

# Capstone: Find the best neighborhood in Berlin to open a Döner Kebab Restaurant

## Introduction:

This project is about providing potential operators of kebab restaurants in Berlin with a data-driven recommendation for the location of their new business. This question is investigated in this project for Döner Kebab restaurants in Berlin (Germany).

Especially this Germanized fast food with Turkish roots is one of the most competitive gastronomic fields in Berlin.

## Some background information about Döner Kebab in Germany:

“Doner kebab (Turkish; "(spinning) grilled meat"), or kebab for short, is one of the most famous dishes in Turkish cuisine. It is similar to the Greek gyros. It consists of slices of meat seasoned with marinade, which are placed in layers on a vertical rotating spit and grilled on the side. The outer, browned layers are then gradually cut off.

In the early 1970s, the Federal Republic of Germany began to offer the meat cut from the roast cone not only on a plate with side dishes, but alternatively in a dumpling and thus also for takeaway. It is unclear when the first German kebab snack bar opened. According to legend, it was in the early 1970s in Berlin on Kottbusser Damm.

While the döner kebab was initially staged with folkloric elements in order to cater to the exoticism of German guests, the 1990s saw a clear transformation of the döner snack to the US-influenced global fast food culture.

The common variant of the doner kebab in pita bread in German-speaking countries differs from the Turkish one mainly by the addition of garden lettuce, sliced tomatoes, cucumbers and onions, white and red cabbage and by the sauces used with

mayonnaise and yogurt, for example in the variants "garlic", "herbs", "hot" and "curry", which do not belong to the traditional Turkish cuisine."

see here:

[https://de.wikipedia.org/wiki/D%C3%B6ner\\_Kebab#Verbreitung\\_im\\_deutschsprachigen\\_Raum](https://de.wikipedia.org/wiki/D%C3%B6ner_Kebab#Verbreitung_im_deutschsprachigen_Raum)



## Business Problem:

For entrepreneurs and especially for restaurateurs, the question of the right or best location always arises before opening a new establishment. To a not inconsiderable extent, this location determines the subsequent business success of the company. Therefore, it should be carefully selected, especially when we are dealing with such a large competition.

Berlin is Germany's capital and has 3.645 million (2019) inhabitants living in its 12 boroughs. These boroughs are each further divided into different numbers of neighborhoods. Of these inhabitants, approximately 200,000 residents have a Turkish migrant background, which is why döner kebab is particularly prized in Berlin.



## Target Audience:

The target group of this evaluation are entrepreneurs who want to open a kebab restaurant in one of the many neighborhoods in Berlin and for the decision would like to trust less on their gut feeling, but rather on the result of a data analysis.

## Data

One city will be analysed in this projects: Berlin

We will be using the below datasets for analysing Berlin:

**Data 1:** Berlin is Germany's capital and has 3.645 million (2019) inhabitants living in its 12 boroughs. These boroughs are each further divided into different numbers of neighborhoods. The required data was obtained with BeautifulSoup from one of the Wikipedia pages of Berlin. There they are available in one of the tables. In addition to the boroughs and the neighborhoods, information about the population density in the respective borough is also given there. These data are the basis for the further analysis.

Link to the data is: [https://de.wikipedia.org/wiki/Verwaltungsgliederung\\_Berlins](https://de.wikipedia.org/wiki/Verwaltungsgliederung_Berlins)

**Data 2:** Secondly, census data from the city of Berlin from 2019 was used. This data is available as a report and can be downloaded at the following address:

[https://download.statistik-berlin-brandenburg.de/d29b001f80353b17/289c7e11acc8/SB\\_A01-11-00\\_2019j01\\_BE.pdf](https://download.statistik-berlin-brandenburg.de/d29b001f80353b17/289c7e11acc8/SB_A01-11-00_2019j01_BE.pdf)

From the available data, the average household income for all boroughs is used. The income at a location can help to choose the right place for a new restaurant. For kebab restaurants, middle incomes are of interest because these people can afford to eat out but cannot always afford upscale restaurants.

**Data 3:** Berlin geographical coordinates data of the neighborhoods will be loaded by Nominatim (geopy.geocoders) to convert an address into latitude and longitude values.

**Data 4:** Berlin geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in Berlin. The below is image of the Foursquare API data.

## Data Wrangling

The data from the various data sources listed above had to be merged into one data set. The following steps were roughly necessary for this, which can also be traced in the associated notebook:

1. load administrative data of the boroughs and neighborhoods with Beautifulsoup.
2. load data into a dataframe with Pandas (`read_html()`)
3. since the data is not in the required form in the columns of the dataframe, it must be separated and stored in separate columns.
4. load the coordinates of the neighborhoods via Nominatim and add the longitudinal and lateral data in the dataframe.
5. load data on the average monthly income of households in the boroughs and add them to the dataframe.
6. Get all the Venues in Berlin
7. Get an overview about the venues in Berlins neighborhoods
8. Only add Restaurants as Venue Categories
9. Use this list to Extract Restaurants and only include Döner Kebab Restaurants in our Data Set
10. OneHot encode and count restaurants
11. Prepare the data for clustering
12. Combine all of those into a working Data Set to cluster and geo spatial map of the results showing the best neighborhood to open Doner Kebab Restaurant

Combining all of these disparate data sets will clearly demonstrate the following:

- which neighborhoods in Berlin have clusters of like Restaurants
- how populated each neighborhoods is
- the average income of these neighborhoods
- which neighborhood should he target to open a new Doner Kebab Restaurant



# Methodology

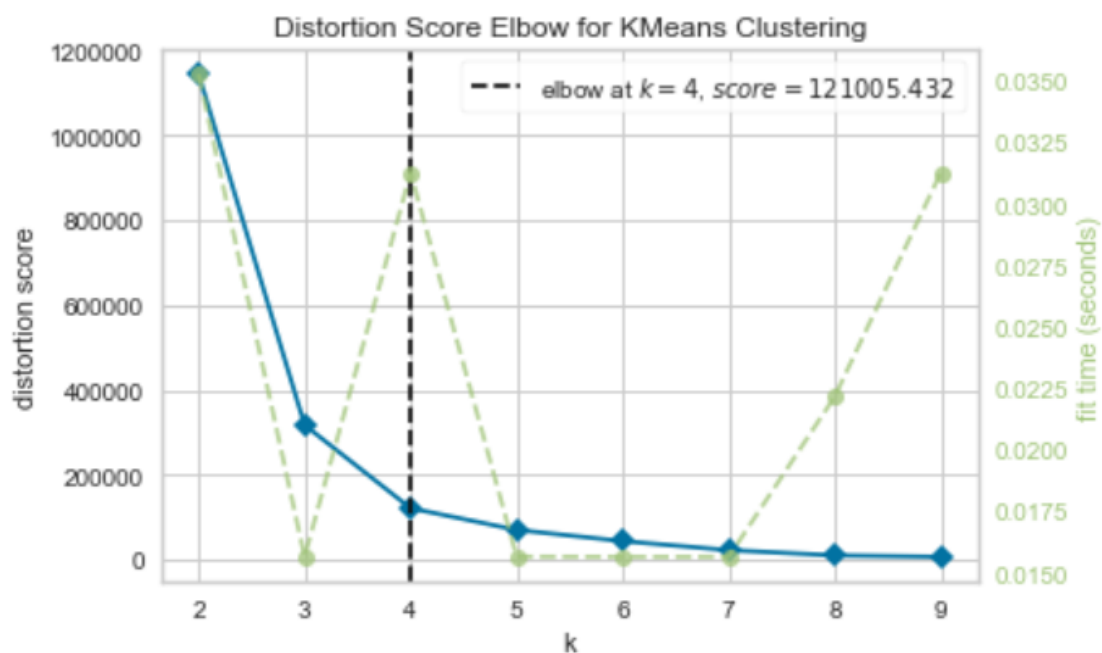
## Choice of Algorithms

I used K-Means Clustering. <https://towardsdatascience.com/clustering-algorithms-for-customer-segmentation-af637c6830ac>

A backgrounder on K-Means clustering “K-means clustering is an iterative clustering algorithm where the number of clusters K is predetermined and the algorithm iteratively assigns each data point to one of the K clusters based on the feature similarity.”

To use a reasonable number of clusters the elbow method was applied. For the neighborhoods in Berlin, this resulted in 4 clusters, taking into account the proportionate number of kebab restaurants, average income and population density.

A backgrounder on elbow method: "The KElbowVisualizer implements the “elbow” method to help data scientists select the optimal number of clusters by fitting the model with a range of values for K. If the line chart resembles an arm, then the “elbow” (the point of inflection on the curve) is a good indication that the underlying model fits best at that point. In the visualizer “elbow” will be annotated with a dashed line." see here: <https://www.scikit-yb.org/en/latest/api/cluster/elbow.html>

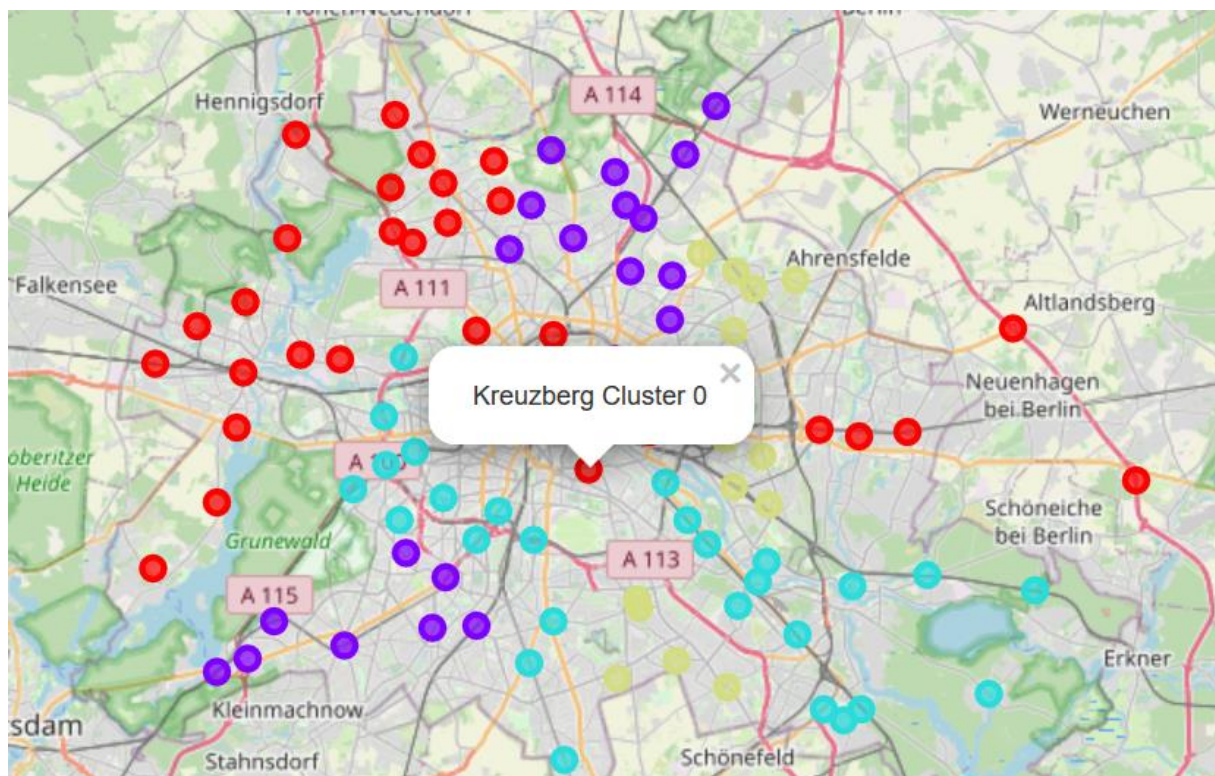


Next Steps:

1. Run K means and segment data into clusters and generate labels
2. Merge the Berlin data with cluster labels
3. Analyze clusters according to the different input dimensions

## Results

Based on the currently available data, the **Kreuzberg** neighborhood is recommended for potential new kebab restaurants to open in Berlin. There, the location factors are the best, so that a successful opening can be recommended here.

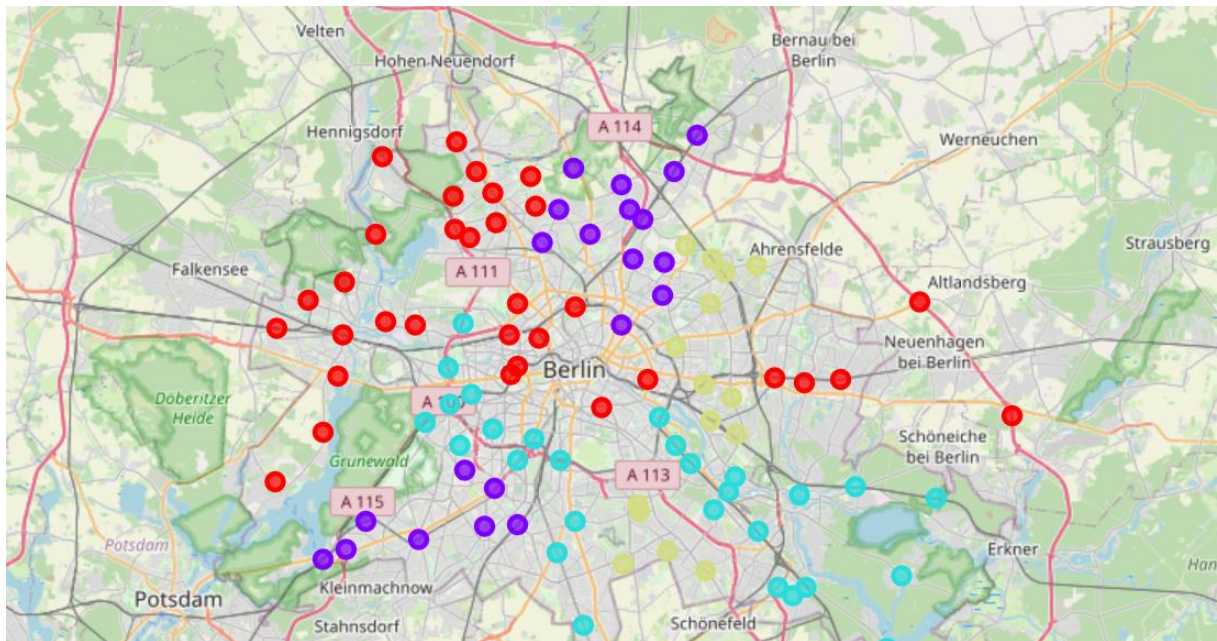


## Discussion

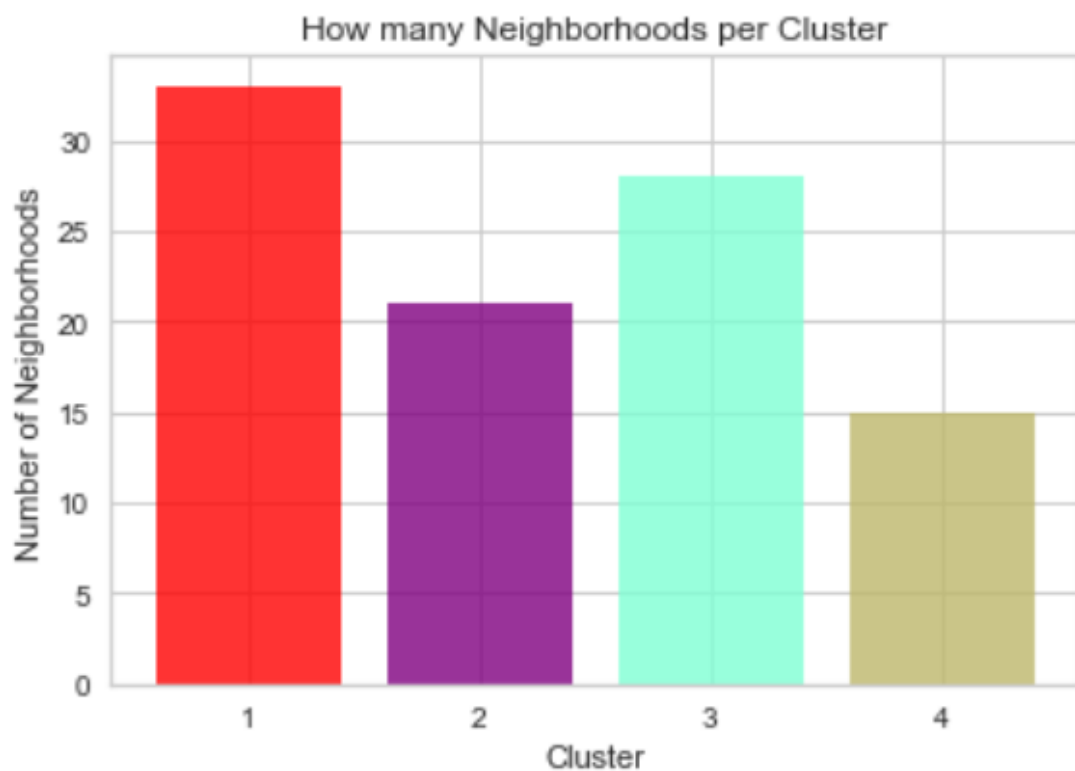
Resulting Dataframe (head()):

	Neighborhood	Latitude	Longitude	Borough	Doner Restaurant	Population_per_sqare_kilometer	Average monthly income in €	Cluster Labels
0	Mitte	52.533216	13.359796	Mitte	0.000000	9.713	2025	0
1	Moabit	52.534679	13.334478	Mitte	0.083333	9.713	2025	0
2	Hansaviertel	52.519123	13.341872	Mitte	0.000000	9.713	2025	0
3	Tiergarten	52.514375	13.336450	Mitte	0.000000	9.713	2025	0
4	Wedding	52.550123	13.341970	Mitte	0.000000	9.713	2025	0

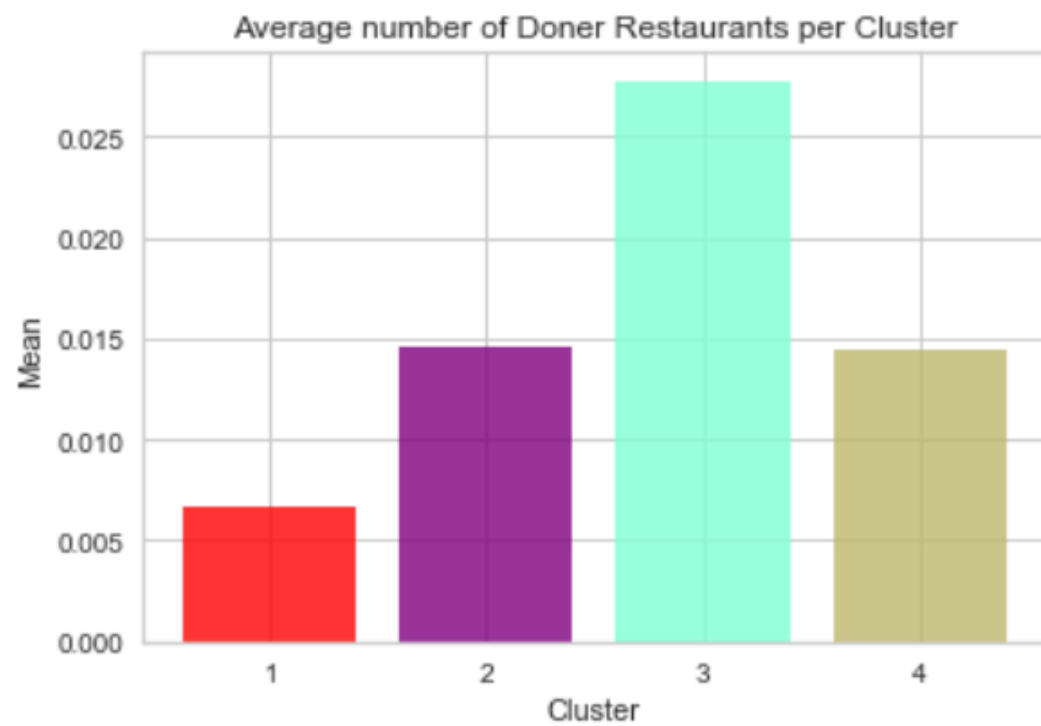
**Plot the clusters on a Map of the Toronto and Super Impose the best location of a Store:**

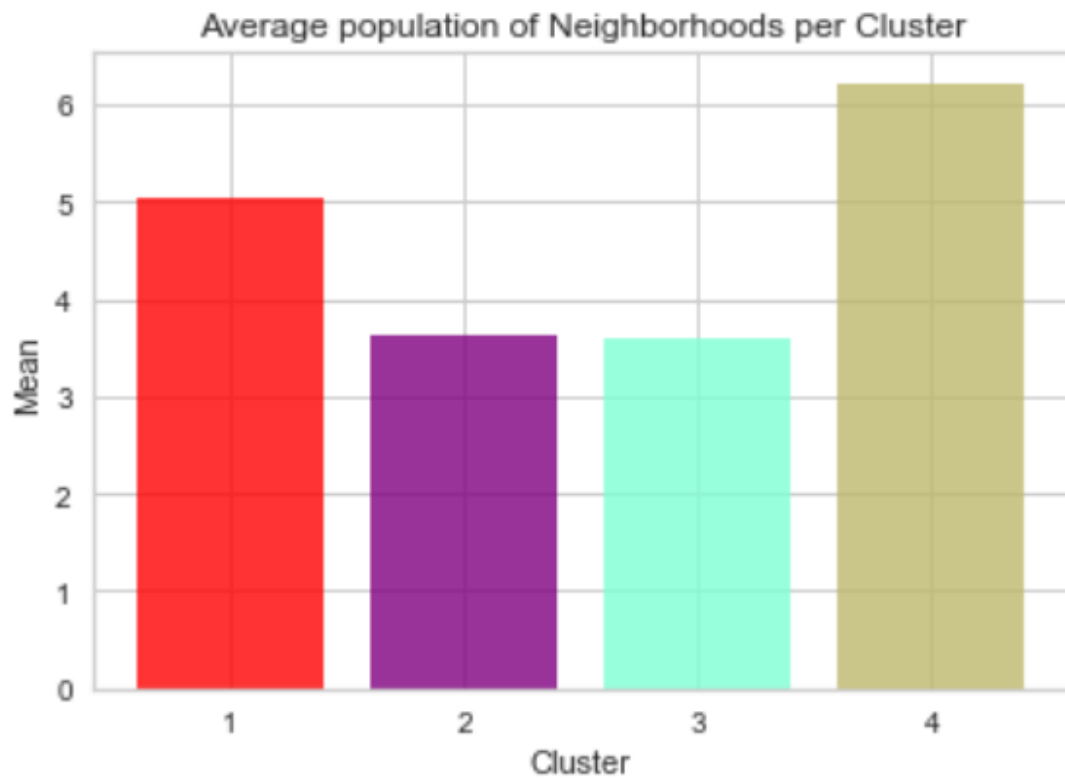


**Analyze clusters according to the different input dimensions:**









Cluster 3 has the most Döner Kebab restaurants and is excluded from further consideration due to the competitive situation in this cluster. Clusters 2 and 4 have about the same number of kebab restaurants and cluster 1 the least. Accordingly, cluster 1 is considered further. Cluster 1 has a medium average income, which fits well with kebab restaurants. In addition, the population density in cluster 1 is the highest, which speaks for a potentially good occupancy rate of a new kebab restaurant in one of these neighborhoods.

## Now let's have a deeper look into the Neighborhoods in Cluster 1:

	Borough	Neighborhood	Doner Restaurant	Population_per_sqare_kilometer	Average monthly income in €	Cluster Labels
0	Mitte	Mitte	0.000000	9.713	2025	0
1	Mitte	Moabit	0.083333	9.713	2025	0
2	Mitte	Hansaviertel	0.000000	9.713	2025	0
3	Mitte	Tiergarten	0.000000	9.713	2025	0
4	Mitte	Wedding	0.000000	9.713	2025	0
5	Mitte	Gesundbrunnen	0.017544	9.713	2025	0
6	Friedrichshain-Kreuzberg	Friedrichshain	0.016129	14.374	2075	0
7	Friedrichshain-Kreuzberg	Kreuzberg	0.000000	14.374	2075	0
28	Spandau	Spandau	0.024390	2.671	2050	0
29	Spandau	Haselhorst	0.000000	2.671	2050	0
30	Spandau	Siemensstadt	0.000000	2.671	2050	0
31	Spandau	Staaken	0.000000	2.671	2050	0
32	Spandau	Gatow	0.000000	2.671	2050	0
33	Spandau	Kladow	0.000000	2.671	2050	0
34	Spandau	Hakenfelde	0.000000	2.671	2050	0
35	Spandau	Falkenhagener Feld	0.000000	2.671	2050	0
36	Spandau	Wilhelmstadt	0.000000	2.671	2050	0
71	Marzahn-Hellersdorf	Marzahn	0.000000	4.434	2100	0
72	Marzahn-Hellersdorf	Biesdorf	0.000000	4.434	2100	0
73	Marzahn-Hellersdorf	Kaulsdorf	0.000000	4.434	2100	0
74	Marzahn-Hellersdorf	Mahlsdorf	0.000000	4.434	2100	0
75	Marzahn-Hellersdorf	Hellersdorf	0.000000	4.434	2100	0
86	Reinickendorf	Reinickendorf	0.000000	2.975	2125	0
87	Reinickendorf	Tegel	0.000000	2.975	2125	0
88	Reinickendorf	Konradshöhe	0.000000	2.975	2125	0
89	Reinickendorf	Heiligensee	0.000000	2.975	2125	0
90	Reinickendorf	Frohnau	0.076923	2.975	2125	0
91	Reinickendorf	Hermisdorf	0.000000	2.975	2125	0
92	Reinickendorf	Waldmannslust	0.000000	2.975	2125	0
93	Reinickendorf	Lübars	0.000000	2.975	2125	0
94	Reinickendorf	Wittenau	0.000000	2.975	2125	0
95	Reinickendorf	Märkisches Viertel	0.000000	2.975	2125	0
96	Reinickendorf	Borsigwalde	0.000000	2.975	2125	0

Cluster 1 sorted by number of kebab restaurants, population density, and income to derive a recommendation for a suitable location:

	Borough	Neighborhood	Doner Restaurant	Population_per_sqare_kilometer	Average monthly income in €	Cluster Labels
7	Friedrichshain-Kreuzberg	Kreuzberg	0.000000	14.374	2075	0
0	Mitte	Mitte	0.000000	9.713	2025	0
2	Mitte	Hansaviertel	0.000000	9.713	2025	0
3	Mitte	Tiergarten	0.000000	9.713	2025	0
4	Mitte	Wedding	0.000000	9.713	2025	0
71	Marzahn-Hellersdorf	Marzahn	0.000000	4.434	2100	0
72	Marzahn-Hellersdorf	Biesdorf	0.000000	4.434	2100	0
73	Marzahn-Hellersdorf	Kaulsdorf	0.000000	4.434	2100	0
74	Marzahn-Hellersdorf	Mahlsdorf	0.000000	4.434	2100	0

**As we can see in the sorted list (sorted by number of kebab restaurants, population density and income), Kreuzberg should be a suitable choice for a new kebab restaurant.**

## **Conclusion**

I feel confident with the recommendation I gave as it is backed up with demonstrated data analysis. There are certainly other data that can be considered in a more in-depth analysis. For example, data on the ethnic composition of the neighborhoods could be obtained. Likewise, data on tourist offerings and accommodations in the neighborhood would be interesting. These data are certainly related to the eating habits and preferences of people in a neighborhood. Nevertheless, the evaluation that has been done allows for a good enough recommendation.