

Coursera Capstone Project

Battle of the Neighbourhoods – Differences between Hamburg Districts



February 2020
Beate Goerig

Contents

1. INTRODUCTION	3
2. DATA	3
2.1 DATASET 1: FOURSQUARE DATA	3
2.2 DATASET 2: DATA OF HAMBURG DISTRICTS	3
3. METHODOLOGY	4
3.1 PART 1 – COMPARISON OF DISTRICTS	4
3.2 PART 2 – LOCATION FOR NEW ITALIAN RESTAURANT	4
4. RESULTS	5
4.1 PART 1 – COMPARISON OF DISTRICTS	5
4.1.1 Analysis of Altona	5
4.1.2 Analysis of Wandsbek	9
4.1.3 Analysis of Harburg	13
4.2 PART 2 – LOCATION FOR NEW ITALIAN RESTAURANT	18
5. DISCUSSION	20
6. CONCLUSION	20

1. INTRODUCTION

Hamburg consists of 7 districts, of which the districts of Altona, Wandsbek and Harburg, among others, came to Hamburg by the Greater Hamburg Act in 1937. All three were independent cities with their own city centre before they were incorporated. These former city centres are still recognisable by the neighbourhood that has the same name as the whole district.

My aim is it to analyse if the centre of activity can still be found in these formerly independent city centres or whether it has moved somewhere else. I aim to do this by examine what are the most and least urban neighbourhoods based on the population density and analysing if they can still be counted as centres of activity by analysing what kind of venues can be found there.

As a second approach I'd like to give a recommendation which neighbourhood in which district would be the best location for opening up an Italian Restaurant because one of the most liked type of restaurant in Germany are Italian Restaurants.

Stakeholders

The main stakeholders of this study are:

- The City of Hamburg because they can get a better view on how the three districts starting from (almost) the same point developed differently over time. This could lead to insights what needs to be done to improve the quality of living in these districts.
- The (imaginary) owner of a successful Italian restaurant in downtown Hamburg who wants to open another restaurant in one of the mentioned districts.

2. DATA

To solve the above-described problem I am going to use 2 different datasets:

2.1 DATASET 1: FOURSQUARE DATA

I am going to use Foursquare data to find all the popular venues of the city. This dataset will give me the name, category, latitude and longitude of each venue within a given radius around the neighbourhoods and districts of interest.

Example of the Data:

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Hamburg-Altstadt	53.54992	9.99893	GOOT - Finest Cuts	53.548509	10.000675	Comfort Food Restaurant
1	Hamburg-Altstadt	53.54992	9.99893	O-ren Ishii	53.548291	9.997890	Vietnamese Restaurant
2	Hamburg-Altstadt	53.54992	9.99893	Picasso	53.549934	9.995627	Spanish Restaurant
3	Hamburg-Altstadt	53.54992	9.99893	Thalia Theater	53.551897	9.998384	Theater
4	Hamburg-Altstadt	53.54992	9.99893	Mi Chii	53.548320	9.998530	Vietnamese Restaurant

2.2 DATASET 2: DATA OF HAMBURG DISTRICTS

The names and geolocation data for the neighbourhoods of Hamburg can be obtained from the German Wikipedia page of the districts and neighbourhoods of Hamburg:

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

It is the second table on this page.

This information can only be gathered from the German page, because it is not included in the English equivalent at https://en.wikipedia.org/wiki/Boroughs_and_quarters_of_Hamburg.

The data provided at the German Wikipedia page is particularly useful, as it provides the density of population of each neighbourhood („Bevölkerungsdichte (Einwohner/km²)“) which I take as indicators if a neighbourhood has a more or less urban character.

Example of the Data:

Stadtteil	Ortsteile	Bezirk	Fläche (km²)	Einwohner	Bevölkerungsdichte (Einwohner/km²)	Koordinaten	Karte
Hamburg-Altstadt	101–102	Hamburg-Mitte	2,4	2272	947	53° 33' 0" N, 10° 0' 0" O	
HafenCity	103–104	Hamburg-Mitte	2,2	4592	2087	53° 32' 28" N, 10° 0' 1" O	
Neustadt	105–108	Hamburg-Mitte	2,3	12.920	5617	53° 33' 7" N, 9° 59' 8" O	
St. Pauli	109–112	Hamburg-Mitte	2,5	22.436	8974	53° 33' 25" N, 9° 57' 50" O	
St. Georg	113–114	Hamburg-Mitte	2,4	11.384	4743	53° 33' 18" N, 10° 0' 44" O	
Hammerbrook	115–118	Hamburg-Mitte	3,0	4323	1441	53° 32' 43" N, 10° 1' 50" O	
Borgfelde	119–120	Hamburg-Mitte	0,8	7696	9620	53° 33' 17" N, 10° 2' 4" O	
Hamm	121–127	Hamburg-Mitte	4,0	38.773	9693	53° 33' 39" N, 10° 3' 28" O	
Horn	128–129	Hamburg-Mitte	5,9	38.799	6576	53° 33' 14" N, 10° 5' 24" O	

However, the geo location data is provided in Degree, Minute, Second format, so that the decimal coordinates are added via geocoder for the sake of simplicity.

3. METHODOLOGY

3.1 PART 1 – COMPARISON OF DISTRICTS

As I am comparing the structure of three different districts, I extracted from the combined data set of foursquare data and Hamburg districts data the individual data sets for Altona, Wandsbek and Harburg, the districts of interest.

For a comprehensive portrait of each district, I created overviews of their venues in form of a map of the venues in the district, a graph for the frequency of venues and a word cloud for the venues.

The information which neighbourhood in each district is the most urban or most rural one, I extracted the lines with highest or lowest density value and analyses them by venue, frequency of venues and word cloud.

3.2 PART 2 – LOCATION FOR NEW ITALIAN RESTAURANT

To find the best neighbourhood to open a new Italian Restaurant I first checked how many Italian restaurants are accounted for in Hamburg in general and in the aforementioned districts of interest.

To find out in which neighbourhoods the market for a new restaurant is not saturated I performed a kmeans clustering based on mean occurrence of venue category.

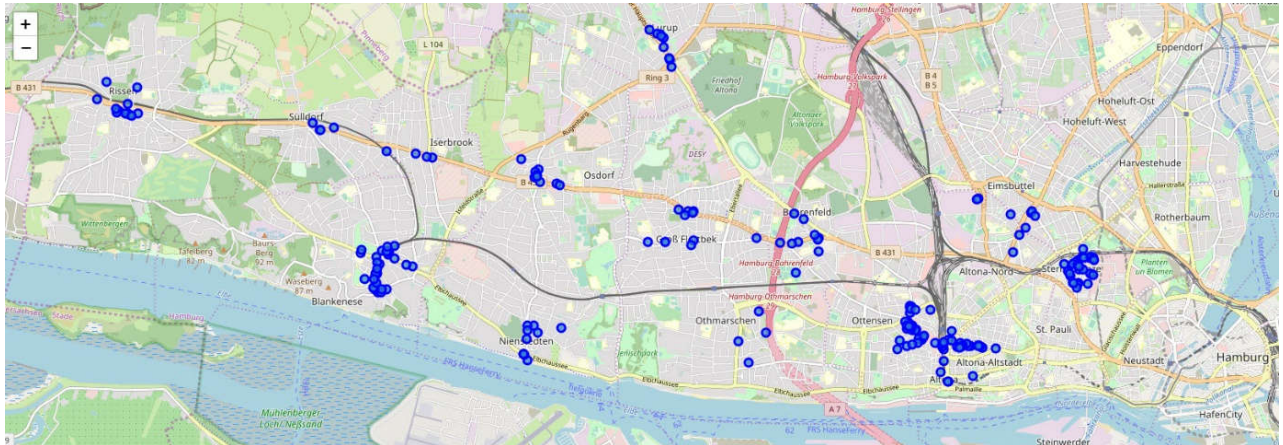
From the resulting cluster indicating a not saturated market the most urban neighbourhoods, that is with population or density above the mean of the cluster, were chosen as recommendations.

4. RESULTS

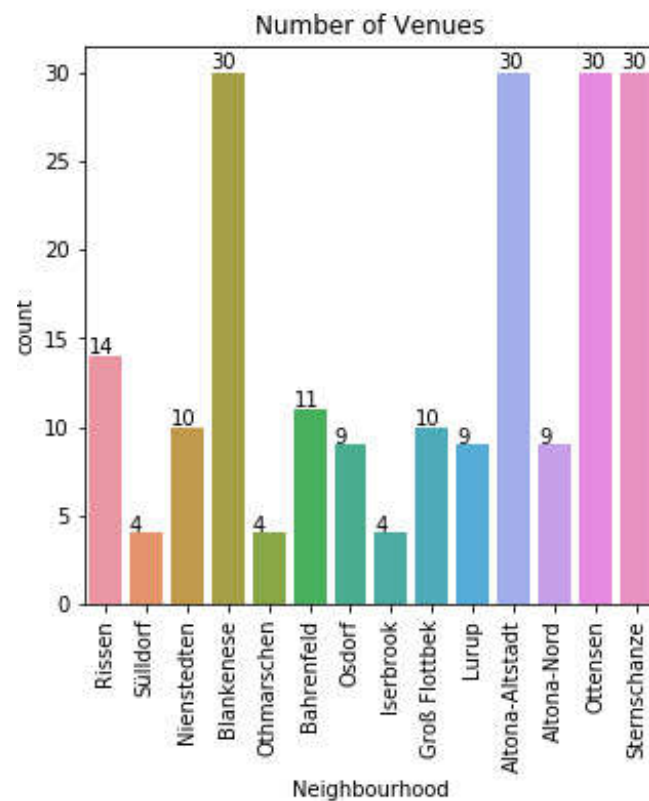
4.1 PART 1 – COMPARISON OF DISTRICTS

4.1.1 Analysis of Altona

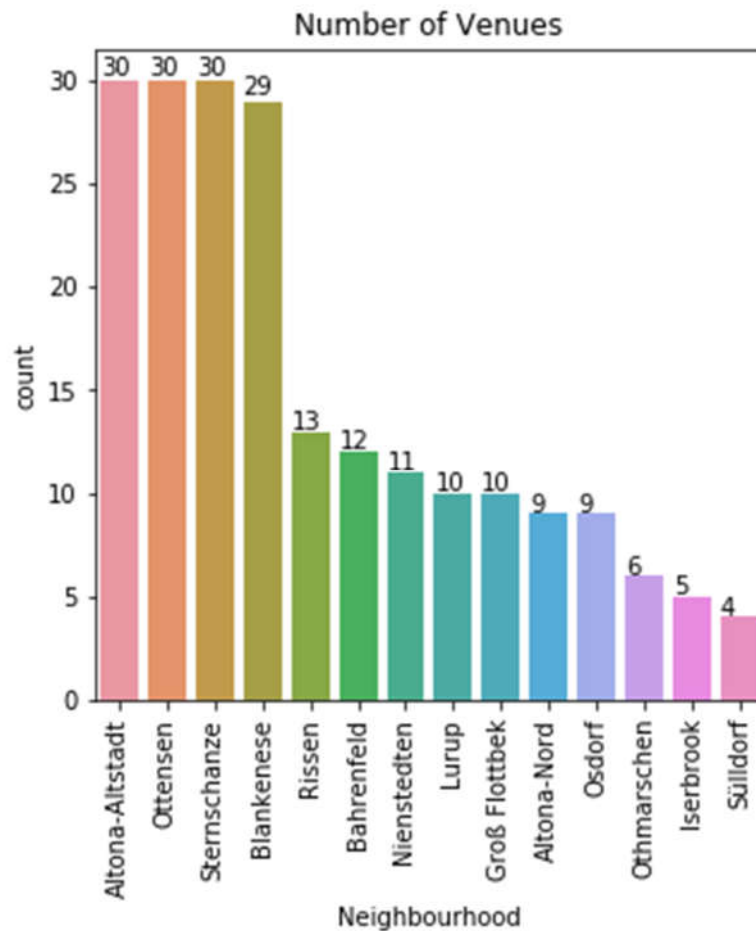
The map of Altona shows a clustering of venues around certain hot spots with a few bigger clusters and several smaller ones:



The distribution graph confirms this impression with 4 neighbourhoods having considerably more venues than the others:



This discrepancy comes even clearer when we look at an ordered version of this frequency graph:



The word cloud for Altona shows a wide variety of venue categories with the biggest categories of Restaurant and Supermarket.



The neighbourhood with the highest density and therefore the most urban neighbourhood is Sternschanze and the one with the lowest density indicating the most rural neighbourhood is Rissen.

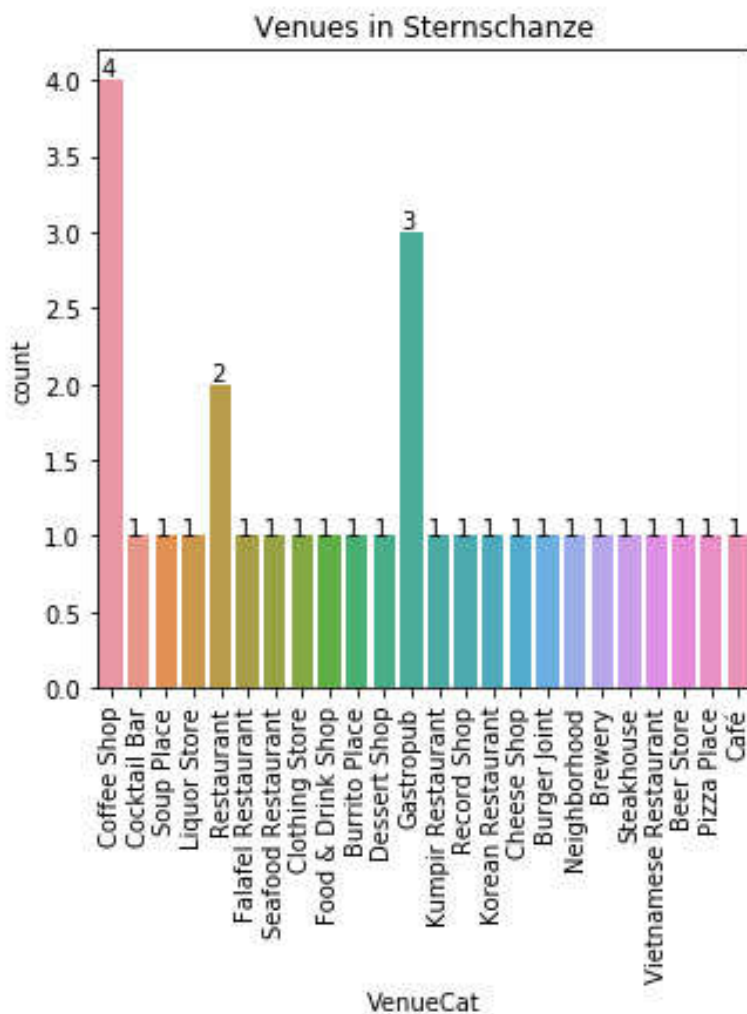
Neighbourhood	Sternschanze	Neighbourhood	Rissen
Population	8095	Population	15.763
Density	16190	Density	944
Name: 304, dtype: object		Name: 477, dtype: object	

As the former centre of the independent city is now the neighbourhood called Altona-Altstadt, this centre seems to have moved to Sternschanze.

The seemingly discrepancy between density and population can be explained by the size of the respective neighbourhoods as Rissen has an area size of 16,7 km² versus Sternschanze a mere 0,5 km².

To analyse these neighbourhoods in more depths let's compare their distribution of venues and their word clouds.

For its small area there are a lot of venues recorded for Sternschanze as the distribution graph and the word cloud show.

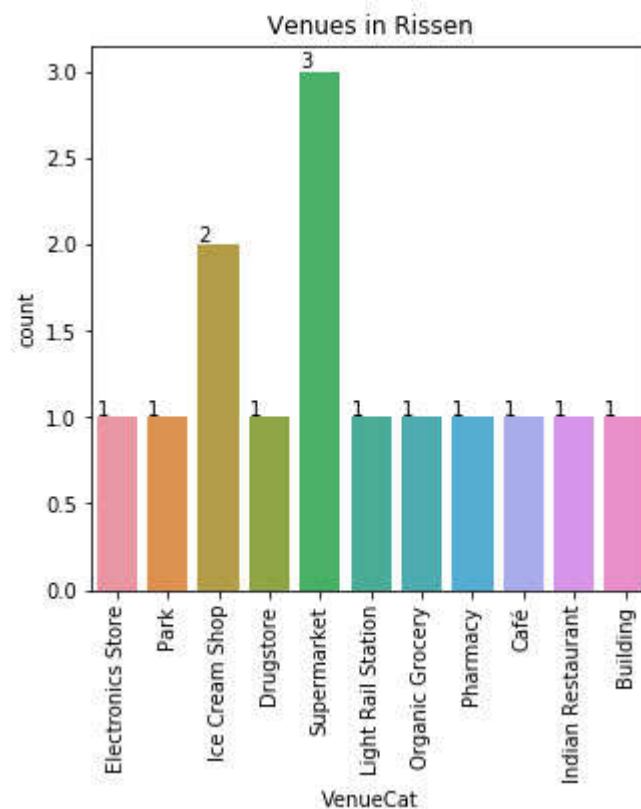


Eye-Catching is the amount of venues who are just once in the area.

According to the word cloud, the most frequent venues are Coffee Shops and Restaurants.



In Rissen however are less venues with a lesser variety recorded.



Remarkable – just as it was in the Data for Sternschanze - is the amount of venues whose frequency is only 1. This points to a possible disadvantage of the foursquare data for Hamburg: there seems not to be enough data for a meaningful analysis.

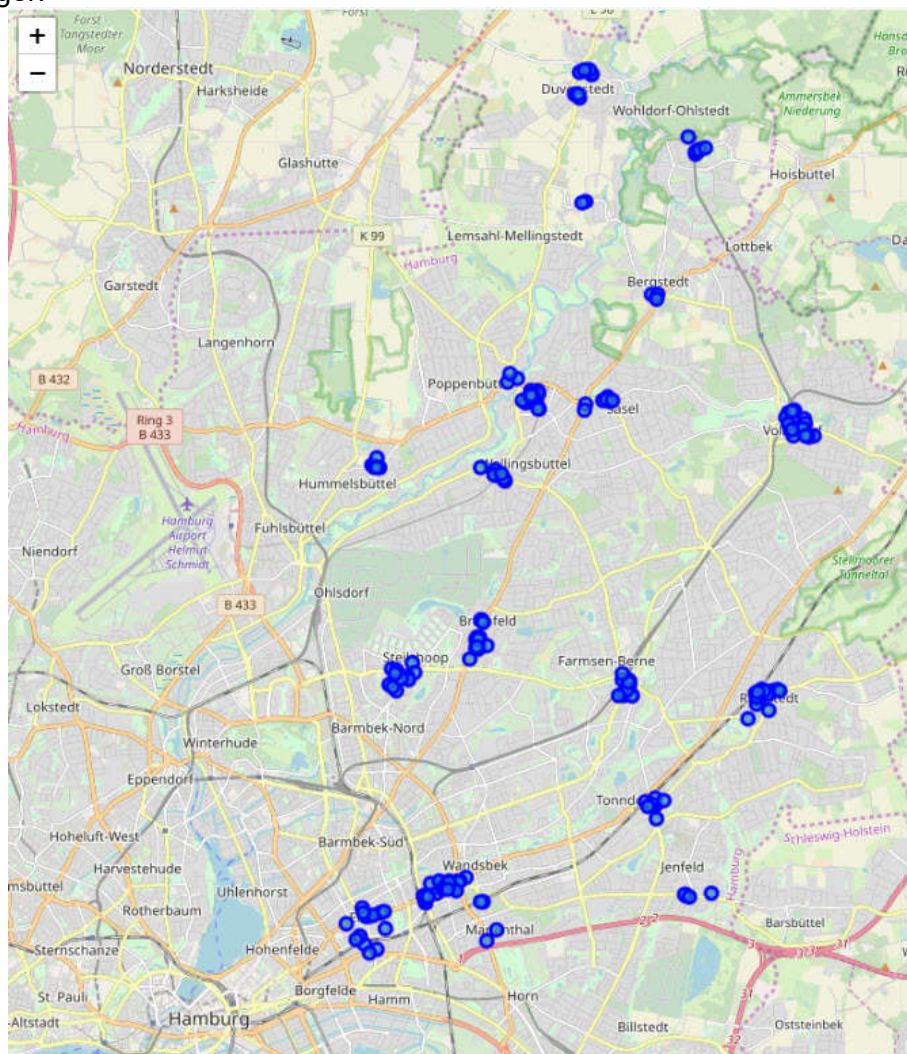
The more rural characteristic of Rissen shows also in the frequency of the venue Supermarket and Ice Cream Shop displayed in the work cloud.



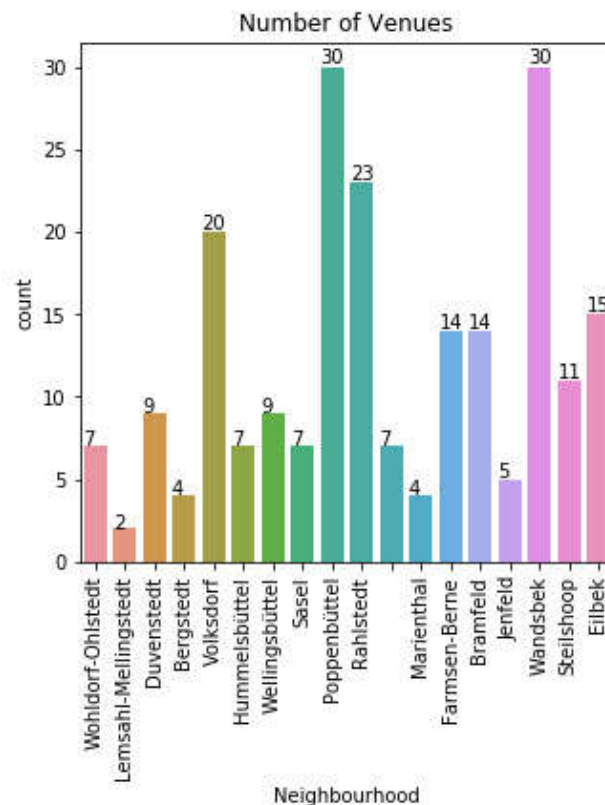
4.1.2 Analysis of Wandsbek

The next district we are looking at is Wandsbek.

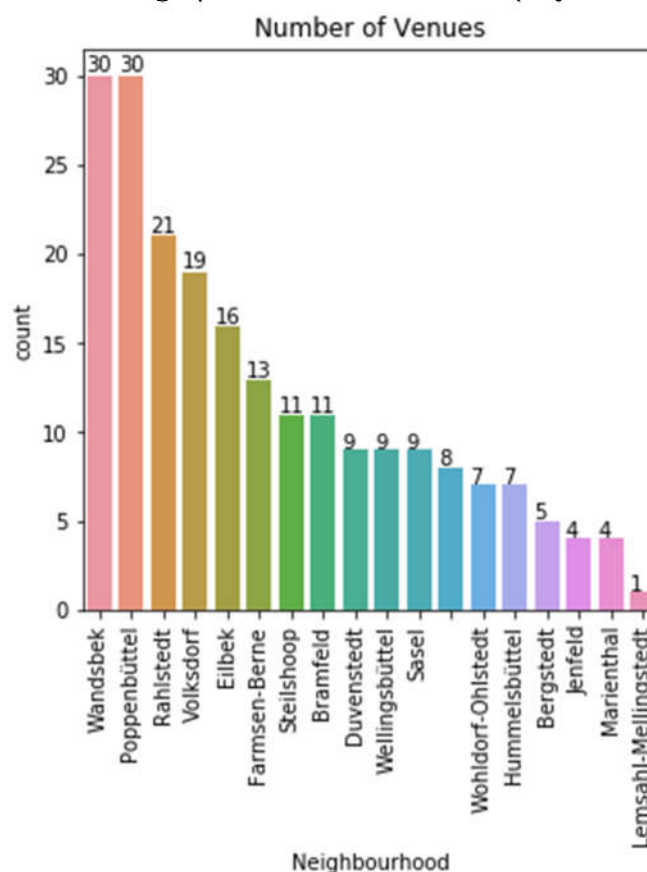
The map of Wandsbek shows the same clustering of venues like the map of Altona, if not even stronger.

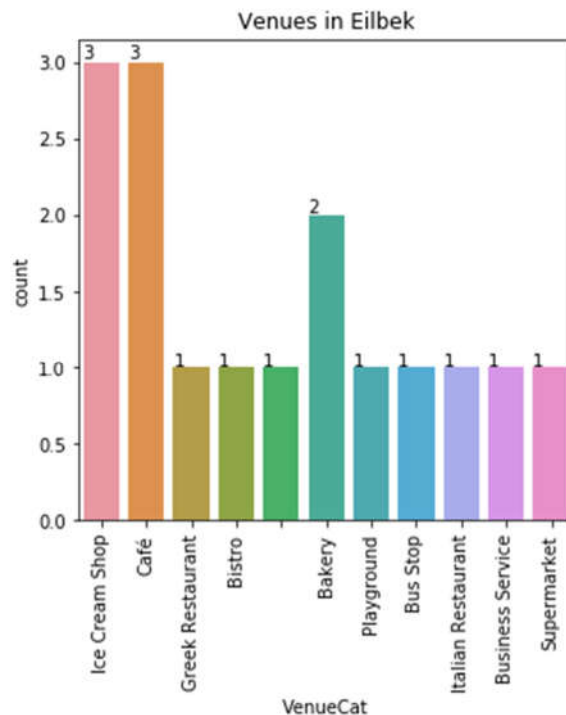


The distribution graph of the venues shows a more mixed distribution. There are certainly neighbourhoods with a lot more venues than the most of the others, but in between there are neighbourhoods with a medium range of venues.



A comparison with the same graph in a more ordered display confirms this impression.

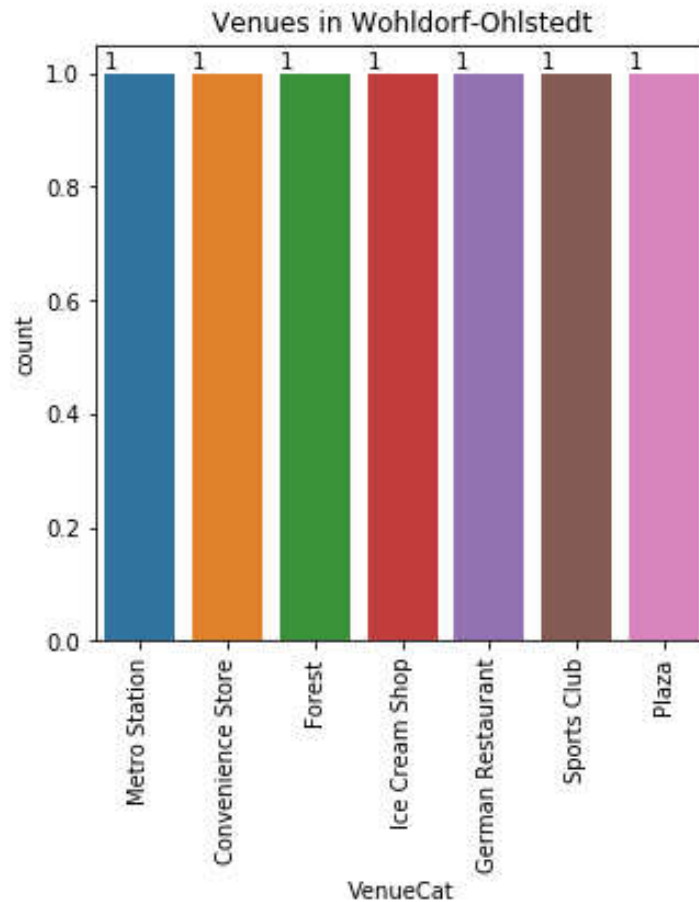




This again leads to a more prominent display of those venues, that are listed more than once in the word cloud: Ice Cream Shop, Café and Bakery.



The venues of Wohldorf-Ohlstedt drive this problem to extremes, as there are equally one each of 7 venues.



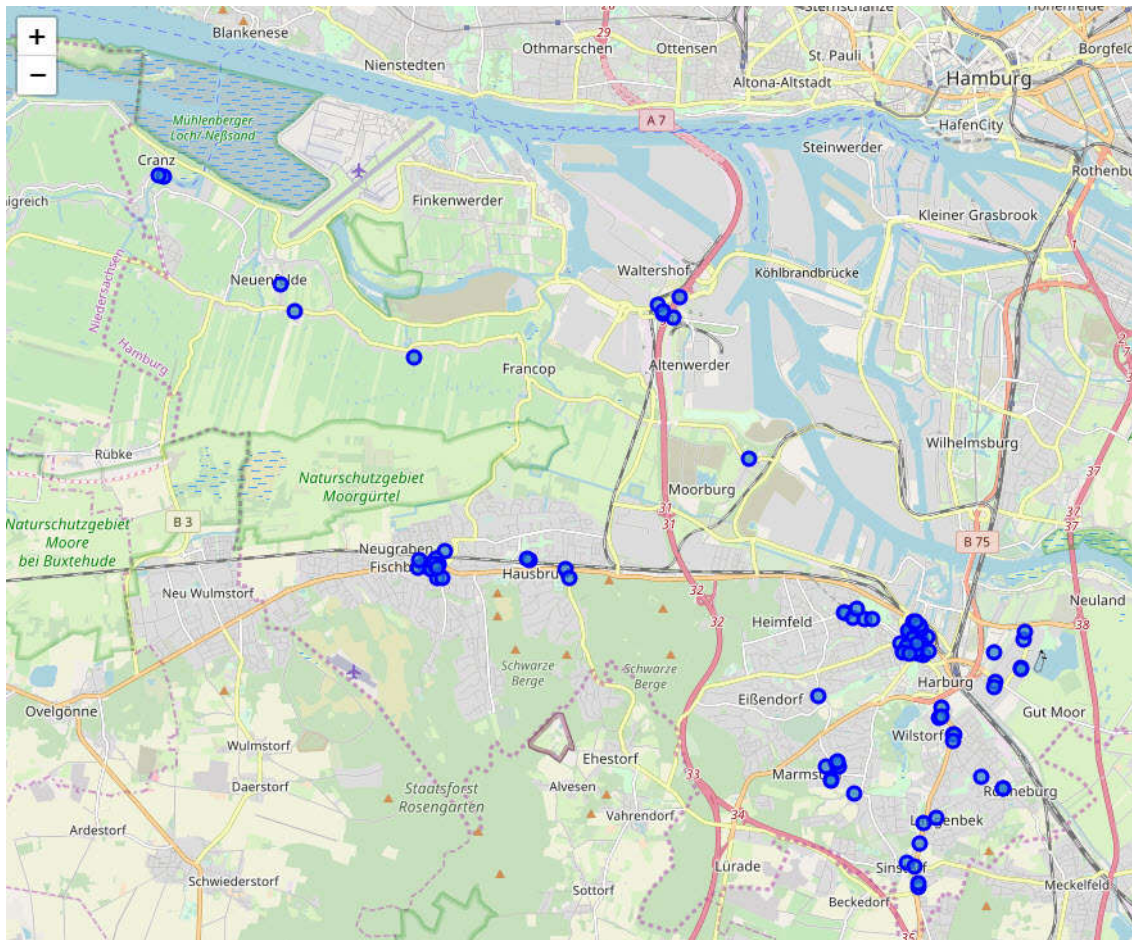
Nevertheless, the lack of venues and the characteristic of those present, underline the rural character of this neighbourhood.



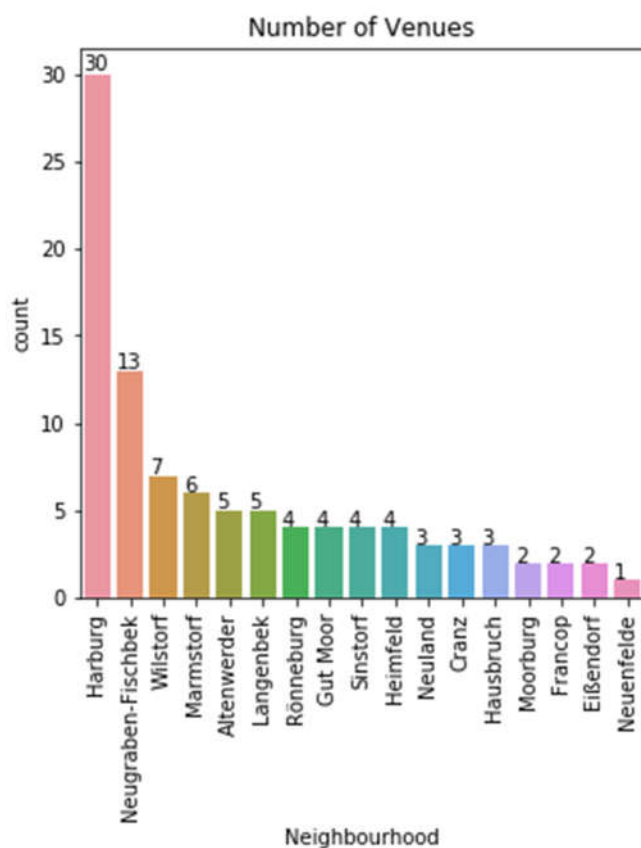
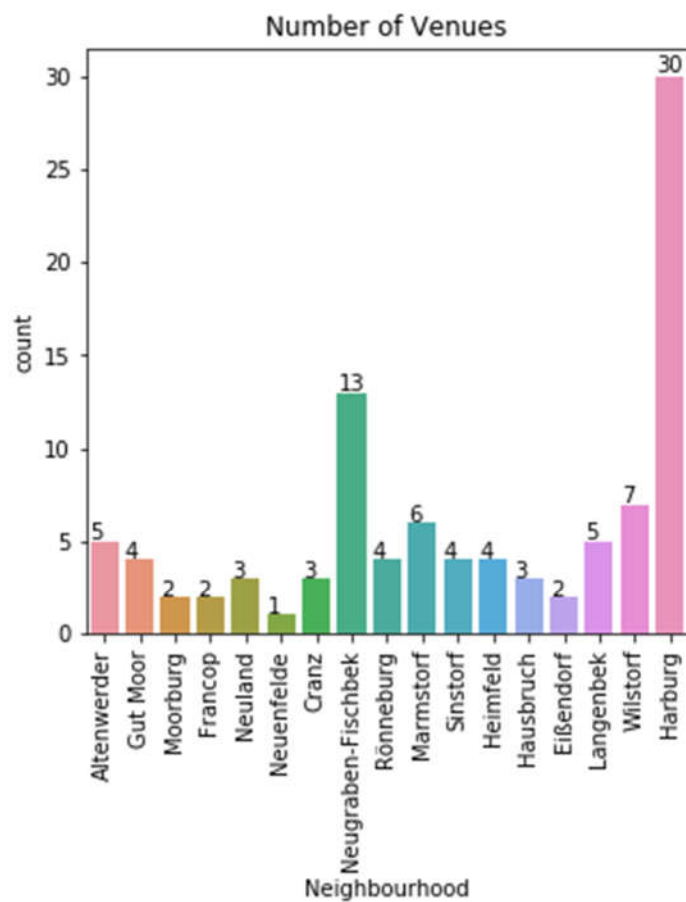
4.1.3 Analysis of Harburg

The last district we are taking a closer look at is Harburg.

The map with its large green areas shows that Harburg is likely to have the biggest discrepancy between urban and rural neighbourhoods. Apart from that also this map shows the familiar clustering of venues around a few hot spots.



The distribution of venues on the frequency graph is similar to the distribution of venues in Altona, one Neighbourhood with a lot of venues and most of the rest in the single-digit range of venues, as can be seen in the graphs on the next side.



Neighbourhood	Harburg	Neighbourhood	Altenwerder
Population	25.912	Population	3
Density	6478	Density	0
Name: 1184, dtype: object		Name: 1250, dtype: object	

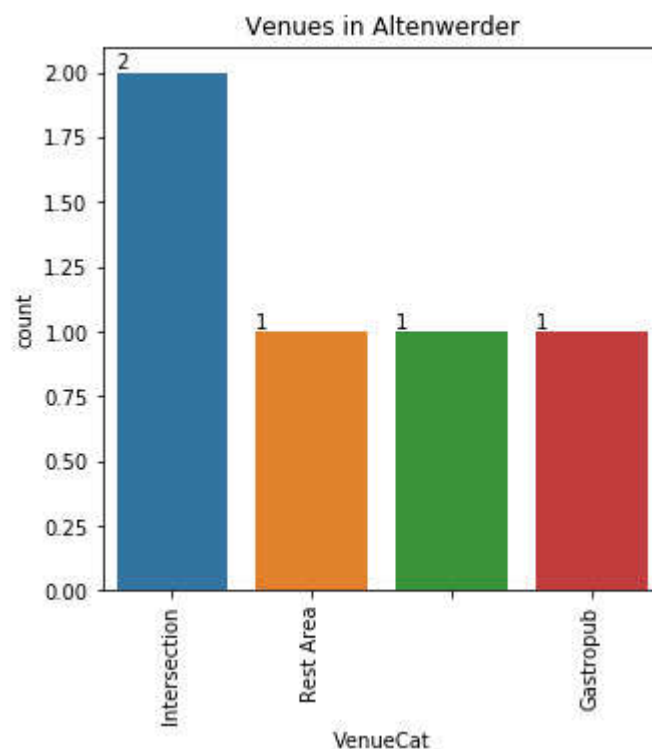
Bar chart titled "Venues in Harburg (Neighbourhood)" showing the count of venues for each VenueCat. The y-axis is labeled "count" and ranges from 0.0 to 3.0. The x-axis is labeled "VenueCat". The bars are color-coded by VenueCat: pink (1), orange (1), brown (2), olive (2), green (3), teal (1), dark teal (1), light teal (2), blue (1), light blue (2), light blue (2), light blue (2), light blue (2), light blue (1), light blue (2), light blue (1), light blue (1), light blue (1), light blue (1).

VenueCat	count
Department Store	1
Supermarket	1
Brazilian Restaurant	1
Drugstore	2
Steakhouse	2
History Museum	2
Mexican Restaurant	1
Bakery	3
Irish Pub	1
Bank	1
Market	1
German Restaurant	2
Restaurant	1
Falafel Restaurant	1
Hotel	2
Italian Restaurant	2
Seafood Restaurant	1
Fast Food Restaurant	2
Cafe	1
Indian Restaurant	1
Shopping Mall	1

The word cloud looks correspondingly diverse with restaurant and bakery in prominent places.



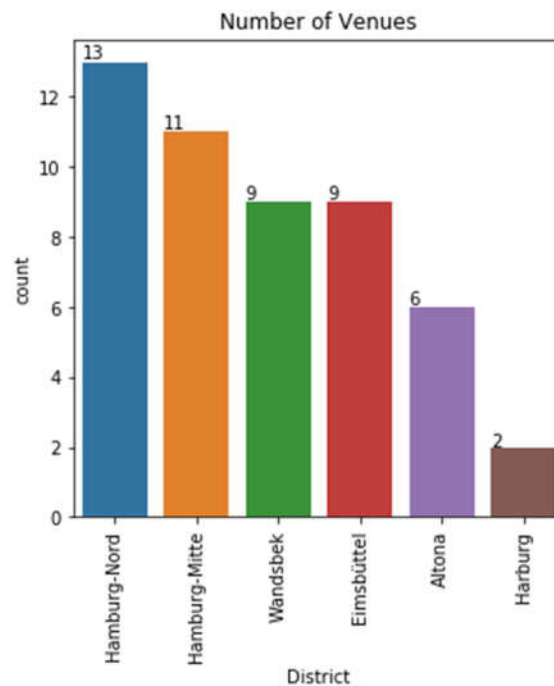
As for a neighbourhood almost without any residents can be expected, the distribution graph of venues shows for Altenwerder only a few non-hospitable venues. As the word cloud reveals, the missing x-label is “Train Station”.



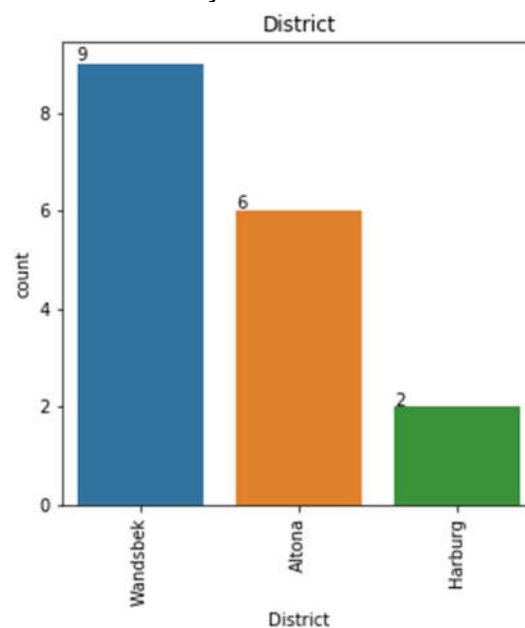
4.2 PART 2 – LOCATION FOR NEW ITALIAN RESTAURANT

To find the best neighbourhood to open a new Italian Restaurant I first checked how many Italian restaurants are accounted for in Hamburg in general and in the aforementioned districts of interest. As the graphs show, the result for this is that Harburg has the least number of restaurants.

Number of Italian Restaurants in Hamburg:

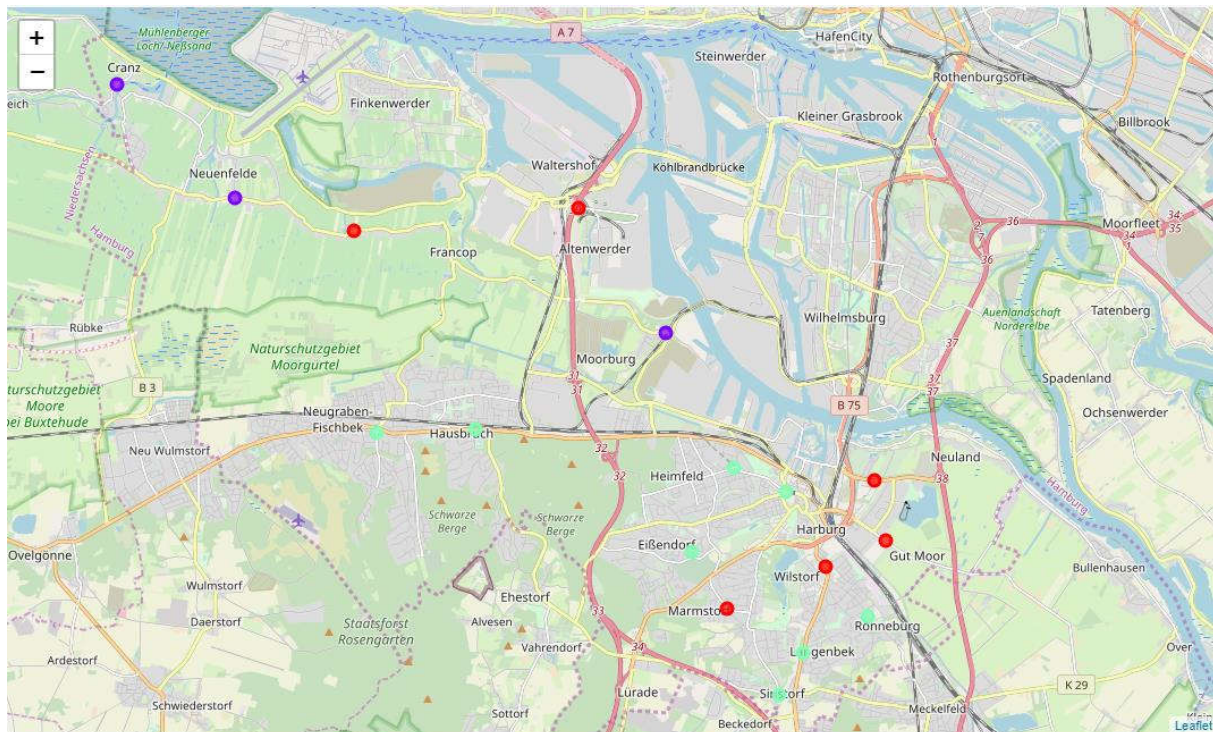


Number of Italian Restaurants in the analyzed Districts:



To cluster the neighbourhoods into three clusters I used the K-Means clustering Algorithm. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. It uses iterative refinement approach.

The below map shows the distribution of the different types of clusters created.



Cluster 0 (red dots): The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated. Actually, no restaurant at all seems to be in this cluster

Cluster 1 (purple dots): The Total and Total Sum of cluster1 has highest value. It shows that the markets are saturated. Number of restaurants are very high.

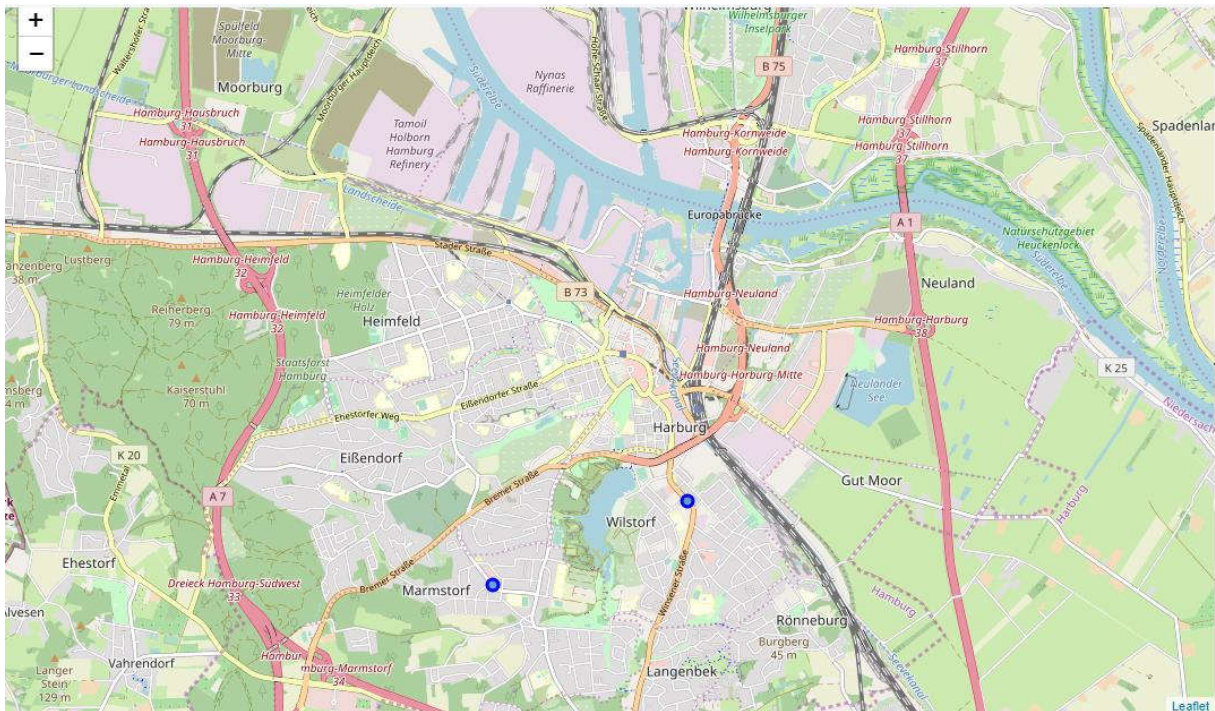
Cluster 2 (sage dots): The Total and Total Sum of cluster1 has medium height. It shows that the markets are not yet saturated.

	Eastern European Restaurant	Falafel Restaurant	Fast Food Restaurant	German Restaurant	Greek Restaurant	Indian Restaurant	Italian Restaurant	Mexican Restaurant	Restaurant	Seafood Restaurant	Szechuan Restaurant	Turkish Restaurant	Total	Total Sum
cluster0	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
cluster1	0.000	0.000000	0.000000	0.888889	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.888889	1.777778
cluster2	0.025	0.004167	0.070833	0.059615	0.09375	0.004167	0.008333	0.004167	0.004167	0.004167	0.009615	0.009615	0.297596	0.595192

Based on this result I concentrated on cluster 0 for further analysis. A visual inspection of the cluster on the map shows that this cluster includes neighbourhoods which seem to be less and more inhabited. In order to take this into account, I decided to select only the neighbourhoods with the highest population density or the largest population for my recommendation.

	Neighbourhood	District	Population	Density	Latitude	Longitude	Total	Cluster_Labels
7	Marmstorf	Harburg	8929.000	1539	53.43991	9.96058	0.0	0
3	Wilstorf	Harburg	17.506	5149	53.44746	9.99021	0.0	0

The following map illustrates where the resulting neighbourhoods Marmstorf and Wilstorf are located.



5. DISCUSSION

In this section I would like to point out some difficulties with the data used.

1. Although the Hamburg data as scrapped from the Wikipedia page had the geolocation data already included, they seemed to be in a format that didn't respond well to the attempt to convert them from Degree, Minute, Second coordinates to decimal coordinates. So, to not waste too much time figuring out a solution, the decimal coordinates were added via geocoder for the sake of simplicity.
2. The analysis of the neighbourhoods in the chosen districts revealed a quite thin data basis for a really meaningful analysis. A possible reason for this might be, that foursquare is not as popular in Germany, as it is in the US and Canada. A more complete dataset, like Google Maps data for example, might lead to completely different results, especially concerning the recommendation for opening a new Italian restaurant.
3. Nevertheless, a further analysis, like comparing the identified most urban and most rural neighbourhoods with the characteristics of the neighbourhoods that were the assumed centres of the independent cities before 1937, might bring further insights.

6. CONCLUSION

My first question to analyse was, if the centres of the three districts of Altona, Wandsbek and Harburg, which were incorporated into the City of Hamburg in 1937, remained as they were before the incorporation.

The data revealed, that, if you want to call the neighbourhood with the highest population density the centre of a district, that the centres of Altona and Wandsbek moved away from the assumed centres before the incorporation into Hamburg. That this was not the case with Harburg might be explained by the separation of Harburg from the centre of Hamburg by the river Elbe and the Hamburg Port located between Harburg and Hamburg centre.

According to the (limited) data based on foursquare, the most promising locations for a new Italian restaurant are the neighbourhoods of Marmstorf and Wilstorf in the Harburg District.