

Segmenting and Clustering Neighborhoods in Munich

March 19, 2021

Tayana Nazareth Langer

1 Introduction

Munich is the capital of Bavaria and the third-largest city in Germany, with a population of 1,558,395 inhabitants in 2020 [1]. Due to its high quality of life, its economic and cultural importance, Munich is a city that attracts investors, workers, families, and students, from Germany and abroad. It has a heated real estate market with high prices.

Our client is a real estate agency that operates in the city of Munich and seeks an effective tool to help its clients find the property they are looking for. The desired product is a model that classifies and maps the areas of the city according to the existing similarities. The classes found will help realtors to focus on the areas that best fit the profile of clients, Based on the service preferences that customers want in their future neighborhood.

2 Data and Methodology

Munich district data postal codes were taken from web page: <https://www.muenchen.de/int/en/living/postal-codes.html> , and RESTful API calls were made to Foursquare API to retrieve the information about the venues. The Foursquare's data was used to find out the types of places in the neighborhood and the frequency with which they were visited. Data preparation was done using Pandas library.

This project used a machine learning model to classify and map similar areas of the city. Specifically, the K-Means Clustering model. Clustering is a technique to divide data into different groups, where the records in each group are similar to one another. So, the goal of clustering is to identify significant and meaningful groups of data. K-means divides the data into K clusters by minimizing the sum of the squared distances of each record to the mean of its assigned cluster [2].

2.1 Data preparation

The following steps were taken to prepare the data for the analysis:

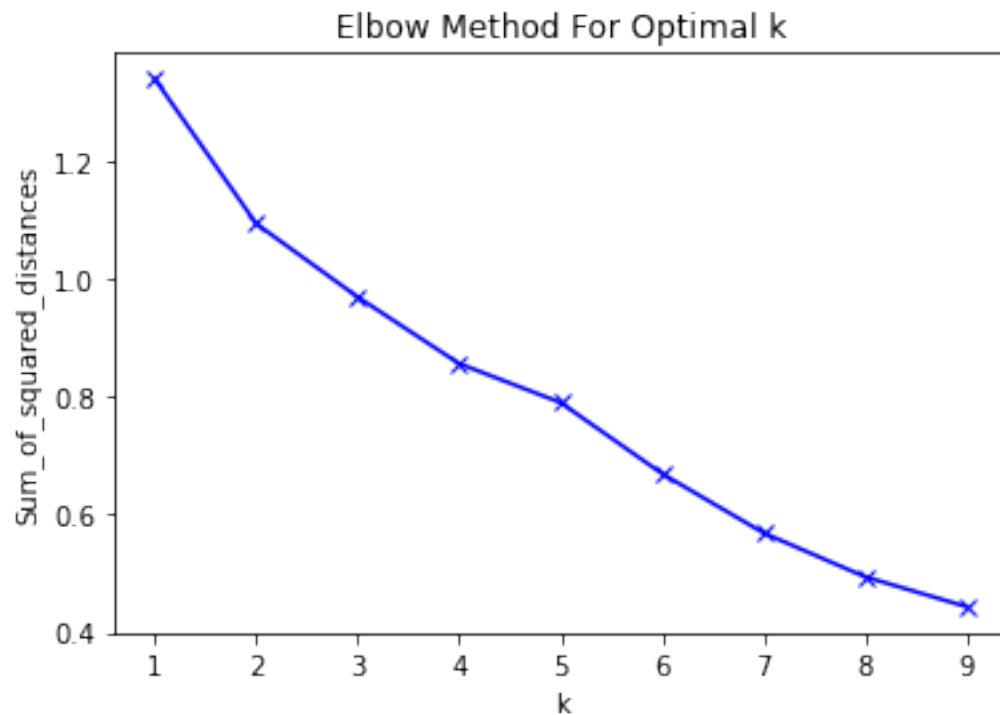
1. First of all, import the necessary libraries for this project;
2. Get the postcodes for the city of Munich;
3. Explode the dataframe to see all Postal codes;
4. Get the latitude and longitude for each postal code;

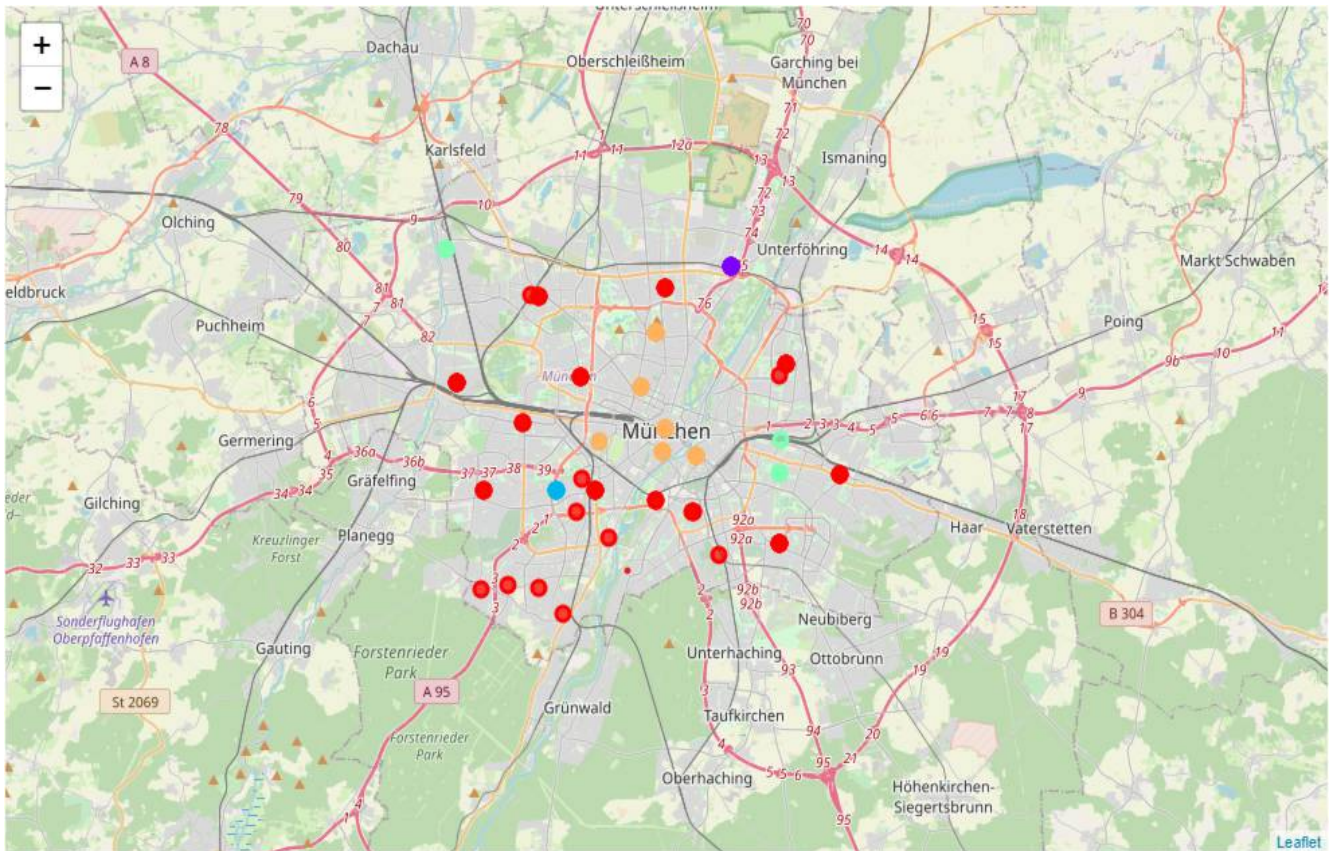
5. Visualization: Let's see all the postal codes on the map;
6. Next, we use the Foursquare API to explore the neighborhoods and segment them.

The Feldmoching-Hasenberg district has been excluded because it does not have enough observations. The venue bus stop was also excluded because it was not relevant for this model.

3 Cluster Neighbourhoods

This section presents the model used to find similar areas in the city of Munich. The Elbow method was used to define the best number of clusters in the model. The best K found was 5, so The city was divided into 5 similar areas.



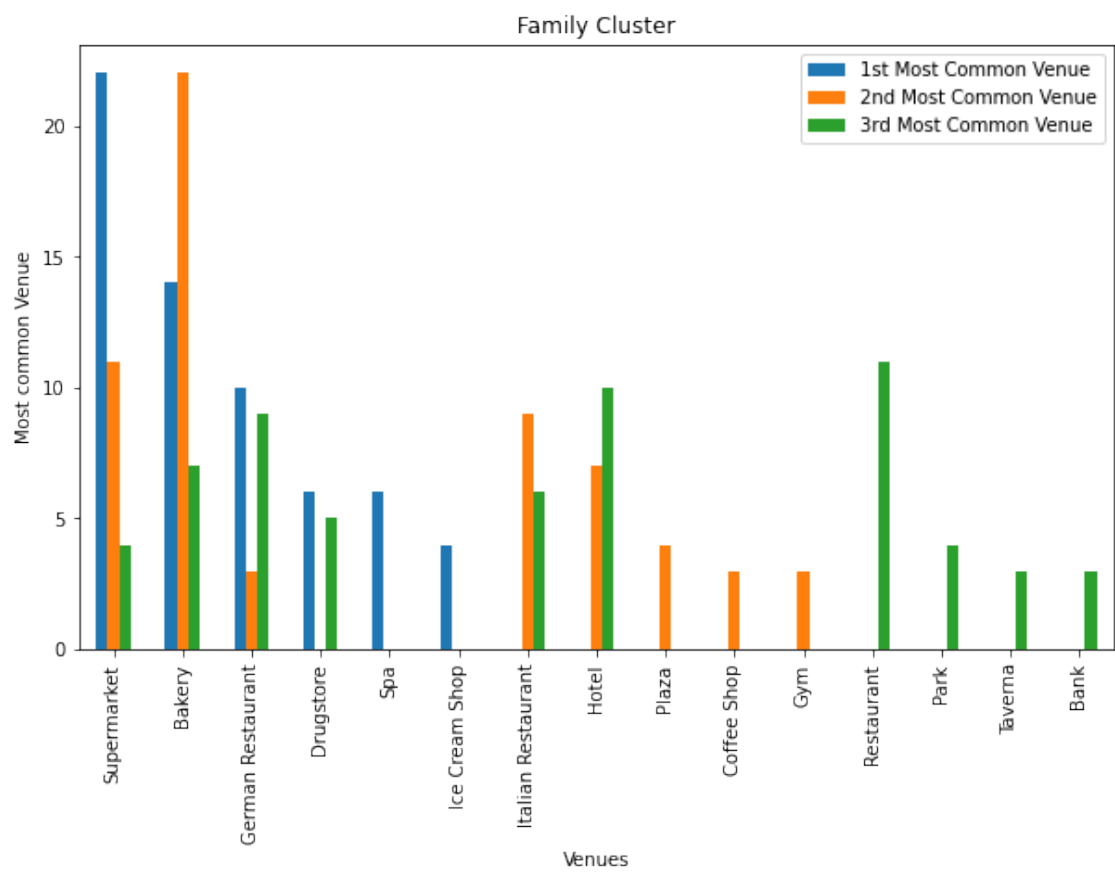


Map 1. Munich Clusters

4 Results and Discussion

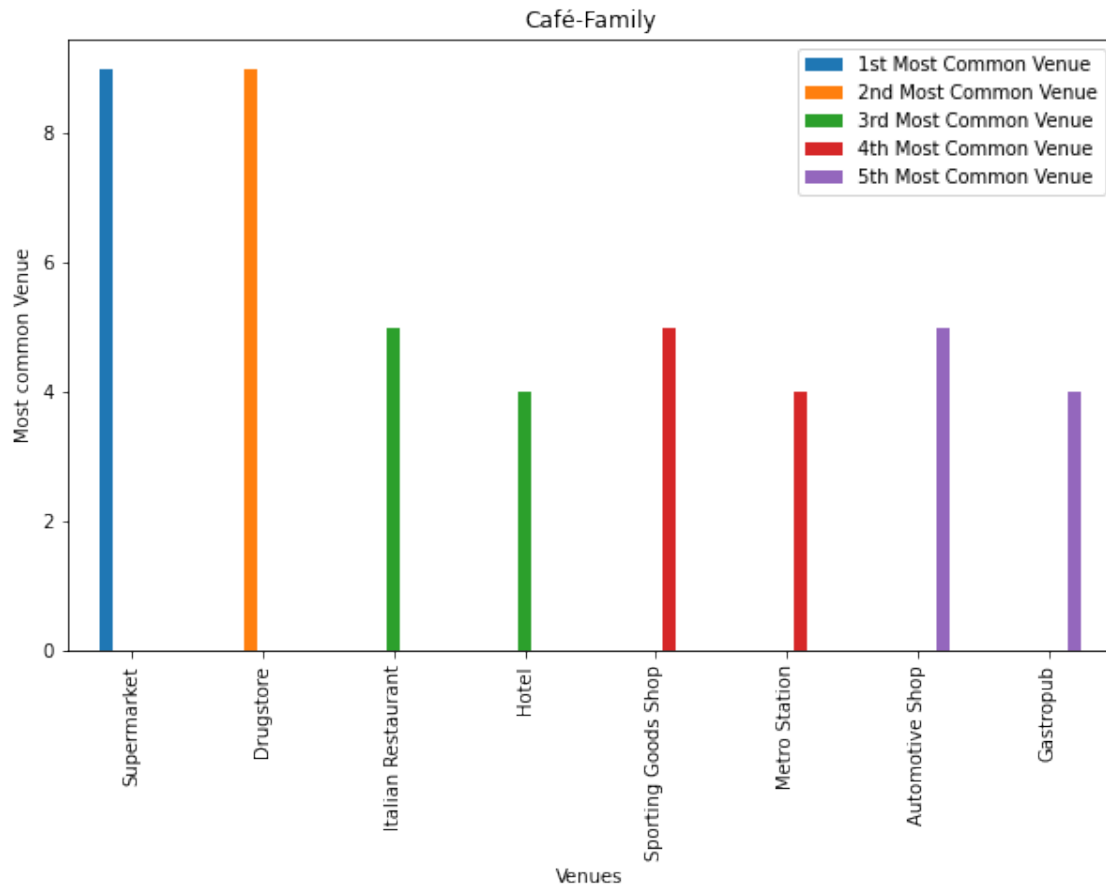
The city of Munich was divided into 5 clusters according to the similarities found in the Foursquare venues data. The city of Munich is somehow homogeneous, with many supermarkets, bakeries, restaurants, cafés, and other facilities.

The first cluster, number 0, is the biggest cluster in the city. This cluster is a good area for families with children, which is why we call it family cluster. It has many supermarkets, bakeries, drugstores, and other facilities like gyms, and ice cream stores, as shown in graph 1.



Graphic 1 Family Cluster

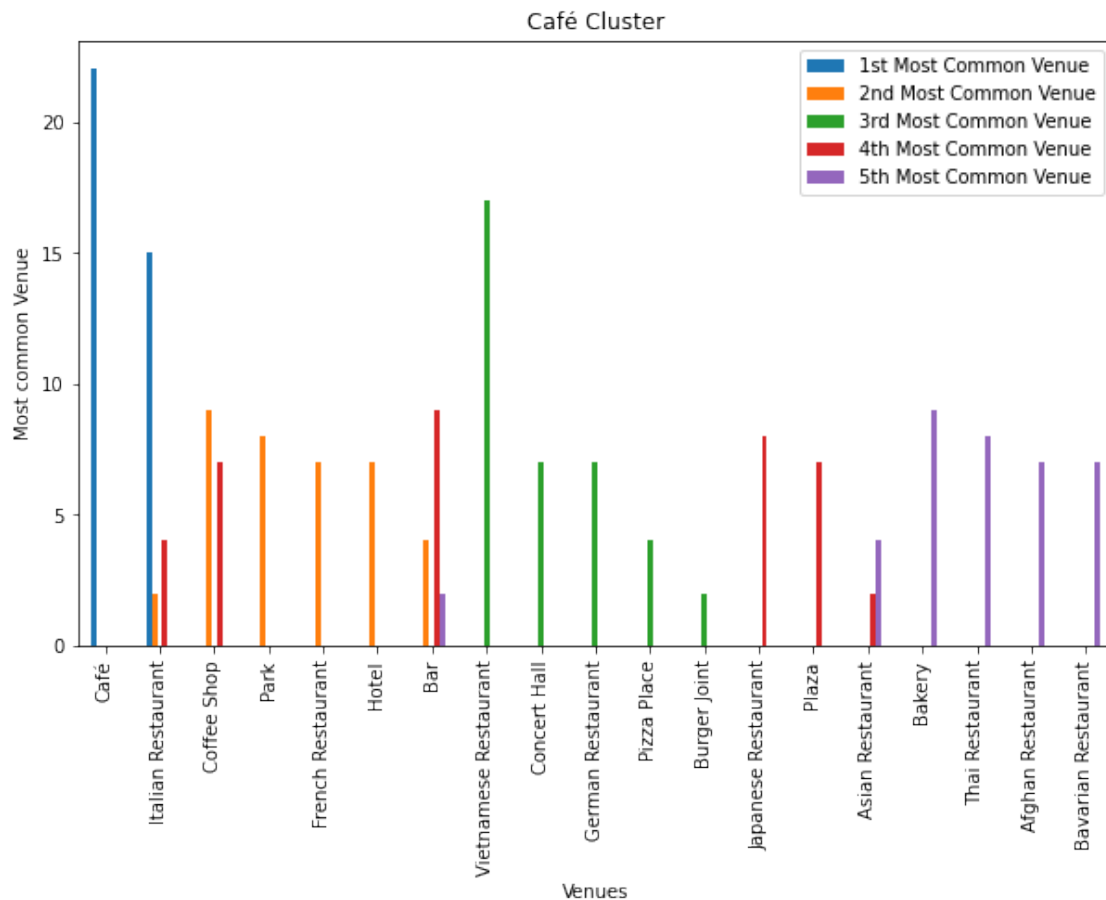
The cluster 3 is a good area for people who do not want to live in the center, but still want restaurants, cafes, hotels, gastropub, nightlife and also good facilities for families like supermarkets, drugstores. These are good areas for young families, we call this cluster Café-Family cluster.



Graphic 2 Café-Family cluster

The central area of the city belongs Café Cluster, these neighborhoods together presented as ‘1st most common place’, 22 cafes, 12 bakeries, 8 taverns, 7 Italian restaurants, 4 ice cream shops, as shown in graph 3. Apparently, this is a perfect cluster for those who like to go out for a bar, concert hall and have an active life.

[



Graphic 3 Café cluster

Two clusters presented only one district: Schwabing-Freimann (Cluster 1) and Sendling Westpark (Cluster 2). Schwabing-Freimann Presented an interesting diversity of venues. The district's '1st Most Common Venue' was fast food restaurant. The district also had gyms fitness centers, hotels, Yoga studios, and food and drink shops. Sendling Westpark also featured interesting types of locations. the most commonplace was the ice cream shop, the neighborhood also has a commercial service, tunnel, cafeteria, gym/fitness center, Italian restaurants, yoga studio, space for events.

5 Conclusion

Despite the limitations of the data, it was possible to have an overview of similar areas existing in the city of Munich. The aim of this study was to facilitate the work of real estate agents, who, according to the clusters found, can focus the search for properties in similar areas based on the preferences of their clients.

6 Bibliography

[1] Wikipedia. <https://en.wikipedia.org/wiki/Munich>

[2] Peter Bruce, Andrew Bruce, Peter Gedeck. “Practical Statistics for Data Scientists, 2nd Edition”. O’Reilly, May 2010.