

IBM Data Science Professional Certificate

Applied Data Science Capstone Project

Gunjan S Makhloga

'SHOPs AND STOREs' SHOPPING IN MUNICH

INTRODUCTION

BACKGROUND

Munich is one of the most prosperous and populous cities of Bavaria, Germany. It has a rich tradition that can be seen in its world-famous Oktoberfest and beer-gardens. Furthermore, being home to people with different ethnic and geographical backgrounds, and tourists from all over the world, it has an interesting mix of different tastes. It would be interesting to understand the kind of shops and stores that are mostly visited by this eclectic set of people in different parts of Munich. It could also help investors understand the 'Shops and Stores' ecosystem here and gauge the demand to open a shop or a store of their interest in this ecosystem.

The shopping ecosystem in a city like Munich is very dynamic in nature. This means that there are fewer gaps between demand and supply. The moment a white space is identified in this ecosystem, an investor is ready to bridge it. The market is quick to respond to changing consumer behaviors and quick to generate demand. Price sensitivity of people is lower given the decent income levels. A combination of these factors makes Munich an attractive investment destination for investors with the right store at the right location and catering to the right audience; all of which can be identified by investing time in analyzing the vast amount of "Data" at our disposal.

BUSINESS PROBLEM

In cities like Munich that offer varied shopping experiences (Both online as well as offline), it is easy for someone to get lost. With a huge chunk of the population working in white collar jobs, "Time" is an important currency and usually a limiting factor for someone seeking an optimal shopping experience. Optimizing for the best shopping experience in a stipulated duration of time becomes critical. This leads us to

seek answers to questions like “Which are the best spots in the city that deliver such optimal shopping experiences?”; “If an investor is looking to open a shop in the city, which location will offer the best returns?”. Such questions have led me into creating an analysis for ‘Shops and Stores shopping in Munich’. The goal of this project is to use Foursquare API to determine the optimal location to open a store or a shop of a particular kind and understand their distribution within the city.

INTEREST

This report could be helpful to investors looking to open a shop or a store of their interest in a city like Munich. It could also help residents and tourist ease their shopping hunt and optimize for the best shopping experience.

DATA

DATA DESCRIPTION

Data collection plays an important role in deciding the nature of data analysis and the insights that could be drawn from it. As required to answer the business problem, we used Munich and its district dataset from publicly available source detailed in next section. The following data were pulled from various data sources:

- Munich districts and their postal codes

	District	PostalCode
0	Allach-Untermenzing	80995, 80997, 80999, 81247, 81249
1	Altstadt-Lehel	80331, 80333, 80335, 80336, 80469, 80538, 80539
2	Au-Haidhausen	81541, 81543, 81667, 81669, 81671, 81675, 81677
3	Aubing-Lochhausen-Langwied	81243, 81245, 81249
4	Berg am Laim	81671, 81673, 81735, 81825
5	Bogenhausen	81675, 81677, 81679, 81925, 81927, 81929
6	Feldmoching-Hasenbergl	80933, 80935, 80995
7	Hadern	80689, 81375, 81377
8	Laim	80686, 80687, 80689
9	Ludwigsvorstadt-Isarvorstadt	80335, 80336, 80337, 80469
10	Maxvorstadt	80333, 80335, 80539, 80636, 80797, 80798, 8079...

- Geographical coordinates in the form of latitudes and longitudes for each postal code

	District	PostalCode	Latitude	Longitude
0	Allach-Untermenzing	80995	48.211665	11.525658
0	Allach-Untermenzing	80995	48.211665	11.525658
1	Altstadt-Lehel	80331	48.136064	11.573450
0	Allach-Untermenzing	80995	48.211665	11.525658
1	Altstadt-Lehel	80331	48.136064	11.573450

- 100 venues that were within 500 meters radius for each postal code
- Venues and their category

DATA SOURCES

- Munich district data and their postal codes were scraped from web page <https://www.muenchen.de/int/en/living/postal-codes.html> using Python GET request and BeautifulSoup packages.
- The latitudes and longitudes coordinate of Munich and each of its district postal code were obtained by using Python Geocoder package.
- RESTful API calls were made to Foursquare API to retrieve the information about the venues for each postal code. It also provided the categories of the venue data. To the interest of this project, we used only 'Shops' and 'Stores' categories to solve the business problem.

METHODOLOGY

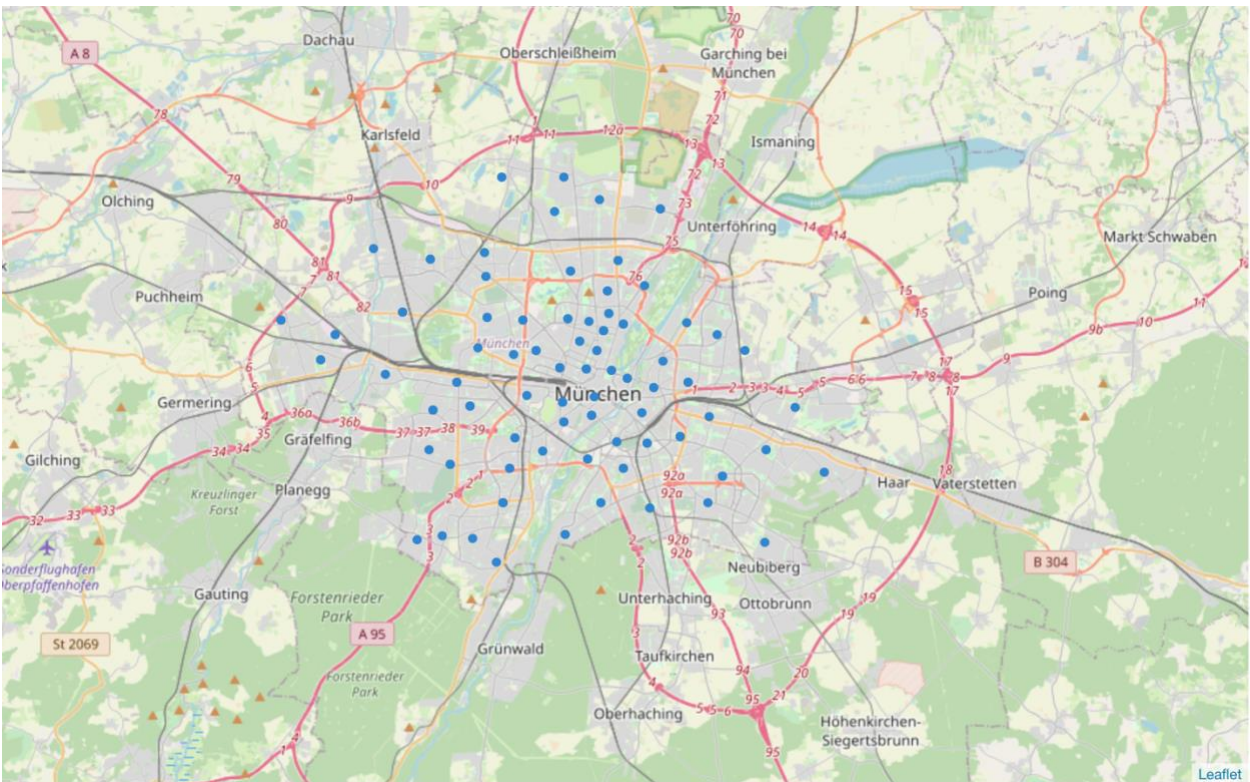
Firstly, we obtained dataset containing list of districts in Munich from the web page <https://www.muenchen.de/int/en/living/postal-codes.html> using GET request to pull necessary data through the webpage. We exploded districts and their respective postal codes to separate postal codes grouped for each district to one postal code per row. This is shown in the figure given below:

	District	variable	PostalCode
0	Allach-Untermenzing	0	80995
1	Altstadt-Lehel	0	80331
2	Au-Haidhausen	0	81541
3	Aubing-Lochhausen-Langwied	0	81243
4	Berg am Laim	0	81671

Secondly, we obtained geographical coordinate of Munich using the geocoder package. We fetched geographical coordinates as latitude and longitude. Thereafter, we created a new dataframe to loop the same procedure over each district to get latitude and longitude of each postal code in a district as shown below:

	District	PostalCode	Latitude	Longitude
0	Allach-Untermenzing	80995	48.211665	11.525658
0	Allach-Untermenzing	80995	48.211665	11.525658
1	Altstadt-Lehel	80331	48.136064	11.573450
0	Allach-Untermenzing	80995	48.211665	11.525658
1	Altstadt-Lehel	80331	48.136064	11.573450

Thirdly, we visualized the location of various postal codes within Munich to have a general understanding of the location which appeared as such



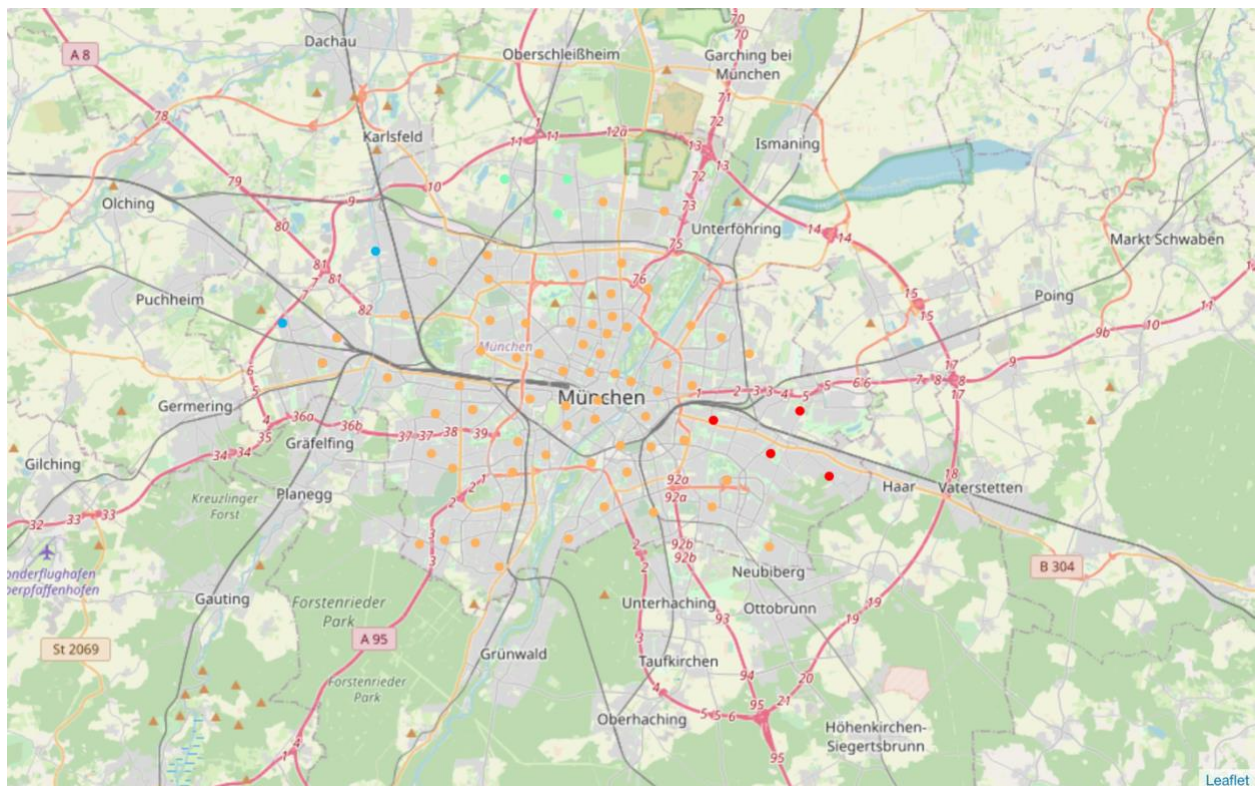
Furthermore, we defined Foursquare credentials and version to make API calls to Foursquare in order to explore each neighborhood and returned the top 100 venues within 500 m for each postal code. The extracted venue categories were encoded using one-hot encoding to handle any categorical values in the dataset.

As for clustering algorithms, we used K-means to cluster data and find solutions to the business problem at hand. This type of algorithm clusters fits every data value to its nearest cluster. Every data point in a cluster are similar to each other and dissimilar with data points in another cluster.

We used Folium library to visualize distribution of clusters within Munich. This plotted clusters helped us visualize 'Shops and Stores' distribution within Munich. These clusters were further analyzed to understand 'Most Common Venues' on a broader spectrum. To ease our venue selection in each cluster, we used '1st most common venue' to plot number of venues per cluster in the end. This bar plot helped us give a name to each of our clusters.

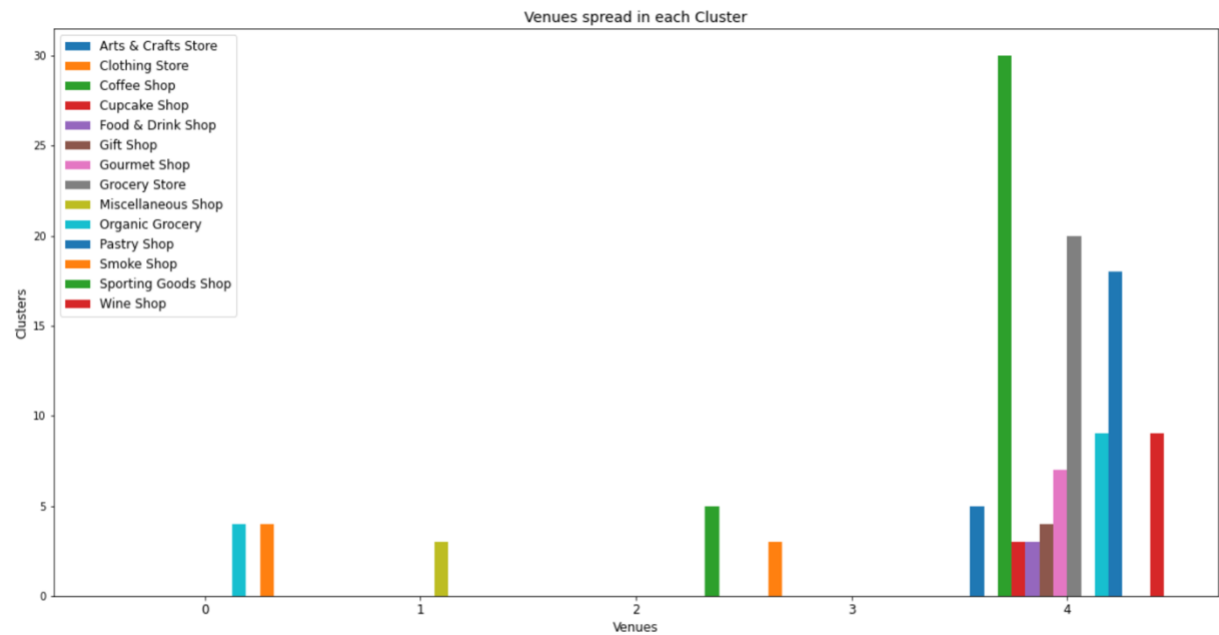
RESULTS

The results from K-means clustering categorized the neighborhoods into 5 clusters based on type of 'Shops and Stores' shopping. This can be visualized in 'cluster map' given below:



Cluster shown in orange dominated most of the city. Clusters in blue, red and green holds a minority at the outskirts of Munich. We will discuss each cluster in detail in 'Discussion' section.

Graphical plot of ‘Most common venues’ in each cluster is as shown:



Cluster (4) has a beautiful spread of different kinds of venues while the other cluster (0-3) has venues spread scantily. Cluster (4) can be visualized as orange dots in the ‘cluster map’ which covered majority of the city.

DISCUSSION

As mentioned in previous section, Cluster in orange color dominated most part of the city while the other clusters were scantily present near the border of Munich. Let’s discuss the 5 clusters shown in the ‘cluster map’ cluster-wise:

Cluster 0:

```
#CLUSTER 0 (Red): 'Organic Grocery and Smoke Shop'  
M_merged.loc[M_merged['Cluster Name'] == 0, M_merged.columns[[1] + list(range(5, M_merged.shape[1]))]]
```

	PostalCode	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	81671	Smoke Shop	Shopping Mall	Shipping Store	Organic Grocery	Arts & Crafts Store	Music Store	Gourmet Shop	Grocery Store	Hardware Store	Liquor Store
23	81735	Organic Grocery	Shopping Mall	Clothing Store	Arts & Crafts Store	Grocery Store	Hardware Store	Liquor Store	Men's Store	Miscellaneous Shop	Mobile Phone Shop
29	81673	Smoke Shop	Shopping Mall	Shipping Store	Organic Grocery	Arts & Crafts Store	Music Store	Gourmet Shop	Grocery Store	Hardware Store	Liquor Store
48	81825	Organic Grocery	Shopping Mall	Clothing Store	Arts & Crafts Store	Grocery Store	Hardware Store	Liquor Store	Men's Store	Miscellaneous Shop	Mobile Phone Shop
54	81735	Smoke Shop	Shopping Mall	Shipping Store	Organic Grocery	Arts & Crafts Store	Music Store	Gourmet Shop	Grocery Store	Hardware Store	Liquor Store
72	81827	Organic Grocery	Shopping Mall	Clothing Store	Arts & Crafts Store	Grocery Store	Hardware Store	Liquor Store	Men's Store	Miscellaneous Shop	Mobile Phone Shop
77	81825	Smoke Shop	Shopping Mall	Shipping Store	Organic Grocery	Arts & Crafts Store	Music Store	Gourmet Shop	Grocery Store	Hardware Store	Liquor Store
92	81829	Organic Grocery	Shopping Mall	Clothing Store	Arts & Crafts Store	Grocery Store	Hardware Store	Liquor Store	Men's Store	Miscellaneous Shop	Mobile Phone Shop

Cluster 0 has a small coverage and is present in the eastern part of Munich. Although it seems to have more of ‘non-food’ shops and stores including a shopping mall, Organic grocery and Smoke shops are most frequented venues. Thus, we labelled this cluster as ‘Organic Grocery and Smoke Shops’. Both of them are equally frequented as shown in graphical plot of ‘Most common venues’ in each cluster.

Cluster 1:

```
#CLUSTER 1 (Violet): 'Miscellaneous Shop'
M_merged.loc[M_merged['Cluster Name'] == 1,M_merged.columns[[1] + list(range(5, M_merged.shape[1]))]]
```

	PostalCode	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
3	81243	Miscellaneous Shop	Arts & Crafts Store	Organic Grocery	Gourmet Shop	Grocery Store	Hardware Store	Liquor Store	Men's Store	Mobile Phone Shop	Music Store
28	81245	Miscellaneous Shop	Arts & Crafts Store	Organic Grocery	Gourmet Shop	Grocery Store	Hardware Store	Liquor Store	Men's Store	Mobile Phone Shop	Music Store
53	81249	Miscellaneous Shop	Arts & Crafts Store	Organic Grocery	Gourmet Shop	Grocery Store	Hardware Store	Liquor Store	Men's Store	Mobile Phone Shop	Music Store

Cluster 1 should have been in violet shade following the rainbow color order. However, it's not shown in ‘cluster map’. This can still be seen in graphical plot of ‘Most common venues’ in each cluster as olive green bar for Miscellaneous Shops. Thus, we labelled this cluster as ‘Miscellaneous Shops’.

Cluster 2:

```
#CLUSTER 2 (Blue): 'Sporting Goods Shop'
M_merged.loc[M_merged['Cluster Name'] == 2,M_merged.columns[[1] + list(range(5, M_merged.shape[1]))]]
```

	PostalCode	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	80995	Sporting Goods Shop	Arts & Crafts Store	Organic Grocery	Gourmet Shop	Grocery Store	Hardware Store	Liquor Store	Men's Store	Miscellaneous Shop	Mobile Phone Shop
25	80997	Sporting Goods Shop	Arts & Crafts Store	Organic Grocery	Gourmet Shop	Grocery Store	Hardware Store	Liquor Store	Men's Store	Miscellaneous Shop	Mobile Phone Shop
50	80999	Sporting Goods Shop	Arts & Crafts Store	Organic Grocery	Gourmet Shop	Grocery Store	Hardware Store	Liquor Store	Men's Store	Miscellaneous Shop	Mobile Phone Shop
74	81247	Sporting Goods Shop	Arts & Crafts Store	Organic Grocery	Gourmet Shop	Grocery Store	Hardware Store	Liquor Store	Men's Store	Miscellaneous Shop	Mobile Phone Shop
93	81249	Sporting Goods Shop	Arts & Crafts Store	Organic Grocery	Gourmet Shop	Grocery Store	Hardware Store	Liquor Store	Men's Store	Miscellaneous Shop	Mobile Phone Shop

Cluster 2 is spread scantily in the north-western border of Munich. Detailed study of this cluster shows a mix of different categories (of venues). The most common venue in this cluster is Sporting Goods Shop followed by Art and Crafts Store. Thus, we labelled this cluster as ‘Sporting Goods Shop’.

Cluster 3:

```
#CLUSTER 3 (Green): 'Clothing Store'
M_merged.loc[M_merged['Cluster Name'] == 3, M_merged.columns[[1] + list(range(5, M_merged.shape[1]))]]
```

	PostalCode	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
6	80933	Clothing Store	Arts & Crafts Store	Pastry Shop	Grocery Store	Hardware Store	Liquor Store	Men's Store	Miscellaneous Shop	Mobile Phone Shop	Music Store
31	80935	Clothing Store	Arts & Crafts Store	Pastry Shop	Grocery Store	Hardware Store	Liquor Store	Men's Store	Miscellaneous Shop	Mobile Phone Shop	Music Store
56	80995	Clothing Store	Arts & Crafts Store	Pastry Shop	Grocery Store	Hardware Store	Liquor Store	Men's Store	Miscellaneous Shop	Mobile Phone Shop	Music Store

Cluster 3 is shown in the extreme north of Munich with Clothing Store being the most frequented venue. Thus, we called this cluster as 'Clothing Store'.

Cluster 4:

```
#CLUSTER 4 (Orange): 'Coffee Shop'
M_merged.loc[M_merged['Cluster Name'] == 4, M_merged.columns[[1] + list(range(5, M_merged.shape[1]))]]
```

	PostalCode	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
1	80331	Coffee Shop	Clothing Store	Cosmetics Shop	Gourmet Shop	Shoe Store	Organic Grocery	Pastry Shop	Grocery Store	Department Store	Food & Drink Shop
2	81541	Gourmet Shop	Coffee Shop	Organic Grocery	Pastry Shop	Shipping Store	Hardware Store	Mobile Phone Shop	Department Store	Cosmetics Shop	Cupcake Shop
5	81675	Wine Shop	Coffee Shop	Pastry Shop	Gourmet Shop	Smoke Shop	Optical Shop	Grocery Store	Hardware Store	Liquor Store	Men's Store
7	80689	Food & Drink Shop	Shop & Service	Miscellaneous Shop	Music Store	Arts & Crafts Store	Organic Grocery	Grocery Store	Hardware Store	Liquor Store	Men's Store
8	80686	Wine Shop	Sporting Goods Shop	Shop & Service	Coffee Shop	Mobile Phone Shop	Music Store	Pastry Shop	Organic Grocery	Optical Shop	Grocery Store
...
122	81379	Grocery Store	Organic Grocery	Clothing Store	Liquor Store	Costume Shop	Optical Shop	Food & Drink Shop	Pastry Shop	Record Shop	Furniture / Home Store
123	80801	Coffee Shop	Gift Shop	Clothing Store	Donut Shop	Arts & Crafts Store	Sausage Shop	Grocery Store	Optical Shop	Organic Grocery	Frozen Yogurt Shop
124	80939	Pastry Shop	Clothing Store	Coffee Shop	Gift Shop	Farmers Market	Arts & Crafts Store	Donut Shop	Organic Grocery	Optical Shop	Liquor Store
125	80809	Coffee Shop	Gift Shop	Pastry Shop	Liquor Store	Donut Shop	Beer Store	Optical Shop	Shipping Store	Farmers Market	Grocery Store
126	80802	Coffee Shop	Gift Shop	Clothing Store	Donut Shop	Arts & Crafts Store	Sausage Shop	Grocery Store	Optical Shop	Organic Grocery	Frozen Yogurt Shop

Cluster 4 has an interesting spread of almost every venue category. As shown in graphical plot of 'Most common venues' in each cluster, cluster 4 also has arts and crafts store, Wine shop, cupcake shop, food and drink shop, gift shop, gourmet shop, organic grocery and pastry shop. For ease in labeling this cluster, let's call it as 'Coffee Shop' on basis of its count.

CONCLUSION

Munich provides optimal shopping experience in most part of the city. It has Coffee shops in abundance. And who doesn't like Coffee shops in vicinity! Beautiful cafes are a common sight in Munich. Further east enjoys Organic Grocery and Smoke shop shopping while 'Clothing Stores' are frequented in the north. 'Sporting Goods Store' seems to be frequented in the north-western part of Munich.

Data analysis and machine learning used in this project helped us understand the spread of 'Shops and Stores' in Munich. On an interesting note, much of 'Shops and Stores' ecosystem is consistent throughout the city.