# *Capstone Project: Where to locate a new Japanese restaurant in Munich?*

Fei Zhao

22/07/2020

# Content

- ❑ *Background*

- ❑ *Project Objective*

- ❑ *Data Acquisition and Cleaning*

- ❑ *Exploratory Analysis*

- ❑ *Methodology: Clustering*

- ❑ *Result*

- ❑ *Conclusion*

# Background

- The size of restaurant industry market is continuously increasing

- Jack, a well-trained Japanese chef, wants to start his own restaurant business at Munich

**Why restaurant business?**

- Eating out is a popular pastime in Germany

- 51.3 million people sometimes doing so

- 11.9 million people stating they often did so

# Project objective

- Help Jack find a location to open his Japanese restaurant in Munich



**About Munich**

- 11th-largest city in the European Union

- 1.5 million population

- more than 530,000 people of foreign background

- 25 districts

**Question: where to locate his new restaurant?**

# Data acquisition and cleaning I

There are three important factors for the location decision making.
We need to scrape the required data from the website:

| Impact Factors | Data |
|---|---|
| Population base | District data [1] |
| Foreigner population ratio | Foreigner ratio data[2] |
| Venues categories within each districts | Venues geographical data[3] |

**Data source:**
*1. District data*
https://www.muenchen.de/int/en/living/postal-codes.html
https://www.all-zipcodes.com/postcode-munchen-in-germany
*2. Foreigner data*
http://www.total-munich.com/20160623888/blog/moving-to-munich/moving-to-munich-introduction-to-munich-s-boroughs.html
*3. Venues data*
https://api.foursquare.com

**Packages:**

- BeautifulSoup

- Folium

- Requests

# Data acquisition and cleaning II

After wrangling data, these two tables will be used for later analysis.

I. *Munich_df*

| | District | latitude | longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 347 | Bogenhausen | 48.1391052689 | 11.6027305035 | Nomiya | 48.131653 | 11.598526 | Sushi Restaurant |
| 359 | Bogenhausen | 48.1391052689 | 11.6027305035 | Sushi Bento | 48.139390 | 11.592329 | Sushi Restaurant |
| 434 | Laim | 48.1321268962 | 11.5122375474 | Umeshu Sushibar | 48.137233 | 11.523225 | Sushi Restaurant |
| 467 | Laim | 48.1321268962 | 11.5122375474 | Ozeansushi | 48.128715 | 11.501577 | Sushi Restaurant |
| 497 | Ludwigsvorstadt-Isarvorstadt | 48.1455583623 | 11.5539521328 | Kaito | 48.149599 | 11.560872 | Sushi Restaurant |

II. *Munich_venues*

| | District | Surface (km²) | Population | Density (Persons/km²) | Foreigners (%) | latitude | longitude | Postal_Code |
|---|---|---|---|---|---|---|---|---|
| 0 | Allach-Untermenzing | 15.45 | 31,353 | 2,029 | 19.5 | 48.217157311 | 11.5149477676 | 80995 |
| 1 | Altstadt-Lehel | 3.15 | 20,806 | 6,614 | 25.8 | 48.1359642378 | 11.5729048503 | 80331 |
| 2 | Au-Haidhausen | 4.22 | 60,756 | 14,399 | 23.4 | 48.1204341448 | 11.5874447961 | 81541 |
| 3 | Aubing-Lochhausen-Langwied | 34.06 | 42,859 | 1,258 | 23.3 | 48.1452756567 | 11.4368110446 | 81243 |
| 4 | Berg am Laim | 6.31 | 44,002 | 6,971 | 29.9 | 48.1219040581 | 11.6182992419 | 81671 |

# Exploratory data analysis I

Analyze the Munich population data, find the Top 10 Density (Persons/km²) District, and put it into a dataframe named: *Density_District_candidate*

```
0                    Schwabing-West
1                   Schwanthalerhöhe
2                      Au-Haidhausen
3                        Maxvorstadt
4      Ludwigsvorstadt-Isarvorstadt
5                               Laim
6                            Sendling
7           Obergiesing-Fasangarten
8             Neuhausen-Nymphenburg
9                 Sendling-Westpark
Name: District, dtype: object
```

*Dataframe: Density_District_candidate*
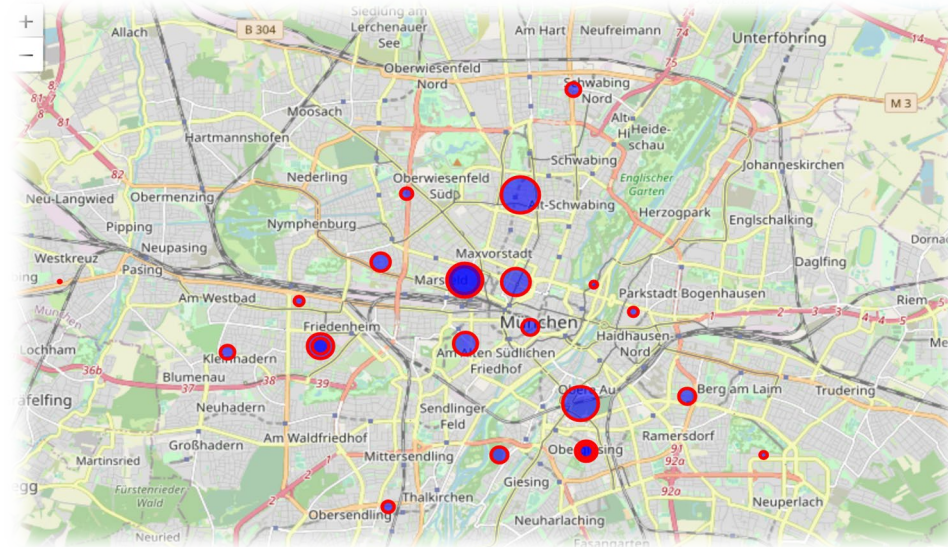


Since each restaurant has its limited radius of service, instead of exploring the total population at each district, we pay more attention to the **population density**. Higher population density means that the restaurant can cover more potential customers within its radius of service.

# Exploratory data analysis II

Analyze the Munich foreigner's ratio data, find the Top 10 Foreigners(%) District, and put it into a dataframe named: *Foreigners_District_candidate*

```
0                    Schwabing-West
1                    Schwanthalerhöhe
2                    Au-Haidhausen
3                    Maxvorstadt
4       Ludwigsvorstadt-Isarvorstadt
5                    Laim
6                    Sendling
7             Obergiesing-Fasangarten
8             Neuhausen-Nymphenburg
9                    Sendling-Westpark
Name: District, dtype: object
```

*Dataframe: Foreigners_District_candidate*



German Food vs Japanese Food

Since Japanese food is an international food for Munich locals, a more diversity district could be more friendly for

an external food culture. Thus, we assume that the higher ratio of foreigner is, the better the location is for Jack's

new restaurant.

# Exploratory data analysis III

Analyze the restaurant industry and Japanese restaurant in the Munich
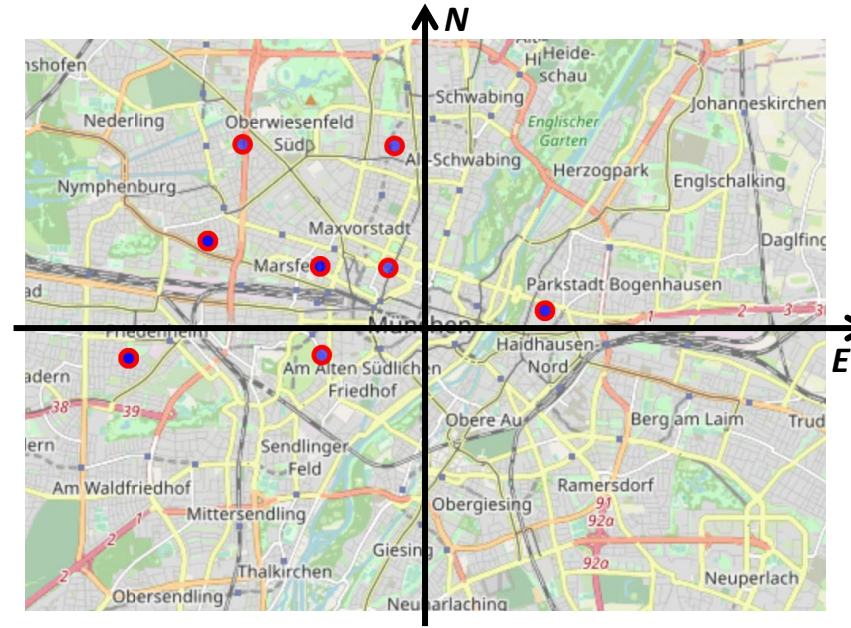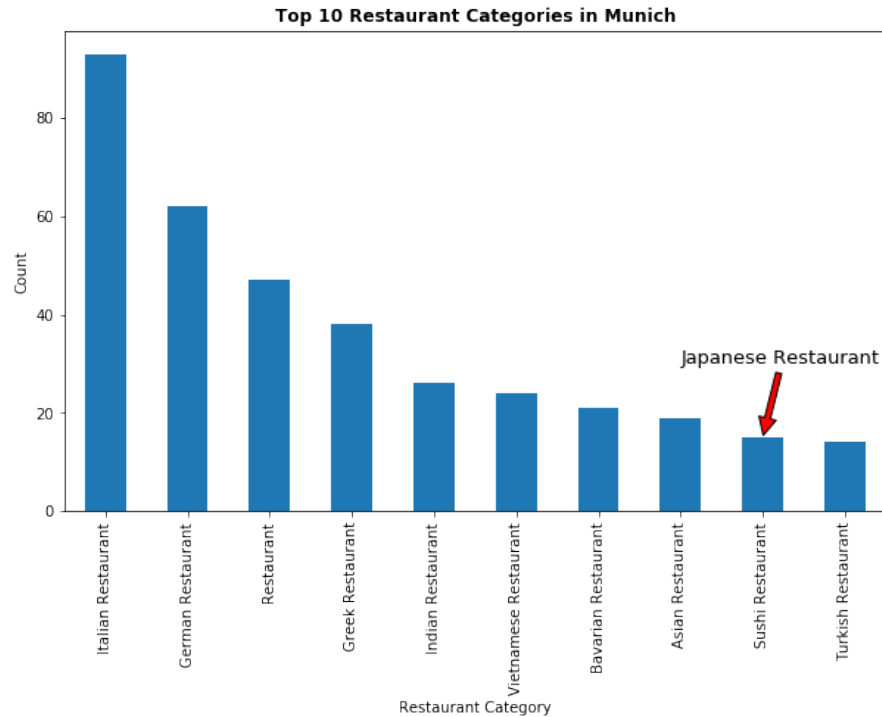




*Figure: The location of existing Japanese restaurant*

- Even though, food industry is competitive in Munich, the Japanese restaurant still has a good growth potential because of its low market share ratio.

- The majority of Japanese restaurant are located at the *northwest* side of Munich.

# Methodology: clustering I

Use clustering method to figure out the pattern of location at which Japanese restaurants are located

- **Package**: sklearn.cluster

- **Data wrangling**:

  1) generate the dummy variable

  2) group the data by its venue's types

  3) Calculate the mean frequency each venue category at each district

| | District | Afghan Restaurant | American Restaurant | Argentinian Restaurant | Art Gallery | Art Museum | Arts & Entertainment | Asian Restaurant | Athletics & Sports | Austrian Restaurant | ... | Trattoria/Osteria | Tunnel | Re |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Allach-Untermenzing | 0.0 | 0.0 | 0.00 | 0.00 | 0.00 | 0.0 | 0.000000 | 0.0 | 0.00 | ... | 0.000000 | 0.000000 | |
| 1 | Altstadt-Lehel | 0.0 | 0.0 | 0.01 | 0.00 | 0.01 | 0.0 | 0.000000 | 0.0 | 0.00 | ... | 0.000000 | 0.000000 | |
| 2 | Au-Haidhausen | 0.0 | 0.0 | 0.00 | 0.00 | 0.00 | 0.0 | 0.010309 | 0.0 | 0.00 | ... | 0.020619 | 0.000000 | |
| 3 | Aubing-Lochhausen-Langwied | 0.0 | 0.0 | 0.00 | 0.00 | 0.00 | 0.0 | 0.068966 | 0.0 | 0.00 | ... | 0.000000 | 0.034483 | |
| 4 | Berg am Laim | 0.0 | 0.0 | 0.00 | 0.02 | 0.00 | 0.0 | 0.060000 | 0.0 | 0.02 | ... | 0.000000 | 0.000000 | |

- **Implement clustering**: using K-means clustering method and set K=3

# Methodology: clustering II

Base on clustering result, there are three patterns of districts where Japanese restaurants are located

**Pattern I**

- Districts have plenty of restaurants

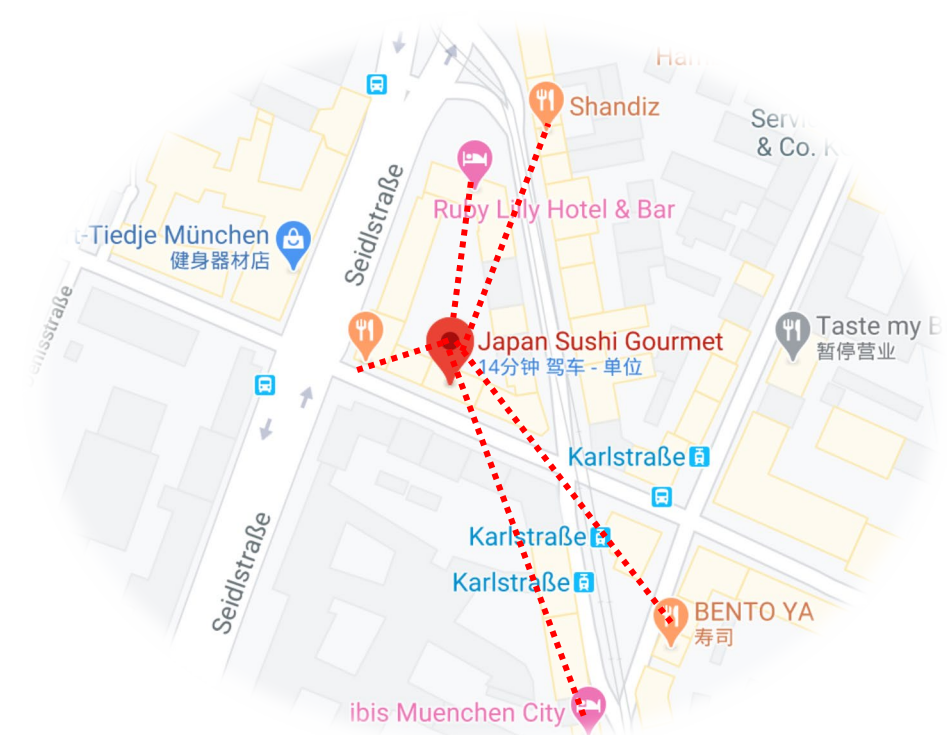| | District | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue | Cluster |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Bogenhausen | Italian Restaurant | Café | Hotel | German Restaurant | Restaurant | Plaza | Bar | Bakery | French Restaurant | Ice Cream Shop | 0 |
| 3 | Moosach | German Restaurant | Italian Restaurant | Café | Bakery | Plaza | Trattoria/Osteria | Gastropub | Restaurant | Concert Hall | Middle Eastern Restaurant | 0 |
| 4 | Neuhausen-Nymphenburg | German Restaurant | Italian Restaurant | Indian Restaurant | Café | Supermarket | Plaza | Sushi Restaurant | Vietnamese Restaurant | Bakery | Hotel | 0 |
| 5 | Schwabing-West | Italian Restaurant | Café | Vietnamese Restaurant | Trattoria/Osteria | Plaza | Greek Restaurant | Pizza Place | Gastropub | Thai Restaurant | Bar | 0 |

**Pattern II**

- Districts have a lot supermarkets and public transportation station

| | 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 | Cluster |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Laim | Supermarket | Bus Stop | Tram Station | Laundromat | Bank | Hotel | Pizza Place | Sushi Restaurant | Sandwich Place | Fast Food Restaurant | 1 |
| 8 | Sendling-Westpark | Supermarket | Bus Stop | Tram Station | Laundromat | Bank | Hotel | Pizza Place | Sushi Restaurant | Sandwich Place | Fast Food Restaurant | 1 |

**Pattern III**

- Districts have a lot hotels and cafe

| | 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 | Cluster |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | Ludwigsvorstadt-Isarvorstadt | Hotel | Café | Italian Restaurant | Vietnamese Restaurant | Middle Eastern Restaurant | German Restaurant | Steakhouse | Burger Joint | Restaurant | Drugstore | 2 |
| 6 | Schwanthalerhöhe | Hotel | Café | Italian Restaurant | Vietnamese Restaurant | Middle Eastern Restaurant | German Restaurant | Steakhouse | Burger Joint | Restaurant | Drugstore | 2 |
| 7 | Sendling | Hotel | Café | German Restaurant | Bavarian Restaurant | Vietnamese Restaurant | Burger Joint | Italian Restaurant | Middle Eastern Restaurant | Asian Restaurant | Hotel Bar | 2 |



*Then, **clustering all districts** to find the district without Japanese restaurant that is highly similar to the district with Japanese restaurant in term of venues and population characteristics.*

# Result

- In the intersection set of Density_District_candidate and Foreigners_District_candidate,

  *"Obergiesing-Fasangarten"* district is the only one ranks in top 10 in both criteria

```
1   district_venues_sorted[district_venues_sorted['District']=='Obergiesing-Fasangarten']
```

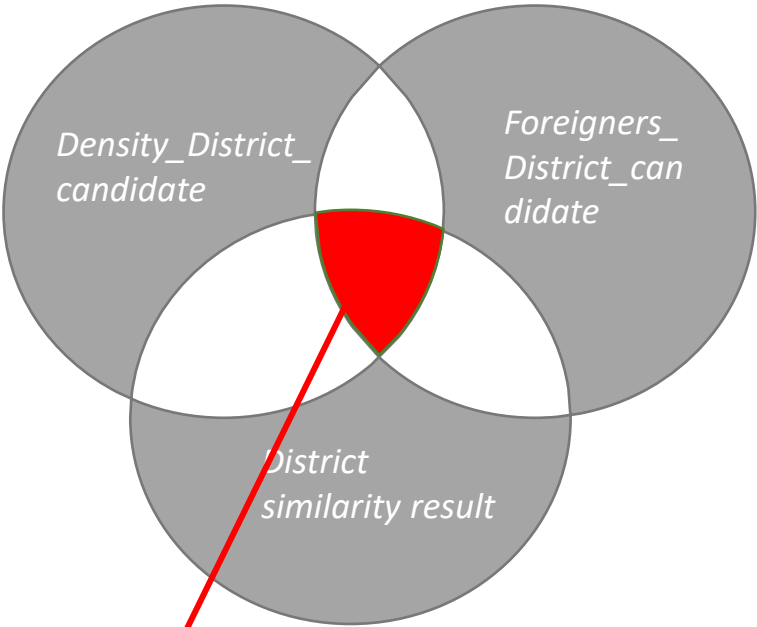| | District | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 14 | Obergiesing-Fasangarten | Supermarket | Hotel | Café | Greek Restaurant | Ice Cream Shop | Italian Restaurant | Bar | German Restaurant | Drugstore | Plaza |

- Compared with location patterns in previous slide, we find that this candidate district

  shares a lot characteristics with the pattern II and pattern III.

| | 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 | Cluster |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Laim | Supermarket | Bus Stop | Tram Station | Laundromat | Bank | Hotel | Pizza Place | Sushi Restaurant | Sandwich Place | Fast Food Restaurant | 1 |
| 8 | Sendling-Westpark | Supermarket | Bus Stop | Tram Station | Laundromat | Bank | Hotel | Pizza Place | Sushi Restaurant | Sandwich Place | Fast Food Restaurant | 1 |

*Pattern II*

| | 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 | Cluster |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | Ludwigsvorstadt-Isarvorstadt | Hotel | Café | Italian Restaurant | Vietnamese Restaurant | Middle Eastern Restaurant | German Restaurant | Steakhouse | Burger Joint | Restaurant | Drugstore | 2 |
| 6 | Schwanthalerhöhe | Hotel | Café | Italian Restaurant | Vietnamese Restaurant | Middle Eastern Restaurant | German Restaurant | Steakhouse | Burger Joint | Restaurant | Drugstore | 2 |
| 7 | Sendling | Hotel | Café | German Restaurant | Bavarian Restaurant | Vietnamese Restaurant | Burger Joint | Italian Restaurant | Middle Eastern Restaurant | Asian Restaurant | Hotel Bar | 2 |

*Pattern III*



*Density_District_candidate*

*Foreigners_District_candidate*

*District similarity result*

"Obergiesing-Fasangarten" district

# Conclusion

- **According to above analysis, we recommend Jack to locate his new restaurant at "Obergiesing-Fasangarten" district.**

    - ✓ 30.1% foreigner population

    - ✓ The 7th population density district in Munich

    - ✓ Without any Japanese restauranut

    - ✓ Similar pattern with other districts that have successfully operated Japanese restaurant

    - ✓ Long distance from existing Japanese restaurants

- **Further application of analysis**

    - ❑ This analytical approach can also be used for select the location for new bars or other business.

- **Limitation**
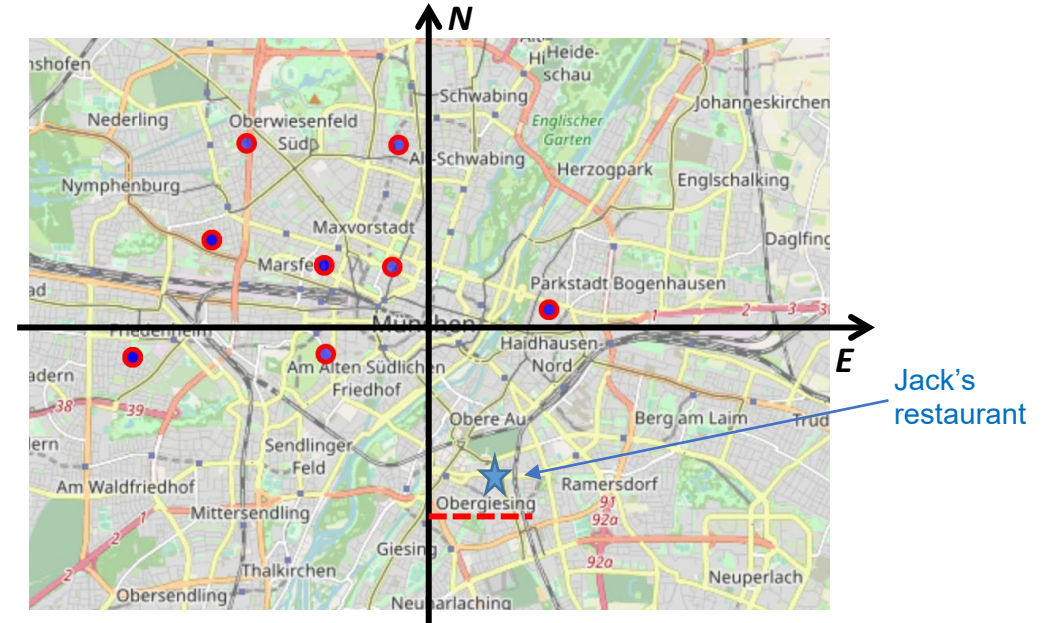    - ❑ Only the range like district is given in our advice instead of a more specific address



*Figure: The location of existing Japanese restaurant*