Applied Data Science Project-Capstone

Finding an optimal place to open a Pakistani restaurant in Karachi

Saifal Talpur

January 4, 2021

Table of Contents

1. Introduction

This project is related to finding an optimal location for opening a Pakistani restaurant in the metropolitan city of Karachi, located in the southern region of Pakistan. The city is also termed as the "city of lights." It consists of over 30 million population with diverse ethnicity. The city is also the financial hub of Pakistan, a south asian nation bordering China, India, Iran and Afghanistan.

Since, it is the major financial attraction and involves a huge population, there exists tremendous business opportunities in Karachi. Majority of the business carried out in the 'city of lights' is either in manufacturing or in the service sector. Investing in the service sector is easier and requires less resources and the capital investment. This project moreover offers the opportunity of availing the potential of applying the data science.

To sum up, the project considered the role of data science towards obtaining an optimal location to open a restaurant. Although, the project focused on a typical case study of opening the restaurant (a typical but profitable business model in many countries including Pakistan), the technique applied is moreover independent of the type of business or the geographical location. Towards obtaining the solution, the project involved following techniques, which are moreover explained in the later sections of the report.

• Data collection

• Data pre-processing

• Data clustering and segmentation

• Data visualisation

2. Background

Opening a new restaurant, particularly in a cosmopolitan city is not only challenging but a concern as well in case to turn it into a successful business. This project thus answers these questions and provide a solution which can assist the entrepreneurs to start their own business. This project focused on finding an optimal location for a restaurant, serving a favourite cuisine for the locality. Although, many restaurants exist in the region, however, the lack of business knowledge and, despite a huge potential of using the data science, there exists no application of applying the data science in the business models. This project thus explores the potential of applying the data science in developing the business model with answering the following questions.

1. How the data regarding existing restaurant venues can be collected by applying the data science?

2. How the collected data can be used to find an optimal restaurant location which is besides accessible to public transport also offers reasonable food price?

3. By applying the machine learning techniques, such as segmentation and clustering, how a choice about most and least favourite venues can be extracted?

3. Data acquisition

First, it was figured out what data is required and then how to obtain the required data provided the convenience to the restaurants is considered as well. Initially, the type and location of restaurants was figured out followed by optimal distance from the public transport and food prices.

3.1 Data sources

Information from **Foursquare API** and **Geocoder** data sources was used in terms of obtaining the data related to location of restaurants in the city within the customised radius besides finding the geographical information like the latitude and longitude of restaurants to rent, respectively.

	id	name	categories	lat	lng	neighborhood
0	52d9009c11d2079ba506dca1	M Clifton	Metro Station	24.830433	67.038430	CC-2, Block 8, Clifton
1	4b5fdbf6f964a520f9ce29e3	M Clifton	Metro Station	24.816201	67.021181	Com 5/1, Sh-e-Firdousi, Boat Basin, Clifton Bl...
2	4b5d5cc1f964a520d85a29e3	M Mhaheed-e-Millat Rd	Metro Station	24.876738	67.065969	Off Shaheed-e-Millat Rd (Delhi Mercantile Hous...
3	4e189aa56284ea7246fa36c6	M Club Road	Metro Station	24.846429	67.026303	Club Road
4	4cb026cd562d224b511a1488	M D.H.A	Metro Station	24.799193	67.046309	Shop 1&2, 15-C, Badar Commercial Area, Phase 5...

3.2 Procedure

Following steps were carried out from data collection to feature selection:

1. A data set was initially produced based on information from Foursquare API, first requiring the client ID, client password, desired category (restaurant in this project), latitude, longitude and radius of favourite zone to open a restaurant in Karachi.

2. The data obtained was saved in a csv file and then read through 'Pandas' library.
3. The data set from step 2 was then merged with another developed data set (including information like **postal code** and **food prices**). Since, the food price is also reflected in the rental price, hence any of the available information can be a valuable parameter in deciding the suitable location based on the budget.

	id	name	categories	lat	lng	neighborhood	postal code	price
0	52d9009c11d2079ba506dca1	The East End	Pakistani Restaurant	24.830433	67.038430	CC-2, Block 8, Clifton	75600	4500
1	4b5fdbf6f964a520f9ce29e3	Bar-B-Q Tonight	Pakistani Restaurant	24.816201	67.021181	Com 5/1, Sh-e-Firdousi, Boat Basin, Clifton Bl...	75600	5000
2	4b5d5cc1f964a520d85a29e3	Shaikh Abdul Ghaffar Kabab House	Pakistani Restaurant	24.876738	67.065969	Off Shaheed-e-Millat Rd (Delhi Mercantile Hous...	75400	3500
3	4e189aa56284ea7246fa36c6	The Pakistani	Indian Restaurant	24.846429	67.026303	Club Road	75530	4500
4	4cb026cd562d224b511a1488	Mirchili	Pakistani Restaurant	24.799193	67.046309	Shop 1&2, 15-C, Badar Commercial Area, Phase 5...	75500	5000

	id	name	categories	lat	lng	neighborhood	postal code	price	categories_simplified
30	54d1e3a4498e26dc9e820cc3	Khalifa Idrees Nihari	Pakistani Restaurant	24.860545	67.013575	Aslam Road	75510	1800	Pakistani
31	4b88a627f964a520b40632e3	Biryani Centre	Pakistani Restaurant	24.807743	67.037277	26th St, Tauheed Commercial Area (Kh-e-Shamsheer)	75500	5000	Pakistani
32	4e2eee46d4c058fdbeee6a70	Special Salateen Restaurant	Pakistani Restaurant	24.810098	67.017595	Block 2, Clifton	75600	4500	Pakistani
33	586013e48f0be461c0b8fb42	Zaoq Resturant	Pakistani Restaurant	24.802140	67.029655	2nd Floor Dolmen City Mall Clifton	75600	4300	Pakistani
34	4d4afd737d36f04d7e4221e5	Daily Dubai Restaurant	Pakistani Restaurant	24.797185	67.044400	Badar Commercial St 10 (Off 26th Street)	75500	4600	Pakistani
35	4c1caaac63750f47bfe1b667	Student Biryani	Pakistani Restaurant	24.881438	67.066382	Bahadurabad Chowrangi (Alamgir St)	75300	3680	Pakistani
36	4db13be44b22ba776d7e216e	Student Biryani	Pakistani Restaurant	24.798461	67.045398	Badar Comm. Area (10th Badar Comm. St)	75500	4600	Pakistani
37	5a30f6b5851de51358a10e1c	Sultan The Royal Eatery	Pakistani Restaurant	24.868716	67.078231	Tipu Sultan Road	75400	4000	Pakistani
38	4dafdf06f7b149e03f3d3b5a	Suzie Wong	Asian Restaurant	24.846505	67.031120	Marriott Hotel, Abdullah Haroon Rd	75520	15000	Asian
39	50c44ea6e4b0781921dd9eae	Student Biryani	Pakistani Restaurant	24.907108	67.071508	University Rd	75270	1500	Pakistani
40	4d1f6a9e8629224baaf20a87	Savor	Pakistani Restaurant	24.785754	67.091148	31/1 Zulfiqar Street 1, D.H.A. Phase 8 Zone A ...	75500	4500	Pakistani
41	5499a6d8498e8b6d491aa47f	the white biryani	Pakistani Restaurant	24.809087	67.069605	16 C 3rd St, D.H.A Phase 6 Rahat Commercial Area	75500	4800	Pakistani
42	4d445f3e4e5d370497b0ea93	Super Salateen Hotel	Pakistani Restaurant	24.901007	67.111699	Gulshan-e-Jamal, (Near Dalmia)	75260	1600	Pakistani
43	4fd387d6e4b09c51fe507c45	Darbar Haleem & Biryani	Pakistani Restaurant	24.928082	67.095909	Near Disco Bakery	75290	1500	Pakistani

4. Methodology

Since involving several factors, the decision to open a Pakistani restaurant in a locality of Karachi must be made, where each factor influences on finding an optimal restaurant location. The locality serves different cuisines for instance, Middle Eastern, Asian, Indian, Street food and BBQ. Since, a variety of food types are served, hence, it is important the subject restaurant will serve the favourite food (particularly, in the optimal vicinity) with amenities like minimum distance to public

restaurant, offering food at reasonable cost (moreover dependent on restaurant's rental prices in the area).

To achieve this goal, this project considers the technique **K-means Clustering** to make segmentations of the types of Pakistani restaurants or, the restaurant with benefits. After that, research about the prices to rent a place for opening a restaurant will be made. In addition, the optimal location should be accessible by public transportation. In the end, the focus will be to find locations that have approximately 300 m from the public transport.
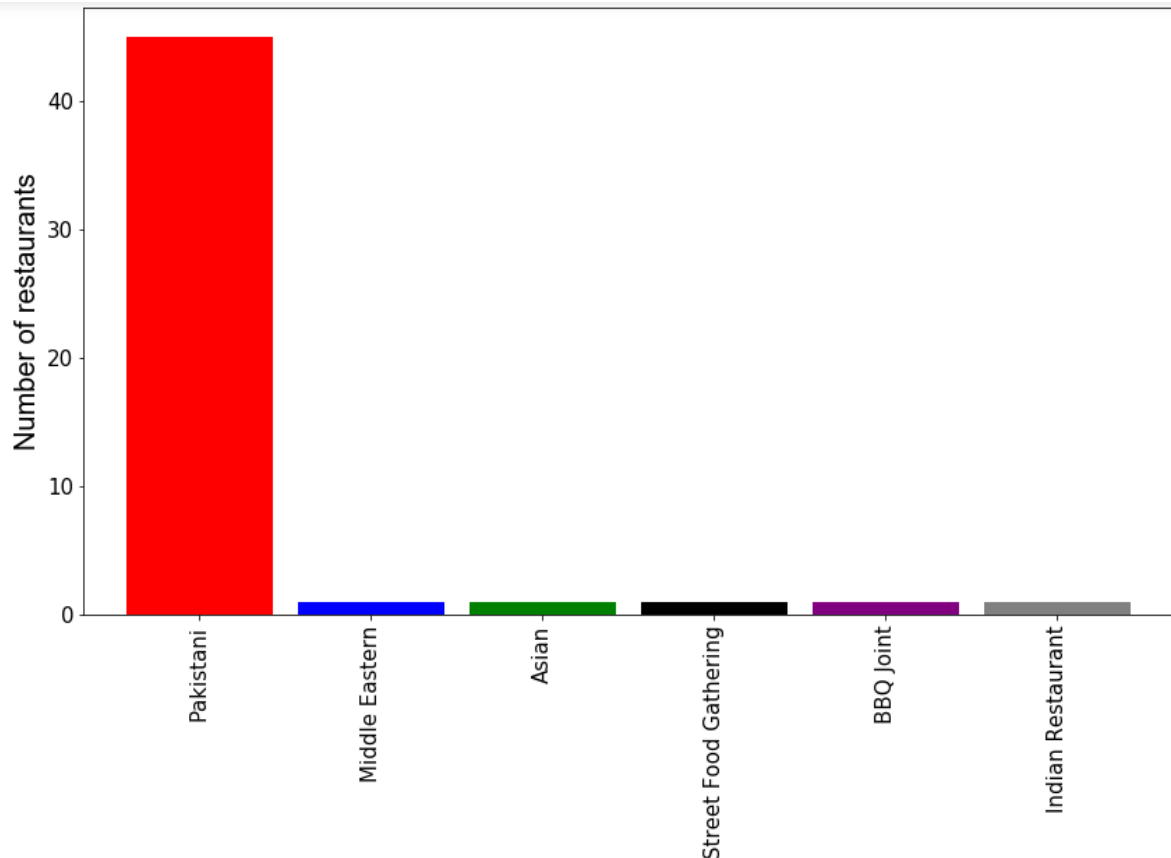

Fig. 1 Type and number of restaurants in Karachi

Following factors are moreover considered in performing the restaurant search:

- number of existing Pakistani restaurants in the selected radius
- prices and locations of restaurants in Karachi to open a restaurant
- segmentation of types of Pakistani restaurants in a neighbourhood
- distance of the available places to rent nearby public transportation

	neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	16 C 3rd St, D.H.A Phase 6 Rahat Commercial Area	Pakistani Restaurant	Street Food Gathering	Middle Eastern Restaurant	Indian Restaurant	BBQ Joint
1	1C Bukhari Commercial Lane 13 Opposite Khayaba...	Pakistani Restaurant	Street Food Gathering	Middle Eastern Restaurant	Indian Restaurant	BBQ Joint
2	26th St, Tauheed Commercial Area (Kh-e-Shamsheer)	Pakistani Restaurant	Street Food Gathering	Middle Eastern Restaurant	Indian Restaurant	BBQ Joint
3	2nd Floor Dolmen City Mall Clifton	Pakistani Restaurant	Street Food Gathering	Middle Eastern Restaurant	Indian Restaurant	BBQ Joint
4	31/1 Zulfiqar Street 1, D.H.A. Phase 8 Zone A ...	Pakistani Restaurant	Street Food Gathering	Middle Eastern Restaurant	Indian Restaurant	BBQ Joint

K-means Clustering is an unsupervised machine learning algorithm, which is used to cluster and/or segment the data types.

	id	neighborhood	postal code	price	categories_simplified	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
38	4dafdf06f7b149e03f3d3b5a	Marriott Hotel, Abdullah Haroon Rd	75520	15000	Asian	0	Asian Restaurant	Street Food Gathering	Pakistani Restaurant	Middle Eastern Restaurant	Indian Restaurant
49	59ea29e66fa81f72c84fa4dd	Khalid bin Waleed Rd.	75100	2600	Middle Eastern	0	Middle Eastern Restaurant	Street Food Gathering	Pakistani Restaurant	Indian Restaurant	BBQ Joint

	name	neighborhood	postal code	price	categories_simplified	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	The East End	CC-2, Block 8, Clifton	75600	4500	Pakistani	1	Pakistani Restaurant	Street Food Gathering	Middle Eastern Restaurant	Indian Restaurant	BBQ Joint
1	Bar-B-Q Tonight	Com 5/1, Sh-e-Firdousi, Boat Basin, Clifton Bl...	75600	5000	Pakistani	1	Pakistani Restaurant	Street Food Gathering	Middle Eastern Restaurant	Indian Restaurant	BBQ Joint
2	Shaikh Abdul Ghaffar Kabab House	Off Shaheed-e-Millat Rd (Delhi Mercantile Hous...	75400	3500	Pakistani	1	Pakistani Restaurant	Street Food Gathering	Middle Eastern Restaurant	Indian Restaurant	BBQ Joint
4	Mirchili	Shop 1&2, 15-C, Badar Commercial Area, Phase 5...	75500	5000	Pakistani	1	Pakistani Restaurant	Street Food Gathering	Middle Eastern Restaurant	Indian Restaurant	BBQ Joint
5	Zahid Nihari	Tariq Road	75401	4500	Pakistani	1	Pakistani Restaurant	Street Food Gathering	Middle Eastern Restaurant	Indian Restaurant	BBQ Joint

	name	neighborhood	postal code	price	categories_simplified	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
3	The Pakistani	Club Road	75530	4500	Indian Restaurant	2	Indian Restaurant	Street Food Gathering	Pakistani Restaurant	Middle Eastern Restaurant	BBQ Joint

	name	neighborhood	postal code	price	categories_simplified	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
15	Ghaffar Kabab House	Port Grand, Native Jetty Bridge (M. T. Khan Rd)	74000	6000	BBQ Joint	3	BBQ Joint	Street Food Gathering	Pakistani Restaurant	Middle Eastern Restaurant	Indian Restaurant

	name	neighborhood	postal code	price	categories_simplified	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
6	Boat Basin	Kh-e-Saadi (Kh-e-Roomi)	75500	5000	Street Food Gathering	4	Street Food Gathering	Pakistani Restaurant	Middle Eastern Restaurant	Indian Restaurant	BBQ Joint

5. Exploratory Data Analysis

Since involving several factors, the decision to open a Pakistani restaurant in a locality of Karachi must be made, where each factor influences on finding an optimal restaurant location. The locality serves different cuisines for instance, Middle Eastern, Asian, Indian, Street food and BBQ. Since, a variety of food types are served, hence, it is important the subject restaurant will serve the

favourite food (particularly, in the optimal vicinity) with amenities like minimum distance to public restaurant, offering food at reasonable cost (moreover dependent on restaurant's rental prices in the area).

Address	index	Postcode	Price	Latitude	Longitude	Score
CC-2, Block 8, Clifton	0	75600	4500	24.830433	67.038430	300.0
University Rd, Block-13A, Gulshan-e-Iqbal	27	75300	2000	24.904386	67.077488	300.0
Club Road (Dr. Ziauddin Ahmed Rd.)	28	75530	4600	24.846723	67.026029	300.0
Nursery, Shahrah-e-Faisal	29	75350	4200	24.860488	67.064217	300.0
Aslam Road	30	75510	1800	24.860545	67.013575	300.0

	cluster_labels	top_venues_categories
0	0	Middle Eastern Restaurant
1	1	Pakistani Restaurant
2	2	Indian Restaurant
3	3	BBQ Joint
4	4	Street Food Gathering

cluster_labels	count	first
0	2	Asian Restaurant
1	45	Pakistani Restaurant
2	1	Indian Restaurant
3	1	BBQ Joint
4	1	Street Food Gathering

6. Results and Discussion

Since involving several factors, the decision to open a Pakistani restaurant in a locality of Karachi must be made, where each factor influences on finding an optimal restaurant location. The locality serves different cuisines for instance, Middle Eastern, Asian, Indian, Street food and BBQ. Since, a variety of food types are served, hence, it is important the subject restaurant will serve the favourite food (particularly, in the optimal vicinity) with amenities like minimum distance to public restaurant, offering food at reasonable cost (moreover dependent on restaurant's rental prices in the area).

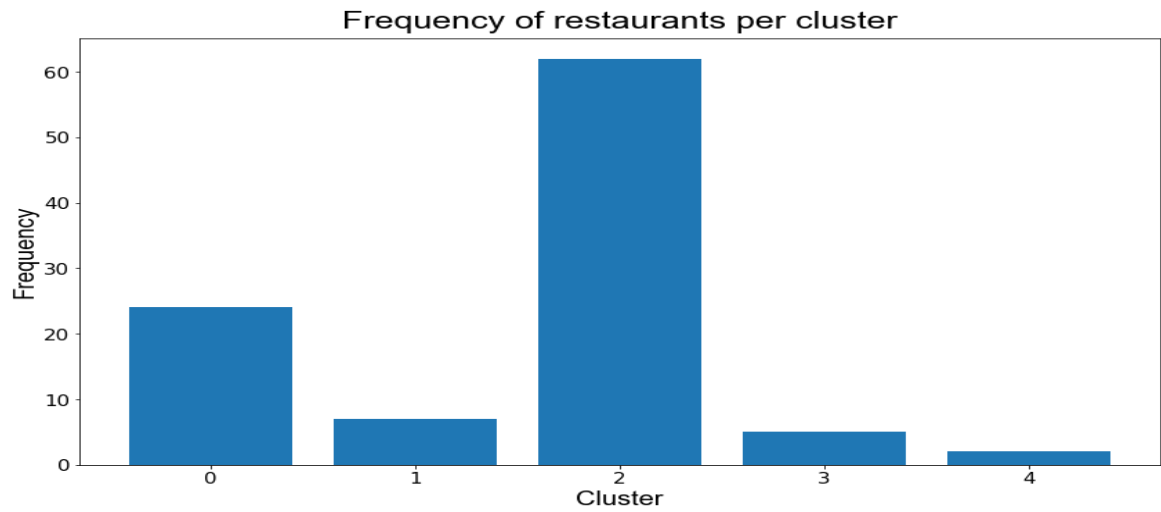|	index	Postcode	Price	Latitude	Longitude	Score
count	50.00000	50.00000	50.000000	50.000000	50.000000	50.0
mean	24.50000	75438.22000	4091.400000	24.857195	67.058065	300.0
std	14.57738	299.92729	2599.775508	0.048561	0.031590	0.0
min	0.00000	74000.00000	1500.000000	24.778282	66.989577	300.0
25%	12.25000	75350.00000	2500.000000	24.811624	67.031392	300.0
50%	24.50000	75500.00000	4250.000000	24.854619	67.057831	300.0
75%	36.75000	75545.00000	4600.000000	24.895193	67.078045	300.0
max	49.00000	75950.00000	15000.000000	24.955090	67.117584	300.0

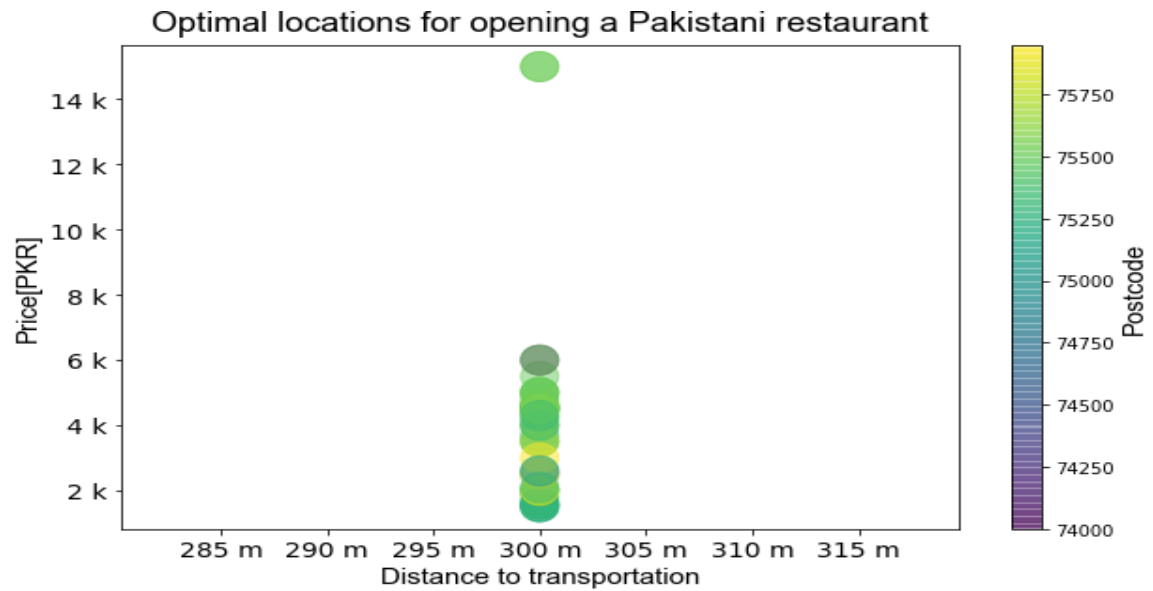Fig.2 Frequency of restaurants per cluster

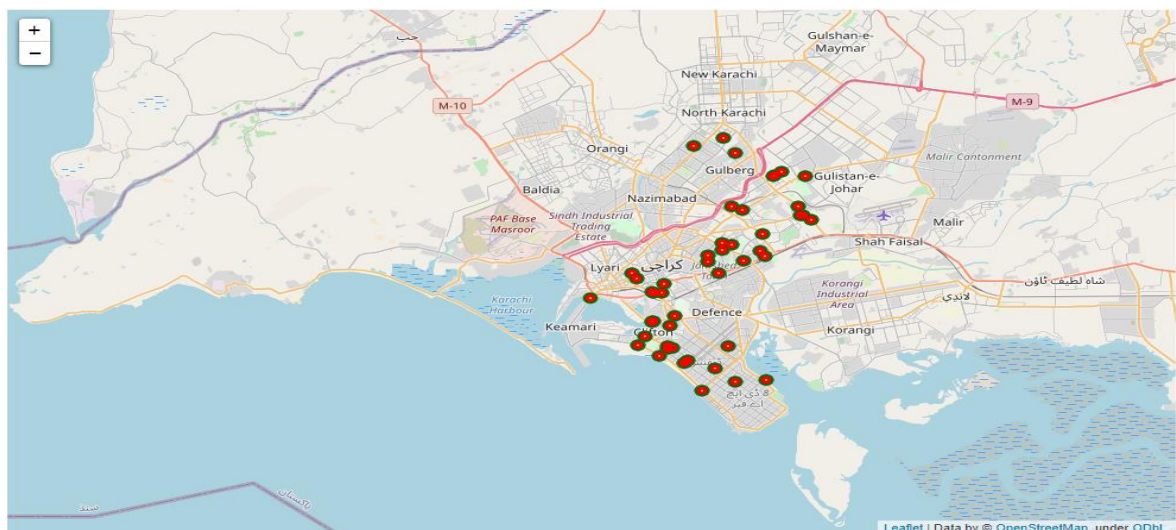Fig. 3 Optimal locations for opening a Pakistani restaurant

Fig. 4 Map showing optimal locations (as indicated in red)

7. Conclusion

Since involving several factors, the decision to open a Pakistani restaurant in a locality of Karachi must be made, where each factor influences on finding an optimal restaurant location. The locality serves different cuisines for instance, Middle Eastern, Asian, Indian, Street food and BBQ. Since, a variety of food types are served, hence, it is important the subject restaurant will serve the favourite food (particularly, in the optimal vicinity) with amenities like minimum distance to public restaurant, offering food at reasonable cost (moreover dependent on restaurant's rental prices in the area).