Capstone Project "Battle of Neighborhoods" (Week 1)

Karima Chakroun

March 9, 2019

1. Introduction

1.1. Background

Hamburg is the second-largest city in Germany with more than 1.8 million inhabitants and one of the major science, research, and education metropoles in Europe [1,2]. The city has several universities, both public and private, with more than 100 000 students in total [2,3]. Notably, the number of Hamburg's students rises every year [3-6] and will continue to grow, since projects like the "Hamburg excellence strategy" [8,9] and the "Science City Bahrenfeld" [10-12] will draw even more young people from all over the world to Hamburg [1]. However, the city's university buildings and campuses offer only very limited space and are often overcrowded, old, and decrepit [7]. Hence, they are not a very attractive place for students to pass the time, learn, or socialize inbetween and after courses. With more students to come in the following years, there is clearly a need for more student venues like cafés and bars nearby these universities. For example, a new venue type called "study café" might be of particular interest for most students, which could offer free WLAN/electricity, low prices for drinks & foods, long opening hours, and enough space/tables for students to work with books and laptops alone or in groups. However, such off-campus study cafés are yet entirely lacking in Hamburg. Private or public investors should size this opportunity and initiate the opening of such new types of student venues.

1.2. Business problem

Before opening a new study café in Hamburg, investors are faced with the question where to best open such a place. To approach this problem, several aspects should be considered:

a) **Where are the universities located?** (since the new venue should be located nearby one or more universities)

b) **What are the characteristics of different districts?** (since the new venue should be located in a lively district with other food & drink venues, but not with too many cafés already)

c) **How high are the rental prices for different districts?** (since the new venue should be located in a district with low to moderate rental prices in order to be able to offer low-priced food & drinks to students)

The present project addresses this problem by using different data about Hamburg (e.g. rental prices, locations data of universities, venues, and districts) to aid investors in finding the optimal location for opening a new study café.

1.3. Interested parties

This report will be of particular interest to any investors, private or public, who might want to open a new student venue in Hamburg and are yet undecided about the best location for such a place. Such investors could even include the universities themselves, which might decide to rent nearby off-campus places to offer more space for their students. Especially non-local investors who are not familiar with the city of Hamburg might profit from this report, as it will give them a nice visual overview of Hamburg's districts and their relevant characteristics.

2. Data description

To solve the problem, datasets from different sources were combined to answer the three above questions.

2.1. Where are the universities located?

To obtain data on this question, a list of Hamburg's main universities was scraped from the website https://www.4icu.org/de/hamburg/, containing 18 name entries. Based on these names, the geopy client (https://pypi.org/project/geopy/) was then used with the Nominatim geolocator service to request the geographical coordinates (latitude and longitude) of each university. Since the request returned no result for 6 of the 18 universities, the latitude and longitude of these 6 universities were searched and added manually using google maps (https://maps.google.de/). All data were combined into a dataframe including the name, latitude, and longitude for each of the 18 universities (shown here are the first 3 rows):

	University	Lat	Lon
0	Universität Hamburg	53.480616	10.240777
1	Technische Universität Hamburg	53.461007	9.969227
2	Hochschule für Angewandte Wissenschaften Hamburg	53.493382	10.200562

2.2. What are the characteristics of different districts?

To obtain data on this question, a geojson file of Hamburg's districts was first downloaded from https://gist.github.com/webtobesocial/935759ba975ffd9f6df6d1059fe5ad82/raw, containing the names and geographical borders of 104 districts. On one hand, this geojson file was used to create a choropleth map of rental prices in Hamburg's districts (see 2.3.). On the other hand, the data in this file were also used to obtain a list of Hamburg's district names. Based on these names, the geopy client was used with the Nominatim service to request the central geographical coordinates (latitude and longitude) of each district. These data were combined into a dataframe containing the name, latitude, and longitude of Hamburg's 104 districts (shown here are the first 3 rows):

	District	Lat	Lon
0	Allermöhe	53.483600	10.125000
1	Alsterdorf	53.610541	10.003889
2	Altengamme	53.429725	10.272787

Next, a list of venues for each district was obtained via the Foursquare API, using the "explore" endpoint with a limit of 100 and a radius of 500 meter around a district's given latitude and longitude. The returned information were combined with the district data into a dataframe showing for each venue the name, latitude, longitude, and venue category (shown here are the first 3 rows):

	District	District Lat	District Lon	Venue	Venue Lat	Venue Lon	Venue Category
0	Alsterdorf	53.610541	10.003889	Eppendorfer Moor	53.613315	10.002277	Nature Preserve
1	Alsterdorf	53.610541	10.003889	REWE	53.607687	10.005800	Supermarket
2	Alsterdorf	53.610541	10.003889	Best Western Premier Alsterkrug Hotel	53.613080	9.999037	Hotel

These venue data were then used in a k-means clustering analysis to cluster Hamburg's districts based on their venue characteristics (see methodology section).

2.3. How high are the rental prices for different districts?

To obtain data on this question, a list of rental prices for Hamurg's districts was scraped from the website https://www.4icu.org/de/hamburg/, containing the names and the average rental prices (€/m²) of 86 districts for the year 2019. Rental prices were in a string format (e.g. "11,21 €/m²") and were converted to floats (e.g. 11.21) for further analysis. Shown here are the first 3 rows:

	District	Price
0	Allermöhe	11.21
1	Alsterdorf	14.04
2	Altona-Altstadt	15.44

To integrate these data with the district's location and venue data from above, it was checked if the spelling of all 86 districts included in this dataframe was identical to the districts' spelling in the above dataframe, i.e. the one based on the geojson file. Three district names needed to be changed to match the spelling in the above dataframe ("Hamb.-Altstadt" to "Hamburg-Altstadt", "St. Georg" to "St.Georg", and "St. Pauli" to "St.Pauli"). Furthermore, the three subdistricts "Hamm-Nord", "Hamm-Süd", and "Hamm-Mitte" needed to be combined to the district "Hamm" and the rental price for this district was calculated as the average rental price of the three combined subdistricts. After these changes, both the districts' location and rental price data were combined into a dataframe containing the name, latitude, longitude, and rental price (if available) of the 104 districts (shown here are the first 3 rows):

	District	Lat	Lon	Price
0	Allermöhe	53.483600	10.125000	11.21
1	Alsterdorf	53.610541	10.003889	14.04
2	Altengamme	53.429725	10.272787	NaN

These data were then used in combination with the geojson file from above to create a choropleth map of rental prices in Hamburg's districts (see methodology section).