Coursera Capstone of Data Science - Data Analysis of Population Density and Venues in Munich

Introduction

Description and Discussion of Background

Munich as the largest city of Bavaria, the third largest city in Germany behind Berlin and Hamburg is one of the rapidest growing cities in Germany. Along with the development of the city Munich is also a top destination for immigrants. The estimated population of Munich in 2019 is about 1.45 million and the the population density is at 4,500 per square kilometer, which is the highest in the country.

The growth of industry and population makes Munich a hot city to do investments. There are 25 boroughs in Munich and every borough has different popular venues and population density. The relationship of population density could be a very interesting point for investors to consider which addresses the problem I want to analyze.

A good way to analyze and visualize the problem is to create a map with information of people density and clustered venues' information.

Data Description

The needed data consist of three parts: geographic data, population data and venues data.

- geographic data can be obtained from google map, wikepedia[1].

- population data can be aquaired from database of government in Munich[2] like https://www.opengov-muenchen.de/dataset/indikatorenatlas-bevoelkerung-natuerlicher-saldo-auf-1000-einwohner-innen-83r65mct and wikipedia[3] https://de.wikipedia.org/wiki/Stadtbezirke_M%C3%BCnchens

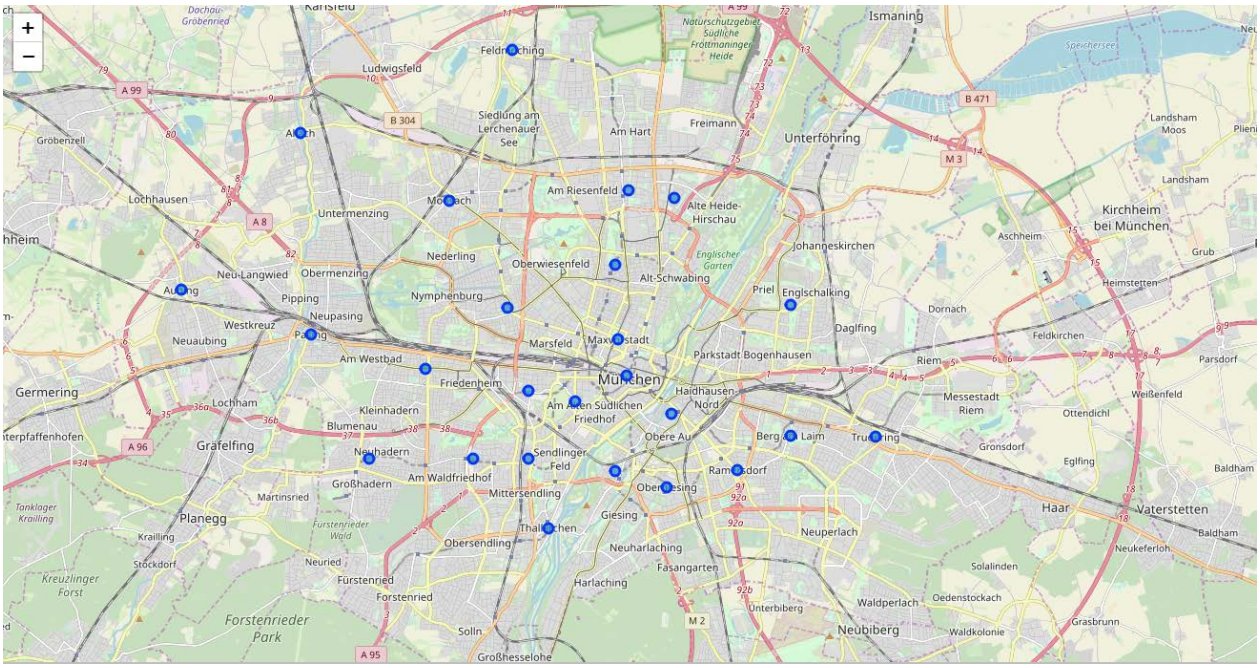
- venues data can be obtained from Forsquare database.

Methodology

The main database consists of five elements: Nr. which means the borough's number, borough's name, population density, latitude and longitude of every borough.

	Nr.	Stadtbezirk	Dichte(Einw./km²)	Latitude	Longitude
0	1	Altstadt-Lehel	6820.0	48.137828	11.574582
1	2	Ludwigsvorstadt-Isarvorstadt	12477.0	48.131771	11.555809
2	3	Maxvorstadt	12435.0	48.146570	11.571445
3	4	Schwabing-West	15908.0	48.164417	11.570364
4	5	Au-Haidhausen	14693.0	48.128753	11.590536

With the help of package FOLIUM the boroughs in Munich can be visualized as blow:



Based on the coordinates of every borough Foursquare API can help us find the most visited venues in the borough. In my case I am interested in the top 100 venues within 1km. After getting the response data from Foursquare API a data frame with information of borough and its top 100 venues can be obtained.

	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Altstadt-Lehel	48.137828	11.574582	Marienplatz	48.137177	11.575561	Plaza
1	Altstadt-Lehel	48.137828	11.574582	Fischbrunnen	48.137211	11.576047	Fountain
2	Altstadt-Lehel	48.137828	11.574582	Alois Dallmayr	48.138554	11.576750	Gourmet Shop
3	Altstadt-Lehel	48.137828	11.574582	St. Peter	48.136530	11.575615	Church
4	Altstadt-Lehel	48.137828	11.574582	Kustermann	48.136242	11.574897	Department Store

Since the types of venues can indicate the type of borough, we can use the top ten most common venues to deduce the type of the borough. To achieve this goal K-means algorithm for clustering is a very good tool to cluster the boroughs because the problem is a typical unsupervised learning case and K-means algorithm is one of the most common used algorithms for that.

The boroughs are clustered into 3 clusters. Based on the most common visited venues of every cluster, which is listed in table below,

Cluster Labels	1st Most Common Venue	Nr.
0	Bakery	2
	German Restaurant	1
	Supermarket	1
1	Bus Stop	7
	Hotel	1
2	Café	3
	Hotel	3
	Italian Restaurant	2
	Supermarket	4
	Zoo Exhibit	1

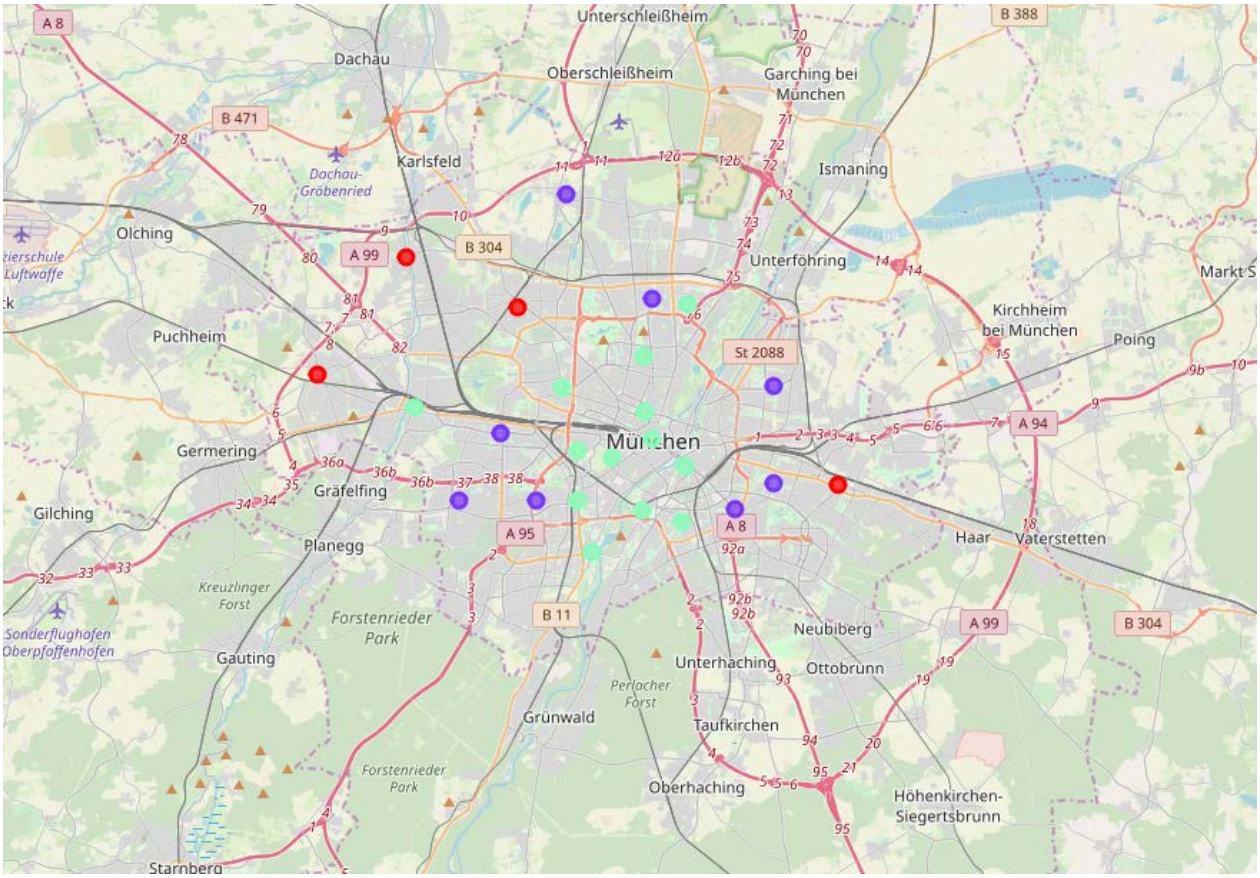
we can define every cluster as:

- Cluster 0: local life venues

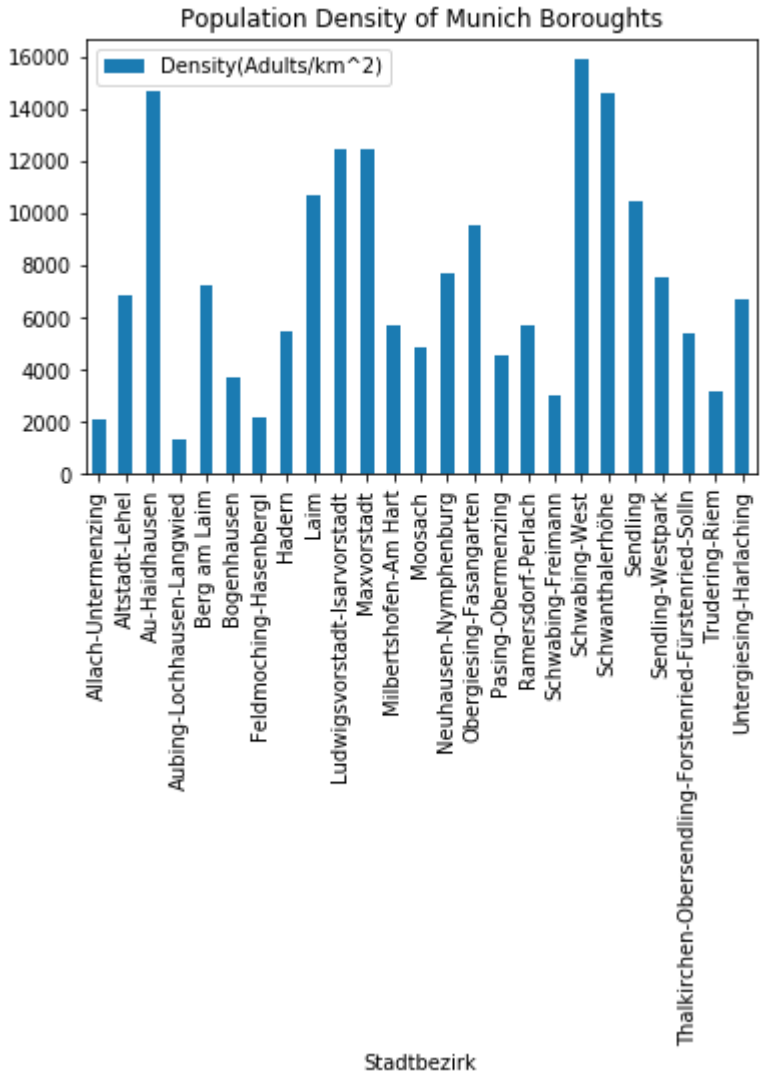
- Cluster 1: business trip/work life venues

- Cluster 2: Casual trip venues

And the clustered boroughs can be visualized on the map:



The population density of every boroughs can be shown here:



Results

The relationship of the population density and type of borough can be concluded as: the higher the population density is, the more commercialized the borough is. For example, borough Aubing-Lochhausen-Langwied has the least population density and the most visited venue in that area is Bakery. While Schwabing-West has the largest population density and the most visited venue is Italian restaurant.

Discussion

Based on the results I would like to recommend the investors to invest in the borough which has a large population density. Large population density in Munich means more requirements and more business.

Conclusion

There will be more business chance if there are more people and usually city center is a very good place to invest.

References

[1] https://de.wikipedia.org/wiki/Stadtbezirke_M%C3%BCnchens

[2] https://www.opengov-muenchen.de/dataset/indikatorenatlas-bevoelkerung-natuerlicher-saldo-auf-1000-einwohner-innen-83r65mct

[3] https://de.wikipedia.org/wiki/Stadtbezirke_M%C3%BCnchens