A new bakery in Vienna

Report on the final project of the IBM data science program

Julian Fesel

February 17, 2019

ABSTRACT: This report describes the analysis that was made to determine which locations in Vienna present themselves for opening a new bakery. Data on the location of existing bakeries and pastry selling supermarkets is collected was collected from Foursquare and is compared to the population in the same district. From this, favourable locations are selected by looking for low bakery densities.

Contents

1 Introduction

Austria is home of a great bread and pastry tradition which is perpetuated by many small bakeries all over the country. Given the competition between them and newly arising competition from bakeries placed directly in larger supermarkets, this reports tries to find favourable locations for opening a new bakery in Vienna, which is the capital of Austria. Put in other words, this report tries to answer the question: *What are good places in Vienna for opening a new bakery?*

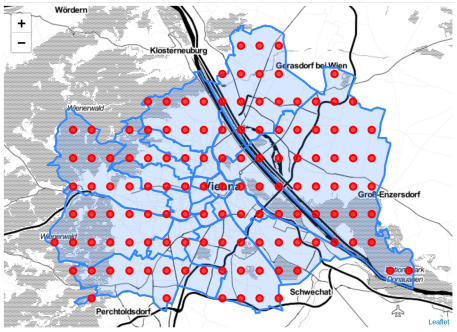

Figure 1: Coordinate grid which was used for querying the data from Foursquare.

2 Data sources

Different data sources were accumulated and used to answer this question. Data on the area and the population in each district of Vienna was gathered from a table on Wikipedia [4]. The latter was done by scraping the page using the `requests` and `BeautifulSoup` packages. Geographical data on the districts was taken from a geojson file which can be found in the references [1]. In addition, a geoJson of the outer boundaries of Vienna was used, which can also be found in the references [3]. The main data on the locations (latitude and longitude) and names of bakeries and supermarkets in Vienna was gathered from Foursquare by using their API [2].

3 Methodology

In order read in the data from Foursquare, a coordinate grid was layed over the city and the API used to make a query search at each point. This way the restriction on the API to 50 results per request was circumvented. The grid is visualised in figure 1. The data was wrangled and cleaned and then visualised using the `Folium` package. The bakeries and supermarkets in each district were counted and compared with the population data. The ratio of population to number of bakeries was determined and visualised in bar charts and on choropleth maps.

4 Results

The results are visualised in figures 2, 3 and 4. Figure 2 shows the locations of bakeries and supermarkets in Vienna and shows that they tend to concentrate around the inner city. Figure 3 visualises the number of people divided by the number of bakeries or the number of people divided by the number of bakeries and supermarket in each district. Again the ratio of number of people and number of pastry sellers is higher in the outer parts of the city. Figure 4 gives the same information as figure 3 in bar chart form.

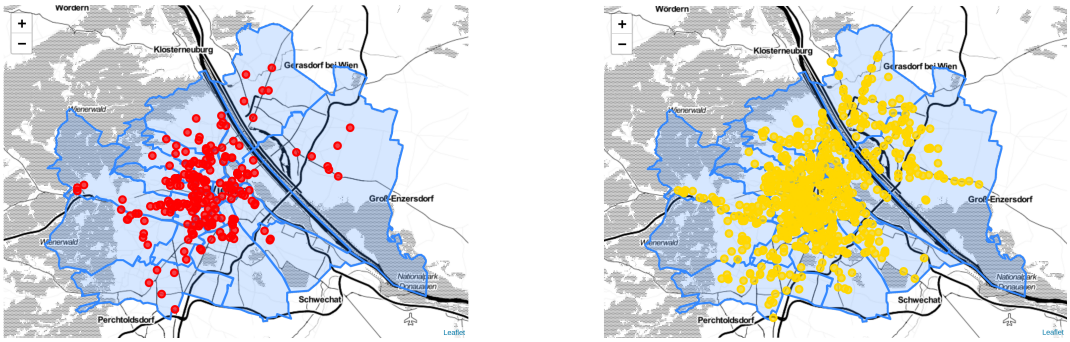

Figure 2: *Left:* Red markers representing the locations of bakeries in Vienna. *Right:* Yellow markers representing the locations of supermarkets in Vienna.

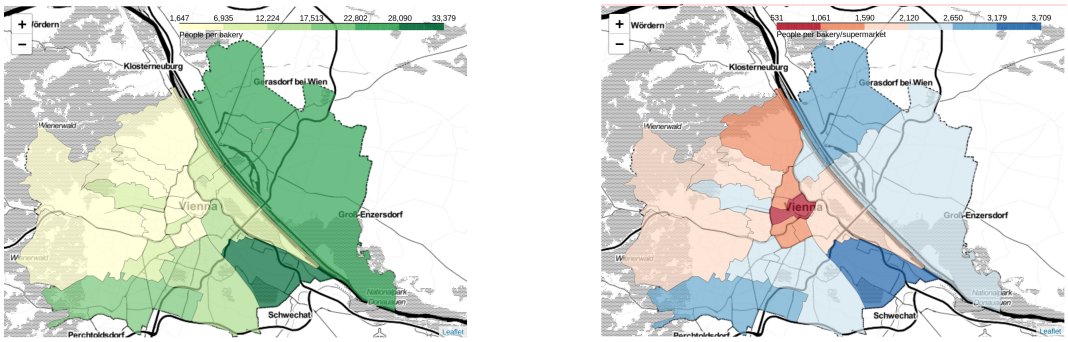

Figure 3: *Left:* Visualisation of the number of people per bakery in each district. *Right:* Visualisation of the number of people per bakery/supermarket in each district.

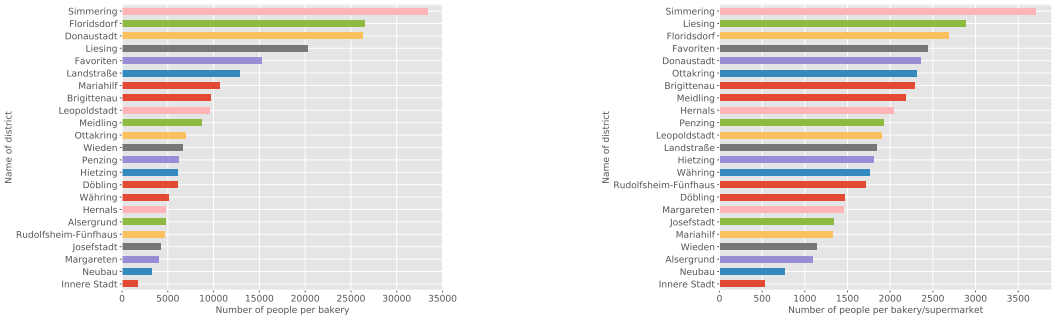

Figure 4: *Left:* Visualisation of the number of people per bakery in each district. *Right:* Visualisation of the number of people per bakery/supermarket in each district.

5 Discussion

As can be seen in figure 4 on the left, the districts with the highest number of people per existing bakeries, therefore making them more favourable as location for a new bakery, are Simmering, Floridsdorf and Donaustadt. This only changes slightly if the supermarkets are also taken into account (figure 4 on the right), with the assumption that they equally sell pastry. A weakness of this analysis is, that the population data is grained very coarsely and therefore does not allow to pinpoint the locations more closely. In addition, this analysis does not take into account other factors like business due to tourism in the inner city, movement of people during the day towards the center due to their jobs etc.

6 Conclusion

As a first recommendation, the outer parts of the city can be taken into consideration, while the analysis should probably be extended and further more finely grained population data collected. Further data on tourism, people movement and shopping habits should also prove useful.

References

[1] *Bezirksgrenzen Wien.* URL: `https://www.data.gv.at/katalog/dataset/2ee6b8bf-6292-413c-bb8b-bd22dbb2ad4b` (visited on 02/15/2019).

[2] *Foursquare location data.* URL: `https://developer.foursquare.com/docs/api/venues/search` (visited on 02/15/2019).

[3] *Stadtgrenzen Wien.* URL: `https://wahlen.strategieanalysen.at/geojson/` (visited on 02/15/2019).

[4] *Wiener Gemeindebezirke.* URL: `https://de.wikipedia.org/wiki/Wiener_Gemeindebezirke#Bezirke_und_Bezirksteile` (visited on 02/17/2019).