LOCATION RECOMMENDATION FOR INDIAN RESTAURANT IN HARROW

1. INTRODUCTION

When you consider the demography of London, you will see a rainbow—a community of multi-color multi-ethnic groups. From the South of the River Thames and across to the other side of it, the diversity of London is the strength every business from across the walks of life has taken opportunity of Growth.

In London Borough of Harrow, 63.8% of its population from the BME (Black and Minority Ethnic) communities, with the largest group being of Indian ethnicity, 26.4% (specifically those from Gujarat and South India). But something that really lack its essence is the availability of a high-end Indian Restaurant that caters beyond the food but comes with the ambience of the service. With the arrays of restaurants in Harrow, there are peculiarly lacking the finest restaurant that can answer the call within certain areas of interest when it comes to Indian cuisines and flavors.

To get a very good location for a restaurant that meet this need, the Harrow area, which has dense Indian population is explored through clustering and segmentation based on the Borough Coordinates & Indian Population in Boroughs and proximity to supplies.

This project will rely on public data from Wikipedia and Foursquare. Libraries—For convenience, all the libraries are presented at the beginning

2. DATA SECTION

2.1. DATA SETS: WIKI PAGES

Wiki page – List of London boroughs

The London Area consists of 32 Boroughs and the “City of London”. Our data – Boroughs, area, Latitude, longitude, etc., will be from the wiki page :
https://en.wikipedia.org/wiki/List_of_London_boroughs

The focus of this project will be the neighborhoods are that are within London Borough of Harrow

Wiki page – Indian community of London

The percentage of Indian population in various London Boroughs are collected from the wiki page: https://en.wikipedia.org/wiki/Indian_community_of_London

Wiki page – List of areas of London

The London area data is scrapped from the wiki - List_of_areas_of_London: https://en.wikipedia.org/wiki/List_of_areas_of_London

The BeautifulSoup package is used to scrap the needed data from Wikipedia.

2.2. DATA SETS: FOURSQUARE API

The Foursquare API will be used to obtain the Harrow Area venues for the geographical location data. These will be used to explore the neighborhoods of Harrow accordingly.

The venues within the neighborhoods of Harrow like the area's restaurants and proximity to amenities would be correlated. Also, accessibility and ease of supplies would be considered as it relates to venues.

2.3. DATA SETS: GEOCODER PACKAGE

In obtaining the location data of the locations, the Geocoder package is used with the ArcGIS geocoder to obtain the latitude and longitude of the needed locations.

These will help to create a new data frame that will be used subsequently for the Harrow areas.

2.4. PYTHON LIBRARY FILES

The Below Libraries are used:

- Pandas - Library for Data Analysis
- NumPy – Library to handle data in a vectorized manner
- JSON – Library to handle JSON files
- Folium – Map rendering Library
- Matplotlib – Python Plotting Module
- Geopy – To retrieve Location Data
- Requests – Library to handle http requests
- Sklearn – Python machine learning Library

2.5. FOLIUM LIBRARY

Python visualization library is used to visualize the neighborhoods cluster distribution of Chicago city over an interactive leaflet map. Extensive comparative analysis of two randomly picked neighborhoods world be carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn.

2.6. DATA PREPARATION

London Indian Population:

	Community	Percentage of Indian population
0	Harrow	26.4
1	Hounslow	19.0
2	Brent	18.6
3	Redbridge	16.4
4	Ealing	14.3

London boroughs Data from Wiki Page

	Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est)[1]	Co-ordinates
0	Barking and Dagenham [note 1]			Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194,352	51°33'39"N 0°09'21"E / 51.5607°N 0.1557°E /...
1	Barnet			Barnet London Borough Council	Conservative	North London Business Park, Oakleigh Road South	33.49	369,088	51°37'31"N 0°09'06"W / 51.6252°N 0.1517°W /...
2	Bexley			Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	236,687	51°27'18"N 0°09'02"E / 51.4549°N 0.1505°E /...
3	Brent			Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	317,264	51°33'32"N 0°16'54"W / 51.5588°N 0.2817°W /...
4	Bromley			Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	317,899	51°24'14"N 0°01'11"E / 51.4039°N 0.0198°E /...

London borough with Max Area & Population

```
Borough with Maximum Population :
Borough       Croydon
Area            33.41
Population     372752
Name: 6, dtype: object

Borough with Maximum Area :
Borough       Bromley
Area            57.97
Population     317899
Name: 4, dtype: object
```

London borough with Max Indian Population

```
Borough with Maximum Indian Population :
Community              Harrow
Indian population %      26.4
Name: 0, dtype: object
```

London borough & Neighborhood Data exploration

	Location	Borough	Postcode	Post-town
0	Abbey Wood	Bexley, Greenwich	SE2	LONDON
1	Acton	Ealing, Hammersmith and Fulham	W3	LONDON
2	Acton	Ealing, Hammersmith and Fulham	W4	LONDON
3	Addington	Croydon	CR0	CROYDON
4	Addiscombe	Croydon	CR0	CROYDON

Data Clean-Up

	Neighborhood	Borough	Postcode	Post-town
0	Belmont	Harrow	HA3	HARROW, STANMORE
1	Belmont	Harrow	HA7	HARROW, STANMORE
2	Harrow	Harrow	HA1	HARROW
3	Harrow on the Hill	Harrow	HA1	HARROW
4	Harrow Weald	Harrow	HA3	HARROW
5	Hatch End	Harrow	HA5	PINNER
6	North Harrow	Harrow	HA1	HARROW

Geocoder - to obtain Coordinate for each area

We are using geocoder to store the location data—latitude and longitude as follows. The obtained coordinates are then joined to create new data frame.

	Neighborhood	Borough	Postcode	Post-town	Latitude	Longitude
0	Belmont	Harrow	HA3	HARROW, STANMORE	45.877912	5.652908
1	Belmont	Harrow	HA7	HARROW, STANMORE	45.877912	5.652908
2	Harrow	Harrow	HA1	HARROW	51.596769	-0.337275
3	Harrow on the Hill	Harrow	HA1	HARROW	51.579270	-0.336656
4	Harrow Weald	Harrow	HA3	HARROW	51.604786	-0.340485
5	Hatch End	Harrow	HA5	PINNER	51.608440	-0.373548
6	North Harrow	Harrow	HA1	HARROW	51.585162	-0.363176
7	North Harrow	Harrow	HA2	HARROW	51.585162	-0.363176
8	Pinner	Harrow	HA5	PINNER	51.596871	-0.377014
9	Rayners Lane	Harrow	HA5	PINNER	51.576714	-0.370300
10	South Harrow	Harrow	HA2	HARROW	51.564652	-0.352221
11	Stanmore	Harrow	HA7	STANMORE	51.617421	-0.309511
12	Wealdstone	Harrow	HA3	HARROW	51.593635	-0.329476
13	West Harrow	Harrow	HA2	HARROW	51.579585	-0.353069

3. METHODOLOGY

3.1. DATA EXPLORATION

Single Neighborhood

An initial exploration of a single Neighborhood within the London area was done to examine the Foursquare workability. The West Harrow is used for our Single Neighborhood exploration.

```
Latitude and longitude values of West Harrow are 51.5795852, -0.3530692.
```

Let's explore the top 100 venues that are within a 2000 meters radius of West Harrow. And then, let's create the GET request URL, and then the URL is named.

Results are shown below:


```
{'meta': {'code': 200, 'requestId': '5cbbec199fb6b776fcf95781'},
 'response': {'headerLocation': 'West Harrow',
  'headerFullLocation': 'West Harrow, London',
  'headerLocationGranularity': 'neighborhood',
  'totalResults': 41,
  'suggestedBounds': {'ne': {'lat': 51.59758521800001,
    'lng': -0.32415766151483744},
   'sw': {'lat': 51.56158518199998, 'lng': -0.3819807384851626}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '4bd1ab0b046076b074307271',
       'name': 'West Harrow Park',
       'location': {'address': 'The Ridgeway',
        'lat': 51.577553917349476,
        'lng': -0.354538592936466,
        'labeledLatLngs': [{'label': 'display',
          'lat': 51.577553917349476,
          'lng': -0.354538592936466}],
        'distance': 247,
        'postalCode': 'HA2 7ED',
        'cc': 'GB',
        'city': 'Harrow on the Hill',
        'state': 'Greater London',
        'country': 'United Kingdom',
```

The result is then cleaned up from json to a structured **pandas** data frame as shown below:

	name	categories	lat	lng
0	West Harrow Park	Park	51.577554	-0.354539
1	Harrow Recreation Ground	Park	51.585777	-0.345683
2	Twist Ice Cream	Ice Cream Shop	51.586708	-0.361651
3	Nando's	Portuguese Restaurant	51.581665	-0.333119
4	Kebab Land	Middle Eastern Restaurant	51.580034	-0.335987
5	the chocolate room	Coffee Shop	51.580970	-0.333788
6	Saravanna Bhavan	Indian Restaurant	51.572855	-0.371405
7	Waitrose & Partners	Supermarket	51.567635	-0.352037
8	The Dolls House	Café	51.571992	-0.338525
9	The Gym	Gym / Fitness Center	51.586388	-0.361284

Top 5 venues in West Harrow

	Count
Indian Restaurant	5
Pub	3
Ice Cream Shop	2
Portuguese Restaurant	2
Supermarket	2

Interestingly, we have Indian restaurant as top venue in West Harrow.

Multiple Neighborhood

Now let's explore (Multiple) Neighborhoods in the Harrow area. To do this, the function getNearbyVenues is used and it's created to repeat the same process for all neighborhoods.

```
print(london_venues.shape)
london_venues.head()
```

```
(357, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Harrow	51.596769	-0.337275	Everest Lounge	51.594233	-0.332192	Indian Restaurant
1	Harrow	51.596769	-0.337275	Waitrose & Partners	51.605329	-0.339664	Supermarket
2	Harrow	51.596769	-0.337275	Harrow Recreation Ground	51.585777	-0.345683	Park
3	Harrow	51.596769	-0.337275	The Bombay Central	51.604540	-0.339608	Indian Restaurant
4	Harrow	51.596769	-0.337275	Shree Krishna Vada Pav (SKVP)	51.587331	-0.332211	Indian Restaurant

The number of venues returned for each neighborhood is then explored as follows:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Harrow	30	30	30	30	30	30
Harrow Weald	29	29	29	29	29	29
Harrow on the Hill	30	30	30	30	30	30
Hatch End	30	30	30	30	30	30
North Harrow	60	60	60	60	60	60
Pinner	30	30	30	30	30	30
Rayners Lane	30	30	30	30	30	30
South Harrow	30	30	30	30	30	30
Stanmore	28	28	28	28	28	28
Wealdstone	30	30	30	30	30	30
West Harrow	30	30	30	30	30	30

There are 48 unique categories.

	Count
Indian Restaurant	41
Coffee Shop	34
Pub	29
Supermarket	23
Park	21
Gym / Fitness Center	18
Café	18
Grocery Store	15
Sandwich Place	14
Fast Food Restaurant	10

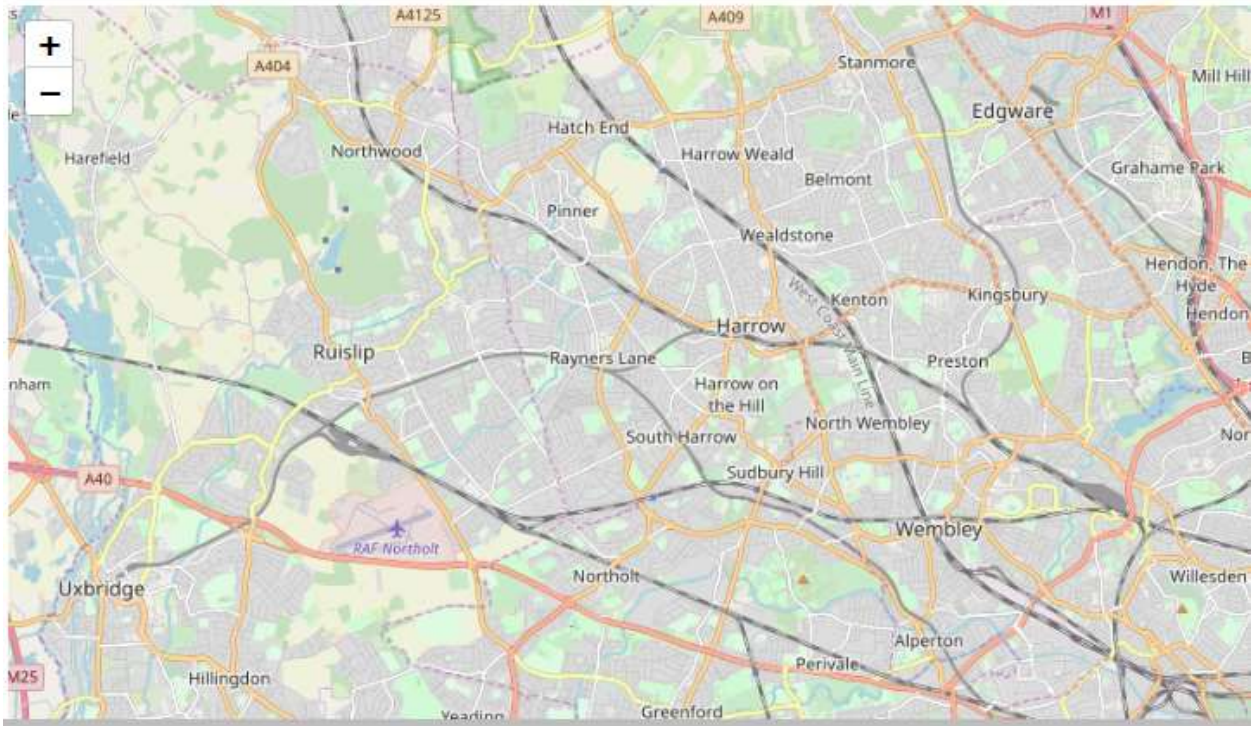
3.2. CLUSTERING

For this section, the neighborhoods in Harrow will be clustered based on the processed data obtained above.

Map Visualization—Using the geopy library, the latitude and longitude values of London is obtained.

```
The geograpical coordinate of Harrow are 51.5967688, -0.33727515543507.
```

```
map_london = folium.Map(location = [latitude, longitude], zoom_start = 12)
map_london
```



Harrow neighborhoods are then superimposed on top as shown below, still using the ‘folium’ library. Please note due to the location of the South East London, you might need to zoom to see the superimposed areas.

```
display(map_london)
```

Analyzing Each Neighborhood

In this section, the objective is to check and explore the venues in each neighborhood.

One Hot Encoding

```
(357, 49)
```

	Neighborhood	Athletics & Sports	Bakery	Bar	Beer Store	Bookstore	Burger Joint	Café	Chinese Restaurant	Coffee Shop	Deli / Bodega	Fast Food Restaurant	Fish & Chips Shop	Forest	French Restaurant	Furniture / Home Store	Gas Station
0	Harrow	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Harrow	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Harrow	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Harrow	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Harrow	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Check Indian Restaurants

```
london_grouped.loc[london_grouped['Indian Restaurant'] != 0]
```

	Neighborhood	Athletics & Sports	Bakery	Bar	Beer Store	Bookstore	Burger Joint	Café	Chinese Restaurant	Coffee Shop	Deli / Bodega	Fast Food Restaurant	Fish & Chips Shop	Forest	French Restaurant
0	Harrow	0.000000	0.033333	0.033333	0.000000	0.033333	0.0	0.066667	0.033333	0.133333	0.0	0.033333	0.000000	0.000000	0.0
1	Harrow Weald	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.034483	0.034483	0.068966	0.0	0.000000	0.000000	0.034483	0.0
2	Harrow on the Hill	0.000000	0.033333	0.066667	0.000000	0.033333	0.0	0.033333	0.033333	0.100000	0.0	0.066667	0.000000	0.000000	0.0
4	North Harrow	0.000000	0.000000	0.000000	0.033333	0.000000	0.0	0.066667	0.000000	0.133333	0.0	0.000000	0.000000	0.000000	0.0
6	Rayners Lane	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.033333	0.000000	0.066667	0.0	0.066667	0.000000	0.000000	0.0
7	South Harrow	0.033333	0.000000	0.033333	0.000000	0.000000	0.0	0.033333	0.000000	0.066667	0.0	0.033333	0.033333	0.000000	0.0
8	Stanmore	0.000000	0.000000	0.035714	0.000000	0.000000	0.0	0.035714	0.035714	0.071429	0.0	0.035714	0.035714	0.000000	0.0
9	Wealdstone	0.000000	0.033333	0.066667	0.000000	0.000000	0.0	0.066667	0.033333	0.133333	0.0	0.066667	0.000000	0.000000	0.0
10	West Harrow	0.000000	0.033333	0.033333	0.000000	0.000000	0.0	0.033333	0.033333	0.066667	0.0	0.033333	0.000000	0.000000	0.0

Regrouping and Category Statistics

We create a new panda data frame with 10 most common venues as shown below:

```
neighborhoods_venues_sorted.head()
```

	Neighborhood	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	Harrow	Coffee Shop	Indian Restaurant	Supermarket	Sandwich Place	Pub	Café	Grocery Store	Bar	Irish Pub	Fast Food Restaurant
1	Harrow Weald	Park	Supermarket	Steakhouse	Grocery Store	Gym	Gym / Fitness Center	Indian Restaurant	Coffee Shop	Museum	Golf Course
2	Harrow on the Hill	Coffee Shop	Pub	Indian Restaurant	Park	Supermarket	Bar	Fast Food Restaurant	Ice Cream Shop	Mediterranean Restaurant	Middle Eastern Restaurant
3	Hatch End	Supermarket	Italian Restaurant	Pub	Café	Grocery Store	Coffee Shop	Pizza Place	Performing Arts Venue	Karaoke Bar	Deli / Bodega
4	North Harrow	Coffee Shop	Indian Restaurant	Gym / Fitness Center	Park	Pub	Café	Pizza Place	Grocery Store	Ice Cream Shop	Italian Restaurant

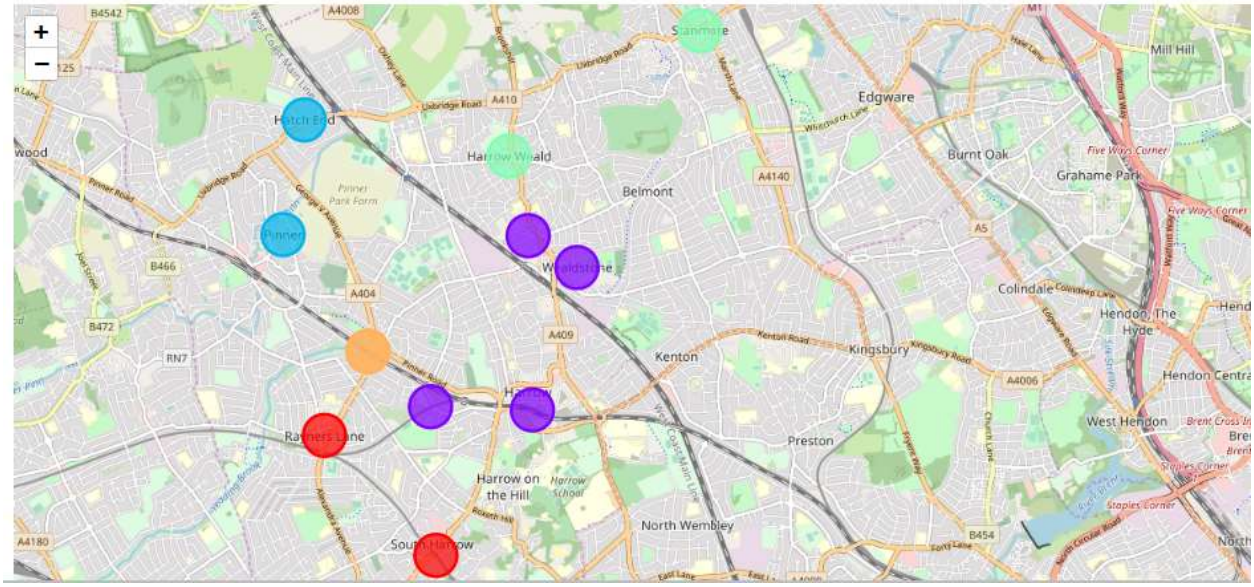
Clustering of Neighborhoods

We create the grouped clustering for the neighborhood as shown below:

	Neighborhood	Borough	Postcode	Post-town	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Harrow	Harrow	HA1	HARROW	51.596769	-0.337275	1	Coffee Shop	Indian Restaurant	Supermarket	Sandwich Place	Pub	Café	Grocery Store	Bar	Irish Pub	Fast Food Restaurant
1	Harrow on the Hill	Harrow	HA1	HARROW	51.579270	-0.336656	1	Coffee Shop	Pub	Indian Restaurant	Park	Supermarket	Bar	Fast Food Restaurant	Ice Cream Shop	Mediterranean Restaurant	Middle Eastern Restaurant
2	Harrow Weald	Harrow	HA3	HARROW	51.604786	-0.340485	3	Park	Supermarket	Steakhouse	Grocery Store	Gym	Gym / Fitness Center	Indian Restaurant	Coffee Shop	Museum	Golf Course
3	Hatch End	Harrow	HA5	PINNER	51.608440	-0.373548	2	Supermarket	Italian Restaurant	Pub	Café	Grocery Store	Coffee Shop	Pizza Place	Performing Arts Venue	Karaoke Bar	Deli / Bodega
4	North Harrow	Harrow	HA1	HARROW	51.585162	-0.363176	4	Coffee Shop	Indian Restaurant	Gym / Fitness Center	Park	Pub	Café	Pizza Place	Grocery Store	Ice Cream Shop	Italian Restaurant

Visualizing the Resulting Clusters

```
display(map_clusters)
```



The individual clusters data frame is mentioned below:

Cluster 1

	Borough	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
7	Harrow	-0.370300	0	Indian Restaurant	Grocery Store	Pub	Gym / Fitness Center	Fast Food Restaurant	Coffee Shop	Park	Supermarket	Sandwich Place	Café
8	Harrow	-0.352221	0	Indian Restaurant	Pub	Supermarket	Coffee Shop	Grocery Store	Vegetarian / Vegan Restaurant	North Indian Restaurant	Bar	Café	Fast Food Restaurant

Cluster 2

	Borough	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Harrow	-0.337275	1	Coffee Shop	Indian Restaurant	Supermarket	Sandwich Place	Pub	Café	Grocery Store	Bar	Irish Pub	Fast Food Restaurant
1	Harrow	-0.336656	1	Coffee Shop	Pub	Indian Restaurant	Park	Supermarket	Bar	Fast Food Restaurant	Ice Cream Shop	Mediterranean Restaurant	Middle Eastern Restaurant
10	Harrow	-0.329476	1	Coffee Shop	Indian Restaurant	Supermarket	Bar	Sandwich Place	Fast Food Restaurant	Café	Pub	Portuguese Restaurant	Ice Cream Shop
11	Harrow	-0.353069	1	Indian Restaurant	Pub	Park	Coffee Shop	Supermarket	Gym / Fitness Center	Ice Cream Shop	Portuguese Restaurant	Gym	Irish Pub

Cluster 3

	Borough	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Harrow	-0.336656	2	Coffee Shop	Clothing Store	Portuguese Restaurant	Gym	Pizza Place	Women's Store	Middle Eastern Restaurant	Bar	Bookstore	Department Store
3	Harrow	-0.373548	2	Deli / Bodega	Chinese Restaurant	Burger Joint	Grocery Store	Furniture / Home Store	Italian Restaurant	Pizza Place	Pub	Greek Restaurant	Seafood Restaurant
6	Harrow	-0.377014	2	Italian Restaurant	Coffee Shop	Wine Bar	Pizza Place	Café	Bus Stop	Sandwich Place	Pub	Bookstore	Supermarket
8	Harrow	-0.352221	2	Furniture / Home Store	Fast Food Restaurant	Supermarket	Portuguese Restaurant	Indian Restaurant	Coffee Shop	Metro Station	Park	Pizza Place	Bakery
9	Harrow	-0.309511	2	Coffee Shop	Gas Station	Italian Restaurant	Pizza Place	Platform	Portuguese Restaurant	Indian Restaurant	Sandwich Place	Supermarket	Middle Eastern Restaurant

Cluster 4

	Borough	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
11	Harrow	-0.353069	3	Café	Park	Indian Restaurant	Grocery Store	Women's Store	Gas Station	Furniture / Home Store	French Restaurant	Food & Drink Shop	Fast Food Restaurant

Cluster 5

	Borough	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Harrow	-0.340485	4	Indian Restaurant	Park	Thai Restaurant	Supermarket	Grocery Store	Beer Store	Deli / Bodega	Gas Station	Furniture / Home Store	French Restaurant
7	Harrow	-0.370300	4	Indian Restaurant	Grocery Store	Coffee Shop	Pizza Place	Sandwich Place	Pub	Café	Lawyer	Fast Food Restaurant	Betting Shop

4. DISCUSSION AND CONCLUSION

The following are the highlights of Harrow Clusters:

1. Coffee Shops, Indian restaurants & Italian restaurants are popular in the Harrow.
2. As for Indian restaurants, we can see the neighborhoods Rayner's Lane, Harrow Weald and South Harrow which top in the Indians visited venues.
3. Also, to note, Harrow & Wealdstone are having second top Indians visited venues which are close to Rayner's Lanes. But Harrow Weald is comparatively far from said 3 neighborhoods.

Considering all the analysis with the available Data - Below are the neighborhoods in the priority order to open Indian Restaurants:

1. Rayner's Lane
2. Harrow Weald
3. South Harrow

Below are the few other factors in terms of data would have given better results with more insight into the best location.

1. Real Estate
2. Crime data
3. Income per capita
4. Traffic Data
5. More venues exploration with the Foursquare
6. Ratings and feedback of the current restaurants within the clusters

Thanks & Regards,

Unesh Alex, Rajendran Rayappan

yuneshalex.r@gmail.com