**Rent Prices & Nearby Venues Data Analysis of Munich**

**Introduction**

Munich is the third-largest city in Germany with about 1.56 million inhabitant. It is divided into 25 districts and it has the most populated density per km2 in Germany (4,500 people per km2 ). Together with London, Paris and Berlin, the Munich property market ranks as one of the most important in Europe. As in 2020 demand exceeded available property, Munich remain Europe's favourite locations among investors[1].

This report intend to provide information to investors that wants to buy an accommodation in Munich considering rent prices and location attractiveness. Based on these two factors, Munich’s district will be segmented and clustered using Python and machine learning k-means. The report includes tables and a map with the plotted clusters.

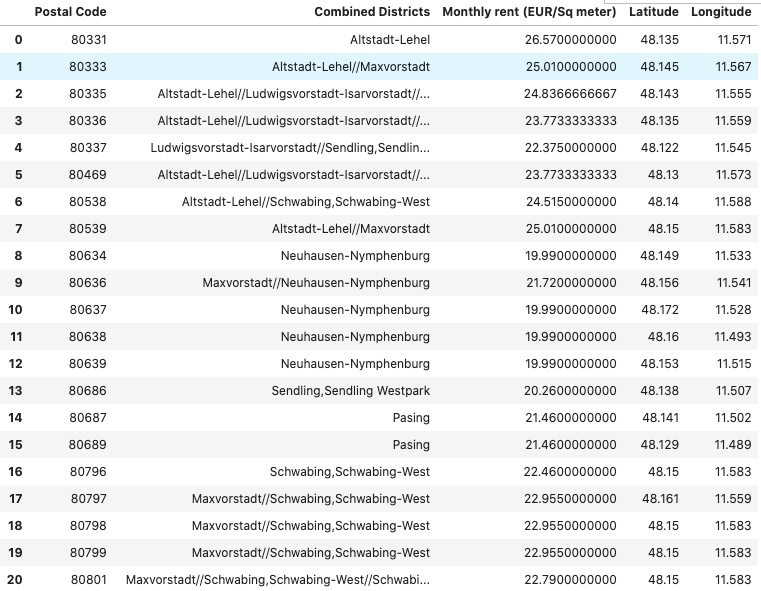
**Data section**

The report contains the following data:

“Bayern Postal Codes” from geonames.org[2] using Pandas library web scrapping function. “Average rent of apartments in Munich, Germany for the first half of 2019, by district”[3] downloaded from Statista and imported into Github project’s repository. “Postal Codes in Munich” [4] from official Munich website imported into Github project’s repository. Foursquare locations API to collect venues in Munich.

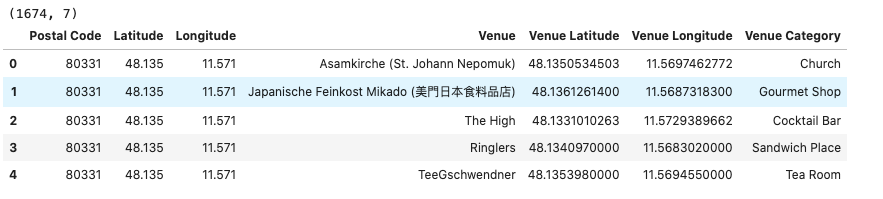
**Methodology section**

GitHub repository was used as the main database for this project. The master data called “dataset 3” was created using two different source from geonames.org, Average\_apart\_rent\_Munich.csv and Postal\_Code\_Munich(3).csv. Dataset 3 contains districts, Average price Monthly Rents, latitude and longitude of each postal code. It contains 50 postal codes.

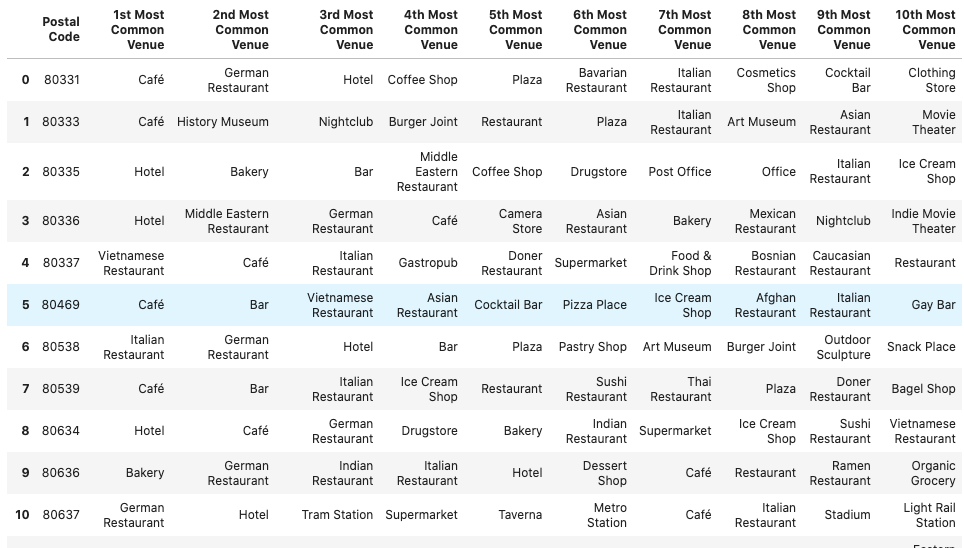


See notebook for all postal codes.

Foursquare API was used to explore venues in a radius of 500meters of each poste code (given their latitude and longitude) and 1674 venues were found.

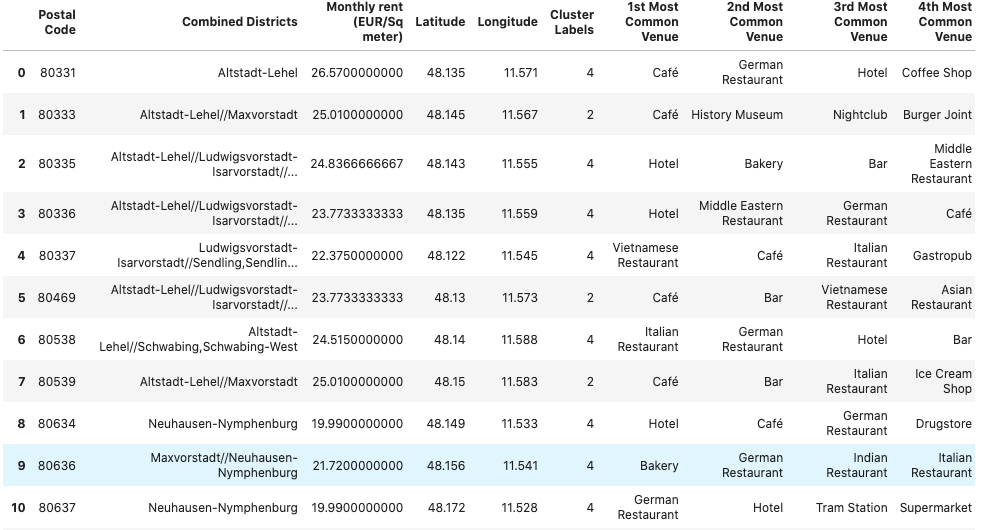


The above data were then grouped by postal codes and the mean of the frequency of occurrence of each category was created. The final data frame displays the top 10 venues for the 50 postal codes.



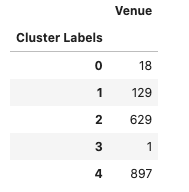
See notebook for all postal codes

The previous table shows that several postal codes have common venues. Therefore, unsupervised learning K-means algorithm will be used to cluster the Munich’s postal codes. The K-means is vastly used for clustering in many data science applications, especially useful to quickly discover insights from data. Below is the merged table with cluster labels for each postal codes.

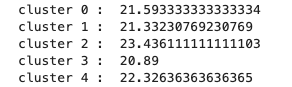


See notebook for all postal codes

The number of venues is calculated for each clusters:



The mean for monthly rents in EUR for each cluster was computed.



Each clusters are analysed and classified as follow:

CLUSTER 0 : // Avg rent price: 21.59EUR (EUR/Sq meter) // 18 venues // Transportations, restaurants, fast foods and farmer markets

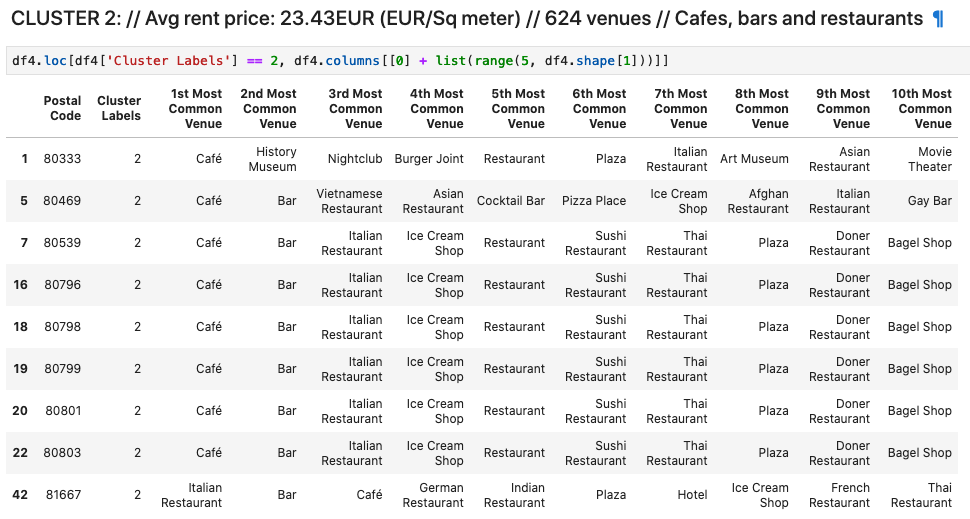
CLUSTER 1: // Avg rent price: 21.33EUR(EUR/Sq meter) // 129 venues // Transportations, restaurants, supermarket, bakery, banks

CLUSTER 2: // Avg rent price: 23.43EUR (EUR/Sq meter) // 624 venues // Cafes, bars and restaurants

CLUSTER 3: Outlier

CLUSTER 4: // Avg rent price: 22.32EUR (EUR/Sq meter) // 898 venues // Multiple Social Venues, Accomodations

Example of Cluster :



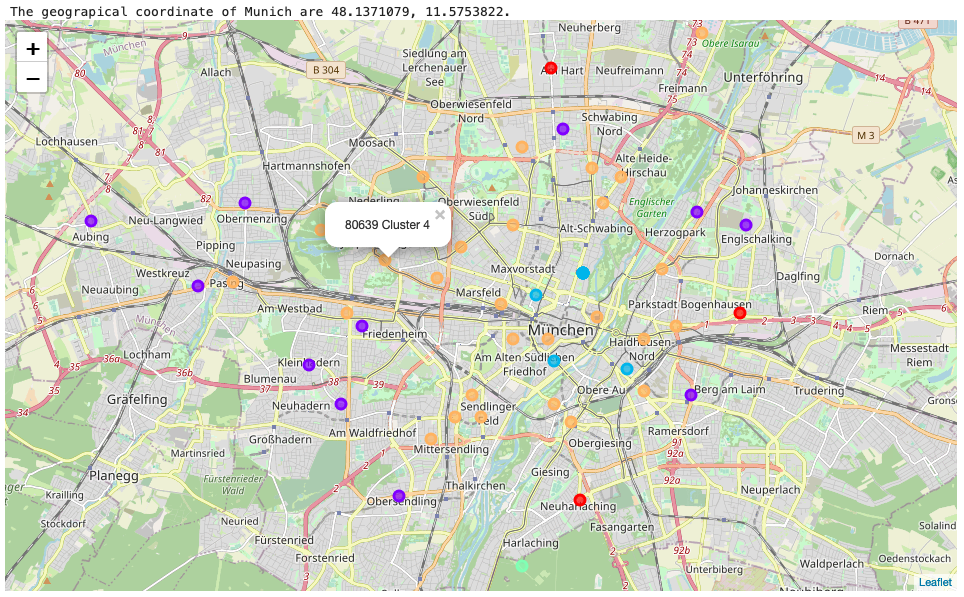
**Results**

Munich Latitude and Longitude was imported with API Geolocator and

Python folium library was used to visualize the city on a map. Clusters were plot on the map.

When selecting a specific cluster on the map the following information appears:

Red 🡪 « Postal Code, Cluster 0 »  
Violet 🡪 « Postal Code, Cluster 1 »  
Blue 🡪 « Postal Code, Cluster 2 »  
Orange 🡪 « Postal Code, Cluster 4 »



**Discussion and recommendations**

Districts with postal codes belonging to cluster 2 and 4 tends to be slightly more expensive than clusters 0 and 1. As shown on the map cluster 2 and 4 tend to be more central. Multiplying previous clusters average prices by 50 could give investors an idea of how much they will be able to rent there accommodations, for example a 50sqm flat. Cluster 2 and 4 respectively, 1171,5EUR/month and 1116EUR/month. Cluster 0 and 1 respectively, 1079,5EUR/month and 1066,5EUR/month.Cluster 2 and 4 have respectively 629 and 897 venues, much more than cluster 1 and 0 i.e. 18 and 129 venues. However clusters have really different categories of venues. Cluster 2 has a more intense density of cafe, bars and restaurants than cluster 4 and might be more appropriate for students. Therefore, in the area corresponding to cluster 2, investors should rather buy small single room flats. Accommodations belonging to Postal codes from cluster 4 will benefit from a more diverse type of venues (café and restaurants but also supermarkets, bakeries, banks, clothing stores, drugstore, offices, yoga studio…) and might be better for young workers or family. In those areas, it might be a better investment to buy accommodations with a bit more rooms. Accommodations in Cluster 1 will have all the necessary venues and will be quieter than accommodations in 2 and 4. Cluster 0 has really few venues but might be the quietest area and the less congested. Cluster 0 and 1 might be more suitable for families, for example small houses with garden.

**Conclusion**

Munich’s district were segmented and clustered using Python and k-means machine learning. Data for districts, average price monthly rents, latitude and longitude of each postal codes were web scrapped . Foursquare API was used to explore venues in a radius of 500meters of each poste code. Unsupervised learning K-means algorithm was used to cluster Munich’s postal codes. Previous analyses help investors to decide where to buy an accommodation. Investors can identify clusters and have a broad idea about the location price and venues density and category.

Further analyses could include to web scrape specialised website such as “ImmobilienScout24” or “Immowelt.de” to gather more data about price and include it as a component for clustering. Furthermore optimizing clustering could be done by finding a better k for the k-means with the Elbow Method.

**References**

[1] <https://www.muenchen.de/rathaus/wirtschaft_en/munich-business-location/economic-data.html>

[2] https://www.geonames.org/postal-codes/DE/BY/bayern.html

[3] <https://www.statista.com/statistics/800552/rental-costs-in-munich-germany-by-district/>

[4] https://www.muenchen.de/int/en/living/postal-codes.html