Introduction/Business Problem

London is a global epicentre for business activities and a ‘must-visit’ city on everybody’s bucket list whether on business or personal trip. Indian software professionals always find themselves relocating to London, at times with family, for short to medium term duration (ranging from few weeks to months). Along with the joy of relocating to a new place, these expats are often overwhelmed by challenges and questions about social connections, racism, food, culture, living standards and safety. To avoid the cultural shock, they need to make an informed decision on where to live. The goal of this project is to provide key information to these short-term expats to help them decide on safe and immigrant friendly neighbourhoods.

People generally tend to prefer to stay closer to their own community for few simple reasons: Emotional support, getting along, food and culture.

This project will explore London Borough based on the following criteria to suggest clusters of Borough that are suitable to an individuals’ preference.

1. Indian or Asian dominated Boroughs
2. Boroughs with least reported crimes of any type
3. Boroughs with large number of Indian restaurants and stores
4. Boroughs popular for Indian restaurants and stores

Ideal audience is short term Indian expats (mostly Indian IT contractors who are relocating to Greater London with or without family)

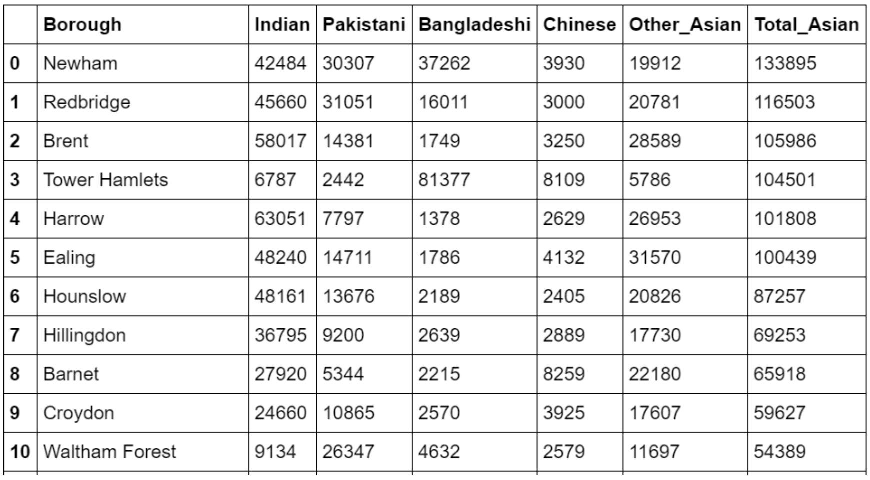
Data

1. Ethnic groups in London - https://en.wikipedia.org/wiki/Ethnic_groups_in_London
2. Population in London Boroughs - https://data.london.gov.uk/download/2011-census-demography/62f62c4d-eb60-4846-9efd-1e1373641452/london-unrounded-data.xls
3. 24 month crime data by London Borough - https://data.london.gov.uk/download/recorded_crime_summary/d2e9ccfc-a054-41e3-89fb-53c2bc3ed87a/MPS%20Borough%20Level%20Crime%20%28most%20recent%2024%20months%29.csv
4. Nearest most popular venues data for the boroughs from FourSquare API

Data Source

Data will be sourced from 4 different websites. Please note that it is very important to understand the data, its completeness, gaps, caveats and any errors to arrive at robust insights.

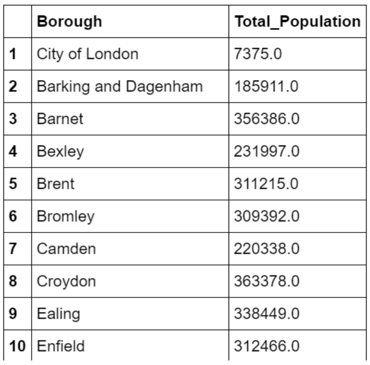
1. Ethnic groups in London – all the ethnic group data is scraped, cleaned and converted into relevant data form (Object to Float for numbers etc.) from this website (https://en.wikipedia.org/wiki/Ethnic_groups_in_London)



	Borough	Indian	Pakistani	Bangladeshi	Chinese	Other_Asian	Total_Asian
0	Newham	42484	30307	37262	3930	19912	133895
1	Redbridge	45660	31051	16011	3000	20781	116503
2	Brent	58017	14381	1749	3250	28589	105986
3	Tower Hamlets	6787	2442	81377	8109	5786	104501
4	Harrow	63051	7797	1378	2629	26953	101808
5	Ealing	48240	14711	1786	4132	31570	100439
6	Hounslow	48161	13676	2189	2405	20826	87257
7	Hillingdon	36795	9200	2639	2889	17730	69253
8	Barnet	27920	5344	2215	8259	22180	65918
9	Croydon	24660	10865	2570	3925	17607	59627
10	Waltham Forest	9134	26347	4632	2579	11697	54389

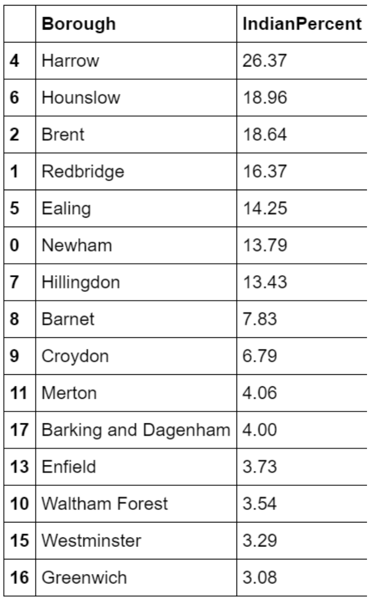
Absolute numbers may not be relevant as the proportion (%) of a particular ethnic group in that population will paint a better picture of concentration for each borough. We need total population by each borough which we will source from a different website in the next section.

2. Population in London Boroughs – We will download and use the following excel file from official London Data website and use 'Persons' sheet for population (https://data.london.gov.uk/download/2011-census-demography/62f62c4d-eb60-4846-9efd-1e1373641452/london-unrounded-data.xls).



	Borough	Total_Population
1	City of London	7375.0
2	Barking and Dagenham	185911.0
3	Barnet	356386.0
4	Bexley	231997.0
5	Brent	311215.0
6	Bromley	309392.0
7	Camden	220338.0
8	Croydon	363378.0
9	Ealing	338449.0
10	Enfield	312466.0

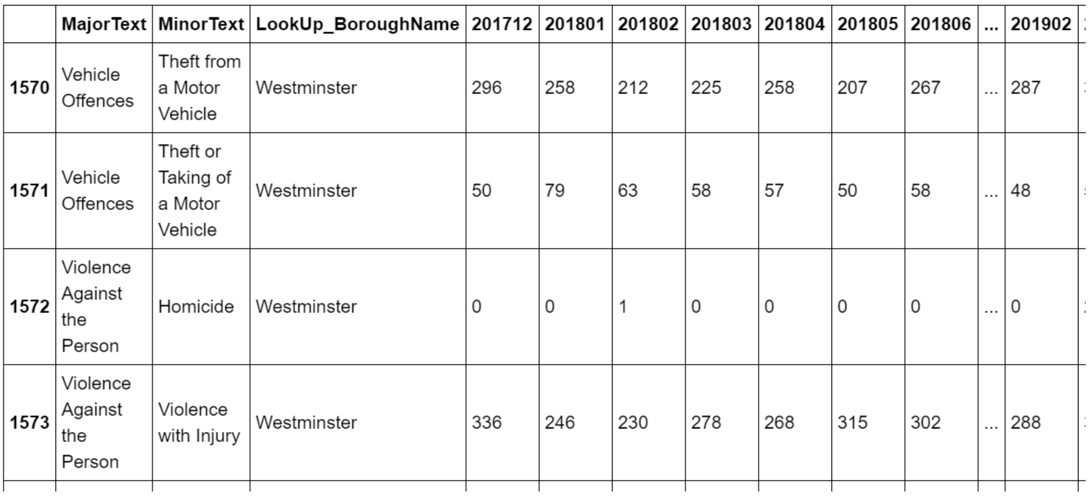
Now we have the population by ethnicity for each borough and total population by borough, so we can calculate Indian concentration (% Indian population).



	Borough	IndianPercent
4	Harrow	26.37
6	Hounslow	18.96
2	Brent	18.64
1	Redbridge	16.37
5	Ealing	14.25
0	Newham	13.79
7	Hillingdon	13.43
8	Barnet	7.83
9	Croydon	6.79
11	Merton	4.06
17	Barking and Dagenham	4.00
13	Enfield	3.73
10	Waltham Forest	3.54
15	Westminster	3.29
16	Greenwich	3.08

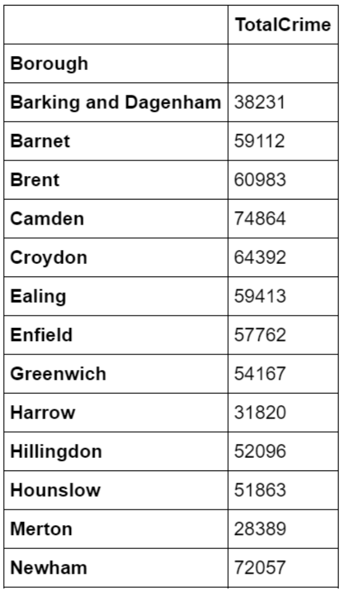
3. Crime data of London Boroughs (24 months) – from official London website https://data.london.gov.uk/download/recorded_crime_summary/d2e9ccfc-a054-41e3-89fb-53c2bc3ed87a/MPS%20Borough%20Level%20Crime%20%28most%20recent%2024%20months%29.csv

   It has data by crime type and subtype for the past 24 months (1 column for each month starting from Dec 2017)



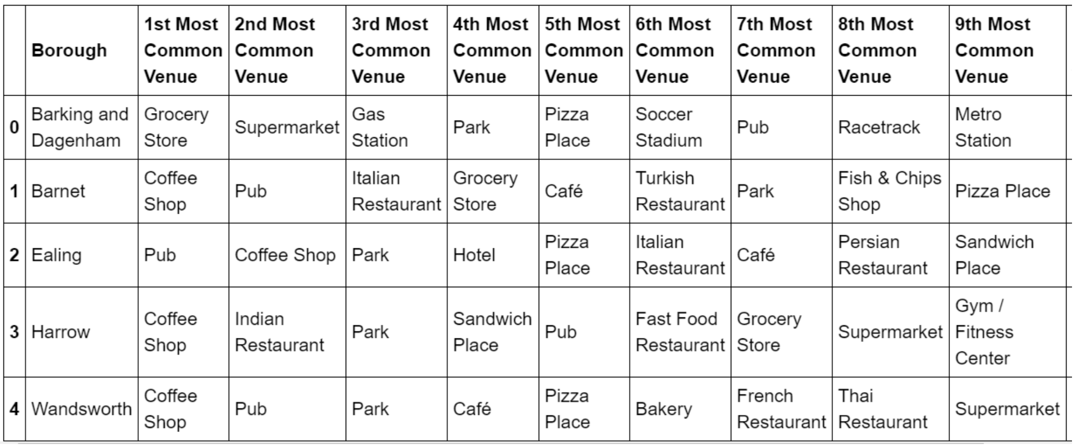
	MajorText	MinorText	LookUp_BoroughName	201712	201801	201802	201803	201804	201805	201806	...	201902
1570	Vehicle Offences	Theft from a Motor Vehicle	Westminster	296	258	212	225	258	207	267	...	287
1571	Vehicle Offences	Theft or Taking of a Motor Vehicle	Westminster	50	79	63	58	57	50	58	...	48
1572	Violence Against the Person	Homicide	Westminster	0	0	1	0	0	0	0	...	0
1573	Violence Against the Person	Violence with Injury	Westminster	336	246	230	278	268	315	302	...	288

   We will summarize the data over past 24 months and by all crime type (Major text and minor text) for each borough.



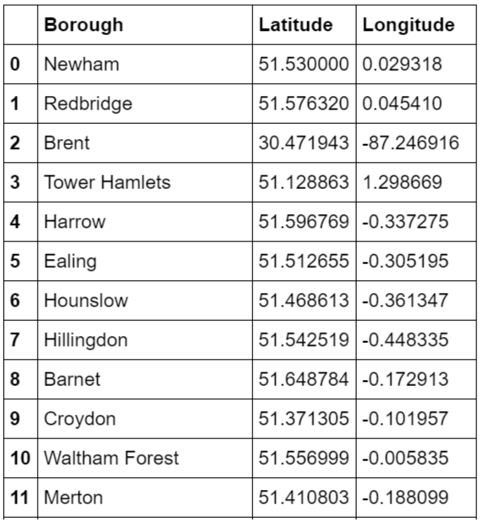
	TotalCrime
Borough	
Barking and Dagenham	38231
Barnet	59112
Brent	60983
Camden	74864
Croydon	64392
Ealing	59413
Enfield	57762
Greenwich	54167
Harrow	31820
Hillingdon	52096
Hounslow	51863
Merton	28389
Newham	72057

4. Now we will use Four Square API to get 200 venue listings for each borough within the radius of 2.5 kms and find the top 10 most popular venue types.



	Borough	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
0	Barking and Dagenham	Grocery Store	Supermarket	Gas Station	Park	Pizza Place	Soccer Stadium	Pub	Racetrack	Metro Station
1	Barnet	Coffee Shop	Pub	Italian Restaurant	Grocery Store	Café	Turkish Restaurant	Park	Fish & Chips Shop	Pizza Place
2	Ealing	Pub	Coffee Shop	Park	Hotel	Pizza Place	Italian Restaurant	Café	Persian Restaurant	Sandwich Place
3	Harrow	Coffee Shop	Indian Restaurant	Park	Sandwich Place	Pub	Fast Food Restaurant	Grocery Store	Supermarket	Gym / Fitness Center
4	Wandsworth	Coffee Shop	Pub	Park	Café	Pizza Place	Bakery	French Restaurant	Thai Restaurant	Supermarket

Finally, to map them out we will also generate latitude and longitudes of all the boroughs

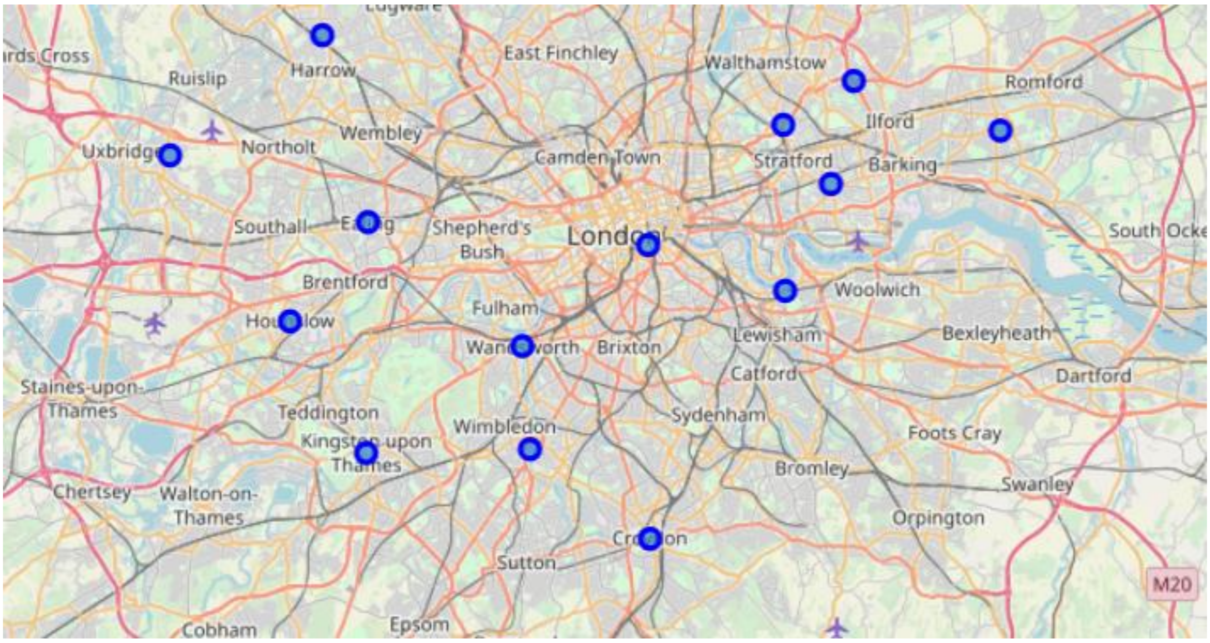


	Borough	Latitude	Longitude
0	Newham	51.530000	0.029318
1	Redbridge	51.576320	0.045410
2	Brent	30.471943	-87.246916
3	Tower Hamlets	51.128863	1.298669
4	Harrow	51.596769	-0.337275
5	Ealing	51.512655	-0.305195
6	Hounslow	51.468613	-0.361347
7	Hillingdon	51.542519	-0.448335
8	Barnet	51.648784	-0.172913
9	Croydon	51.371305	-0.101957
10	Waltham Forest	51.556999	-0.005835
11	Merton	51.410803	-0.188099

Methodology

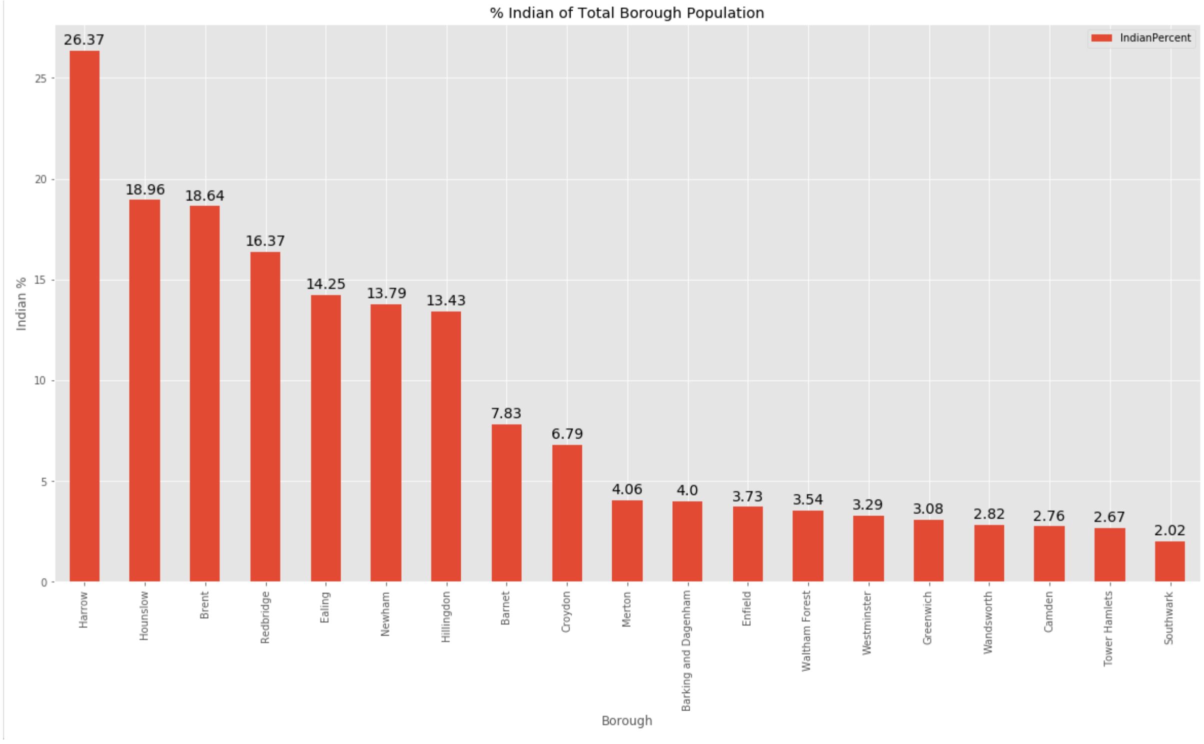
Initially, we will do *Exploratory Data analysis* to look at all London Boroughs. It is very important to understand the data for feature selection and modelling.

These are all the London boroughs that we will be analysing.



Then, we will analyse and segment the boroughs by the selection criteria provided (ethnicity, number of crimes and population).

Since our target audience is short term Indian expats, let’s look at the concentration of Indian population in these boroughs. Data source and calculations are already discussed in the ‘Data’ section of the report.

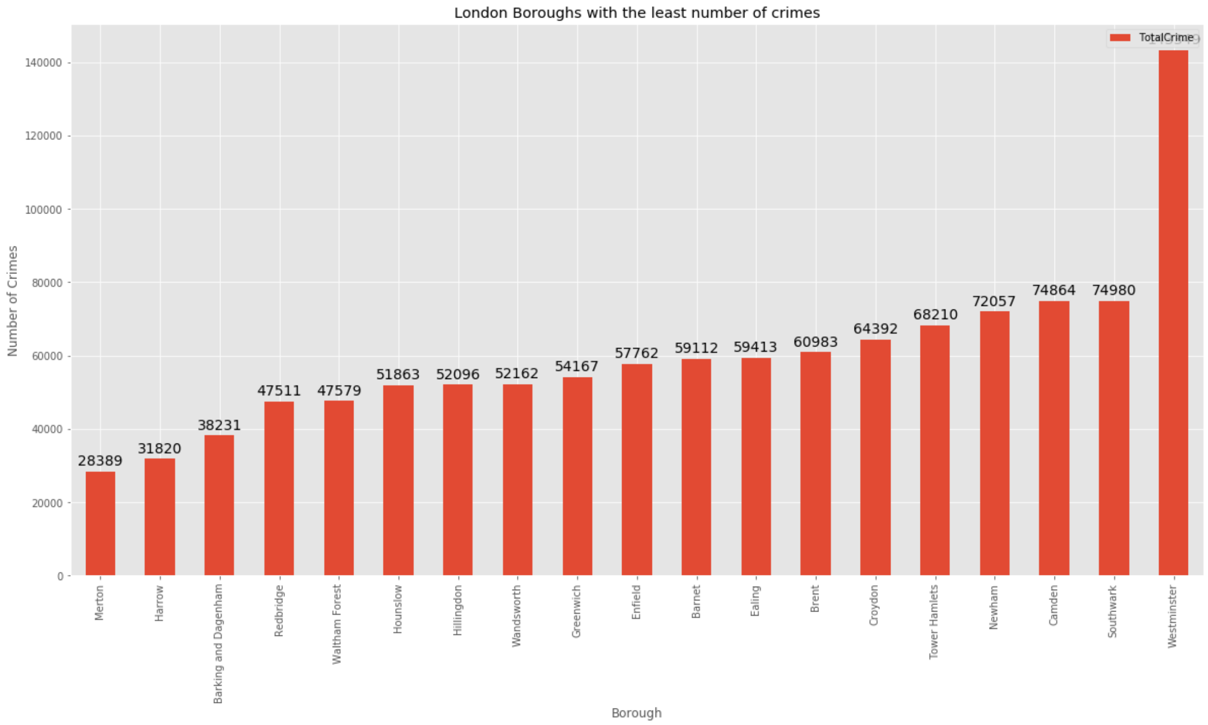


% Indian Concentration

We would be interested in boroughs with high concentration of Indian population. Here we are looking at % and not the whole numbers as we want to focus on concentration and some broughs are densely populated than the others, so even a higher number of Indian populations would not suggest higher concentration.

Harrow, Hunslow, Brent, Ealing etc have the highest concentration of Indian population.

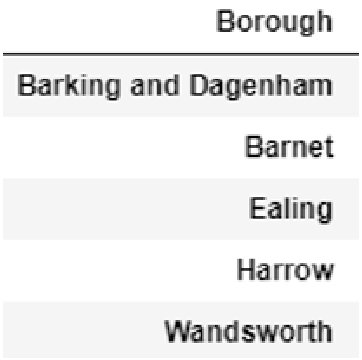
Next, we will analyse the crime rates in the boroughs. Here, we are interested in the absolute number as the borough with least number of reported crimes will be the safest.



Crime by borough

Merton with 28,369 reported crime over a 24-month period has lowest crime and would be interesting neighbour but if we look at '% Indian Concentration' graph, Merton doesn't have high concentration of Indian population.

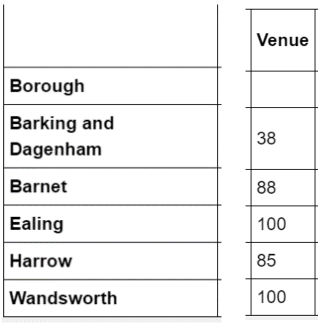
Combining the insights from the above two graphs, we can narrow down 5 boroughs that are of interest to us and would require further analysis on Indian restaurant and popularity.



Borough
Barking and Dagenham
Barnet
Ealing
Harrow
Wandsworth

Top 5 finalist Boroughs

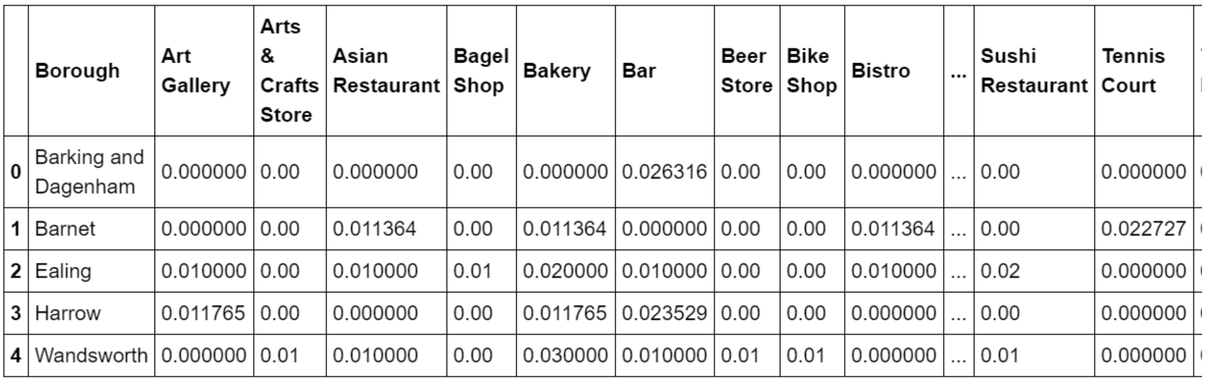
Now using Foursquare API we will look for top 200 venues within a radius of 2.5 kms in our selected boroughs. We look at top 10 venue by choice in each borough to identify if Indian restaurants are among the popular choice of the boroughs.



	Venue
Borough	
Barking and Dagenham	38
Barnet	88
Ealing	100
Harrow	85
Wandsworth	100

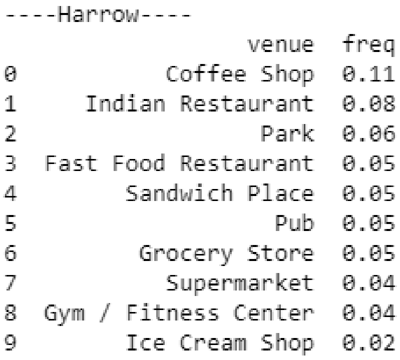
Wandsworth and Ealing have most venues, Barnet and Harrow are not far behind.

We will use one hot encoding to find the most popular venues.

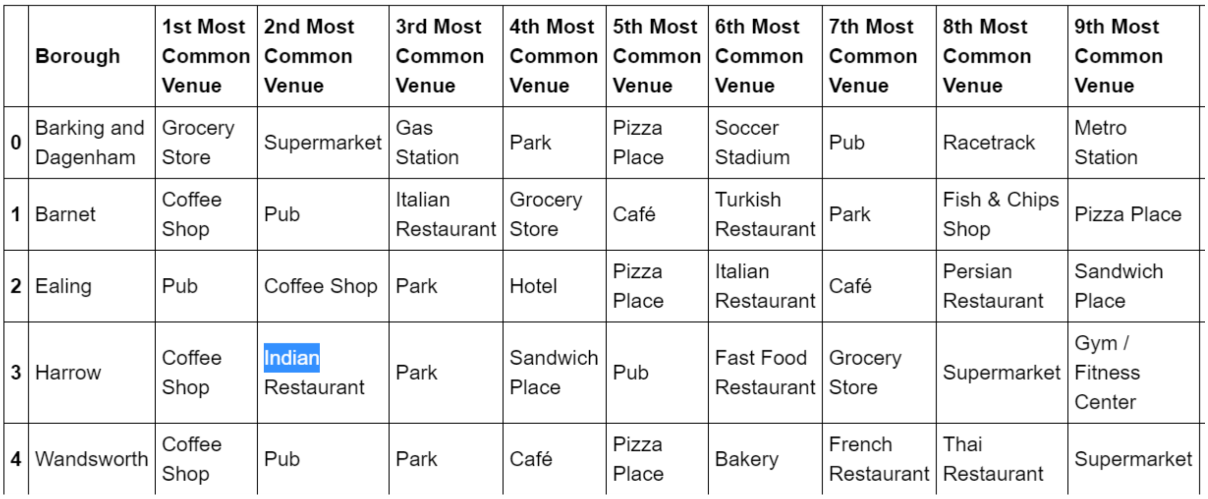


	Borough	Art Gallery	Arts & Crafts Store	Asian Restaurant	Bagel Shop	Bakery	Bar	Beer Store	Bike Shop	Bistro	...	Sushi Restaurant	Tennis Court
0	Barking and Dagenham	0.000000	0.00	0.000000	0.00	0.000000	0.026316	0.00	0.00	0.000000	...	0.00	0.000000
1	Barnet	0.000000	0.00	0.011364	0.00	0.011364	0.000000	0.00	0.00	0.011364	...	0.00	0.022727
2	Ealing	0.010000	0.00	0.010000	0.01	0.020000	0.010000	0.00	0.00	0.010000	...	0.02	0.000000
3	Harrow	0.011765	0.00	0.000000	0.00	0.011765	0.023529	0.00	0.00	0.000000	...	0.00	0.000000
4	Wandsworth	0.000000	0.01	0.010000	0.00	0.030000	0.010000	0.01	0.01	0.000000	...	0.01	0.000000

We will use this to find the most popular categories in these boroughs.



```
----Harrow----
                    venue  freq
0             Coffee Shop  0.11
1       Indian Restaurant  0.08
2                    Park  0.06
3    Fast Food Restaurant  0.05
4          Sandwich Place  0.05
5                     Pub  0.05
6           Grocery Store  0.05
7             Supermarket  0.04
8    Gym / Fitness Center  0.04
9          Ice Cream Shop  0.02
```



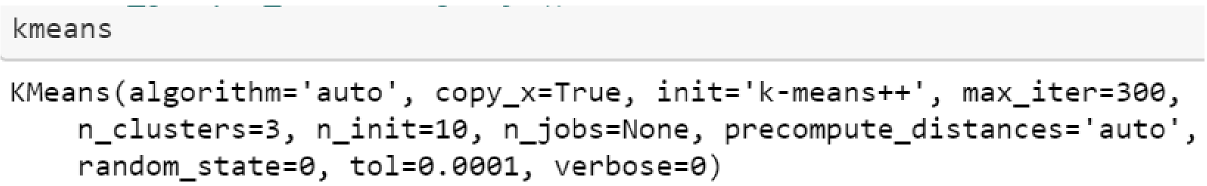
	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	Barking and Dagenham	Grocery Store	Supermarket	Gas Station	Park	Pizza Place	Soccer Stadium	Pub	Racetrack	Metro Station
1	Barnet	Coffee Shop	Pub	Italian Restaurant	Grocery Store	Café	Turkish Restaurant	Park	Fish & Chips Shop	Pizza Place
2	Ealing	Pub	Coffee Shop	Park	Hotel	Pizza Place	Italian Restaurant	Café	Persian Restaurant	Sandwich Place
3	Harrow	Coffee Shop	Indian Restaurant	Park	Sandwich Place	Pub	Fast Food Restaurant	Grocery Store	Supermarket	Gym / Fitness Center
4	Wandsworth	Coffee Shop	Pub	Park	Café	Pizza Place	Bakery	French Restaurant	Thai Restaurant	Supermarket

Since we do not have a labelled dataset, we cannot use supervised learning, we will be using unsupervised Machine Learning Algorithm. We will use K-Means clustering algorithm to cluster the boroughs based on the similarity within the clusters (and differences with other clusters).

Results

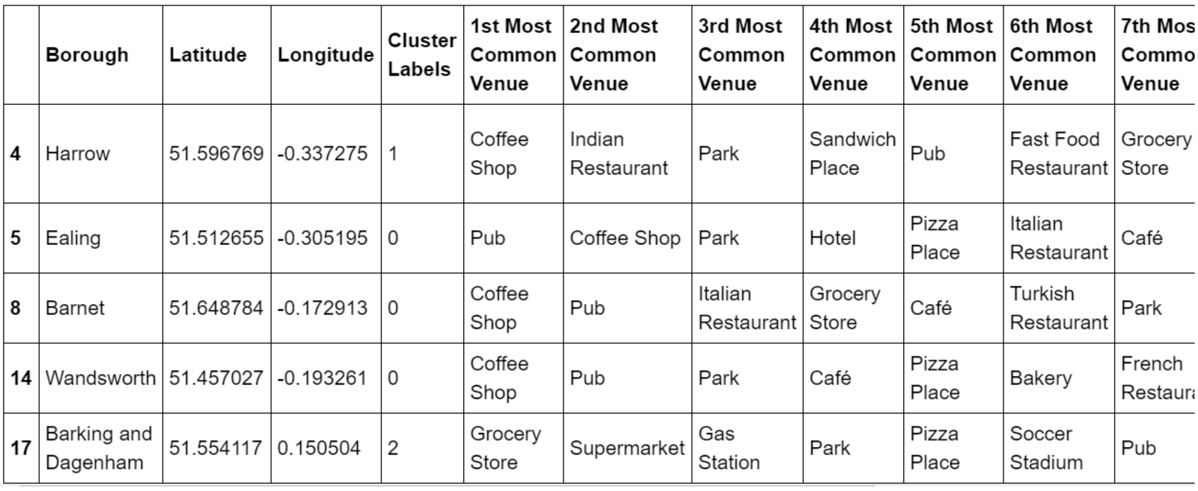
Cluster analysis (K-means)

We will create 3 clusters



```
kmeans
```

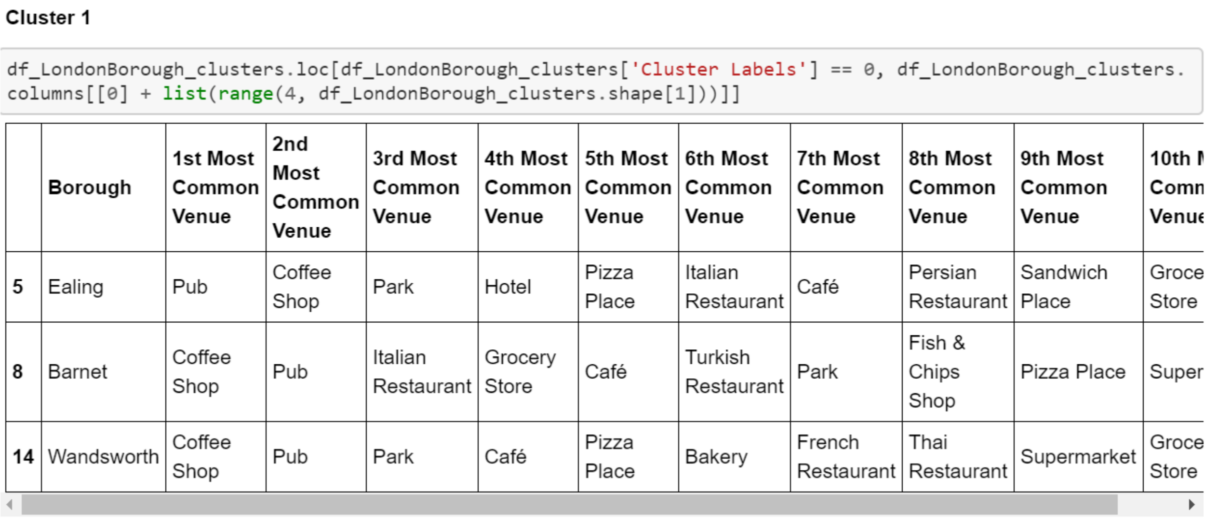
```
KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
    n_clusters=3, n_init=10, n_jobs=None, precompute_distances='auto',
    random_state=0, tol=0.0001, verbose=0)
```



	Borough	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 Mos Commo Venue
4	Harrow	51.596769	-0.337275	1	Coffee Shop	Indian Restaurant	Park	Sandwich Place	Pub	Fast Food Restaurant	Grocery Store
5	Ealing	51.512655	-0.305195	0	Pub	Coffee Shop	Park	Hotel	Pizza Place	Italian Restaurant	Café
8	Barnet	51.648784	-0.172913	0	Coffee Shop	Pub	Italian Restaurant	Grocery Store	Café	Turkish Restaurant	Park
14	Wandsworth	51.457027	-0.193261	0	Coffee Shop	Pub	Park	Café	Pizza Place	Bakery	French Restaura
17	Barking and Dagenham	51.554117	0.150504	2	Grocery Store	Supermarket	Gas Station	Park	Pizza Place	Soccer Stadium	Pub

None other borough had Indian restaurants rated in top 10 categories except Harrow.

(I did cluster analysis for all the boroughs also to understand how the clusters look like and what can we derive from the clusters. The results were very similar.)



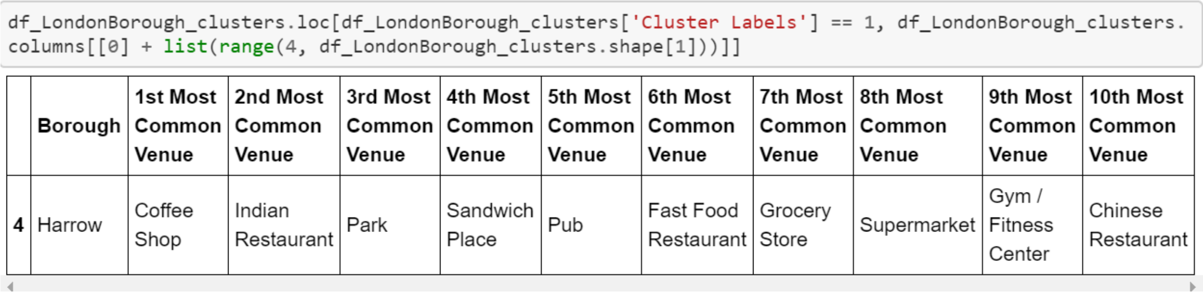
Cluster 1

```
df_LondonBorough_clusters.loc[df_LondonBorough_clusters['Cluster Labels'] == 0, df_LondonBorough_clusters.
columns[[0] + list(range(4, df_LondonBorough_clusters.shape[1]))]]
```

	Borough	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 M Comm Venue
5	Ealing	Pub	Coffee Shop	Park	Hotel	Pizza Place	Italian Restaurant	Café	Persian Restaurant	Sandwich Place	Groce Store
8	Barnet	Coffee Shop	Pub	Italian Restaurant	Grocery Store	Café	Turkish Restaurant	Park	Fish & Chips Shop	Pizza Place	Super
14	Wandsworth	Coffee Shop	Pub	Park	Café	Pizza Place	Bakery	French Restaurant	Thai Restaurant	Supermarket	Groce Store

Cluster 1 presents 3 boroughs with a mix of moderate crime rates, Indian concentration and other amenities. This would be interest to somebody who enjoys different cuisine (Thai, Chinese, Italian).

Cluster 2

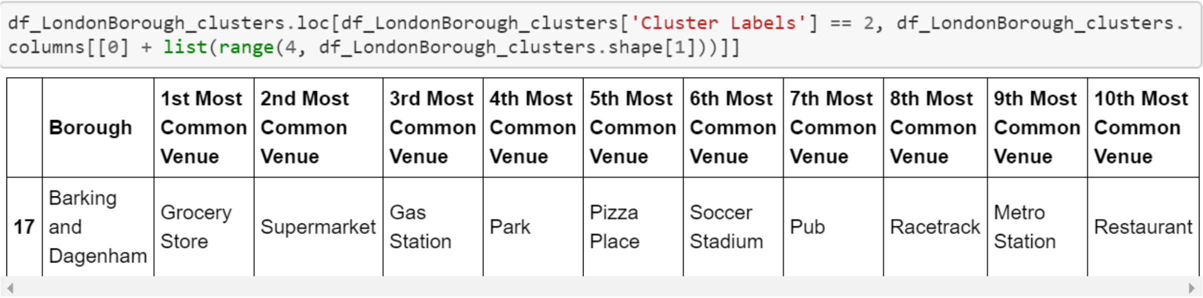


```
df_LondonBorough_clusters.loc[df_LondonBorough_clusters['Cluster Labels'] == 1, df_LondonBorough_clusters.
columns[[0] + list(range(4, df_LondonBorough_clusters.shape[1]))]]
```

	Borough	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
4	Harrow	Coffee Shop	Indian Restaurant	Park	Sandwich Place	Pub	Fast Food Restaurant	Grocery Store	Supermarket	Gym / Fitness Center	Chinese Restaurant

Cluster 2 presents only Harrow as an option. Harrow has the lowest crime rate, highest Indian concentration and has most popular Indian restaurant category and other amenities such as gym and grocery stores.

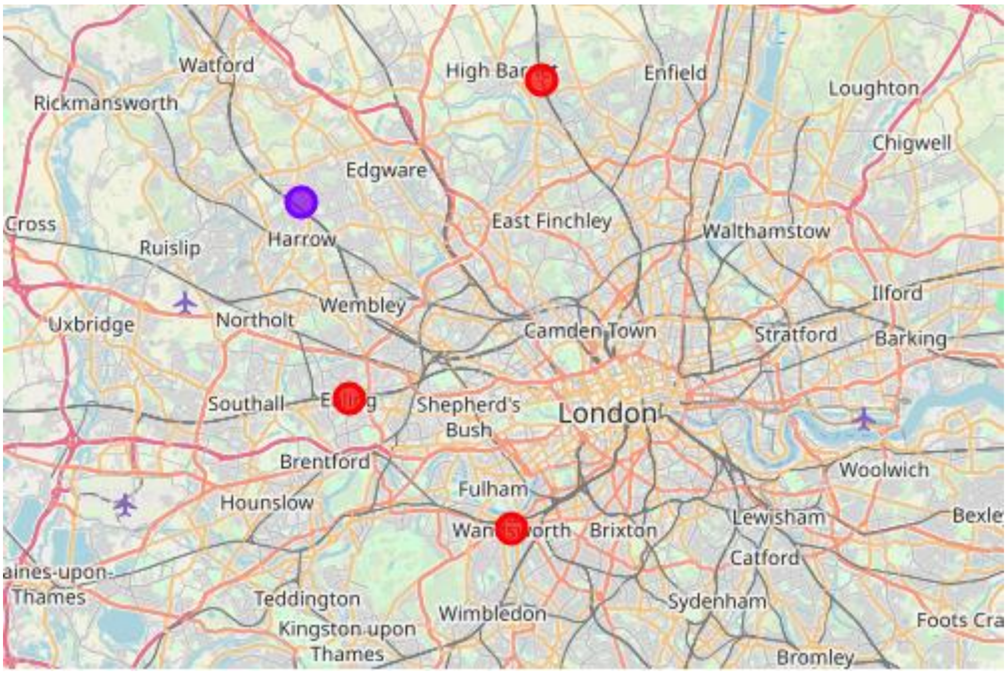
Cluster 3



```
df_LondonBorough_clusters.loc[df_LondonBorough_clusters['Cluster Labels'] == 2, df_LondonBorough_clusters.
columns[[0] + list(range(4, df_LondonBorough_clusters.shape[1]))]]
```

	Borough	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
17	Barking and Dagenham	Grocery Store	Supermarket	Gas Station	Park	Pizza Place	Soccer Stadium	Pub	Racetrack	Metro Station	Restaurant

Cluster 3 presents Barking and Dagenham as an option. Barking and Dagenham also has one of the lowest crimes (3rd lowest), modest Indian concentration and good selection of other amenities such as grocery stores and metro station.



Cluster Analysis

Discussion

Every individual has different needs and priorities. The three clusters provide a snapshot of what each borough provides in its vicinity. If an individual like different cuisines (Cafes, Thai, Chinese, Italian, Turkish) then he may choose an option from cluster 1. If train commute is a deciding factor for an individual then cluster 3 option would be a good option for him as Metro station is listed as 9th top venue in this cluster. Boroughs in these clusters have moderate Indian concentration and crime rates.

If an individual prefers great Indian restaurant, along with other features like gym, park, super market, then cluster 2 would be ideal for him. Borough in this cluster has the highest Indian concentration and is the safest.

Please note all the options provided in these clusters are in safe areas, but one may have to look at crime rate analysis, Indian population concentration graphs and popularity of venues together before making the final choice.

Conclusion

With Park, Gym and Grocery store, coupled with high Indian density, lowest crime rates – it looks like Harrow is a good suggestion according to the criteria provided by a short-term Indian expat.

Until now many Indian expats have relied on word of mouth and basic internet search as an alternative approach to identify safe areas to consider but that approach was always very subjective based on the liking and bias of the recommending person. This project provides a data driven approach to recommendations and insights to make an informed decision.

Future work to do

It would be nice to add following different data to provide more robust recommendations

1. Travel time to work place
2. Housing availability (short term house listings)
3. Housing prices
4. Cost of living index