Decision support for starting a new business in Hamburg

Jannis Willwater

24.07.2020

1. Introduction

1.1. Background

Hamburg is with a population of 1.9 million citizens the second largest city in Germany. Especially in the last 10 years the city changes due to gentrification and an overall real estate boom. In this fast-changing city, many new businesses try to establish themselves. For that a precise analysis of the market situation is necessary in order to have the greatest possible chance of success.

1.2. Problem

The districts of Hamburg are very different with different main age groups. There are districts more inhabited by families, students or pensioners. Many shops or restaurants aim for customers in a special age group. Even a club with the best concept will not survive in an area with mainly pensioners. Additionally, the already existing venues should be taken into account. How is the competition? Where could be a gap in the market? Especially for foreign founders or companies that don't know the city, a close look on the age distribution and the competition is critical to success. As an example, we will analyze the data in order to find a good district for a new children's amusement park.

1.3. Objectives

This analysis will investigate the age distribution and businesses in the districts of Hamburg. To find the right location for a new business, i. e., a children amusement park. Therefore, we will classify the districts twice. Once based on the age distribution and on the existing venues. By this we will be able to make a precise recommendation where new businesses should be located.

2. Data sources

The following data where used to solve the problem:

1. The geographical information come from esri [1] in the form of a geojson-file
2. The information about the age distribution can be found on the state website of the "Statistikamt Nord" [2] as an xlsx-file.
3. To find the existing venues in Hamburg the Forsquare API were used [3].

3. Methodology

At the beginning the age distribution of the different districts is examined. For that the data from the "Statistikamt Nord" is used [2]. In the xlsx-file a lot of data from the year 2018 is shown. For our analysis only the columns with the information about the percentage of people under 18 and over 65 are from interested. This information is used to cluster the districts. For that a K-means algorithm is used. The data did not have to be normalized since all data were in percent.

The second step is the analysis of the venues in Hamburg. For that the Forsquare-API is used to find all venues in the districts of Hamburg. Before the districts are clustered based on this data, all venues that are age unspecific like supermarkets, hotels and some restaurants are deleted. The remaining data is sorted after the districts and normalized. Then the K-means algorithm is used again for clustering the data. It shows that there are two main clusters.

4. Results

4.1. Age-distribution

In Figure 1 the results of the clustering depending on the age distribution are pictured.

Cluster Labels	Under_18	Over_65
0	17.677989	21.488167
1	23.582843	12.340095
2	13.307644	13.024794

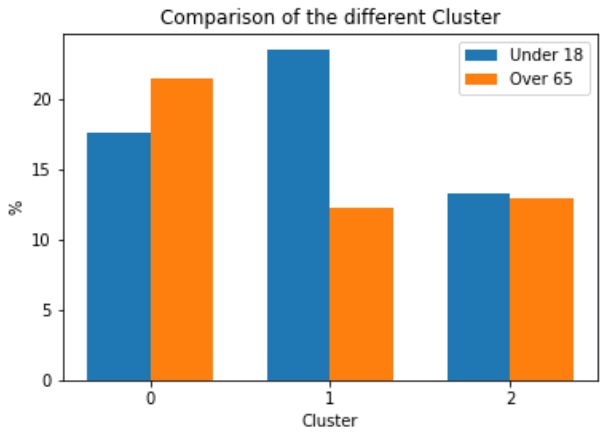

Figure 1: Clustering of the districts of Hamburg depending on the population under 18 and over 65.

As a result, we get the following three clusters:

- Cluster 0: High percentage of people over 65 (21.5 %) and a reasonable high percentage of people under 18 (17.7 %). We will call that cluster *Pensioners*.
- Cluster 1: Very high percentage of people under 18 (23.6 %) and a low percentage of people over 65 (12.3 %). That cluster will be called cluster *Family*.
- Cluster 2: This cluster shows a low percentage of people under 18 (13.3 %) and over 65 (13.0 %). That cluster will be called *Bachelors*.

For our special problem, the districts belonging to the cluster *Family* are from great interest.

4.2. Venue-analysis

The second step is the analysis of the venues in Hamburg by clustering the districts again. The figures 2 and 3 show the 10 most common venues in a few of the district of the same cluster. We can see

that in cluster 0 the most common venues are drugstores, bakeries, shopping malls and golf courses. So more for an older generation and is therefore called *pensioners-venues*. While cluster 1 has as most common venues more ice cream shops, parks and bakeries. Is therefore much more child friendly (*family-venues*).

	District	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	Allermöhe	Drugstore	Golf Course	Gym	Clothing Store	Park	Palace	Mediterranean Restaurant	Lake	Furniture / Home Store	Bakery
2	Altengamme	Shopping Mall	Bakery	Golf Course	Drugstore	Palace	Park	Clothing Store	Fast Food Restaurant	Beach	Steakhouse
8	Bergedorf	Golf Course	Clothing Store	Drugstore	Hardware Store	Bakery	Big Box Store	Train Station	Fast Food Restaurant	Gym	Furniture / Home Store
11	Billstedt	Bakery	Clothing Store	Electronics Store	Fast Food Restaurant	Park	Ice Cream Shop	Mediterranean Restaurant	Bar	Train Station	Big Box Store
12	Billwerder	Ice Cream Shop	Golf Course	Park	Drugstore	Fast Food Restaurant	Bakery	Gym	Mediterranean Restaurant	Nature Preserve	Hot Dog Joint
17	Curslack	Shopping Mall	Golf Course	Park	Event Space	Gym	Bakery	Clothing Store	Beach	Palace	River
29	Hamm	Drugstore	Fast Food Restaurant	Ice Cream Shop	Bakery	Big Box Store	Clothing Store	Steakhouse	Gym / Fitness Center	Gas Station	Hookah Bar
38	Jenfeld	Clothing Store	Electronics Store	Fast Food Restaurant	Bakery	Ice Cream Shop	Train Station	Drugstore	Park	Bar	Discount Store
39	Kirchwerder	Drugstore	Bakery	Train Station	Big Box Store	Golf Course	Fast Food Restaurant	Liquor Store	Clothing Store	Hardware Store	Burger Joint
40	Langenbek	Bakery	Drugstore	Intersection	Golf Course	Gas Station	Park	History Museum	Garden Center	Fast Food Restaurant	Steakhouse
43	Lohbrügge	Drugstore	Golf Course	Furniture / Home Store	Gym	Park	Clothing Store	Bakery	Hardware Store	Palace	Mobile Phone Shop
47	Marmstorf	Intersection	Bakery	Golf Course	Park	Gas Station	Drugstore	History Museum	Fast Food Restaurant	Garden Center	Train Station

Figure 2: Cluster 0 after the clustering from the Hamburger districts depending on the existing venues.

	District	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
1	Alsterdorf	Drugstore	Airport Lounge	Airport Service	Gym / Fitness Center	Park	Bistro	Pool	Beer Garden	Rental Car Location	Snack Place
3	Altenwerder	Park	Ice Cream Shop	Pub	Bakery	Cruise	Beach Bar	Beach	Harbor / Marina	Scenic Lookout	Trail
4	Altona north	Electronics Store	Park	Gym / Fitness Center	Sandwich Place	Grocery Store	Train	Fast Food Restaurant	Gym	Pub	Convenience Store
5	Bahrenfeld	Bakery	Ice Cream Shop	Park	Gym / Fitness Center	Beach Bar	BBQ Joint	Deli / Bodega	Coffee Shop	Pub	Candy Store
6	Barmbek-Nord	Vietnamese Restaurant	Park	Bakery	Ice Cream Shop	Gym / Fitness Center	Pool	Lake	Pizza Place	Nature Preserve	Museum
7	Barmbek-Süd	Drugstore	Ice Cream Shop	Burger Joint	Pizza Place	Theater	Bowling Alley	Bar	Trattoria/Osteria	Thai Restaurant	Clothing Store
9	Bergstedt	Ice Cream Shop	Bakery	Farmers Market	Farm	Nature Preserve	Clothing Store	Coffee Shop	Fast Food Restaurant	Pizza Place	Steakhouse
10	Billbrook	Park	Bakery	Ice Cream Shop	Drugstore	Fast Food Restaurant	Nature Preserve	Mediterranean Restaurant	IT Services	Volleyball Court	Rental Car Location
13	Blankenese	Ice Cream Shop	Shopping Mall	Beach	Clothing Store	Bakery	Drugstore	Taverna	Tea Room	Gym / Fitness Center	Garden
14	Borgfelde	Ice Cream Shop	Sandwich Place	Vegetarian / Vegan Restaurant	Vietnamese Restaurant	Park	Chinese Restaurant	Coffee Shop	Bakery	Indian Restaurant	Wine Shop
15	Bramfeld	Gym / Fitness Center	Bakery	Drugstore	Ice Cream Shop	Vietnamese Restaurant	Thai Restaurant	Plaza	Athletics & Sports	Furniture / Home Store	Taverna
16	Cranz	Pizza Place	Coffee Shop	Bakery	Beach	Outdoor Sculpture	Grocery Store	Modern European Restaurant	National Park	Gym Pool	Gym

Figure 3: Cluster 1 after the clustering from the Hamburger districts depending on the existing venues.

4.3. Comparison of both cluster methods

With both clusters of the district of Hamburg a investor could make a pronounced decision where to open the new business. For our example with the children's amusement park it should be located in a district with a lot of children. So, from the district *family* from chapter 3.1. Regarding the second clustering, there are good arguments to locate the park in a district of both clusters. Either in the cluster *pensioners-venues*, because there is a low competition or in the district in the *family-venues* cluster, because there are already interesting for families and could therefore attract more tourist. We will concentrate here on the first case. So, we need all districts that are in the cluster *family* and *pensioners-venues.* There are three districts that fulfill these criteria: Billwerder, Curslack and Sinstorf

The three districts are shown in the following map (created with the folium package). In the map the blue circle is Billwerder, the green is Curslack and the red is Sinstorf. We can see that all of these districts are located in the north of Hamburg. Billwerder is closer to the city center and Curslack and Sinstorf are more rural.

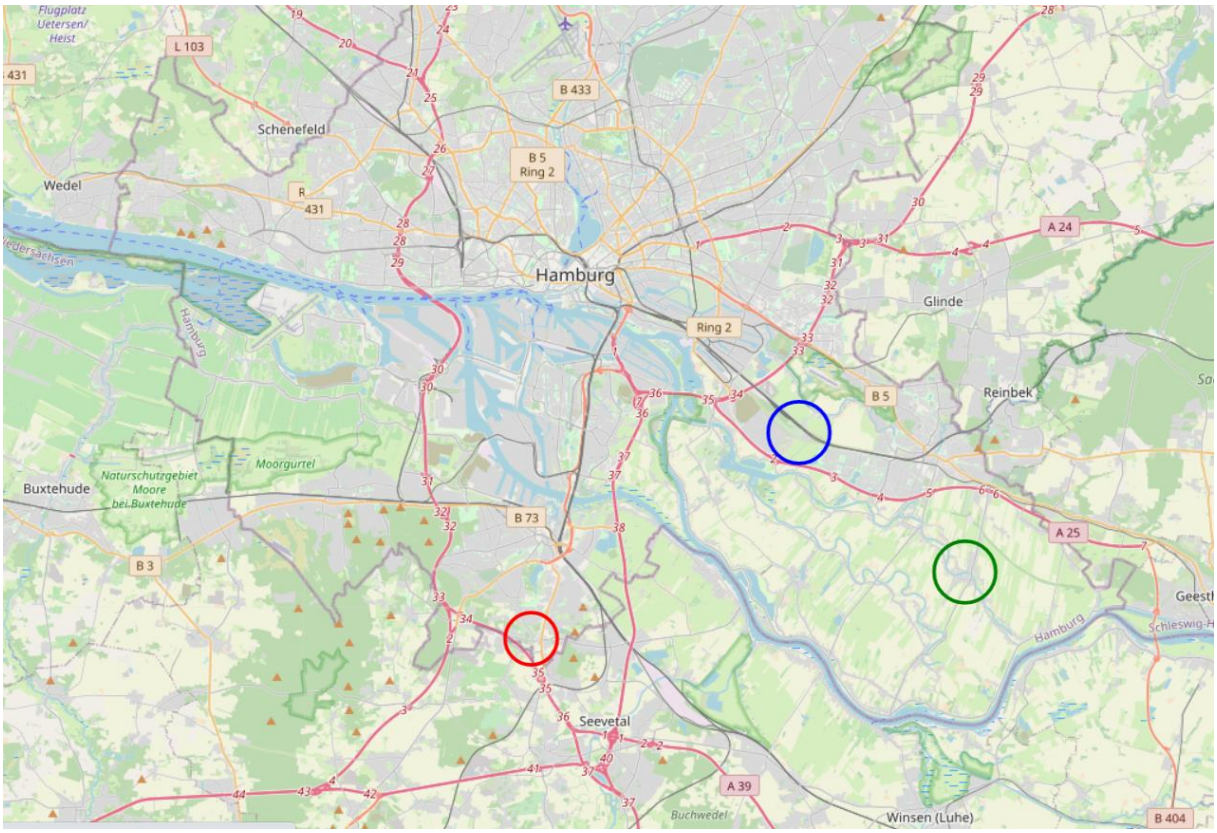

Figure 4: Map of Hamburg with circles marking the following districts: Billwerder (blue), Curslack (green) and Sinstorf (red)

Additionally, in figure 5 we show the venue categories with more than one entry in these districts. It shows, that Billwerder has already three parks, ice cream shops and fast food restaurants, while Curslack and Sinstorf have much less child friendly venues.

Now we have a variety of information to help with the decision where to locate a children's amusement park. We will discuss this question in the next chapter.

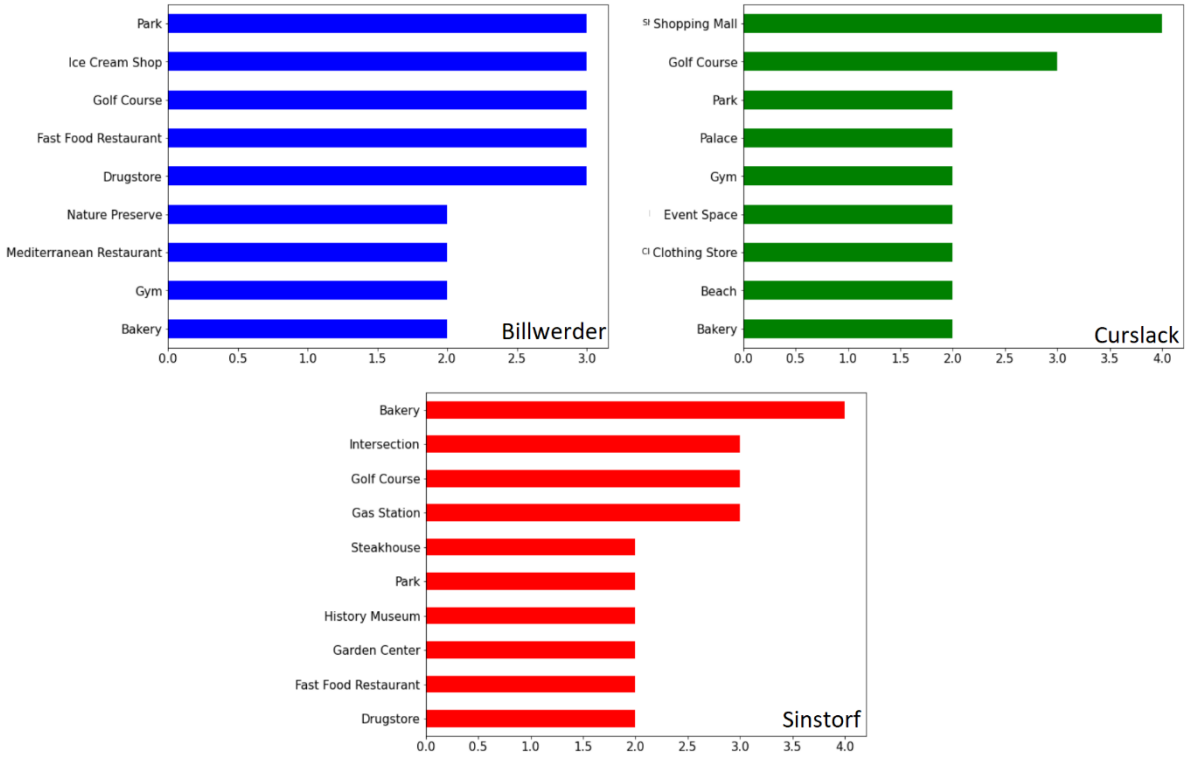

Figure 5: Venue categories with more than one entry in the districts Billwerder, Curslack and Sinstorf.

5. Discussion

With our data analysis we were able to limit the possible distractions for a children's amusement park to three. For that we only used the age distribution and already existing venues. The main selection criteria were a high proportion of under 18-year old's and districts, where there is not so much amusement for children so far. The final decision should be made on additional information. For example, is the district of Billwerder better, because it is closer to the city and easier to reach. But therefore, the rent will be much higher.

6. Conclusion

In this project I developed a strong tool that can be used for all new businesses in Hamburg. It is based on the age distribution and existing venues and it can be easily adjusted for each project and has therefore diverse application possibilities.

As an example, I used a children's amusement park and were able to limit it to only three possible districts. Additional information about the chosen districts like the location can be collected and used for the decision.

7. References

[1] esri Deutschland; https://opendata-esri-de.opendata.arcgis.com/datasets/8437e52c5e2d4963b6098accf571a891_0

[2] Statistikamt Nord; https://www.statistik-nord.de/zahlen-fakten/regionalstatistik-datenbanken-und-karten/hamburger-stadtteil-profile-und-interaktive-karten

[3] Forsquare API; https://developer.foursquare.com/