Exploring food venues nearby Berlin underground metro stations (U-Bahn)

by Minh Mai
Berlin, June 2020

1. Introduction

Berlin is the capital and largest city of Germany by both area and population. It hosts an exciting art and nightlife scene and, best of all, it is very affordable to live in. Whether you're a *Nachtschwärmer* or not, you definitely are looking for a place to fill your stomach, and you can get easily lost in the decision of which place suits you best. This project aims to explore and analyse food venues surrounding the public transport stations in Berlin and describe different clusters of the stations.

The Berlin-all-inclusive public transport system consists of U-Bahn (underground railway), S-Bahn (elevated railway), buses, and trams, which are managed by The Berliner Verkehrsbetriebe (BVG) and the Deutsche Bahn (DB). Within the scope of this project I only focused on exploring and clustering food venues surrounding the U-Bahn with 173 stations. The results can be useful for those like going out to eat via the U-Bahn in Berlin. They provide the idea of what types of food surrounding each station.

2. Data

In order to segment the Berlin U-Bahn stations, a dataset is required that contains the name of the stations, with respective latitude and longitude coordinates. This dataset is downloaded from Wikipedia: Liste der Berliner U-Bahnhöfe.

By using *BeautifulSoup* library the needed table and column were scraped from the Wikipedia page. The retrieved dataset is then transformed into a pandas dataframe, by looping through the data and filling the dataframe rows one at a time. Then I removed duplicate stations since one station can have multiple lines going through. After that, I converted the original coordinates from degrees/minutes/seconds to decimal for later mapping. As a result, a dataframe called *df_ubahn* is created with columns Station, Latitude and Longitude details of every station.

	Station	Latitude	Longitude
0	Adenauerplatz	52.499722	13.307222
1	Afrikanische Straße	52.560556	13.334167
2	Alexanderplatz	52.521389	13.413333
3	Altstadt Spandau	52.539167	13.205556
4	Alt-Mariendorf	52.439722	13.387500

Figure 1. The first 5 rows of cleaned dataset.

3. Methodology

In order to have an overview of the U-Bahn stations' location, I created a map using *Folium* library. First, I started finding geographical coordinates for the city Berlin by using geopy library. Then I visualized the U-Bahn stations on a map with latitudes and longitudes of every stations from the dataframe *df_ubahn*.

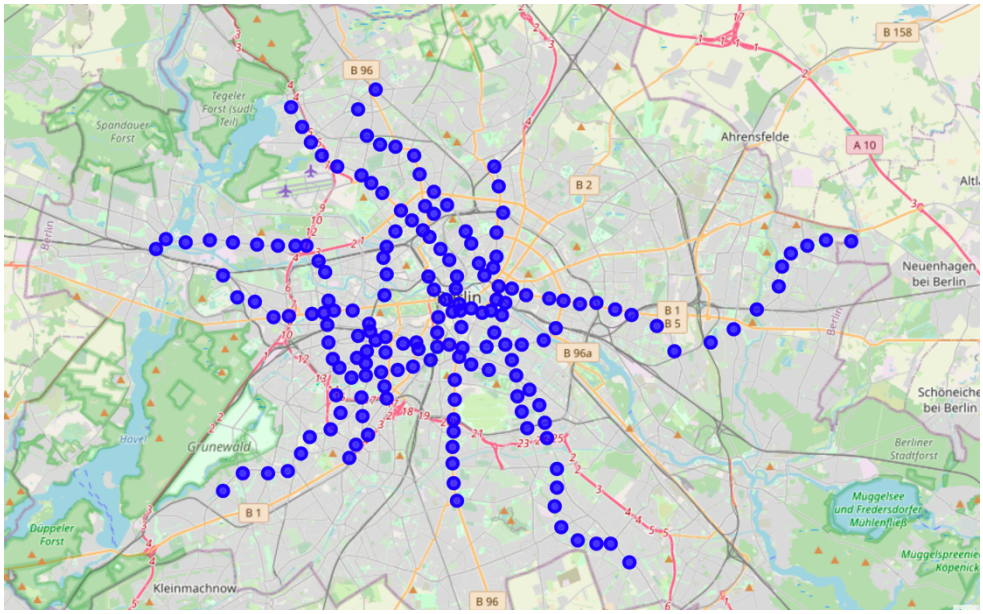

Figure 2. Berlin U-Bahn stations map.

It's time to explore the food venues around every U-Bahn station. Firstly, I defined Foursquare credentials and version. Then I used them to fetch Foursquare Venue Category Hierarchy to get CategoryId of Food Venues. Next, I created a function to extract food venues around all the U-Bahn stations, their coordinates and then transformed retrieved data to a pandas dataframe named *food_ubahn*.

	Station	U-Bahn Latitude	U-Bahn Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adenauerplatz	52.499722	13.307222	Bellucci	52.499430	13.306800	Italian Restaurant
1	Adenauerplatz	52.499722	13.307222	Frau Behrens Torten	52.501653	13.307663	Café
2	Adenauerplatz	52.499722	13.307222	SAVU	52.499323	13.305033	Modern European Restaurant
3	Adenauerplatz	52.499722	13.307222	Block House	52.499846	13.306933	Steakhouse
4	Adenauerplatz	52.499722	13.307222	Kurpfalz Weinstuben	52.500898	13.307718	German Restaurant

Figure 3. The first 5 rows of dataframe food_ubahn.

The resulting dataframe *food_ubahn* has a total of 4630 venues and 122 unique categories. I realized that there are not relevant categories (e.g. Café, Bakery), or not clear (too generalized) categories (e.g. Restaurant, Diner), so I decided to get rid of them. Now the dataframe shrank to 101 unique food categories.

After using one hot encoding and taking the mean of the frequency for each food venue category, I used K-means clustering, an unsupervised learning algorithm used to create K clusters of data points based on feature similarity.

	Station	Afghan Restaurant	African Restaurant	American Restaurant	Argentinian Restaurant	Asian Restaurant	Austrian Restaurant	BBQ Joint	Bagel Shop	Bavarian Restaurant
0	Adenauerplatz	0.0	0.0	0.036364	0.000000	0.054545	0.018182	0.000000	0.018182	0.000000
1	Afrikanische Straße	0.0	0.0	0.000000	0.333333	0.000000	0.000000	0.000000	0.000000	0.000000
2	Alexanderplatz	0.0	0.0	0.000000	0.000000	0.050000	0.000000	0.016667	0.000000	0.016667
3	Alt-Mariendorf	0.0	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	Alt-Tegel	0.0	0.0	0.000000	0.000000	0.043478	0.000000	0.000000	0.000000	0.000000

Figure 4. U-Bahn grouped dataframe showing mean of the frequency for the first 9 categories.

4. Analysis

Observing the elbow method and silhouette score, I found that the optimal number of clusters is K = 4.

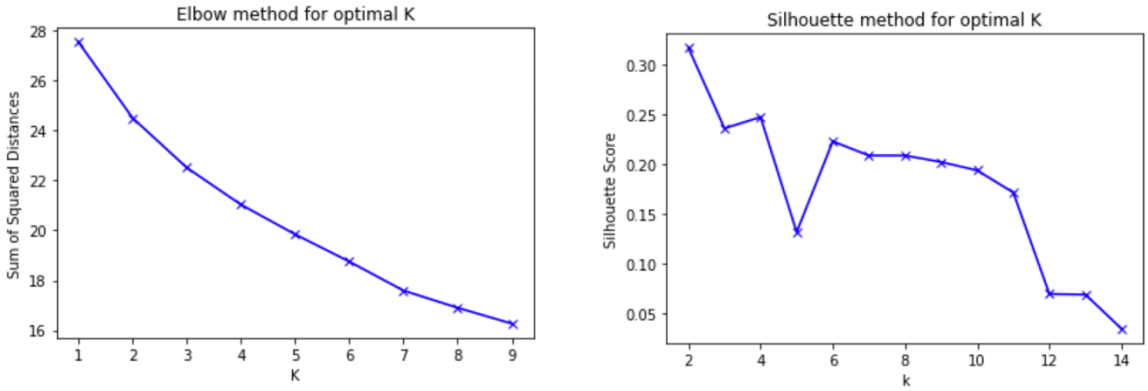

Figure 5. Elbow and silhouette method for optimal K.

I then ran k-means to cluster the stations into 4 clusters and created a new dataframe that include the cluster as well as the top 10 venues for each station. Finally, I visualized the resulting clusters into a map:

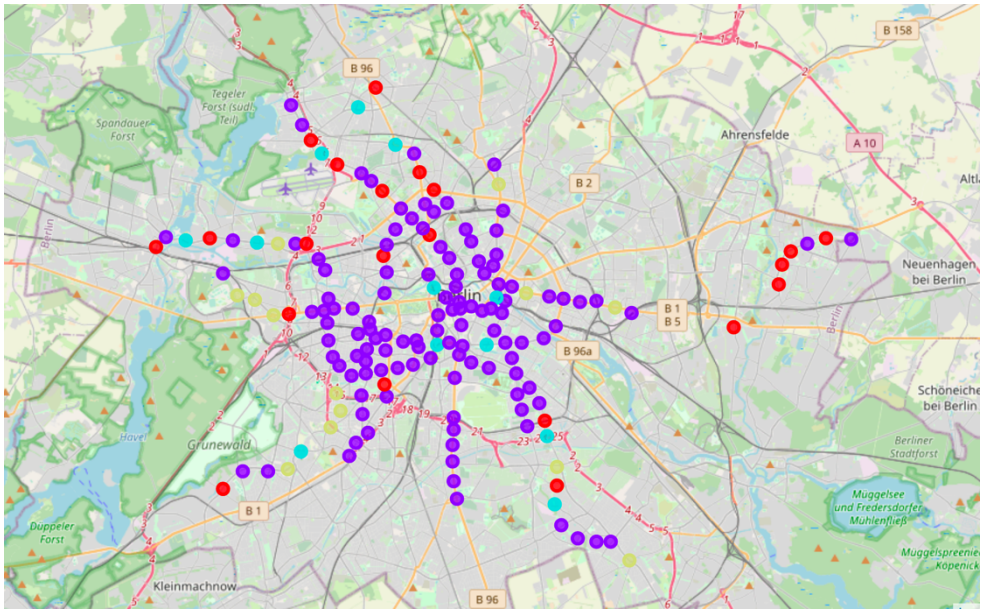

Figure 6. Clusters map for the U-Bahn stations.

5. Results

By reaching the end of the analysis, we got a sneak peak of a total of 173 U-Bahn stations and food venues surrounding them. I've scraped data from web resources like Wikipedia, python libraries like Geopy, and Foursquare Venue API, to set up a realistic data-analysis scenario. I could generalize the clusters as following:

- **Cluster 1** (*red*): Mostly Döner restaurants can be found at these stations, which are located outside the Ringbahn .There are also Italian and Asian Restaurants nearby.

	Station	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
12	Birkenstraße	Doner Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Taverna	French Restaurant	German Restaurant	Austrian Restaurant	Turkish Restaurant	Burger Joint	Egyptian Restaurant
24	Cottbusser Platz	Doner Restaurant	Italian Restaurant	Snack Place	Yemeni Restaurant	Franconian Restaurant	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant
29	Elsterwerdaer Platz	Doner Restaurant	Italian Restaurant	Snack Place	Asian Restaurant	French Restaurant	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant
34	Franz-Neumann-Platz	Doner Restaurant	Pizza Place	French Restaurant	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fish & Chips Shop
48	Haselhorst	Doner Restaurant	BBQ Joint	Fried Chicken Joint	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fish & Chips Shop

- **Cluster 2** (*purple*): This cluster has the most restaurants surrounding U-Bahn stations, which are located mostly in the city centre. Dominating around those stations are Italian and Vietnamese Restaurants (both 1st and 2nd venues). German Restaurants also have a significant presence in this cluster.

	Station	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
0	Adenauerplatz	Italian Restaurant	Vietnamese Restaurant	Steakhouse	Asian Restaurant	German Restaurant	Chinese Restaurant	American Restaurant	Sushi Restaurant	Mediterranean Restaurant	Greek Restaurant
1	Afrikanische Straße	Argentinian Restaurant	Indian Restaurant	Seafood Restaurant	Yemeni Restaurant	Franconian Restaurant	Doner Restaurant	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant
2	Alexanderplatz	Vietnamese Restaurant	Italian Restaurant	German Restaurant	Burger Joint	Sushi Restaurant	Sandwich Place	Asian Restaurant	Indian Restaurant	Thai Restaurant	Burrito Place
3	Altstadt Spandau	Italian Restaurant	German Restaurant	Seafood Restaurant	Pizza Place	Doner Restaurant	Thai Restaurant	Turkish Restaurant	Vietnamese Restaurant	Sushi Restaurant	Dumpling Restaurant
4	Alt-Mariendorf	Steakhouse	Pizza Place	Italian Restaurant	Chinese Restaurant	Greek Restaurant	German Restaurant	Yemeni Restaurant	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant

- **Cluster 3** (*cyan*): Surrounded mostly by German Restaurants.

	Station	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
20	Britz-Süd	Vegetarian / Vegan Restaurant	German Restaurant	Yemeni Restaurant	French Restaurant	Doner Restaurant	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant
23	Bundestag	German Restaurant	Burger Joint	BBQ Joint	Modern European Restaurant	Pizza Place	French Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant
25	Dahlem-Dorf	German Restaurant	Burger Joint	Asian Restaurant	Pizza Place	Fried Chicken Joint	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant
40	Gleisdreieck	German Restaurant	Yemeni Restaurant	French Restaurant	Doner Restaurant	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant
43	Grenzallee	Pizza Place	German Restaurant	Yemeni Restaurant	French Restaurant	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant

- **Cluster 4** (*olive*): Italian Restaurants are again in the lead.

	Station	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
14	Blaschkoallee	Italian Restaurant	Eastern European Restaurant	Pizza Place	Yemeni Restaurant	French Restaurant	Dumpling Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant
19	Breitenbachplatz	Italian Restaurant	Mexican Restaurant	Yemeni Restaurant	French Restaurant	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant
36	Freie Universität	Italian Restaurant	Asian Restaurant	Yemeni Restaurant	French Restaurant	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant
51	Heidelberger Platz	Italian Restaurant	Yemeni Restaurant	French Restaurant	Doner Restaurant	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant
83	Magdalenenstraße	Italian Restaurant	Doner Restaurant	Pizza Place	French Restaurant	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant

6. Conclusion

To conclude, I used K-means clustering which created 4 different clusters for Berlin U-Bahn stations having food venues surrounding around them. Whenever you have a chance to visit Berlin, I hope this map will give you an imagination of where to eat nearby the U-Bahn stations.

If time allows, I'd like to repeat the clustering for the S-Bahn which has 166 stations, since it's also relevant in the context of food venues around public transportation.

This project is a part of the capstone project within IBM Data Science Professional Specialization. Python code with more details about the data and methodology are available on NBViewer.

Feel free to find me on LinkedIn and GitHub.