönhausen	48
Alt-Treptow	100
Altglienicke	17
Baumschulenweg	23
...	...
Wilhelmsruh	38
Wilhelmstadt	23
Wilmersdorf	100
Wittenau	29
Zehlendorf	44

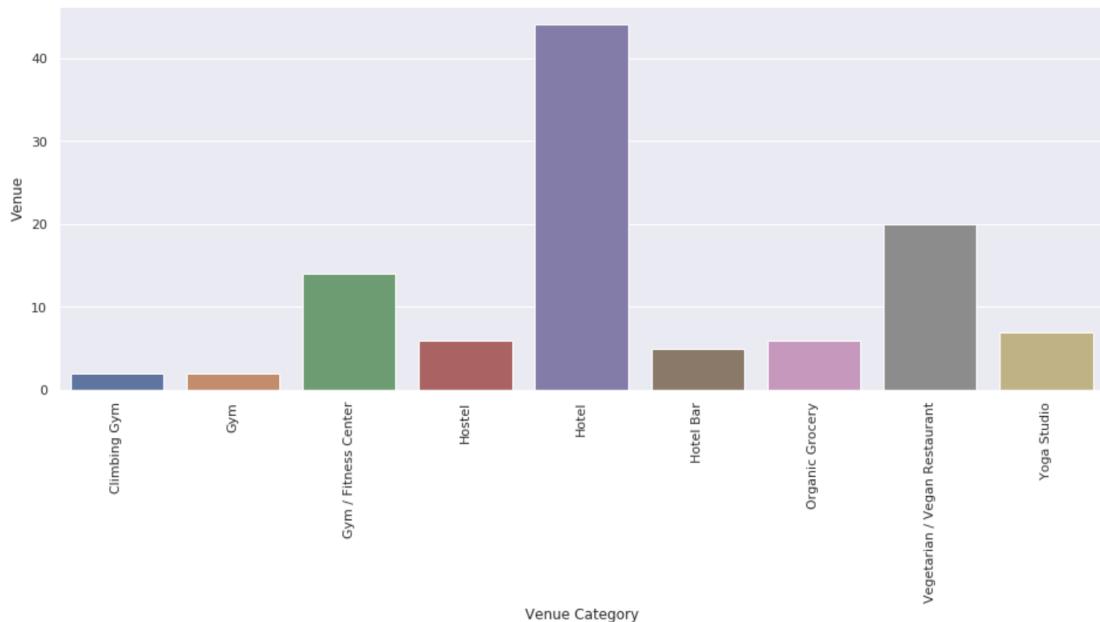
Neighborhood

Alt-Treptow	100
Friedrichshain	100
Hansaviertel	100
Kreuzberg	100
Mitte	100
Moabit	100
Neukölln	100
Prenzlauer Berg	100
Schöneberg	100
Tiergarten	100

By checking the amount of Vegan/Vegetarian Restaurants per Neighborhood, one can see that Friedrichshain has the highest number of restaurants in this category, while Hansaviertel does not have a single Vegan Restaurant within the top100 venues.

Neighborhood	
Alt-Treptow	2
Friedrichshain	7
Kreuzberg	2
Mitte	1
Moabit	2
Neukölln	2
Prenzlauer Berg	1
Schöneberg	1
Tiergarten	2

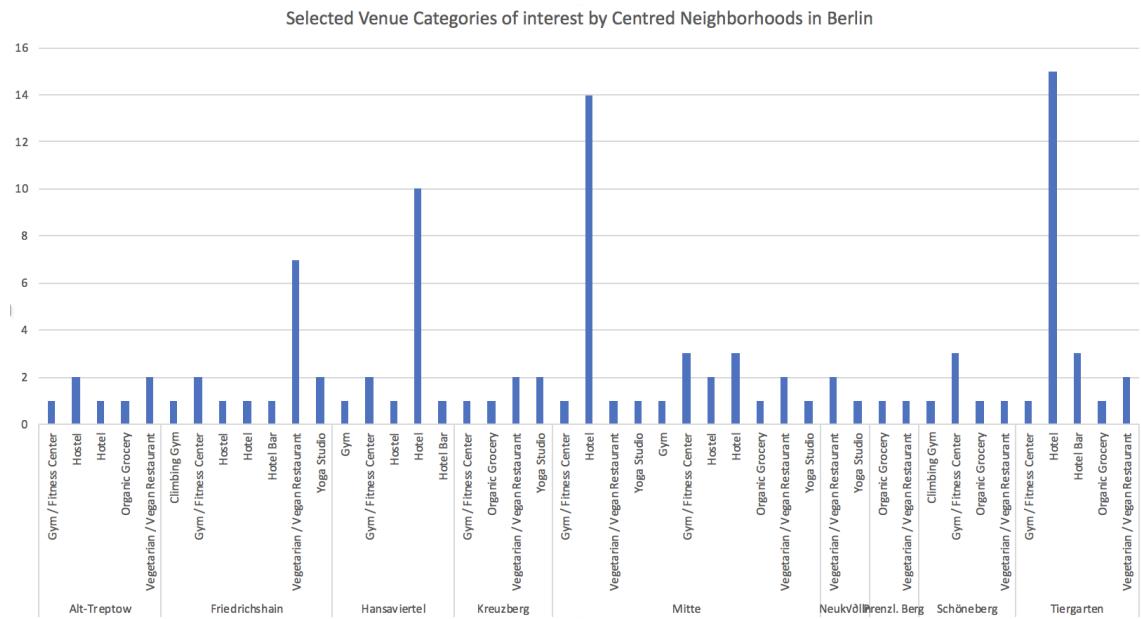
Within the selected neighborhoods, these were the amount of pre-selected categories (relevant for our business) and how often they appeared. For example, there were 20 Vegan restaurants among the neighborhoods and 44 hotels.



To deep dive more into the analysis/data and to prepare the data for the algorithm, I will do a "one hot encoding" with the data. That means that I will be able to determine the most common venues (top15) within the neighborhood. Having a closer look at the table, only Friedrichshain and Alt-Treptow are the neighborhoods that have a Vegan/Vegetarian Restaurants under the top15 venues.

Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue	12th Most Common Venue	13th Most Common Venue	14th Most Common Venue
0 Alt-Treptow	Café	Nightclub	Bar	Park	Rock Club	Coffee Shop	Ice Cream Shop	Pizza Place	Bakery	Beer Garden	Indian Restaurant	Hostel	Vegetarian / Vegan Restaurant	Vietnamese Restaurant
1 Friedrichshain	Café	Vegetarian / Vegan Restaurant	Coffee Shop	Nightclub	Ice Cream Shop	Bar	Thai Restaurant	Record Shop	Falafel Restaurant	Middle Eastern Restaurant	German Restaurant	Italian Restaurant	Cocktail Bar	Gym / Fitness Center
2 Hansaviertel	Zoo Exhibit	Hotel	Art Museum	Café	Beer Garden	Park	Coffee Shop	Italian Restaurant	Vietnamese Restaurant	Gym / Fitness Center	Indian Restaurant	Restaurant	Bar	Bakery
3 Kreuzberg	Coffee Shop	Bar	Café	German Restaurant	Art Gallery	Wine Bar	Pizza Place	Turkish Restaurant	Cocktail Bar	Ice Cream Shop	Sushi Restaurant	Music Store	Indie Movie Theater	Gourmet Shop
4 Mitte	Hotel	History Museum	Coffee Shop	Café	Plaza	Art Gallery	Museum	Concert Hall	Bookstore	Restaurant	Chocolate Shop	Dance Studio	Park	Art Museum
5 Moabit	Café	Bar	Vietnamese Restaurant	Turkish Restaurant	Italian Restaurant	Hotel	Cocktail Bar	Park	Coffee Shop	Indian Restaurant	Restaurant	Gym / Fitness Center	Trattoria/Osteria	Pastry Shop
6 Neukölln	Bar	Coffee Shop	Café	Cocktail Bar	Italian Restaurant	Indie Movie Theater	Pizza Place	Bakery	Beer Bar	Bistro	Breakfast Spot	Dive Bar	Food & Drink Shop	German Restaurant
7 Prenzlauer Berg	Café	Vietnamese Restaurant	Cocktail Bar	Ice Cream Shop	Bar	Breakfast Spot	Beer Bar	Bakery	Falafel Restaurant	Park	Japanese Restaurant	Chinese Restaurant	Korean Restaurant	Coffee Shop
8 Schöneberg	Café	Pizza Place	Bakery	Bistro	Vietnamese Restaurant	Gym / Fitness Center	Cocktail Bar	French Restaurant	Indie Movie Theater	Doner Restaurant	Park	Korean Restaurant	Italian Restaurant	Ice Cream Shop
9 Tiergarten	Hotel	Art Gallery	Art Museum	Concert Hall	Italian Restaurant	Zoo Exhibit	Hotel Bar	Plaza	Restaurant	Bar	Café	Cocktail Bar	Modern European Restaurant	Monument / Landmark

When looking for the business relevant categories per neighborhood, one can see that Friedrichshain has the highest density of vegan restaurants.



What I did next, was to group the venues into categories such as Bars, Restaurants, Cafe, Pub, Nightclub and count them by neighborhood. I considered those categories as "Entertainment" or "Going out" activities among the top15 venues. Friedrichshain and Prenzlauer Berg have a total of 10 out of 15 venues that can be associated with the "Going out" factor. Interestingly, only Friedrichshain and Alt-Treptow have Vegan Restaurants among the top15 venues. Prenzlauer Berg, Schöneberg and Moabit also have a high amount of Restaurants among the top15 but none are vegan.

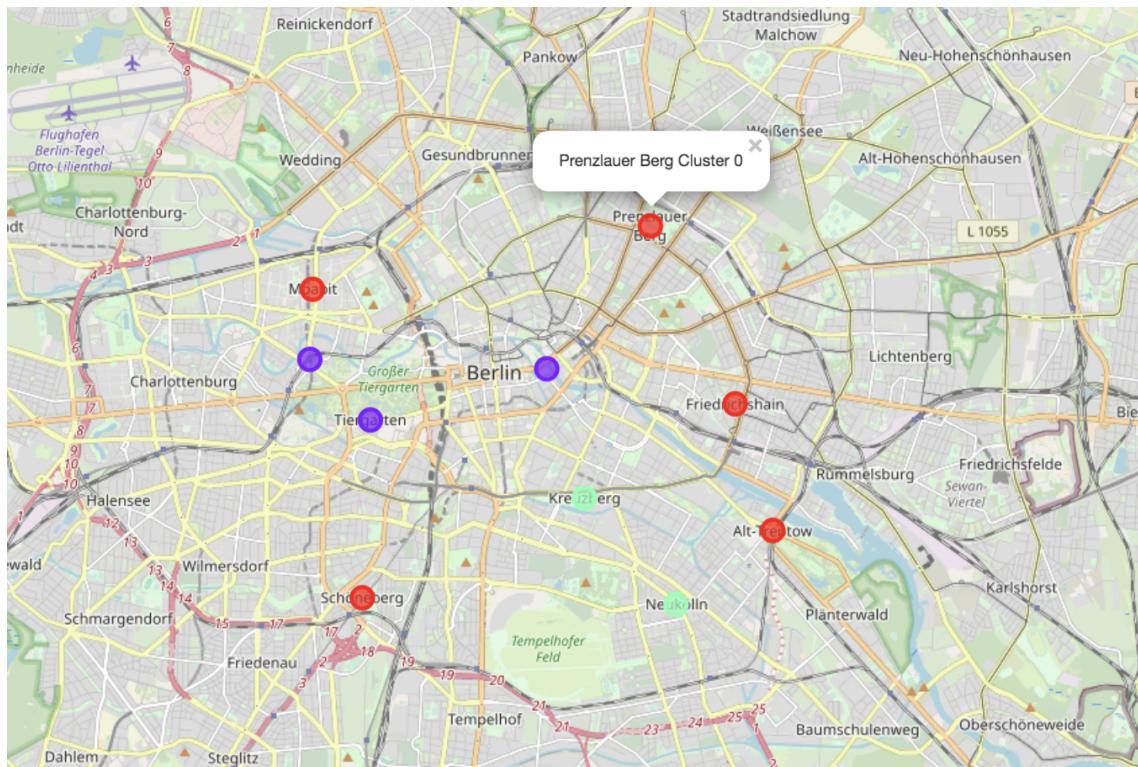
	Neighborhoods	Venues_Con_Bar	Venues_Con_Restaurant	Venues_Con_Cafe	Venues_Con_Pub	Venues_Con_Nightclub	Venues_Vegan_Rest	Sum Venues	Rest Bar Cafe
0	Alt-Treptow	1	3	1	1	1	1	7	7
1	Friedrichshain	2	6	1	0	1	1	10	
2	Hansaviertel	1	4	1	0	0	0	6	
3	Kreuzberg	3	3	1	0	0	0	7	
4	Mitte	0	1	1	0	0	0	2	
5	Moabit	2	5	1	0	0	0	8	
6	Neukölln	4	3	1	0	0	0	8	
7	Prenzlauer Berg	3	6	1	0	0	0	10	
8	Schöneberg	1	5	1	0	0	0	7	
9	Tiergarten	3	3	1	0	0	0	7	

Predictive Modeling

Cluster the neighborhoods using k-means partitioning clustering algorithm

Finding neighborhoods with similar characteristics based on the venue data from the Foursquare API and clustering them, can be of importance for our business decision, where to set up the vegan centre. This information can be used to focus more on our targeted audience and where they can be found. For that, I used the unsupervised learning algorithm k-means for clustering the neighborhoods, because it is one of the most popular and commonly used algorithm for partitioning the neighborhoods. It clusters the neighborhoods into groups that share the mostly frequent used venues of interest. This can be of interest when looking for little or no competition in a neighborhood or places of interest nearby that help acquire new customers. The positive aspect of the algorithm is that objects that are within a cluster are very lookalike while the objects in other clusters are very different. So, once found a cluster that suits best, it's very likely that it is not interfering with another one, where it might harm our business.

I start running k-means to cluster the neighborhoods into **3 clusters**. With the most frequent categories at hand, I will be able to categorize the cluster into meaningful segments to look for the best place that covers most of the categories that match our criterias for preventing too much competition and to set up the business.



Sum Venues GoingOut Venues_Con_Restaurant Venues_Vegan_Rest

Cluster Labels

Cluster Labels	0	8.4	5.000000	0.4
1	5.0	2.666667	0.0	
2	7.5	3.000000	0.0	

Findings for k=3

By looking at the map and the table, one can see the following:

Cluster 0: High amount of GoingOut Venues (high number of Restaurants, Cafe, Bar, Nightclub and Vegan Restaurants among the top15 venues)

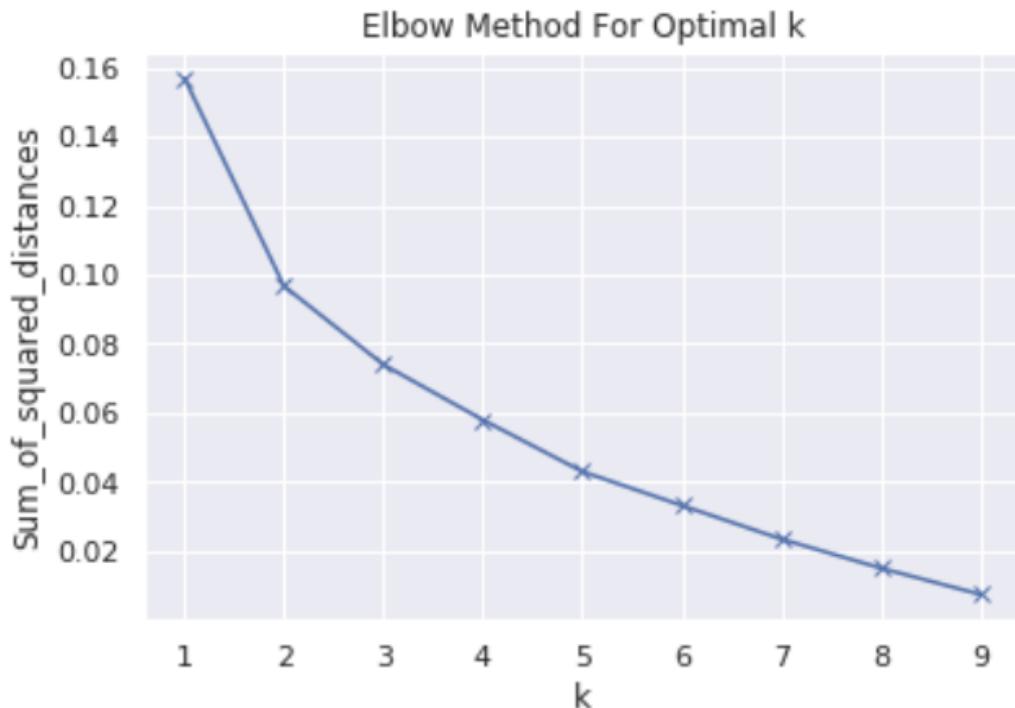
Cluster 1: Fair amount of GoingOut Venues (Lowest number for GoingOut Venues including the lowest number for Restaurants and zero Vegan restaurants among the top15)

Cluster 2: Medium amount of GoingOut Venues (high number for GoingOut Venues, but far less restaurants on average and zero Vegan restaurants among the top15)

In this case, Cluster 1 (the purple bubbles) is probably the best choice, since competition is lower for vegan restaurants and the amount of restaurants is the smallest. But let's find out if k=3 is the optimum or if a smaller or higher number of clusters is more suitable.

Elbow Method to determine the optimal k

Determining the best number of k in the k-Means clustering algorithm is one of the key aspect of how successfully the outcome of the algorithm will be. Therefore, it is important to find the best k in order to generate the optimum sphere-like clusters shaped around the centroids. There are many other ways to determine the optimum k, like The Silhouette Method. In this case study, I'll focus on the widely known and most popular approach, the Elbow Method. It calculates the sum of squared errors within each cluster. To minimize the error, one needs to find the right spot, where the error will start to diminish.



By looking at the graph, it is not 100% clear which k is the optimum, but one can clearly say it will be between 2 or 3. Therefore, we'll run the algorithm again, setting k=2.

Clustering k=2

Map k=2

Cluster Labels	Sum	Venues_GoingOut	Venues_Con_Restaurant	Venues_Vegan_Rest
0	8.142857	4.428571	0.285714	
1	5.000000	2.666667	0.000000	

Findings k=2

Cluster 0: High amount of GoingOut Venues (high number of Restaurants as well as Cafe, Bar, Nightclub and Vegan Restaurants among the top15 venues)

Cluster 1: Fair amount of GoingOut Venues (incl. restaurants but zero Vegan restaurants among the top15 venues)

Our choice would be Cluster 1 again.

Results

After checking the numbers of clusters of k=2 and k=3, I found it easier to choose from k=2, when it comes to where to locate the vegan lifestyle centre, because the choices with no vegan restaurants within the cluster 1 for the top15 common venues and less restaurants, means that there is lower competition for us and that is what we were looking for in the first place. When using k=3, the choice is less obvious but still Cluster 1 would be probably the best fit out of the three since the amount of restaurants is lowest, there is no vegan restaurant and the amount of GoingOut venues is lowest in the cluster too. By comparing the two maps (k=2 vs k=3), one can see that Cluster 2 (k=3) falls into Cluster 1 of k=2. That means that when we reduce the cluster by one, all the values from that cluster would fall into the Cluster 1, the one that we wouldn't choose to set up our business. This in turn means that having set k=3 does not give us an extra value or business opportunity and the extra cluster is redundant

Looking at the map, one can see that Cluster 1 for k=2 and k=3, is right in the centre of Berlin with all the tourist hotspots and main attraction. Prices for renting out a place to run a business there, might be very high and not optimum. That could also be the reason why there are only exist a few places to eat and drink. This brings us right to the conclusion and discussion section.

Conclusion/Discussion/Outlook

Conclusion

Given the data from Wikipedia and the Foursquare API, we could use the unsupervised clustering algorithm k-means to determine the best spot to set up our vegan lifestyle centre business. For k=2 and k=3, the cluster 1 (very centrally located around the borough Berlin Mitte) would be the best fit. While k=3 did not result in an obvious choice, running the elbow method, brought us to the conclusion that the values from the extra cluster fall into the cluster that we wouldn't choose. This means that k=2 was more helpful in figuring out that cluster 1 would be our best fit, since it lead us to a more obvious result to base our decision on.

Discussions on:

The Business Problem:

- We know that the Business model does not only depend on the location (based on clustering), but it has an impact on it. Information, such as rental prices for the cluster as well as the amount of vegan restaurants, which are already there, but not among the most common one, or the activities vegan people prefer, are also relevant and might play a huge roll in setting up the business there or not. This holds also true for yoga studios and hotels nearby.

Data: Based on the data from the API, the following questions arise:

- how good is the quality of it? how much emphasize can be given on the most common venues in terms of who uses the Swarm, former Foursquare App? Does our target group use the App, are they over represented or under represented in the App. How important are the not so common venues?
- How good is the data on new restaurants? Will they appear instantly? Are new places over or underrated?
- How accurate is the data from Foursquare? One should check the app with the actual venues in a particular area of interest.
- How well did I chose the radius? Is it too big or too small? What impact does it have on our data and cluster?
- How well chosen were the neighborhoods? Do we need more or even less neighborhoods to look at? The k-means algorithm works best with fairly big data. Therefore, it might be useful to feed the algorithms with more neighborhoods.
- Is there other data available to base our decision on (sources for well-being, income, rental prices, population, socio demographic data)?
- Should we pay more attention to other aspects like accomodations and yoga studios?
- Are all restaurants are competitors? Maybe, it's good to have some restaurants nearby, because not everyone might be interested in the restaurant, but also in other aspects of the centre (animal free butchery, yoga studio, clothing shop, etc.)

A more general question that arises:

- Are 80.000 vegan living people, tourists and people, who are interested, enough?

Outlook:

Clustering Algorithm:

- We know that the algorithm's outcome also depends on the distance measurement used, eg. using a different measurement from Euclidian Distance like the Manhattan Distance might lead to another clustering that impacts our business decision. Why not investigate it by trying it out? Maybe, this will change where the values from one cluster would fall into another one if we reduce k by 1.
- We should also use an alternative measurement like the silhouette method to find if the optimum is still k=2.
- Furthermore, we should compare neighborhoods in other cities for example New York, Moscow, London or Tokyo with our data. This will be very interesting when opening or starting a vegan lifestyle centre chain to let our business grow worldwide once set up successfully locally.