Capstone Project - The Battle of Neighborhoods (Week 2)

1. Introduction

In this section a description of the problem and a discussion of the background are provided.

1.1 Background information

Moving from one place to another is something that is done very frequently in our times and can both be very exciting and challenging. There are many reasons for it, e.g., beginning a new job, studying in another city or moving together with your partner. People often want to find a similar place to the one they lived before, especially when they got used to their former neighborhood and felt very comfortable there. Sometimes they also want to change a few things and put more emphasis on different venues nearby, e.g., a park or a certain kind of shop/restaurant. This problem can be tackled using data science methods.

1.2 Problem statement

We are exploring neighbourhoods in the city of Munich, Germany, to find a new apartment for Anna that meets her requirements. Anna is 30 years old and at the moment she is living in the borough 'Maxvorstadt' where she feels very comofortable. However, she is only allowed to stay in her current apartment until the end of the year and there are a few things she does not like about it, e.g., that it is very expensive (19.50 €/m²), has no park nearby and that the surrounding area is a little bit too crowded. As a result, she is looking for a new apartment in another neighbourhood that is ideally very similar to Maxvorstadt. In addition, Anna would enjoy having the following venues nearby:

Venue	Reason	Priority
Park	Anna likes to take a walk in the park after work.	1
Tram Station	Anna requires a tram station nearby to get to work easily.	2
German Restaurant	Anna often comes home late from work and has no time to cook on her own. She is very much into German food.	3
Thai Restaurant	Anna also likes Thai food.	4
Cupcake Shop	Anna loves cupcakes and is thinking about opening her own shop.	5

We will use data science methods to identify the most promising neighborhoods based on these criteria.

1.3 Target audience

Solving this problem can be interesting for everyone who is moving to another city, not exclusively for Anna. The method can be easily transferred to other cities and different people with different needs and preferences. It would also be possible for real estate agents to use similar algorithms to find the perfect home for their customers.

2. Data

In this section a description of the data and how it will be used to solve the problem is provided.

2.1 Data description

In order to find the most promising neighborhood for Anna the following data is needed:

- A list of existing boroughs in Munich, including additional information like Area in km^2 and Inhabitants per km^2. This data is gathered through web scraping from the following webpage: 'https://en.wikipedia.org/wiki/Population_growth_of_Munich#Age_distribution'.
- Geospatial data (latitude, longitude) for each borough. This data is gathered using the Geocoder API based on Open Street Map data and is used as an input for the Foursquare API.
- Average rental prices for all boroughs in Munich. This data is gathered through web scraping from the following webpage: 'https://www.tz.de/leben/wohnen/uebersicht-muenchner-mieten-preise-nach-postleitzahlen-tz-6133643.html'. To work with this data, the postal codes need to be attributed to the corresponding borough via 'https://www.muenchen.de/leben/service/postleitzahlen.html'.
- Information about nearby venues for all boroughs of Munich. This data is gathered using the Foursquare API.

2.2 Data usage

We will start by preparing the data from the websites so that we have one overview table that contains all the information (borough, longitude, latitude, rental price, inhabitants per km^2) we need. In a next step we perform an exploratory data analysis (using the describe method and a histogram plot) to get insights in the average rental prices for different boroughs in Munich, to which we can compare Anna's current rent and the average rent in her current borough, Maxvorstadt. The data is visualized in a Folium map.

We will prepare a second table using Foursquare data, which contains the top 100 venues within a radius of 1.5 km for each borough. We use one hot encoding to get the 10 most common venues for each borough. Clustering the neighborhood with k-means will limit the boroughs to the ones that are similar to Maxvorstadt (e.g., are in the same cluster).

Using the data from the first table we will limit our further analysis to boroughs where the rent is cheaper than Anna's current rent and which are less crowded (inhabitants per km^2) than Maxvorstadt.

Finally, we will visualize Anna's favourite venues color coded on the map and count their occurrence within the list of top 100 venues of each borough. Combining our results gives us the possibility to recommend the best borough for Anna.

2.3 Data preparation

The data is prepared like described below.

2.3.1 Import Python libraries

Python libraries that are required for the data analysis are imported first.

2.3.2 Get a list of existing boroughs in Munich, Germany

Data from the Wikipedia page below is read into the dataframe and cleansed afterwards: 'https://en.wikipedia.org/wiki/Population_growth_of_Munich#Age_distribution'
The first 10 lines are shown in the following table:

	Borough	Area in km²	Inhabitant Count	Inhabitants per km²	Longitude	Latitude
0	Allach-Untermenzing	15.45	27730	1795	NaN	NaN
1	Altstadt-Lehel	3.16	18876	5973	NaN	NaN
2	Aubing-Lochhausen-Langwied	34.06	37857	1111	NaN	NaN
3	Au-Haidhausen	4.22	54382	12887	NaN	NaN
4	Berg am Laim	6.31	39009	6182	NaN	NaN
5	Bogenhausen	23.71	75657	3191	NaN	NaN
6	Feldmoching-Hasenbergl	28.71	54245	1889	NaN	NaN
7	Hadern	9.23	44993	4875	NaN	NaN
8	Laim	5.29	50082	9457	NaN	NaN
9	Ludwigsvorstadt-Isarvorstadt	4.39	45736	10418	NaN	NaN

We now have a dataset containing the 25 boroughs of Munich. It also includes additional information like the number of inhabitants per km² to judge whether a borough is less crowded than Maxvorstadt.

2.3.3 Add coordinates to each borough (using Geocoder API)

A function is defined that requests the coordinates of each borough using the Geocoder API. The coordinates for the city of Munich are: 11.5753822 E, 48.1371079 N.

We are looping through all boroughs and append the received coordinates to the dataframe presented above. The first 10 lines of this dataframe are shown below:

	Borough	Area in km²	Inhabitant Count	Inhabitants per km²	Longitude	Latitude
0	Allach-Untermenzing	15.45	27730	1795	11.462973	48.195157
1	Altstadt-Lehel	3.16	18876	5973	11.574582	48.137828
2	Aubing-Lochhausen-Langwied	34.06	37857	1111	11.400221	48.165059
3	Au-Haidhausen	4.22	54382	12887	11.590536	48.128753
4	Berg am Laim	6.31	39009	6182	11.633600	48.133971
5	Bogenhausen	23.71	75657	3191	11.633484	48.154782
6	Feldmoching-Hasenbergl	28.71	54245	1889	11.541275	48.213804
7	Hadern	9.23	44993	4875	11.481842	48.118064
8	Laim	5.29	50082	9457	11.503835	48.144352
9	Ludwigsvorstadt-Isarvorstadt	4.39	45736	10418	11.573366	48.130340

2.3.4 Add rental price per squaremeter for each borough

In the first step we gather a table of all postal codes in Munich from a webpage ("https://www.muenchen.de/leben/service/postleitzahlen.html") and attribute them to the corresponding borough so that we end up with a table containing each single postal code and its borough. The first 10 lines are shown below:

	Borough	PostalCode
0	Allach-Untermenzing	80995
1	Allach-Untermenzing	80997
2	Allach-Untermenzing	80999
3	Allach-Untermenzing	81247
4	Allach-Untermenzing	81249
5	Altstadt-Lehel	80331
6	Altstadt-Lehel	80333
7	Altstadt-Lehel	80335
8	Altstadt-Lehel	80336
9	Altstadt-Lehel	80469

In the next step we gather rental and buy prices sorted by postal code from the following webpage and print the first 10 lines of the table:
'https://www.tz.de/leben/wohnen/uebersicht-muenchner-mieten-preise-nach-postleitzahlen-tz-6133643.html'.

	PostalCode	Rental price	Rental trend	Buy price	Buy trend
2	80995	1410	1,1%	5000	6,8%
3	80997	1325	-1,9%	5430	12,4%
4	80999	1305	5,2%	5880	11,4%
5	81247	1455	2,5%	6520	1,4%
6	81249	1325	5,6%	5100	3,4%
8	80331	2230	3,5%	k.A.	k.A.
9	80333	1910	1,9%	9120	20,8%
10	80335	1955	2,9%	8690	5,8%
11	80336	1815	0,0%	8960	9,0%
12	80469	206	4,8%	8370	0,5%

Next, we group the values by boroughs and calculate an average rental price per borough.

	Rental price
Borough	
Allach-Untermenzing	1370.555556
Altstadt-Lehel	1877.555556
Au-Haidhausen	1708.571429
Aubing-Lochhausen-Langwied	1378.333333
Berg am Laim	1478.750000
Bogenhausen	1729.375000
Feldmoching-Hasenbergl	1401.250000
Hadern	1456.666667
Laim	1521.428571
Ludwigsvorstadt-Isarvorstadt	1784.307692

We end up with the dataframe 'df_nb' containing all the necessary information for each borough, including name, area, inhabitant count, longitude, latitude and average rent per m².

	Borough	Area in km²	Inhabitant Count	Inhabitants per km²	Longitude	Latitude	Rent per m²
0	Allach-Untermenzing	15.45	27730	1795	11.462973	48.195157	13.71
1	Altstadt-Lehel	3.16	18876	5973	11.574582	48.137828	18.78
2	Aubing-Lochhausen-Langwied	34.06	37857	1111	11.400221	48.165059	17.09
3	Au-Haidhausen	4.22	54382	12887	11.590536	48.128753	13.78
4	Berg am Laim	6.31	39009	6182	11.633600	48.133971	14.79
5	Bogenhausen	23.71	75657	3191	11.633484	48.154782	17.29
6	Feldmoching-Hasenbergl	28.71	54245	1889	11.541275	48.213804	14.01
7	Hadern	9.23	44993	4875	11.481842	48.118064	14.57
8	Laim	5.29	50082	9457	11.503835	48.144352	15.21
9	Ludwigsvorstadt-Isarvorstadt	4.39	45736	10418	11.573366	48.130340	17.84
10	Maxvorstadt	4.29	46058	10736	11.562418	48.151092	19.98
11	Milbertshofen-Am Hart	13.37	66992	5011	11.575043	48.182385	14.73
12	Moosach	11.09	47754	4306	11.506122	48.180166	15.11
13	Neuhausen-Nymphenburg	12.92	84604	6548	11.531517	48.154222	16.44
14	Obergiesing	5.71	47007	8232	11.596084	48.111130	15.78
15	Pasing-Obermenzing	16.50	63763	3864	11.461767	48.149956	14.46
16	Ramersdorf-Perlach	19.90	102689	5160	11.633371	48.100894	15.17
17	Schwabing-Freimann	25.67	62430	2432	11.608583	48.189278	18.66
18	Schwabing-West	4.37	59553	13628	11.569873	48.168271	19.27
19	Schwanthalerhöhe	2.07	26103	12610	11.541057	48.133782	18.97
20	Sendling	3.94	37146	9428	11.512817	48.124588	16.14
21	Sendling-Westpark	7.81	50903	6518	11.519333	48.118031	15.18
22	Thalkirchen-Obersendling-Forstenried-Fürstenri...	17.75	80701	4547	11.508051	48.084213	14.43
23	Trudering-Riem	22.45	53915	2401	11.663338	48.126036	14.78
24	Untergiesing-Harlaching	8.06	48075	5965	11.570189	48.114963	15.50

2.3.5 Using Foursquare to explore the area around the boroughs

First, we use geopy to get the latitude and longitude values of Munich, which are 48.1371079° N, 11.5753822 °E. Then we define Foursquare credentials and version as well as a limit for requesting nearby venues (which is 100 at maximum).

A function based on the Foursquare API is created that finds the top 100 venues within a radius of 1.5 km for each borough.

We run this function on all 25 boroughs and get 1845 venues with 224 unique venue categories. This also means that we do not find 100 venues in each borough. A list is shown below:

```
Neighborhood
Allach-Untermenzing                                          21
Altstadt-Lehel                                              100
Au-Haidhausen                                               100
Aubing-Lochhausen-Langwied                                   12
Berg am Laim                                                 70
Bogenhausen                                                  98
Feldmoching-Hasenbergl                                       21
Hadern                                                       41
Laim                                                         91
Ludwigsvorstadt-Isarvorstadt                                100
Maxvorstadt                                                 100
Milbertshofen-Am Hart                                       100
Moosach                                                      54
Neuhausen-Nymphenburg                                        99
Obergiesing                                                 100
Pasing-Obermenzing                                           82
Ramersdorf-Perlach                                           64
Schwabing-Freimann                                           58
Schwabing-West                                              100
Schwanthalerhöhe                                            100
Sendling                                                     63
Sendling-Westpark                                            75
Thalkirchen-Obersendling-Forstenried-Fürstenried-Solln       58
Trudering-Riem                                               38
Untergiesing-Harlaching                                     100
Name: Venue, dtype: int64
```

Finally, we end up with the dataframe 'munich_venues', which contains the top 100 venues of all boroughs in Munich. The first 5 lines are shown below.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Allach-Untermenzing	48.195157	11.462973	Bäckerei Schuhmair	48.197175	11.459016	Bakery
1	Allach-Untermenzing	48.195157	11.462973	Sport Bittl	48.191447	11.466553	Sporting Goods Shop
2	Allach-Untermenzing	48.195157	11.462973	Trattoria Olive	48.189905	11.466970	Trattoria/Osteria
3	Allach-Untermenzing	48.195157	11.462973	Zur Allacher Mühle	48.198411	11.457869	Bavarian Restaurant
4	Allach-Untermenzing	48.195157	11.462973	Würmtalhof	48.188834	11.460680	German Restaurant

2.4 Summary of datasets

Further data analysis can now be performed on the two data sets resulting from our data preparation:

- 'df_nb' (borough, inhabitant count, longitude, latitude, rent)
- 'munich_venues' (top 100 venues in each of the borough within a radius of 1.5 km).

3. Methodology

This section represents the main component of the report.

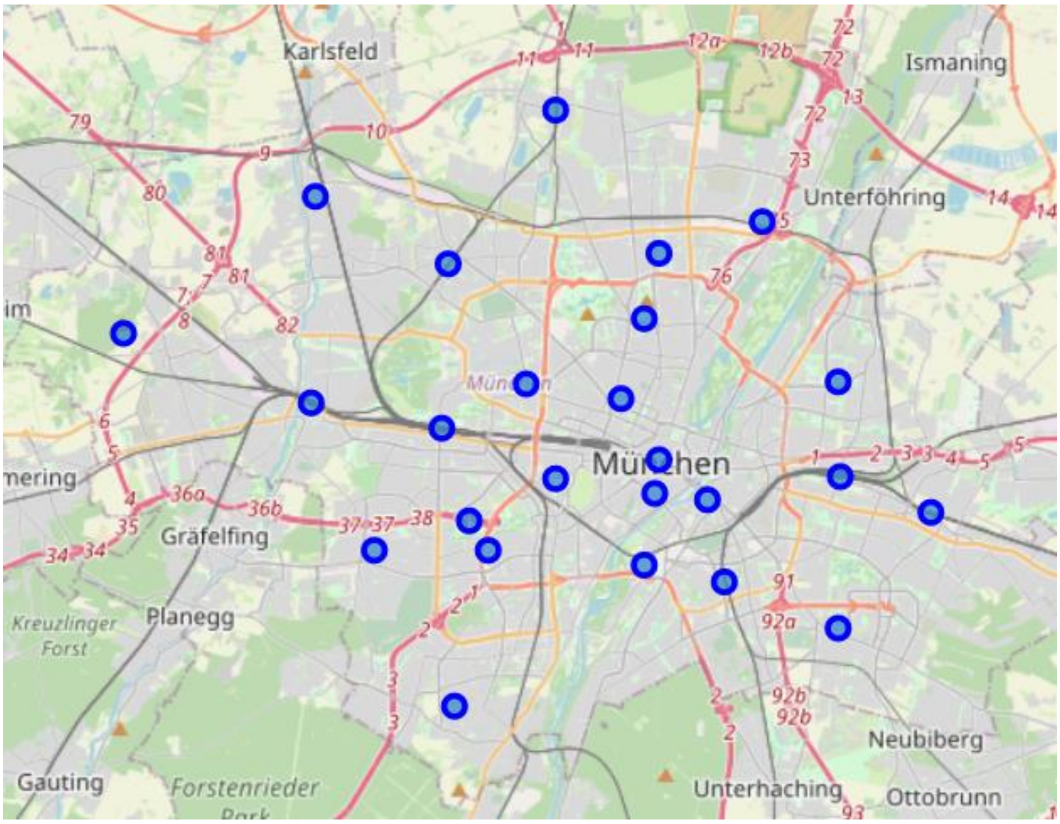
3.1 Exploratory Data Analysis

As mentioned in the previous section, we have the two following datasets that can be used to solve the question, which borough is the best one for Anna's new apartment:

- The dataframe 'df_nb' is used to find a borough which is less crowded than Maxvorstadt and where the average rental price is less than what Anna is paying now (19.50 €/m²).
- The dataframe 'munich_venues' is used to explore and cluster neighborhoods, inclduing venues of interest nearby that meet Anna's expectations.

3.1.1 Exploring boroughs of Munich and show them on map

We use geopy and Folium to create a map of Munich with boroughs superimposed on top:



3.2 Analyze Foursquare dataset via one hot encoding

For further analysis we use one hot encoding to gather the frequency of all venues for each borough:

	Neighborhood	Accessories Store	Afghan Restaurant	American Restaurant	Arcade	Art Gallery	Art Museum	Asian Restaurant	Athletics & Sports	Austrian Restaurant	...	Tram Station	Trattoria/Osteria
0	Allach-Untermenzing	0	0	0	0	0	0	0	0	0	...	0	0
1	Allach-Untermenzing	0	0	0	0	0	0	0	0	0	...	0	0
2	Allach-Untermenzing	0	0	0	0	0	0	0	0	0	...	0	1
3	Allach-Untermenzing	0	0	0	0	0	0	0	0	0	...	0	0
4	Allach-Untermenzing	0	0	0	0	0	0	0	0	0	...	0	0

5 rows × 225 columns

Then we group the rows by neighborhood and by taking the mean of the frequency of occurrence of each category:

	Neighborhood	Accessories Store	Afghan Restaurant	American Restaurant	Arcade	Art Gallery	Art Museum	Asian Restaurant	Athletics & Sports	Austrian Restaurant	...	Tram Station	Trattoria/Osteria
0	Allach-Untermenzing	0.000000	0.00	0.000000	0.0	0.00	0.00	0.000000	0.000000	0.0	...	0.000000	0.095238
1	Altstadt-Lehel	0.000000	0.00	0.000000	0.0	0.00	0.01	0.000000	0.000000	0.0	...	0.000000	0.010000
2	Au-Haidhausen	0.000000	0.02	0.000000	0.0	0.01	0.00	0.000000	0.000000	0.0	...	0.000000	0.010000
3	Aubing-Lochhausen-Langwied	0.000000	0.00	0.000000	0.0	0.00	0.00	0.000000	0.000000	0.0	...	0.000000	0.000000
4	Berg am Laim	0.014286	0.00	0.000000	0.0	0.00	0.00	0.028571	0.000000	0.0	...	0.042857	0.000000
5	Bogenhausen	0.000000	0.00	0.000000	0.0	0.00	0.00	0.040816	0.020408	0.0	...	0.020408	0.010204
6	Feldmoching-Hasenbergl	0.000000	0.00	0.000000	0.0	0.00	0.00	0.000000	0.000000	0.0	...	0.000000	0.000000
7	Hadern	0.000000	0.00	0.000000	0.0	0.00	0.00	0.000000	0.000000	0.0	...	0.024390	0.024390
8	Laim	0.000000	0.00	0.010989	0.0	0.00	0.00	0.000000	0.010989	0.0	...	0.021978	0.010989
9	Ludwigsvorstadt-Isarvorstadt	0.000000	0.02	0.000000	0.0	0.01	0.00	0.010000	0.000000	0.0	...	0.000000	0.000000

10 rows × 225 columns

Based on this dataframe we can define the top 10 most common venues for each borough. For the first borough it looks like this:

```
----Allach-Untermenzing----
                 venue  freq
0                Hotel  0.14
1         Shopping Mall  0.10
2               Bakery  0.10
3   Bavarian Restaurant  0.10
4     Trattoria/Osteria  0.10
5            Drugstore  0.10
6          Supermarket  0.10
7     German Restaurant  0.05
8   Sporting Goods Shop  0.05
9                 Park  0.05
```

We put the results into the dataframe ‘df_nb’:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Allach-Untermenzing	Hotel	Drugstore	Bavarian Restaurant	Supermarket	Bakery	Shopping Mall	Trattoria/Osteria	Italian Restaurant	Light Rail Station	German Restaurant
1	Altstadt-Lehel	Plaza	Café	Bavarian Restaurant	German Restaurant	Hotel	Coffee Shop	Restaurant	Gourmet Shop	Department Store	Cocktail Bar
2	Au-Haidhausen	Café	Plaza	German Restaurant	Hotel	Coffee Shop	Bavarian Restaurant	Pub	Beer Garden	Cocktail Bar	Beach
3	Aubing-Lochhausen-Langwied	Soccer Field	Bus Stop	Bakery	Bistro	Supermarket	Light Rail Station	German Restaurant	Hotel	Post Office	Pharmacy
4	Berg am Laim	Hotel	Italian Restaurant	Supermarket	Bakery	Drugstore	Tram Station	German Restaurant	Asian Restaurant	Gym	Gastropub
5	Bogenhausen	Italian Restaurant	Bus Stop	Supermarket	Drugstore	Hotel	Bakery	Asian Restaurant	Beer Garden	German Restaurant	Plaza
6	Feldmoching-Hasenbergl	Supermarket	Plaza	Drugstore	Bakery	Italian Restaurant	Gastropub	German Restaurant	Greek Restaurant	Intersection	Café
7	Hadern	Supermarket	Bus Stop	Bakery	Bank	German Restaurant	Drugstore	Ice Cream Shop	Sushi Restaurant	Residential Building (Apartment / Condo)	Sandwich Place
8	Laim	Supermarket	Greek Restaurant	Plaza	Italian Restaurant	Drugstore	Pizza Place	Bakery	Hotel	Park	Gastropub
9	Ludwigsvorstadt-Isarvorstadt	Café	Coffee Shop	Italian Restaurant	Bavarian Restaurant	German Restaurant	Plaza	Cocktail Bar	Ice Cream Shop	Pizza Place	Bookstore
10	Maxvorstadt	Café	Plaza	Bakery	Vietnamese Restaurant	Bar	Ice Cream Shop	Italian Restaurant	Hotel	Steakhouse	Trattoria/Osteria
11	Milbertshofen-Am Hart	Hotel	Italian Restaurant	Bakery	Drugstore	Greek Restaurant	Supermarket	Café	Gastropub	Restaurant	Museum
12	Moosach	Supermarket	Bakery	Drugstore	Hotel	Bus Stop	Park	Trattoria/Osteria	German Restaurant	Italian Restaurant	Food
13	Neuhausen-Nymphenburg	Italian Restaurant	Café	Plaza	German Restaurant	Hotel	Supermarket	Drugstore	Greek Restaurant	Sushi Restaurant	Restaurant
14	Obergiesing	Supermarket	Greek Restaurant	German Restaurant	Drugstore	Italian Restaurant	Plaza	Hotel	Restaurant	Bar	Bakery
15	Pasing-Obermenzing	Supermarket	German Restaurant	Drugstore	Italian Restaurant	Ice Cream Shop	Bus Stop	Bakery	Coffee Shop	Tram Station	Asian Restaurant
16	Ramersdorf-Perlach	Supermarket	Bakery	Clothing Store	German Restaurant	Ice Cream Shop	Asian Restaurant	Shopping Mall	Italian Restaurant	Gym / Fitness Center	Bus Stop
17	Schwabing-Freimann	Supermarket	Gym / Fitness Center	Fast Food Restaurant	Greek Restaurant	Asian Restaurant	Furniture / Home Store	Beer Garden	Italian Restaurant	Bus Stop	Bakery
18	Schwabing-West	Italian Restaurant	Greek Restaurant	Restaurant	Vietnamese Restaurant	Plaza	Thai Restaurant	Bakery	Park	Asian Restaurant	Museum
19	Schwanthalerhöhe	Café	Hotel	Asian Restaurant	Italian Restaurant	Bavarian Restaurant	German Restaurant	Vietnamese Restaurant	Music Venue	Bistro	Lounge
20	Sendling	Bakery	Bus Stop	Supermarket	Greek Restaurant	German Restaurant	Ice Cream Shop	Italian Restaurant	Restaurant	Beer Garden	Park
21	Sendling-Westpark	Supermarket	Hotel	Bakery	Drugstore	Greek Restaurant	German Restaurant	Italian Restaurant	Ice Cream Shop	Gym / Fitness Center	Beer Garden
22	Thalkirchen-Obersendling-Forstenried-Fürstenri...	Supermarket	Hotel	Gym / Fitness Center	Drugstore	German Restaurant	Bus Stop	BBQ Joint	Italian Restaurant	Gas Station	Organic Grocery
23	Trudering-Riem	Supermarket	Hotel	Drugstore	Bakery	German Restaurant	Accessories Store	Indian Restaurant	Paper / Office Supplies Store	Pub	Sandwich Place
24	Untergiesing-Harlaching	Italian Restaurant	Café	German Restaurant	Bar	Greek Restaurant	Park	Ice Cream Shop	Pub	Beach	Plaza

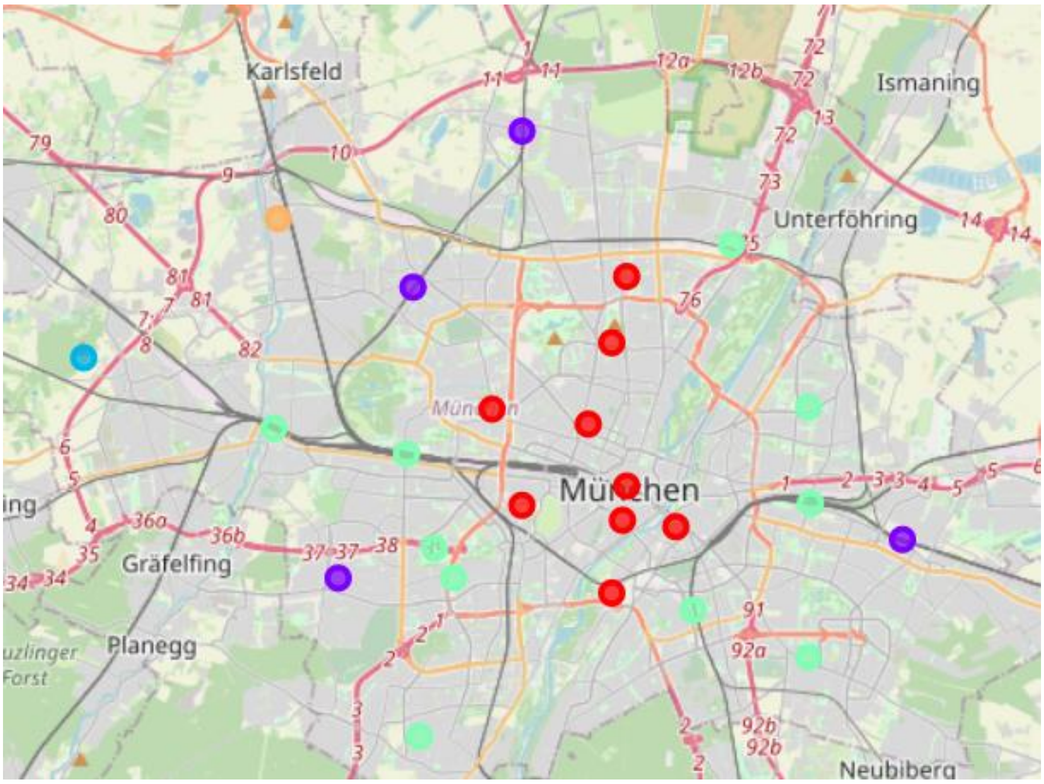
3.3 Cluster neighborhoods

To find neighbourhoods that are similar to Maxvorstadt we run k-means clustering algorithm with k = 5.

Then we create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood:

	Borough	Area in km²	Inhabitant Count	Inhabitants per km²	Longitude	Latitude	Rent per m²	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Allach-Untermenzing	15.45	27730	1795	11.462973	48.195157	13.71	4	Hotel	Drugstore	Bavarian Restaurant	Supermarket	Bakery
1	Altstadt-Lehel	3.16	18876	5973	11.574582	48.137828	18.78	0	Plaza	Café	Bavarian Restaurant	German Restaurant	Hotel
2	Aubing-Lochhausen-Langwied	34.06	37857	1111	11.400221	48.165059	17.09	2	Soccer Field	Bus Stop	Bakery	Bistro	Supermarket
3	Au-Haidhausen	4.22	54382	12887	11.590536	48.128753	13.78	0	Café	Plaza	German Restaurant	Hotel	Coffee Shop
4	Berg am Laim	6.31	39009	6182	11.633600	48.133971	14.79	3	Hotel	Italian Restaurant	Supermarket	Bakery	Drugstore
5	Bogenhausen	23.71	75657	3191	11.633484	48.154782	17.29	3	Italian Restaurant	Bus Stop	Supermarket	Drugstore	Hotel
6	Feldmoching-Hasenbergl	28.71	54245	1889	11.541275	48.213804	14.01	1	Supermarket	Plaza	Drugstore	Bakery	Italian Restaurant
7	Hadern	9.23	44993	4875	11.481842	48.118064	14.57	1	Supermarket	Bus Stop	Bakery	Bank	German Restaurant
8	Laim	5.29	50082	9457	11.503835	48.144352	15.21	3	Supermarket	Greek Restaurant	Plaza	Italian Restaurant	Drugstore
9	Ludwigsvorstadt-Isarvorstadt	4.39	45736	10418	11.573366	48.130340	17.84	0	Café	Coffee Shop	Italian Restaurant	Bavarian Restaurant	German Restaurant

The results are visualized on a Folium map by color-coding the boroughs based on their cluster:



To examine the 5 different clusters, a list of all boroughs included is shown below. The boroughs with cluster label 0 are similar to Maxvorstadt, where Anna is currently living.

	Borough	Inhabitants per km²	Longitude	Latitude	Rent per m²	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Altstadt-Lehel	5973	11.574582	48.137828	18.78	0	Plaza	Café	Bavarian Restaurant	German Restaurant	Hotel
3	Au-Haidhausen	12887	11.590536	48.128753	13.78	0	Café	Plaza	German Restaurant	Hotel	Coffee Shop
9	Ludwigsvorstadt-Isarvorstadt	10418	11.573366	48.130340	17.84	0	Café	Coffee Shop	Italian Restaurant	Bavarian Restaurant	German Restaurant
10	Maxvorstadt	10736	11.562418	48.151092	19.98	0	Café	Plaza	Bakery	Vietnamese Restaurant	Bar
11	Milbertshofen-Am Hart	5011	11.575043	48.182385	14.73	0	Hotel	Italian Restaurant	Bakery	Drugstore	Greek Restaurant
13	Neuhausen-Nymphenburg	6548	11.531517	48.154222	16.44	0	Italian Restaurant	Café	Plaza	German Restaurant	Hotel
18	Schwabing-West	13628	11.569873	48.168271	19.27	0	Italian Restaurant	Greek Restaurant	Restaurant	Vietnamese Restaurant	Plaza
19	Schwanthalerhöhe	12610	11.541057	48.133782	18.97	0	Café	Hotel	Asian Restaurant	Italian Restaurant	Bavarian Restaurant
24	Untergiesing-Harlaching	5965	11.570189	48.114963	15.50	0	Italian Restaurant	Café	German Restaurant	Bar	Greek Restaurant

	Borough	Inhabitants per km²	Longitude	Latitude	Rent per m²	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
6	Feldmoching-Hasenbergl	1889	11.541275	48.213804	14.01	1	Supermarket	Plaza	Drugstore	Bakery	Italian Restaurant
7	Hadern	4875	11.481842	48.118064	14.57	1	Supermarket	Bus Stop	Bakery	Bank	German Restaurant
12	Moosach	4306	11.506122	48.180166	15.11	1	Supermarket	Bakery	Drugstore	Hotel	Bus Stop
23	Trudering-Riem	2401	11.663338	48.126036	14.78	1	Supermarket	Hotel	Drugstore	Bakery	German Restaurant

	Borough	Inhabitants per km²	Longitude	Latitude	Rent per m²	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	Aubing-Lochhausen-Langwied	1111	11.400221	48.165059	17.09	2	Soccer Field	Bus Stop	Bakery	Bistro	Supermarket

	Borough	Inhabitants per km²	Longitude	Latitude	Rent per m²	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Berg am Laim	6182	11.633600	48.133971	14.79	3	Hotel	Italian Restaurant	Supermarket	Bakery	Drugstore
5	Bogenhausen	3191	11.633484	48.154782	17.29	3	Italian Restaurant	Bus Stop	Supermarket	Drugstore	Hotel
8	Laim	9457	11.503835	48.144352	15.21	3	Supermarket	Greek Restaurant	Plaza	Italian Restaurant	Drugstore
14	Obergiesing	8232	11.596084	48.111130	15.78	3	Supermarket	Greek Restaurant	German Restaurant	Drugstore	Italian Restaurant
15	Pasing-Obermenzing	3864	11.461767	48.149956	14.46	3	Supermarket	German Restaurant	Drugstore	Italian Restaurant	Ice Cream Shop
16	Ramersdorf-Perlach	5160	11.633371	48.100894	15.17	3	Supermarket	Bakery	Clothing Store	German Restaurant	Ice Cream Shop
17	Schwabing-Freimann	2432	11.608583	48.189278	18.66	3	Supermarket	Gym / Fitness Center	Fast Food Restaurant	Greek Restaurant	Asian Restaurant
20	Sendling	9428	11.512817	48.124588	16.14	3	Bakery	Bus Stop	Supermarket	Greek Restaurant	German Restaurant
21	Sendling-Westpark	6518	11.519333	48.118031	15.18	3	Supermarket	Hotel	Bakery	Drugstore	Greek Restaurant
22	Thalkirchen-Obersendling-Forstenried-Fürstenri...	4547	11.508051	48.084213	14.43	3	Supermarket	Hotel	Gym / Fitness Center	Drugstore	German Restaurant

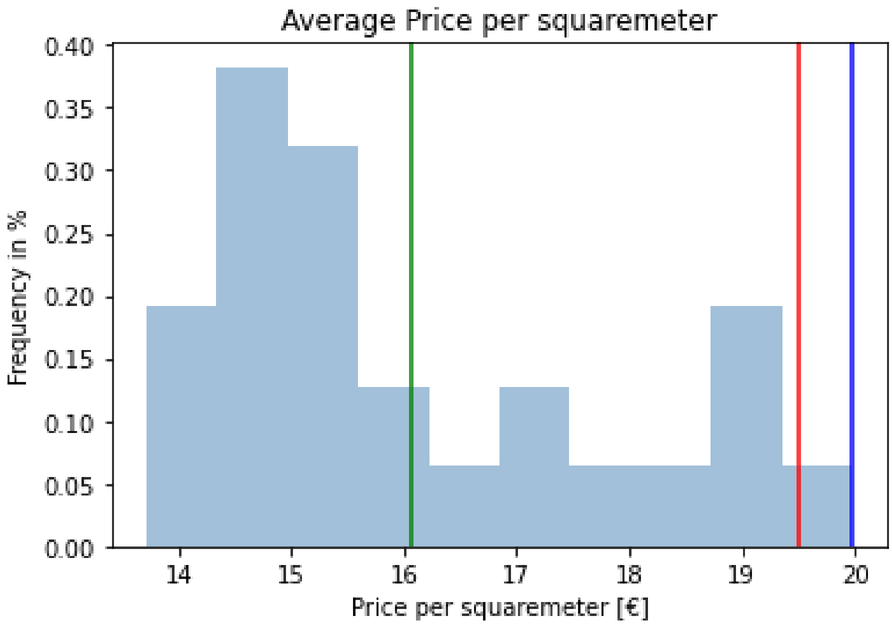
	Borough	Inhabitants per km²	Longitude	Latitude	Rent per m²	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Allach-Untermenzing	1795	11.462973	48.195157	13.71	4	Hotel	Drugstore	Bavarian Restaurant	Supermarket	Bakery

3.4 Find a borough where the average rental price is less than what Anna is paying now

To compare rental prices, we perform statistics using the describe method. We can see that the average rent in Munich is (16.07 +/- 1.88) €/m².

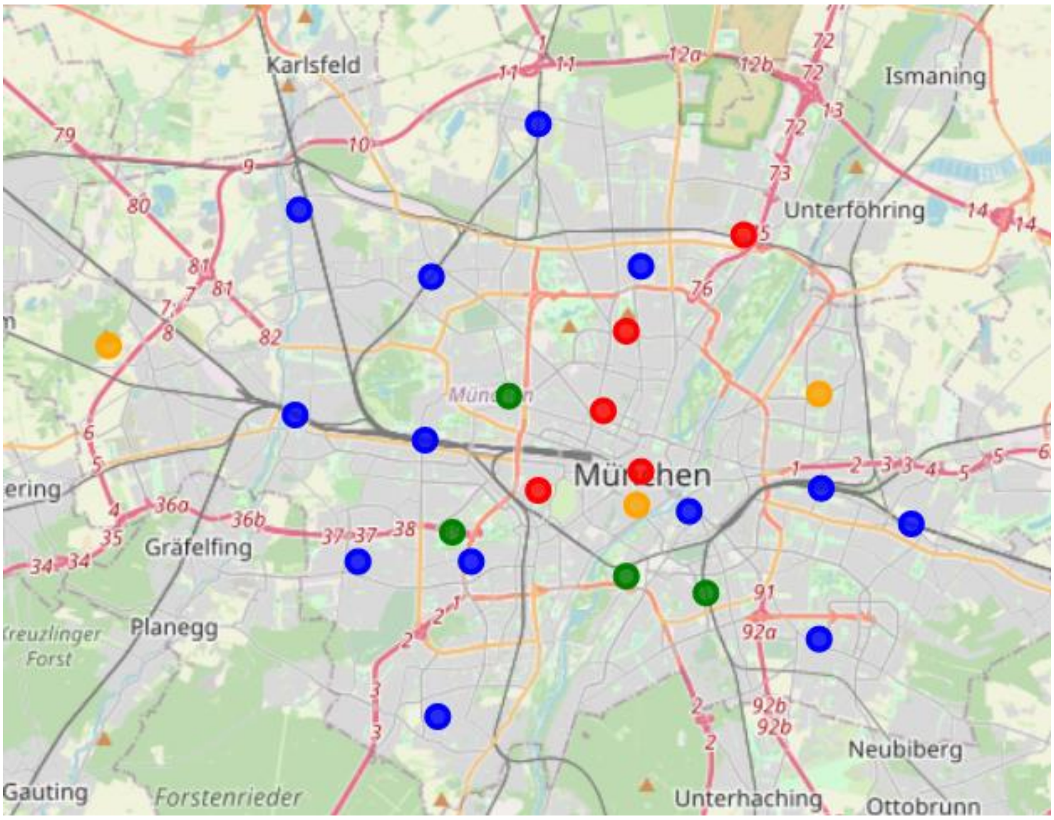
	Rent per m²
count	25.000000
mean	16.066800
std	1.883624
min	13.710000
25%	14.730000
50%	15.210000
75%	17.290000
max	19.980000

In the next step we visualize the average rental prices per m² in a histogram:



Anna's rent (19.50 €/m², red line) is slightly below the average rental price in Maxvorstadt (19.98 €/m², blue line) but still very expensive compared with the average rental price in Munich (16.07 €/m², green line).

The results are visualized on a Folium map, which shows that the most expensive boroughs (red circles) including Maxvorstadt are located close to the city center whereas rental prices in the outskirts (blue circles) are cheaper on average.



3.5 Find borough with average rental price smaller than current rent within cluster 0

We filter the list of boroughs so that only the ones with an average rental price smaller than Anna’s current rent and an inhabitant count per km² smaller than for Maxvorstadt are remaining:

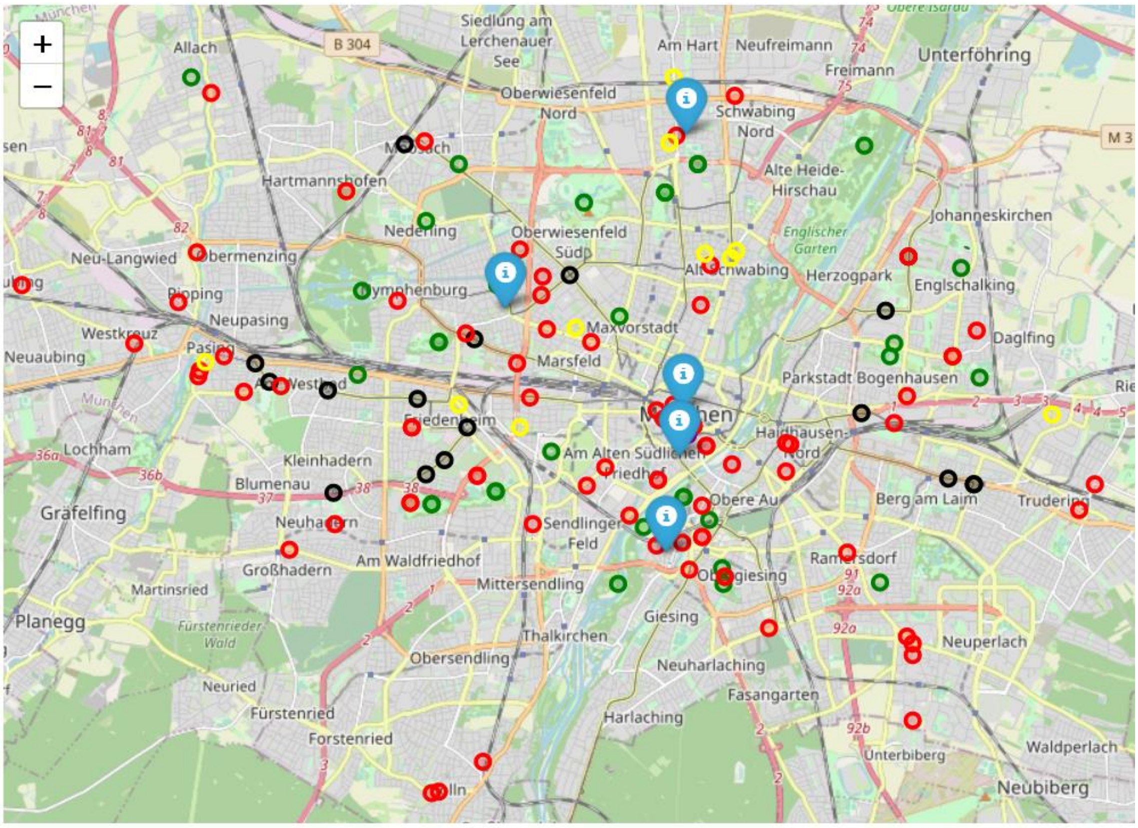
	index	Borough	Area in km²	Inhabitant Count	Inhabitants per km²	Longitude	Latitude	Rent per m²	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	1	Altstadt-Lehel	3.16	18876	5973	11.574582	48.137828	18.78	0	Plaza	Café	Bavarian Restaurant	German Restaurant	Hotel
1	9	Ludwigsvorstadt-Isarvorstadt	4.39	45736	10418	11.573366	48.130340	17.84	0	Café	Coffee Shop	Italian Restaurant	Bavarian Restaurant	German Restaurant
2	11	Milbertshofen-Am Hart	13.37	66992	5011	11.575043	48.182385	14.73	0	Hotel	Italian Restaurant	Bakery	Drugstore	Greek Restaurant
3	13	Neuhausen-Nymphenburg	12.92	84604	6548	11.531517	48.154222	16.44	0	Italian Restaurant	Café	Plaza	German Restaurant	Hotel
4	24	Untergiesing-Harlaching	8.06	48075	5965	11.570189	48.114963	15.50	0	Italian Restaurant	Café	German Restaurant	Bar	Greek Restaurant

3.6 Narrow results based on Anna's favourite venues

Based on the list presented in the beginning, Anna likes to have nearby:

- Park (green)
- Tram Station (black)
- German Restaurant (red)
- Thai Restaurant (yellow)
- Cupcake Shop or possibility to open one (purple)

To visualize the distribution of Anna's favorurite venues along with the remaining boroughs from above:



For a more detailed analysis, we look for Anna's favourite venues within the top 100 venues of each borough and count their occurrence. For this we define a function and loop through all boroughs using this function. The result is shown in the table below:

	Borough	Inhabitants per km²	Rent per m²	# Park	# Tram Station	# German Restaurant	# Thai Restaurant	# Cupcake Shop
0	Altstadt-Lehel	5973	18.78	0	0	5	0	1
1	Ludwigsvorstadt-Isarvorstadt	10418	17.84	1	0	4	0	1
2	Milbertshofen-Am Hart	5011	14.73	2	0	2	2	0
3	Neuhausen-Nymphenburg	6548	16.44	2	2	6	0	0
4	Untergiesing-Harlaching	5965	15.50	4	0	7	0	0

4. Results and Discussion

By performing an exploratory data analysis, using one hot encoding and clustering algroithms we have produced a dataset with all neighborhoods that are similar to Maxvorstadt where Anna is living at the moment. In the next step we have limited them to the ones where the rent is cheaper than 19.50 €/m² and the inhabitant count per km² is less than for Maxvorstadt.

From the final table we can see that there are 5 boroughs which are similar to Maxvorstadt and meet the basic criteria:

- A rent less than what Anna is paying at the moment (19.50 €).
- An inhabitant count per km² less than for Maxvorstadt.

Looking at Anna's favourite venues we can see that:

- Each of the boroughs except for Altstadt-Lehel as a park nearby. There are even 4 parks in Untergiesing-Harlaching.
- It seems that only Neuhausen-Nyphenberg has a tram station close by, which is of course not very relistic (there are many according to Google Maps) and is probably due to the limit of 100 venues per borough returned by the Fourquare API. This column will be ignored in the following.
- There are a lot of German restaurants in each borough with a maximum of 7 in Untergiesing-Harlaching.
- Thai restaurants are only found in Milbertshofen-Am Hart.
- Cupcake stores are present in Altstadt-Lehel and in Ludwigsvorstadt-Isarvorstadt.

4.1 Recommendation based on the analysis

As for Anna it is more import to pay a lower rent and to have a Thai restaurant nearby than a cupcake shop, she will probably decide to move to Milbertshofen-Am Hart. Here she can take a walk in one of the two parks after work and go to one of the two German or Thai restaurants. In her free time she can enjoy her favourite hobby, which is baking. Since there is now cupcake shop in this neighborhood she can work on her dream of opening her own shop.

5. Conclusions

The aim of this project was to identify a borough that is similar to Annas current one (Maxvorstadt) and has as many of her favourite venues nearby. In addition, both the rent per m^2 and the inhabitant count per km^2 should be less than in her current borough.

By just taking into account the similarity to her current neighbourhood, rental price and inhabitant count we identified five boroughs that are generally suitable.

After combining these results with Anna's preferences on venues nearby, we identified one single borough, that is most likely the best choice for Anna to move to: Milbertshofen-Am Hart. This option is also preferred because of the lowest average rental price among the five boroughs.

This data analysis is of course very basic and one should consider a finer division of the neighborhoods so that the limit of 100 venues by the Foursquare API is not biasing the data analysis. There might also be additional factors like availability of apartments, noise, proximity to friends and so on, which should be considered. However, the data analysis presented above should be appropriate to get a first orientation.