

# Segmenting and Clustering Neighborhoods in Berlin

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

## Table of Contents

[1. Introduction](#)

[2. Data](#)

[3. Methodology](#)

[4. Results](#)

[5. Discussion](#)

[6. Conclusion](#)

## 1. Introduction

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.

An interesting question, both, from a historical and sociological view is, whether there are still quantifiable differences between East and West, e.g., in the neighborhoods of Berlin?

More precisely:

**1. Is there (still) a difference between neighborhoods located in East and West Berlin?**

**2. If yes, what are the key characteristics of neighborhoods located in of East and West Berlin?**

To answer this questions, location data of the 12 borrows and 96 neighborhoods of Berlin are used to get information about the top 10 venues for each neighborhood. All neighborhoods will be clustered (*k*-means) based on similarities in their venue categories. Theses clusters will visualized on a map. Finally, the distribution of East and West Berlin neighborhood across cluster will be analyzed.

The target audience of this projects are historians and sociologists as well as local politicians and city planners. This project will provide them valuable insights in potential structural differences in the Berlin neighborhoods allowing them to assess the re-unification process and help make local decision regarding city planning.

## **2. Data**

The project will use the following data and sources:

1. List of all neighborhoods and boroughs in Berlin (Wikipedia: [https://de.wikipedia.org/wiki/Verwaltungsgliederung\\_Berlins](https://de.wikipedia.org/wiki/Verwaltungsgliederung_Berlins) ([https://de.wikipedia.org/wiki/Verwaltungsgliederung\\_Berlins](https://de.wikipedia.org/wiki/Verwaltungsgliederung_Berlins)))
1. Former boroughs of East Berlin (Wikipedia: [https://en.wikipedia.org/wiki/East\\_Berlin](https://en.wikipedia.org/wiki/East_Berlin) ([https://en.wikipedia.org/wiki/East\\_Berlin](https://en.wikipedia.org/wiki/East_Berlin)))

1. Newly created and renamed boroughs after re-unification

(Wikipedia:[https://de.wikipedia.org/wiki/Verwaltungsgliederung\\_Berlins](https://de.wikipedia.org/wiki/Verwaltungsgliederung_Berlins)  
([https://de.wikipedia.org/wiki/Verwaltungsgliederung\\_Berlins](https://de.wikipedia.org/wiki/Verwaltungsgliederung_Berlins)))

Friedrichshain-Kreuzberg: Two neighborhoods - Friedrichshain (East) and Kreuzberg (West)

Mitte: Six neighborhoods - Mitte (East), Tiergarten (West), Wedding (West), Gesundbrunnen (West), Moabit (west), Hansaviertel (West)

Marzahn-Hellersdorf: Fusion of two boroughs - Marzahn (East) and Hellersdorf (East)

Treptow-Köpenick: Fusion of two boroughs - Treptow (West) and Köpenick (West)

Pankow: Fusion of three boroughs - Pankow (East), Prenzlauer Berg (East), Weissensee (East)

Lichtenberg: Fusion of two boroughs - Lichtenberg (East), Hohenschönhausen

1. Location data of all Berlin neighborhoods (geopy) (e.g., latitude, longitude: 52.530102, 13.342542)

1. Location data (geojson file) of all Berlin boroughs ([https://tsb-opendata.s3.eu-central-1.amazonaws.com/ortsteile/lor\\_ortsteile.geojson](https://tsb-opendata.s3.eu-central-1.amazonaws.com/ortsteile/lor_ortsteile.geojson) ([https://tsb-opendata.s3.eu-central-1.amazonaws.com/ortsteile/lor\\_ortsteile.geojson](https://tsb-opendata.s3.eu-central-1.amazonaws.com/ortsteile/lor_ortsteile.geojson)))

1. The top 10 venues and its categories for all 96 neighborhoods in Berlin within a radius of 500 meters (FourSquare)

(e.g., Restaurants, Museum, Theater -

<https://developer.foursquare.com/docs/resources/categories>

(<https://developer.foursquare.com/docs/resources/categories>))

## 3. Methodology

### 3.1. Create dataframe of all Berlin neighborhoods

Create a dataframe of all Berlin neighborhoods - including information about borough, location (longitude and latitude) as well as the part of Berlin (West or East). The dataframe is called **berlin\_df**.

First, extract neighborhood and borough data from data source 1 (see Data section) and check data: should cover 96 neighborhoods and 12 boroughs

Second, get list of boroughs in former East Berlin from data source 2. Extract neighborhood data and rename columns. Add column 'Berlin' to indicate part of Berlin - in this case 'East'. Rename former East Berlin boroughs according to current names of Berlin boroughs. Delete old names and remove former boroughs, which are now neighborhoods of a new borough (data source 3).

Third, using the dataframe with the names of all 96 neighborhoods, the coordinates of Berlin neighborhoods are determined with the geolocator function. To avoid time-outs requests are split into 5 packages.

The neighborhood location data is merged with the list of East Berlin boroughs; West Berlin neighborhoods are labelled with 'West'. Finally, the labeling for the two East Berlin neighborhoods now belonging to a West Berlin borough is adjusted to 'East'.

### **3.2. Create a map of the Berlin neighborhoods**

Add column 'Color' to the dataframe for coloring East and West on map differently. Download Berlin neighborhood geojson file (data source 5). Then generate a choropleth map using the part of Berlin ('East', 'West') and superimpose the location of all 96 neighborhoods on top.

### **3.3. Get top venues for all neighborhoods**

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. Create a new dataframe called `berlin_venues`. Run function and check how many venues are returned for each neighborhood and how many unique categories are there. Next, analyze each neighborhood and create a profile along the unique categories extracted. Group rows by neighborhood and by taking the mean of the frequency of occurrence of each category. Finally, **print each neighborhood along with the top 10 most common venues**.

Write a function to sort the venues in descending order and create a new dataframe and display the top 10 venues for each neighborhood:  
**neighborhoods\_venues\_sorted**.

### 3.4. Cluster neighborhoods using *k-means*

Run *k-means* to cluster the neighborhood into 5 clusters. Then create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood: **berlin\_merged**. Check the cluster labels and assign neighborhoods with Cluster Label 'NaN' to new cluster 5. Finally calculate the ratio of East neighborhoods per cluster (count of East neighborhoods per cluster divided by total count of neighborhoods (East and West) per cluster)

### 3.5. Visualize the resulting clusters on a map

## 4. Results

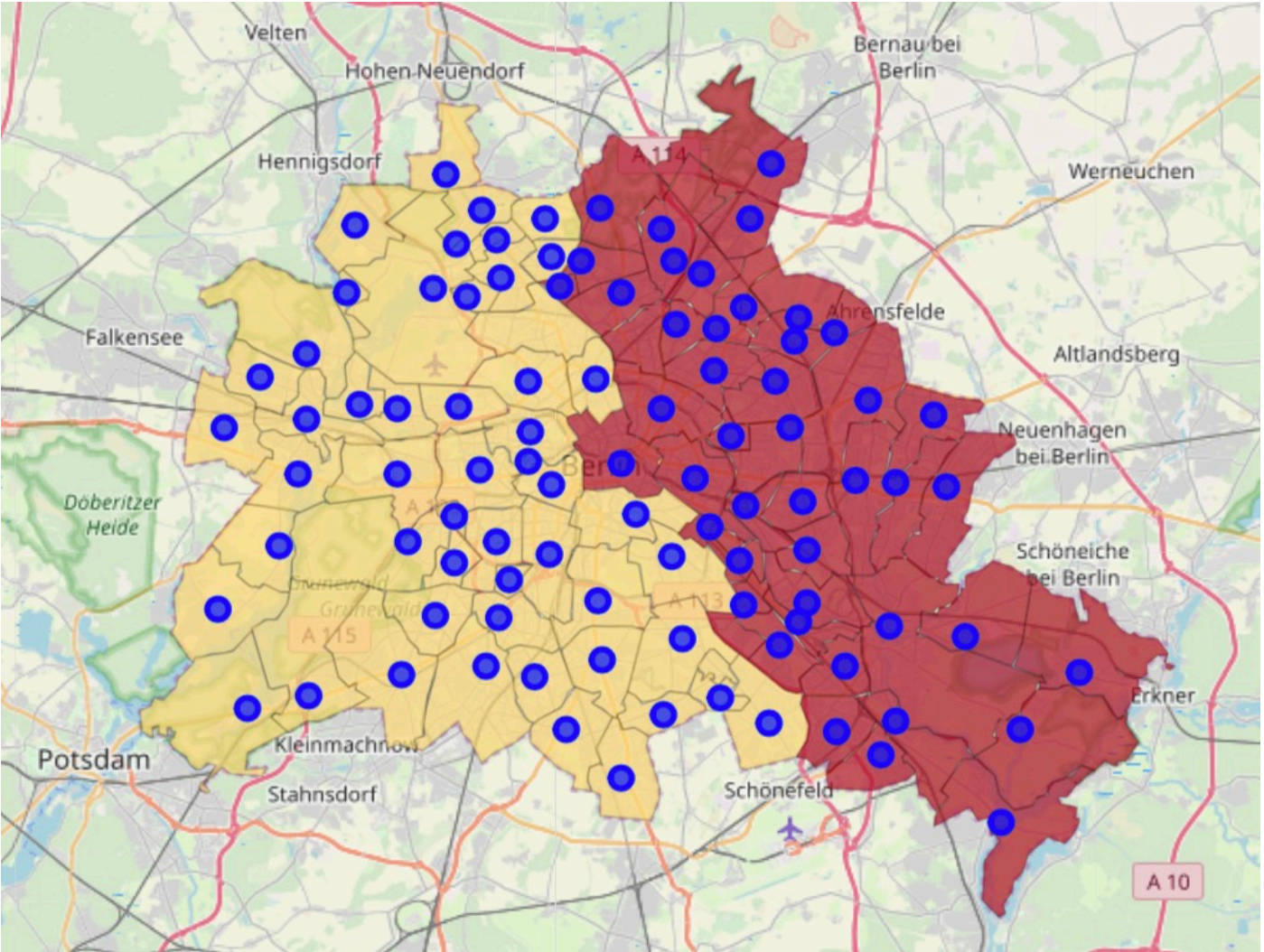
### 4.1. Dataframe of all Berlin neighborhoods

The resulting dataframe `berlin_df` comprises all relevant information for the 96 neighborhoods and 12 boroughs of Berlin. There are 45 neighborhoods in East Berlin belonging to 6 boroughs. There are 51 neighborhoods in West Berlin belonging to 8 boroughs. Below the 'count'-analysis of the **berlin\_df** dataframe is shown:

		Neighborhood	Latitude	Longitude
Borough	Berlin			
Charlottenburg-Wilmersdorf	West	7	7	7
Friedrichshain-Kreuzberg	East	1	1	1
	West	1	1	1
Lichtenberg	East	10	10	10
Marzahn-Hellersdorf	East	5	5	5
Mitte	East	1	1	1
	West	5	5	5
Neukölln	West	5	5	5
Pankow	East	13	13	13
Reinickendorf	West	11	11	11
Spandau	West	9	9	9
Steglitz-Zehlendorf	West	7	7	7
Tempelhof-Schöneberg	West	6	6	6
Treptow-Köpenick	East	15	15	15

#### 4.2. Map of the Berlin neighborhoods

The map shows the 96 neighborhoods across East (red) and West (yellow) Berlin.



4.3. Top 10 venues for all neighborhoods

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

----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

The top 10 venues for each neighborhood: **neighborhoods\_venues\_sorted**. First 5 rows:



	Neighborhood	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	Adlershof	Steakhouse	Insurance Office	Supermarket	Trattoria/Osteria	Italian Restaurant	Greek Restaurant	Bank	Flower Shop	Food & Drink Shop	Food Court
1	Alt-Hohenschönhausen	Supermarket	Discount Store	Tram Station	Coffee Shop	Indian Restaurant	Drugstore	Greek Restaurant	Farmers Market	Fast Food Restaurant	Fish Market
2	Alt-Treptow	Café	Bakery	Italian Restaurant	Gift Shop	Platform	Sandwich Place	Snack Place	Deli / Bodega	Big Box Store	Beer Garden
3	Baumschulenweg	Supermarket	Ice Cream Shop	Bus Stop	Asian Restaurant	Italian Restaurant	Drugstore	Farmers Market	Fried Chicken Joint	French Restaurant	Fountain
4	Biesdorf	Bakery	Park	Palace	Supermarket	Outdoor Event Space	Light Rail Station	Yoga Studio	Fountain	Food Truck	Food Court

#### 4.4. Neighborhood clusters - using *k-means*

Using *key-means* 5 clusters of the Berlin neighborhoods are created (Cluster labels 0-4). One additional cluster (5) is defined for 3 neighborhoods, where no venue information was available (resulting in a 'NaN' cluster label). A new dataframe is created that includes the cluster as well as the top 10 venues for each neighborhood: **berlin\_merged**. First 5 rows:

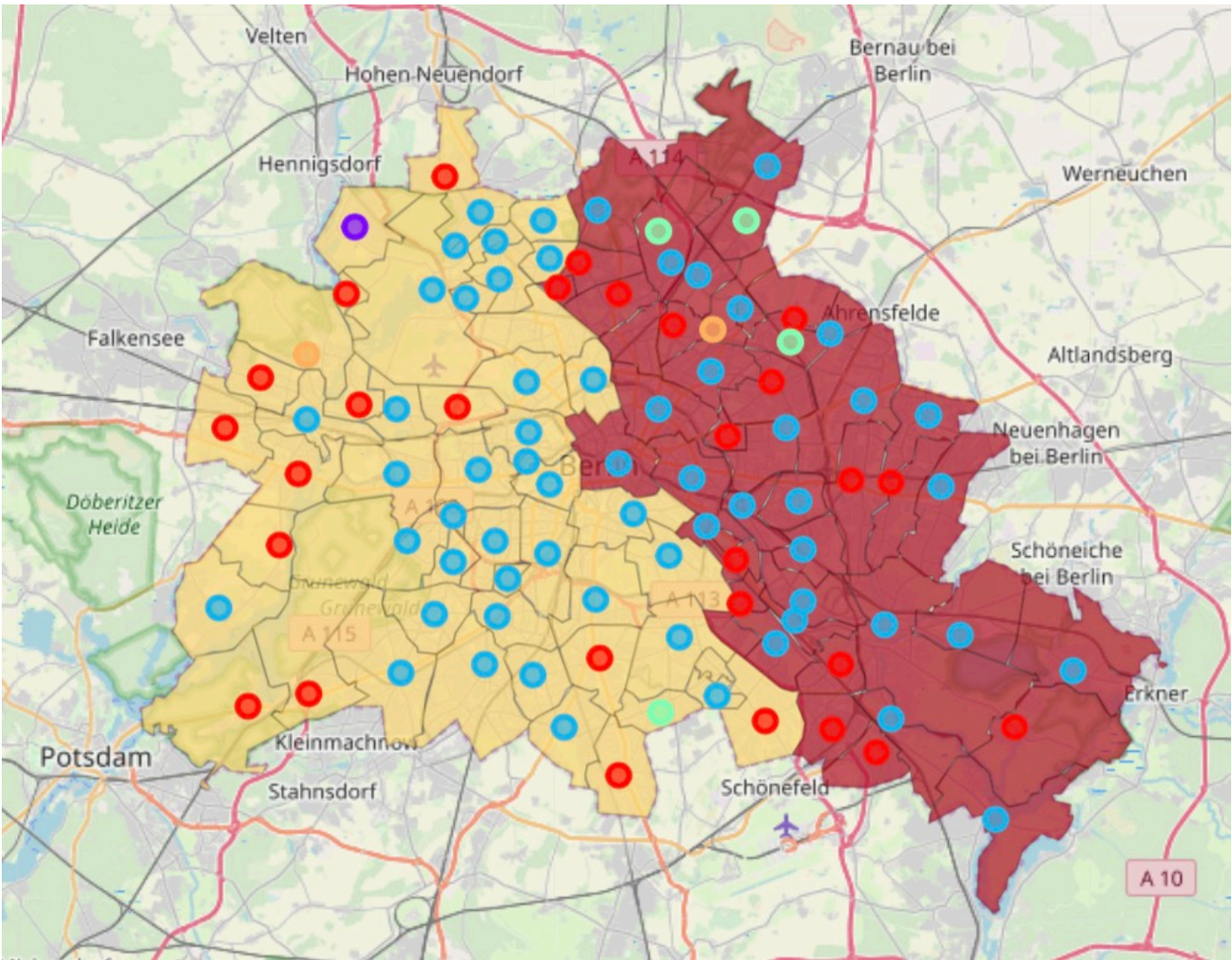
	Borough	Neighborhood	Latitude	Longitude	Berlin	Color	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Mitte	Mitte	52.517690	13.402376	East	1	2.0	History Museum	Museum	Hotel	Ice Cream Shop	Bookstore	Scenic Lookout	Café	Candy Store	Plaza	Concert Hall
1	Mitte	Moabit	52.530102	13.342542	West	0	2.0	Café	Hotel	Cocktail Bar	Bar	Hostel	Drugstore	Burger Joint	Breakfast Spot	Pub	German Restaurant
2	Mitte	Hansaviertel	52.519123	13.341872	West	0	2.0	Café	Art Museum	Irish Pub	Mediterranean Restaurant	Rental Car Location	Liquor Store	Plaza	Boat or Ferry	Playground	Bistro
3	Mitte	Tiergarten	52.509778	13.357260	West	0	2.0	Hotel Bar	German Restaurant	Hotel	Park	Coffee Shop	Scandinavian Restaurant	Café	Sculpture Garden	Memorial Site	Garden
4	Mitte	Wedding	52.550123	13.341970	West	0	2.0	Ice Cream Shop	Bakery	Big Box Store	Gas Station	Food & Drink Shop	Supermarket	Tennis Court	Bookstore	Pharmacy	Park

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%. With 96 neighborhoods in total and 45 East Berlin neighborhoods, the expectation value for the 'EastRatio'is 47%. There is one outlier (Cluster Label 1) with only one (West) neighborhood. Cluster 3 shows a higher EastRatio, however only comprises 4 neighborhoods. Besides these clusters, there are no significant East or West clusters.

	East Neighborhood	West Neighborhood	Total Neighborhood	EastRatio_Percentage
Cluster Labels				
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



## 4.5. Map of resulting clusters



## 5. 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.

## 6. 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?