An analysis of neighborhoods in Hamburg for couples with children

Xiaxia Xu

23.06.2021

1. Introduction

1.1 Background

Hamburg is a beautiful and modern city. According to Numbeo, Hamburg is the top 20 out of 87 cities in Europa ranked by high Quality of Life Index (https://www.numbeo.com/quality-of-life/in/Hamburg). The city of Hamburg is divided into seven boroughs with 104 neighborhoods. Each of the neighborhoods has its own specific characters. Whether you are a worker moving from outside of the city, or a student who wants to rent a flat close to his college at an economical price or if you are a new parent and want to move to a quiet place that offers plenty of activities for your kids, finding a neighborhood that meets the personal requirements the most is challenging. This project aims to give you an overview of the neighborhoods in Hamburg that could fit your personal requirements by clustering the neighborhoods into several groups based on the features (rental price, the most common venue categories, etc.) of each neighborhood.

1.2 Problem statement: which neighborhoods are most suitable for a couple with 3 children?

More specifically, in this project, I'm trying to find a neighborhood in Hamburg for my friend Betty, who is going to give birth to twin babies at the end of 2021. Betty is currently living in the neighborhood of Eimsbüttel with her partner and a 4-year-old daughter. Because of the coming of the twins, she needs to find a bigger apartment for 2 adults with 3 children. She actually quite enjoys her current living area. Every morning, after she sends her daughter to the kindergarten, she grabs a coffee and then sits on the bench in the park along the Alster River near her living area. There is also a big playground nearby. In the afternoon, after she picks up the daughter, she goes to the

playground. Usually, there are some other couples with children in the playground too, so she can easily have chats with them. She also enjoys the convenience of shopping because the current living area has several supermarkets and grocery stores. When she is busy and don't have time to cook, she can easily get food from the restaurants downstairs of her apartment. However, there is also something she does not like about her current living area. For example, there are usually many people walking, running, cycling in the park nearby, which makes it a little bit too crowed in the park. It is also crowed in the restaurants too. She usually needs to wait quite long time to get the food she ordered. She would like to have an apartment in a quieter neighborhood. Besides, she doesn't like driving. So, it will be great if the new living area has convenient public transportation.

After talking with her, I summarized her key requirements for a neighborhood with the following characters:

- Less people
- Playground
- Park
- Supermarket
- Bus stops or U ban stations
- Restaurants
- Coffee shops

I further discussed with her about how important of the above characters and we reached to the agreements about the importance of each character:

- Playground, **important ratio=0.25**;
- Nature, including parks, rivers, etc., **important ratio: 0.25**;
- Supermarkets, stores to buy groceries, etc., **important ratio: 0.15**;
- Public transportation, including bus stops, U ban, S ban, train stations, etc., **important ratio: 0.15**;
- Food, including restaurants, fast food shops, etc., **important ratio: 0.1**;
- Coffee shops, **important ratio: 0.1**;

Other considerations:

- **Less population density** than the current neighborhood;

- Other couples with children, so she can have chats with the other parents;
- **Lower rental price** is better;

Overall, Betty would like to keep the things she likes about the current neighborhood (playground, parks/rivers nearby, restaurants and coffee shops, easy public transportation, parents with children that she can talk with) and avoid the things she dislikes (too many people). I will help her to identify the most promising neighborhoods based on her above wishes.

1.3 Target audience

Even though the analysis is customized to the needs of Betty (2 adult with 3 children), the approach is applicable to the other people who have the needs to find the most suitable neighborhood in Hamburg. Furthermore, combined with the buying price of the apartment/houses in each neighborhood, this approach can even be used to find the perfect neighborhood for the people who want to buy properties.

2. Data

2.1 Information needed

After identifying Betty's wishes for the new neighborhood, I decide to collect the following information:

- Information of the neighborhoods in Hamburg, including names etc.;
- Population density of each neighborhood;
- Percentage of families with children living in each neighborhood;
- Rental price of each neighborhood;
- Venues in each neighborhood;

2.2 Data source

To get the needed information, the following datasets are collected through internet:

- Information about each neighborhood, including names and coordinates, population density.
  https://de.zxc.wiki/wiki/Liste_der_Bezirke_und_Stadtteile_Hamburgs

- Percentage of families with children in each neighborhood **(31.12.2020)**. This information is downloaded from: https://www.govdata.de/web/guest/suchen/-/details/statistisches-jahrbuch-hamburg-2020-2021

- Rental price of each neighborhood **(06/2021)**. This information is downloaded from: https://www.wohnungsboerse.net/mietspiegel-Hamburg/3195

- Information about the venues in each neighborhood **(06/2021)**. This information is gathered through **Foursquare API**, using which to get the most common venue categories in each neighborhood.

3. Methodology

- **Collect and clean the data**. I will first start to collect all the needed information, clean the dataset, dealing with missing values, etc..

- **Explore the data**. After I get the cleaned data, I will do an exploratory data analysis. From describe analysis, I can get valuable insights for the distributions of the important features, like 'population density', 'rental price' and 'percentage of families with children' living in each neighborhood.

- **Set criterion to filter out the unwanted neighborhoods**. Betty really does not like the crowed neighborhoods and she would like to live in a neighborhood with many couples with children. Based on these two wished, I first filter out the neighborhoods with high density of population and less percentage of families with children.

- **Visualize the neighborhoods that she might be interested**. Folium map is used to visualize the distribution of the neighborhoods that she might be interested (green) and not interested (gray). This will give her a first impression about the locations that might fit her wishes.

- **Get the venue categories in each neighborhood**. Foursquare API is used to get the most common venue categories in each neighborhood. One hot encoding is used to put all the venue categories of each

neighborhood into one data frame. Then, I will summarize the venues into the categories Betty wishes to have (see the list in the introduction part).

- **Get interested index**. Based on her wishes, the interested index for each neighborhood is calculated based on the importance of the features. **Interested index = Playground * 0.25 + Nature * 0.25 + Stores * 0.15 + transportation * 0.15 + Food * 0.1 + Coffee * 0.1**

- **Cluster the neighborhoods**. k means is used to cluster the neighborhoods into 5 groups. The features I will use are 'Interested index', 'rental price', 'percentage of families with children' of each neighborhood. Folium map is then used to visualize the locations of the 6 clustered groups. This will allow Betty to have an overview about the 6 clusters.

- **Check each cluster and pick up the one fits Betty's wishes most**. Based on the principle of lower 'rental price', higher 'percentage of families with children' and higher 'interested index', I will pick up the cluster that fits the wishes of Betty most. Folium map is then used to visualize the neighborhoods that fits her wished most to give her an impression about the locations of these neighborhoods.

4. Analysis

4.1 Data preparation

There are several steps needed to be done to merge all the information into one data frame.

- The rental price of each neighborhood was got from wohnungsboerse.net in a format of pdf. I saved the file into an excel format and then manually organized the data in excel.

Stadtteil	€ / m²	Stadtteil	€ / m²
Eppendor	17,82 €	Hamm-N	12,80 €
Hoheluft-	17,51 €	Poppenbü	12,76 €
Barmbek-	17,36 €	Langenho	12,75 €
Winterhud	17,28 €	Sülldorf	12,66 €
Hohenfeld	17,28 €	Sasel	12,65 €
Ottensen	17,26 €	Schnelsen	12,55 €
Nienstedt	16,97 €	Rahlstedt	12,51 €
Altona-Al	16,96 €	Hamm-Sü	12,39 €
Altona-No	16,90 €	Eidelstedt	12,33 €

	Neighborhood	RentAppartmentPrice
0	Kirchwerder	10.52
1	Langenbek	10.54
2	Wilstorf	10.63
3	Steilshoop	11.10
4	Heimfeld	11.11

Data frame before (left) and after (after) data cleansing

- The percentage of the family with children for each neighborhood was got from govdata.de. This dataset was clean and I did not do further manipulations.

	Neighborhood	FamilyWithChild_Percent
0	Hamburg-Altstadt	0.121
1	HafenCity	0.263
2	Neustadt	0.114
3	St. Pauli	0.128
4	St. Georg	0.105

- The information of coordinates and population density of each neighborhood was got from wiki.

	Neighborhood	Borough	PopulationDensity	Coordinates
0	Hamburg-Altstadt	Hamburg-center	979	53 ° 33 ′ 0 ″ N, 10 ° 0 ′ 0 ″ E
1	HafenCity	Hamburg-center	2239	53 ° 32 ′ 28 "N, 10 ° 0 ′ 1" E
2	Neustadt	Hamburg-center	5549	53 ° 33 ′ 7 ″ N, 9 ° 59 ′ 8 ″ E
3	St. Pauli	Hamburg-center	8839	53 ° 33 '25 "N, 9 ° 57' 50" E
4	St. Georg	Hamburg-center	4733	53 ° 33 ′ 18 "N, 10 ° 0 ′ 44" E

The name of the neighborhood in this dataset is quite different from the above two datasets. For example, the name of the neighborhood 'Hamburg-Altstadt' is named as 'Hamburg old town' in this dataset. This will introduce errors when later I want to merge the dataset into one. For consistency, I replaced the name in this dataset to the name in the above two datasets.

```
## correct the name of the neighborhood
df_neighborhood_coordinates.at[0,'Neighborhood']='Hamburg-Altstadt'
df_neighborhood_coordinates.at[8,'Neighborhood']='Horn'
df_neighborhood_coordinates.at[4,'Neighborhood']='St. Georg'
df_neighborhood_coordinates.at[19,'Neighborhood']='Altona-Altstadt'
df_neighborhood_coordinates.at[21,'Neighborhood']='Altona-Nord'
df_neighborhood_coordinates.at[44,'Neighborhood']='Groß Borstel'
df_neighborhood_coordinates.at[51,'Neighborhood']='Barmbek-Nord'
df_neighborhood_coordinates.at[49,'Neighborhood']='Barmbek-Süd'
df_neighborhood_coordinates.at[99,'Neighborhood']='Hausbruch'
df_neighborhood_coordinates.at[54,'Neighborhood']='Langenhorn'
df_neighborhood_coordinates.at[73,'Neighborhood']='Lohbrügge'
df_neighborhood_coordinates.at[94,'Neighborhood']='Marmstorf'
df_neighborhood_coordinates.at[32,'Neighborhood']='Rissen'
df_neighborhood_coordinates.at[62,'Neighborhood']='Steilshoop'
df_neighborhood_coordinates.at[81,'Neighborhood']='Allermöhe'
```

Then I calculated the 'Latitude' and 'Longitude' data from the coordinates.

	Neighborhood	Borough	PopulationDensity	Latitude	Longitude
0	Hamburg-Altstadt	Hamburg-center	979	53.550000	10.000000
1	HafenCity	Hamburg-center	2239	53.541111	10.000278
2	Neustadt	Hamburg-center	5549	53.551944	9.985556
3	St. Pauli	Hamburg-center	8839	53.556944	9.963889
4	St. Georg	Hamburg-center	4733	53.555000	10.012222

- I merged the above three dataset into one data frame and correct the data type for the 'PopulationDensity'.

	Neighborhood	RentAppartmentPrice	FamilyWithChild_Percent	Borough	PopulationDensity	Latitude	Longitude
0	Kirchwerder	10.52	0.233	Bergedorf	313	53.419722	10.201667
1	Langenbek	10.54	0.197	Harburg	5048	53.437222	9.986111
2	Wilstorf	10.63	0.195	Harburg	5194	53.443611	9.984167
3	Steilshoop	11.10	0.223	Wandsbek	7976	53.610278	10.059167
4	Heimfeld	11.11	0.182	Harburg	1725	53.463889	9.956111

```
Before correction:                              After correction:
 Neighborhood               object               Neighborhood               object
RentAppartmentPrice        float64              RentAppartmentPrice        float64
FamilyWithChild_Percent    float64              FamilyWithChild_Percent    float64
Borough                     object              Borough                     object
PopulationDensity           object              PopulationDensity          float64
Latitude                   float64              Latitude                   float64
Longitude                  float64              Longitude                  float64
dtype: object                                   dtype: object
```

- The venues within each neighborhood were gathered using the Foursquare API. I chose a radius of 1000m and 100 as limit of number of venues returned by Foursquare API.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Langenbek	53.437222	9.986111	Lidl	53.442411	9.993219	Supermarket
1	Langenbek	53.437222	9.986111	Bäckerei Weiss	53.438949	9.987133	Bakery
2	Langenbek	53.437222	9.986111	H Freudenthalweg	53.439816	9.988239	Bus Stop
3	Langenbek	53.437222	9.986111	Sky	53.432509	9.986515	Shopping Mall
4	Langenbek	53.437222	9.986111	H Buchholzer Weg	53.431969	9.983450	Bus Stop

Then venues are grouped by neighborhood and I quickly check the top 6 popular venues for each neighborhood. The difference between two neighborhoods could be seen from the difference of the top 6 venues in each neighborhood. For example, Alsterdorf has 5 bus stops, 4

supermarkets and 3 restaurants, which means this neighborhood is easy public transportation accessible, easy shopping and ordering food (Betty would like this neighborhood). However, Betty would not like Hummelsbuttel, because none of the top 6 venues in this neighborhood is in her wish-list.

```
    ----Alsterdorf----              ----Hummelsbüttel----
              venue  freq                          venue  freq
0          Bus Stop   5.0       0                  Bakery   3.0
1       Supermarket   4.0       1  Furniture / Home Store   2.0
2             Hotel   3.0       2               Drugstore   1.0
3        Restaurant   3.0       3                     Spa   1.0
4         Drugstore   2.0       4             Burger Joint   1.0
5    Ice Cream Shop   2.0       5        Electronics Store   1.0
```

- I further calculate the 'Sum_interest' for each neighborhood based on Betty's wishes.

  Sum_interest = Playground * 0.25 + Nature * 0.25 + Stores * 0.15 + transportation * 0.15 + Food * 0.1 + Coffee * 0.1

	Neighborhood	Sum_interest
0	Allermöhe	0.70
1	Alsterdorf	6.90
2	Bergedorf	5.45
3	Billstedt	6.40
4	Blankenese	11.80
5	Duvenstedt	1.60
6	Eidelstedt	2.10
7	Eißendorf	3.60
8	Farmsen-Berne	4.85
9	Groß Borstel	0.70

- Finally, I merge all the dataset into one, which later will be used for clustering.

|	Neighborhood	RentAppartmentPrice	FamilyWithChild_Percent	Sum_interest
0	Langenbek	10.54	0.197	1.90
1	Wilstorf	10.63	0.195	3.85
2	Steilshoop	11.10	0.223	2.70
3	Heimfeld	11.11	0.182	3.05
4	Eißendorf	11.39	0.193	3.60

4.2 Data exploration

- Betty complained that her current living area was super crowed. So, I started with an exploration on the population density of each neighborhood

in Hamburg. From the plot, we can see that no wonder Betty complained about too many people in Eimsbüttel, because Eimsbüttel is almost the top 2 neighborhoods with the highest population density.
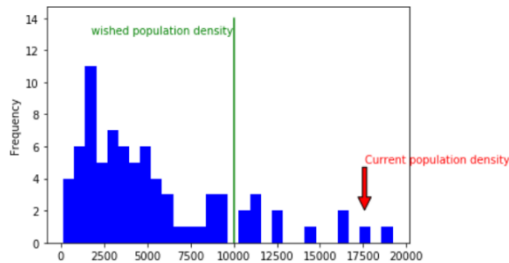
- Then I did the exploration on the rental price. Actually, Betty paid much less rental price compared to the mean value in Eimsbüttel. This is because she has lived in the current apartment for 12 years and due to the legal protection of renters, her rent has not raised much since she moved in 12 years ago. For the new apartment, she could accept a rental price lower than 20 per m^2. Of course, lower is better.
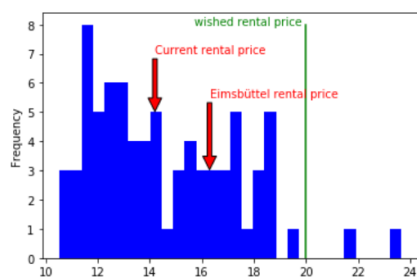

- Then I did the exploration on the percentage of families with children. Her current living area Eimsbüttel has quite low percentage of families with children. Betty would like to live in a new neighborhood with at least 17.5% families with children.
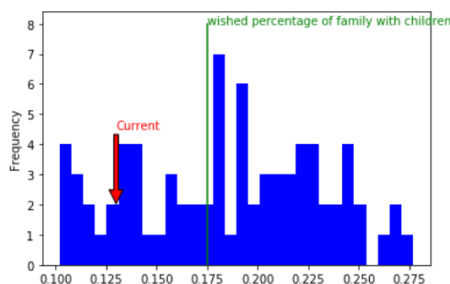
- Based on the above exploration, I filtered out some neighborhoods that Betty would definitely not like. These neighborhoods either have

population density higher than 10000 (Km^2), or rental price is higher than 20 per m^2. The number of neighborhoods that she might be interested is 46.

- I created a map of Hamburg with the neighborhoods she might be interested shown in green and not interested shown in gray. This will give her the first impression about the locations of the neighborhoods that might fit her wishes. It is quite obvious that her interested neighborhoods are not located in the center of the Hamburg city.
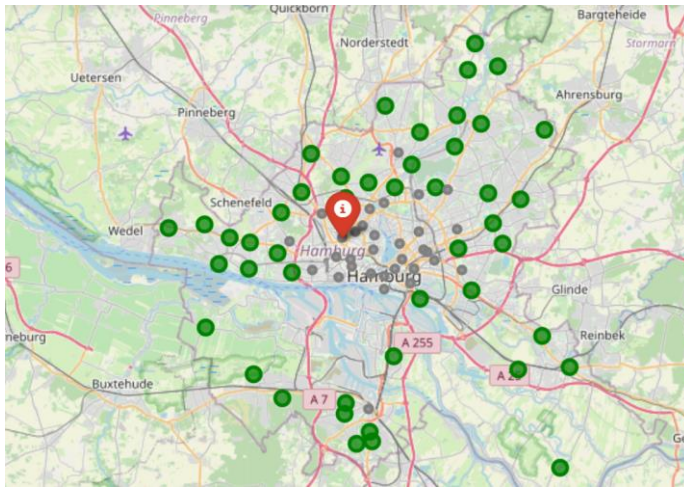

4.3 Cluster the neighborhoods into 6 groups

- K means was used to cluster the neighborhoods into 6 groups based on the features of rental price, percentage of families with children and sum_interest.

	Cluster Labels	Neighborhood	RentAppartmentPrice	FamilyWithChild_Percent	Sum_interest	Borough	PopulationDensity	Latitude	Longitude
0	1	Langenbek	10.54	0.197	1.90	Harburg	5048.0	53.437222	9.986111
1	1	Wilstorf	10.63	0.195	3.85	Harburg	5194.0	53.443611	9.984167
2	2	Steilshoop	11.10	0.223	2.70	Wandsbek	7976.0	53.610278	10.059167
3	1	Heimfeld	11.11	0.182	3.05	Harburg	1725.0	53.463889	9.956111
4	1	Eißendorf	11.39	0.193	3.60	Harburg	3012.0	53.455833	9.954444
5	1	Marmstorf	11.47	0.192	0.30	Harburg	1545.0	53.435833	9.968611
6	2	Neuenfelde	11.69	0.244	0.30	Harburg	316.0	53.514722	9.795556
7	1	Lohbrügge	11.70	0.185	1.45	Bergedorf	3078.0	53.509444	10.182222
8	1	Allermöhe	11.71	0.181	0.70	Bergedorf	160.0	53.486111	10.154167
9	2	Neugraben-Fischbek	11.83	0.251	0.15	Harburg	1404.0	53.483333	9.850000
10	0	Ohlsdorf	11.83	0.181	7.80	Hamburg North	2315.0	53.625833	10.031389
11	0	Bergedorf	11.95	0.197	5.45	Bergedorf	3475.0	53.488056	10.212500

- **I created a map to show the 6 clusters.**
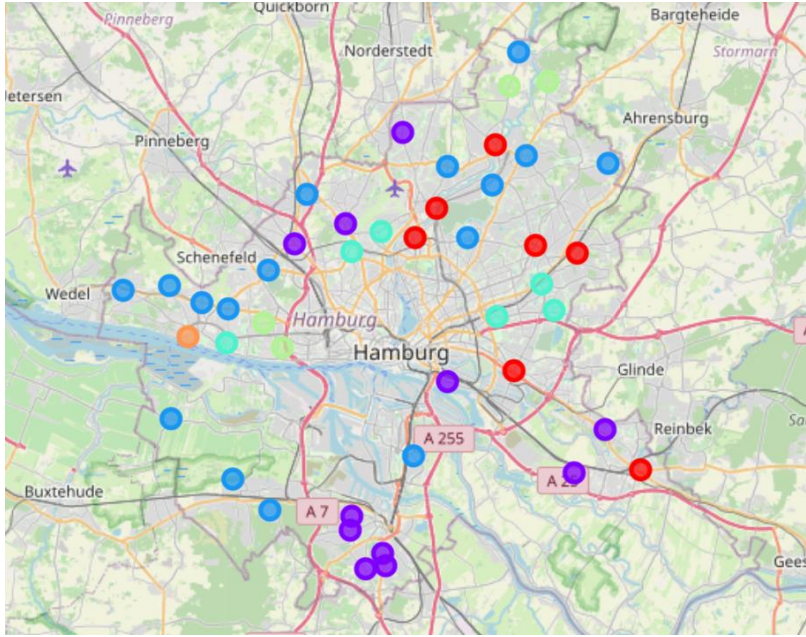
- **Cluster1**

	Neighborhood	RentAppartmentPrice	FamilyWithChild_Percent	Sum_interest	Borough	PopulationDensity	Latitude	Longitude
10	Ohlsdorf	11.83	0.181	7.80	Hamburg North	2315.0	53.625833	10.031389
11	Bergedorf	11.95	0.197	5.45	Bergedorf	3475.0	53.488056	10.212500
14	Billstedt	12.19	0.222	6.40	Hamburg-center	4142.0	53.540556	10.101111
17	Rahlstedt	12.51	0.206	5.80	Wandsbek	3462.0	53.601944	10.156667
22	Poppenbüttel	12.76	0.210	8.65	Wandsbek	2951.0	53.659167	10.084722
31	Farmsen-Berne	14.04	0.214	4.85	Wandsbek	4324.0	53.606389	10.119722
38	Alsterdorf	15.29	0.181	6.90	Hamburg North	4751.0	53.610833	10.013056

- **Cluster2**

	Neighborhood	RentAppartmentPrice	FamilyWithChild_Percent	Sum_interest	Borough	PopulationDensity	Latitude	Longitude
0	Langenbek	10.54	0.197	1.90	Harburg	5048.0	53.437222	9.986111
1	Wilstorf	10.63	0.195	3.85	Harburg	5194.0	53.443611	9.984167
3	Heimfeld	11.11	0.182	3.05	Harburg	1725.0	53.463889	9.956111
4	Eißendorf	11.39	0.193	3.60	Harburg	3012.0	53.455833	9.954444
5	Marmstorf	11.47	0.192	0.30	Harburg	1545.0	53.435833	9.968611
7	Lohbrügge	11.70	0.185	1.45	Bergedorf	3078.0	53.509444	10.182222
8	Allermöhe	11.71	0.181	0.70	Bergedorf	160.0	53.486111	10.154167
12	Rothenburgsort	12.05	0.176	2.50	Hamburg-center	1241.0	53.535000	10.040833
16	Eidelstedt	12.33	0.206	2.10	Eimsbüttel	3938.0	53.606944	9.906111
21	Langenhorn	12.75	0.204	3.40	Hamburg North	3444.0	53.665556	10.001389
23	Niendorf	12.87	0.181	4.20	Eimsbüttel	3293.0	53.617778	9.950278

- **Cluster 3**

	Neighborhood	RentAppartmentPrice	FamilyWithChild_Percent	Sum_interest	Borough	PopulationDensity	Latitude	Longitude
2	Steilshoop	11.10	0.223	2.70	Wandsbek	7976.0	53.610278	10.059167
6	Neuenfelde	11.69	0.244	0.30	Harburg	316.0	53.514722	9.795556
9	Neugraben-Fischbek	11.83	0.251	0.15	Harburg	1404.0	53.483333	9.850000
13	Hausbruch	12.10	0.242	2.60	Harburg	1738.0	53.466667	9.883333
15	Duvenstedt	12.27	0.267	1.60	Wandsbek	906.0	53.708056	10.104444
18	Schnelsen	12.55	0.229	4.20	Eimsbüttel	3289.0	53.633333	9.916667
19	Sasel	12.65	0.245	3.75	Wandsbek	2827.0	53.653889	10.111944
20	Sülldorf	12.66	0.220	1.60	Altona	1655.0	53.585000	9.794167
24	Rissen	12.93	0.210	1.85	Altona	951.0	53.583056	9.753611
25	Lurup	13.14	0.238	3.75	Altona	5811.0	53.593056	9.882778
26	Wilhelmsburg	13.19	0.227	2.80	Hamburg-center	1512.0	53.495000	10.011111
27	Wellingsbüttel	13.48	0.225	3.35	Wandsbek	2646.0	53.637778	10.081111
28	Iserbrook	13.52	0.222	0.60	Altona	4231.0	53.576389	9.823333
29	Volksdorf	13.66	0.238	2.30	Wandsbek	1808.0	53.649722	10.184167
32	Osdorf	14.04	0.229	1.70	Altona	3690.0	53.572778	9.846667
33	Hummelsbüttel	14.30	0.215	2.25	Wandsbek	2021.0	53.647778	10.041389

- ## Cluster 4

	Neighborhood	RentAppartmentPrice	FamilyWithChild_Percent	Sum_interest	Borough	PopulationDensity	Latitude	Longitude
37	Lokstedt	15.13	0.193	3.35	Eimsbüttel	5994.0	53.603056	9.956389
39	Tonndorf	15.56	0.191	2.60	Wandsbek	3884.0	53.586111	10.124722
40	Marienthal	15.83	0.180	2.40	Wandsbek	4225.0	53.568889	10.085833
42	Groß Borstel	16.42	0.194	0.70	Hamburg North	2321.0	53.613611	9.982500
43	Nienstedten	16.97	0.235	2.90	Altona	1670.0	53.555000	9.844722
44	Jenfeld	18.17	0.213	2.75	Wandsbek	5425.0	53.572222	10.136111

- ## Cluster 5

	Neighborhood	RentAppartmentPrice	FamilyWithChild_Percent	Sum_interest	Borough	PopulationDensity	Latitude	Longitude
30	Lemsahl-Mellingstedt	13.69	0.277	2.3	Wandsbek	857.0	53.690000	10.096389
34	Groß Flottbek	14.32	0.246	4.0	Altona	4630.0	53.565278	9.877500
36	Wohldorf-Ohlstedt	15.09	0.267	2.4	Wandsbek	269.0	53.692778	10.131111
41	Othmarschen	15.91	0.250	4.7	Altona	2623.0	53.552778	9.894444

- ## Cluster6

	Neighborhood	RentAppartmentPrice	FamilyWithChild_Percent	Sum_interest	Borough	PopulationDensity	Latitude	Longitude
35	Blankenese	14.85	0.21	11.8	Altona	1783.0	53.558333	9.811111

5. Results and Discussion

Cluster labels =0 and 5 are the groups of neighborhoods that I will suggest to Betty, because these neighborhoods have the **highest interest index**, the **median rent level and the proportion of families with children**. There are total 8 neighborhoods in these two groups.

	Neighborhood	RentAppartmentPrice	FamilyWithChild_Percent	Sum_interest	Borough	PopulationDensity	Latitude	Longitude
10	Ohlsdorf	11.83	0.181	7.80	Hamburg North	2315.0	53.625833	10.031389
11	Bergedorf	11.95	0.197	5.45	Bergedorf	3475.0	53.488056	10.212500
14	Billstedt	12.19	0.222	6.40	Hamburg-center	4142.0	53.540556	10.101111
17	Rahlstedt	12.51	0.206	5.80	Wandsbek	3462.0	53.601944	10.156667
22	Poppenbüttel	12.76	0.210	8.65	Wandsbek	2951.0	53.659167	10.084722
31	Farmsen-Berne	14.04	0.214	4.85	Wandsbek	4324.0	53.606389	10.119722
35	Blankenese	14.85	0.210	11.80	Altona	1783.0	53.558333	9.811111
38	Alsterdorf	15.29	0.181	6.90	Hamburg North	4751.0	53.610833	10.013056

Let's visualize these 8 neighborhoods that I will suggest to Betty (green) and the other neighborhoods (gray)
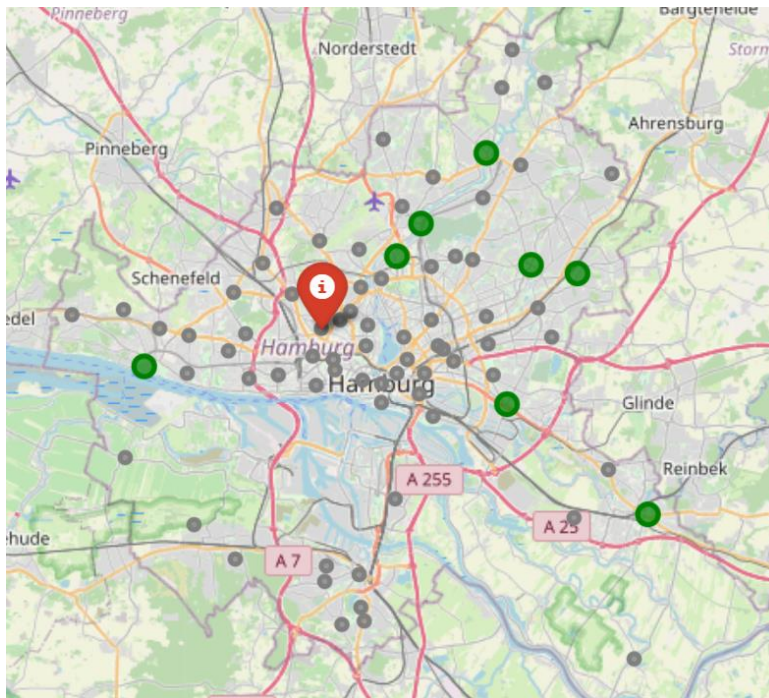
Let's go to the details about the 8 suggested neighborhoods for Betty.

● The mean value of the population density of the suggested neighborhood is much lower than the current neighborhood. It is also a little bit lower than the median level of all the neighborhoods.
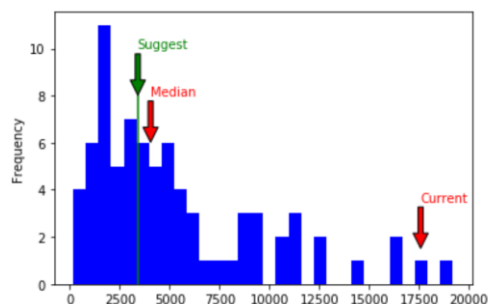

● The mean value of the rental price of the suggested neighborhood is much lower than the average value in Eimsbüttel. It is also a little bit lower than

her current rental price and the median rental price of all the neighborhoods.
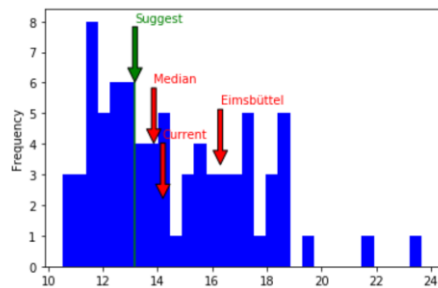
- The percentage of the families with children of the suggested neighborhood is much higher than the average value in Eimsbüttel. It is also a little bit higher than the median level of all the neighborhoods.
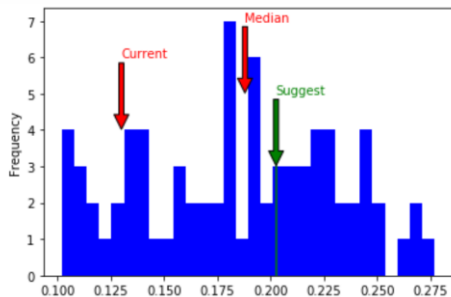
- Finally, let's check the venues in each suggested neighborhood. Betty would really like to have a playground nearby. However, just Ohlsdorf has playground. This is definitely not true. The reason of the low number of playground might be Foursquare API did not return this venue correctly. So to improve, the data of the number of playground in each neighborhood is needed.

----Ohlsdorf----

	venue	freq
0	Food	5.40
1	Supermarket	0.45
2	Transportation	0.45
3	Coffee	0.40
4	PlayGround	0.25
5	Nature	0.00

----Bergedorf----

	venue	freq
0	Food	4.40
1	Transportation	0.30
2	Coffee	0.20
3	Supermarket	0.15
4	PlayGround	0.00
5	Nature	0.00

----Billstedt----

	venue	freq
0	Food	6.1
1	Supermarket	0.3
2	PlayGround	0.0
3	Nature	0.0
4	Transportation	0.0
5	Coffee	0.0

----Rahlstedt----

	venue	freq
0	Food	4.10
1	Supermarket	0.45
2	Transportation	0.30
3	Coffee	0.10
4	PlayGround	0.00
5	Nature	0.00

----Poppenbüttel----

	venue	freq
0	Food	6.10
1	Supermarket	0.75
2	Nature	0.50
3	Transportation	0.30
4	Coffee	0.20
5	PlayGround	0.00

----Farmsen-Berne----

	venue	freq
0	Food	5.10
1	Supermarket	0.45
2	Transportation	0.15
3	PlayGround	0.00
4	Nature	0.00
5	Coffee	0.00

----Blankenese----

	venue	freq
0	Food	10.40
1	Nature	1.00
2	Coffee	0.70
3	Supermarket	0.30
4	Transportation	0.15
5	PlayGround	0.00

----Alsterdorf----

	venue	freq
0	Food	6.10
1	Transportation	0.75
2	Supermarket	0.60
3	Nature	0.25
4	Coffee	0.10
5	PlayGround	0.00

6. Conclusion and outlook

In this report, **I use K means to cluster the neighborhoods of Hamburg into 6 groups based on the features of each neighborhoods** (rental price, population density, percentage of the couples with children, venues categories). I identified 8 neighborhoods for Betty, who wants to move to a quiet neighborhood with playgrounds, parks, easy transportation, supermarkets, restaurants and coffee shops at an economical price. Even though the analysis is customized to the needs of Betty, **the approach is applicable to the other people who have other specific needs**. Furthermore, combined with the buying price of the apartment/houses in each neighborhood, this approach can even be used to find the perfect neighborhood if you want to buy properties.

7. References

- Foursqurare API
- https://de.zxc.wiki/wiki/Liste_der_Bezirke_und_Stadtteile_Hamburgs
- https://www.govdata.de/web/guest/suchen/-/details/statistisches-jahrbuch-hamburg-2020-2021
- https://www.wohnungsboerse.net/mietspiegel-Hamburg/3195