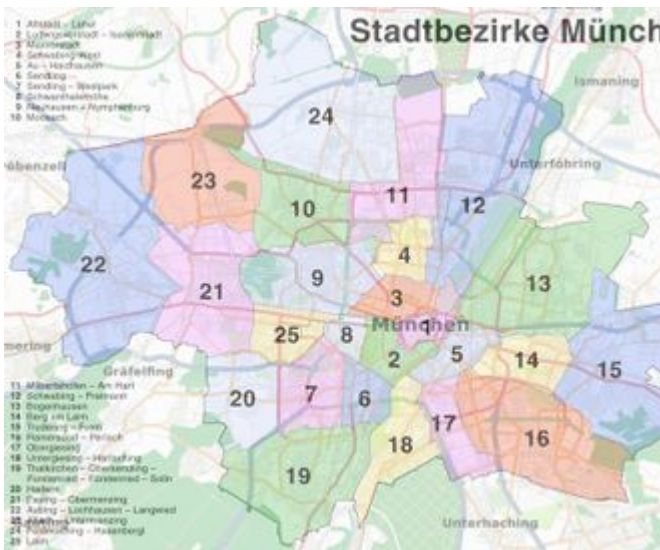


Munich Bakery Placement



Insights to Munich's neighborhoods inhabitants, location and venues with a special focus on bakeries and consideration of best strategic investment for bakery stores.

Table of contents

1. Introduction Section

Description of the problem & discussion of the background (who would be interested in this project)

2. Data Section

Description of the data that will be used to solve the problem and the source of the data

3. Methodology section

3.1 Getting the Data

3.2 Visualizing the data

3.3 Utilizing the Foursquare API to explore the neighborhoods venues and segment them

3.4 Inhabitants/bakery calculation

3.5 Clustering and evaluating the venue data

4. Results section

Discussion of results.

5. Discussion section

Discussion of any observations you noted and any recommendations you can make based on the results.

6. Conclusion section

Introduction

Client

My client is a large bakery chain.

Goal

They want to enter the market of southern Germanys capital Munich.

Background

As a bakery, they don't particularly think it is best to have the stores in the center of the city – they consider it more important to be close to people's homes as they found out in their market research that Germans like to get their fresh "Brötchen" from the bakery in the morning to eat them for breakfast at home. Additionally, competition plays a role of course: the client wants to build their stores in the neighborhood with the least competitors in the "bakery field". Finally, they also want to consider all other venue categories. Many other venues might be favorable, as people will have a motivation other than the bakery to come to the area and then conveniently also buy at the bakery but could also be a threat to the bakery (if the customer for example decided to go to a Cafe or Restaurant instead). Thus, they wish a segmentation of the neighborhoods overall venues to include this information in their final investment decision.

Task

The bakery chain wants me as a Data Scientist to find a neighborhood in Munich, Germany where there are many inhabitants but least bakeries/ #inhabitants and present a segmentation of the neighborhoods venues in general as additional decision support. All the information needs to be visualized in a management like manner.

Data

Data needed to examine the best neighborhood to build a bakery store in Munich

- **inhabitants per neighborhood of Munich** (source: html table at <https://suedbayerische-immobilien.de/Einwohner-Muenchen-Stadtteile>)
- **location (longitude & latitude) of Munich** (source: geopy library)
- **location (longitude & latitude) of Munich's neighborhoods centroids** (source: <https://www.gps-latitude-longitude.com/address-to-longitude-latitude-gps-coordinates> & <https://www.google.com/maps/>)
- **amount and location of bakeries in Munich - parsing them to the neighborhoods** (source: Foursquare)
- **amount and location of all other venues in Munich - parsing them to the neighborhoods** (source: Foursquare)
- **map data to visualize the venues and neighborhoods for management** (source: Folium, which is a great visualization library. Its possible to zoom into the maps and click on each circle mark to reveal the name of the respective neighborhood.)

All this data is available open source and without any cost for the client.

Methodology

Retrieving Munich's neighborhoods location and inhabitant's data

The inhabitant data of Munich's neighborhoods was scraped from an html table of a public website utilizing BeautifulSoup in Python. The geodata of each neighborhood's location was looked up on *google maps* and *gps-latitude-longitude.com* and manually stored in an excel sheet which was read and transformed to a pandas data frame. Finally, these two information sources were merged.

Visualizing the data

The geopy library was used to get the latitude and longitude values of the city Munich itself. Based on this data a map of Munich with the neighborhoods displayed on it was created to get a first overview of the distribution.

Retrieving Munich's neighborhoods venue data

The Foursquare API was utilized to get information about each neighborhood's venues (the venues names, location and category). A venue within a radius of 500m around a neighborhood's centroid was defined to be part of that neighborhood.

Analysis of the venues

Calculation of how many venues were returned for each neighborhood and how many unique venue categories exist, to decide if it makes sense to transfer their categorical values to numerical for clustering or if the data frame would become too high dimensional. There are 191 unique venue categories in the retrieved data. This was considered an acceptable amount for one-hot encoding, so the categorical data was changed to numerical with the goal of applying k-means clustering afterwards.

Calculation of inhabitant per bakery and neighborhood

The core of the store location evaluation is the comparison of inhabitants per neighborhood with the bakeries per neighborhood. Thus, first the bakery occurrences for each neighborhood were counted and this information merged with the inhabitants per neighborhood. The number of inhabitants divided by the number of bakeries per neighborhood creates an objective criterion showing where the biggest supply gap in bakery goods exists.

Displaying top venues

In order to gain more insights from the vast amount of venue information a table with the 10 most common venues per neighborhood was created. This was done by grouping rows by neighborhood and calculating the mean occurrence frequency for each category.

Clustering the venue data

After changing the categorical data of "venue category" to numerical for clustering, the clustering could be performed. Clustering was chosen because it is a good way to segment a collection of data points into smaller groups with similar attributes. This way the information is aggregated and can be displayed in a management ready format. While it would not be realistic for most cases to have a different marketing strategy for each data point, it is relatively common to have different approaches for each segment. Thus, it is very important to know how to differentiate your customers!

The neighborhoods were clustered based on their venue categories using the *k*-means algorithm with $k = 5$ clusters and the resulting clusters were visualized on a map.

Results

Insights from the bakery information

The table below shows the inhabitant per bakery for each neighborhood, including the neighborhoods that don't have a bakery (listed in Foresquare). The neighborhoods "Thalkirchen - Obersendling - Forstenried - F.", "Bogenhausen" and "Milbertshofen - Am Hart" all have over 70000 inhabitants per bakery.

	Neighborhood	NumberBakeries	Inhabitants	inh_per_bakery
18	3 Thalkirchen - Obersendling - Forstenried - F...	NaN	90.790	90.790000
12	4 Bogenhausen	1.0	82.138	82.138000
15	5 Milbertshofen - Am Hart	1.0	73.617	73.617000
19	7 Schwabing - Freimann	NaN	69.676	69.676000
20	9 Schwabing West	NaN	65.892	65.892000
21	11 Feldmoching - Hasenberg	NaN	59.391	59.391000
16	12 Sendling - Westpark	1.0	55.405	55.405000
11	1 Ramersdorf - Perlach	2.0	108.244	54.122000
17	14 Untergiesing - Harlaching	1.0	51.937	51.937000
22	17 Obergiesing - Fasanengarten	NaN	51.499	51.499000
14	19 Hadern	1.0	48.945	48.945000
5	2 Neuhausen - Nymphenburg	2.0	95.906	47.953000
23	21 Aubing - Lochhausen - Langwied	NaN	42.305	42.305000
10	6 Pasing - Obermenzing	2.0	70.783	35.391500
9	8 Trudering - Riem	2.0	67.009	33.504500
13	23 Allach - Untermenzing	1.0	30.737	30.737000
6	15 Maxvorstadt	2.0	51.642	25.821000
8	18 Ludwigsvorstadt - Isarvorstadt	2.0	50.620	25.310000
7	20 Berg am Laim	2.0	43.068	21.534000
24	25 Altstadt - Lehel	NaN	20.422	20.422000
3	10 Au - Haidhausen	3.0	59.752	19.917333
2	13 Laim	3.0	54.030	18.010000
1	16 Moosach	4.0	51.537	12.884250
4	24 Schwanthalerhöhe	3.0	29.663	9.887667
0	22 Sendling	5.0	39.953	7.990600

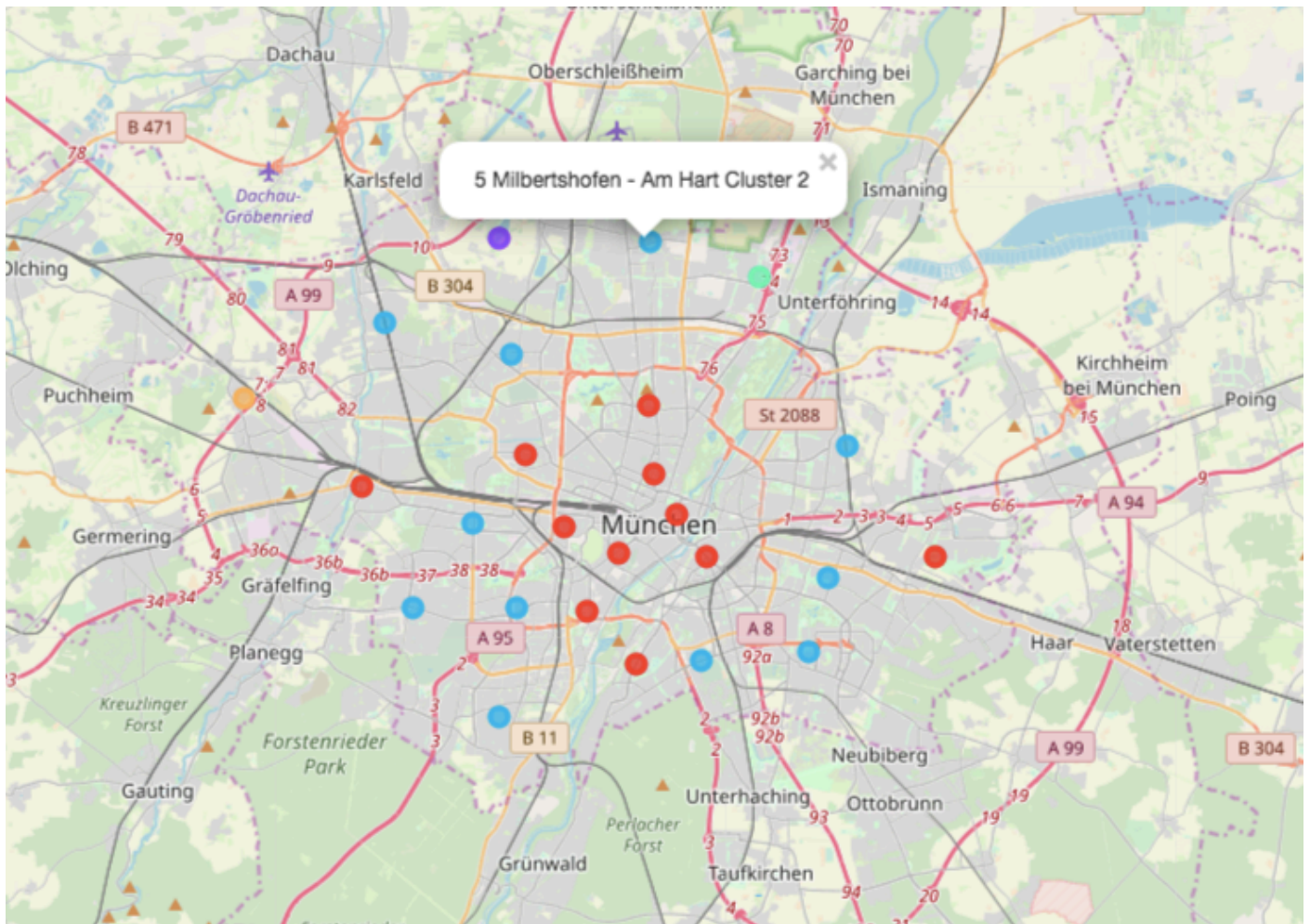
Table: Overview of Munich inhabitants per bakery in each neighborhood

Insights from the venue cluster information

The resulting clusters show, that the venue structure is different depending on the centricity of the neighborhood: the most central neighborhoods/best locations are therefore within one cluster (red), while the "second best" neighborhoods form a circle like cluster around it (blue). The 3 outer most neighborhoods are unlike the others in their venue representation and thus each form their own cluster.

How does this information help the client?

If the client decides to build a bakery store in one of the red/center neighborhoods they can have a look at the most common venues there and adjust their product portfolio accordingly. In the red neighborhoods for example are much more Cafes, Restaurant and Bars listed, then in the other clusters. The outer clusters on the other hand have supermarkets and transportation facilities listed under the most common venues. If the client decides to build a store in the red areas they should be aware that there is more distraction from other food venues and thus, the clients products should have a USP over these places. If the client decides to build more on the outer skirts of the city they could seek alliances with the supermarkets present there or adjust their products to the customers' needs who seems to spend some time traveling into the city. So, bakery goods that are easy to consume while traveling might be a good proposal there.



Picture: Munich neighborhoods cluster visualization of venues

Cluster Labels	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
2	1 Ramersdorf - Perlach	Supermarket	German Restaurant	Bakery	Hotel	Bus Stop	Italian Restaurant	Ice Cream Shop	Plaza	Bus Line	Market
0	10 Au - Haidhausen	Italian Restaurant	Café	German Restaurant	Plaza	French Restaurant	Bakery	Ice Cream Shop	Beach	Bar	Gourmet Shop
1	11 Feldmoching - Hasenberg	Lake	Zoo Exhibit	Event Service	Food & Drink Shop	Food	Flower Shop	Fish Market	Field	Fast Food Restaurant	Farmers Market
2	12 Sendling - Westpark	Bus Stop	Supermarket	Greek Restaurant	Tunnel	Ice Cream Shop	Brewery	Metro Station	Liquor Store	German Restaurant	Coffee Shop
2	13 Laim	Supermarket	Bakery	Bus Stop	Gastropub	Restaurant	Greek Restaurant	Tram Station	Bank	Plaza	Hotel
0	14 Untergiesing - Harlaching	Zoo Exhibit	Soccer Field	Tram Station	Sports Club	Supermarket	German Restaurant	Bus Stop	Café	Lawyer	Beer Garden
0	15 Maxvorstadt	Café	Bar	Italian Restaurant	Vietnamese Restaurant	Art Museum	Burger Joint	Mediterranean Restaurant	Plaza	Pizza Place	French Restaurant
2	16 Moosach	Supermarket	Bakery	Plaza	Tram Station	Drugstore	Hotel	Metro Station	Light Rail Station	Lawyer	Gastropub
2	17 Obergiesing - Fasanengarten	Hotel	Supermarket	Toy / Game Store	Office	Museum	Bus Stop	German Restaurant	Pie Shop	Gym	Ethiopian Restaurant
0	18 Ludwigsvorstadt - Isarvorstadt	Café	Italian Restaurant	German Restaurant	Burger Joint	Vietnamese Restaurant	Hotel	Ice Cream Shop	Asian Restaurant	Bar	Greek Restaurant
2	19 Haderm	Bus Stop	Hotel	Asian Restaurant	Sandwich Place	Garden	Supermarket	Sushi Restaurant	German Restaurant	Bank	Bakery
0	2 Neuhausen - Nymphenburg	Italian Restaurant	Supermarket	Park	Bakery	Restaurant	Shipping Store	Tram Station	Light Rail Station	Department Store	Flower Shop
2	20 Berg am Laim	Tram Station	Bakery	Metro Station	Hotel	Bus Stop	Pool	Drugstore	Café	Supermarket	Fast Food Restaurant
4	21 Aubing - Lochhausen - Langwied	Soccer Field	Dog Run	Bus Stop	Fast Food Restaurant	Event Space	Food & Drink Shop	Food	Flower Shop	Fish Market	Field
0	22 Sendling	Italian Restaurant	Bakery	German Restaurant	Café	Supermarket	Gastropub	Ice Cream Shop	Hotel	Bus Stop	Metro Station
2	23 Allach - Untermenzing	Supermarket	Sporting Goods Shop	Drugstore	Trattoria/Osteria	German Restaurant	Bakery	Food & Drink Shop	Gym	Hotel	Bus Stop
0	24 Schwanthalerhöhe	Café	Hotel	Italian Restaurant	Doner Restaurant	Bar	Pizza Place	Bakery	Asian Restaurant	Burger Joint	Ice Cream Shop
0	25 Altstadt - Lehel	German Restaurant	Boutique	Hotel	Plaza	Bavarian Restaurant	Café	Cocktail Bar	Italian Restaurant	Japanese Restaurant	Restaurant
2	3 Thalkirchen - Obersendling - Forstenried - F...	Bus Stop	Supermarket	Fast Food Restaurant	Gym / Fitness Center	Trattoria/Osteria	BBQ Joint	Beer Store	Lottery Retailer	Garden Center	Trail
2	4 Bogenhausen	Drugstore	Italian Restaurant	Café	Supermarket	Light Rail Station	Bakery	Zoo Exhibit	Falafel Restaurant	Food & Drink Shop	Food
2	5 Milbertshofen - Am Hart	Bus Stop	Supermarket	Metro Station	Pharmacy	Bakery	Shipping Store	Shoe Store	Shopping Mall	Lottery Retailer	Fried Chicken Joint
0	6 Pasing - Obermenzing	Drugstore	Café	Coffee Shop	German Restaurant	Supermarket	Italian Restaurant	Plaza	Organic Grocery	Bakery	Gym / Fitness Center
3	7 Schwabing - Freimann	Greek Restaurant	Outdoor Sculpture	Concert Hall	Pub	Rock Climbing Spot	German Restaurant	Electronics Store	Fish Market	Field	Fast Food Restaurant
0	8 Trudering - Riem	Clothing Store	Italian Restaurant	Bakery	German Restaurant	Plaza	Pizza Place	Furniture / Home Store	Coffee Shop	Electronics Store	Shoe Store
0	9 Schwabing West	Vietnamese Restaurant	Plaza	Italian Restaurant	Café	Bar	Supermarket	Greek Restaurant	Austrian Restaurant	Park	Drugstore

Table: Munich neighborhoods, their most common venues and cluster labels

Discussion

Foresquare Database

The venues were retrieved by foresquare which has a vast collection of venues worldwide but of course is not complete. Thus, the venues and calculations made in this exercise are only a rough direction for a decision but should not be considered to be very precise. Personally, living in Munich, I am sure there is more than one bakery in Bogenhausen. So, the above calculations are to be seen as an example of how to solve such a use case but for a real client another venue retrieving service should be taken instead or additional to Foresquare.

Neighborhood Area

One assumption of the calculation is that the neighborhoods of Munich are circular and all have the radius r . Of course, this is not the case but due to the difficulty in even retrieving the centroid geodata for each neighborhood it was not possible to display the neighborhoods geodata even closer to reality. But due to this assumption some venues of a neighborhood might haven't been considered because they were outside of the specified r around the centroid and other venues that belong to a neighboring neighborhood might have been included in more than one neighborhood if they lay within r of more than one centroid.

Clustering

The number k of Clusters is randomly chosen. This might not be the best segmentation of the neighborhoods.

Conclusion

In Conclusion to both calculation sections, I would recommend the client to build their stores in the neighborhoods "Thalkirchen - Obersendling - Forstenried - F.", "Bogenhausen" and "Milbertshofen - Am Hart" which all have over 70000 inhabitants per bakery.

Concerning the Clustering the results of these 3 neighborhoods: they are all in Cluster 2. Cluster 2 can be characterized by being the second circle (around the core of munich), with mainly supermarkets, hotels and public transport venues.

Thus, it seems that the best location for the bakery stores are the neighborhoods that lie around the core of munich, not the city center itself but also not the most desolate places outside. The competition is not too strong here and there's options to start alliances with the existing supermarkets and/or hotels in the areas.

By adjusting the product portfolio, a little to the needs of their customers in these areas I am optimistic that the investment will be worthwhile!