

The Battle of Neighborhoods Week 1

Dirk Belger

06.12.2020

Introduction

In this project I will explore the culinary possibilities in Cologne. It is a city which has a diverse offering of different Bars Restaurants and Cafes and it is nearly impossible to get to know all of them, not mentioning that there are closing and opening up new venues from time to time. As a person living in or visiting this city, there is so much to explore and a goal of this project is to help identifying those venues that are worth visiting. As a way of achieving this, I tried to model a basic scoring system, that takes into account the rating, the price and the distance of the venue depending on the current location, while I limit the radius to 750 meters. It then recommends you a list of venues with the best scoring, independent of the actual category of the venue. Therefore this method is not really suitable for all user, for instance those who look for a specific kind of venue or cuisine. However, for users that are new to the city it provides venues that are objectively worth visiting.

Data

I use public libraries and API's in this project. I use Foursquare API and some common Python Libraries like folium, pandas and matplotlib.

The available FourSquare location Data via the Foursquare API provide us with the necessary information to visualize the restaurants that are spread throughout the city. On a more detailed point of view, we can dive into specific neighbourhood or streets and find the best rated venues, the distance to the venue and the respective price category. All postal codes from the city of Cologne are available on "<https://www.koeln.de/postleitzahlen>" and the necessary location from the respective Cologne postal codes can be found on "<http://www.fa-technik.adfc.de/code/opengeodb/PLZ.tab>", which is a csv-file, similar to the Toronto example we had throughout the course. With those data sources we can collect the information that we need to model the scoring system and recommend the places that scored the best.

Methodology

Postal code data will be collected from <https://www.koeln.de/postleitzahlen>, cleaned and processed into a dataframe. The respective location data is read into a dataframe from the mentioned csv-file and merged with the postal code data including the neighbourhood names. With the available location data we will use our Foursquare Developer account to fetch the detailed venue information such as category, price, rating etc. Then we create a score with the available venue information to output those venues that are ranked highest with our applied scoring system. These are the venues that are recommended to visit.

The latitude- and longitude information of each postal code in Cologne is collected by the following statement

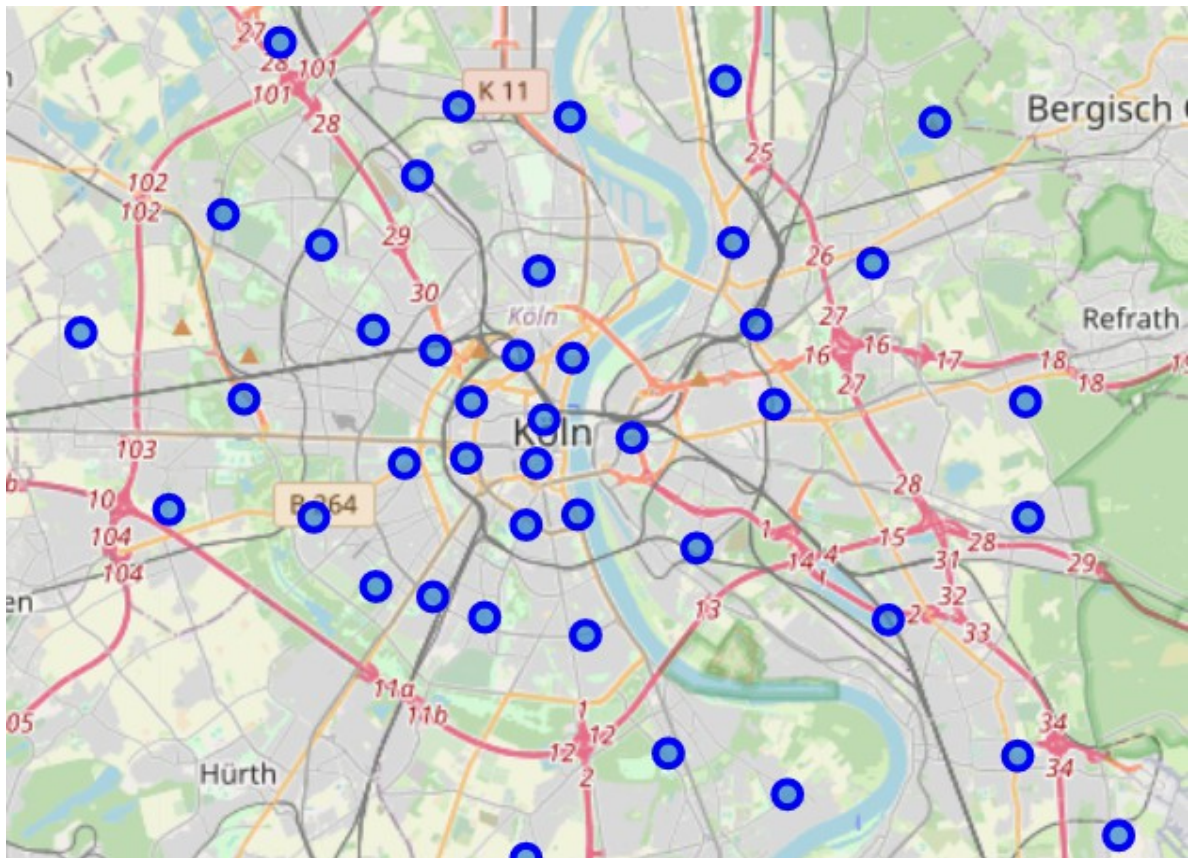
```
df_geo = pd.read_csv("http://www.fa-technik.adfc.de/code/opengeodb/PLZ.tab", sep='t', dtype=str)
```

and then merged with a dataframe that contains all postal codes of Cologne. These information are available on '<https://www.koeln.de/postleitzahlen>' and processed into the dataframe using the library BeautifulSoup4. See here the 10 first entry of the resulting dataframe.

Out[7]:

	plz	#loc_id	lon	lat	Ort
0	50667	8321	6.95768491832349	50.9400886235176	Köln
1	50668	8322	6.96545005139455	50.950695435874	Köln
2	50670	8323	6.95095743959049	50.9511722093173	Köln
3	50672	8309	6.9384761893034	50.9434409895397	Köln
4	50674	8324	6.93704081933882	50.933583841904	Köln
5	50676	8325	6.95555962327279	50.9329782103951	Köln
6	50677	8326	6.95323649210595	50.9225551674677	Köln
7	50678	8327	6.96648245395115	50.9240361563136	Köln
8	50679	8328	6.98103226157441	50.9374228290525	Köln
9	50733	8329	6.95639738632809	50.9651646910743	Köln
10	50735	8330	6.96489890149431	50.9913171833546	Köln

Using the folium library we can visualize the location points/neighbourhoods on a the map.



I utilized the Foursquare API to explore the boroughs and segment them. I limited the radius to 750 meter for each borough from their given latitude and longitude informations, which I think is still a reasonable amount of meter walk to a cool venue. I also filtered the results in a way, that only „Food Places“ are displayed.

Here is a head of the list Venues name, category, latitude and longitude informations from Forsquare API as well as the distance to the venue and the adress.

Out[21]:

	ID	Name	Category	Distance	PostalCode	Address
5	5c084427061b51002c127368	Tapeo & Co.	Tapas Restaurant	81	50674	Lindenstr. 38
4	57645427498e7a5746a35766	Tanica	Italian Restaurant	142	50674	Engelbertstraße 31
79	5a3a15691f8ed64718884d30	Tigermilch	Peruvian Restaurant	176	50674	Brüsseler Str. 12
0	583ec82c19b1ad33a488ae04	mikoto	Sushi Restaurant	196	59674	Hohenstaufenring 55
2	4b05886bf964a52082c422e3	Café Wahlen	Café	206	50674	Hohenstaufenring 64
...
80	4ee268668231cd14a32afac3	Shaka Zulu	African Restaurant	716	50672	Limburger Str. 29
84	5159889ee4b0655021dd3bd5	Hot Point	Chinese Restaurant	719	50667	Salierring 44
82	4fb50c37e4b0186ba39cc480	Caveedel	Café	724	50672	Brüsseler Str. 69
94	4d91eae49acaa143a55af2f0	Lakshmi	Indian Restaurant	738	50676	Thieboldsgasse 101-103
73	541c3701498ed1e0072d104f	Häppchen	Café	750	50674	Otto-Fischer-Str. 9

Results

This table is then complemented with the rating and the price range and by the following formula the score is calculated:

$$\text{Score} = (1 - (\text{Distance}/\text{max_distance})) * 0,15 + (1 - (\text{Price}/\text{max_price})) * 0,15 + \text{Rating}/\text{max_Rating} * 0,7$$

With weighing the scoring I tried to put the importance of the rating in the middle and the distance and the price range a less important factor. For me personally, this is a very good way to account for these characteristics, it is highly subjective though.

This leads to the following table:

Out[78]:

	ID	Name_x	Category	Distance	PostalCode	Address	Name_y	Price	Rating	Score
3	583ec82c19b1ad33a488ae04	mikoto	Sushi Restaurant	196	59674	Hohenstaufenring 55	mikoto	1	8.6	0.825833
0	5c084427061b51002c127368	Tapeo & Co.	Tapas Restaurant	81	50674	Lindenstr. 38	Tapeo & Co.	1	8.0	0.819097
26	4ba20b35f964a52001d837e3	Meister Gerhard no 008	Tapas Restaurant	357	50674	Rathenauplatz 8	Meister Gerhard no 008	1	8.8	0.785486
15	4b05886af964a52045c422e3	Orlando	Café	316	50764	Engelbertstr. 9	Orlando	1	8.6	0.784167
10	54d7540d498ed907b5f12ed4	Bangkok	Thai Restaurant	285	50674	Lindenstraße 81	Bangkok	1	8.2	0.763819
12	4b05886ff964a520e9c522e3	El Inca	Peruvian Restaurant	294	50674	Görresstr. 2	El Inca	1	8.2	0.760694
13	51d506fe498e56324b31353c	Gernys Schnelleinkauf	Soup Place	296	50676	Schaafenstr. 53-55	Gernys Schnelleinkauf	1	7.9	0.736667
29	4b0e556ff964a520e35623e3	cafecafe	Café	401	50674	Aachener Str. 45	cafecafe	1	8.3	0.731319
1	57645427498e7a5746a35766	Tanica	Italian Restaurant	142	50674	Engelbertstraße 31	Tanica	2	8.1	0.730694

We now see a list of suggested venues in the area sorted by the calculated score.

Discussion

In this project I tried to introduce a model to help using available Foursquare data in order to offer venues, that are within walking distance, are ranked reasonable high and offer a good price range. While this model tried to quantify these characteristics, we must keep in mind that it is a simplification and putting more time in the project could to more complex model that involves additional characteristics of the venue that might be available on Foursquare or even on other platforms in the web. It is also important to state, that the suggested model might be a helpful tool for people that are within the area and looking for venues of a rather broad spectrum. However, for instance if someone is looking for a great italian restaurant, this project is not modelled for such a person.

While looking at the results and the calculated score of the above example, the venues that are suggested are indeed also venues that I personally very like and therefore I am confident that this very simple model is a good way to incorporate the characteristics in a quantitative way.

Conclusion

This small project showed in a very understandable way, how to use public available data to help decision making with limited information. Especially for a big City like Cologne with a variety of venues it is nearly impossible to have a complete overview over the available location. For the completeness, it is to say that this project can be conducted for any other city with entries on FourSquare.