

Coursera - IBM

Applied Data Science Capstone Project

The Battle of Neighborhoods in Munich, Germany

by Xingyuan Tan

1. Introduction

Munich is the capital and most populous city of Bavaria. With a population of 1,558,395 inhabitants as of July 31, 2020, it is the third-largest city in Germany, after Berlin and Hamburg, and thus the largest which does not constitute its own state, as well as the 11th-largest city in the European Union. The city's metropolitan region is home to 6 million people.

Today, Munich is a global centre of art, science, technology, finance, publishing, culture, innovation, education, business, and tourism and enjoys a very high standard and quality of living, reaching first in Germany and third worldwide according to the 2018 Mercer survey, and being rated the world's most liveable city by the Monocle's Quality of Life Survey 2018. According to the Globalization and World Rankings Research Institute, Munich is considered an alpha-world city, as of 2015. It is one of the most prosperous and fastest growing cities in Germany.

2. Business Problem

As Munich is becoming a very attractive city to live in, one individual wants to start his own business in the city by opening a restaurant. However the city is too big for him to decide which neighborhood would be the best choice for a start-up.

3. Data

The postal code data is coming from Muechen.de, being published via <https://www.muenchen.de/int/en/living/postal-codes.html>; and the geo data is retrieving from GeoNames.org and being stored in git-hub for further usage. Since two data contains different part of data, the combination is being done in the project.

3.1 Neighborhoods Data

Postal codes in Munich



A large city, several dozen postal codes are used for Munich: From 80331 to 81929, a broad range of five-digit codes is used to ensure that each and every Munich resident receives letters and packages. Our filter search, below, helps you find out which postal code goes with which district of Munich. You can also filter by district to see which postal codes are used for each one.

Munich's postal codes by district

Search Postal Code or District	
District	Postal Code
Allach-Untermenzing	80995, 80997, 80999, 81247, 81249
Altstadt-Lehel	80331, 80333, 80335, 80336, 80469, 80538, 80539
Au-Haidhausen	81541, 81543, 81667, 81669, 81671, 81675, 81677
Aubing-Lochhausen-Langwied	81243, 81245, 81249

3.2 Geo Data (from GeoNames.Org)

Germany - postal codes

Either enter a postal code (eg. "9011", "AB1", "9980-999") or a city (eg. "London")

	Place	Code	Country	Admin1	Admin2	Admin3	Admin4
1	München	80333	Germany	Bayern	Upper Bavaria	Kreisfreie Stadt München	München, Landeshauptstadt
		48.145/11.567					
2	München	80331	Germany	Bayern	Upper Bavaria	Kreisfreie Stadt München	München, Landeshauptstadt
		48.135/11.571					
3	München	80335	Germany	Bayern	Upper Bavaria	Kreisfreie Stadt München	München, Landeshauptstadt
		48.143/11.555					
4	München	80336	Germany	Bayern	Upper Bavaria	Kreisfreie Stadt München	München, Landeshauptstadt
		48.135/11.559					
5	München	80337	Germany	Bayern	Upper Bavaria	Kreisfreie Stadt München	München, Landeshauptstadt
		48.122/11.545					
6	München	80339	Germany	Bayern	Upper Bavaria	Kreisfreie Stadt München	München, Landeshauptstadt
		48.139/11.534					
7	München	80469	Germany	Bayern	Upper Bavaria	Kreisfreie Stadt München	München, Landeshauptstadt
		48.13/11.573					
8	München	80538	Germany	Bayern	Upper Bavaria	Kreisfreie Stadt München	München, Landeshauptstadt
		48.14/11.588					
9	München	80539	Germany	Bayern	Upper Bavaria	Kreisfreie Stadt München	München, Landeshauptstadt
		48.15/11.583					
10	München	80636	Germany	Bayern	Upper Bavaria	Kreisfreie Stadt München	München, Landeshauptstadt
		48.156/11.541					

3.3 Data Combination

After data is retrieved from different, it is needed to be combined in the project to get the neighborhoods data with geo info.

```
latitudes = [] # Initializing the latitude array
longitudes = [] # Initializing the longitude array
postal_codes = muc['Postal Code']
```

```
#Mapping postal code between two data frames
for postal_code in postal_codes :
    lat = df[df['zipcode']== int(postal_code)]['latitude'].iloc[0]
    lon = df[df['zipcode'] == int(postal_code)]['longitude'].iloc[0]

    latitudes.append(lat)
    longitudes.append(lon)
```

```
muc['Latitude'] = latitudes
muc['Longitude'] = longitudes
```

```
muc.head()
```

	District	Postal Code	Latitude	Longitude
0	Allach-Untermenzing	80995	48.1976	11.5181
1	Allach-Untermenzing	80997	48.1834	11.4784
2	Allach-Untermenzing	80999	48.1853	11.4643
3	Allach-Untermenzing	81247	48.1662	11.4673
4	Allach-Untermenzing	81249	48.1500	11.5833

3.4 Venue Data

```
print(muc_venues.shape)
muc_venues.head()
```

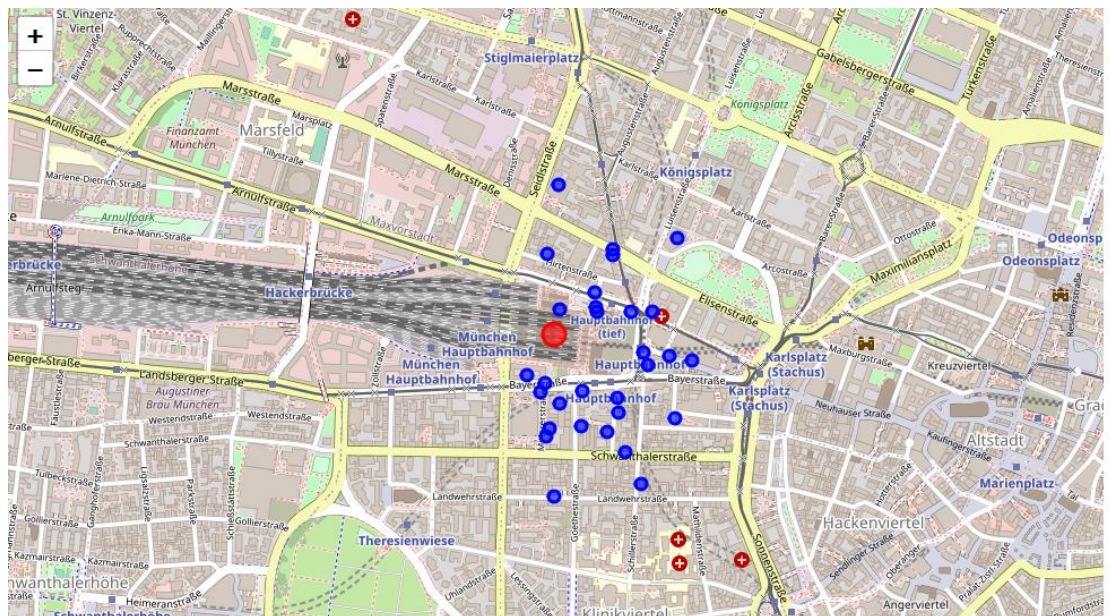
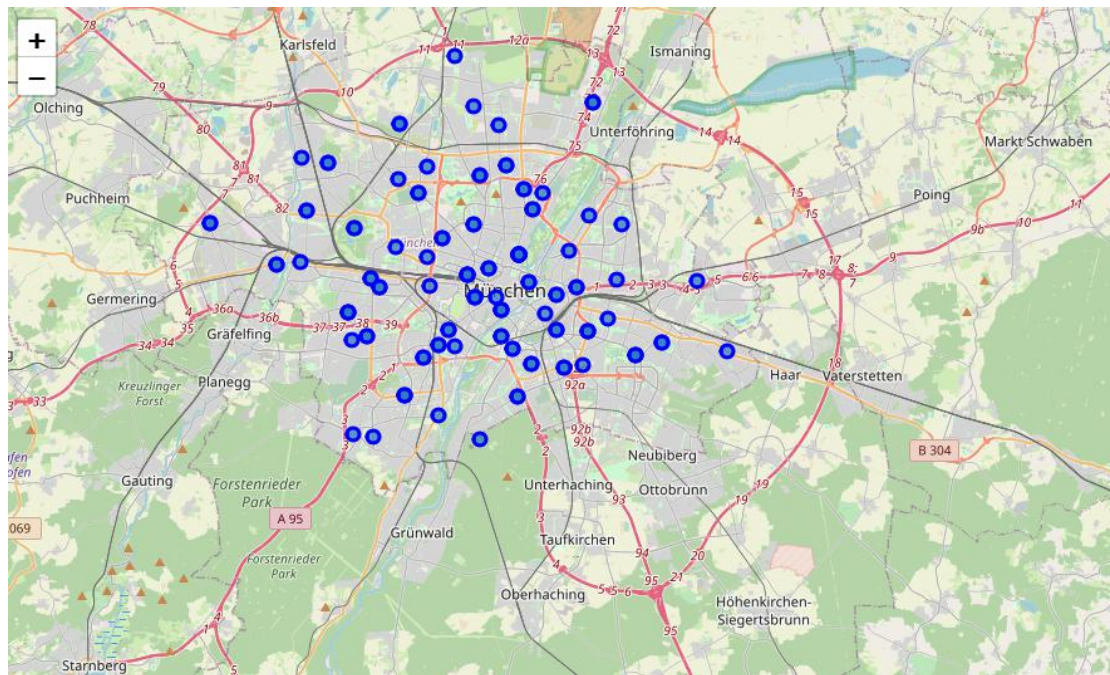
```
(2414, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Allach-Untermenzing	48.1976	11.5181	Indian Village	48.197932	11.524083	Indian Restaurant
1	Allach-Untermenzing	48.1976	11.5181	EDEKA Fratzl	48.197393	11.524481	Supermarket
2	Allach-Untermenzing	48.1976	11.5181	Pizza Fasanerie	48.197816	11.524199	Pizza Place
3	Allach-Untermenzing	48.1976	11.5181	H Fasanerie Bahnhof	48.196895	11.524417	Bus Stop
4	Allach-Untermenzing	48.1834	11.4784	NORMA	48.183488	11.478840	Supermarket

4. Methodology

4.1 Folium

Folium is being used to get the map of City Munich and the venues around Muechen Hauptbahnhof.



4.2 K-Means Clustering

By Clustering and visualize the data, it is very easy for users to check where is the better place to start a restaurant.

```
# set number of clusters
kclusters = 5

muc_grouped_clustering = muc_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(muc_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]

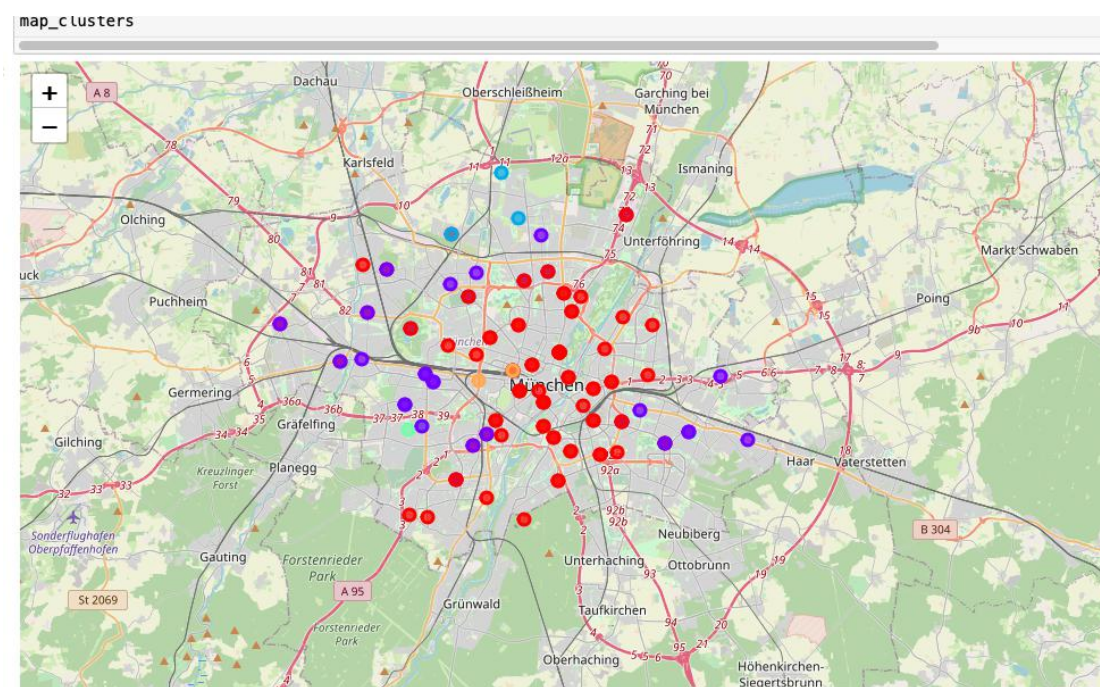
array([0, 0, 0, 0, 1, 0, 2, 3, 1, 0], dtype=int32)

# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels00', kmeans.labels_)

muc_merged = muc

# merge manhattan_grouped with manhattan_data to add latitude/longitude for each neighborhood
muc_merged = muc_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='District')

muc_merged.head() # check the last columns!
```



5. Results

The neighborhoods are divided into n clusters where n is the number of clusters found using the optimal approach. The clustered neighborhoods are visualized using different colors so as to make them distinguishable.

6. Discussion

The result shows that in the Altstadt (old city) of Munich, the most common venues are ice-cream shops, museums, and Restaurant, which reflects the reality that it is currently one of the most popular tourist place in the city. In cluster2, the most common venues changes from restaurant to Supermarket, which indicates that those are more liveing areas for the citizens. Outer than cluster2, the most common venues change to Bus Stop and Hotel, which gives the sign that those places are where most companies are located so there're more visits in Bus Stop (for working) and Hotel (business trips). The analyse around München Hauptbahnhof isn't very well as expected though, which could be due to less of information and limited area.

The critical part of the project is where to find the geo data and how to combine different source of data together for further analysis.

7. Conclusion

So now the neighborhoods in Munich are being clustered. For users who want to open a restaurant, it is better to open it in the inner city or near to the living communities, where the potential visits could be higher due to tourism and the citizen people. However, the analysis is only based on single source, there may have other factors which could affect the decision as well.

Cluster 1

```
muc_merged.loc[muc_merged['Cluster Labels00'] == 0, muc_merged.columns[[1] + list(range(5, muc_merged.shape[1]))]]
```

	Postal Code	Cluster Labels01	Cluster Labels	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	0	0	Italian Restaurant	Café	Bakery	Bus Stop	Supermarket	Bar	Sporting Goods Shop	German Restaurant	Plaza	Ice Cream Shop
1	80997	0	0	Italian Restaurant	Café	Bakery	Bus Stop	Supermarket	Bar	Sporting Goods Shop	German Restaurant	Plaza	Ice Cream Shop
2	80999	0	0	Italian Restaurant	Café	Bakery	Bus Stop	Supermarket	Bar	Sporting Goods Shop	German Restaurant	Plaza	Ice Cream Shop
3	81247	0	0	Italian Restaurant	Café	Bakery	Bus Stop	Supermarket	Bar	Sporting Goods Shop	German Restaurant	Plaza	Ice Cream Shop
4	81249	0	0	Italian Restaurant	Café	Bakery	Bus Stop	Supermarket	Bar	Sporting Goods Shop	German Restaurant	Plaza	Ice Cream Shop
5	80331	0	0	Café	Hotel	Italian Restaurant	Plaza	Bar	Middle Eastern Restaurant	Cocktail Bar	Art Museum	Burger Joint	Restaurant

Cluster 2

```
muc_merged.loc[muc_merged['Cluster Labels00'] == 1, muc_merged.columns[[1] + list(range(5, muc_merged.shape[1]))]]
```

	Postal Code	Cluster Labels01	Cluster Labels	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
22	81671	1	1	Supermarket	Bakery	Drugstore	Hotel	Metro Station	Bus Stop	Asian Restaurant	Italian Restaurant	Soccer Field	Bavarian Restaurant
23	81673	1	1	Supermarket	Bakery	Drugstore	Hotel	Metro Station	Bus Stop	Asian Restaurant	Italian Restaurant	Soccer Field	Bavarian Restaurant
24	81735	1	1	Supermarket	Bakery	Drugstore	Hotel	Metro Station	Bus Stop	Asian Restaurant	Italian Restaurant	Soccer Field	Bavarian Restaurant
25	81825	1	1	Supermarket	Bakery	Drugstore	Hotel	Metro Station	Bus Stop	Asian Restaurant	Italian Restaurant	Soccer Field	Bavarian Restaurant

Cluster 3

```
muc_merged.loc[muc_merged['Cluster Labels00'] == 2, muc_merged.columns[[1] + list(range(5, muc_merged.shape[1]))]]
```

	Postal Code	Cluster Labels01	Cluster Labels	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
32	80933	2	2	Bus Stop	Supermarket	Indian Restaurant	Pizza Place	Gastropub	Business Service	Food & Drink Shop	Beer Garden	Bakery	Greek Restaurant
33	80935	2	2	Bus Stop	Supermarket	Indian Restaurant	Pizza Place	Gastropub	Business Service	Food & Drink Shop	Beer Garden	Bakery	Greek Restaurant
34	80995	2	2	Bus Stop	Supermarket	Indian Restaurant	Pizza Place	Gastropub	Business Service	Food & Drink Shop	Beer Garden	Bakery	Greek Restaurant