Capstone Project: restaurant opening in Munich

Leonardo Mangold

December 2020

Introduction

Munich is the capital of Baviera, a region in Germany. Being the third biggest city in the country, is also the eleventh biggest city in Europe. Having multiple high value industries and millions of tourists yearly, it has a vibrant economic activity.

The purpose of this project is to analyze the neighborhoods of Munich in order to determine possible location for opening a restaurant. I intend to analyze relevant data and provide value insights.

- Data

We will be needing the following data:

1. District data of Munich which we can find at: https://www.muenchen.de/int/en/living/postal-codes.html
2. Geographical coordinates of Munich and each of its neighborhoods
3. Venue data for neighborhoods in Munich

- Process

I intend to do the following:

1. Import all necessary libraries
2. Find Munich's Districts data and reading it in Jupyter
3. Creating a Dataframe that presents District, Latitude and Longitude for each Postal Code
4. Define a function to access Munich most common venues in Foursquare
5. Cluster Districts according to their most common venues categories
6. Visualize clusters in a map and analyze each cluster most common venues categories
7. Determine which clusters are a good viable option for opening a restaurant, and which are not

Importing dataset and creating Dataframe

- Import libraries and dataset

After importing every necessary library, we import our dataset using Pandas read_html function. The result is the following dataset:

	District	Postal Code
0	Allach-Untermenzing	80995, 80997, 80999, 81247, 81249
1	Altstadt-Lehel	80331, 80333, 80335, 80336, 80469, 80538, 80539
2	Au-Haidhausen	81541, 81543, 81667, 81669, 81671, 81675, 81677
3	Aubing-Lochhausen-Langwied	81243, 81245, 81249
4	Berg am Laim	81671, 81673, 81735, 81825

We now separate each Postal Code, having a single row for each one and indicating the corresponding District. We use a unique function to see that Munich presents 25 different Districts:

```
Munich presents 25 districts
```

	District	Postal Code
0	Allach-Untermenzing	80995
1	Allach-Untermenzing	80997
2	Allach-Untermenzing	80999
3	Allach-Untermenzing	81247
4	Allach-Untermenzing	81249

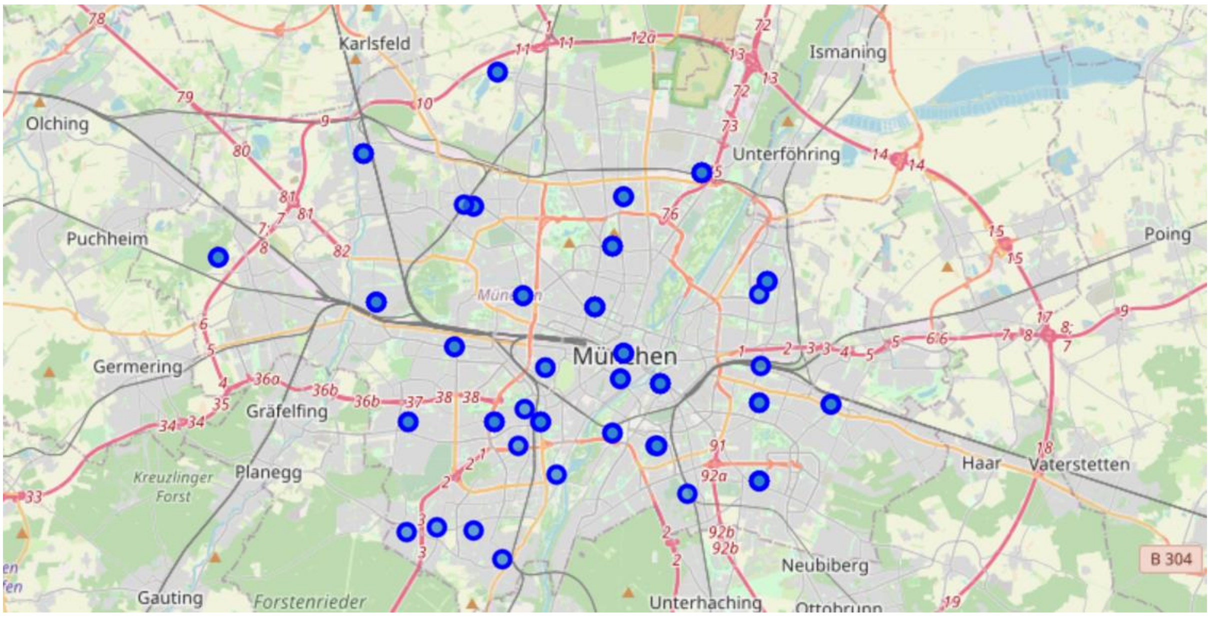
- Adding Latitude and Longitude

The following step is using the geolocator library to add the Latitude and Longitude of each Postal Code to our Dataframe:

	District	Postal Code	Latitude	Longitude
0	Allach-Untermenzing	80995	48.195157	11.462973
1	Allach-Untermenzing	80997	48.195157	11.462973
2	Allach-Untermenzing	80999	48.195157	11.462973
3	Allach-Untermenzing	81247	48.195157	11.462973
4	Allach-Untermenzing	81249	48.195157	11.462973

Visualizing Data

We are now ready to present the 25 unique Districts on a Munich map.



Foursquare

- Venues function

We define a function that brings Foursquare's venue category for each Postal Code, having a total of 3.387 venues categories in our Dataframe.

	District	Latitude	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	dm-drogerie markt	48.194118	11.465640	Drugstore
3	Allach-Untermenzing	48.195157	11.462973	Sicilia	48.193331	11.459387	Italian Restaurant
4	Allach-Untermenzing	48.195157	11.462973	Lidl	48.194428	11.465612	Supermarket

- Count

By using the count function, we determine the sum of venues categories of each District. This helps us identify which districts present various venues, indicating commercial activities separating them from more residential Districts whose venues are limited.

	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
District						
Allach-Untermenzing	40	40	40	40	40	40
Altstadt-Lehel	700	700	700	700	700	700
Au-Haidhausen	266	266	266	266	266	266
Berg am Laim	29	29	29	29	29	29
Bogenhausen	72	72	72	72	72	72
Feldmoching-Hasenbergl	6	6	6	6	6	6
Hadern	33	33	33	33	33	33
Laim	63	63	63	63	63	63
Ludwigsvorstadt-Isarvorstadt	400	400	400	400	400	400
Maxvorstadt	387	387	387	387	387	387
Milbertshofen-Am Hart	76	76	76	76	76	76
Moosach	132	132	132	132	132	132

- One hot encoding

Our current Dataframe has 182 unique venues. We use one hot encoding to determine how many of each venues appear in each Postal Code.

	District	ATM	Afghan Restaurant	American Restaurant	Arcade	Art Gallery	Art Museum	Asian Restaurant	Athletics & Sports	Austrian Restaurant	Auto Dealership	Automotive Shop	BBQ Joint	Bagel Shop	Bakery
0	Allach-Untermenzing	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1	Allach-Untermenzing	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Allach-Untermenzing	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Allach-Untermenzing	0	0	0	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	0	0	0

❖ Mean

Using the mean function, we determine the venues presence in each District.

	District	ATM	Afghan Restaurant	American Restaurant	Arcade	Art Gallery	Art Museum	Asian Restaurant	Athletics & Sports	Austrian Restaurant	Auto Dealership	Automotive Shop	BBQ Joint	Bagel Shop	Bake
0	Allach-Untermenzing	0.0	0.000000	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.125	0.0	0.0	0.1250
1	Altstadt-Lehel	0.0	0.000000	0.0	0.0	0.0	0.01	0.0	0.0	0.0	0.0	0.000	0.0	0.0	0.0000
2	Au-Haidhausen	0.0	0.026316	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.000	0.0	0.0	0.0000
3	Berg am Laim	0.0	0.000000	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.000	0.0	0.0	0.1034
4	Bogenhausen	0.0	0.000000	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.000	0.0	0.0	0.1388

❖ 10 most common venues in each District

It is time to create a new Dataframe showing the 10 most common venues of each District, in descending order.

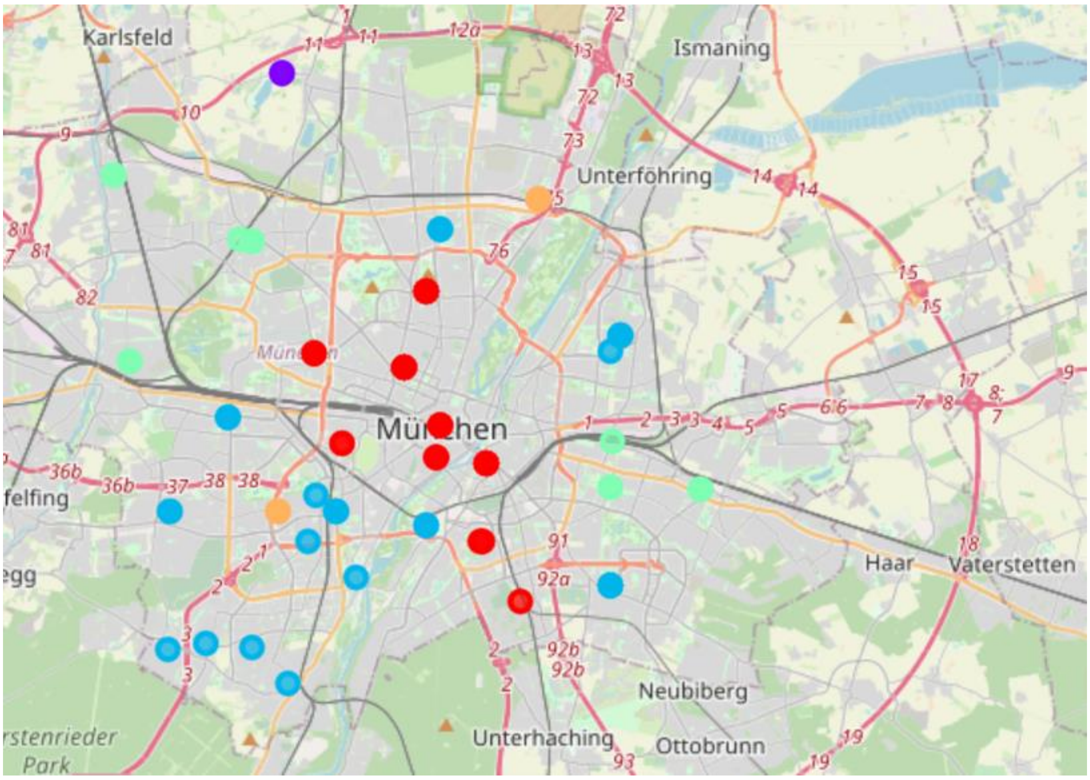
	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Allach-Untermenzing	Supermarket	Drugstore	Bakery	Sporting Goods Shop	Automotive Shop	Italian Restaurant	Yoga Studio	Fish Market	Fast Food Restaurant	Farmers Market
1	Altstadt-Lehel	Bavarian Restaurant	Café	Plaza	Hotel	German Restaurant	Restaurant	Coffee Shop	Cocktail Bar	Church	Clothing Store
2	Au-Haidhausen	Italian Restaurant	Concert Hall	Coffee Shop	French Restaurant	Thai Restaurant	Gourmet Shop	Doner Restaurant	Bistro	Movie Theater	Rock Club
3	Berg am Laim	Supermarket	Drugstore	Café	Metro Station	Hotel	Gastropub	Bakery	Cafeteria	Light Rail Station	Eastern European Restaurant
4	Bogenhausen	Bus Stop	Drugstore	Bakery	Italian Restaurant	Greek Restaurant	Park	Bank	Supermarket	Pharmacy	Water Park

Clustering

- We choose to classify Munich's Districts into 5 clusters and add the cluster number to the Dataframe.

	Cluster Labels	District	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	3	Allach-Untermenzing	Supermarket	Drugstore	Bakery	Sporting Goods Shop	Automotive Shop	Italian Restaurant	Yoga Studio	Fish Market	Fast Food Restaurant	Farmers Market
1	0	Altstadt-Lehel	Bavarian Restaurant	Café	Plaza	Hotel	German Restaurant	Restaurant	Coffee Shop	Cocktail Bar	Church	Clothing Store
2	0	Au-Haidhausen	Italian Restaurant	Concert Hall	Coffee Shop	French Restaurant	Thai Restaurant	Gourmet Shop	Doner Restaurant	Bistro	Movie Theater	Rock Club
3	3	Berg am Laim	Supermarket	Drugstore	Café	Metro Station	Hotel	Gastropub	Bakery	Cafeteria	Light Rail Station	Eastern European Restaurant
4	2	Bogenhausen	Bus Stop	Drugstore	Bakery	Italian Restaurant	Greek Restaurant	Park	Bank	Supermarket	Pharmacy	Water Park

- Visualizing the Districts as clusters on a map



- Most common venues of each cluster

Finally, lets see which venues are the most common in each cluster:

```python
# Cluster 0 Value Counts
cluster0['1st Most Common Venue'].value_counts()
```

```
Café                     15
Vietnamese Restaurant     8
Bavarian Restaurant       7
Italian Restaurant        7
Bakery                    5
Park                      4
```

```python
# Cluster 1 Value Counts
cluster1['1st Most Common Venue'].value_counts()
```

```
Motorcycle Shop    3
```

```python
# Cluster 2 Value Counts
cluster2['1st Most Common Venue'].value_counts()
```

```
Supermarket         22
German Restaurant   10
Bus Stop             6
```

```python
# Cluster 3 Value Counts
cluster3['1st Most Common Venue'].value_counts()
```

```
Supermarket    13
Spa             6
Bakery          5
```

```python
# Cluster 4 Value Counts
cluster4['1st Most Common Venue'].value_counts()
```

```
Fast Food Restaurant    8
Ice Cream Shop          5
```

Conclusion

By analyzing the five clusters we see that some of them are more suited for opening a restaurant than others.

- Cluster 0

This cluster is spread across Munich, including its centre and most tourist locations. As a result, there should be a strong demand, but rent would be higher than in other Districts.

Its most common venues include Cafes, different types of restaurants (Italian, Vietnamese, Asian, Stakehouse) as well as Cocktail Bars. On the one hand this Districts presents lots of potential customers but, on the other hand, competence is numerous and diverse. In order to operate in this cluster, we must offer a highly differentiated product in order to attract customers attention and convince them to try our restaurant.

- Cluster 1

Districts in this cluster are away from the centre of Munich and do not present restaurants as most common venues. Therefore, we ought to discard this cluster.

- Cluster 2

Also spreading across Munich but excepting the centre, these Districts present lots of supermarkets, German restaurants and bus stops. Looks like residential Districts (lots of supermarkets and bus stops). People in these Districts seem to prefer traditional food (German restaurants) over Fast Food, Italian, Asian, Falafel, Stakehouse, etc. So, it would be an interesting option to consider if our target are Munich's residents and not tourists.

- Cluster 3

This cluster represents a single District which is far away from the others. Although it has some food offers (Greek, Doner, Fast Food and Falafel restaurants), there aren't many. We can assume these Districts present more accessible prices to consumers, as well as lower costs for companies (rent for example). Might be a good opportunity if we do not have too much capital to invest.

- Cluster 4

In this case we have Fast Food as the most common venues. Considering their low prices and how difficult it is to compete with them, we ought to discard this cluster.