Business Opportunities of opening a New Restaurant in Cologne

By: Sandeepan Mukherjee

The Restaurant Battle of Neighborhoods in Cologne

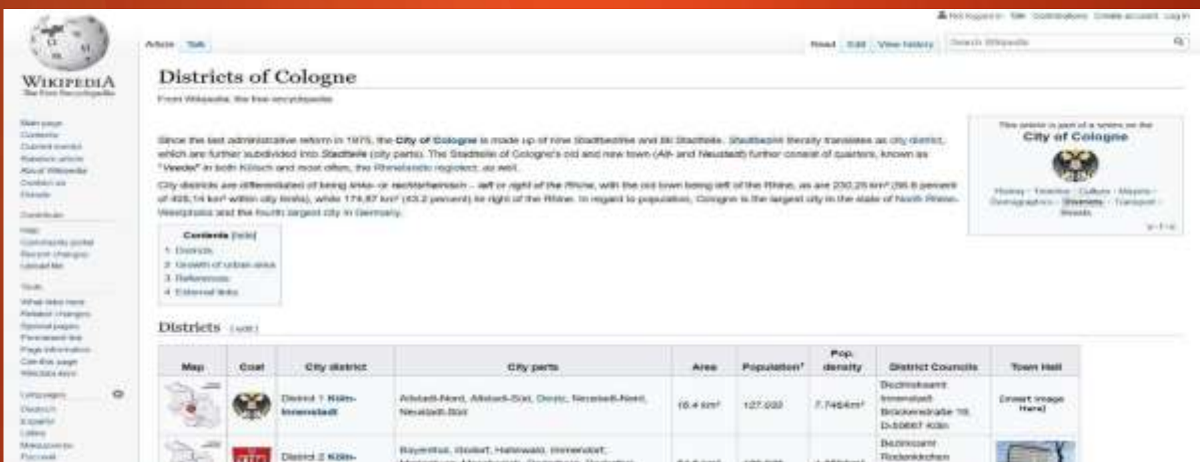
Cologne is a city near which the author lives nearby, and it is well-known for its cathedral, trade fairs and conferences, shopping boulevard, and lively party scene. Since it draws a massive amount of visitors, there is a large business potential for restaurants because there is never a lack of people 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

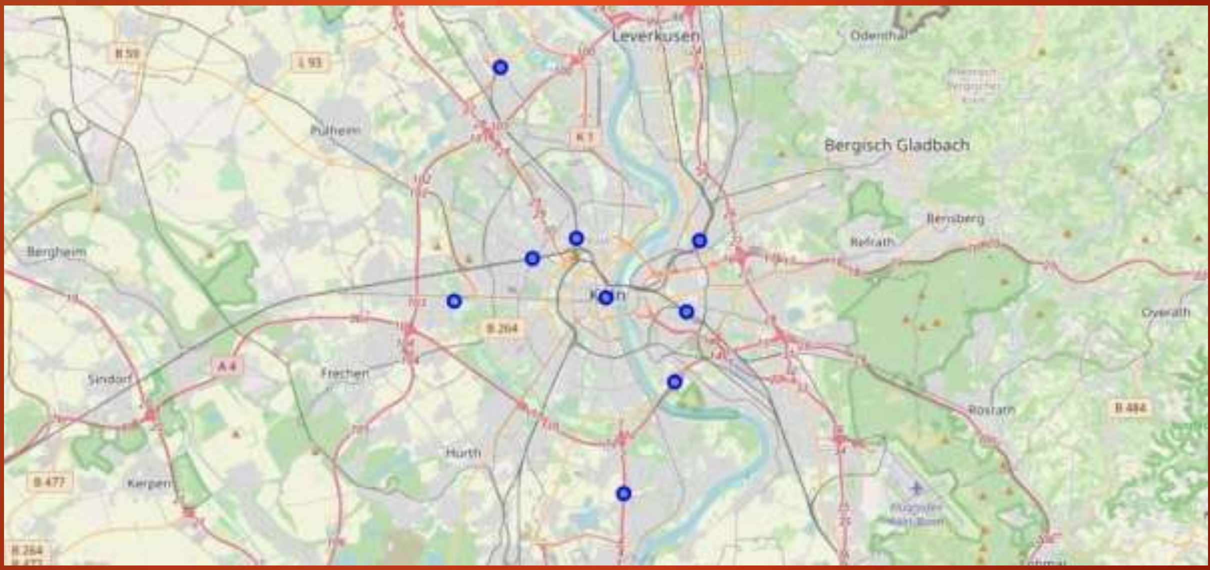
Remove any Boroughs that are not Assigned

	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 km²	100.936	1.850/km²	Bezirksamt Rodenkirchen Hauptstraße 85, D-5099...
2	Köln-Lindenthal	Braunsfeld, Junkersdorf, Klettenberg, Lindenth...	41.6 km²	137.552	3.308/km²	Bezirksamt Lindenthal Aachener Straße 220, 509...
3	Köln-Ehrenfeld	Bickendorf, Bocklemünd/Mengenich, Ehrenfeld, N...	23.8 km²	103.621	4.348/km²	Bezirksamt Ehrenfeld Venloer Straße 419 – 421,...
4	Köln-Nippes	Bilderstöckchen, Longerich, Mauenheim, Niehl, ...	31.8 km²	110.092	3.462/km²	Bezirksamt NippesNeusser Straße 450,D-50733 Köln
5	Köln-Chorweiler	Blumenberg, Chorweiler, Esch/Auweiler, Fühling...	67.2 km²	80.870	1.204/km²	Bezirksamt Chorweiler Pariser Platz 1, D-50765...
6	Köln-Porz	Eil, Elsdorf, Ensen, Finkenberg, Gremberghoven...	78.8 km²	106.520	1.352/km²	Bezirksamt PorzFriedrich-Ebert-Ufer 64–70, D-5...
7	Köln-Kalk	Brück, Höhenberg, Humboldt/Gremberg, Kalk, Mer...	38.2 km²	108.330	2.841/km²	Bezirksamt KalkKalker Hauptstraße 247–273,D-51...
8	Köln-Mülheim	Buchforst, Buchheim, Dellbrück, Dünnwald, Flit...	52.2 km²	144.374	2.764/km²	Bezirksamt Mülheim Wiener Platz 2a,D-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-Innenstadt	Altstadt-Nord, Altstadt-Süd, Deutz, Neustadt-N...	16.4 km²	127.033	7.746/km²	Bezirksksamt Innenstadt Brückenstraße 19, D-50...	50.937328	6.959234
1	Köln-Rodenkirchen	Bayenthal, Godorf, Hahnwald, Immendorf, Marien...	54.6 km²	100.936	1.850/km²	Bezirksamt Rodenkirchen Hauptstraße 85, D-5099...	50.865622	6.969718
2	Köln-Lindenthal	Braunsfeld, Junkersdorf, Klettenberg, Lindenth...	41.6 km²	137.552	3.308/km²	Bezirksamt Lindenthal Aachener Straße 220, 509...	50.935935	6.871246
3	Köln-Ehrenfeld	Bickendorf, Bocklemünd/Mengenich, Ehrenfeld, N...	23.8 km²	103.621	4.348/km²	Bezirksamt Ehrenfeld Venloer Straße 419 – 421,...	50.951502	6.916529
4	Köln-Nippes	Bilderstöckchen, Longerich, Mauenheim, Niehl, ...	31.8 km²	110.092	3.462/km²	Bezirksamt NippesNeusser Straße 450,D-50733 Köln	50.958994	6.941777
5	Köln-Chorweiler	Blumenberg, Chorweiler, Esch/Auweiler, Fühling...	67.2 km²	80.870	1.204/km²	Bezirksamt Chorweiler Pariser Platz 1, D-50765...	51.021167	6.898034
6	Köln-Porz	Eil, Elsdorf, Ensen, Finkenberg, Gremberghoven...	78.8 km²	106.520	1.352/km²	Bezirksamt PorzFriedrich-Ebert-Ufer 64–70, D-5...	50.906705	6.999129
7	Köln-Kalk	Brück, Höhenberg, Humboldt/Gremberg, Kalk, Mer...	38.2 km²	108.330	2.841/km²	Bezirksamt KalkKalker Hauptstraße 247–273,D-51...	50.931923	7.005806
8	Köln-Mülheim	Buchforst, Buchheim, Dellbrück, Dünnwald, Flit...	52.2 km²	144.374	2.764/km²	Bezirksamt Mülheim Wiener Platz 2a,D-51065 Köln	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.

Out[17]:

	City district	Bavarian Restaurant	Chinese Restaurant	Doner Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Greek Restaurant	Italian Restaurant	Kebab Restaurant	Lebanese Restaurant	Restaurant	Scand Res
0	Köln-Chorweiler	0.00	0.0	0.0	0.00	0.00	0.083333	0.00	0.000000	0.00	0.05	0.0	
1	Köln-Ehrenfeld	0.00	0.0	0.0	0.05	0.05	0.000000	0.00	0.050000	0.05	0.00	0.1	
2	Köln-Innenstadt	0.05	0.0	0.0	0.00	0.00	0.050000	0.00	0.000000	0.00	0.05	0.0	
3	Köln-Kalk	0.00	0.0	0.0	0.00	0.00	0.000000	0.25	0.000000	0.00	0.00	0.0	
4	Köln-Lindenthal	0.00	0.0	0.0	0.00	0.00	0.000000	0.00	0.111111	0.00	0.00	0.0	

STEP 2: After grouping

	Neighborhood	American Restaurant	Asian Restaurant	Austrian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Doner Restaurant	Eastern European Restaurant	Falafel Restaurant	Fast Food Restaurant	French Restaurant	German Restaurant	Greek Restaurant	Indian Restaurant	Italian Restaurant	Japanese Restaurant	Kebab Restaurant	Kurdish Restaurant
0	Köln-Chorweiler	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.333333	0.000000	0.000000	0.000000	0.000000	0.333333	0.000000	0.000000	0.000000
1	Köln-Ehrenfeld	0.000000	0.000000	0.000000	0.038462	0.000000	0.000000	0.000000	0.000000	0.000000	0.038462	0.076923	0.038462	0.000000	0.153846	0.038462	0.038462	0.000000
2	Köln-Innenstadt	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.041667	0.000000	0.083333	0.083333	0.000000	0.000000	0.291667	0.000000	0.000000	0.000000
3	Köln-Kalk	0.000000	0.041667	0.000000	0.000000	0.041667	0.041667	0.041667	0.000000	0.000000	0.000000	0.041667	0.083333	0.041667	0.083333	0.000000	0.000000	0.041667
4	Köln-Lindenthal	0.041667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.041667	0.333333	0.083333	0.041667	0.083333	0.041667	0.000000	0.000000
5	Köln-Mülheim	0.000000	0.095238	0.000000	0.000000	0.047619	0.000000	0.000000	0.000000	0.000000	0.000000	0.047619	0.047619	0.000000	0.095238	0.000000	0.000000	0.000000
6	Köln-Nippes	0.000000	0.000000	0.064516	0.000000	0.000000	0.032258	0.000000	0.000000	0.000000	0.096774	0.064516	0.064516	0.000000	0.193548	0.000000	0.032258	0.000000
7	Köln-Porz	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.035714	0.071429	0.250000	0.071429	0.000000	0.178571	0.000000	0.000000	0.035714
8	Köln-Rodenkirchen	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.333333	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

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..

]:

	City district	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-Chorweiler	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Scandinavian Restaurant	Restaurant	Lebanese Restaurant	Kebab Restaurant	Italian Restaurant	Greek Restaurant	Falafel Restaurant	Ethiopian Restaurant
1	Köln-Ehrenfeld	Restaurant	Kebab Restaurant	Italian Restaurant	Falafel Restaurant	Ethiopian Restaurant	Vegetarian / Vegan Restaurant	Scandinavian Restaurant	Lebanese Restaurant	Greek Restaurant	Fast Food Restaurant
2	Köln-Innenstadt	Vegetarian / Vegan Restaurant	Lebanese Restaurant	Fast Food Restaurant	Bavarian Restaurant	Scandinavian Restaurant	Restaurant	Kebab Restaurant	Italian Restaurant	Greek Restaurant	Falafel Restaurant
3	Köln-Kalk	Greek Restaurant	Vegetarian / Vegan Restaurant	Scandinavian Restaurant	Restaurant	Lebanese Restaurant	Kebab Restaurant	Italian Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant
4	Köln-Lindenthal	Italian Restaurant	Vegetarian / Vegan Restaurant	Scandinavian Restaurant	Restaurant	Lebanese Restaurant	Kebab Restaurant	Greek Restaurant	Fast Food Restaurant	Falafel 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.

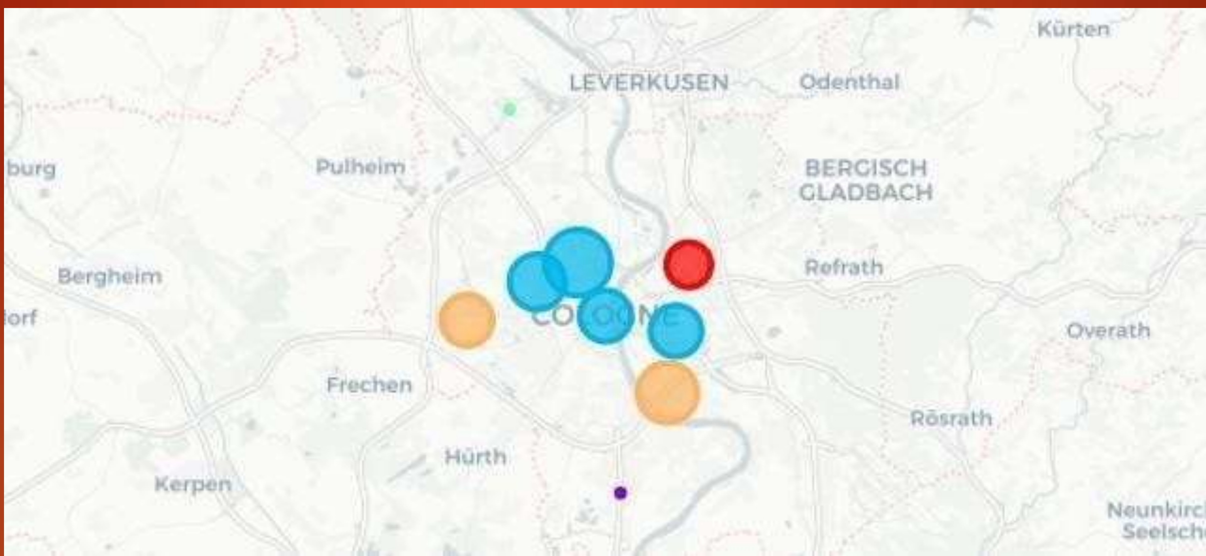
	Cluster Labels	City district	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	2	Köln-Chorweiler	Sushi Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Kebab Restaurant	Chinese Restaurant
1	3	Köln-Ehrenfeld	Italian Restaurant	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	Köln-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:

In [22]:

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 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 I Com V
2	Braunsfeld, Junkersdorf, Klettenberg, Lindenth...	Bezirksamt Lindenthal Aachener Straße 220, 509...	50.935935	6.871246	0.0	Italian Restaurant	Vegetarian / Vegan Restaurant	Scandinavian Restaurant	Restaurant	Lebanese Restaurant	Kebab Restaurant	Greek Restaurant	Fast F Restau

Second cluster result shows Vegan Cuisine Cluster in InnenStadt

	City parts	District Councils	Latitude	Longitude	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 Con
0	Altstadt-Nord, Altstadt-Süd, Deutz, Neustadt-N...	Bezirksamt 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 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 Mo Comm Ven
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	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th I 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 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	8t Co
5	Blumenberg, Chorweiler, 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	Gre 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.