A guide to the neighborhoods of Düsseldorf, DE

Oliver M.

January 14, 2020

1 Introduction

1.1 Background

Düsseldorf is the 7th largest city of Germany and according to a study of the consulting company Mercer in 2019, one of the global top 10 cities it is worth living in.

Today, Düsseldorf is divided into 10 departments and 50 neighborhoods. Originally it was founded in the 12th/13th century next to the rivers Rhine and Düssel and has since grown into a city with a total population of ~620 thousands. In 2018 the population increased by ~14 thousands.

Due to its heritage and expansion over more than eight centuries, it is coined by a rather urbanized city center with a mix of historic and modernized buildings, accompanied by diverse neighborhoods surrounding the city centre.

1.2 Problem

Imagine you would like to move to Düsseldorf. One of the first questions is, to decide where you would like to live within the city. So the following two questions might be of relevance for you: Which neighborhoods would suit your preferences with regards to urban design? Which neighborhoods are more popular, i.e. in higher demand than others?

This project aims to provide a clustering of the different neighborhoods based on selected social and urban design specific indicators (e.g. % of urbanised and non-urbanised area, population per square kilometre), as well as the type of venues located in the neighborhoods. Next to the clustering of neighborhoods, an analysis of the change of population per neighborhood over the last five years aims to provide an additional perspective on the neighborhoods and their development.

1.3 Target Group

Anybody who is interested in the design and variety of the different neighborhoods of Düsseldorf is a potential interest in this report and might accompany any offical city guide. Thereby it does not matter, if you would like to move to Düsseldorf or if you just would like to visit the city and plan your city tour.

2   Data Description

2.1   Data sources

The primary source of this project is a dedicated resource for datasets of Düsseldorf:

The city of Düsseldorf is publishing a large variety of datasets via the webpage
http://opendata.duesseldorf.de. For this project I will focus on the datasets that provide the geometrical coordinates of the neighborhoods and social areas of Düsseldorf, as well as the data on population changes and composition of each social area.

The secondary source of this project is Foursquare and its API, which allows to perform requests on recommended venues in a specific radius around a specific geo location. Each venue is thereby categorized in a hierarchical category structure and thereby allows to cluster the different locations by their type.

In the iPythonNotebook of this project, for each data source the relevant link is provided.

2.2   Data Examples

Neighborhoods Geo Location Data

	Quelle	Stand	Stadtbezirk	Stadtteil	Name	geometry
0	Amt 12/2	31.12.2017	5	54	Wittlaer	POLYGON ((6.76623 51.31355, 6.76618 51.31359, ...
1	Amt 12/2	31.12.2017	5	55	Angermund	POLYGON ((6.80952 51.34292, 6.80948 51.34302, ...
2	Amt 12/2	31.12.2017	4	43	Lörick	POLYGON ((6.74610 51.25203, 6.74602 51.25205, ...
3	Amt 12/2	31.12.2017	4	42	Heerdt	POLYGON ((6.73560 51.23767, 6.73522 51.23780, ...
4	Amt 12/2	31.12.2017	4	41	Oberkassel	POLYGON ((6.76868 51.23195, 6.76868 51.23196, ...

Description of columns:

Column	Translation	Description
Quelle	Source	Name of the city department that has created the data.
Stand	Status	Date of the data.
Stadtbezirk	Department	ID of the department.
Stadtteil	Neighborhood ID	ID of the neighborhood.
Name	Neighborhood	Name of the neighborhood.
geometry	geometry	Description of the borders of each neighborhood as polygon geo coordinates.

Content:

The data of the neighboorhoods has been created by the department 'Amt 12/2' of the citiy of Duesseldorf with status of December 31st, 2017. The city area of Duesseldorf is divided into 10 departments and 50 neighboorhoods. Each neighborhood is assigned to one department. Departments

only have an ID and no name, whereas neighborhoods have an ID and a name. The neighborhood ID is a merged key based on the ID of the department and a digit for the neighborhood.

As the analysis will be based on the neighborhoods only the columns 'Neighborhood ID', 'Neighborhood and 'geometry' will be used.

Population

	Stadtteilnummer	Stadtteilname	2012	2013	2014	2015	2016	2017	2018
0	11	Altstadt	2232	2253	2221	2264	2244	2358	2404
1	12	Carlstadt	2290	2323	2315	2491	2588	2279	2285
2	13	Stadtmitte	13900	13953	14388	14680	15008	14840	14954
3	14	Pempelfort	29931	30345	30814	31449	31897	32626	33064
4	15	Derendorf	19740	19927	20005	20232	20401	20651	20610

Description of columns:

Column	Translation	Description
Stadtteilnummer	Neighborhood ID	ID of the neighborhood.
Stadtteilname	Neighborhood	Name of the neighborhood.
2012-2018	2012-2018	Size of population per neighborhood at the end of the year.

Content:

The data includes the total population of each neighborhood by the end of the years 2012 to 2018. The population data is provided for all 50 neighborhoods.

For the analysis the data for the past five years (2013 to 2018) will be used.

Social Areas (Geo Location Data)

	OBJEKT	STADTBEZ	Sozialraum7	Sozialraum4	Quelle	geometry
0	Region	5	506053	506	Amt 12/2	POLYGON ((6.73981 51.30221, 6.73912 51.30235, ...
1	Region	4	409043	409	Amt 12/2	POLYGON ((6.72762 51.24037, 6.72785 51.24064, ...
2	Region	4	411043	411	Amt 12/2	POLYGON ((6.72426 51.24696, 6.72440 51.24771, ...
3	Region	6	604063	604	Amt 12/2	POLYGON ((6.82758 51.26707, 6.82614 51.26850, ...
4	Region	6	607064	607	Amt 12/2	POLYGON ((6.80817 51.25976, 6.80728 51.26016, ...

Description of columns:

Column	Translation	Description
OBJEKT	Object	Details not available.
STADTBEZ	Department	ID of the Department.
Sozialraum7	Social Area 7	Combined ID of Social Area ID and Neighborhood ID.
Sozialraum4	Social Area	ID of the Social Area.
Quelle	Source	Name of the city department that has created the data.
geometry	geometry	Description of the borders of each neighborhood as polygon geo coordinates.

Content:

The data includes the geo location data for the 179 social areas in Düsseldorf with status 2018. The social areas are mapped to the departments and with three exceptions, also to neighborhoods. The key in column "Sozialraum7" is a merged key, in which the last three digits represent the corresponding Neighborhood ID and all digits before that represent the Social Area ID ('Sozialraum4').

As transparent in the 'Description' of the DataFrame, the value '0' is appearing three times in the column 'Sozialraum7'. All other values are unique. The three '0' entries need to be further analyzed and decided on how to cope with those.

Social Areas (Geographical and Environmental Data)

	Stadtbezirk	Sozialraum	Fläche in ha	Einwohner je km²	Wohnfläche Anteil in %	Grünfläche Anteil in %	Von Verkehrslärm betroffene Bevölkerung in %
0	1	101	56,95	13440,85	29,82	0,31	47,95
1	1	102	43,62	5144,30	26,67	2,65	0,49
2	1	103	45,19	5727,21	22,84	9,37	8,94
3	1	104	116,03	2300,26	12,68	21,54	30,27
4	1	105	67,90	10501,97	18,79	3,07	32,86

Description of columns:

Column	Translation	Description
Stadtbezirk	Department	ID of the department.
Sozialraum	Social Area	ID of the social area.
Fläche in ha	Area in ha	Size of area in hectare.
Einwohner je km2	Population per sqkm	Population per square kilometre.
Wohnfläche Anteil in %	Perc. of living space	Percentage of areas with living spaces.
Grünfläche Anteil in %	Perc. of green space	Percentage of areas with green spaces.
Von Verkehrslärm betroffene Bevölkerung in %	Perc. affected by traffic noise	Percentage of population affected by traffic noise.

Content:

The data includes the geographic and environmental data for the 179 social areas in Düsseldorf with status 2016. Besides the general information of the corresponding department, social area ID and total size in hectare, the data comprises of four different figures that can be summed up in three social characteristics:

1) Population density: Population per square kilometre.

2) Land use: Percentage of area used as living spaces or green spaces.

3) Noise pollution: Percentage of population affected by traffic noise.

This data will be aggregated on neighborhood level and combined with the information about recommended venues in the neighborhood (provided by foursquare) to cluster the neighborhoods.

Recommended Venues

	Neighborhood ID	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category ID
0	101	51.144373	6.898046	S Düsseldorf-Garath	51.143952	6.897350	4bf58dd8d48988d1fc931735
1	101	51.144373	6.898046	Netto Filiale	51.143584	6.896548	52f2ab2ebcbc57f1066b8b46
2	101	51.144373	6.898046	ALDI SÜD	51.143110	6.896143	52f2ab2ebcbc57f1066b8b46
3	101	51.144373	6.898046	Forst Garath	51.148080	6.902555	52e81612bcbc57f1066b7a23
4	102	51.133328	6.910560	EDEKA Rath	51.131841	6.907200	52f2ab2ebcbc57f1066b8b46

Description of columns:

Column	Translation	Description
Neighborhood ID	Neighborhood ID	ID of the neighborhood.
Neighborhood Latitude	Neighborhood Latitude	Latitude of the centre of the neighborhood.
Neighborhood Longitude	Neighborhood Longitude	Longitude of the centre of the neighborhood.
Venue	Venue	Name of a venue recommended by foursquare.
Venue Latitude	Venue Latitude	Latitude of the venue.
Venue Longitude	Venue Longitude	Longitude of the venue.
Venue Category ID	Venue Category ID	ID of the category that has been assigned to the venue.

Content:

The data consists of the Neighborhood ID and the geo coordinates of its centre as well as the venue name, its geo coordinates and the ID of the category it has been assigned to. This data will be

aggregated on neighborhood and category level to get the count of the venues per category per neighborhood. This will provide a set of features that allows a clustering of similar neighborhoods based on the recommended venues.

As foursquare provides a hierarchical tree for the venue categories, it needs to analysed, if all venues have been categorized on the same hierarchical level, before performing the aggregation. This should provide sligthly less accurate but more consistent results, as e.g. a chinese restaurant could be categorized in one case as 'Asian' (level 2) and in another case as 'Chinese' (level 3). Both categorizations would be correct, but the clustering model would not recognize them as being the same. After the analysis on the categories provided, it will be decided on which level of the hierarchy all venues will be harmonized.

2.3 Data Format

The data was provided as .json, .csv or .geojson. Depending on the data type and structure, it needed to imported and unpacked accordingly.

3 Data Cleaning

The following major cleaning tasks of the data were neccessary:

Population

In March 2014, the neighborhood 'Hubbelrath' got split into 'Hubbelrath' and the new neighborhood 'Knittkuhl'. Therefore the data for 2013 was not including comparable numbers for the population of both neighborhoods in the following years. Therefore the population of 'Hubbelrath' in 2013 was split according to the ratio of the population distribution across those two neighborhoods in 2014.

Social Areas

The 179 social areas needed to be assigned to each of the 50 neighorhoods, as the analysis was performed on the neighborhoods. The 7-digit ID of the social area was a combined key of the 'Social Area ID' and the 'Neighborhood ID'. With this it provided a mapping of each social area to the neighboorhoods. In three cases, however, the 7-digit ID was '0'. When comparing the geo location data of these social areas with the neighborhoods, minor overlaps with multiple neighborhoods were identified. To not loose the information of these social areas, they were assigned to the neighborhood, they had the most overlap with.

Not all figures of the features of the social area data set were provided due to the three possible reasons:

1) The value would logically not be appropriate, due to a known change impactful change in this social area, e.g. replacing a former industrial area with new living space

2) The value was not known or confidential

3) The value was zero

All three placeholders have been replaced with a '0' value in order to be able to leverage the existing data for the analysis. A prediction of values was not possible and would have an unknown impact on the clustering of the data.

4   Data Preparation

Two major tasks for the data preparation had to be performed:

1) Harmonization of venue categories of the recommended venues data of foursquare

Each recommended venue provided by foursquare is assigned to a venue category. The category structure of foursquare, however, is providing a 5-level deep hierarchy of categories. The fifth level thereby is the most detailed one, whereas the first level is the most generic one.

The following table provides the number of categories available per level:

```
Number of categories per category level:
Level 1_ID: 10
Level 2_ID: 459
Level 3_ID: 369
Level 4_ID: 92
Level 5_ID: 14
```

As an example: A chinese restaurant could be assigned to category 'China' (level 3) or to category 'Asian' (level 2). Both categories are correct, however, they are not treated as equal when using these two categories in a clustering. To cluster on comparable feature values, all category assignments have been checked and reassigned to the correct level 2 category.

2) Formatting and consolidating data per neighborhood

As the focus of this project is on the neighborhoods of Düsseldorf, all datasets had to be prepared to provide the features per 'Neighborhood ID'. Based on this, they could be joined and prepared for the data model.

5   Exploratory Data Analysis

5.1 Neighborhoods

Since 2014, Düsseldorf is divided into 50 neighborhoods. This data set provides the geo location data of each neighborhood. Figure 1 shows the area and borders of each neighborhood on the map.

5.2 Population

From the end of 2013 until the end of 2018, the total population of Düsseldorf increased by around 30,000 residents, from approximately 614 thousand to 643 thousand (figure 2). This represents an increase of 4.7%.
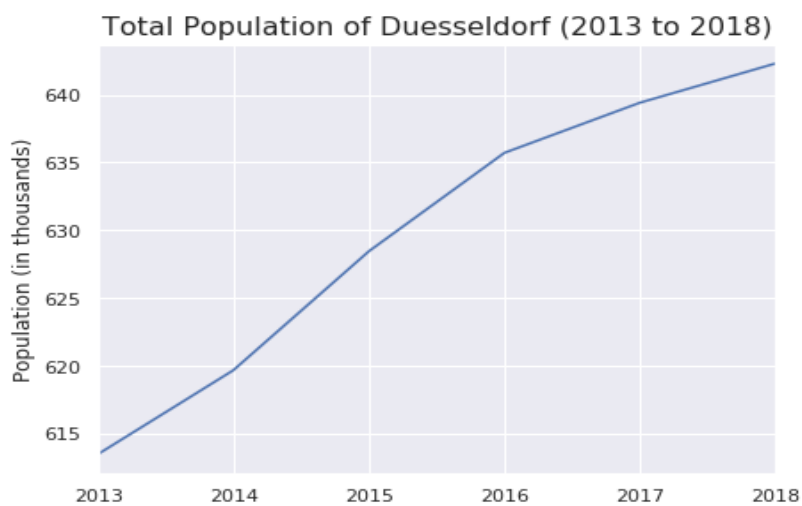

Figure 1: Development of total population 2013-2018

The boxplot of the population per neighborhood per year shows a range of below 1,000 and around 40,000. A spread of 40 times. 50% of the neighborhoods have a population size between 5,000 and 20,000 (figure 3).
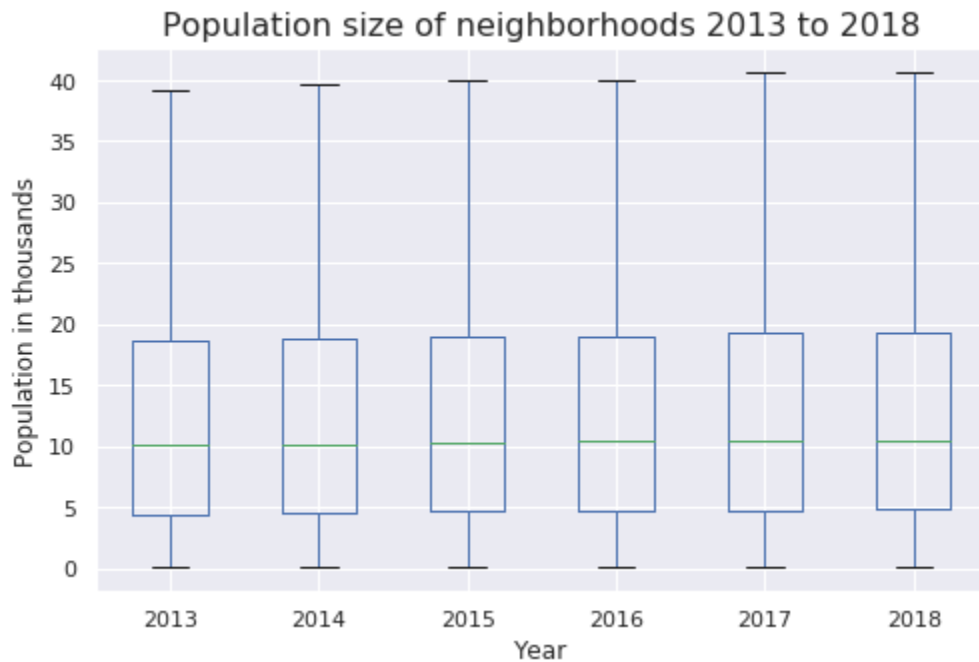
Figure 2: Boxplot of population per neighborhood per year

The distribution of the population size shows this in greater detail. As the distribution is right skewed, the mean is at 12,500 slightly higher than the median at 10,000.
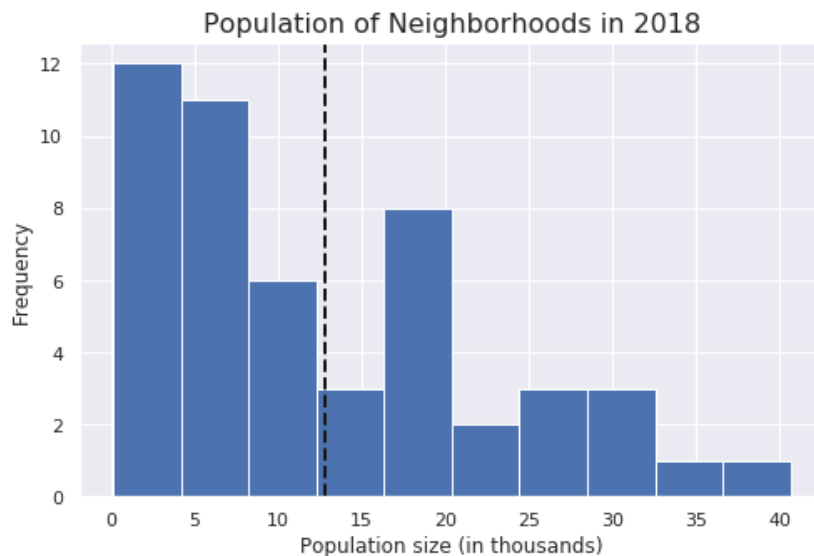

Figure 3: Histogram of popoulation sizes of neighborhoods in 2018

Part of the objective of this project is to identify the neighborhoods with the strongest growth in population and by that providing and indicator for popularity. This provides basis for further analysis

and questions, as e.g. if this leads to an increase in prices or if this is only driven by a greater offering due to new build living spaces? However, answering those questions is out of the scope of this project.

According to the data, Mörsenbroich, Heerdt and Flingern Nord are the only neighborhoods appearing in the Top5 growth in total and relative figures. This separates them from the rest and defines them as most popular neighborhoods.

Top 5: Change of population year-on-year:

Neighborhood ID	Neighborhood	2013	2014	2015	2016	2017	2018	Total
064	Mörsenbroich	14803	620	928	1056	595	124	3323
022	Flingern Nord	22261	613	475	502	766	494	2850
014	Pempelfort	30345	469	635	448	729	438	2719
042	Heerdt	9968	134	213	713	370	287	1717
036	Bilk	39165	543	191	139	514	99	1486

Top 5: Relative change of population year-on-year:

Neighborhood ID	Neighborhood	2013	2014	2015	2016	2017	2018	Total
064	Mörsenbroich	14803	4.2	6.0	6.5	3.4	0.7	22.4
042	Heerdt	9968	1.3	2.1	6.9	3.4	2.5	17.2
022	Flingern Nord	22261	2.8	2.1	2.1	3.2	2.0	12.8
034	Hamm	4066	3.0	2.8	2.2	1.0	1.4	10.8
097	Itter	2174	6.0	1.8	0.5	0.8	0.8	10.2

Figure 4: Total and relative growth of population 2013 to 2018 (Top 5)

Reviewing the Top 5 neighborhoods (figure 6), with a decrease / least increase in population, the population of 'Hafen' decreased by 30% over the course of the five years, whereas the total number is with 48 extremely low. It is also in terms of population the smallest of all neighborhoods. This specific development therefore does not need to be further analyzed and is a good example, why it is important to compare the total and the relative change of each neighbhorhood.

Only two additional neighborhoods, 'Hellerhof' and 'Carlstadt' have seen a slight reduction in population. The significant decrease of population in Carlstadt in 2017 could be a potential target for a further analysis, but is outside of the scope of this project, as the given data will not provide any additional insights into the reasons. Also it seems to have been a one year event, as the population remained stable in 2018.

The development of Hellerhof was not constantly negative over the five years, as it was +2.2% in 2016. But before choosing Hellerhof as a potential target area to move to, it makes sense to get additional information on the negative development over the last years.

Neighborhood ID	Neighborhood	2013	2014	2015	2016	2017	2018	Total
102	Hellerhof	6071	-91	-67	128	-72	-87	-189
033	Hafen	157	-18	-8	-1	-14	-7	-48
012	Carlstadt	2323	-8	176	97	-309	6	-38
094	Reisholz	3711	27	98	-99	40	-56	10
075	Knittkuhl*	2066	-11	-7	30	13	6	31

Bottom 5: Relative change of population year-on-year:

Neighborhood ID	Neighborhood	2013	2014	2015	2016	2017	2018	Total
033	Hafen	157	-11.5	-5.8	-0.8	-10.8	-6.0	-30.6
102	Hellerhof	6071	-1.5	-1.1	2.2	-1.2	-1.5	-3.1
012	Carlstadt	2323	-0.3	7.6	3.9	-11.9	0.3	-1.6
094	Reisholz	3711	0.7	2.6	-2.6	1.1	-1.5	0.3
083	Vennhausen	10533	-0.2	-0.7	0.7	0.2	0.6	0.6

Figure 5: Total and relative growth of population 2013 to 2018 (Bottom 5)

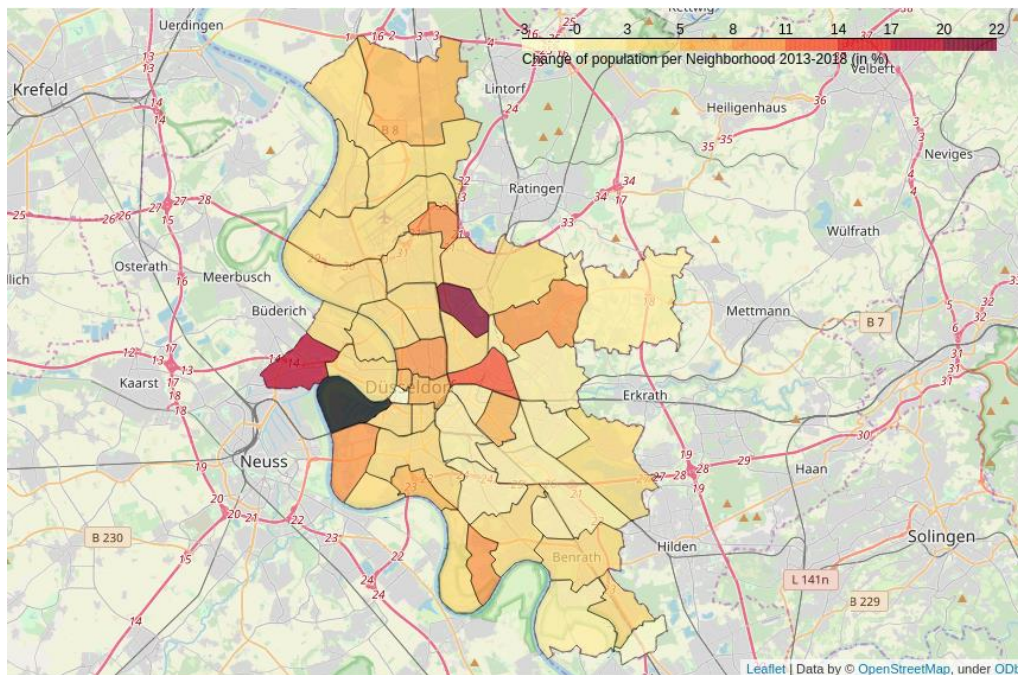Figure 7 provides a visual representation of the relative change of population 2013 to 2018.

Figure 6: Relative change of population per neighborhood 2013 to 2018

5.3 Social areas

The social areas, as defined by the city of Düsseldorf, divide Düsseldorf into 179 areas. Figure 8 provides a visual comparison of the social areas vs. the neighborhoods.
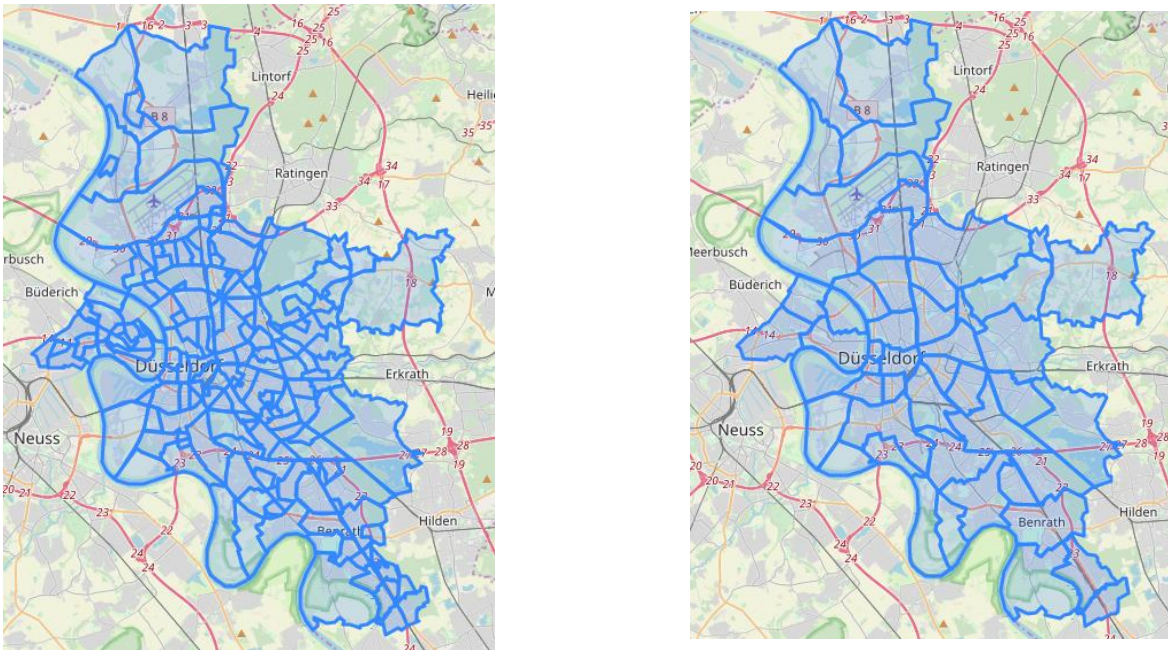

Figure 7: Social areas and neighborhoods

Figure 9 provides a detailed overview and summary of the features of the social area dataset and the spread of the data points. Besides the percentage of living space of the overall space of the social area, all datasets show a broad spread of data with outliers in the upper area.
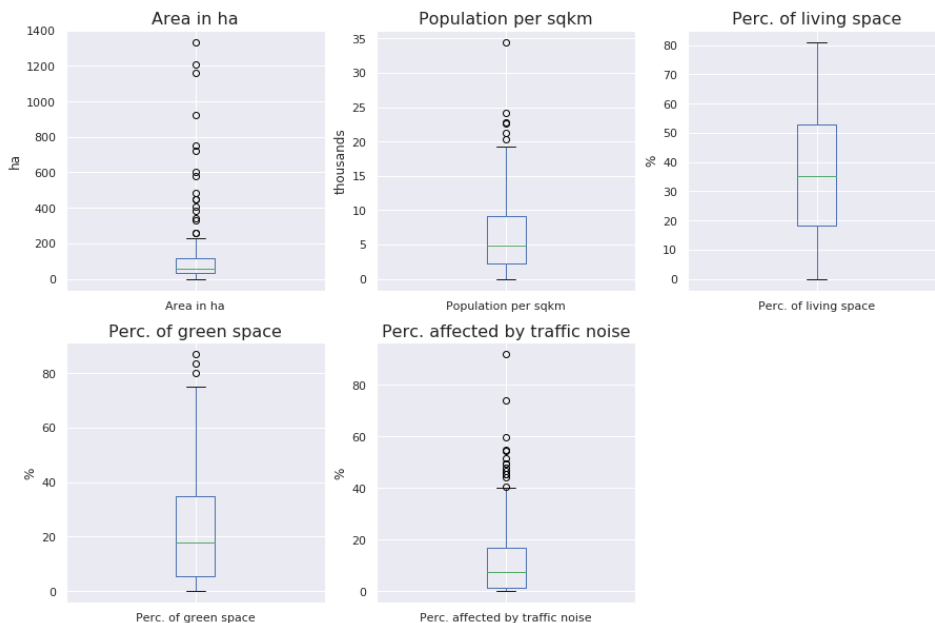
Figure 8:

Features of social areas

5.4 Venues

Foursquare returned a total of 1,091 for all 48 of the 50 neighborhoods. For Lohausen and Itter, foursquare did not provide recommended venues in a radius of 700m around the centre of the neighorbood.

As shown in figure 10, around 22 venues have been returned in average per neighborhood. For five of the 50, more than 60 recommended revenues were returned. Thereby it needs to be considered, that the maximum number of venues per neighborhood was limited to 100.
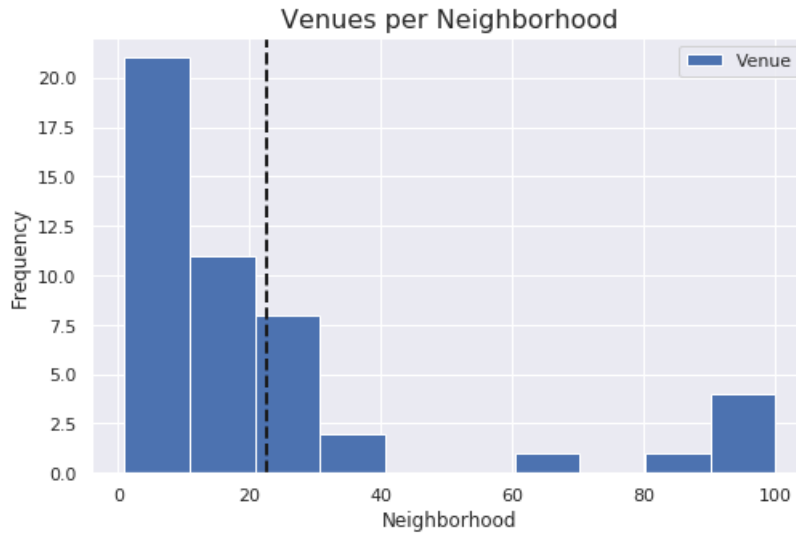

Figure 9: Number of venues per neighborhood

6   Modeling – Clustering of Neighborhoods

To prepare the data for the Kmeans cluster model, the features of the social areas were standard scaled, i.e. the mean of the data was removed and scaled to the unit variance.

For the venue data, the venues where grouped by category per neighborhood to calculate the percentage of each venue category for each neighborhood. The data features thereby were limited between 0 and 1.

All features were combined into one DataFrame. With the help of the elbow method, the ideal number of clusters of five can be identified.
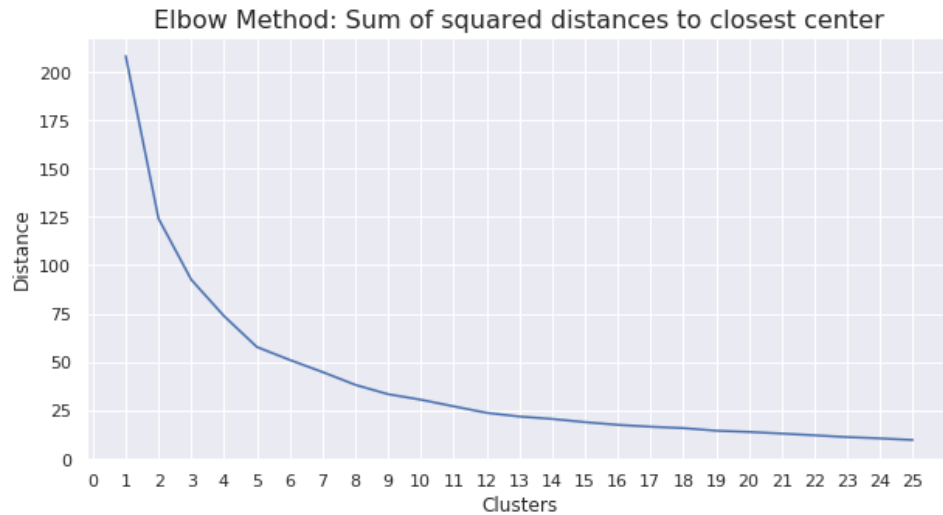
Figure 10: Elbow method

7   Conclusion

The following table provides a description of the five clusters identified, based on their attributes:

Cluster	Description
0	Assumably lots of stores, businesses and even industry, as living and green space only make for about 35% of the overall space. Top venues do include only very minor public / cultural sights.
1	Very green with lots of green spaces, Beaches and Sport Areas. Only few venues for going out, such as bars or restaurants.
2	Very crowded, most dense population per sqkm and highest share of living space. Lots of venues for Food & Drink and Bars.
3	Very noisy and assumably lots of stores, businesses and Hotels.
4	Around two thirds of area is made of living and green space, while medium density of population per sqkm. Potentially more spaceous houses and flats than in Cluster 2. Lots of venues for Food & Drink and Italian restaurants, but also natural sites.

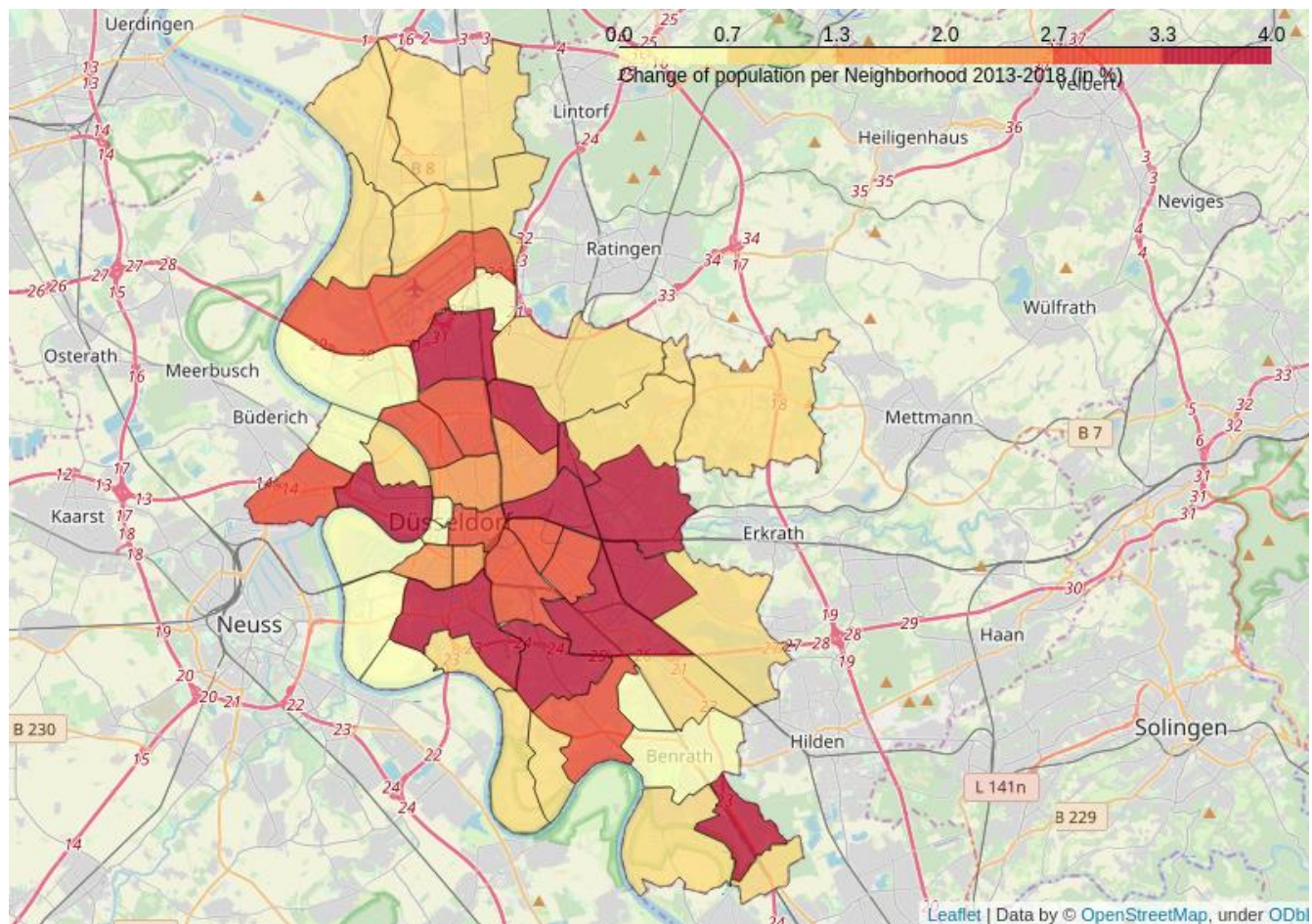Figure 12 shows a visual representation of the clusters across Düsseldorf.

Figure 11: Neighborhood Clusters

Mörsenbroich and Flingern Nord, two of the three strongest growing neighorhoods of Düsseldorf are part of Cluster 4, providing a potentially interesting mix of living and green space. It will be interesting to see, how especially the percentage of green space will develop over time, as a growing population demands additional living spaces. Herdt being less crowded but stronger affected by traffic noise, is part of cluster 3.

Hellerhof, being the neighborhood with the strongest absolute decrease in population is part of Cluster 1, which might still be an interesting living area judging from the facts at hand: near to the city, but still green and very low traffic noises.

The objective of this project is to provide an initial idea of the clustering of the different neighborhoods of Düsseldorf and support this by an analysis of the development of the population in the last five years. It should not be understood as a complete guide or advice on where to move to, but raise additional questions and provide a first starting point to find an answer, on where to move to in Düsseldorf. It definitly also depends on the personal taste and expectations.