A new court in town

Guillermo García-Escribano

November 10, 2020

1. Introduction: Business Problem

A small Spanish investment fund has spotted a growing interest in racquet sports among Europeans. One of such sports, padel, currently most popular in Spain and Hispanic American countries, is now beginning to spread rapidly across Europe and other continents. Padel is easier to play than tennis, so it is easy to convince and engage potential players. However, this sport requires a special court to play and it is still hard to find one in most European cities.

Figure 1: A professional padel player (Source: WPT)

The managers of the fund have analysed different European cities and they have noticed that padel is becoming particulary popular in Munich, Germany, where the scarce infrastractures available for this sport seem to be always booked in spite of their high prices. Nevertheless, the managers have never played the sport and they have never been in Munich either. As a result, they ignore what kind of neighborhood would be fitting for a padel court or where in Munich they would find such a neighborhood. They suspect that good transport connections are essential. But what would padel players rather have in the proximity? Sports stores or other sports facilities? Supermarkets or restaurants? And once that is settled, which neighborhoods of Munich would be the best match for the padel court?

One of the requirements of the investment fund is for the court to be profitable in the short term, so expensive central areas should be avoided at all costs. Also, they want to minimize risks and spare costs by locating their new infrastructure in a neighborhood similar to those where an equivalent sports facility has proven to be successful but at the same time they want to keep clear of competitors in the area. This report aims to use data science methods to provide a handful of neighborhood candidates to the investment fund.

2. Data acquisition and cleaning

2.1 Data sources

Based on definition of our problem, factors that will influence our decision are:

- Location of Munich neighborhoods.
- Public venues in each neighborhood.
- Venue categories near sports facilities.

The city of Munich is divided into 25 boroughs ("Bezirke"), which in turn can be further broken down into 108 neighborhoods ("Bezirksteile"). The neighborhood is in this case the unit selected for exploration.

Following data sources were needed to extract/generate the required information:

- An online public list with the names of the different neighborhoods (available at www.opengov-muenchen.de).
- Centers of candidate neighborhoods were generated algorithmically and approximate addresses of centers of those areas will be obtained using GeoPy's Nominatim reverse geocoding.
- Number of public venues and their type and location in every neighborhood were obtained using Foursquare API.

2.2 Data cleaning

The required data concerning the names of the 108 neighborhoods of Munich was extracted from a CSV file in the open data portal of the city. All unnecessary data in that file was dropped and possible duplicates discarded.

Unlike other big cities, in Munich there is no relation one to one of neighborhoods and postal codes, and there is also no public data source assigning geographical coordinates to the center of the neighborhoods. For this reason, GeoPy's Nominatim geolocator was used to obtain latitude and longitude of each of the neighborhood centers by means of reverse geocoding. In exceptional cases, the geolocator was not able to find a neighborhood just by the name. These cases (only two) were considered negligible and although it is not ideal, due to the lack of alternative data sources they were dropped without a major impact on the analysis. The result of this process can be best visualized with a map marking each neighborhood center of Munich with a blue bubble (Figure 2).

Once the geodata of the neighborhoods is established, the Foursquare API was used to explore the available public venues in a 500m radius of each of the neighborhood centers. A total of 2925 venues for 106 neighborhoods were found, with 279 unique venue categories. An example of some of the venues found in the neighborhood of Graggenau is displayed here:

	Bezirksteil	Bezirksteil Latitude	Bezirksteil Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Graggenau	48.139563	11.580182	Bayerische Staatsoper	48.139639	11.578933	Opera House
1	Graggenau	48.139563	11.580182	Nationaltheater München	48.139599	11.579207	Opera House
2	Graggenau	48.139563	11.580182	Hotel Vier Jahreszeiten Kempinski	48.138918	11.581775	Hotel
3	Graggenau	48.139563	11.580182	Pusser's	48.138500	11.580426	Cocktail Bar
4	Graggenau	48.139563	11.580182	Louis Vuitton München Residenzpost	48.139248	11.577647	Boutique

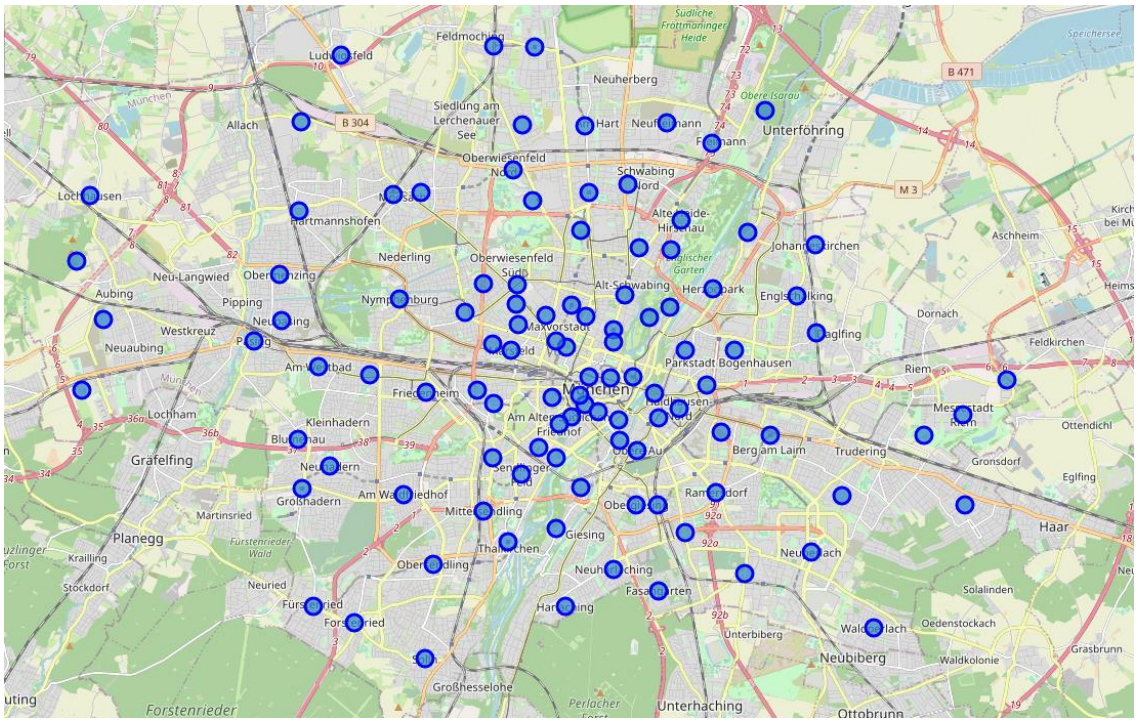

Figure 2: Munich neighborhoods

Once this information was retrieved, all data was set to start the analysis.

3. Methodology

In this project we tried to spot the neighborhoods in Munich that are most suitable for the construction of a new padel court. For that matter, we first found neighborhoods with sports facilities and studied their characteristics. Our new padel court should perform well in a similar environment.

In the first step we collected the required data: location of every neighborhood in Munich. We also identified all public venues in a 500m radius of their center (according to Foursquare).

Second step in our analysis was detecting downtown neighborhoods. These are deemed too expensive by the managers of the fund and their typical activity and structure are usually not compatible with the operation of a racquet sports court due to its size.

Thirdly, we tried to find neighborhoods with sports facilities that we can use as benchmarks for our analysis.

In the fourth and final step we used k-means clustering to identify similar neighborhoods that should give a hint regarding potential successful location areas and which should be a starting point for final 'street level' exploration and search for optimal venue location by the managers.

3. Exploratory analysis and predictive modeling

3.1 Most common venues

In order to be able to compare neighborhoods numerically, we needed to one hot encode the venue categories first. We calculated then the frequency of each venue category. With this information we produced a ranking of the ten most most common venues in each neighborhood. This helps us to understand easily what kind of neighborhoods we are dealing with. An example is given here, where we see that Alt Moosach is a dense neighborhood while Altaubing is almost empty:

	Bezirksteil	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	Alt Moosach	Bakery	Italian Restaurant	Asian Restaurant	Fast Food Restaurant	Burrito Place	Food	Light Rail Station	Drugstore	Supermarket	German Restaurant
1	Altaubing	Design Studio	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Altbogenhausen	Bakery	Italian Restaurant	Plaza	Indian Restaurant	Restaurant	Japanese Restaurant	Vietnamese Restaurant	Café	Sporting Goods Shop	NaN
3	Alte Heide - Hirschau	Soccer Field	Pastry Shop	Athletics & Sports	Bus Stop	Café	Stadium	Supermarket	Gastropub	Italian Restaurant	Track
4	Alte Kaserne	Austrian Restaurant	Asian Restaurant	Post Office	Italian Restaurant	German Restaurant	Bar	Drugstore	Bakery	Greek Restaurant	Indian Restaurant

3.2 K-means clustering

Downtown neighborhoods usually have similar characteristics, for instance large numbers of cafés and restaurants. We could expect that a k-means clustering algorithm would possibly find such a pattern. If this happens, we can eliminate these neighborhoods from our pool of candidates right away. We tested different k parameters and obtained similar results: neighborhoods in the center area show indeed a pattern. Clusters are shown in different colors in Figure 3.

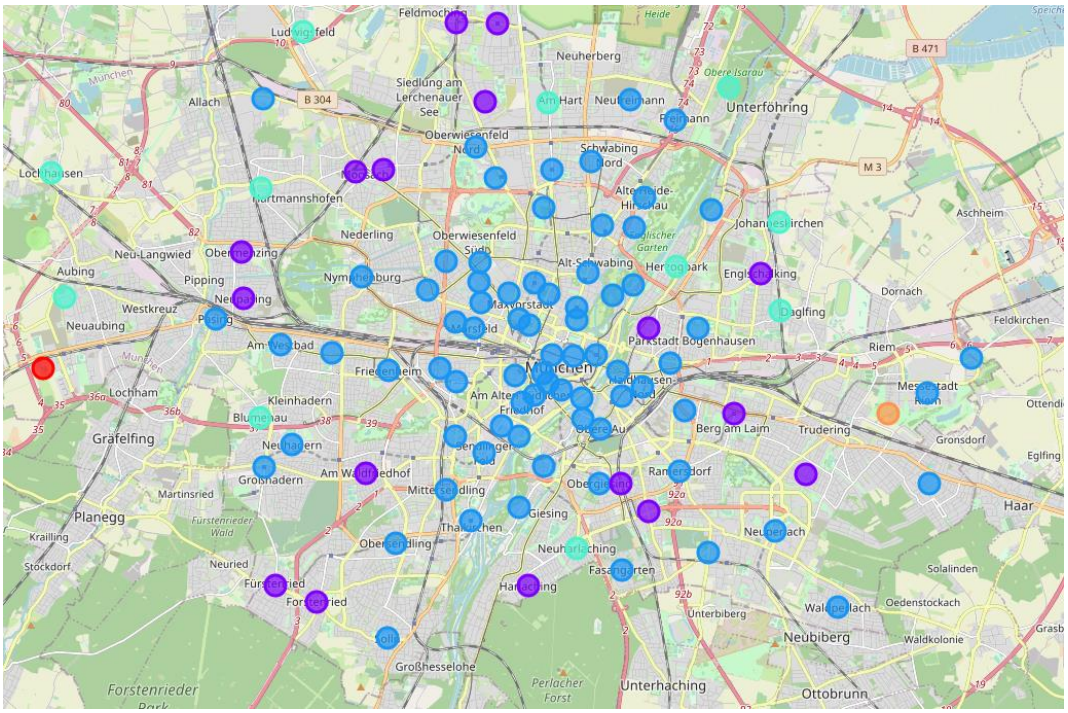

Figure 3: Downtown cluster in blue

After discarding this cluster of neighborhoods, only 71 unique venue categories remain. Fortunately, one of them is “Tennis Court”. This is extremely convenient, since our padel court

will have a similar type of users. We tried to identify if neighborhoods with this venue category present similar characteristics:

	Bezirksteil	Bezirksteil Latitude	Bezirksteil Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Cluster Labels
2629	Harlaching	48.088293	11.565078	Tennis Kail Harlaching	48.088041	11.569694	Tennis Court	1
2823	Obermenzing	48.162570	11.469372	Tennis-Club Blutenburg e.V.	48.162969	11.467291	Tennis Court	1
2881	Lerchenau West	48.196093	11.550703	MSC Hockeyclub	48.197683	11.556074	Tennis Court	1

Since all neighborhoods are in the same cluster, we can deduce that there could be a similar set of venues in them. We see for instance that in the proximity of tennis courts, users can find bus stops and small drink and food stores like bakeries and cafés. This feature could be essential to attract new users to our padel courts. Just to draw a comparison, we could see that a neighborhood with a different sports venue (a soccer field) got assigned to a different cluster. Its neighborhood is much more focused on transports, a gym and a restaurant.

Since the tennis court cluster had way too many candidates, we kept exploring the tennis court cluster and further divided it to obtain the most similar neighborhoods:

	Bezirksteil	Cluster Labels	Latitude	Longitude	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	Alt Moosach	1	48.181071	11.516728	Bakery	Italian Restaurant	Asian Restaurant	Fast Food Restaurant	Burrito Place	Food	Light Rail Station	Drugstore	Supermarket	German Restaurant
1	Altbogenhausen	8	48.145785	11.605498	Bakery	Italian Restaurant	Plaza	Indian Restaurant	Restaurant	Japanese Restaurant	Vietnamese Restaurant	Café	Sporting Goods Shop	NaN
2	Am Waldfriedhof	4	48.113498	11.510758	Bus Stop	Bakery	Plaza	Rental Car Location	Supermarket	Asian Restaurant	NaN	NaN	NaN	NaN
3	Balanstraße-West	1	48.104872	11.605344	Drugstore	Bakery	Supermarket	Lounge	IT Services	Asian Restaurant	Café	Office	Bus Stop	Museum
4	Englschalking	0	48.157923	11.642630	Bakery	Bus Stop	Drugstore	Pharmacy	Supermarket	Park	Italian Restaurant	Greek Restaurant	Light Rail Station	NaN
5	Feldmoching	0	48.213804	11.541275	Ice Cream Shop	Supermarket	Bakery	Bus Stop	Greek Restaurant	Farmers Market	Taverna	Bus Line	Lottery Retailer	Plaza
6	Forstenried	3	48.084835	11.494477	Bakery	Greek Restaurant	BBQ Joint	Ramen Restaurant	NaN	NaN	NaN	NaN	NaN	NaN
7	Fürstenried-West	0	48.088366	11.480830	Bakery	Bus Stop	Plaza	Market	Supermarket	Drugstore	Farmers Market	Pool	Metro Station	Pharmacy
8	Gartenstadt Trudering	2	48.113139	11.657732	Bakery	Bus Stop	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	Giesing	1	48.111130	11.596084	Bakery	Park	Gym / Fitness Center	Shipping Store	Burger Joint	Drugstore	Supermarket	Greek Restaurant	Ice Cream Shop	Hookah Bar
10	Harlaching	5	48.088293	11.565078	Bakery	Café	Plaza	Tennis Court	NaN	NaN	NaN	NaN	NaN	NaN
11	Hasenbergl-Lerchenau Ost	6	48.213429	11.554914	Supermarket	Plaza	Bakery	Gas Station	Clothing Store	NaN	NaN	NaN	NaN	NaN
12	Josephsburg	1	48.126573	11.633809	Bakery	Supermarket	Asian Restaurant	Drugstore	Café	Bavarian Restaurant	Hotel	Metro Station	Italian Restaurant	Gastropub
13	Lerchenau West	7	48.196093	11.550703	Food & Drink Shop	Trattoria/Osteria	Bus Line	Bus Stop	Café	Bakery	Tennis Court	NaN	NaN	NaN
14	Moosach-Bahnhof	1	48.180399	11.507340	Bakery	Drugstore	Supermarket	Hotel	Bus Stop	Big Box Store	Gastropub	Sandwich Place	Asian Restaurant	Hardware Store
15	Neupasing	9	48.152271	11.469903	Spa	Bakery	Doner Restaurant	Italian Restaurant	Supermarket	Restaurant	NaN	NaN	NaN	NaN
16	Obermenzing	0	48.162570	11.469372	Bakery	Bus Stop	Drugstore	Discount Store	Italian Restaurant	Supermarket	Tennis Court	Hostel	NaN	NaN

The tennis court neighborhood of Obermenzig is the one that presents the highest number of similar neighborhoods, three. In the map in Figure 4 they appear in red. Discarding the Obermenzig neighborhood to avoid competition, we are left with three potential candidates in that cluster: Englschalking, Feldmoching and Fürstenried-West. If we study the cluster, we see that all neighborhoods have a bakery, a bus stop and a supermarket. They all happen to be neighborhoods in the outskirts of the city with good enough transport connections and available land space.

	Bezirksteil	Cluster Labels	Latitude	Longitude	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
4	Englschalking	0	48.157923	11.642630	Bakery	Bus Stop	Drugstore	Pharmacy	Supermarket	Park	Italian Restaurant	Greek Restaurant	Light Rail Station	NaN
5	Feldmoching	0	48.213804	11.541275	Ice Cream Shop	Supermarket	Bakery	Bus Stop	Greek Restaurant	Farmers Market	Taverna	Bus Line	Lottery Retailer	Plaza
7	Fürstenried-West	0	48.088366	11.480830	Bakery	Bus Stop	Plaza	Market	Supermarket	Drugstore	Farmers Market	Pool	Metro Station	Pharmacy
16	Obermenzing	0	48.162570	11.469372	Bakery	Bus Stop	Drugstore	Discount Store	Italian Restaurant	Supermarket	Tennis Court	Hostel	NaN	NaN

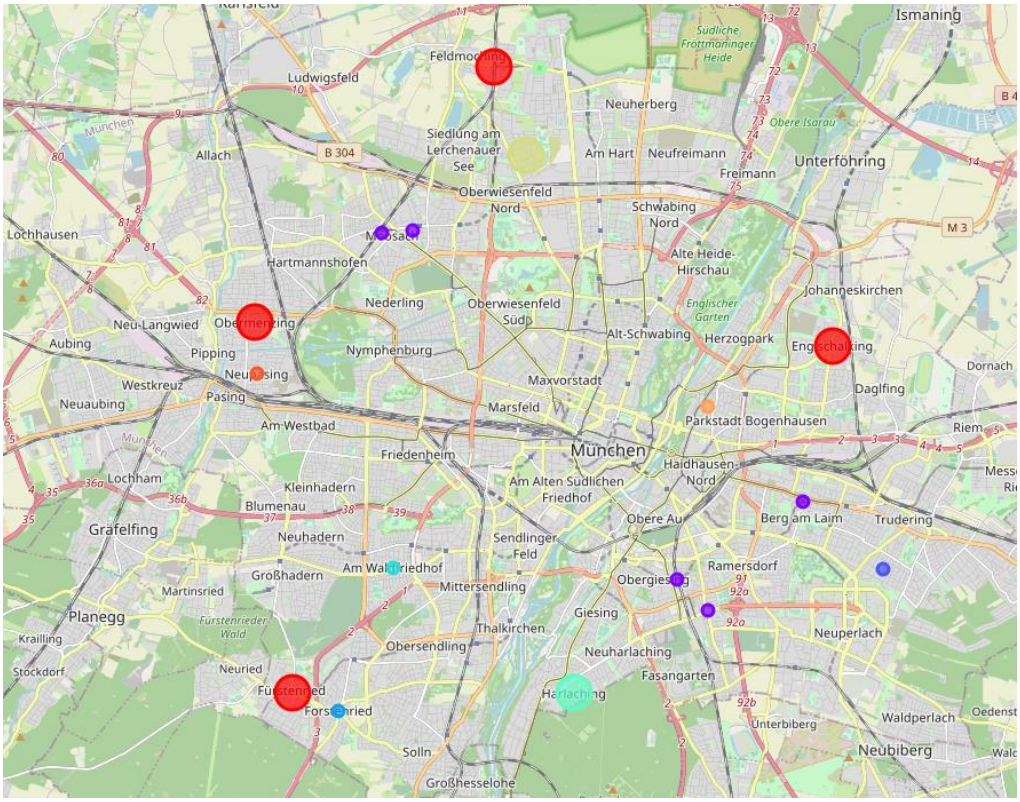

Figure 4: Obermenzing cluster in red

4. Results and discussion

Our analysis shows that there is a large number of neighborhoods (14) that present similarities to those that already have tennis courts. They are neighborhoods out of the city centre with good transport connection and food stores. Most of them do not even reach 10 venues in Foursquare, so the venue density is low.

In order to make the evaluation of candidates easier for the investment fund, the clustering was refined and as a result we found a pool of three candidates (Englschalking, Feldmoching and Fürstenried-West) with a high similarity to Obermenzig, the tennis court neighborhood with the most venue categories. This could make the location more attractive for our users.

We recommend starting the evaluation of potential locations in these three neighborhoods, as they seem the most promising to host a padel court. This, of course, does not imply that those zones are actually optimal locations. The purpose of this analysis was to only provide info on areas out of the city center but attractive enough for potential users (due to transport, shops) - it is entirely possible that there is a very good reason that the suggested neighborhoods have no tennis court yet, reasons which would make them unsuitable for a tennis or padel court

regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually lead to evaluating other factors and conditions or exploring other neighborhoods.

5. Conclusions and future directions

Purpose of this project was to identify Munich areas out of the city center with similar conditions to already existing neighborhoods with equivalent sports facilities for optimal location for a new padel court. By collecting public venue data from Foursquare and using it in a k-means clustering algorithm we have first identified a common pattern for downtown neighborhoods which needed to be eliminated from the candidate pool due to project requirements. The same clustering of neighborhoods showed that all tennis courts (similar sport to padel) were located in neighborhoods with alike characteristics. In order to reduce the pool of candidates to be evaluated by the managers, the clustering was refined and three possible locations were suggested: Englschalking, Feldmoching and Fürstenried-West. They are all neighborhoods in the outskirts of the city, well connected by bus and with bakeries and supermarkets where padel court users can casually shop before or after playing.

The final decision on optimal location will be made by the managers after a further evaluation of the recommended areas, taking into consideration additional factors like real estate availability, prices, social and economic dynamics of every neighborhood, attractiveness of the area, etc.