TOBIAS STURM 2021-01-23

Neighborhood Analysis of a German City for Young Families

Coursera Final Project

Content

- Description of the problem
- Description of the data
- Methodology
- Results
- Discussion
- Conclusion

Description of the problem

Young families with children or plans for some are frequently in the situation to find a new place to call home, that will give their changing life as a family the best possible neighborhood. Identification is not an easy task, as there are multiple factors to be included and not all information readily available. As a support for their decision making, we want to provide a geographical analysis of children-friendly neighboorhoods based on the distance to desired venues. For example the new home must be near a school, but also provide a playground for leasure time. We'll focus on the German City Wermelskirchen out of curiosity.

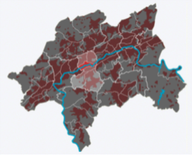
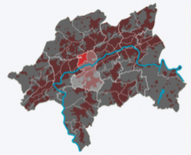
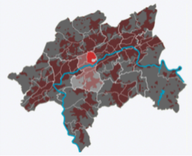
Description of the data

As dataset we're using a publicly available dataset of the German City Wuppertal, including it's districts and several population metrics.

https://de.wikipedia.org/wiki/Liste_der_Stadtbezirke_und_Stadtteile_von_Wuppertal

In the process of preperation we'll be translating and transforming the data. Please bear with me for now, but the relevant data is the following:

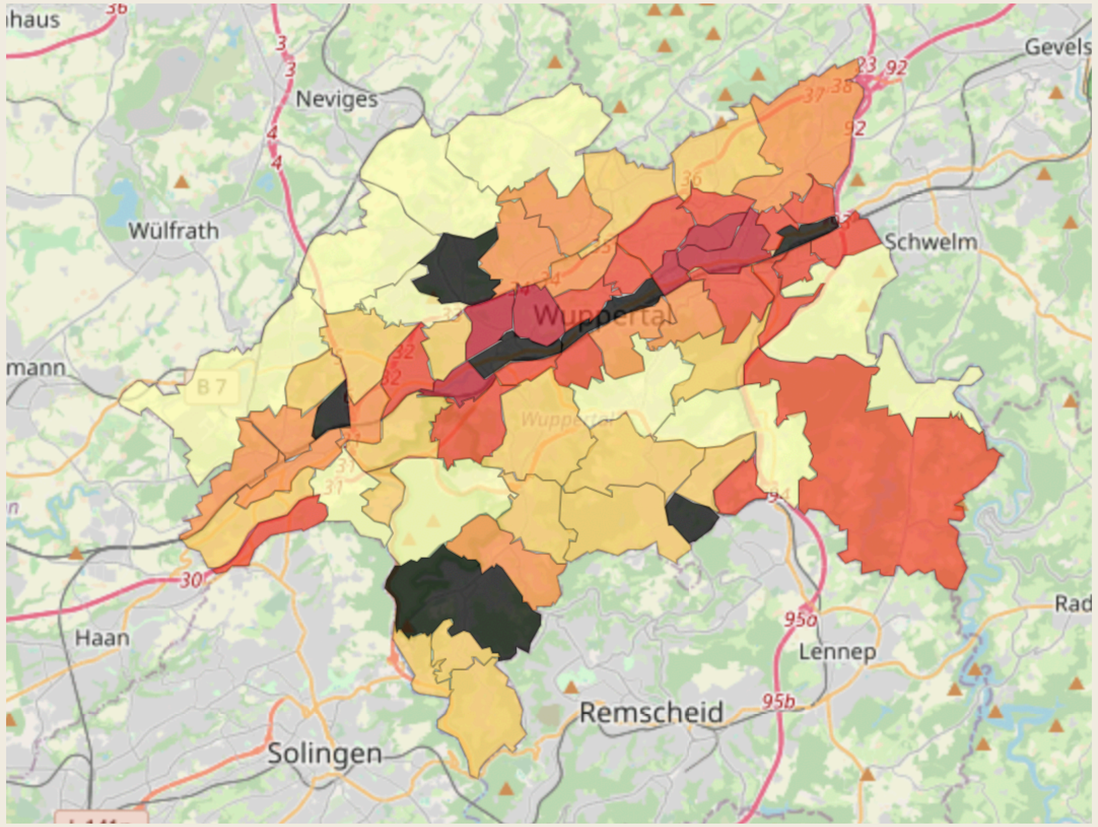
- ‚Neighborhood'
- 'Borough'
- 'Residents'
- 'Size'
- 'Population_Density'
- 'Foreigner_Percentage'
- 'Unemployment_Rate'
- 'Livinghouses'
- 'Flats_thereof'
- 'Schools(Elementary_Schools)'
- 'Private_Cars'

Karte[4]	Nr.	Statistisches Wohnquartier	Stadtbezirk	Kommunale Zuordnung vor der Eingemeindung	Einwohner-zahl	Fläche (km²)
	00	Elberfeld-Mitte	Elberfeld	Stadt Elberfeld (bis 1929)	5780	1,08
	01	Nordstadt	Elberfeld	Stadt Elberfeld (bis 1929)	17.269	1,18
	02	Ostersbaum	Elberfeld	Stadt Elberfeld (bis 1929)	14.919	1,38

Description of the data

Additionally we'll use the also publicy available GEOJSON data for these districts including their geographic boundaries.

Url: http://daten.wuppertal.de/Infrastruktur_Bauen_Wohnen/Quartiere_EPSG4326_JSON.json



Description of the data

Furthermore we'll connect to the foursquare database and use data for venues from there.

Methodology

The imported wikipedia is slightly formatted and prepared for further data manipulation.

eighborhood	Borough	Residents	Size	Population_Density	Foreigner_Percentage	Unemployment_Rate	Livinghouses	Flats_thereof	Schools	Private_Cars	Elementary_Schools	
0	Elberfeld-Mitte	Elberfeld	5780	1.08	5352	2.51	0.0913	651	3718	2	1764	0
1	Nordstadt	Elberfeld	17269	1.18	14635	2.77	0.0903	1637	10675	8	4926	3
2	Ostersbaum	Elberfeld	14919	1.38	10811	2.46	0.0967	1416	8807	4	4877	3
3	Südstadt	Elberfeld	9640	0.59	16339	1.85	0.0766	771	6048	1	2977	1
4	Grifflenberg	Elberfeld	11696	4.45	2628	1.01	0.0321	1557	6289	1	5181	1
5	Friedrichsberg	Elberfeld	6449	2.39	2698	1.44	0.0678	654	3591	2	2396	2
6	Sonnborn	Elberfeld-West	4008	2.39	3929	1.33	0.0594	545	2360	1	1722	1
7	Varresbeck	Elberfeld-West	4376	2.59	1690	1.45	0.0375	804	2228	0	2215	0
8	Nützenberg	Elberfeld-West	5590	1.48	3777	1.75	0.0615	1034	3295	4	2491	3
9	Brill	Elberfeld-West	4414	1.22	3618	0.70	0.0283	693	2773	1	2470	0
10	Arrenberg	Elberfeld-West	5449	0.86	6336	3.21	0.0991	516	2983	1	1679	1

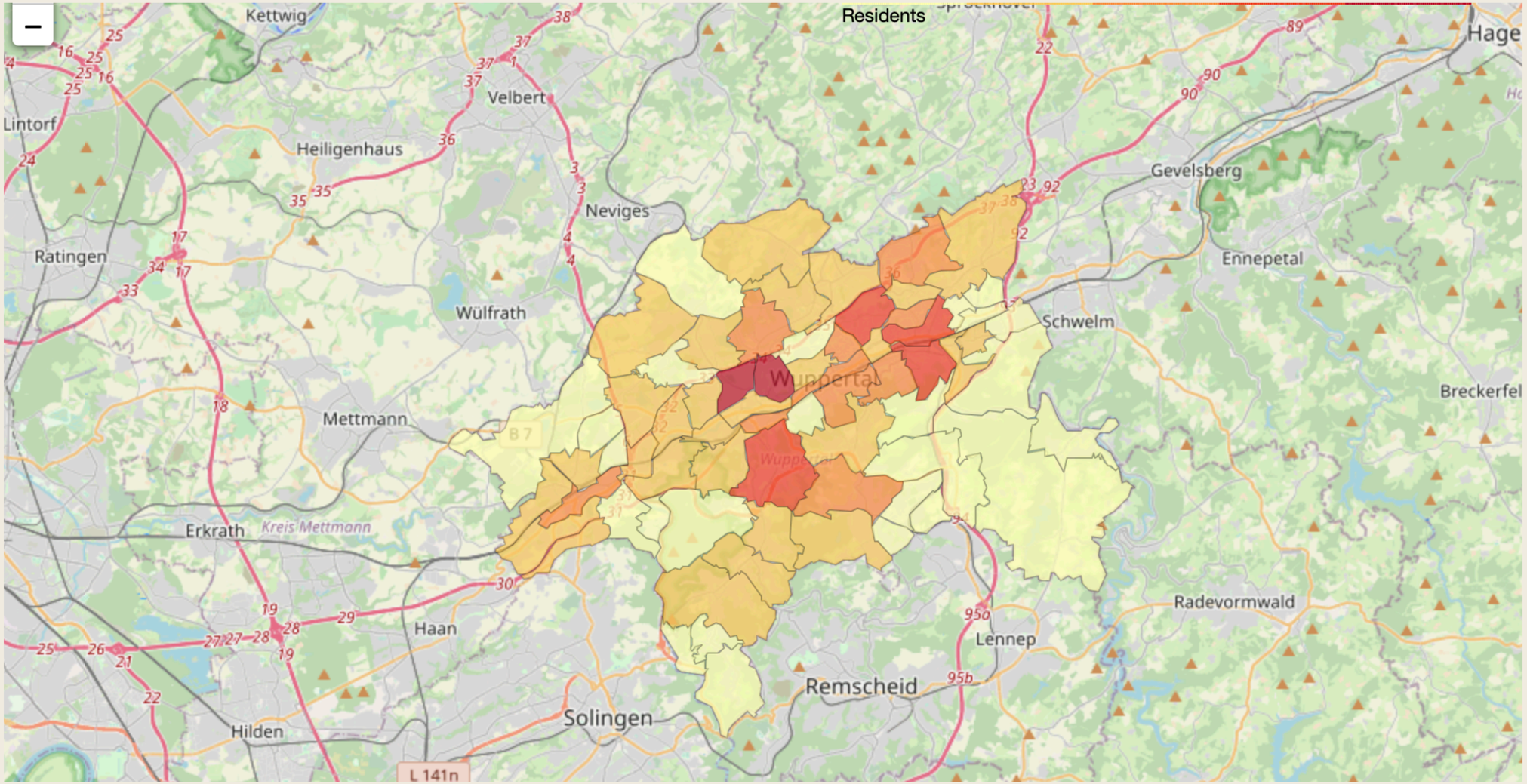
Methodology

We're also importing geojson data from the city of Wuppertal, that contains the boundaries of districts. From these the center coordinates are calculated geometrically and prepared in a frame as the following.

Neighborhood	Longitude	Latitude	
0	Herbringhausen	7.271301	51.231740
1	Schöller-Dornap	7.053574	51.248908
2	Dönberg	7.152916	51.300403
3	Nächstebreck-Ost	7.246319	51.301531
4	Cronenberg	7.130258	51.203025
5	Eckbusch	7.097525	51.276781
48	Kohlfurth	7.116513	51.189941
49	Elberfeld	7.146831	51.255741
50	Rauental	7.237412	51.267165
51	Fr.-Engels-Allee	7.181979	51.264625
52	Hesselnberg	7.170063	51.253498
53	Sonnborn	7.097854	51.242142
54	Loh	7.171771	51.266405
55	Rehsiepen	7.223867	51.226991
56	Wichlinghausen-Nord	7.215522	51.286920
57	Schenkstr.	7.207330	51.220951
58	Buchenhofen	7.103682	51.225656
59	Arrenberg	7.130278	51.249188
60	Löhrerlen	7.248506	51.287162
61	Wichlinghausen-Süd	7.215810	51.281064
62	Rott	7.185731	51.272906
63	Industriestr.	7.086556	51.243415
64	Langerfeld-Mitte	7.243585	51.274339
65	Jesinghauser Str.	7.245921	51.280918
66	Südstadt	7.151344	51.252192
67	Barmen-Mitte	7.200724	51.272681
68	Hilgershöhe	7.235965	51.282507

Methodology

The following map shows all districts with a color coding based on their resident count, to get an initial overview of the city.



Methodology

From Foursquare the venues in 500 meter radius of the district centers are fetched. These do contain the category of each venue.

Unnamed: 0	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	
0	0	Elberfeld	51.255741	7.146831	Peek & Cloppenburg	51.257811	7.145993	Clothing Store
1	1	Elberfeld	51.255741	7.146831	Milia's Coffee	51.256723	7.147150	Café
2	2	Elberfeld	51.255741	7.146831	Café Venezia	51.257475	7.146048	Café
3	3	Elberfeld	51.255741	7.146831	Mazzino	51.255883	7.146583	Coffee Shop
4	4	Elberfeld	51.255741	7.146831	Food Brother	51.257393	7.143594	Burger Joint
5	5	Elberfeld	51.255741	7.146831	mangimangi	51.259548	7.147168	Restaurant

Methodology

For the k-mean clustering, the number of categories are reduced in order to receive meaningful results.

There are more than 50 different categories in the foursquare data, which are reduced to three:

- Activity
- Unfriendly
- Rest

At the same time the field House per resident is calculated to give an indication of wealth in the district for lack of better measurement.

Results

Districts do differ substantially in all their criteria.

	Neighborhood	Size	Residents	Foreigner_Percentage	Unemployment_Rate	Livinghouses	Flats_thereof	Schools	Elementary_Schools	Private_Cars
Borough										
Elberfeld	6	11.07	65753	2.006667	0.075800	6686	39128	18	10	22121
Barmen	10	15.44	59410	1.466000	0.059770	7403	34015	21	9	24162
Oberbarmen	5	12.57	42910	1.460000	0.066640	5511	22830	13	8	17372
Uellendahl-Katernberg	7	25.91	38192	0.524286	0.025857	7713	19619	9	7	21174
Vohwinkel	9	20.42	31578	0.920000	0.037956	5328	15732	12	6	14652
Elberfeld-West	7	11.74	27774	1.372857	0.050100	4243	15895	8	6	12784
Langerfeld-Beyenburg	9	29.40	25517	0.955556	0.058256	4560	13447	7	5	12552
Cronenberg	7	21.50	21846	0.490000	0.025729	4996	11102	7	4	12881
Ronsdorf	6	16.05	21776	0.588333	0.035267	4141	11296	6	5	11819
Heckinghausen	3	5.66	21261	1.163333	0.054267	2555	12150	4	3	9014

Results

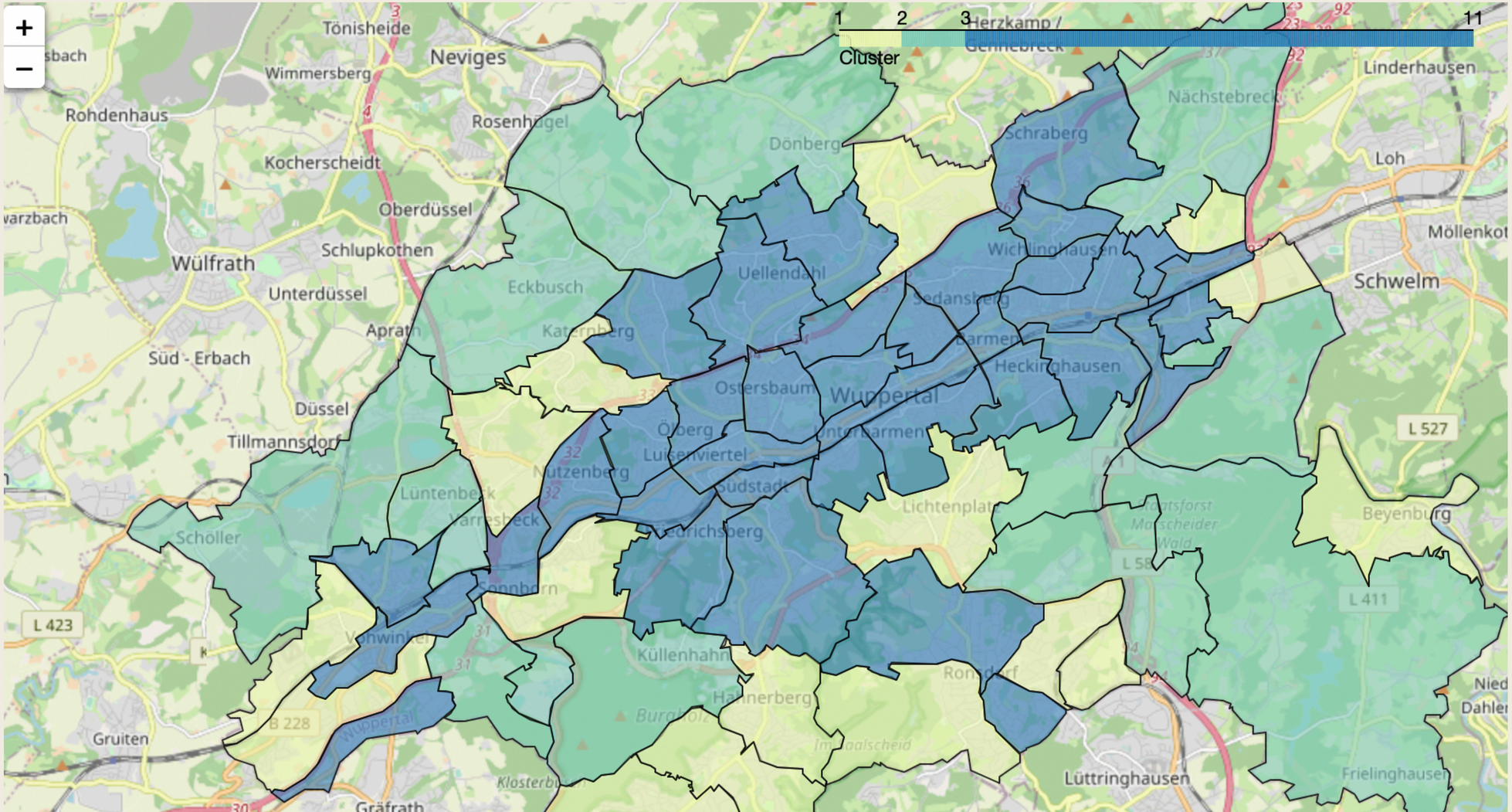
The k-means-alghorythm finds different groups of districts, while the means per measurement are shown below. As shown there are significant differences in between clusters.

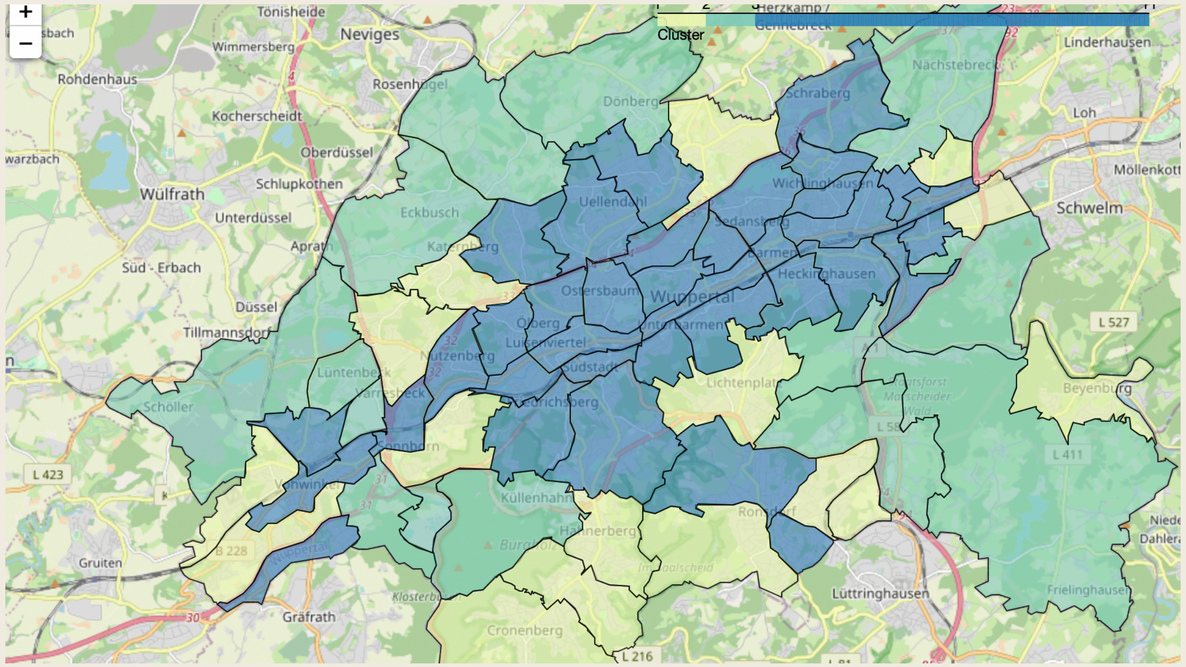
	Unemployment_Rate	Schools	Activity	Unfriendly	Population_Density	Houses_per_resident
Cluster Labels						
0	0.047700	1.500000	0.050000	0.100000	2632.200000	0.153769
1	0.082200	5.000000	0.154762	0.083333	12795.500000	0.107729
2	0.091400	2.500000	0.088816	0.062500	8736.500000	0.087302
3	0.025531	0.625000	0.020833	0.083333	482.500000	0.267042
4	0.073488	2.500000	0.020833	0.099811	6637.250000	0.115429
5	0.083450	4.500000	0.000000	0.200000	15487.000000	0.087387
6	0.050217	1.333333	0.155556	0.111111	3723.500000	0.165549
7	0.035959	0.882353	0.102941	0.017157	1666.294118	0.195487
8	0.068200	1.800000	0.364935	0.009091	4825.600000	0.116707
9	0.096700	4.000000	0.600000	0.000000	10811.000000	0.094913

Discussion

After ranking each cluster by weighted sum, introducing a personal bias, the map looks like this.

The Unemployment is weighted the highest, while schools and children unfriendly venues are reprioritized.



Conclusion

We can conclude, that cluster 7 is the most suited for young families, based on our criteria. This seem to be the neighborhoods in between the center and the outmost districts. This makes sense, as we'd expect the best balance of kids friendly activities and distance from nightlife, industrial zones and high unemployment neighborhoods.