Applied Data Science - Capstone Project

Analysis of gyms in Nuremberg, Germany

Introduction

This report analyses the gyms in the city of Nuremberg, Germany based on the data that is available from Foursquare.

This analysis would be interesting for two types of businesses:
- An entrepreneur who wants to open a new gym in order to assess the competitors
- The owners of existing gyms in order to set themselves apart from the competition

Data

All data was extracted from the Foursquare database. This database contains location-based data worldwide. Using center coordinates or addresses, venues in the vicinity can be queried together with a lot of information about the venue, like opening hours, ratings or tips.
For this analysis the following data was extracted from Foursquare.
- Venue Id
- Venue Name
- Venue Latitude
- Venue Longitude
- Venue Category
- Rating
- Number of Tips
- Number of Likes
- Number of Dislikes

The data was extracted on the 19th of November 2020.
The foursquare API version used was '20180605'.

Methodology

The center coordinates for Nuremberg, Germany were fetched via geopy.geocoder Nominatim. Based on those coordinates all gyms within a radius of 5000 meters were queried. In order to get the gyms, the Venues/Search Endpoint was used together with a categoryId. This catagoryId was the top categoryId for gyms which contains lots of sub catagories. An extract is shown below.

Gym / Fitness Center
4bf58dd8d48988d175941735

Boxing Gym
52f2ab2ebcbc57f1066b8b47

Climbing Gym
503289d391d4c4b30a586d6a
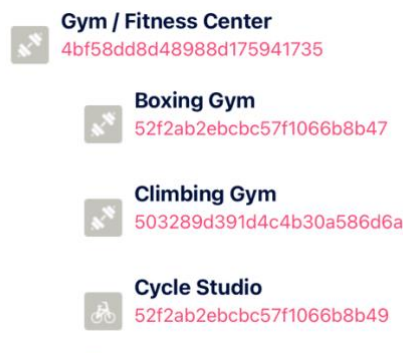
Cycle Studio
52f2ab2ebcbc57f1066b8b49

Figure 1: Top and sub categories as shown on the Foursquare webpage

The results of this query were collected and assembled like shown below.

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Id
0	Nürnberg	49.453872	11.077298	FIT ONE	49.449467	11.074092	Gym / Fitness Center	586ca64e18dc536fecf25667
1	Nürnberg	49.453872	11.077298	FIT STAR Fitnessstudio Nürnberg-Zentrum	49.448212	11.083224	Gym / Fitness Center	577d6a44498eadda3146ab2d
2	Nürnberg	49.453872	11.077298	Schwimmbad Ebensee	49.467580	11.139481	Gym Pool	52dd6b8c498e35e77df9f275
3	Nürnberg	49.453872	11.077298	FIT STAR Fitnessstudio Fürth-Südstadt	49.459940	11.006402	Gym / Fitness Center	577d68b0498eea8e9db20e71
4	Nürnberg	49.453872	11.077298	clever fit Fitnessstudio Nürnberg	49.423984	11.068681	Gym	50a2da52f2e7ce3e7a58855e

Figure 2: Gyms returned by Foursquare query from center coordinates

Based on the Venue Id obtained in the first request, the popularity data (Rating, Number of Tips, Number of Likes, Number of Dislikes) was queried in a second request to Foursquare.

	Venue Id	Rating	No of Tips	No of Likes	No of Dislikes
0	586ca64e18dc536fecf25667	7.8	1	8	NaN
1	577d6a44498eadda3146ab2d	NaN	0	1	NaN
2	52dd6b8c498e35e77df9f275	NaN	0	1	NaN
3	577d68b0498eea8e9db20e71	NaN	0	1	NaN
4	52652161498efba58d4440e0	NaN	0	1	NaN
5	50a2da52f2e7ce3e7a58855e	7.3	2	6	NaN
6	5310437c11d2a01b7c45b5b6	NaN	0	1	NaN
7	4bd9a7ece914a593034557fa	8.6	3	34	NaN

Figure 3: Rating data from Foursquare

Kmeans from sklearn was used in order to cluster the gyms based on their popularity data. Folium was used to display the results on a map.

Results

The foursquare database returned 43 gyms within a radius of 5000m of the city center.
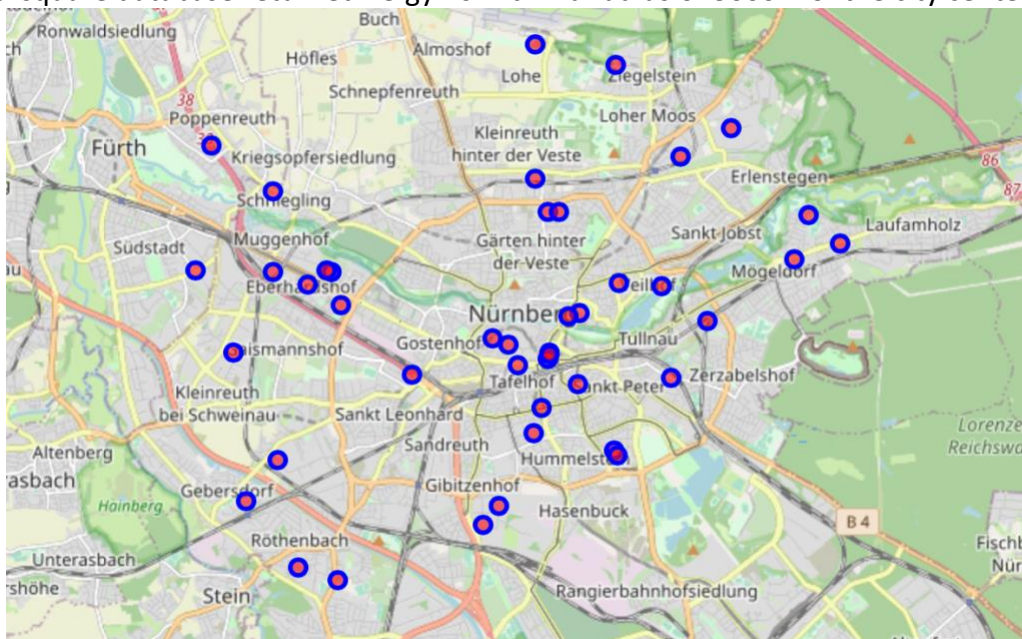

Figure 4: Gyms within 5km radius from Nuremberg city center

Of those 43 gyms,

- only 11 were rated on Foursquare, range 6.7 – 8.9 (min:0, max:10)
- 13 had tips (range 1 – 8 tips)
- 28 had likes (range 1- 51 tips)
- None had dislikes.

Of the 11 rated gyms,

- All had a medium of high number of likes (range: 6 – 51)
- All but one, had tips

Of the 32 gyms without a rating,

- 3 had tips (range: 2-3 tips)
- 17 has likes (range: 1 – 5 likes)

With kmeans the gyms were clustered into four groups based on the rating data:

- Cluster 0 (red) was the biggest one with 32 gyms. It contained all the gyms where no rating was available.
- Cluster 1 (violet) contained 3 high rated gyms with a high number of likes
- Cluster 2 (blue) contained 5 gyms with medium number of likes
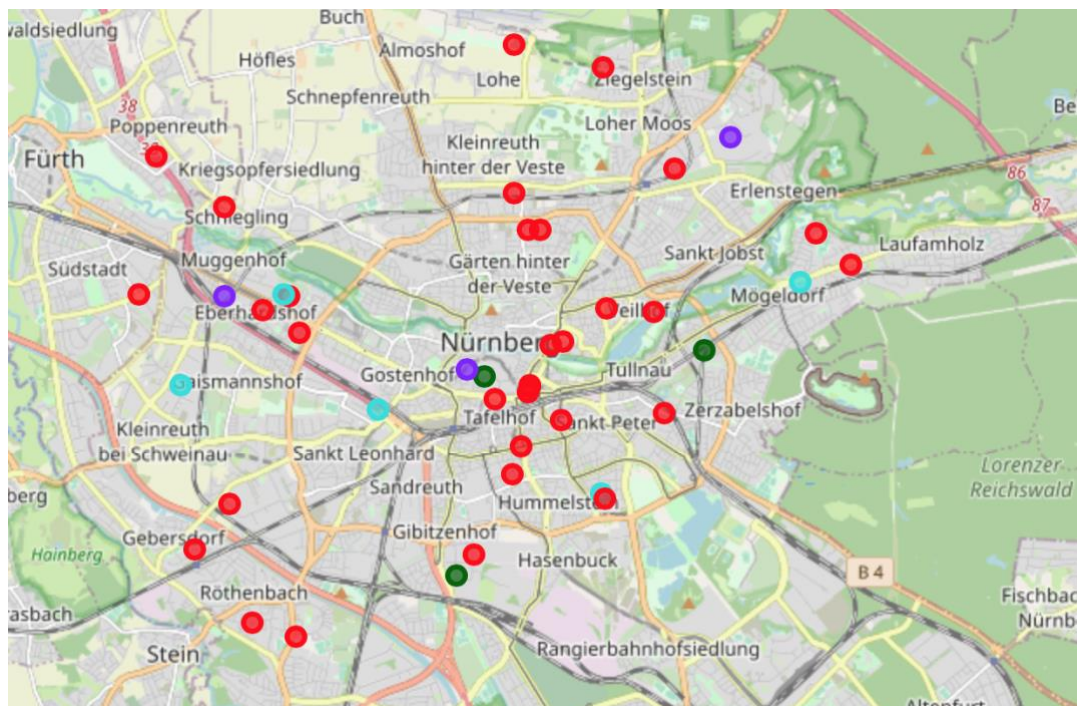- Cluster 3 (green) contained 3 gyms with medium rating low number of likes

Figure 5: Gyms clustered

Discussion

Looking at how the gyms are scattered in the city of Nuremberg one can spot several areas where no gym is nearby. These areas would be a good starting point to investigate more in depth. Like checking the population structure (density, age, income, mobility) and the rent prices.

Looking at the available popularity data of the gyms one can see much room for improvement. With only 11 gyms being rated in total, there is a big opportunity for existing business owners to make their gym more visible and attractive to new customers. They could make a contest and give away free course hours to people who enter ratings, tips or likes on Foursquare. Newcomers could also use the low availability of popularity data to their advantage. By making sure that their gym is listed, rated and has tips they are up for a good start.

Some gyms have a quite good user interaction on Foursquare, much of the popularity information is provided. It could be worth a further investigation to find out why this is the case.

In general, data availability is an issue for this analysis. From personal experience I know that there are gyms missing. i.e. they are not listed on Foursquare. And it is unclear how current the data is.

Conclusion

This report provides an initial overview of how the gyms in Nuremberg, Germany are represented on Foursquare. Based on the data extracted from Foursquare several areas with no gym nearby could be found. The venue data that is available on Foursquare is quite sketchy. While for some gyms good popularity information is available, it is missing for most of them. There is much room for improvement regarding completeness. This could be leveraged by both, new entrants and existing players to make their gym more visible and attractive to digital natives.

Learning

While doing this analysis, I came across several surprising or unexpected issues which I want to share here:

- When searching for venues in a not English-speaking country one needs to be extra careful and double check the results. For example, searching for gyms in Germany using the Venues-Search endpoint with "query" for "gym" returns a list of schools (the German "Gymnasium" is a type of school and not a gym!). To overcome this, the Venues-Search endpoint needs to be used with "CategoryId" and the corresponding top category Id for "Gym / Fitness Center".
- Sometimes the Foursquare API returns unexpected venues. For example, when searching for "CategoryId" '4bf58dd8d48988d175941735' ("Gym/Fitness Center") a lot of venues of the expected category have been returned but also one "Food and Drink Shop"

```python
name = "Nürnberg"
lat = 49.453872
lng = 11.077298
search_category = '4bf58dd8d48988d175941735' # category ID for "Gym/Fitness Center"
radius = 300
url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&v={}&categoryId={}&r
print("Url: ", url)
results = requests.get(url).json()["response"]['venues']
results
```

```
Url:  https://api.foursquare.com/v2/venues/search?client_id=1ASA1R5SJNI4KKZDE0VT02FRVCZPTLGRT35LWIICH2VGSSYW&c
nt_secret=AQEUCQQX3WF40RDYNTMNFCVRNNV5OY1CC3MCVOSD212UVEKL&ll=49.453872,11.077298&v=20180605&categoryId=4bf58d
48988d175941735&radius=300&limit=100
[{'id': '50ec2c28e4b0c0cf858fc47b',
  'name': 'Bodyvision',
  'location': {'address': 'Königstr. 13',
   'lat': 49.45150611743534,
   'lng': 11.077941656112671,
   'labeledLatLngs': [{'label': 'display',
     'lat': 49.45150611743534,
     'lng': 11.0779416561126712}],
   'distance': 267,
   'postalCode': '90402',
   'cc': 'DE',
   'city': 'Nürnberg',
   'state': 'Bayern',
   'country': 'Deutschland',
   'formattedAddress': ['Königstr. 13', '90402 Nürnberg', 'Deutschland']},
  'categories': [{'id': '4bf58dd8d48988d1f9941735',
    'name': 'Food & Drink Shop',
    'pluralName': 'Food & Drink Shops',
    'shortName': 'Food & Drink',
    'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/shops/foodanddrink_',
     'suffix': '.png'},
    'primary': True}],
  'referralId': 'v-1605519281',
  'hasPerk': False}]
```

Figure 6: Unexpected results from Foursquare

  Similarly, when searching for 'Fitness' in the venue name (API:Venues-Search-query), also one venue that does not have "Fitness" in its name has been returned.
- Depending on the radius given in the search query, Foursquare returns different venues. I assumed that it would just return more venues with an increasing radius. But this is not the case. This was quite surprising and unexpected as the total limit of 100 venues was not reached even with the biggest radius. See the jupyter notebook for details.
- Reduce the foursquare queries to a minimum and extract much information in each query. The number of queries that can be made per day is limited depending on the Foursquare account type used. I use the free account, and during development, it happened that I exceeded my daily quota. So, I had to continue the next day.