

# Opening up a restaurant in Berlin – Analysis of venues by using KMeans Clustering

---

Published: 23.01.2021

Author: Daniel V.

## Opening up an restaurant in Berlin

### Introduction

The German capital Berlin has a unique history with it's separation after the second World War, lasting for decades. Since then it has become one of the most visited and populated cities of Europe, influenced by a lot of different cultures around the world. This eventually leads to a lot of potential for setting up an restaurant. In this project I want to determine the best location to open a restaurant by using Data Science

### Business Problem

There are many factors to be considered for opening up an restaurant. The main goal of this Data Science project is to determine an optimal location to set up an restaurant. This might help future restaurant owners and current ones who want to set up a new restaurant in Berlin. By using unsupervised machine learning to create clusters of districts a list of areas will show up to choose from for opening up a restaurant. The restaurant should be close to gastronomical centres and touristic hotspots.

### Data

To solve the business problem following data will be required:

1. A list of the districts and boroughs of Berlin  
([https://en.wikipedia.org/wiki/Boroughs\\_and\\_neighborhoods\\_of\\_Berlin](https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin))
2. Geo-coordinates of districts and boroughs in Berlin (Geocoder tool in the notebook)

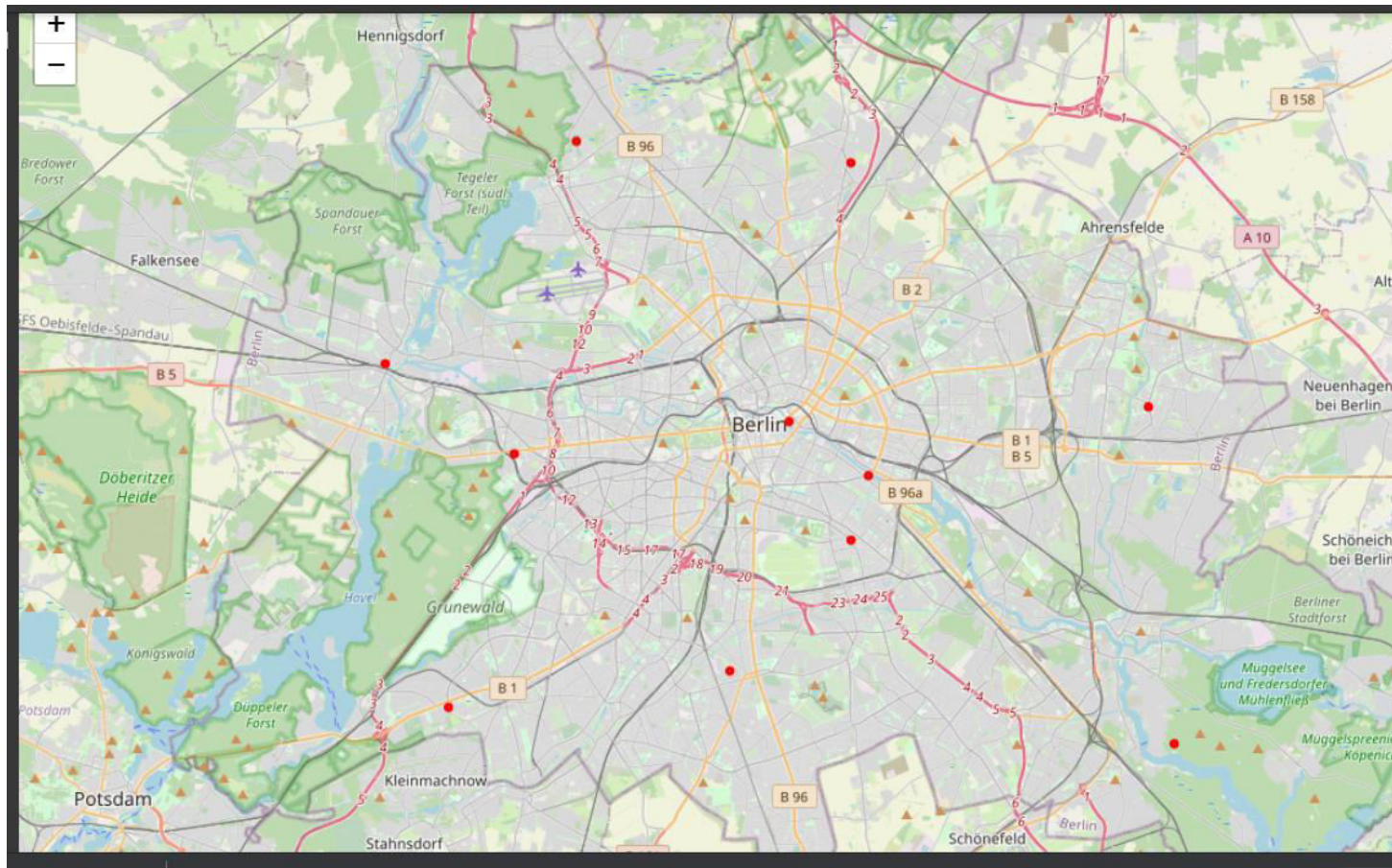
3. Top venues of districts (Foursquare API). This will help us to determine the quality of each cluster later on in the analysis.

## Methodology

Google Colab will be used to create the notebook. First we need to extract the information containing the boroughs of Berlin from Wikipedia. Afterwards we clean the data and append the geographical coordinates of each borough's centre coordinates by using Geocoder. This results in the following data frame.

	Name	Latitude	Longitude
0	Charlottenburg-Wilmersdorf	52.507856	13.263952
1	Friedrichshain-Kreuzberg	52.501115	13.444285
2	Lichtenberg	48.921296	7.481227
3	Marzahn-Hellersdorf	52.522523	13.587663
4	Mitte	52.517885	13.404060
5	Neukölln	52.481150	13.435350
6	Pankow	52.597917	13.435316
7	Reinickendorf	52.604763	13.295287
8	Spandau	52.535788	13.197792
9	Steglitz-Zehlendorf	52.429205	13.229974
10	Tempelhof-Schöneberg	52.440603	13.373703
11	Treptow-Köpenick	52.417893	13.600185

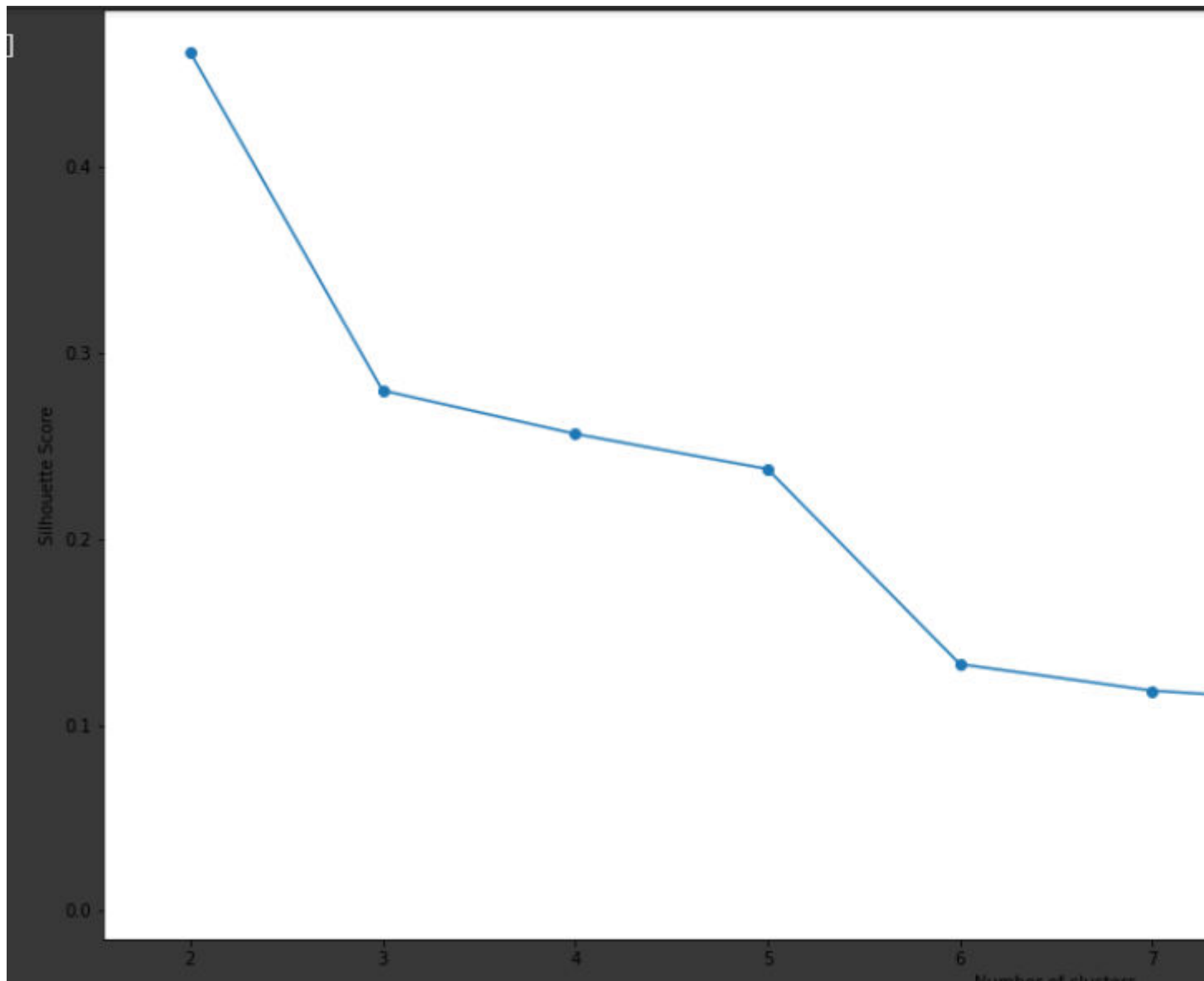
By using the Folium package we can visualize the boroughs of Berlin as red dots.



By creating a dataframe with the top 10 most common venues it is easier to determine information about each neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	
0	Charlottenburg-Wilmersdorf	Café	Italian Restaurant	Bus Stop	Plaza	Hotel	P
1	Friedrichshain-Kreuzberg	Bar	Café	Ice Cream Shop	Bakery	Hotel	
2	Lichtenberg	Hostel	Historic Site	Wine Shop	Doner Restaurant	Fast Food Restaurant	
3	Marzahn-Hellersdorf	Supermarket	Drugstore	Plaza	Trail	Shopping Mall	
4	Mitte	Hotel	Plaza	History Museum	Coffee Shop	Café	

To prepare for step using machine learning, we one hot encode the types of venues into dummies. By using the machine learning method of KMeans clustering, an unsupervised machine learning algorithm, we separate the neighborhoods into appropriate clusters. According to the graph 2 clusters would be the optimum amount. Since it created a cluster with only one borough, I've decided to choose 3 clusters, which was the second highest value in the graph.



The images below show each cluster calculated by KMeans, ranging from cluster 1 to cluster 3.

Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
Marzahn-Hellersdorf	52.522523	13.587663	0	Supermarket	Drugstore	Plaza	
Pankow	52.597917	13.435316	0	Tram Station	Supermarket	Cooking School	Restaurant

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Co
0	Charlottenburg-Wilmersdorf	52.507856	13.263952	1	Café	Italian Restaurant	Bus Stop	
1	Friedrichshain-Kreuzberg	52.501115	13.444285	1	Bar	Café	Ice Cream Shop	B
4	Mitte	52.517885	13.404060	1	Hotel	Plaza	History Museum	C
5	Neukölln	52.481150	13.435350	1	Bar	Coffee Shop	Café	Co
7	Reinickendorf	52.604763	13.295287	1	Zoo Exhibit	Bakery	Ice Cream Shop	
8	Spandau	52.535788	13.197792	1	Fast Food Restaurant	Clothing Store	Bakery	Drug
9	Steglitz-Zehlendorf	52.429205	13.229974	1	Italian Restaurant	Bus Stop	History Museum	
10	Tempelhof-Schöneberg	52.440603	13.373703	1	Bus Stop	Supermarket	Bakery	S
11	Treptow-Köpenick	52.417893	13.600185	1	Gastropub	River	Plaza	S



	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
2	Lichtenberg	48.921296	7.481227	2	Hostel	Historic Site	Wine Shop	Don Restaurants

## Results

The majority of the most common venues in the first cluster aren't related to gastronomy. In fact, in those boroughs most people use traffic stops or visit supermarkets. The second cluster seems to be the most appropriate one to open up a restaurant in, as several neighborhoods contain restaurants in the top 5 most common venues. Those neighborhoods are Charlottenburg-Wilmersdorf, Friedrichshain-Kreuzberg, Neukölln and Steglitz-Zehlendorf for example. The third cluster contains a few restaurants in the top 10 most common venues, but a hostel and fast food type of restaurants are dominating. It would be rather difficult to maintain a successful restaurant in this area.

## Discussion

For future research on this topic a different way of determining the number of clusters should get tested. Perhaps two and three different clusters are not enough, as the biggest cluster still showed differences in itself. Nevertheless, future restaurant owners in Berlin might want to avoid Lichtenberg, Pankow and Marzahn-Hellersdorf, as those neighborhoods are heavily contested by supermarkets or used only as traffic stops to move quickly away.

## Conclusion



Hopefully this report can give a little insight into the gastronomy locations of Berlin despite it's flaws and support future restaurant owners their decisions.