Coursera - IBM Data Science Capstone Project

1. Introduction

Munich is famous for its beergardens (german "Biergarten"). A beergarden is an outdoor place in which beer and food is served at shared tables. Drinks have to be bought at place but food can be brought. Especially during summer evenings beergardens are a nice place to stay and enjoy a drink with friends and family.

The business problem discussed in the following is aiming to find the best district in Munich to open a new beergarden. Data science techniques are used to determine the right spot.

2. Data

2.1. Data

In order to find the best district to open a beergarden the first step is to gather information on the districts of Munich. This can be done by scraping the Wikipedia page of Munich's districts (https://de.wikipedia.org/wiki/Stadtbezirke_M%C3%BCnchens). Here we find the name and the number of inhabitants per each district.

2.2. Geospatial data

For a better graphical interpretation of the results a map is drawn using the district borders of Munich. These borders are derived using GeoJSON data from https://stekhn.carto.com/tables/munich/public/map.

2.3. Venue data

In order to get an impression of the restaurant venues of each district of Munich and to get an impression of the number an location of beergardens in Munich the Foursquare API is queried.

3. Methodology

In order to find the optimal spot to open a new beergarden the following two separate logics are applied:

- K-Means Clustering: K-Means Clustering is applied based on the venue data retrieved from Foursquare. The type of venues in a certain district is used as input clustering districts having similar restaurants. It is assumed that this is an indicator for the restaurant preference of the inhabitants of a certain district.
- Relative frequency: In order to calculate the coverage of beergardens per district a KPI is calculated representing the number of beergardens per 1000 inhabitants of a district.

4. Result

The following overview shows the density of beergardens per inhabitants. The density is simply calculated as a ratio between the number of beergardens and the inhabitants of a district.

	District	Cluster Labels	Inhabitants	Count_beergardens	Beergaden_per_10K
0	Aubing-Lochhausen-Langwied	2	47.813	1	0.2
1	Hadern	2	49.898	1	0.2
2	Ramersdorf-Perlach	1	116.327	2	0.2
3	Milbertshofen-Am Hart	0	75.094	2	0.3
4	Pasing-Obermenzing	3	74.625	3	0.4
5	Thalkirchen-Obersendling-Forstenried-Fürstenri...	3	96.714	4	0.4
6	Obergiesing-Fasangarten	3	54.256	2	0.4
7	Trudering-Riem	0	73.206	3	0.4
8	Feldmoching-Hasenbergl	3	61.774	3	0.5
9	Bogenhausen	0	87.950	5	0.6
10	Allach-Untermenzing	2	33.355	2	0.6
11	Neuhausen-Nymphenburg	0	98.814	6	0.6
12	Berg am Laim	0	46.098	3	0.7
13	Laim	0	56.546	4	0.7
14	Moosach	2	54.223	4	0.7
15	Sendling-Westpark	4	59.643	6	1.0
16	Schwabing-Freimann	0	77.936	9	1.2
17	Untergiesing-Harlaching	3	53.184	10	1.9
18	Schwabing-West	1	68.527	13	1.9
19	Sendling	4	40.983	8	2.0
20	Au-Haidhausen	1	61.356	17	2.8
21	Maxvorstadt	1	51.402	22	4.3
22	Ludwigsvorstadt-Isarvorstadt	1	51.644	24	4.6
23	Schwanthalerhöhe	1	29.743	16	5.4
24	Altstadt-Lehel	1	21.100	20	9.5

A beergarden serves Bavarian / german food. If we analyse the clusters generated by the K-Means algorithm it can be seen that cluster 1 shows the highest similarity to german / bavarian food. This means that districts belonging to Cluster 1 have most commonly German and Bavarian restaurants. It is assumed that the inhabitants of these districts prefer german or Bavarian food. Please find below a list of all districts of Cluster 1

	District	Inhabitants	District_Geopy	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Altstadt-Lehel	21.100	München Altstadt-Lehel	1	Café	German Restaurant	Italian Restaurant	Bavarian Restaurant	Restaurant
1	Ludwigsvorstadt-Isarvorstadt	51.644	München Ludwigsvorstadt-Isarvorstadt	1	Café	German Restaurant	Bavarian Restaurant	Italian Restaurant	Pizza Place
2	Maxvorstadt	51.402	München Maxvorstadt	1	Café	German Restaurant	Italian Restaurant	Restaurant	Bakery
3	Schwabing-West	68.527	München Schwabing-West	1	Café	Italian Restaurant	Vietnamese Restaurant	Greek Restaurant	Restaurant T
4	Au-Haidhausen	61.356	München Au-Haidhausen	1	Café	German Restaurant	Italian Restaurant	Bavarian Restaurant	Restaurant
7	Schwanthalerhöhe	29.743	München Schwanthalerhöhe	1	Café	German Restaurant	Asian Restaurant	Italian Restaurant	Bavarian Restaurant
15	Ramersdorf-Perlach	116.327	München Ramersdorf-Perlach	1	Greek Restaurant	German Restaurant	Italian Restaurant	Café	Bavarian Restaurant

5. Discussion

It is clear that this analysis is strongly simplified and shows some bias. Nevertheless it is meant to apply data science techniques in practice. Simplifications include:

- Venue information from Foursquare is only derived in a radius around the center of the districts and not all venues of that district
- The cluster generated by the K-Means algorithm show a most common venue of "german restaurant" in many
- An investments would most probably also depend on the prices of each district

6. Conclusion

If we combine the results of both analyses it can be a good idea to open a new beergarden in "Ramersdorf-Perlach" because the density of existing beergardens is low but inhabitants prefer german or Bavarian food. The density of beergardens per inhabitants is very low for this district so it could make sense for investors to open a beergarden in this district.