Finding the Optimal Location for an Indian Restaurant in London

Ismael Siffon

April 5th, 2020

1 Introduction

Indian cuisine, one of the most flavorful cuisine in the world, has found its way to the street of the biggest cities in the US and in Europe. In London, the so-called "curry houses" have gained a tremendous influence and became a popular place to eat. This fast-growing sector of the restaurant industry offers many opportunities for new restaurateurs to invest.

Even with the success of Indian cuisine in London, many other factors should be considered when opening a new restaurant. One of the most important factors is location. The choice of location is very important as it can be the determining factor of success, regardless of the quality of the food or the service.

Data-driven decision making, or data science, can be very helpful in finding an adequate location. Information about potential customers, the existing demand for a particular menu, the competitiveness among restaurants, or the safety, can be analyzed in order to pick the best location.

In this project, we are going to help a restaurateur pick the location of his new Indian restaurant in London. With the increasing rate at which Indian cuisine is becoming popular in the UK, our restaurateur has some specific requirements for his location:

- Safety: the area should be safe! Safety is the primary concern of any business and thus places with high crime should be avoided.
- Costumer profile: Our restaurateur wants to have a customer base that already has an acquired taste for Indian cuisine. The location should then have a high Indian population, or other nationalities that have similar cuisine.
- Competition: Our restaurateur has a marketing plan that will benefit him if the location already has some Indian restaurants around. However, to avoid too much competition, it should not be crowded with Indian restaurants.
- Customer purchasing power: Because Indian food is rich in condiments, it is expensive to make, thus not affordable to everybody. To make sure his restaurant keeps a high customer influx, our restaurateur wants to be in an area where most people have a middle-class income.
- Location affordability: Because our restaurateur hasn't built a name yet, he does not want most of invested money to go to rent, instead of advertising. Rent should then be affordable.

These requirements are going to guide this project and by the end of it, the restaurateur will have the list of neighborhoods that fit his needs.

2 Data gathering and data cleaning

2.1 Description

The data used for this project comes from multiple sources is defines as below:
Demographic data of all boroughs of London: This data comes from Kaggle and is the first step of the project. For each of the 34 boroughs in London, it contains information like:

- Population density
- Average age
- Largest, second and third migrant population and percentage
- Employment rate
- Gross annual pay
- Crime rate, among others.

Crucial information can be revealed from the demographic analysis of London. For example, we can determine where the majority of Indians are clustered in London, or what places should be avoided because of safety concerns.

List of neighborhoods and their coordinates: The data from Kaggle provides information for each borough in London, but boroughs are too large areas for this project. We need to work on a deeper level of neighborhoods. The list of neighborhoods of each borough can be found on wikipedia.com. Also, the coordinates for each neighborhood can be found using the Google API.

Foursquare data: Using Foursquare, we will be able to gather information about restaurants surrounding the neighborhoods in London. This part of the data is important when it comes to evaluating the competition around an area. We can retrieve information about other Indian restaurants such as:

- The number of Indian restaurants
- The number of likes
- The rating

2.2 Data gathering

The demographic data was retrieved from Kaggle and is in a 'csv' format. Its size is 38 rows by 85 columns, and most of the rows are not relevant for our project. The table below shows some of the information from this file.

Code	New code	Area name	Inner/ Outer London	GLA Population Estimate 2016	GLA Household Estimate 2016	Inland Area (Hectares)	Population density (per hectare) 2016	Average Age, 2016	Proportion of population aged 0-15, 2016	...	Happiness score 2011-14 (out of 10)	Anxiety score 2011-14 (out of 10)
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN
E09000001	E09000001	City of London	Inner London	8,548	5,179	290.4	28.9	42.9	27.2	...	5.99	5.57
E09000002	E09000002	Barking and Dagenham	Outer London	205,773	76,841	3,610.8	57.3	32.9	21.0	...	7.05	3.05
E09000003	E09000003	Barnet	Outer London	385,108	149,147	8,674.8	44.5	37.2	21.0	...	7.37	2.75
E09000004	E09000004	Bexley	Outer London	243,303	97,233	6,058.1	39.9	38.9	20.8	...	7.21	3.29

Table 1

The lists of neighborhoods in reach borough were retrieved from Wikipedia using web-scrapping.

2.3 Data cleaning

The first step is to remove the columns and rows we do not need and only keep the data needed for our analysis. Then, we remove the rows that do not represent a borough and adjust the type of each column. The columns that we will keep are:

- Borough,
- Population density (per hectare) 2016,
- Largest migrant population by country of birth (2011)
- % of largest migrant population (2011)
- Second largest migrant population by country of birth (2011)
- % of second largest migrant population (2011)
- Third largest migrant population by country of birth (2011)
- % of third largest migrant population (2011)
- Gross Annual Pay, (2015)
- Crime rates per thousand population 2014/15
- Median House Price, 2014

The table below shows the first 5 rows from the new table:

Borough	Population density (per hectare) 2016	Largest migrant population by country of birth (2011)	% of largest migrant population (2011)	Second largest migrant population by country of birth (2011)	% of second largest migrant population (2011)	Third largest migrant population by country of birth (2011)	% of third largest migrant population (2011)	Gross Annual Pay, (2015)	Crime rates per thousand population 2014/15	Median House Price, 2014
City of London	28.9	United States	2.8	France	2.0	Australia	1.9	NaN	NaN	765000
Barking and Dagenham	57.3	Nigeria	4.7	India	2.3	Pakistan	2.3	28428	83.4	215000
Barnet	44.5	India	3.1	Poland	2.4	Iran	2.0	33084	62.7	400000
Bexley	39.9	Nigeria	2.6	India	1.5	Ireland	0.9	32040	51.8	250000
Brent	76.1	India	9.2	Poland	3.4	Ireland	2.9	29777	78.8	385000

Table 2

3 Methodology

In this project, we are going to use the data collected in order to find a location that fits the requirements imposed by our restaurateur. These requirements will then guide the direction of our workflow.

The first step was to gather and clean data. The data collected provides demographic information about every borough in London. We decided to only keep the columns related to our objective such as crime rate, gross pay, median house price, and migrant population data.

- Crime rate will be used to tackle the safety concern
- Gross pay will be used to tackle the customer's purchasing power concern
- House price will be used to compare how expensive renting a local can be.
- Migrant population data will be used to reveal the areas with a large population of Indians.

Now that our data is ready to be used, we will first observe its behavior using statistical numbers and some plots for visualization. We will then transform the data in order to have meaningful numbers. One of the requirements is to have an Indian cuisine lover customer base. For that, we will calculate the density of the Indian/Pakistan/Sri Lanka population for the boroughs where they appear in the list of the three largest migrant population. With this information, we will be able to locate clusters of migrants from Indian/Pakistan/Sri Lanka.
The next step will be to filter out the boroughs that do not meet the requirements. Here are the conditions that need to be met:

- Crime rate < 75th percentile
- Density from India/Pakistan/Sri Lanka > 2 people per hectare
- Gross pay > £30,000
- Median house price < £500,000

After filtering out the boroughs that do not fit the requirement, we will focus on the remaining boroughs and look into their neighborhoods. After importing the neighborhoods geospatial coordinates from the Google API, we will explore all the Indians restaurants within them using

Foursquare. The purpose of that is to be able to compare the competitiveness and the customer response to Indian cuisine for each neighborhood.

Finally, we will use the clustering method to group the neighborhoods based on their similarities when it comes to Indian restaurants. Our restaurateur will be able to decide where he wants his new restaurant to be based on the competition and the customer response to Indian cuisine.

4 Analysis

4.1 Data visualization

Our analysis starts with a look at some statistical information about the dataset. The table below shows a summary of some basic statistical information of each column. We can see that the gross pay varies between £27,000 and £42,800 and the median house price between £215,000 and £1,195,000. On Fig.1, we can see the distribution of each parameter. The densest boroughs are Kensington and Chelsea, Tower Hamlets, and Hackney, while the most expensive in terms of real estate are City of London, Camden, Kensington and Chelsea, and Westminster. Westminster has a very high rate of crime compared to the other borough.

	Population density (per hectare) 2016	% of largest migrant population (2011)	% of second largest migrant population (2011)	% of third largest migrant population (2011)	Gross Annual Pay, (2015)	Crime rates per thousand population 2014/15	Median House Price, 2014
count	33.000000	33.000000	33.000000	33.000000	30.000000	32.000000	3.300000e+01
mean	73.609091	4.696970	2.860606	2.060606	33622.000000	84.868750	4.290288e+05
std	38.615422	3.098436	1.430459	0.913078	3687.779128	31.129329	2.061033e+05
min	21.700000	1.100000	1.000000	0.700000	27174.000000	50.400000	2.150000e+05
25%	44.500000	2.800000	1.800000	1.500000	31479.500000	63.800000	3.070000e+05
50%	58.800000	3.600000	2.500000	1.900000	32938.000000	77.500000	3.850000e+05
75%	108.200000	5.100000	3.500000	2.400000	35617.750000	99.675000	4.330000e+05
max	153.000000	15.300000	6.800000	5.300000	42798.000000	212.400000	1.195000e+06

Table 3

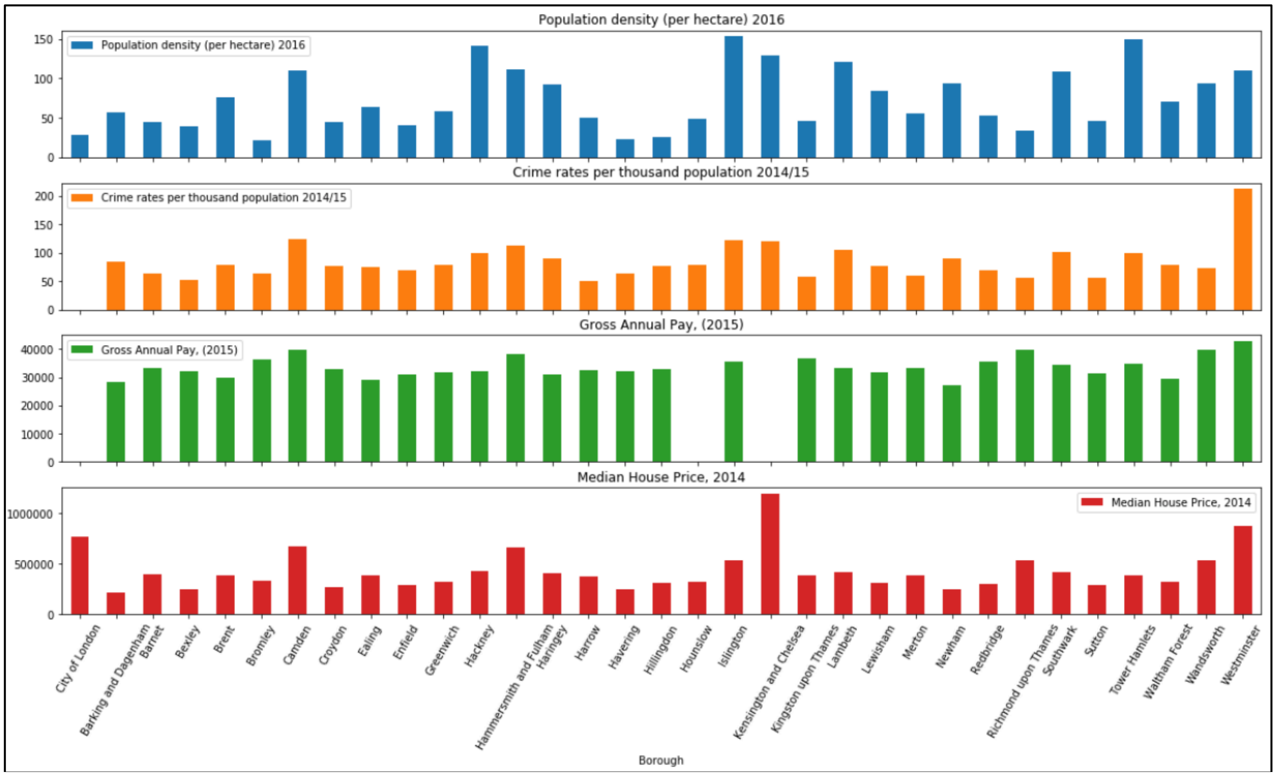

Fig. 1

Our restaurateur specified that he wants his restaurant in an area that has an Indian community. We said earlier that the Indian community in the UK is important. Let's see how it compares to other migrant nationalities in Fig 2. India is by far the largest migrant population in London, followed by Ireland and Poland. Pakistan and Sri Lanka come respectively in the 7th and 8th place.

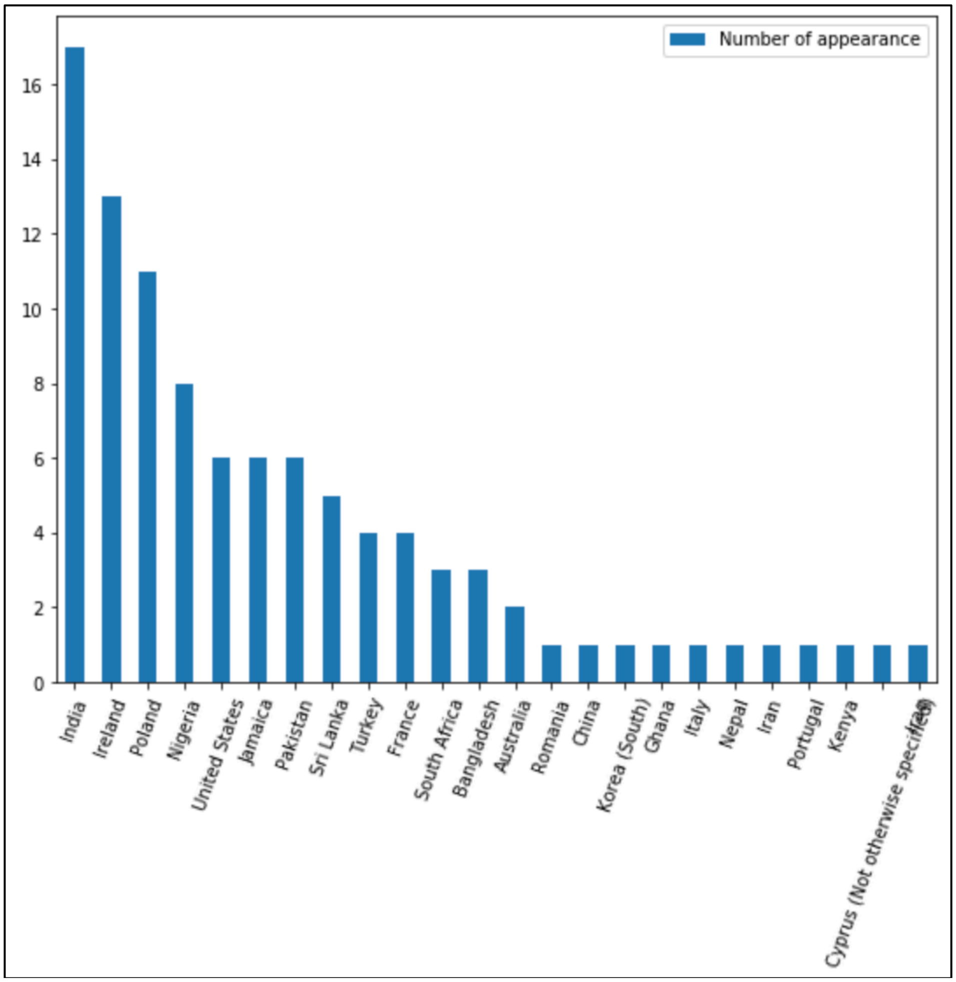

Fig. 2

4.2 Data filtering

For our analysis, we need to have a meaningful parameter that we can use to compare the Indian population between the boroughs. We decided to calculate the density of the Indians population for each borough. Also, we decided to include the population of Pakistan and Sri Lanka, because their cuisine similarities make them potential customers. The formula we used is as follows:

$$\begin{aligned}\textbf{\textit{Density of India/Pakistan/Sri Lanka population}} \\ = \textbf{\textit{Population Density}} * (\textbf{\textit{percentage of migrants from India}} \\ + \textbf{\textit{percentage of migrants from Pakistan}} \\ + \textbf{\textit{percentage of migrants from Sri Lanka}})\end{aligned}$$

Now that we have the data we need, we can start our selection by filtering out the boroughs that do not fit the requirement requested by our restaurateur. We decided to filter based on the following intervals:

- Crime rate < 75th percentile

- Density from India/Pakistan/Sri Lanka > 2 people per hectare
- Gross pay > £30,000
- Median house price < £500,000

We end up with 2 boroughs: Harrow and Redbridge.

Borough	Population density (per hectare) 2016	Gross Annual Pay, (2015)	Crime rates per thousand population 2014/15	Median House Price, 2014	Population density from India/Pakistan/Sri Lanka
Harrow	49.8	32529.0	50.4	370000.0	6.6234
Redbridge	53.3	35665.0	69.7	301500.0	8.2615

Table 4

4.3 Data clustering

Now that we filtered out the boroughs that do not meet our conditions, we can focus on the neighborhoods within the remaining boroughs. There are 45 neighborhoods between Harrow and Redbridge. Their locations were found using the Google API. Here is a map with the location of each neighborhood:

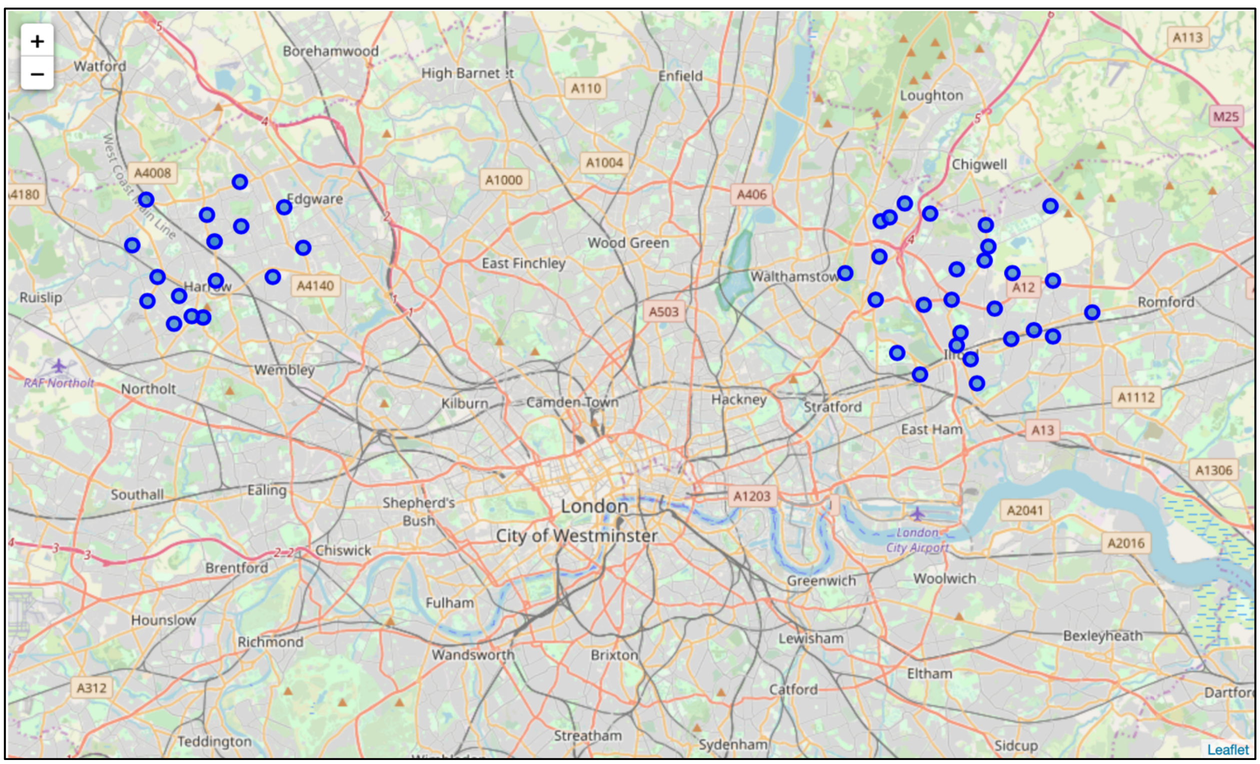

Fig. 3

Now that we have the neighborhoods of interest, we can find information about all the Indian restaurants within each neighborhood using Foursquare. Foursquare has a lot of information about restaurants, but the ones we are interested in are:

- The number of Indian restaurants
- The restaurants' ratings
- The restaurants' number of likes

The number of Indian restaurants and their rating will can be used to examine the competition, and the number of likes can be used to evaluate the population response to Indian cuisine, or the population level of interest. After importing retrieving this information from Foursquare, we can group the restaurants by their neighborhoods. Doing so, we decided to consider the highest rating and the total number of likes for each neighborhood. We ended up with the table below (only the first 7 rows):

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Number of restaurants	Max rating	Total number of likes
Aldborough Hatch	51.585525	0.098766	1	0.0	0
Aldersbrook	51.551474	0.048841	13	7.8	64
Barkingside	51.589925	0.083874	4	0.0	4
Belmont	51.601249	-0.319275	6	0.0	3
Canons Park	51.607674	-0.296236	5	7.1	16
Chadwell Heath	51.572228	0.141995	4	0.0	1
Clayhall	51.586974	0.068774	5	0.0	2

Table 5

Finally, we can cluster the neighborhoods based on the parameters mentioned earlier. We decide to have 6 clusters. A summary of each cluster can be found in table 6. Four our restaurateur, we can describe each cluster as in table 7. We can visualize how the cluster are distributed over the map of London in Fig. 4. We can see cluster '0', red on the map, is the biggest with 16 neighborhoods. The next biggest is cluster '2', with 14 neighborhoods.

	Number of restaurants		Max rating		Total number of likes	
	max	min	max	min	max	min
Cluster number						
0	7	1	0.0	0.0	7	0
1	20	16	7.6	7.0	39	27
2	12	1	8.4	5.8	27	4
3	20	20	8.1	8.1	165	165
4	16	10	0.0	0.0	9	3
5	23	13	7.8	7.6	81	64

Table 6

Cluster label	Color on map	Description
0	Red	Low competition & poor level of interest
1	Purple	High competition & good level of interest
2	Blue	Medium competition & good level of interest
3	Turquoise	Very High competition & great level of interest
4	Green	Medium competition & poor level of interest
5	Orange	High competition & great level of interest

Table 7

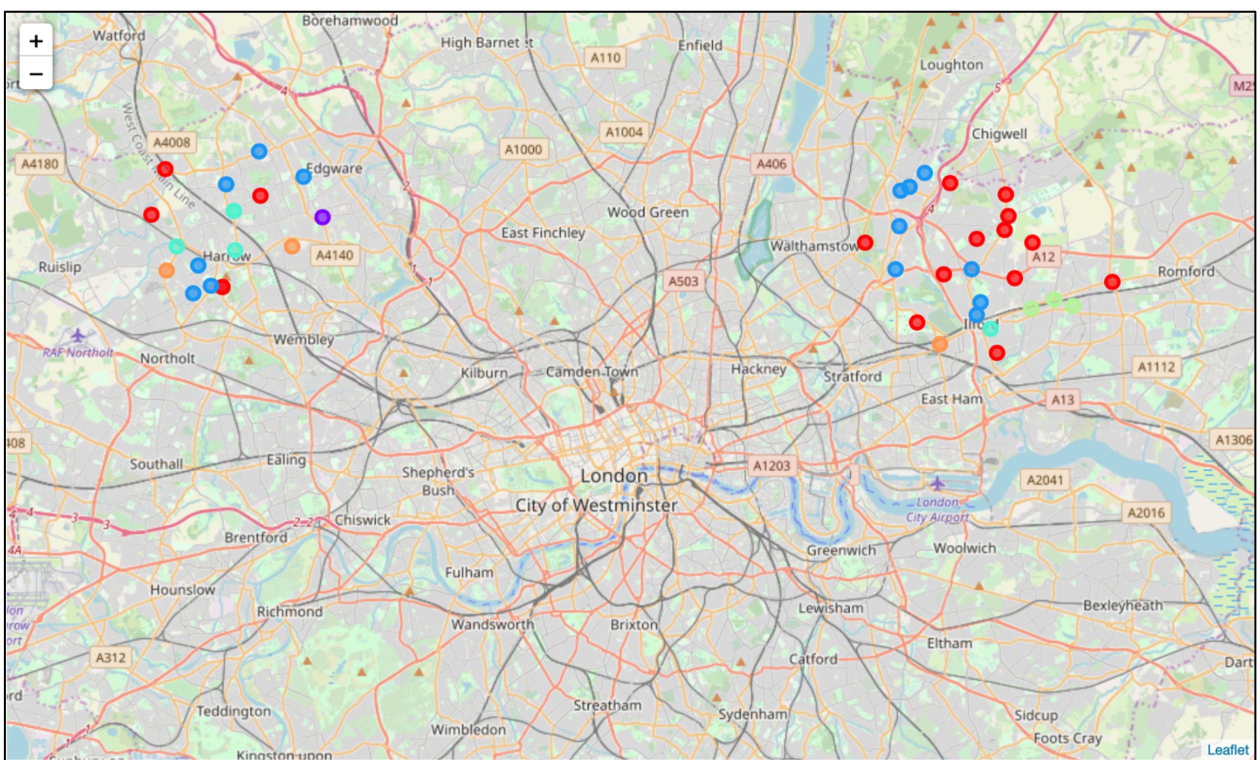

Fig. 4

5 Results and Discussion

Very useful information was retrieved from the demographic data for all 34 boroughs in London. We could see that the largest migrant population was from India, which is a good sign that the Indian cuisine was already introduced to most people in London. Also, this Indian community represents a potential customer base that will not require much advertisement to attract. After calculating the density of the India/Pakistan/Sri Lanka population for each borough, we could see which borough in London regroups the most migrants.

Given the requirements from our restaurateur, we filtered the boroughs based on crime data, gross pay, median house price, and density of migrant from India/Pakistan/Sri Lanka. As requested, our filtering conditions suggested low crime rate for safety, high gross pay for customer's purchasing power, low median house price for affordable rent, and high migrant density for proximity to customer base. Only 2 boroughs satisfied the conditions: Harrow and Redbridge.

The focus being now on Harrow and Redbridge, we found all the neighborhoods of these boroughs along with their geospatial coordinates. We use Foursquare to retrieve information about all the Indian restaurants within each neighborhood. We particularly focused on the number of restaurants, their number of likes, and their rating. The purpose being to examine the competition, represented by the number of restaurants and their ratings, and the population response to Indian cuisine, or population level of interest, represented by the number of likes.

We could finally cluster the neighborhoods based on the mentioned parameters. We chose to have a total of 6 clusters, each having a different profile depending on their competition level and their population level of interest. A lot of the neighborhoods are in cluster '0', which has low competition, but also poor level of interest. We will not recommend these neighborhoods to our restaurateur because he mentioned that he would like to locate his restaurant in an area where people already show interest to Indian cuisine. Plus, the marketing strategy will work better when there is some moderate level of competition. We would recommend the neighborhoods in cluster '2', which fit better our restaurateur conditions. They are 14 neighborhoods in cluster '2', which is plenty to choose from.

6 Conclusion

In this project, we used demographic data and data from Foursquare in order to help a restaurateur decide the location of his future Indian restaurant in London. Using the demographic data for all boroughs in London, we filtered out the boroughs that did not meet our restaurateur's conditions with respect to safety, customer's purchasing power, real estate affordability, and Indian population presence. We then focused on the neighborhoods inside the remaining boroughs. Using Foursquare, we retrieved the information about all Indian restaurants within each neighborhood. We then performed the clustering method to group the neighborhood based on their similarities of their Indian restaurant’s profiles.

Each cluster has a different profile regarding competition or population interest to Indian cuisine. Our restaurateur would be able to make an intelligent decision on where to locate his new restaurant.