# Meat-Free Kebab Shop Location Report

Joshua Stevenson-Hoare 2019-09-26

## 1 Introduction

### 1.1 Background

Berlin is known as one of the vegan food capitals of the world, with nearly any kind of food available in meat-free variants. However, like many German cities, it also has an obsession with Döner Kebabs. This represents a unique business opportunity that has been largely unexplored. At present there are very few entirely meat-free kebab shops in Berlin. Those that do exist cater largely to the vegan, vegetarian, and environmentally conscious markets. There is an untapped market however.

The late-night customer often leaves a social event in the late evening or early hours of the morning seeking greasy salty food to sate their hunger. Often tired and inebriated, this customer wants fast and friendly service with a minimum of frills. Current meat-free kebab offerings can provide some of these, but are not typically open late enough to take advantage of this customer base.

This is potentially interesting to owners of current vegan restaurants/cafes, looking to expand into the late-night market. It is also potentially interesting to owners of late-night fast food outlets looking to attract customers who do not eat meat.

### 1.2 Problem Specifics

The business problem at hand is to identify the optimal areas of Berlin to open a late-night meat-free kebab shop. A good area to open a meat-free kebab shop is one that fulfils the following criteria.

- An area with lots of late-night establishments such as bars or clubs, as these are our target customer base
- An area where late-night fast food places already exist, as these will have good footfall

There is also a question of whether to place the meat-free kebab shop near other vegan, vegetarian, or meat-free food outlets. This may help with attracting customers, but if the location is far from bars or clubs then it will hurt footfall. So there are three questions.

1) What areas of Berlin have a high number of both late-night drinking venues and kebab outlets?

2) Are these areas overlapping with areas with high numbers of vegan (etc.) venues?

3) What is the optimal area(s) to open a meat-free döner kebab shop in Berlin?

### 1.3 Interest

This is potentially interesting to owners of current vegan restaurants/cafes, looking to expand into the late-night market. It is also potentially interesting to owners of late-night fast food outlets looking to attract customers who do not eat meat.

# 2 Data

## 2.1 Data Sources

Firstly, location data will be acquired using the Wikipedia library for Python. This will obtain a list of all areas in Berlin, and then use this to access the coordinates of these areas.

Secondly, the venue data for bars and kebab shops will be acquired using the FourSquare API. This will use the coordinates from the location data to search for venues in each region of Berlin. The venue data for vegan and vegetarian venues will use a list from HappyCow, which will then be input into FourSquare to get the location data. The total venue data will be refined into the categories of interest to generate regional information.

## 2.2 Data Analysis

Exploratory analysis will be carried out to plot all the venues fitting one of our three categories (bars,kebabs,vegan) in Berlin. This will give us an idea of the distribution of these venues.

Correlations will be performed to see if areas with higher numbers of each category also have higher numbers of the other category. This will allow us to see any linear relationships between the venue categories.

Then, clusters of similar areas will be generated using k-means clustering. This will find areas that are similar and areas that are dissimilar. Using our critera we can then suggest what areas of Berlin are optimal in which to open a meat-free kebab shop.

# 3 Methodology

## 3.1 Data Acquisition & Wrangling

### 3.1.1 Location Data

Regional data for Berlin were acquired by scaping the Wikipedia page for the boroughs and neighbourhoods of Berlin. This was used to a create a dataframe containing all localities (equivalent to an American neirhbourhood) within the Berlin region of Germany.

This list of localities names was then populated with the coordinates for the centre of each localities by using the name to find listed coordinates on the Wikipedia page for each locality. It was necessary at this point to alter the listed names of some localities as some shared a name with either their parent borough (eg. Mitte), an unrelated term (eg. Wedding), or with locations elsewhere in Berlin or Northern Europe (eg. Falkenberg). The correct coordinates were verified by visualising all points on a map and ensuring that they all fell within the boundaries of Berlin.

### 3.1.2 Venue Data

The data on venues came from two sources – FourSquare and HappyCow. Foursquare was used for all bar and kebab venues. HappyCow was used for all vegan/vegetarian venues as it was revealed during piloting work that FourSquare has poor labelling for the types of products sold by each venue.

Using the FourSquare API was relatively simple. Venues were collected within a given 750m radius of each locality coordinate pair, using search queries that contained terms relevant to either bars or kebab shops. These were then refined to remove erroneous results (such as venues whose name included the letters ‘bar’ but were not alcohol serving venues).

Using the information from HappyCow was somewhat more convoluted. Initially a search was performed on the HappyCow website for vegan and vegetarian venues within the Berlin area. This search was then saved as a HTML file, and further processed into a text list of venues and addresses.

This list was then sorted into two separate lists of names and addresses that could be paired up. A geocoding package was then used to the find the coordinates for each listed address. These were repaired with the names of each venue.

The two venue databases were then combined together to create one database of all venues to be used in this analysis.

### 3.1.3 Data Wrangling

The resulting venue dataframe was sorted by the type of venue that each represented. These were then rearranged to form another database that could be used for further analysis.

## 3.2 Data Visualisation

A folium map was created displaying all the venues in the dataframe, overlaid on a map of Berlin (Figure 1). It can be see than the bars and kebab shops are more spread around Berlin, whereas the vegan and vegetarian venues show more clustering. However, all types of venues can be found across the majority of localities in Berlin.

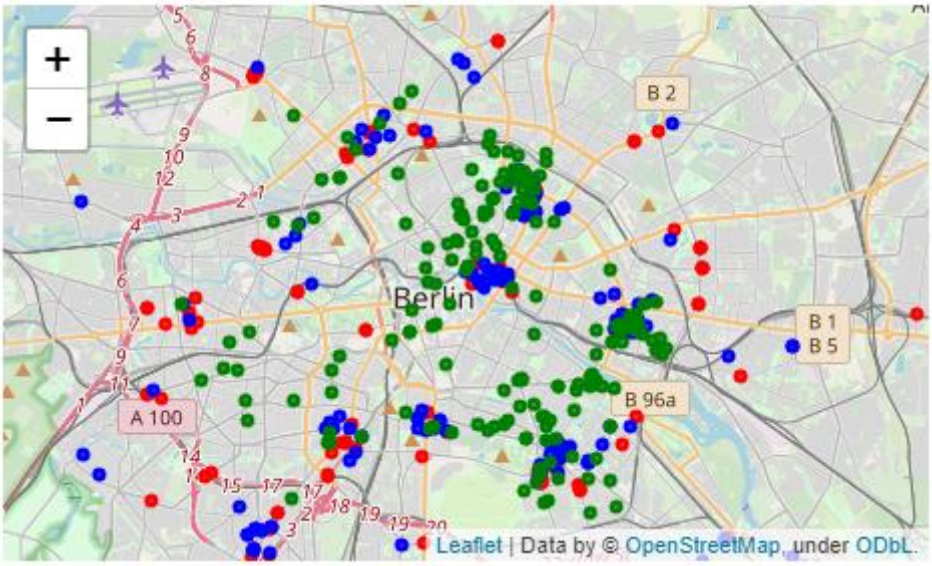


Figure 1. All venue locations overlaid on a map of Berlin. Blue dots are bar venues, red dots are kebab venues, and green dots are vegan/vegetarian venues.

## 3.3 Correlation of Regional Data

Correlations were performed between the number of each venue type within each locality, in an effort to determine how these venue types related to each other. Positive correlations were found between all venue types, indicating that the more of any venue there is, the more of the other types will also appear. This is not unexpected as this can be taken as a marker of venue density as a whole. More populous areas should be expected to have more venues of all kinds.

The strongest correlations were between Bars and Vegan venues ($r$ = 0.917), and the weakest between Kebab venues and Vegan venues ($r$ = 0.634). The correlation between Bars and Kebab venues was slightly strong than that between Kebab and Vegan ($r$ = 0.647). All correlations were significant, $p$ <0.001.

## 3.4 Clustering Analysis

A k-Means model was chosen for this analysis. This is because this model allows unsupervised categorisation of data, based on a relatively small number of features. Here the chosen features were the number of each venue type within each locality.

The Elbow method was employed to determine the best value of *k* to use, this can be seen in Figure 2. This method uses increasing values of *k*, and creates a series of models. The cost for each model (the distance between each point and its assigned centroid) is calculated and plotted on a graph. This should have the effect of a line which decreases rapidly at first, and then tends towards asymptote. A bilinear fit shows where the 'elbow' is, and this is the value of *k* above which there is diminishing returns in cost by increasing the number of clusters.

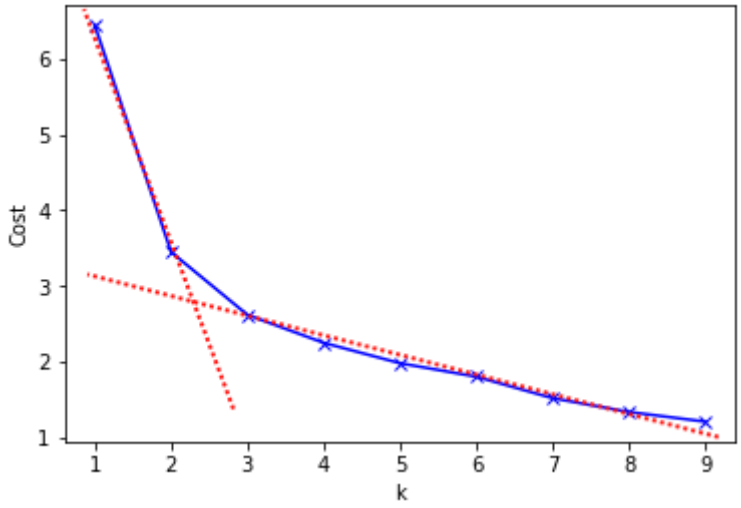


Figure 2. The results of using the Elbow Method to determine *k*. The red lines indicate the limbs of the bilinear fit.

The value of *k* suggested from the Elbow Method was between 2 and 3, so three clusters were used as this was the larger of the pair, and so deemed to be less likely to underfit the data.

# 4 Results

## 4.1 Cluster Description

The k-Means model was successfully fit to our data, generating three clusters of localities based on their number of venues in each category within each locality. These clusters were initially examined based on the average feature value per cluster.

| Cluster # | Mean # Bars | Mean # Kebabs | Mean # Vegan |
|---|---|---|---|
| 0 | 1.2 | 2.0 | 0.4 |
| 1 | 20.5 | 9.3 | 27.0 |
| 2 | 5.6 | 4.2 | 9.4 |

Based off this information the clusters found were assigned the following descriptions:

0) Low number of all venues
1) Many bars and vegan, some kebabs
2) Moderate number of all venues

## 4.2 Cluster Visualisation

These clusters were then visualised on a folium map (Figure 3). This showed that the clusters also had some geographical correspondence. Clusters 1 and 2 were mainly centrally located within

Berlin, whereas cluster 0 consisted mostly of the outlying areas of Berlin – though there were some cluster 0 areas within central Berlin too.

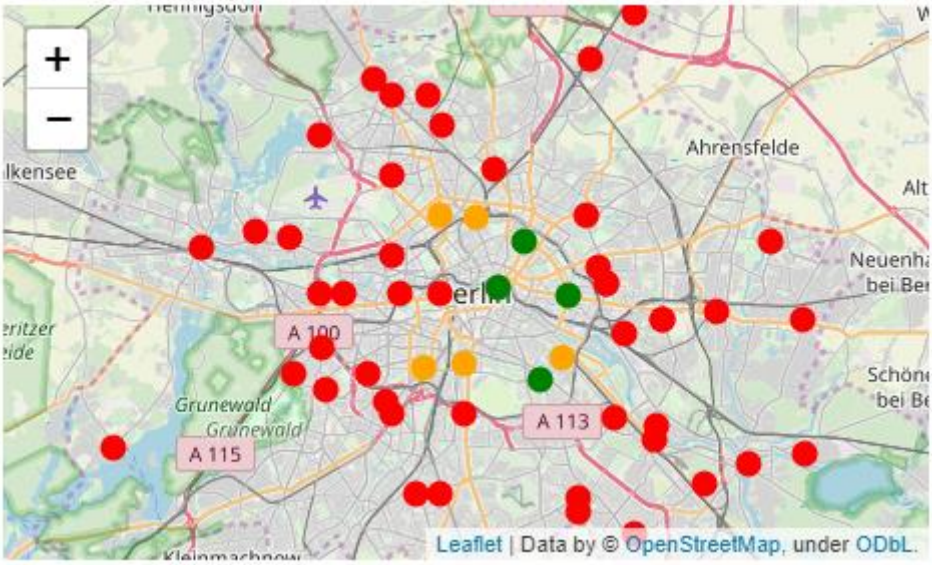


Figure 3. Clusters visualised on a map of Berlin. Red dots are cluster 0, green dots are cluster 1, and orange dots are cluster 2.

# 5 Discussion

## 5.1 Answering the Questions

Our aim was to answer three questions:

> 1) What areas of Berlin have a high number of both late-night drinking venues and kebab outlets?
>
> 2) Are these areas overlapping with areas with high numbers of vegan (etc.) venues?
>
> 3) What is the optimal area(s) to open a meat-free döner kebab shop in Berlin?

From our results we can now provide informed answers to these questions.

Areas in clusters 1 and 2 contains large numbers of bars/other drinking venues that are open late at night, and also kebab shops. These are primarily located around central Berlin.

These areas do have overlap with areas containing high numbers of vegan venues. This is more true of areas in cluster 1, but there is still a moderate amount of vegan venues in areas in cluster 2.

Our results suggest that the optimal place to open a meat-free döner kebab shop in Berlin would be areas within cluster 1. These are Prenzlauer Berg, Mitte, Friedrichshain, and Neukölln. These areas have a high concentration of late-night drinking venues, providing a large potential customer base. They do have some kebab shops, ensuring that there is footfall, and a high concentration of vegan venues which is evidence that consumers in these areas have an interest in meat-free alternatives.

## 5.2 Notes on this Analysis

**Acquisition of Information on Vegan Venues**

Information on vegan/vegetarian venues was sourced not from FourSquare, but from HappyCow. This resource is a free to access crowd-sourced repository of information on venues that are solely aimed at, or offer options for, vegan and vegetarian consumers. The choice to use this resource instead of FourSquare was based on the sparsity of information available from FourSquare. Initial exploration provided only 4 venues on the FourSquare database that were categorised as vegan/vegetarian. Although certainly more existed there was no easy way to determine the content of a venue's selection.

There was some complications in accessing the information from HappyCow as they do not provide an API. Fortunately, there is a searchable website, whose output can be parsed using html-reading methods. For ease of use and replicability (the URL not being stable), this information was preprocessed slightly. This caused a small bottleneck in the creation of this document, but was not sufficient to counteract using more complicated means.

**Variability in k-Means output**

As the k-Means method by default uses random starting points, without some control the method's output is not guaranteed to be the same each time. Furthermore, different values of *k* (the number of clusters) can generate different clusters too. I opted to use the elbow method to provide some guidance on what value to assign to *k* here. This suggested an elbow of between 2 and 3. As *k* must be a positive integer, I opted for 3 as 2 seemed too low to be of informational value. I could have reduced the distances to each cluster mean by using higher values of *k* but this would have been subject to diminishing returns and may have led to overfitting.

**Knowledge of the local area**

It should be pointed out that I am not German, nor have I ever been to Berlin. The only relevant local knowledge that I had prior to this analyis was that doner kebabs are popular in Germany in general, and that Berlin has a large number of vegan restaurants. Other than a short piece of research to determine what regionalisation is present in Berlin, I did not need to know any more information about the area to perform this method of analysis. If you look at the maps generated the clusters are arranged very geographically. Given that the data input had no location information whatsoever, this is just an outcome of the similarity between the localities within Berlin.

# 6 Conclusion

This report looked at a fictional scenario where a client seeks information on where to open a restaurant specialising in meat-free doner kebabs, in Berlin. In order to do this, location and venue data was acquired from multiple sources to generate a profile of the localities within Berlin. From this, clustering classification was performed to provide a suggestion as to an area or areas which would provide a large base of customers who fit the target demographic (late-night, drinking, interest in meat-free alternatives).

The results of the report suggested four areas, all within central Berlin, where the profile of venues in that locality would suggest that it would be an optimal location to open a meat-free doner kebab shop. These areas are Prenzlauer Berg, Mitte, Friedrichshain, and Neukölln.

The analysis employed here was very successful as it produced a concrete suggestion for the client, using a wholly naive system. A client would be able to use this information to leverage the consumer base within Berlin without requiring local knowledge. This method could be extended to other cities or other fast-food products with relative ease.