Business Opportunities of opening a New Restaurant in Cologne

By: Sandeepan Mukherjee



The Restaurant Battle of Neighborhoods in Cologne

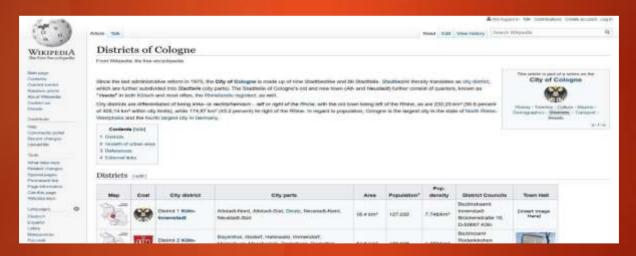
Cologne is a city near which the author lives nearby, and it is well-known for its cathedral, trade fairs and conferences, shopping boulevard, and lively party scene. Since it draws a massive amount of visitors, there is a large business potential for restaurants because there is never a lack of people from all over Europe and the world visiting the beautiful region. So ther is wide variety of cousines and restaurants one can chose to start their business venture into. Thus, the goal I want to reach with this exercise is to give a simple recommendation to businesses and stakeholders who are looking to open a new restaurant in Cologne and solve issues like in which district of the city will you find a large number or even concentration of which types of restaurants? Where to open a Mediterranean food, German food, or where to get fast food and the list of competitors in that area. The target audience is Food Entrepreneurs and Business owners.

Data Preparation

As mentioned in the project requirements, I will use foursquare data about restaurants in Cologne. Foursquare is a US tech company from New York focusing on location data. Their technology and data powers apps such as Apple's Maps, Uber, Twitter and many other household names. Here is an example of a restaurants in Cologne on foursquare: https://de.foursquare.com/v/sattgr%C3%BCn/5c33306cc824ae002c2b414c. I will use foursquare data such as the restaurant name, ID, location and category of food (vegetarian, Italian etc.).

Also, I will use the overview of districts/city parts of Cologne from Wikipedia: https://en.wikipedia.org/wiki/Districts_of_Cologne

Based on this criterion, we will use Data science Python libraries to create a few promising neighborhoods. The benefits of each region can then be specifically articulated so that stakeholders will choose the best possible final spot.



Data acquisition and Preproceesing

We will do data preparation and collection starting with scraping of the Wikipedia page districts of Cologne: https://en.wikipedia.org/wiki/Districts of Cologne. To create a pandas dataframe and then perform data cleaning task to make the results look presentable and drop and remove inconsistencies in the dataframe

Remove any Boroughs that are not Assigned

- 1	City district	City parts	Area	Population1	Pop. density	District Councils
0	Köln-innenstadt	Altstadt-Nord, Altstadt-Süd, Deutz, Neustadt-N	16.4 km²	127.033	7.746/km²	Bezirksksamt Innenstadt Brückenstraße 19, D-50
1	Köln-Rodenkirchen	Bayenthal, Godorf, Hahnwald, Immendorf, Marien	54.6 km*	100.936	1.850/km/*	Bezirksamt Rodenkirchen Hauptstraße 85, D-5099
2	Koin-Lindenthai	Braunsfeld, Junkersdorf, Klettenberg, Lindenth	41.6 km²	137.552	3.30B/km²	Bezirksamt Lindenthal Aachener Straise 220, 509
3	Kdin-Ehrenfeld	Bickendorf, Bocklemund/Mengenich, Ehrenfeld, N.,	23 8 km²	103.621	4.34B/km²	Bezirksamt Ehrenfeld Veniber Straße 419 – 421
4	Kom-Nippes	Bilderstöckchen, Longerich, Mauenheim, Niehl,	31.8 km²	110.092	3.462/km²	Bezirksamt NippesNeusser Straße 450 D-50733 Kolin
5	Koln-Chorweller	Blumenberg, Chorweiler, Esch/Auweiler, Fühling	67.2 km²	80.870	1.204/km*	Bezirksami Chorweiler Pariser Platz 1, D-50765
6	Köin-Porz	Ell, Elsdorf, Ensen, Finkenberg, Gremberghoven	78.8 km²	106 520	1.352/km²	Bezirksamt PorzFriedrich-Ebert-Lifer 64–70, D-5
7	Köln-Kalk	Brück, Höhenberg, Humboldl/Gremberg, Kalk, Mer.	38.2 km²	108,330	2.841/km²	Bezirksamt KalkKalker Hauptstraße 247–273,0-51
8	Köln-Mülheim	Buchforst, Buchheim, Delbrück, Dünnwald, Filt	52.2 km²	144.374	2.764/km²	Bezirksamt Mülheim Wiener Platz 2a,0-51065 Köln

Then, we enabled geopy and used the nominatim function to add geospatial data to the data frame, that is the latitude and the longitude seen on the right side of the following table.

	City district	City parts	Area	Population1	Pop. density	District Councils	Latitude	Longitude
0	Köln-Innenstadl	Altstadt-Nord, Altstadt-Süd, Deutz, Neustadt-N.	16.4 km²	127,033	7.746/km/	Bezirksksamt innenstadt Brückenstraße 19, D-50	50:937328	6.959234
1	Kom-Rodenkirchen	Bayenthat Godorf, Hahrweld, Immendorf, Marien	54.6 km²	100 936	1.850/km²	Bezinsamt Rodentischen Hauptstraße 85, D-5099	50.865622	6.969718
2	Kön-Lindenthai	Braunsfeld Juniumsdorf, Wettenberg, Lindenth	41.6 km²	137.562	3.306/km²	Bezitisant Lindenthal Aacherer Strade 220, 509	50.935935	6.871246
3	Köln-Ehrenfeld	Bickendorf, Bocklemand/Mengenich, Ehrenfeld, N.,	23.6 km²	103.621	4:348/km²	Bearksant Ehrenfeld Verrioer Straße 419 – 421	50.951502	6.916529
4	Köln-Nippes	Bilderstöckchen, Longerich, Mauenheim, Niehl,	31.6 km²	110.092	3.462/km²	Bearksant NeppesNeusser Strate 450.0-50733 Koln	50.958994	6.941777
	Köln-Charweller	Blurtenberg, Chorweller, EschiAuweiler, Fühling	67.2 km²	80.470	1.204/km²	Bezirksant Chorweler Parser Platz 1, 0-50765.	51.021167	6.695034
8	Köln-Parz	Eil Elsdorf, Erisen, Finkenberg, Gremberghaven	78.8 km²	106,520	1.352/km²	Sezirksamt PorzFnedrich-Ebert-Ufer 64–70, D-5	50.906705	6.999129
7	Krin-Kalk	Brück, Höhenberg, Humboldt/Gremberg, Kalk, Mer.	38.2 km²	108.330	2.841/km²	Bezirksamt KalkKalker Hauptstraße 247–273,D-61	56.931923	7.005806
	Köln-Mülheim	Buchforst, Buchheim, Deitbrück, Dünnwald, Filt.	52.2 km²	144.374	2.764/km²	Bezirksant Müthem Wener Platz 2a,0-51065 Köm	50.958147	7.013526

Visualization of city parts



Data Analysis

Now, foursquare data comes into play and first we found the nearby and common venues, grouped them and performed Clustering Using Kmeans

STEP 1: Specifically to find clusters of restaurant types in the different city districts we need to first transformed the data frame with the restaurant venues, associated to city districts, by one-hot encoding (0/1), as seen in the picture below.

17]:		district	Bavarian Restaurant	Restaurant	Res									
		Koks- Chorweller	0.00	1.0	8.0	0.00	0.00	0.080333	0.00	0.000000	6.00	8.00	0.0	
	,	Köln- Ehrenfeld	0.00			0.08	0.06	0.000000	0.00	0.080000	0.06	0.00	0.1	
	2	KSks- trenerotadi	0.08	0.0	0.0	0.00	0.00	0.050000	0.00	0.000000	0.00	0.00	0.0	
	1	KömiKale	0.00	30.0	0.0	0.00	0.00	0.000000	0.85	0.000000	11.00	0.00	0.0	
	4	Köln- Lindenthal	0.00		0.0	0.00	0.00	0.000000	0.00	0.111111	11.00	0.00	0.0	
	- 31													

STEP 2: After grouping

		American	Anima	Acceptance	Chinese	Comfort	Doner	Eastern	Palatal	Fast Food	Francis	German	Greek	Indian	Anthro		in the	Manager	
	Neighborhood	Restaurant	Restaurest	Restaurast	Restaurant	Pand Restaurant	Restaurant	Rectivitant	Restaurant	Restaurant	Restaurant	Restaurant	Revisorant	Restaurant	Restaurant	Restaurant	Restaurant	Restaurant	Re.
.0	NON-ENCOREM	0.000000	0.000000	0.900000	0.000000	8.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.335533	5,000000	0.000000	0.800000	0
1	tom-Ehrenbei	0,000000	0.000000	0.000000	0.006403	6,000000	0,000000	0.000000	0.000000	1.000000	0.030462	0.076900.	0.038462	0.000000	0.103846	0.09863	3.031.62	0.800068	6. 1
2	ISSN Werestart	0.000000	0.000000	9 930000	0.000000	8.009000	0.000093	0.000090	0.041907	4 300000	0.085555	0.065533	0.000000	0.000000	0.291667	0.000000	0.000000	0.00000	
3	HOW-KIRK	0.000000	0.041687	0.000000	E 0000000	BOASONT	0.041667	DOWNER	0.000000	0 900000	0.900000	0.041007	0.083333	0.041007	0.063333	0.000000	0.000000	0.941667	100
*	Witn Lindsoftwe	199110.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.006000	1 100000	0.041667	0.000000	0.001333	0.041007	0.001123	0.001657	9.000000	0.000000	6 1
	NON ANDRESS	0.008800	0.095258	0.000000	0.000000	8.047919	0.000000	0.000891	0.006900	0.000000	0.000000	0.047610	0.047610	9 (090000)	0.005238	(1000000)	E 000000	0.00000	
	KING Hopes	E 000000	0.00000	0.004616	81000000	8199000	0.030258	0.000000	0.000000	1 1000H	0.090774	0.084816	0.064516	4.103300	0.193546	0.000000	1112298	0.00000	1. 1
	Kittle-Press	0.008800	0.000000	0.000000	0.000000	8.000000	0.000000	0.000000	0.000000	0.000714	0.071429	0.200000	0.071420	0.000000	0.176671	0.000000	0.000000	0.000714	6 1
	Roser Author	0.000000	0.00000	13.000000	810000000	£ 000000	0.000683	0.000890	0.000000	0 000000	0.800000	0.000333	0.000000	0.000000	0.000000	0.000000	2 (900000)	0.100000	
43																			

STEP 3: After using the information we create a data frame in which you can see the most common restaurant venue types for each city district..

1]:		City	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	Köln- Chonweiler	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Scandinavian Restaurant	Restaurant	Lebanese Restaurant	Kebab Restaurant	talan Restaurant	Greek Restaurant	Fatafel Restaurant	Ethiopian Restaurant
8	3	Köln- Etvenfeld	Restaurant	Ketrab Restaurant	Italian Restaurant	Fatafel Restaurant	Ethiopian Restaurant	Vegetarian / Vegan Restaurant	Scandinavian Restaurant	Lebanese Restaurant	Greek Restaurant	Fast Food Restaurant
À	2	Köin- innenstadi	Vegetarian / Vegan Restaurant	Lebanese Restaurant	Fast Food Restaurant	Bavarian Restaurant	Scandinavian Restaurant	Restaurant	Kebab Restaurant	Italian Restaurant	Greek Restaurant	Fatafel Restaurant
	3	Kdo-Kak	Green Restaurant	Vegetarian / Vegan Restaurant	Scandinavian Restaurant	Restaurant	Lebanese Restaurant	Kebab Restaurant	Italian Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant
	4	Kötn- Lindenthai	Italian Restaurant	Vegetarian / Vegan Restaurant	Scandinavian Restaurant	Restaurant	Lebanese Restaurant	Kebab Restaurant	Greek Restaurant	Fast Food Restaurant	Falafei Restaurant	Ethiopian Restaurant

STEP 4: After this we would run a k-means clustering algorithm from the scikit-learn package. One could use the ellbow method to systematically define the k value, but I simply chose k to be 5, having been inspired by one of the coursera courses to do so.

	Cluster Labels	City district	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	2	Kötn- Chorweiler	Sushi Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Kebab Restaurant	Chinese Restaurant
1	3	Köln- Ehrenfeld	Italian Restaurant	Restaurant	Tapas Restaurant	Kebab Restaurant	Portuguese Restaurant
2	4	Köln- Innenstadt	Italian Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant	Mexican Restaurant	Chinese Restaurant
3	.1	Köln-Kalk	Greek Restaurant	Vegetarian / Vegan Restaurant	Turkish Restaurant	Chinese Restaurant	Doner Restaurant
4	0	Kotn- Lindenthal	Italian Restaurant	Vegetarian / Vegan Restaurant	Turkish Restaurant	Chinese Restaurant	Doner Restaurant

The table includes the city districts and their most popular venues, which have now been allocated one of five distinct cluster labels ranging from 0 to 4.

We can now use the cluster labels to show the city districts marked with a cluster-specific color on a map using the library folium:

Clusters view





What you see above is the nine bubbles for the nine city districts, with five different colors for the five different clusters.

Examination of Results

First cluster result shows Italian Cuisine Cluster in Lindenthal

	City parts	District Councils	Latitude	Longitude	Cluster	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 I Com Ve
2	Braunsfeld, Junkersdorf,	Bezirksamt Lindenthal Aachener Straße 220, 509	50.935935	6.871246	0.0	Italian Restaurant	/ Vegan	Scandinavian Restaurant	Restaurant	Lebanese Restaurant	Kebab Restaurant	Greek Restaurant	Fast F Restai

Second cluster result shows Vegan Cuisine Cluster in InnenStadt

	City parts	District Councils	Latitude	Longitude	Cluster Labels	1st Most Common Venue			4th Most Common Venue		6th Most Common Venue	7th Most Common Venue	8th Con
0	Altstadt- Nord, Altstadt- Süd, Deutz, Neustadt- N	Bezirksksamt Innenstadt Brückenstraße 19, D-50	50.937328	6.959234	1.0	Vegetarian / Vegan Restaurant	Lebanese Restaurant	Fast Food Restaurant	Bavarian Restaurant	Scandinavian Restaurant	Restaurant	Kebab Restaurant	Italiar Resta

Third cluster result shows Greek Cuisine Cluster in KalkKalker

	City parts	District Councils	Latitude	Longitude	Cluster	1st Most Common Venue	2nd Most Common Venue		Common	5th Most Common Venue	6th Most Common Venue	Commi
7	Brück, Höhenberg, Humboldt/Gremberg, Kalk, Mer	Bezirksamt KalkKalker Hauptstraße 247–273,D- 51	50.931923	7.005806	2.0	Greek Restaurant	Vegetarian / Vegan Restaurant	Scandinavian Restaurant	Restaurant	Lebanese Restaurant	Kebab Restaurant	Italian Restaura

Fourth cluster result shows Turkish Cuisine Cluster in Ehrenfeld

	City parts	District Councils	Latitude	Longitude	Cluster Labels	1st Most Common Venue	626	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th / Com Ve
3	Bickendorf, Bocklemünd/Mengenich, Ehrenfeld, N	Bezirksamt Ehrenfeld Venloer Straße 419 – 421,	50.951502	6.916529	3.0	Restaurant	Kebab Restaurant	Italian Restaurant	Falafel Restaurant	Ethiopian Restaurant	Vegetarian / Vegan Restaurant	Scandina Restaura

Fifth cluster result shows Thai Cuisine Cluster in Chorweiler

	City parts	District Councils	Latitude	Longitude	Cluster	1st Most Common Venue		3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	Co
5	Blumenberg, Chorweller, Esch/Auweiler, Fühling	Bezirksamt Chorweiler Pariser Platz 1, D- 50765	51.021167	6.898034	4.0	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Scandinavian Restaurant	Restaurant	Lebanese Restaurant	Kebab Restaurant	Italian Restaurant	Gree Res

Discussion and Conclusion

The aim of this project was to find Cologne areas near the center with a low number of restaurants in order to assist stakeholders in narrowing their quest for an optimal location for a new restaurant. We defined analysis (köln Innenstadt) and then created an extensive set of locations that satisfy some specific requirements about existing nearby restaurants by measuring restaurant density distribution from Foursquare data. Clustering of those areas was then conducted in order to establish major zones of interest, and addresses for those zone centers were generated to be used as starting points for stakeholders' final discovery.

Answer to Business Question:

Cluster 1 is relects that most preffered place where customers prefer Italian Cuisine but opening an Italian Restaurant In cluster 1 would also lead to sheer level of high competition. Whereas Cluster 2,3 and 5 have less footfall for Italian Cuisine but there is a huge business opportunity for good Italian Cuisine Restaurant.

Cluster 2 reflects that Vegetarian/Vegan Restaurant is most preffered by customers and visitors there and opening Vegan Cuisine restaurant in this cluster would lead to lot of competition but Cluster 4 would be the best place for stakeholders and business owners to open a good Vegan cuisine restaurant.

Cluster 3 reflects that Greek Restaurant is most preffered by customers and visitors there and opening Greek Cuisine restaurant in this cluster would lead to lot of competition but Cluster 2 would be the best place for stakeholders and business owners to open a good Vegan cuisine restaurant and Cluster 1,4 and 5 also looks to be a good place as well.

Cluster 4 reflects that Turkish Restaurant is most preffered by customers and visitors there and opening Turkish Cuisine restaurant in this cluster would lead to lot of competition but Cluster 2 would be the best place for stakeholders and business owners to open a good Turkish cuisine restaurant and Cluster 1, 3 and 5 also looks to be a good place as well.

Cluster 5 reflects that Thai Restaurant is most preffered by customers and visitors there and opening Thai Cuisine restaurant in this cluster would lead to lot of competition but Cluster 4 would be the best place for stakeholders and business owners to open a good Turkish cuisine restaurant.

Stakeholders will make the final decision on optimal restaurant position based on unique characteristics of communities and areas in each recommended district, taking into account additional factors such as attractiveness of each location, levels of noise / proximity to major connection roads, real estate, costs, social and economic complexities of each neighborhood.