IBM Data Science Capstone Project
Problem Definition and Data

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.

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 som other outdoor takeaway stuff like cafes or icecream shops etc. An other option could be also that there are several metro stations nearby.

Target Audience

The Target audience are businessman evaluating opening a new restaurant as well as real estate agents and investors. Those people care about the value of a specific location for opening an open air restaurant. Depending on their focus it could be expected revenue of a beergarden or the amount of value an object has for getting a fair amount of rent.

Data:

1. Wikipedia for getting the Districts of Stuttgart
2. Geocoder to get Locationdata of Districts
3. Foursquare API to get venues in that area

1. Wikipedia

https://de.wikipedia.org/wiki/Liste_der_Stadtbezirke_und_Stadtteile_von_Stuttgart

Liste der Stadtbezirke [Bearbeiten | Quelltext bearbeiten]

Die 23 Stuttgarter Stadtbezirke lassen sich in 5 innere und 18 äußere Bezirke einordnen. Dabei entspricht das innere Stadtgebiet weitgehend der Stuttgarter Markun widerspiegelt. Die Stadtbezirke wiederum sind in einzelne Stadtteile untergliedert. Dabei gibt es Bezirke mit nur einem Stadtteil bis zu Bezirken mit achtzehn Stadttei

Nr.	Stadtbezirk	Einwohner	Fläche (ha)	Bevölkerungsdichte (Einwohner/km²)	Anzahl der Stadtteile	Lage
Innerer Stadtbezirk Nr. 1	Stuttgart-Mitte	23.956	380,8	6.294	10	
Innerer Stadtbezirk Nr. 2	Stuttgart-Nord	27.629	681,5	4.054	11	
Innerer Stadtbezirk Nr. 3	Stuttgart-Ost	48.730	903,5	5.393	8	
Innerer Stadtbezirk Nr. 4	Stuttgart-Süd	44.050	958,6	4.594	7	
Innerer Stadtbezirk Nr. 5	Stuttgart-West	52.668	1864,3	2.825	9	
Äußerer Stadtbezirk Nr. 1	Bad Cannstatt	71.285	1571,3	4.537	18	
Äußerer Stadtbezirk Nr. 2	Birkach	7.149	308,9	2.383	3	
Äußerer Stadtbezirk Nr. 3	Botnang	13.165	213,5	6.166	4	
Äußerer Stadtbezirk Nr. 4	Degerloch	16.686	802,1	2.081	5	

	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

The relevant feature in the list of districts is the name this data is used for getting all districts of Stuttgart

2. Geocoder

	Name	longitude	latitude
0	Stuttgart-Mitte	9.179800	48.775900
1	Stuttgart-Nord	9.176252	48.796661
2	Stuttgart-Ost	9.207365	48.776972
3	Stuttgart-Süd	9.132492	48.753021
4	Stuttgart-West	9.151351	48.777659
5	Bad Cannstatt	9.214680	48.804883
6	Birkach	9.203406	48.728574

The relevant Feature here is the location (longitude and latitude) for each prior extracted district to have a center point per district for fetching afterwards venue data.

3. Foursquare API

	Name	Venue	Latitude	Longitude	Category
0	Stuttgart-Mitte	Schlossplatz	48.778549	9.179855	Plaza
1	Stuttgart-Mitte	Old Bridge	48.774173	9.179274	Ice Cream Shop
2	Stuttgart-Mitte	Markthalle	48.776145	9.179335	Market
3	Stuttgart-Mitte	Bix Jazzclub	48.773178	9.179495	Jazz Club
4	Stuttgart-Mitte	Feinkost Böhm	48.778077	9.176317	Gourmet Shop

From the Foursquare API there are many possible features in that analyis I extract the Category of the venues which are in a specific radius of a distrcit center so relevant features are name of district and category.

Idea to solve the Problem:

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