

IBM Data Science Capstone Project

Final Report

1. Introduction

1.1 Problem Description

For this project I have chosen to investigate the districts in Stuttgart to find out where there are good places to place an open air restaurant.

1.2 Background

A good place to place a open air restaurant (german Biergarten) might be a location with high throughput of people locals as well as tourists and might be marked as a place with several parks in urban area and some other outdoor takeaway stuff like cafes or icecream shops etc. Another option could be also that there are several metro stations nearby.

2. Data

1. Wikipedia for getting the Districts of Stuttgart
https://de.wikipedia.org/wiki/Liste_der_Stadtbezirke_und_Stadtteile_von_Stuttgart
2. Geocoder to get Locationdata of Districts
3. Foursquare API to get venues in that area
<https://api.foursquare.com/v2/venues/explore>

3. Methodology

3.1 Short description

Fetching the top venues per district and then clustering the districts by applying K-Means algorithm. Then interpret the clusters and draw a conclusion which cluster might fit best.

3.2 Analysis

Fetching Districts from Wikipedia

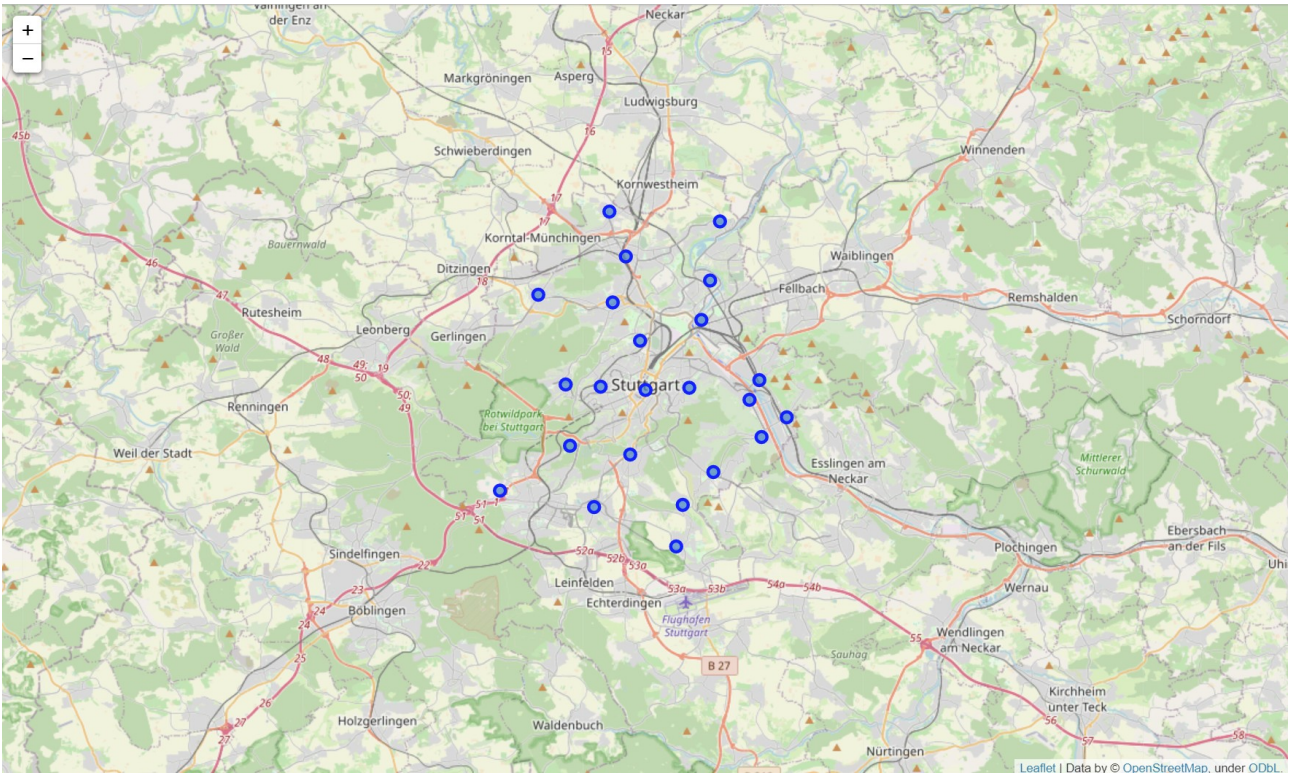
	Number	Name	People	Size	People/Size	Districts
0	Innerer Stadtbezirk Nr. 1	Stuttgart-Mitte	23.956	3808	6.294	10
1	Innerer Stadtbezirk Nr. 2	Stuttgart-Nord	27.629	6815	4.054	11
2	Innerer Stadtbezirk Nr. 3	Stuttgart-Ost	48.730	9035	5.393	8
3	Innerer Stadtbezirk Nr. 4	Stuttgart-Süd	44.050	9586	4.594	7
4	Innerer Stadtbezirk Nr. 5	Stuttgart-West	52.668	18643	2.825	9
5	Äußerer Stadtbezirk Nr. 1	Bad Cannstatt	71.285	15713	4.537	18
6	Äußerer Stadtbezirk Nr. 2	Birkach	7.149	3089	2.383	3
7	Äußerer Stadtbezirk Nr. 3	Botnang	13.165	2135	6.166	4
8	Äußerer Stadtbezirk Nr. 4	Degerloch	16.686	8021	2.081	5
9	Äußerer Stadtbezirk Nr. 5	Feuerbach	30.417	11554	2.645	8
10	Äußerer Stadtbezirk Nr. 6	Hedelfingen	10.296	7324	1.406	4
11	Äußerer Stadtbezirk Nr. 7	Möhringen	33.094	15044	2.200	9
12	Äußerer Stadtbezirk Nr. 8	Mühlhausen	25.675	9120	2.815	5
13	Äußerer Stadtbezirk Nr. 9	Münster	6.796	2210	3.075	1
14	Äußerer Stadtbezirk Nr. 10	Obertürkheim	8.648	5462	1.583	2
15	Äußerer Stadtbezirk Nr. 11	Plieningen	13.397	13075	1.025	5
16	Äußerer Stadtbezirk Nr. 12	Sillenbuch	24.067	7457	3.230	3
17	Äußerer Stadtbezirk Nr. 13	Stammheim	12.502	4334	2.881	2
18	Äußerer Stadtbezirk Nr. 14	Untertürkheim	16.694	6055	2.757	8
19	Äußerer Stadtbezirk Nr. 15	Vaihingen	48.835	20893	2.194	12
20	Äußerer Stadtbezirk Nr. 16	Wangen	9.255	3426	2.701	1
21	Äußerer Stadtbezirk Nr. 17	Weilimdorf	31.873	12585	2.530	6
22	Äußerer Stadtbezirk Nr. 18	Zuffenhausen	38.676	11963	3.234	11

There are 5 inner districts and 18 outer districts next step is to enrich those data with the coordinates with Geocoder. At first try there have been some district mislocated because of name duplicates in Germany. Solution was to prepend the outer areas with „Stuttgart“ which led to the following Dataframe.

	Number	Name	People	Size	People/Size	Districts	longitude	latitude
0	Innerer Stadtbezirk Nr. 1	Stuttgart-Mitte	23.956	3808	6.294	10	9.179800	48.775900
1	Innerer Stadtbezirk Nr. 2	Stuttgart-Nord	27.629	6815	4.054	11	9.176252	48.796661
2	Innerer Stadtbezirk Nr. 3	Stuttgart-Ost	48.730	9035	5.393	8	9.207365	48.776972
3	Innerer Stadtbezirk Nr. 4	Stuttgart-Süd	44.050	9586	4.594	7	9.132492	48.753021
4	Innerer Stadtbezirk Nr. 5	Stuttgart-West	52.668	18643	2.825	9	9.151351	48.777659
5	Äußerer Stadtbezirk Nr. 1	Bad Cannstatt	71.285	15713	4.537	18	9.214680	48.804883
6	Äußerer Stadtbezirk Nr. 2	Birkach	7.149	3089	2.383	3	9.203406	48.728574
7	Äußerer Stadtbezirk Nr. 3	Botnang	13.165	2135	6.166	4	9.129532	48.778495
8	Äußerer Stadtbezirk Nr. 4	Degerloch	16.686	8021	2.081	5	9.170345	48.749597
9	Äußerer Stadtbezirk Nr. 5	Feuerbach	30.417	11554	2.645	8	9.159031	48.812305
10	Äußerer Stadtbezirk Nr. 6	Hedelfingen	10.296	7324	1.406	4	9.252665	48.756436
11	Äußerer Stadtbezirk Nr. 7	Möhringen	33.094	15044	2.200	9	9.147792	48.727816
12	Äußerer Stadtbezirk Nr. 8	Mühlhausen	25.675	9120	2.815	5	9.226179	48.845610
13	Äußerer Stadtbezirk Nr. 9	Münster	6.796	2210	3.075	1	9.220002	48.821296
14	Äußerer Stadtbezirk Nr. 10	Obertürkheim	8.648	5462	1.583	2	9.268515	48.764789
15	Äußerer Stadtbezirk Nr. 11	Plieningen	13.397	13075	1.025	5	9.198991	48.711395
16	Äußerer Stadtbezirk Nr. 12	Sillenbuch	24.067	7457	3.230	3	9.222216	48.742184
17	Äußerer Stadtbezirk Nr. 13	Stammheim	12.502	4334	2.881	2	9.157305	48.849798
18	Äußerer Stadtbezirk Nr. 14	Untertürkheim	16.694	6055	2.757	8	9.250900	48.780200
19	Äußerer Stadtbezirk Nr. 15	Vaihingen	48.835	20893	2.194	12	9.088648	48.734429
20	Äußerer Stadtbezirk Nr. 16	Wangen	9.255	3426	2.701	1	9.244927	48.771985
21	Äußerer Stadtbezirk Nr. 17	Weilimdorf	31.873	12585	2.530	6	9.112444	48.815276
22	Äußerer Stadtbezirk Nr. 18	Zuffenhausen	38.676	11963	3.234	11	9.167440	48.831299

A first check on the map now shows the all districts in the Stuttgart area.

Visual Representation



Fetching Venues

After that fetching of the venues with a radius of 3km is done from Foursquare API with a maximum of 100 Venues per district which results in the following number of venues

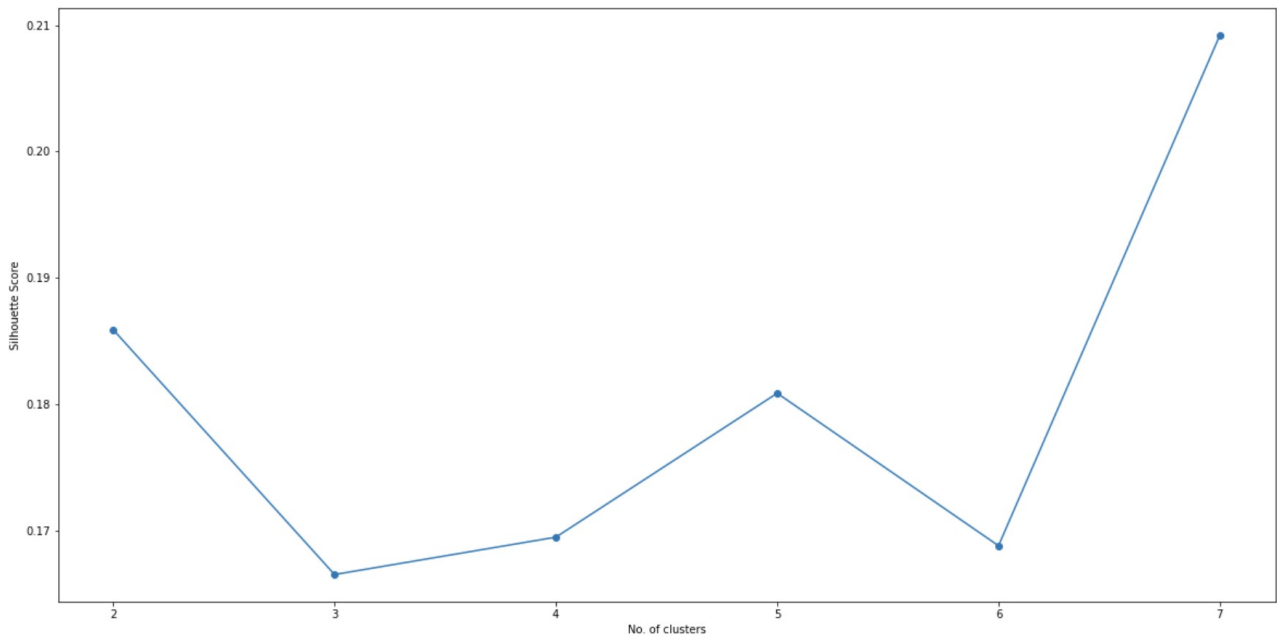
	Venue	Latitude	Longitude	Category
Name				
Bad Cannstatt	100	100	100	100
Birkach	36	36	36	36
Botnang	40	40	40	40
Degerloch	100	100	100	100
Feuerbach	69	69	69	69
Hedelfingen	51	51	51	51
Möhringen	94	94	94	94
Mühlhausen	31	31	31	31
Münster	76	76	76	76
Obertürkheim	47	47	47	47
Plieningen	58	58	58	58
Sillenbuch	39	39	39	39
Stammheim	61	61	61	61
Stuttgart-Mitte	100	100	100	100
Stuttgart-Nord	100	100	100	100
Stuttgart-Ost	100	100	100	100
Stuttgart-Süd	86	86	86	86
Stuttgart-West	100	100	100	100
Untertürkheim	62	62	62	62
Vaihingen	65	65	65	65
Wangen	69	69	69	69
Weilimdorf	35	35	35	35
Zuffenhausen	52	52	52	52

Calculating Most common Venues

	Name	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	Bad Cannstatt	Park	Café	Zoo Exhibit	Hotel	German Restaurant	Restaurant	Pool	Italian Restaurant	Beer Garden	Taverna
1	Birkach	Metro Station	Hotel	Italian Restaurant	German Restaurant	Garden	Office	Restaurant	Coffee Shop	Climbing Gym	Chinese Restaurant
2	Botnang	Italian Restaurant	Supermarket	Park	Plaza	Café	Gym / Fitness Center	Sushi Restaurant	Mountain	Cable Car	Business Service
3	Degerloch	Café	Bar	German Restaurant	Cocktail Bar	Hotel	French Restaurant	Ice Cream Shop	Italian Restaurant	Park	Bakery
4	Feuerbach	Italian Restaurant	Bakery	Burger Joint	Greek Restaurant	Chinese Restaurant	Plaza	Supermarket	Beer Garden	Museum	Scenic Lookout

Calculating ideal number of Clusters

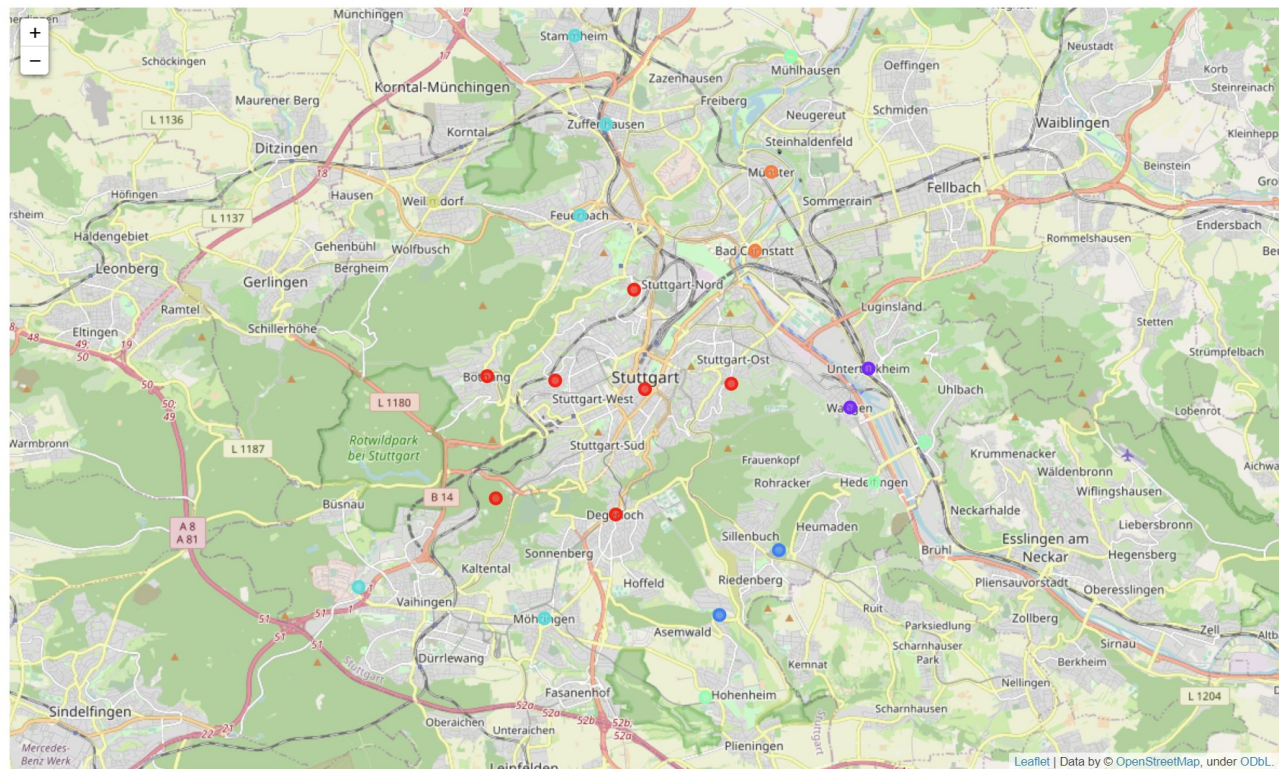
To Calculate the ideal number of clusters I use silhouette_score which results in ideal number of clusters as 7.



Applying K-Means

	Cluster Labels	Name	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	6	Bad Cannstatt	Park	Café	Zoo Exhibit	Hotel	German Restaurant	Restaurant	Pool	Italian Restaurant	Beer Garden	Taverna
1	2	Birkach	Metro Station	Hotel	Italian Restaurant	German Restaurant	Garden	Office	Restaurant	Coffee Shop	Climbing Gym	Chinese Restaurant
2	0	Botnang	Italian Restaurant	Supermarket	Park	Plaza	Café	Gym / Fitness Center	Sushi Restaurant	Mountain	Cable Car	Business Service
3	0	Degerloch	Café	Bar	German Restaurant	Cocktail Bar	Hotel	French Restaurant	Ice Cream Shop	Italian Restaurant	Park	Bakery
4	3	Feuerbach	Italian Restaurant	Bakery	Burger Joint	Greek Restaurant	Chinese Restaurant	Plaza	Supermarket	Beer Garden	Museum	Scenic Lookout
5	4	Hedelfingen	Supermarket	Hotel	German Restaurant	Italian Restaurant	Gym / Fitness Center	Metro Station	Fast Food Restaurant	Concert Hall	Convenience Store	Burger Joint

Visual Representation of Clusters



4. Results

Cluster 1

Cluster 1 represents mostly parks, cafes and other leisure time activities like ice-cream shops and bars this is typical urban area and an ideal location for a biergarten.

Cluster Labels	Name	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
2	0 Botnang	Italian Restaurant	Supermarket	Park	Plaza	Café	Gym / Fitness Center	Sushi Restaurant	Mountain	Cable Car	Business Service
3	0 Degerloch	Café	Bar	German Restaurant	Cocktail Bar	Hotel	French Restaurant	Ice Cream Shop	Italian Restaurant	Park	Bakery
13	0 Stuttgart-Mitte	Café	Park	Bar	Plaza	Ice Cream Shop	Italian Restaurant	Hotel	Cocktail Bar	Scenic Lookout	Coffee Shop
14	0 Stuttgart-Nord	Café	Plaza	Park	Italian Restaurant	Hotel	Bar	Restaurant	Scenic Lookout	Ice Cream Shop	Concert Hall
15	0 Stuttgart-Ost	Park	Bar	Plaza	Cocktail Bar	Hotel	Italian Restaurant	Café	Ice Cream Shop	Restaurant	Nightclub
16	0 Stuttgart-Süd	Supermarket	Italian Restaurant	German Restaurant	Park	Restaurant	Bakery	Ice Cream Shop	Bar	Thai Restaurant	Plaza
17	0 Stuttgart-West	Café	Plaza	Ice Cream Shop	Italian Restaurant	Park	Sushi Restaurant	Bar	Cocktail Bar	Coffee Shop	Hotel

Cluster 2

Cluster 2 is some kind of suburban area with Hotels and Restaurants as well as Sports locations.

Cluster Labels	Name	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
18	1 Untertürkheim	German Restaurant	Hotel	Soccer Field	Greek Restaurant	Fast Food Restaurant	Stadium	Restaurant	Supermarket	Pool	Concert Hall
20	1 Wangen	German Restaurant	Hotel	Supermarket	Bakery	Restaurant	Museum	Fast Food Restaurant	Lounge	Soccer Field	Pool

Cluster 3

Cluster 3 is also outside of the city with metro stations and hotels but with the offices etc more likely an area with many business people.

	Cluster Labels	Name	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	2	Birkach	Metro Station	Hotel	Italian Restaurant	German Restaurant	Garden	Office	Restaurant	Coffee Shop	Climbing Gym	Chinese Restaurant
11	2	Sillenbuch	Italian Restaurant	Metro Station	Supermarket	Hotel	Soccer Field	German Restaurant	Bakery	Music Venue	Climbing Gym	Middle Eastern Restaurant

Cluster 4

Cluster 4 is also suburban area in which many people live indicated by supermarkets.

	Cluster Labels	Name	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	3	Feuerbach	Italian Restaurant	Bakery	Burger Joint	Greek Restaurant	Chinese Restaurant	Plaza	Supermarket	Beer Garden	Museum	Scenic Lookout
6	3	Möhringen	Hotel	Supermarket	Italian Restaurant	Gym / Fitness Center	German Restaurant	Bakery	Greek Restaurant	Metro Station	Drugstore	Asian Restaurant
12	3	Stammheim	Hotel	Supermarket	Metro Station	Bakery	Drugstore	Burger Joint	Coffee Shop	Garden Center	Italian Restaurant	Chinese Restaurant
19	3	Vaihingen	German Restaurant	Hotel	Bakery	Asian Restaurant	Italian Restaurant	Indian Restaurant	Gastropub	Restaurant	Sandwich Place	Supermarket
22	3	Zuffenhausen	Drugstore	Bakery	Burger Joint	Italian Restaurant	Chinese Restaurant	German Restaurant	Hotel	Greek Restaurant	Plaza	Pet Store

Cluster 5

Cluster 5 is also an area in which the people live but less tourists travel to thats an area mainly for locals also indicated by relativ common supermarkets

neighborhoods_venues_sorted[neighborhoods_venues_sorted['Cluster Labels']==4]												
	Cluster Labels	Name	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
5	4	Hedelfingen	Supermarket	Hotel	German Restaurant	Italian Restaurant	Gym / Fitness Center	Metro Station	Fast Food Restaurant	Concert Hall	Convenience Store	Burger Joint
7	4	Mühlhausen	Hotel	Supermarket	Metro Station	German Restaurant	Drugstore	Restaurant	Gym / Fitness Center	Gastropub	Bakery	Fruit & Vegetable Store
9	4	Obertürkheim	German Restaurant	Supermarket	Hotel	Gym / Fitness Center	Italian Restaurant	Fast Food Restaurant	History Museum	Shopping Mall	Lounge	Spa
10	4	Plieningen	Hotel	Supermarket	German Restaurant	Airport Service	Italian Restaurant	Theater	Café	Garden	Multiplex	Electronics Store

Cluster 6

Also a cluster for locals who travel to their workplaces with metro and shop in the supermarkets.

	Cluster Labels	Name	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
21	5	Weilimdorf	Supermarket	Bakery	Italian Restaurant	Drugstore	Metro Station	Mobile Phone Shop	Big Box Store	Bowling Alley	Shopping Mall	Discount Store

Cluster 7

This cluster is also for leisure time activities with Zoos, Parks etc.

	Cluster Labels	Name	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	6	Bad Cannstatt	Park	Café	Zoo Exhibit	Hotel	German Restaurant	Restaurant	Pool	Italian Restaurant	Beer Garden	Taverna
8	6	Münster	German Restaurant	Zoo Exhibit	Metro Station	Park	Greek Restaurant	Taverna	Grocery Store	Hotel	German Pop-Up Restaurant	Miscellaneous Shop

Discussion and Recommendation

Based on the clusters I would have two recommendations if it is mainly a beergarden for locals and families I would place it in cluster 7. If it is mainly a beergarden with focus on business people and tourists I would place it in cluster 1 because of the high throughput in city center and the parks. The zoos etc attract more the families with children.

References:

The Jupyter Notebook of the analysis can be found on Github.

https://github.com/tobiasgrott/Coursera_Applied_Data_Science_Capstone/blob/master/final_project/FinalProject.ipynb