Segmenting and Clustering Neighborhoods in Berlin

East and West Berlin -
30 years after the Berlin Wall came down

Christian Böhm
December 2019

Overview

Introduction

Data

Methodology

Results

Discussion

Conclusion

Introduction

Background

- It is now 30 years ago, that the Berlin Wall came down and the two parts of Berlin, East and West were re-unified again
- The re-unification of Germany brought together two very different political systems, societies and infrastructure leading to a long period of growing together

Business problem and key questions:

- Is there (still) a difference between neighborhoods located in East and West Berlin?
- If yes, what are the key characteristics of neighborhoods located in of East and West Berlin?

Data

Key data sources:

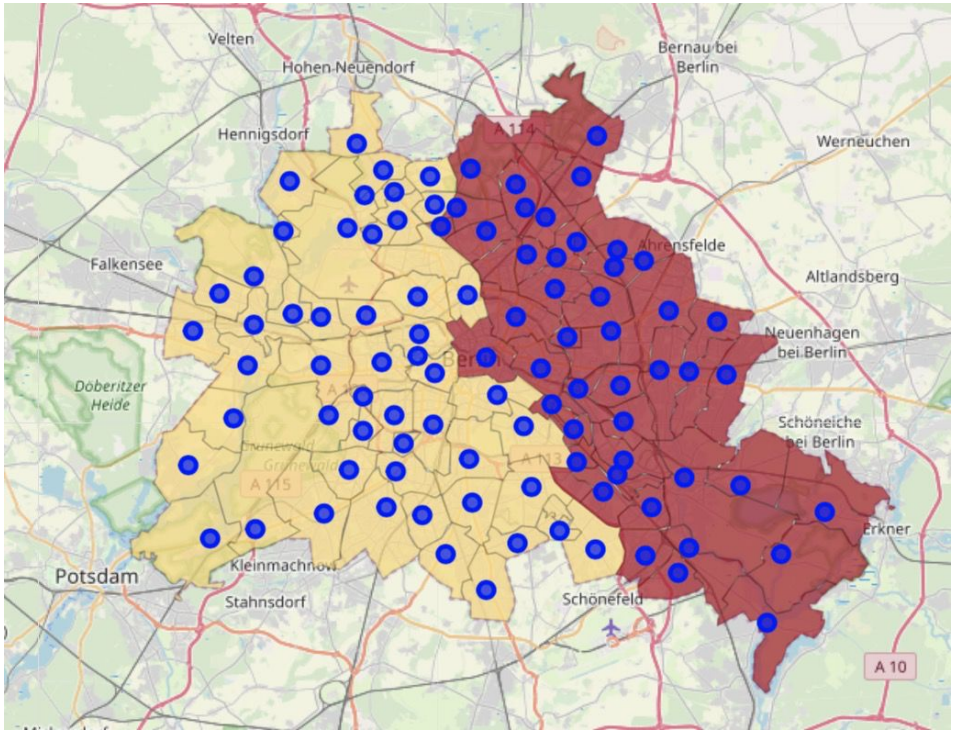
- List of all neighborhoods and boroughs in Berlin (Wikipedia: https://de.wikipedia.org/wiki/Verwaltungsgliederung_Berlins)
- Former boroughs of East Berlin (Wikipedia: https://en.wikipedia.org/wiki/East_Berlin)
- Location data (geojson file) of all Berlin boroughs (https://tsb-opendata.s3.eu-central-1.amazonaws.com/ortsteile/lor_ortsteile.geojson)
- The top 10 venues and its categories for all 96 neighborhoods in Berlin within a radius of 500 meters (FourSquare) https://developer.foursquare.com/docs/resources/categories)

Methodology

- Create dataframe of all Berlin neighborhoods including information about borough, location (longitude and latitude) as well as the part of Berlin (West or East)
- Create a map of the Berlin neighborhoods (folium)
- Create a function to process all the neighborhoods in Berlin using FourSquare: The top 10 venues and their categories for all 96 neighborhoods in Berlin within a radius of 500 meters
- Run *k*-means to cluster the neighborhood into 5 clusters.
- Visualize the resulting clusters on a map

Results - Geo data preparation

- There are 45 neighborhoods in East Berlin (red) belonging to 6 boroughs
- There are 51 neighborhoods in West Berlin (yellow) belonging to 8 boroughs



Results - Venue information

- Using FourSquare, 1170 venues can be downloaded covering 215 different venue categories
- The TOP 10 for each neighborhood are identified (2 examples shown right)

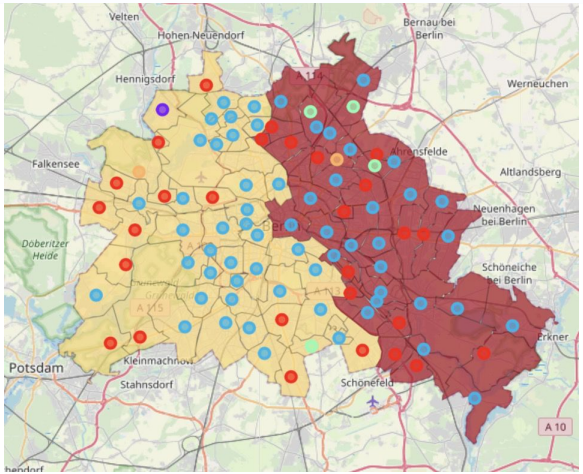
```
----Adlershof----
                venue  freq
0                Bank  0.14
1   Trattoria/Osteria  0.14
2    Insurance Office  0.14
3  Italian Restaurant  0.14
4    Greek Restaurant  0.14
5          Steakhouse  0.14
6         Supermarket  0.14
7         Opera House  0.00
8           Nightclub  0.00
9     Nature Preserve  0.00


----Alt-Hohenschönhausen----
                  venue  freq
0       Discount Store   0.2
1          Supermarket   0.2
2         Tram Station   0.2
3    Indian Restaurant   0.1
4             Drugstore   0.1
5     Greek Restaurant   0.1
6          Coffee Shop   0.1
7  Outdoor Event Space   0.0
8        Movie Theater   0.0
9            Multiplex   0.0
```

Results - *k-means* clustering

- The **majority of the neighborhoods belong to two cluster** (Cluster Labels 0 and 2 with 88 neighborhoods)
- The ratio of East neighborhoods per cluster for these is 51.9% or 42.6%
- There is one outlier (Cluster Label 1) with only one (West) neighborhood
- Cluster 3 shows a higher EastRatio, however only comprises 4 neighborhoods

Cluster Labels	East Neighborhood	West Neighborhood	Total Neighborhood	EastRatio_Percentage
0	14	13	27	51.851852
1	0	1	1	0.000000
2	26	35	61	42.622951
3	3	1	4	75.000000
4	1	1	2	50.000000
5	1	0	1	100.000000



Discussion

- Analysing the ratio of East Berlin neighborhoods per cluster clearly suggest, that there is no correlation between the part of Berlin (East or West) and assignment to a cluster

- The clustering analysis provides neither a pure East nor a pure West cluster (excluding outlier cluster with only one neighborhood)

- This finding is clearly visible on the map of clusters: All cluster cover East and West of Berlin.

Conclusion

- 30 years after the Berlin Wall came down, the city of Berlin seems to be united
- There are no significant differences between East and West Berlin neighborhoods
- An interesting question for further analysis is: How would the results have looked like using data for the venues 20 or 30 years ago?