Capstone Project – The Battle of Neighborhoods

Finding a good location to open an Italian restaurant in Berlin, Germany

Final Assignment for IBM Data Science Coursera Specialization

Erin Bedri

18.05.2020

Introduction

Background

Berlin is the capital and largest city of Germany by both area (30,000 sq. km) and population (ca. 3.8 million).

Being world city of culture, politics, media and science, Berlin attracts many tourists all year round. In 2014, the city recorded 28.7 million overnight hotel stays and 11.9 million hotel guests. Moreover, tourism figures have more than doubled within the last ten years and Berlin has become the third most visited city destination in Europe. The largest visitor groups are from Germany, the United Kingdom, the Netherlands, Italy, Spain and the United States.

Additionally, in 2019, Berlin was ranked 13th in the Quality of living city ranking prepared by Mercer.

Problem

Being a well known cosmopolitan city with people from all around the world, Berlin offers a wide variety in the gastronomy sector, especially Italian restaurants, with new ones are opening all the time.

Searching for an optimal location to open an Italian restaurant in Berlin, however, could be challenging. Although it is logical to think that locations where there are no Italian restaurants, or at least not many of them, would be better, the problem is that interested customers tend to prefer popular and busy neighborhoods with many options available to choose from for them. From an investor point of view, on the one hand that means more movement of people and therefore more customers, but on the other, higher rents and stronger competition.

Interest

To analyze locations, ratings, and competition is a plus for those who want to be part of this industry. Therefore, the project would be perfect fit for investors who wants to open a new Italian restaurant in the city of Berlin.

Data Description

To help solving the problem of finding the optimal location for a new Italian restaurant, the following data will be searched for and considered during the project:

a) Number of already existing Italian restaurants in a given borough;

b) Different boroughs and their respective average rental prices to open a restaurant;

Additionally, the tools that will be used in this project are:

a) Foursquare API to select the number of Italian restaurants and their respective location in any given borough in Berlin;

b) Geocoder to get the latitudes and longitudes of places of interest.

Methodology

The dataset that will be used in this Capstone project was obtained through the Foursquare API, exploring only one type of venue – an Italian restaurant or the so-called Trattoria/Osteria. Further details such as venue ID, coordinates and postal code are also explored. Figure 1 shows the five first lines of the dataset created.

	ID	Name	Category	Latitude	Longitude	Postal Code
0	56bf2966cd1043d2f2179233	Mani in pasta	Trattoria/Osteria	52.502167	13.431717	10997
1	5919dd9c6e465058f788753e	Esse Osteria	Trattoria/Osteria	52.535791	13.422438	10405
2	4bc0c28ff8219c748f22b210	Osteria Ribaltone	Trattoria/Osteria	52.496061	13.342587	10777
3	52c4840d11d221d9ea8d6277	Trattoria Zoe	Trattoria/Osteria	52.541842	13.423225	10437
4	5057650de4b0d070fdec8fe0	Amici Amici	Trattoria/Osteria	52.492960	13.387429	10961

Figure 1: The main dataset created using the Foursquare API exploring all Italian restaurants in Berlin, Germany.

It was found that there are a total of 92 Italian restaurants in Berlin.

Working on this data set, a cleaner data set was created. It includes all the postal codes in Berlin which have an Italian restaurant located there. Figure 2 shows the top five first most populated postal codes with Italian restaurants and their respective total count.

	Postal Code	counts
0	14057	5
1	14059	5
2	10437	4
3	10623	3
4	10178	3

Figure 2: The dataset exploring all Italian restaurants in Berlin, Germany and their total count by postal code.

It was found that there are 54 neighborhoods with at least one Italian restaurant located in them. Two of those neighborhoods have 5 Italian restaurants and one has 4. The rest have 3 or less.

Results and Discussion

By using the matplotlib library, a bar chart representing the Top 5 Neighborhoods with the most Italian restaurants (by postal code) was created. The bar chart can be seen in Figure 3.
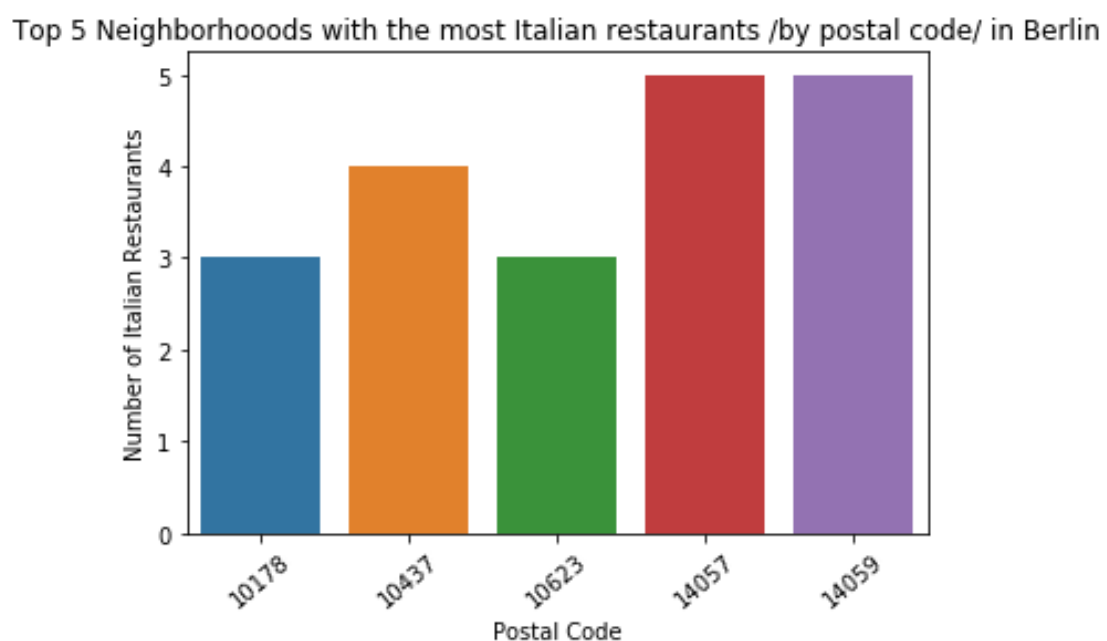

Figure 3: Top 5 neighborhoods with the most Italian restaurants (by postal code) in Berlin.

Furthermore, all 92 Italian restaurants were plotted on an interactive folium map.
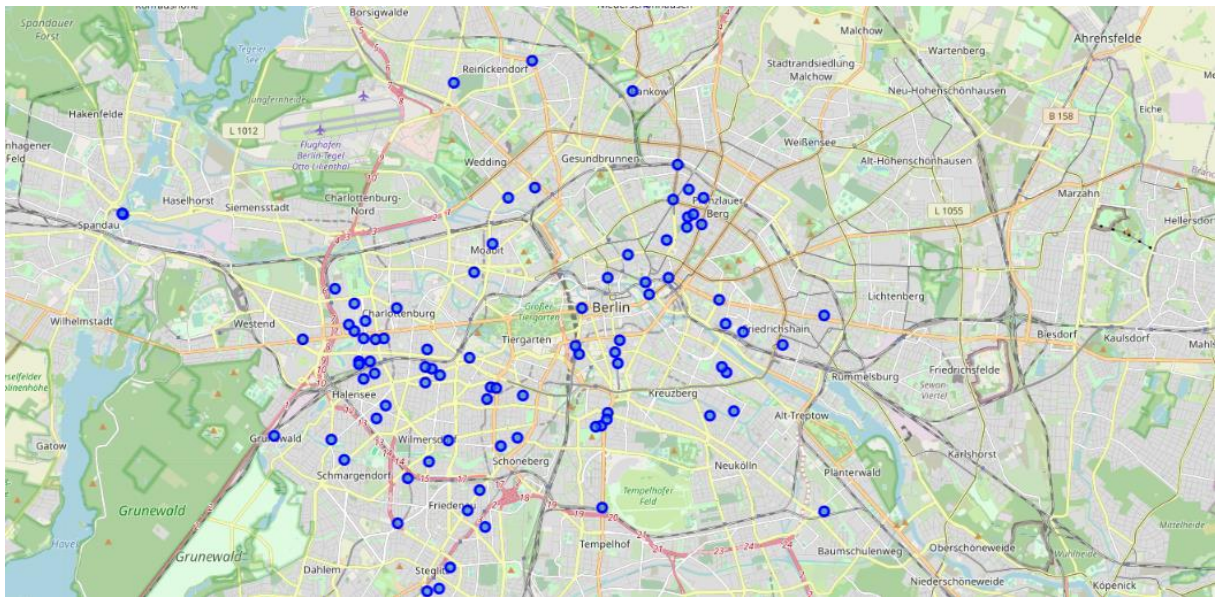

Figure 4: All 92 Italian restaurants in Berlin, Germany.

The next figure represents a choropleth map of Berlin where the city is divided into its neighborhoods. The dark grey regions represent neighborhoods without an Italian restaurant located in them. The rest are neighborhoods with Italian restaurants. The darker the purple color gets; the more restaurants are present in a given neighborhood.
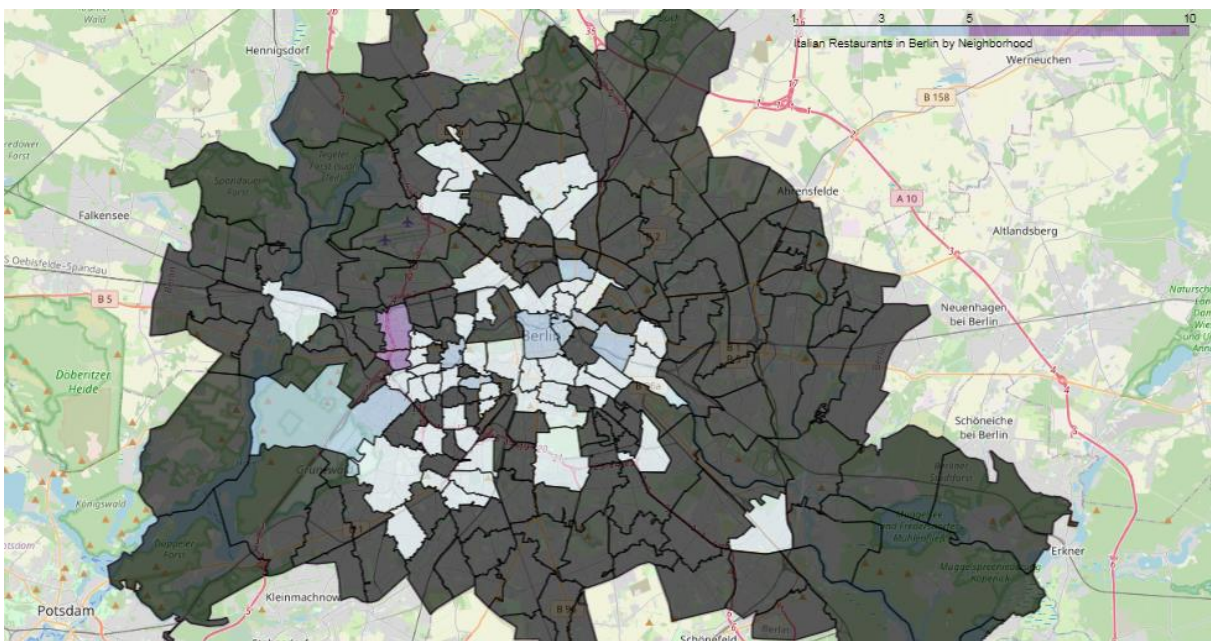

Figure 5: A choropleth map showing Italian restaurants by neighborhood in Berlin.

The next map is a zoomed in version of Figure 5. One can easily see the two most populous neighborhoods (with 5 Italian restaurants in each of them) colored in dark purple.
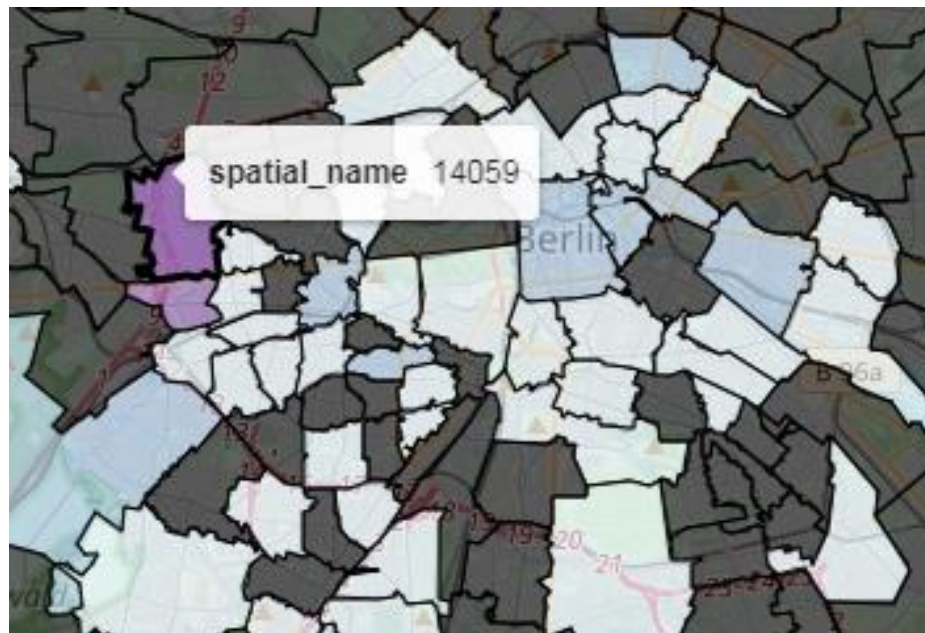

Figure 6: A zoomed in choropleth map showing the most populous neighborhood with Italian restaurants in Berlin.

Conclusion

In this data science project, I showed how to explore venues using Foursquare API and how to get latitudes and longitudes using Geocoder. I chose the Italian restaurants category to explore Foursquare venues in the city of Berlin, Germany.

It was found out that there are a total of 92 Italian restaurants in Berlin spread out to 54 neighborhoods. Two of those neighborhoods had 5 Italian restaurants, making them the two richest neighborhoods with such venues.