Berlin&Hamburg Battle of the neighborhoods

IBM DATA SCIENCE PROFESSIONAL CERTIFICATE, CAPSTONE PROJECT

Table of contents

Berlin vs. Hamburg

Berlin and Hamburg are, seemingly, two very different cities. While Berlin is considered the capital of culture and arts, Hamburg has a very corporate feel to it.

We will try to analyse the neighbozrhoods by ZIP-Codes, to find areas with similarities.

This analysis could be helpfull for corporations thinking of expanding or moving from one city to the other, or for individuals thinking of moving in either direction.

Data acquisition

ZIP-Code and coordinate data is imported from a json file that can also be found on github https://github.com/zauberware.

Venue data is scraped using the Foursquare developer API.

In total, 16481 rows and 11 columns in dataset.

Irrelevant ZIP-Codes were dropped.

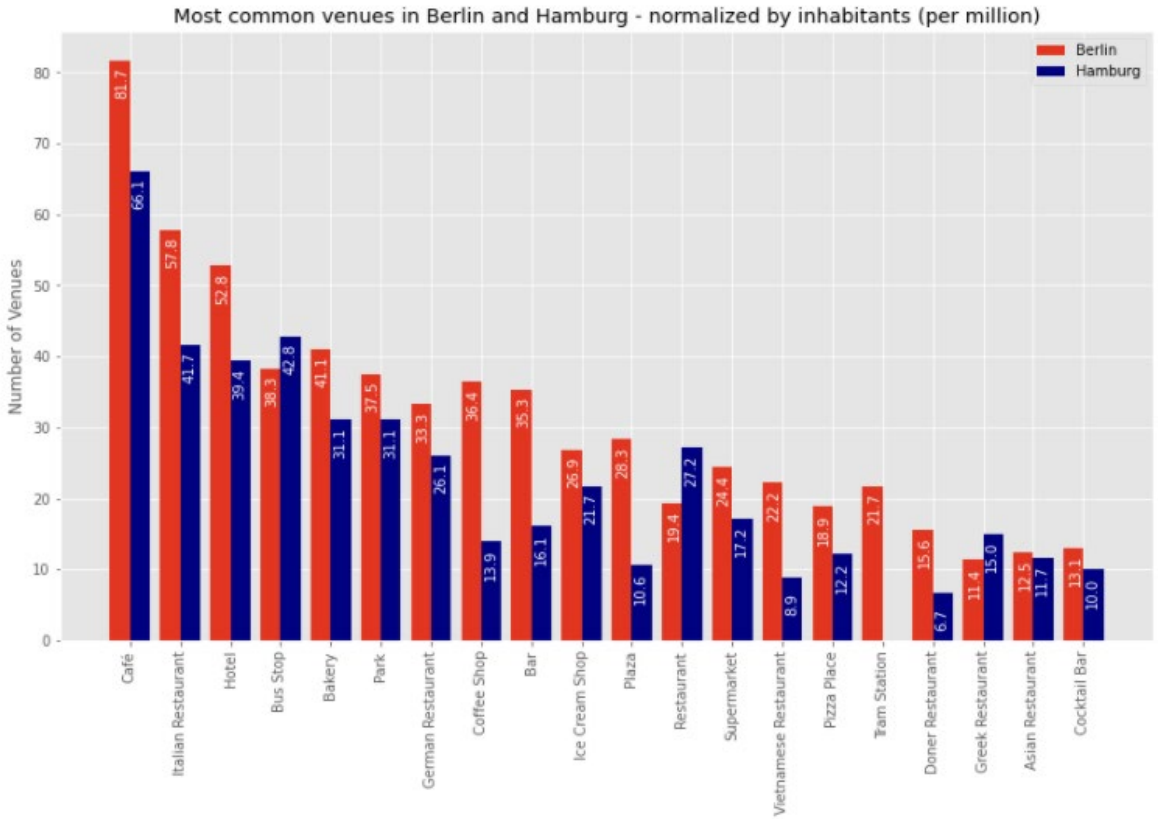
Cleaned dataset (Berlin and Hamburg), 296 rows and 4 columns.

Example of cleaned dataset

	Zipcode	State	Zipcode Latitude	Zipcode Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	10115	Berlin	52.5323	13.3846	Hotel i31	52.531107	13.384270	Hotel
1	10115	Berlin	52.5323	13.3846	Hotel ULTRA Concept Store	52.529362	13.396969	Furniture / Home Store
2	10115	Berlin	52.5323	13.3846	Deutsches Theater	52.523970	13.382001	Theater
3	10115	Berlin	52.5323	13.3846	Ackerstadtpalast	52.529721	13.396777	Performing Arts Venue
4	10115	Berlin	52.5323	13.3846	Sammlung Boros	52.523352	13.384213	Art Gallery
5	10115	Berlin	52.5323	13.3846	Factory Kitchen	52.537449	13.394714	Restaurant
6	10115	Berlin	52.5323	13.3846	Hamburger Bahnhof – Museum für Gegenwart	52.528513	13.372067	Art Museum
7	10115	Berlin	52.5323	13.3846	Du Bonheur	52.536310	13.397558	Pastry Shop
8	10115	Berlin	52.5323	13.3846	Bandol sur Mer	52.528992	13.395436	French Restaurant
9	10115	Berlin	52.5323	13.3846	H Gedenkstätte Berliner Mauer	52.535750	13.390708	Tram Station

Foursquare API scraped data

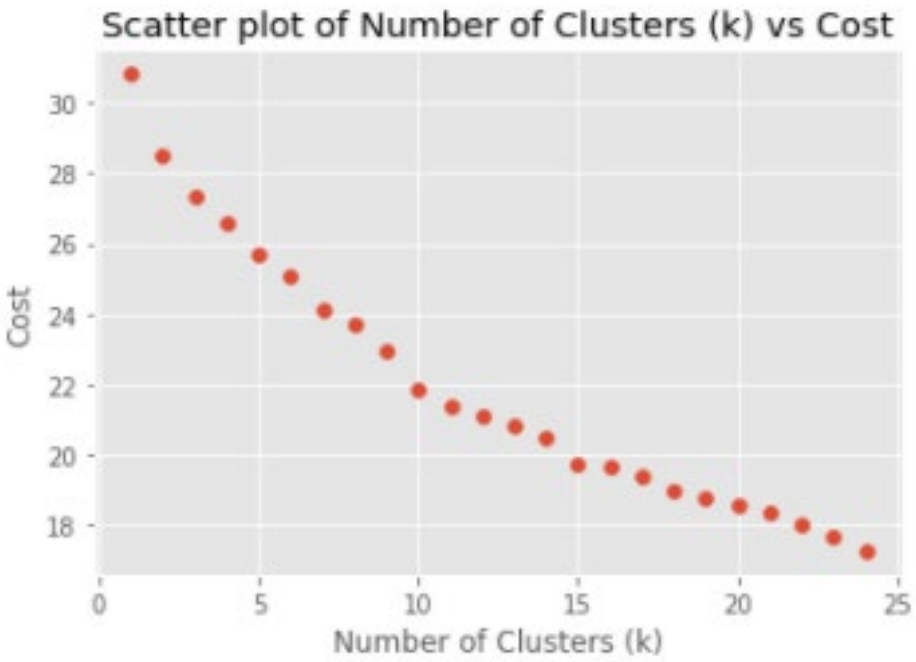
Analyzing venue similarity in both cities



Top 20 most common venues in both Berlin and Hamburg.

- More bars and cafés in Berlin.
- Bus stops more prominent in Hamburg, since other means of transport are less common.
- Almost all restaurant categories more common in Berlin.

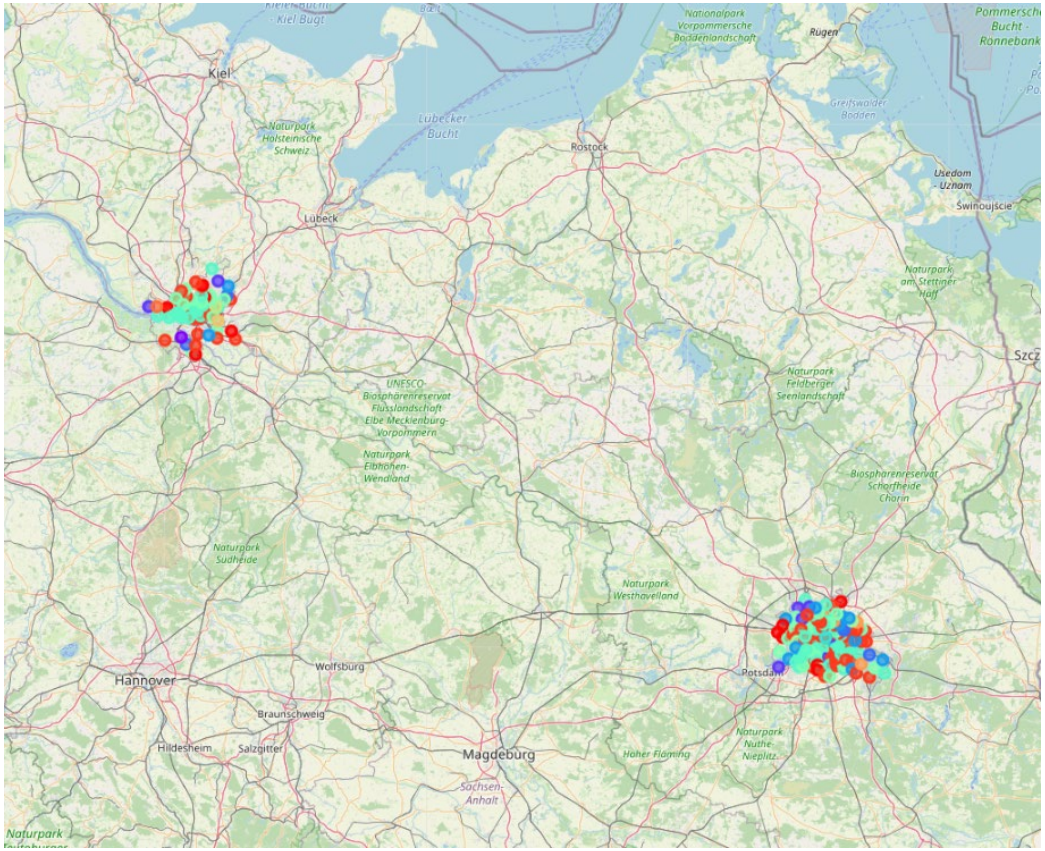
Clustering using k-means



Cluster formation between areas that share venue structure.

- Using elbow method to find that the most effective number of clusters forms when k=15.

Mapping clustered neighborhoods



By mapping the clustered neighborhoods, we can assume that:

- Similarities can be found in areas around the city center.
- Some peripheral neighborhoods probably share similarities like almost exclusive housing areas.

Conclusion

1. Built a model to analyze areas between different cities.
2. Accuracy is dependent on clustering. Greater room for errors.
3. Venue analysis could be used for:
   a) Residents thinking about moving.
   b) Companies thinking of relocation / expansion.
4. Ideas for better execution:
   a) Climate analysis
   b) Commute time analysis