Analyse eines Hydraulikkontaktierungsprozesses mittels Data Mining

Coursera
Applied Data Science Capstone by IBM
Presentation Battle of Neighborhoods
(week 2)

Willi Gwildies

Introduction and Business Problem

Looking for restaurant locations in Berlin:

- Berlin is relativily big an wide broad
- Twelve boroughs made up of 96 neighborhoods

➔ How do they know what kind of food they get in the different neighborhoods of Berlin?

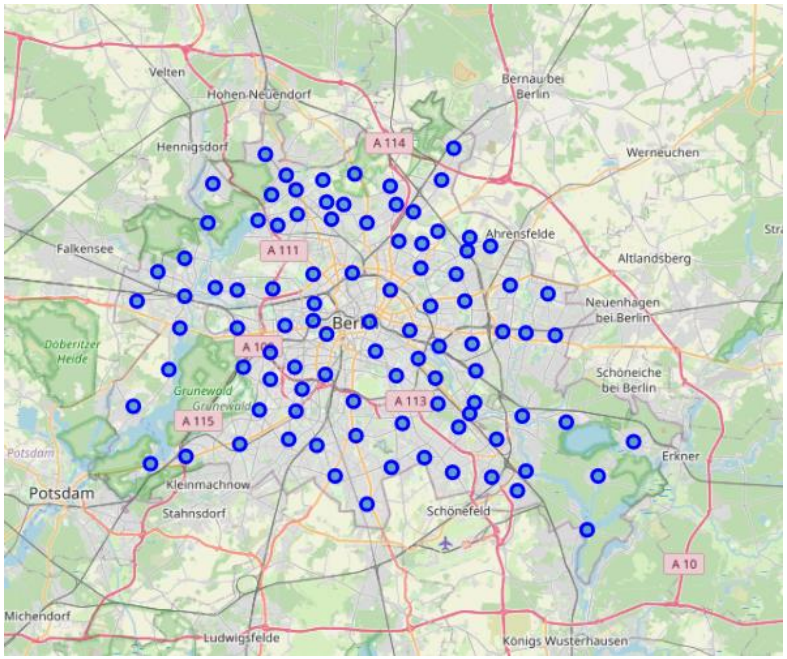

Figure 1: Map of Berlin with it‘s 96 neighborhoods

Data Acquisition and Approach

Data Required:

- Utilizing the Foursquare API to pull the following location data on restaurants in Berlin, Germany:
  - Venue Name
  - Venue ID
  - Venue Location
  - Venue Category
  - Rating
  - Price

Data Acquisition Approach:

- To acquire the data mentioned above, I will need to do the following:
  - Get geolocator lat and long coordinates for neighborhoods
  - Get a list of all venues
  - Get venue IDs for each venue
  - Pull venue name, location, category, rating, price

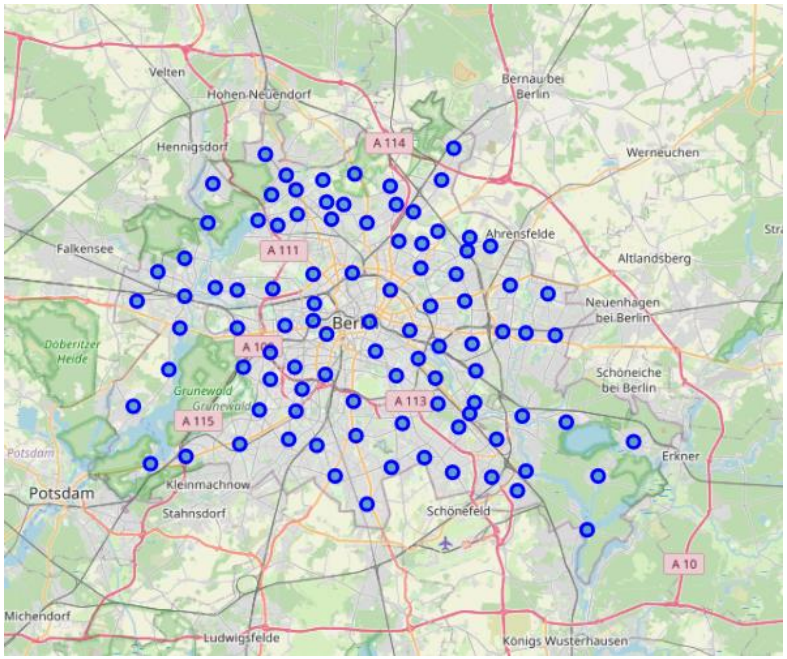

Figure 1: Map of Berlin with it‘s 96 neighborhoods

Methodology

- Using the geodata to get all restaurant for each neighborhood:

Table 1: Table of restaurants per neighborhood

	Neighborhoods	Neighborhoods Latitude	Neighborhoods Longitude	Venue ID	Venue	Venue Latitude	Venue Longitude	Venue Category
15	Mitte	52.517885	13.404060	5a958ec0e4c459472938359f	Wilde Matilde	52.517475	13.405384	German Restaurant
17	Mitte	52.517885	13.404060	584c882dd702824c51e2be9a	Balthazar	52.515913	13.406160	Restaurant
27	Mitte	52.517885	13.404060	59eb78736bdee6069ed97d77	EL COLMADO	52.519412	13.409681	Spanish Restaurant
33	Tiergarten	52.509778	13.357260	4cc6c1d5c844721ea24ef601	Kantine im Felleshus	52.508555	13.350813	Scandinavian Restaurant
38	Tiergarten	52.509778	13.357260	4cdae7f2930af04dfbb08797	Eventlocation Alte Pumpe	52.505481	13.358203	German Restaurant

- Using one hot encoding, for each neighborhood the most common categories of restaurant are calculated:

Table 2: Table of top 5 restaurants per neighborhood

	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Adlershof	Mediterranean Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Oriental Restaurant
1	Alt-Hohenschönhausen	Mediterranean Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Oriental Restaurant
2	Alt-Treptow	Seafood Restaurant	Mediterranean Restaurant	Italian Restaurant	Central and South American Food	Asian Restaurant
3	Blankenburg	Mediterranean Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Oriental Restaurant	Italian Restaurant
4	Bohnsdorf	Italian Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Oriental Restaurant	Mediterranean Restaurant

Methodology

- Using this data the best number of clusters is determined. Using the "elbow method" the number of clusters is set to k=4

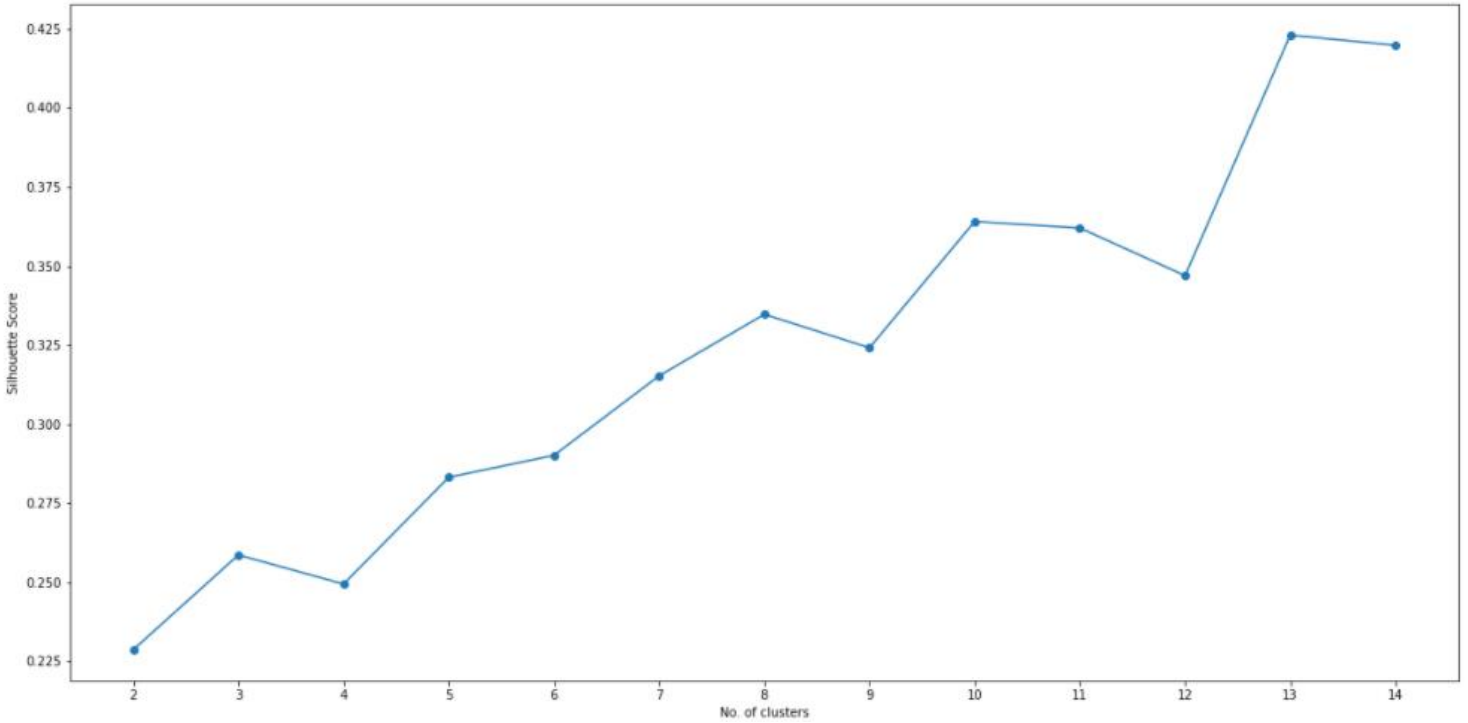

Figure 2: Silhouette Score over number of clusters

Results

- First cluster mainly Asian Food (red marks)
- Second cluster mainly Oriantal Resaurants or Doner Places (purple marks)
- Third cluster mainly Italian and Mediterranean Restaurants (turquoise marks)
- Fourth cluster mainly German cuisine or vegetarian/vegan food (green marks)

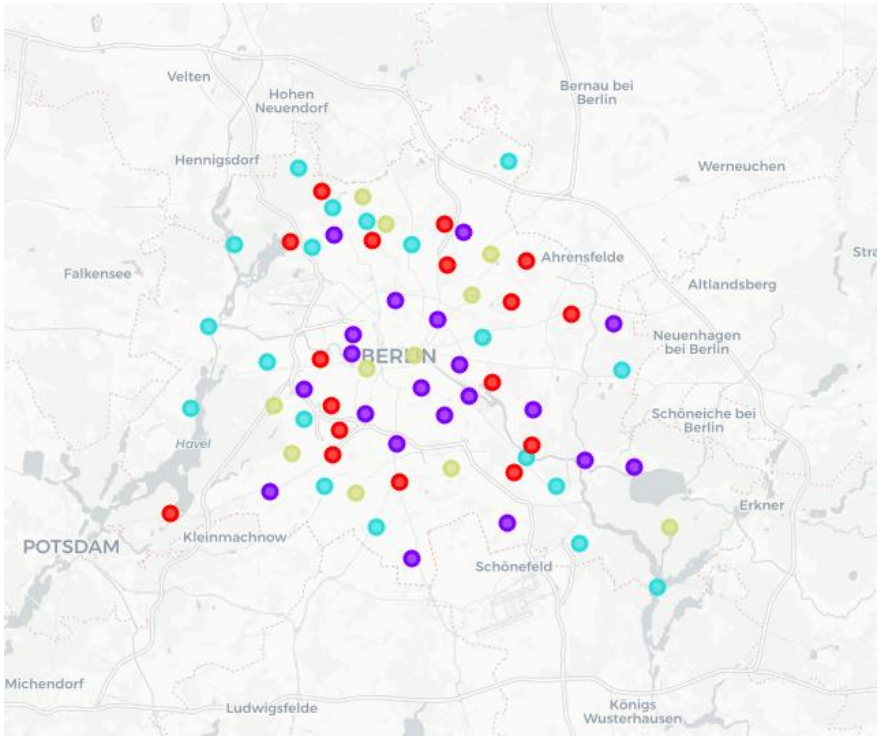

Figure 3: Map of found clusters in Berlin

Coursera
Applied Data Science Capstone by IBM
Presentation Battle of Neighborhoods
(week 2)

Thank you