Capstone Project

–

Tourist Map for Hamburg

By Max Emanuel Feucht

**Introduction / Business Problem**

When visiting a new city, most people use some kind of Tourist Guide; Be it a Guidebook, websites like Tripadvisor, or a person that is already familiar with the city. Means like these help with orienting in a city based on experience from others, which poses a good starting point for discovering a city on one’s own. However, vibrant metropols like Amsterdam, Paris or Berlin are characterized by constant change in many domains of public life: Bars, Clubs, Restaurants, Cafés and many more localities are opening and closing with a rapid frequency. Along with this (and possibly caused by this), city districts continously change their appearance, their character, and the target audience they attract. To keep track of a city’s transformation, an up-to-date overview over localities in a citys districts would be helpful, allowing tourists to specifically target those areas that fulfill their wishes regarding gastronomic establishments. Websites like Tripadvisor offer an overview over different locations of establishments in the form of city maps, however requiring to look at every single establishment to check, whether the respective establishment is a café, a bar, a restaurant, etc. The sheer amount of information on sites like Tripadvisor might further be overwhelming, leaving tourists wishing for a quick overview over district characteristics reagarding the number of establishments from various categories located there. In order to easen the process of getting such an overview, a map covering all that information would come in handy for many tourists. Along with the information about the number of establishments from various categories, the density of inhabitants might be of interest in such a context as well; Some “business districts” might have many restaurants and takeaways, which might be located there mainly to provide nearby businesses with lunch opportunities. Such districts would probably not be of interest for most tourists. Contrary to that, “scene districts” showing the same amount of establishments but a much higher population denisty are likely to satisfy many tourists in their search for amusement and consumption. The population density therefore seems to serve as a appropriate variable to enrich the information of establishment numbers to ultimately provide an accurate and precise guidance for tourists. To allow getting a real quick impression over the characteristic of a district, the single districts of a city are further clustered and assigned to categories, depending on the number of establishments of various categories and the density of inhabitants.

Creating such a helpful, easy-to-use and up-to-date map for tourists could be of great value for stakeholders like tourist websites such as Tripadvisor and Yelp, or online Guidebooks such as “marcopolo”, targeting especially gastronomy-oriented and tech-affine tourists. The creation of the proposed map is the goal of the Capstone project at hand. This approach is tested for the city of Hamburg, as it is the second largest city in Germany, shows a wide variety in districts and happens to be my beautiful hometown.

**Data**

The project at hand specifically focuses on the characteristics of city districts. I therefore looked for socioeconomic data for city districts as well as the corresponding Geodata. The socioeconomic profile of city districts was obtained from:

[https://www.statistik-nord.de/zahlen-fakten/regionalstatistik-datenbanken-und-karten/hamburger-stadtteil-profile-staedtestatistik-fuer-hamburg](https://www.statistik-nord.de/zahlen-fakten/regionalstatistik-datenbanken-und-karten/hamburger-stadtteil-profile-staedtestatistik-fuer-hamburg%22)

From this source, I used the district names, their area size, the inhabitant number and economic data such as price per sqaure meter for a flat. Those economic variables were later dropped from the analysis beacaus of the high missing rates.

The Geodata could be obtained by ESRI Germany’s API:

<https://opendata-esri-de.opendata.arcgis.com/datasets/8437e52c5e2d4963b6098accf571a891_0/geoservice?geometry=7.392%2C53.395%2C11.347%2C53.965>

This API delivers a simple GeoJSON file for all city districts of Hamburg containing information about the districts name as well as the polygon shape of the respective districts in World Geodetic System 1984 (WGS84) coordinates.

Further, the establishment data was retrieved from the Foursquare API. With the data within the GeoJSON file mentioned before, I computed the center coordinates for each district, to obtain the coordinates to pass into the Foursquare API for establishment retrieval. I used a search request and passed those coordinates as well as a limit of 1000 establishments per neighborhood in. The radius to search in was determined by the area size of each Neighborhood. For further details, see the Methodology section. The variables used from the Foursquare API were the venue names, the venue category, as well as the latitude and longitude of each venue. All of this information was stored in a dataframe along with the name and coordinates of the neighborhood whose coordinated were used to make the search request. Specifically, the dataframe from the API retrieval looked like this:

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

The radius was computed by dividing the area size of each district by Pi and then taking the sqaureroot of the result. This yielded an approximate radius which covered the area that was given in the fist source. As not every district is nearly circle-shaped, this approach is only an approximation; especially very long or very thin districts might therefore be underrepresented in the given analysis. Given that the resulting circles might overlap and some of the uncovered areas might by covered by nearby retrieval radii, this is mostly a problem concerning the border districts. To account for this, I calculated the distance of each Neighborhood center to the city center coordinates. When the distance was below 5 km and districts I added only 10% to the retrieval radius. If the distance was over 5km an, I added 40% to the radius, to account for the wider and less dense city districts. Below you can see a depicition of areas covered with this approach; on the left side, you see the coverage with the unaltered radii, on the right you see the adapted radii.

As a result, some of the single radii overlapped, which is why some restaurants were retrieved for two districts. To determine to which district the respective establishment belongs, the distance to the district centers by which the establishment was retrieved was calculated. When an establishment was listed twice or more, only the entry with the smallest distance was kept, and the establishment was assigned to the neighborhood that was closest by.

