Capstone Project- The Battle of Neighbourhoods

Finding an optimal location to open an Asian restaurant in London

Rashmi Bhandarkar

25-04-2021

Contents:

1 Introduction

1.1 Background

As increasing numbers of consumers want to dine out or take prepared food home, the demand for restaurants increased rapidly from 155,000 to nearly 960,000 today in about 40 years. Owning and operating a restaurant business is a dream of many people, but the hard reality is many restaurants fail during their first year, frequently due to lack of planning. There is still room in the market for restaurant business with decent planning. A restaurant's location is as crucial to its success as great food and service. While choosing your restaurant’s location, it is important to identify where your intended customers are located.

1.2 Business Problem

London is the capital and largest city of England and the United Kingdom. Opening a restaurant in a capital city like London can be challenging. One may need to make a huge investment but, before making such investments you want to be certain about the place to enjoy maximum patrons. London has a large population of people from different foreign countries from Asia, Australia, America, Middle east. The 2011 census recorded that 2,998,264 people or 36.7% of London's population are foreign-born making London the city with the second-largest immigrant population, behind New York City. Ethnicity is one of the many factors that play a role in food choices so factors such as the kind of demographics who live there (Racial make-up, ethnic groups) can give investors a good start off. In this project, we aim to find an ideal location for opening Asian ethnic concept restaurant in London through analysis of demographics of London to choose the best borough and explore neighborhoods of that borough.

1.3 Target Audience

This report mainly targeting stakeholders interested in opening an Asian restaurant in London, United Kingdom. Others who are interested in opening ethnic concept restaurant based on the population of ethnic group by borough can refer to this analysis.

2 Description of the data

To solve the problem, data showing demographical representation of London, list of boroughs and neighborhood location and its geographical coordinates will be used in the analysis.

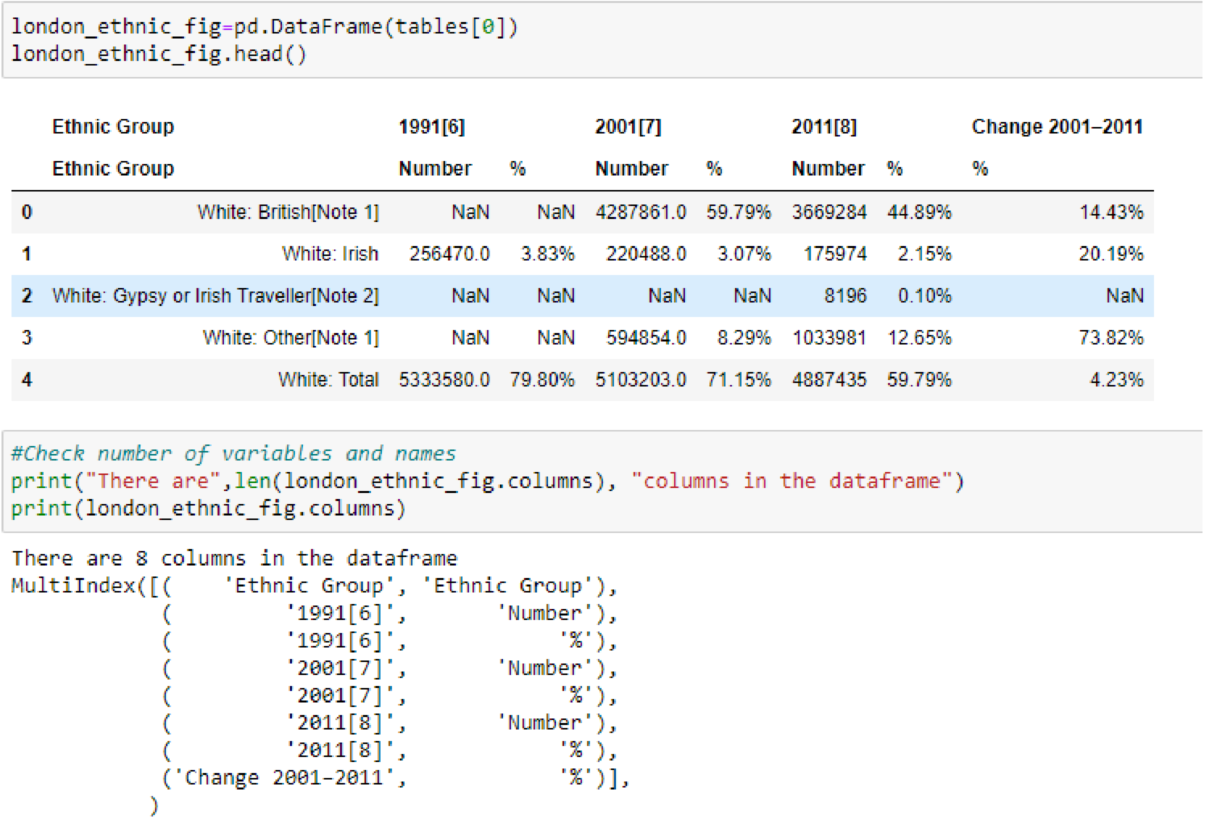
2.1 Data Sources

1. To demonstrate the **Ethnic make-up of London**(2011 Census). The data is scraped from Wikipedia: https://en.wikipedia.org/wiki/Demography_of_London
2. The List of all **boroughs of London** is scraped from the Wikipedia page: https://en.wikipedia.org/wiki/London_boroughs
3. Demography of London giving more details about **Racial make-up of London** boroughs (2011 Census) was obtained by scraping web page: https://en.wikipedia.org/wiki/Demography_of_London
4. **Neighborhoods of Newham** was obtained by web scraping the list available on the page https://en.wikipedia.org/wiki/London_Borough_of_Newham#Districts
5. Geographical co-ordinates of Boroughs of London and Neighborhoods of Newham obtained using Geopy Library (Geocoding Web Services).
6. Foursquare location data (Foursquare API) used to explore neighborhoods of Newham and find the optimal location for opening an Asian restaurant within a defined radius of each neighborhood.

2.2 Exploratory Data Analysis

2.2.1 Ethnic make-up of London(2011 Census)

To demonstrate the ethnic make-up of London, the table showing “ethnic-group of respondents in the 2011 census” is scraped from Wikipedia page using BeautifulSoup library. The scraped table is then transformed into pandas dataframe using “read_html” method. The table scraped from Wikipedia was a multi-index table after reading it into pandas dataframe it became into multi-index column dataframe with imprecise column names (Fig.1).



```
london_ethnic_fig=pd.DataFrame(tables[0])
london_ethnic_fig.head()
```

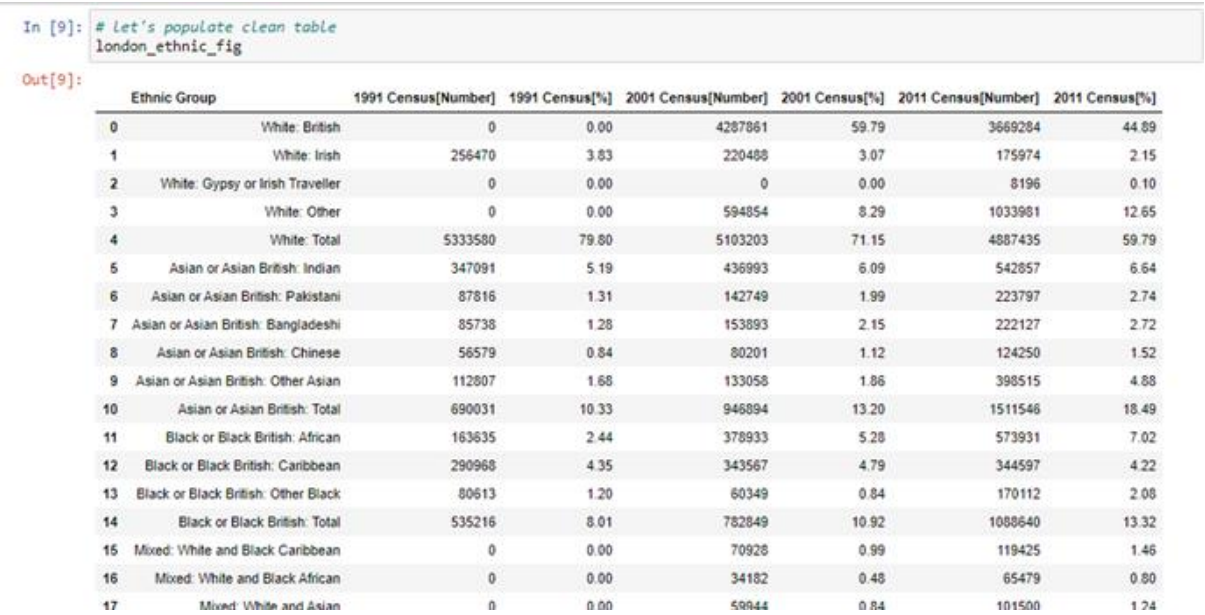
	Ethnic Group	1991[6]		2001[7]		2011[8]		Change 2001–2011
	Ethnic Group	Number	%	Number	%	Number	%	%
0	White: British[Note 1]	NaN	NaN	4287861.0	59.79%	3669284	44.89%	14.43%
1	White: Irish	256470.0	3.83%	220488.0	3.07%	175974	2.15%	20.19%
2	White: Gypsy or Irish Traveller[Note 2]	NaN	NaN	NaN	NaN	8196	0.10%	NaN
3	White: Other[Note 1]	NaN	NaN	594854.0	8.29%	1033981	12.65%	73.82%
4	White: Total	5333580.0	79.80%	5103203.0	71.15%	4887435	59.79%	4.23%

```
#Check number of variables and names
print("There are",len(london_ethnic_fig.columns), "columns in the dataframe")
print(london_ethnic_fig.columns)
```

```
There are 8 columns in the dataframe
MultiIndex([(    'Ethnic Group', 'Ethnic Group'),
            (         '1991[6]',       'Number'),
            (         '1991[6]',            '%'),
            (         '2001[7]',       'Number'),
            (         '2001[7]',            '%'),
            (         '2011[8]',       'Number'),
            (         '2011[8]',            '%'),
            ('Change 2001–2011',            '%')],
           )
```

Figure 1. Multi-index columns before Cleaning

Some data cleaning steps performed on the original dataset to make it simpler to create visualization such as, transformation of multi-index columns into single index columns, removed ‘%’ symbol from some columns using regular expression, strip method is used to remove other unnecessary characters from the “ethnic group” column, replaced “NaN” values directly with 0 which were mostly existing in “1991 Census” column, renamed and removed columns that are not informative to us for visualization. (Fig.2)



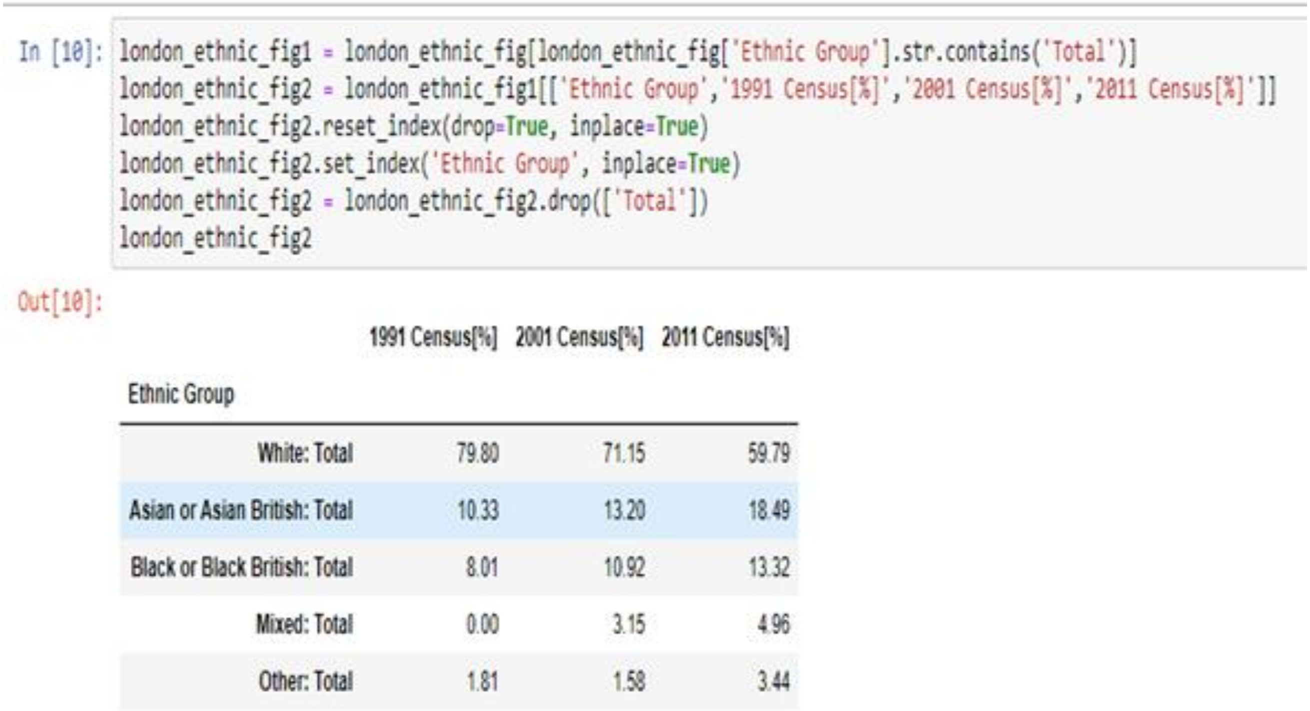
```
In [9]: # Let's populate clean table
        london_ethnic_fig
```

Out[9]:

	Ethnic Group	1991 Census[Number]	1991 Census[%]	2001 Census[Number]	2001 Census[%]	2011 Census[Number]	2011 Census[%]
0	White: British	0	0.00	4287861	59.79	3669284	44.89
1	White: Irish	256470	3.83	220488	3.07	175974	2.15
2	White: Gypsy or Irish Traveller	0	0.00	0	0.00	8196	0.10
3	White: Other	0	0.00	594854	8.29	1033981	12.65
4	White: Total	5333580	79.80	5103203	71.15	4887435	59.79
5	Asian or Asian British: Indian	347091	5.19	436993	6.09	542857	6.64
6	Asian or Asian British: Pakistani	87816	1.31	142749	1.99	223797	2.74
7	Asian or Asian British: Bangladeshi	85738	1.28	153893	2.15	222127	2.72
8	Asian or Asian British: Chinese	56579	0.84	80201	1.12	124250	1.52
9	Asian or Asian British: Other Asian	112807	1.68	133058	1.86	398515	4.88
10	Asian or Asian British: Total	690031	10.33	946894	13.20	1511546	18.49
11	Black or Black British: African	163635	2.44	378933	5.28	573931	7.02
12	Black or Black British: Caribbean	290968	4.35	343567	4.79	344597	4.22
13	Black or Black British: Other Black	80613	1.20	60349	0.84	170112	2.08
14	Black or Black British: Total	535216	8.01	782849	10.92	1088640	13.32
15	Mixed: White and Black Caribbean	0	0.00	70928	0.99	119425	1.46
16	Mixed: White and Black African	0	0.00	34182	0.48	65479	0.80
17	Mixed: White and Asian	0	0.00	59944	0.84	101500	1.24

Figure 2. Single- index columns post some cleaning steps.

The statistics above is providing some information about the number of ethnic groups and percentages (%) of ethnic group as per 1991, 2001 and 2011 Census data.Next, Selecting only columns that are essential for visualization in another dataframe (Fig.3).



```
In [10]: london_ethnic_fig1 = london_ethnic_fig[london_ethnic_fig['Ethnic Group'].str.contains('Total')]
         london_ethnic_fig2 = london_ethnic_fig1[['Ethnic Group','1991 Census[%]','2001 Census[%]','2011 Census[%]']]
         london_ethnic_fig2.reset_index(drop=True, inplace=True)
         london_ethnic_fig2.set_index('Ethnic Group', inplace=True)
         london_ethnic_fig2 = london_ethnic_fig2.drop(['Total'])
         london_ethnic_fig2
```

Out[10]:

	1991 Census[%]	2001 Census[%]	2011 Census[%]
Ethnic Group			
White: Total	79.80	71.15	59.79
Asian or Asian British: Total	10.33	13.20	18.49
Black or Black British: Total	8.01	10.92	13.32
Mixed: Total	0.00	3.15	4.96
Other: Total	1.81	1.58	3.44

Figure 3. Ethnic group by Percentage (%)

Then, plotted "% of ethnic group" for 1991, 2001 and 2011 Census data on the bar graph using Matplotlib visualization library (Fig.4).

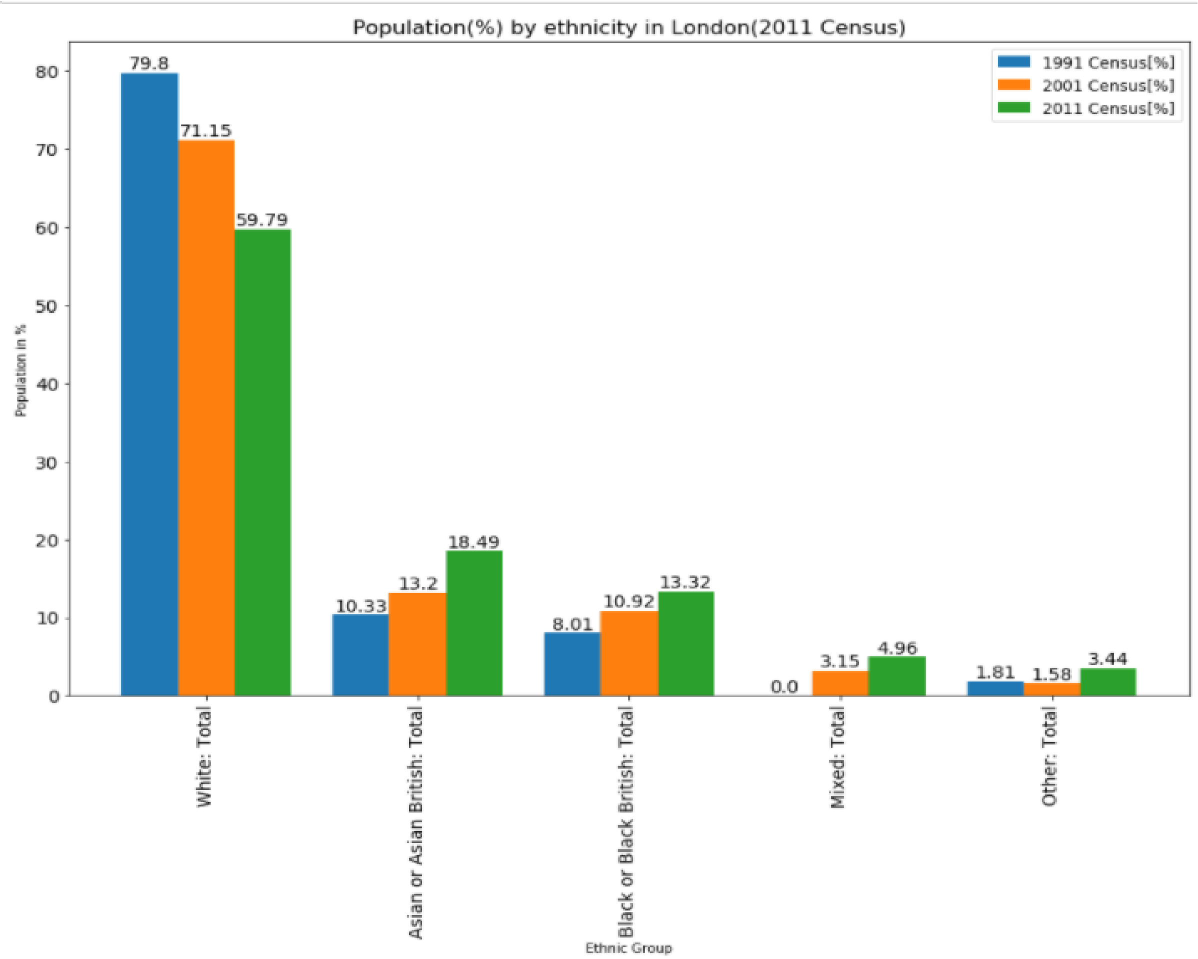
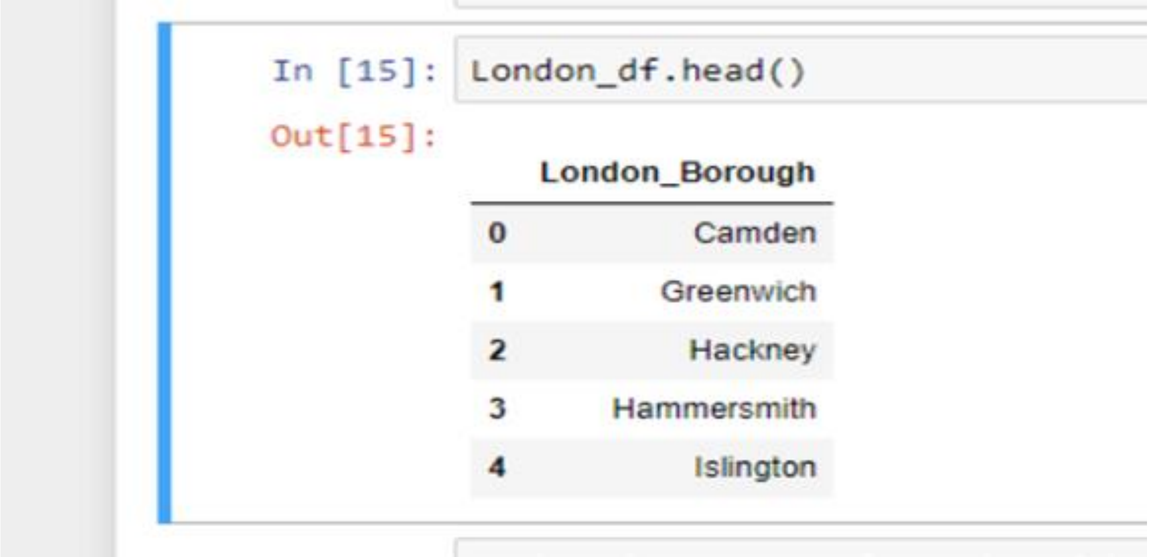


Figure 4. Population by ethnicity in London (percentages)

From the above bar chart, it can be observed the White proportion of the population is highest in London but interestingly there is a sharp fall from 71.15% to 59.79% in 2011 census on the other hand Asian proportion of the population increased from 13.2% to 18.49%. Although there is a huge difference between proportion of the White and Asian population it is evident that 2nd largest non-white proportion of population are Asians. Therefore, opening an Asian restaurant in London would be a great choice considering an increasing number of Asian populations as compared to other ethnic groups. In the further analysis, we will find out boroughs in London with the highest proportion of Asian population. The key factor to make more profit by restaurant owner/investor is to target location where their intended customers reside in majority.

2.2.2 To find a list of all boroughs of London

To find list of all London boroughs, I used BeautifulSoup library and scraped the list from Wikipedia page. Some string manipulation performed using regular expression to extract exact names of the boroughs in the pandas dataframe. (Fig.5)

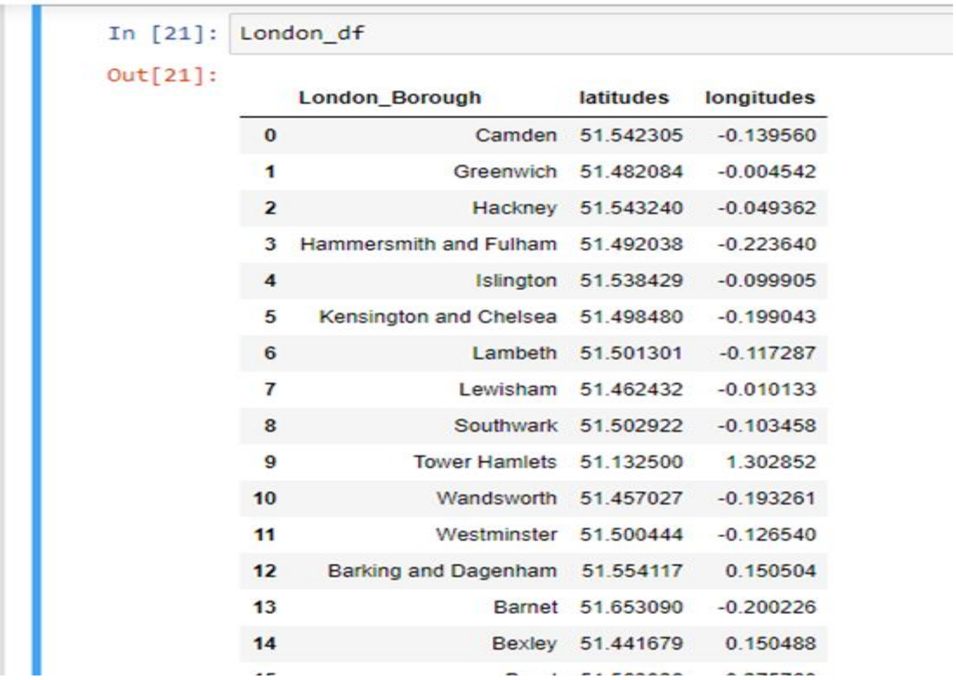


```
In [15]: London_df.head()
Out[15]:
```

	London_Borough
0	Camden
1	Greenwich
2	Hackney
3	Hammersmith
4	Islington

Figure 5. Boroughs of London

I used Geopy library to obtain the geographical coordinates for all boroughs of London. At quick glance on the outcome, I found the coordinates of borough "Tower Hamlet" incorrect so replaced it with correct co-ordinates (Fig.6).



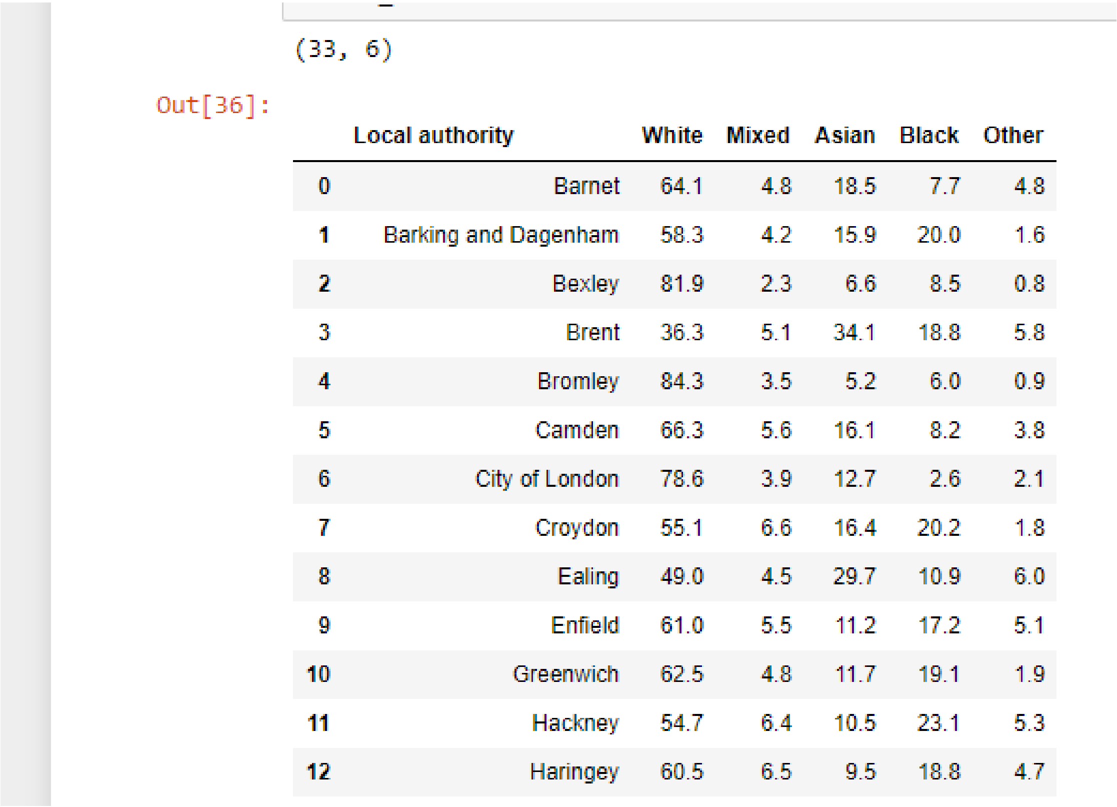
```
In [21]: London_df
Out[21]:
```

	London_Borough	latitudes	longitudes
0	Camden	51.542305	-0.139560
1	Greenwich	51.482084	-0.004542
2	Hackney	51.543240	-0.049362
3	Hammersmith and Fulham	51.492038	-0.223640
4	Islington	51.538429	-0.099905
5	Kensington and Chelsea	51.498480	-0.199043
6	Lambeth	51.501301	-0.117287
7	Lewisham	51.462432	-0.010133
8	Southwark	51.502922	-0.103458
9	Tower Hamlets	51.132500	1.302852
10	Wandsworth	51.457027	-0.193261
11	Westminster	51.500444	-0.126540
12	Barking and Dagenham	51.554117	0.150504
13	Barnet	51.653090	-0.200226
14	Bexley	51.441679	0.150488

Figure 6. Geographical coordinates of boroughs of London

2.2.3 **Racial make-up of London boroughs (2011 Census)**

Now we have list of boroughs of London and their geographical coordinates in the dataframe. The next step is to analyse the racial make-up of London. I scraped the required data table from the Wikipedia page using, BeautifulSoup library. This table shows the proportion of different races by London borough, as found in the 2011 census data. To transform the data into the pandas dataframe “read_html” method is used. Some string manipulation is done to remove whitespaces from the dataframe (Fig.7).



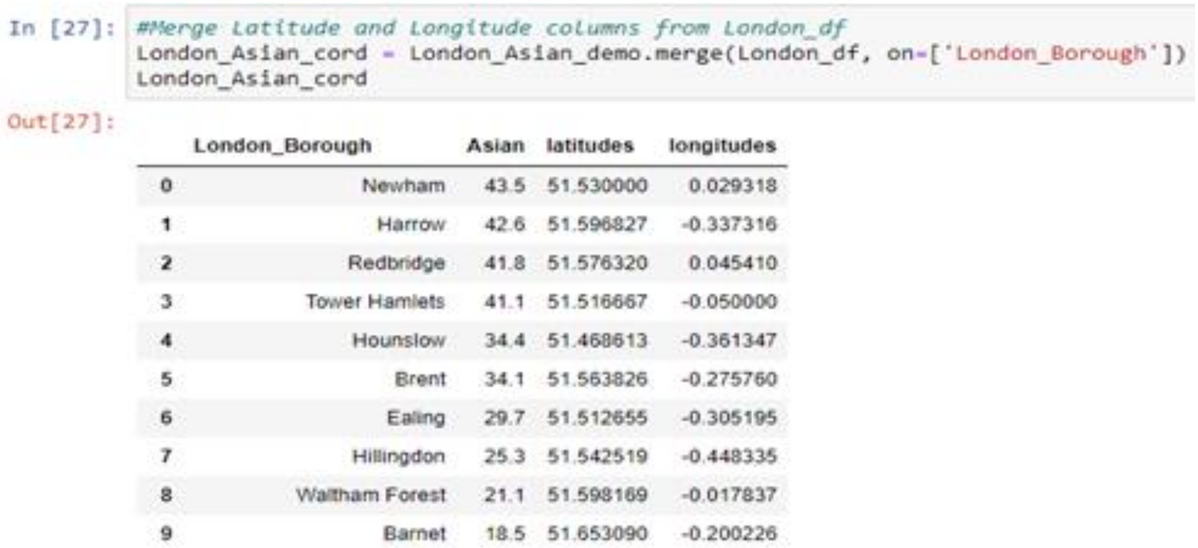
```
(33, 6)
Out[36]:
```

	Local authority	White	Mixed	Asian	Black	Other
0	Barnet	64.1	4.8	18.5	7.7	4.8
1	Barking and Dagenham	58.3	4.2	15.9	20.0	1.6
2	Bexley	81.9	2.3	6.6	8.5	0.8
3	Brent	36.3	5.1	34.1	18.8	5.8
4	Bromley	84.3	3.5	5.2	6.0	0.9
5	Camden	66.3	5.6	16.1	8.2	3.8
6	City of London	78.6	3.9	12.7	2.6	2.1
7	Croydon	55.1	6.6	16.4	20.2	1.8
8	Ealing	49.0	4.5	29.7	10.9	6.0
9	Enfield	61.0	5.5	11.2	17.2	5.1
10	Greenwich	62.5	4.8	11.7	19.1	1.9
11	Hackney	54.7	6.4	10.5	23.1	5.3
12	Haringey	60.5	6.5	9.5	18.8	4.7

Figure 7. Racial make-up of London borough (2011 Census)

The list of boroughs scraped from the Wiki page contains 33 boroughs but in London, there are only 32 boroughs (Inner and Outer) so one additional borough appearing is "City of London" which is part of Greater London so I removed row "City of London" from dataframe.

In the next step, merging of “Racial make-up dataframe” with “Boroughs of London dataframe” is done to visualize the Asian race proportion on the map of London. Below is the output of merged dataframe (Fig.8).



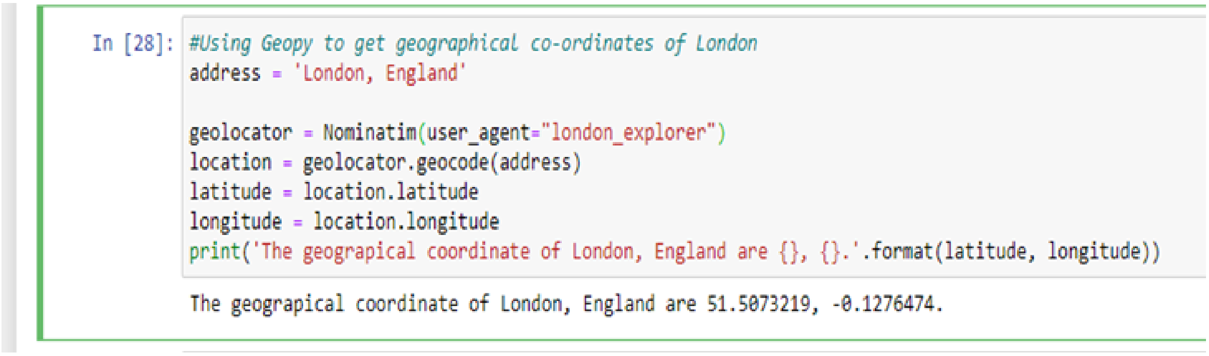
```
In [27]: #Merge Latitude and Longitude columns from London_df
         London_Asian_cord = London_Asian_demo.merge(London_df, on=['London_Borough'])
         London_Asian_cord
```

Out[27]:

	London_Borough	Asian	latitudes	longitudes
0	Newham	43.5	51.530000	0.029318
1	Harrow	42.6	51.596827	-0.337316
2	Redbridge	41.8	51.576320	0.045410
3	Tower Hamlets	41.1	51.516667	-0.050000
4	Hounslow	34.4	51.468613	-0.361347
5	Brent	34.1	51.563826	-0.275760
6	Ealing	29.7	51.512655	-0.305195
7	Hillingdon	25.3	51.542519	-0.448335
8	Waltham Forest	21.1	51.598169	-0.017837
9	Barnet	18.5	51.653090	-0.200226

Figure 8. Asian race make-up proportion in London

Geographical co-ordinates of London are obtained using Geopy library (Fig.9).



```
In [28]: #Using Geopy to get geographical co-ordinates of London
         address = 'London, England'

         geolocator = Nominatim(user_agent="london_explorer")
         location = geolocator.geocode(address)
         latitude = location.latitude
         longitude = location.longitude
         print('The geograpical coordinate of London, England are {}, {}.'.format(latitude, longitude))

         The geograpical coordinate of London, England are 51.5073219, -0.1276474.
```

Figure 9. Geographical coordinates of London, UK

After obtaining geographical coordinates of London, superimposing Asian race proportion obtained from Racial-make up table on the map of London, using geospatial data and folium visualization library (Fig.10).

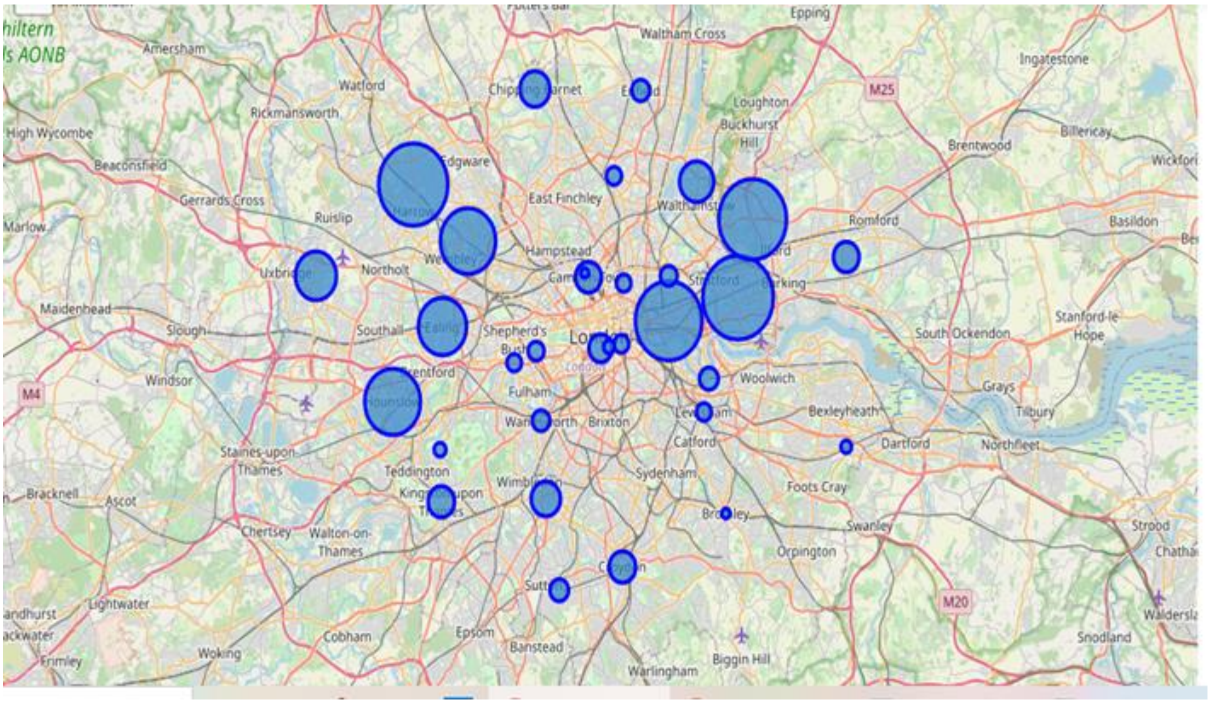


Figure 10. The proportion of Asian race population in London

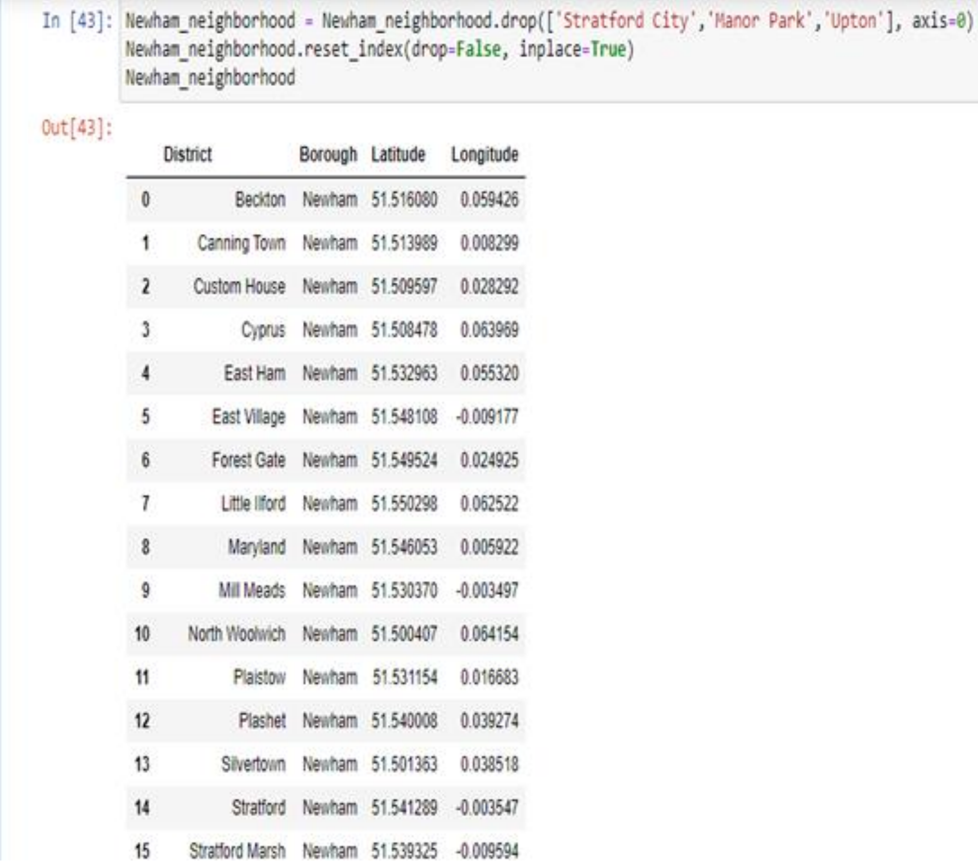
In the above map, the size of each circle indicates the proportion of different races by London borough. It can be noted there are 5-6 boroughs with a fairly good number of Asian race populations in London. It is wonderful news for someone who is looking to open a restaurant in London as they have quite a few options in terms of selecting boroughs.

Among all boroughs of London, Newham has had the largest Asian community for many decades as per 2011 census data presented above. Upon some more research, I found that Newham has the largest total population of Asian origin in London, and it is the 20th most populous borough of all English districts of the United Kingdom. So, through the analysis of the demography of London, I narrowed down my search for the best borough to Newham.

2.2.4 **Neighborhoods of Newham**

We selected best borough for opening Asian restaurant the next step is to find ideal neighborhood/district in the borough of Newham. So, I extracted district of Newham borough by web scraping Wikipedia page which contains this data using BeautifulSoup library.

The geographical coordinates of Newham are obtained from Geopy library. Upon checking dataset, I found the coordinates for “Stratford City” and “Stratford” were similar, so I decided to remove data for “Stratford City”. The coordinates of Manor park and Upton were also incorrect, I replaced them with correct coordinates. It is very important to cross-check coordinates received from Geopy as sometimes due to places with similar names it may provide incorrect coordinates (Fig.11).



```
In [43]: Newham_neighborhood = Newham_neighborhood.drop(['Stratford City','Manor Park','Upton'], axis=0)
         Newham_neighborhood.reset_index(drop=False, inplace=True)
         Newham_neighborhood
```

Out[43]:

	District	Borough	Latitude	Longitude
0	Beckton	Newham	51.516080	0.059426
1	Canning Town	Newham	51.513989	0.008299
2	Custom House	Newham	51.509597	0.028292
3	Cyprus	Newham	51.508478	0.063969
4	East Ham	Newham	51.532963	0.055320
5	East Village	Newham	51.548108	-0.009177
6	Forest Gate	Newham	51.549524	0.024925
7	Little Ilford	Newham	51.550298	0.062522
8	Maryland	Newham	51.546053	0.005922
9	Mill Meads	Newham	51.530370	-0.003497
10	North Woolwich	Newham	51.500407	0.064154
11	Plaistow	Newham	51.531154	0.016683
12	Plashet	Newham	51.540008	0.039274
13	Silvertown	Newham	51.501363	0.038518
14	Stratford	Newham	51.541289	-0.003547
15	Stratford Marsh	Newham	51.539325	-0.009594

Figure 11. Neighborhoods of Newham with geographical coordinates

Neighborhoods/ district of Newham borough are then superimposed on the map of Newham.

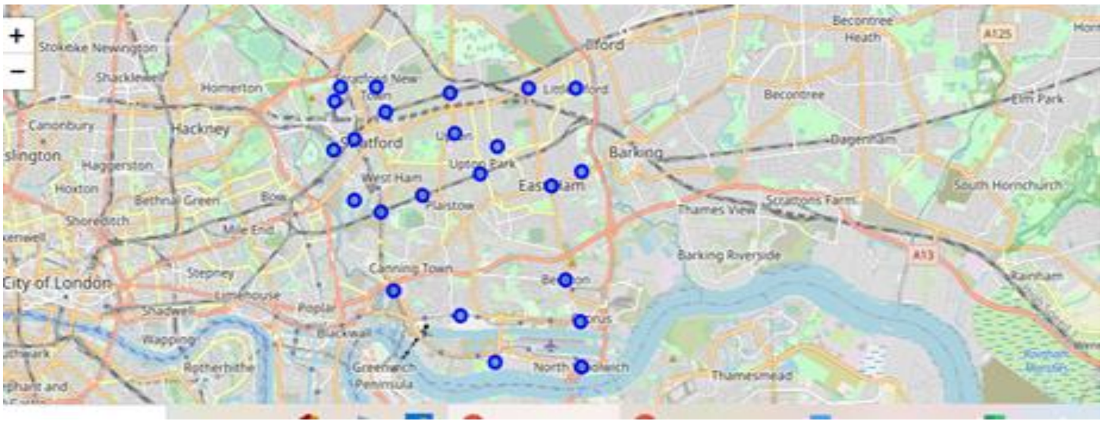
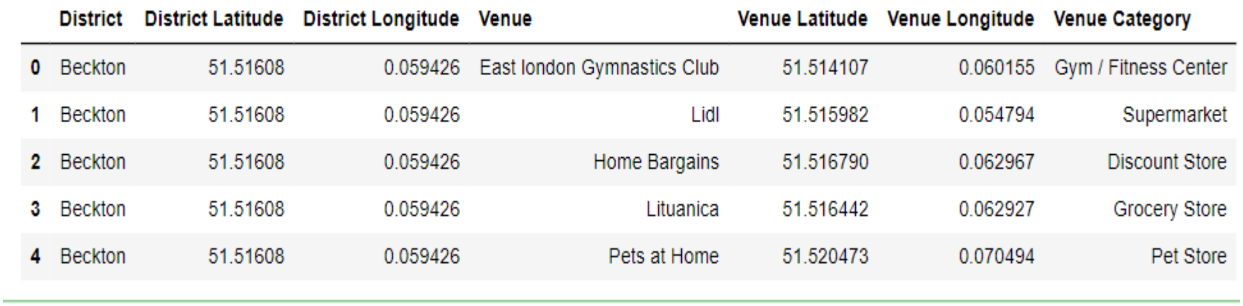


Fig.12. Neighborhood’s map of Newham

3 Methodology

3.1 Foursquare API data analysis

I utilized Foursquare's Places API, to explore neighborhoods/districts in the Newham borough and segment them. The Places API offers real-time access to Foursquare's global database of rich venue data. The function "getNearbyVenues" is created to loop through all the neighborhood/ districts of Newham and created API request URL. To get top venues I set the limit to **100 venues** and radius to **1000 meters** for each neighborhood from their given latitude and longitude. GET request is then made to Foursquare API and only relevant information for each nearby venue is extracted from it. The data is then appended to a python 'list' and lastly python 'list' is flattened to append it to the dataframe being returned by function. The returned dataframe with top 5 row is as follows: [Fig. 13]



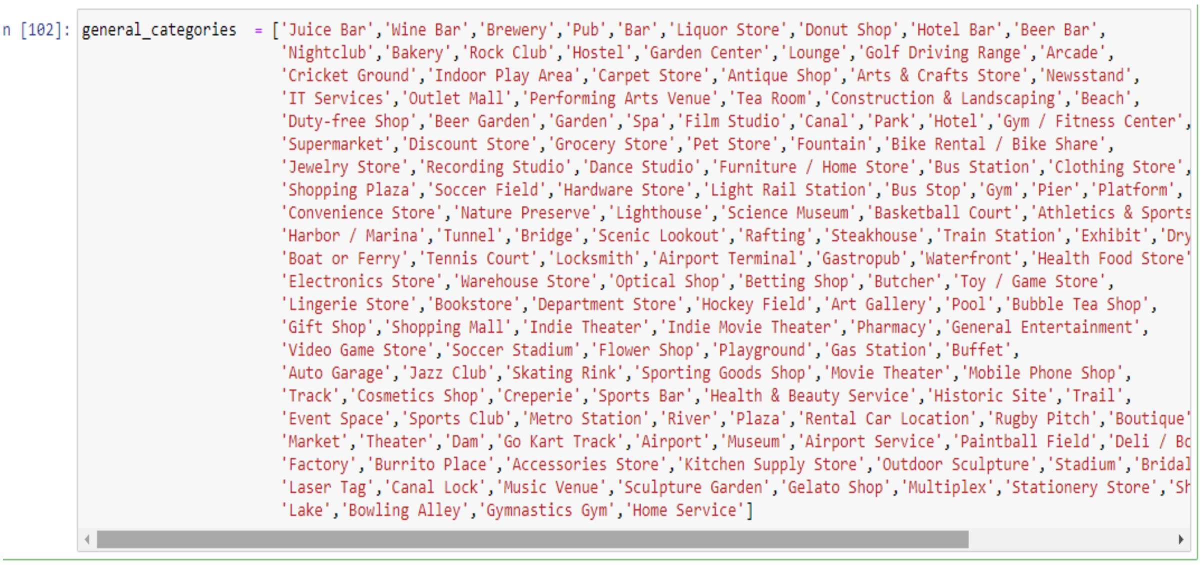
	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Beckton	51.51608	0.059426	East london Gymnastics Club	51.514107	0.060155	Gym / Fitness Center
1	Beckton	51.51608	0.059426	Lidl	51.515982	0.054794	Supermarket
2	Beckton	51.51608	0.059426	Home Bargains	51.516790	0.062967	Discount Store
3	Beckton	51.51608	0.059426	Lituanica	51.516442	0.062927	Grocery Store
4	Beckton	51.51608	0.059426	Pets at Home	51.520473	0.070494	Pet Store

Figure 13. Venues nearby neighborhoods of Newham

The merged dataframe has "District", "District Latitude", "District Longitude", "Venue", "Venue Latitude", "Venue Longitude", "Venue Category" columns and total 1043 nearby venues returned by foursquare. Out of total 1043 venues there are 163 unique "venues categories".

The venue categories returned by foursquare included several general venues categories such as Gym / Fitness Center, Park, Bar, Basketball Court, etc. which are not much useful for the analysis. As our main aim is to segment neighborhoods/ districts based on food venues categories to understand neighborhood's food culture and type of restaurant/ food places already exist in the localities.

So, in the next step, I first created a list and then added all the unique categories (163) that were returned by “getNearbyVenues” function in that list. Then, I manually curated a list with general venue categories which I found insignificant for the analysis as described above.The Decision of choosing general categories depends upon the type of analysis you are performing, and this list can be changed/modified as per your analysis requirement. Following categories considered general in this analysis [Fig. 14].



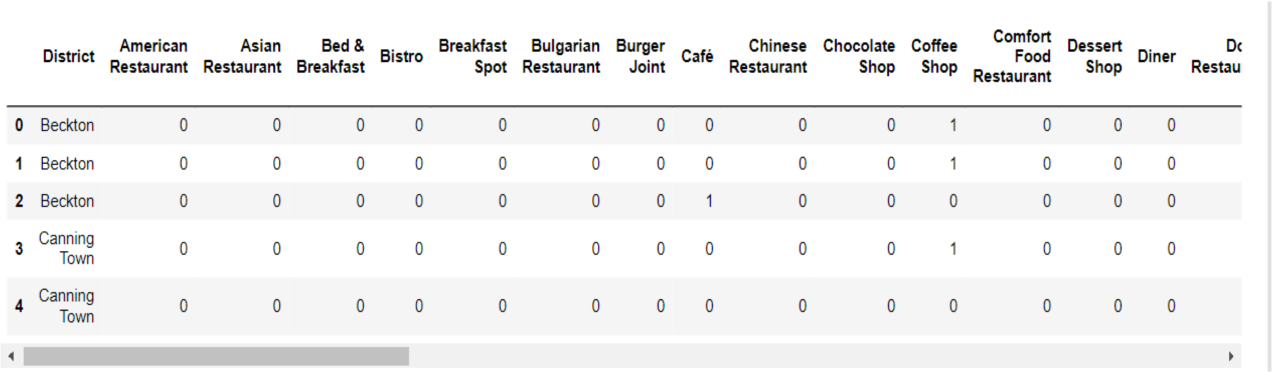
```
n [102]: general_categories  = ['Juice Bar','Wine Bar','Brewery','Pub','Bar','Liquor Store','Donut Shop','Hotel Bar','Beer Bar',
                         'Nightclub','Bakery','Rock Club','Hostel','Garden Center','Lounge','Golf Driving Range','Arcade',
                         'Cricket Ground','Indoor Play Area','Carpet Store','Antique Shop','Arts & Crafts Store','Newsstand',
                         'IT Services','Outlet Mall','Performing Arts Venue','Tea Room','Construction & Landscaping','Beach',
                         'Duty-free Shop','Beer Garden','Garden','Spa','Film Studio','Canal','Park','Hotel','Gym / Fitness Center',
                         'Supermarket','Discount Store','Grocery Store','Pet Store','Fountain','Bike Rental / Bike Share',
                         'Jewelry Store','Recording Studio','Dance Studio','Furniture / Home Store','Bus Station','Clothing Store',
                         'Shopping Plaza','Soccer Field','Hardware Store','Light Rail Station','Bus Stop','Gym','Pier','Platform',
                         'Convenience Store','Nature Preserve','Lighthouse','Science Museum','Basketball Court','Athletics & Sports
                         'Harbor / Marina','Tunnel','Bridge','Scenic Lookout','Rafting','Steakhouse','Train Station','Exhibit','Dry
                         'Boat or Ferry','Tennis Court','Locksmith','Airport Terminal','Gastropub','Waterfront','Health Food Store'
                         'Electronics Store','Warehouse Store','Optical Shop','Betting Shop','Butcher','Toy / Game Store',
                         'Lingerie Store','Bookstore','Department Store','Hockey Field','Art Gallery','Pool','Bubble Tea Shop',
                         'Gift Shop','Shopping Mall','Indie Theater','Indie Movie Theater','Pharmacy','General Entertainment',
                         'Video Game Store','Soccer Stadium','Flower Shop','Playground','Gas Station','Buffet',
                         'Auto Garage','Jazz Club','Skating Rink','Sporting Goods Shop','Movie Theater','Mobile Phone Shop',
                         'Track','Cosmetics Shop','Creperie','Sports Bar','Health & Beauty Service','Historic Site','Trail',
                         'Event Space','Sports Club','Metro Station','River','Plaza','Rental Car Location','Rugby Pitch','Boutique'
                         'Market','Theater','Dam','Go Kart Track','Airport','Museum','Airport Service','Paintball Field','Deli / Bo
                         'Factory','Burrito Place','Accessories Store','Kitchen Supply Store','Outdoor Sculpture','Stadium','Bridal
                         'Laser Tag','Canal Lock','Music Venue','Sculpture Garden','Gelato Shop','Multiplex','Stationery Store','Sh
                         'Lake','Bowling Alley','Gymnastics Gym','Home Service']
```

Figure 14. List of general categories

‘General categories’ were then subtracted from ‘unique categories’ which gives the ‘list of categories’ that are relevant for further analysis. Out of 163 unique categories, we are now left with 46 unique “venue categories” which indicates that we removed more than 65% of irrelevant data.

3.2 Analyse each neighborhood/district

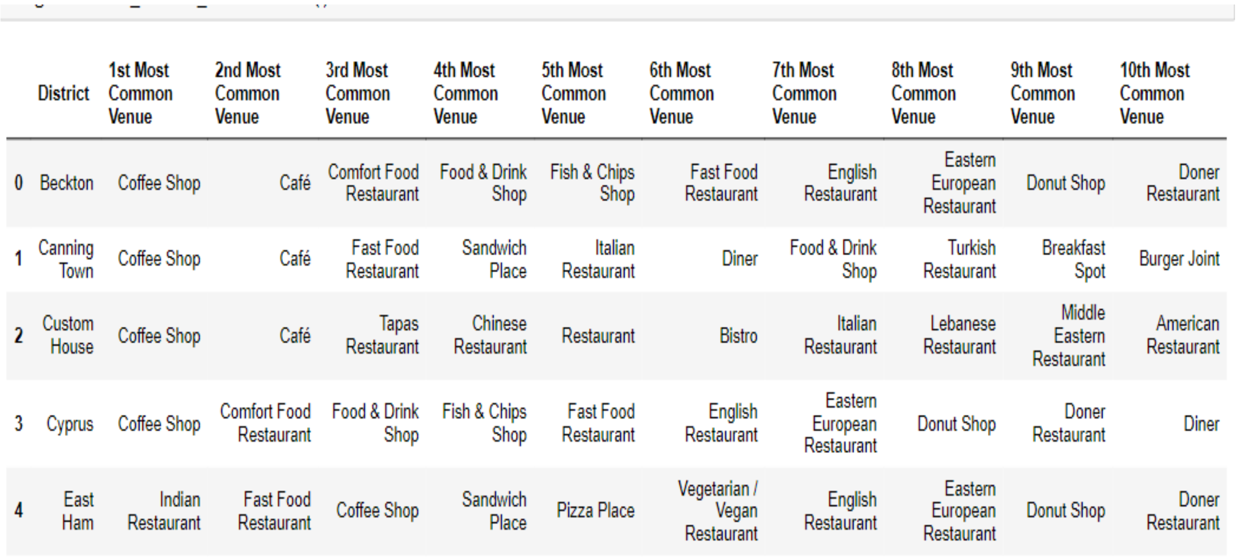
Each neighborhood analysed to understand the most common food venue/places within its 1000 meters of vicinity. “Venue category” is a categorical variable and ML algorithm cannot work directly on categorical data so one hot encoding is performed to convert it into a form that can be provided to Machine learning algorithms. Upon conversion of categorical variable “District” column is then added back and size of the new dataframe is examined [Fig. 15]



	District	American Restaurant	Asian Restaurant	Bed & Breakfast	Bistro	Breakfast Spot	Bulgarian Restaurant	Burger Joint	Café	Chinese Restaurant	Chocolate Shop	Coffee Shop	Comfort Food Restaurant	Dessert Shop	Diner	Dc Restau
0	Beckton	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
1	Beckton	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
2	Beckton	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
3	Canning Town	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
4	Canning Town	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Figure. 15 Dataframe after One- hot encoding

In the next step, grouped rows by “Districts” and by calculating mean of the frequency of occurrence of each category. Based on the mean values,the top 10 venues for each neighborhood can be found out. This top 10 most common venues are then added to the dataframe [Fig 16]



	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	Beckton	Coffee Shop	Café	Comfort Food Restaurant	Food & Drink Shop	Fish & Chips Shop	Fast Food Restaurant	English Restaurant	Eastern European Restaurant	Donut Shop	Doner Restaurant
1	Canning Town	Coffee Shop	Café	Fast Food Restaurant	Sandwich Place	Italian Restaurant	Diner	Food & Drink Shop	Turkish Restaurant	Breakfast Spot	Burger Joint
2	Custom House	Coffee Shop	Café	Tapas Restaurant	Chinese Restaurant	Restaurant	Bistro	Italian Restaurant	Lebanese Restaurant	Middle Eastern Restaurant	American Restaurant
3	Cyprus	Coffee Shop	Comfort Food Restaurant	Food & Drink Shop	Fish & Chips Shop	Fast Food Restaurant	English Restaurant	Eastern European Restaurant	Donut Shop	Doner Restaurant	Diner
4	East Ham	Indian Restaurant	Fast Food Restaurant	Coffee Shop	Sandwich Place	Pizza Place	Vegetarian / Vegan Restaurant	English Restaurant	Eastern European Restaurant	Donut Shop	Doner Restaurant

Fig.16 top 10 Most common venues in each neighborhood

3.3 K-Means Machine learning algorithm

k-means is an unsupervised machine learning algorithm that creates clusters of data points aggregated together because of certain similarities. This algorithm will be used to count neighborhoods for each cluster label for variable cluster size.

To implement this algorithm, it is very important to determine the optimal number of clusters (i.e., k). The Elbow method was then run on the data to find the optimal cluster number of clusters.

Elbow Method:

The Elbow Method calculates the sum of squared distances of samples to their closest cluster centre for different values of ‘k’. The optimal number of clusters is the value after which there is no significant decrease in the sum of squared distances.

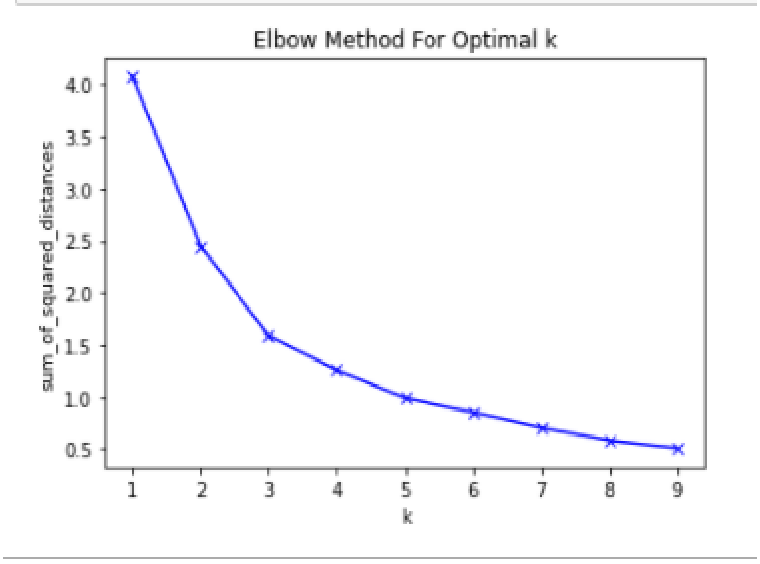
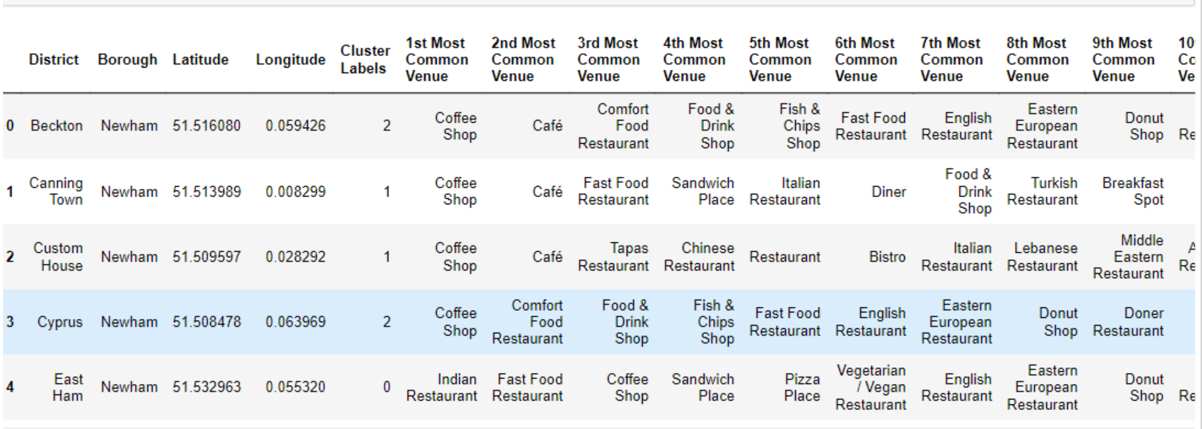


Figure 17. Elbow Method

From the result of K- means it observed that K=4 is the best choice for clustering.

The neighborhoods can be grouped into 4 clusters based on their most common venues. The K- Means algorithm is then applied with K = 4 and clustering labels were added and finally “Newham_neighborhood” dataframe is added to “Newham_merged” dataframe [Fig. 17]



	District	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	10 Cc Ve
0	Beckton	Newham	51.516080	0.059426	2	Coffee Shop	Café	Comfort Food Restaurant	Food & Drink Shop	Fish & Chips Shop	Fast Food Restaurant	English Restaurant	Eastern European Restaurant	Donut Shop	Re
1	Canning Town	Newham	51.513989	0.008299	1	Coffee Shop	Café	Fast Food Restaurant	Sandwich Place	Italian Restaurant	Diner	Food & Drink Shop	Turkish Restaurant	Breakfast Spot	
2	Custom House	Newham	51.509597	0.028292	1	Coffee Shop	Café	Tapas Restaurant	Chinese Restaurant	Restaurant	Bistro	Italian Restaurant	Lebanese Restaurant	Middle Eastern Restaurant	A Re
3	Cyprus	Newham	51.508478	0.063969	2	Coffee Shop	Comfort Food Restaurant	Food & Drink Shop	Fish & Chips Shop	Fast Food Restaurant	English Restaurant	Eastern European Restaurant	Donut Shop	Doner Restaurant	
4	East Ham	Newham	51.532963	0.055320	0	Indian Restaurant	Fast Food Restaurant	Coffee Shop	Sandwich Place	Pizza Place	Vegetarian / Vegan Restaurant	English Restaurant	Eastern European Restaurant	Donut Shop	Re

Figure. 17. Final Merged table with cluster labels for each district

4 Result

Resulted clusters then visualized on the map of Newham using ‘Folium’ library Fig [18].

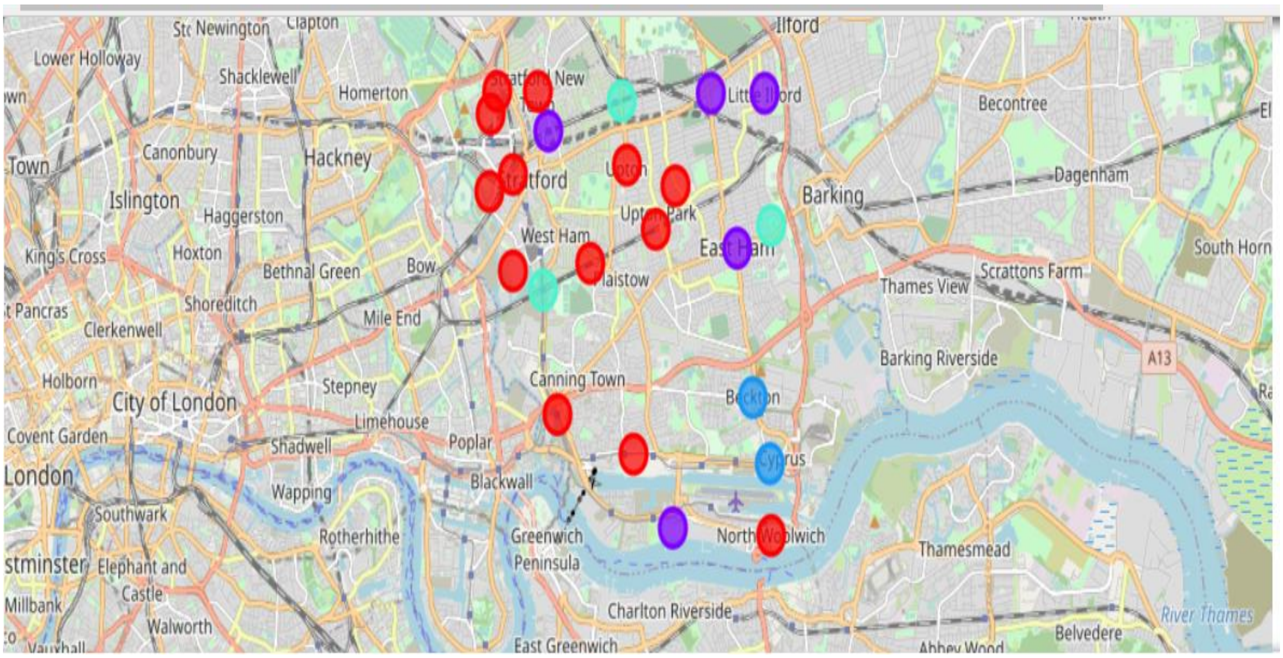
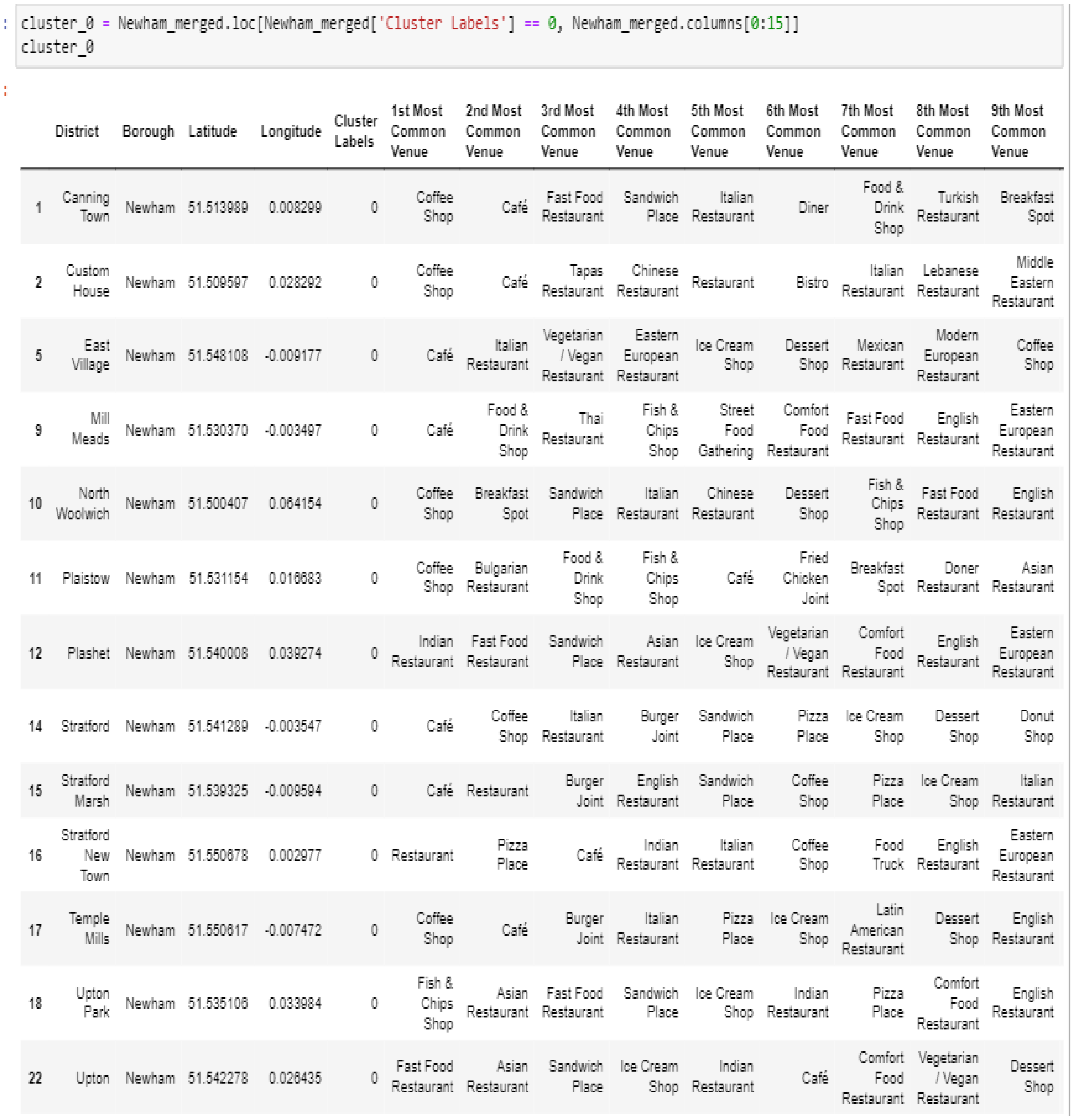


Figure. 18 Neighborhoods of Newham clustering

Examine each cluster

Cluster 0:

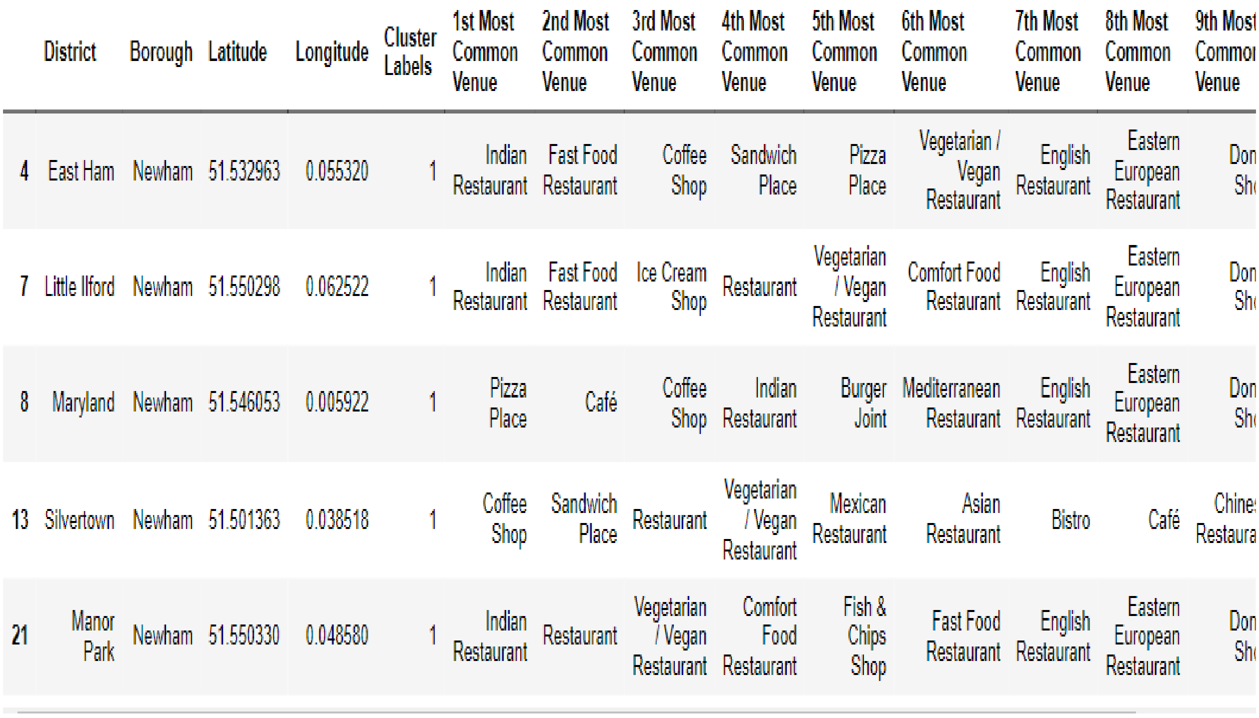


```
cluster_0 = Newham_merged.loc[Newham_merged['Cluster Labels'] == 0, Newham_merged.columns[0:15]]
cluster_0
```

	District	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue
1	Canning Town	Newham	51.513989	0.008299	0	Coffee Shop	Café	Fast Food Restaurant	Sandwich Place	Italian Restaurant	Diner	Food & Drink Shop	Turkish Restaurant	Breakfast Spot
2	Custom House	Newham	51.509597	0.028292	0	Coffee Shop	Café	Tapas Restaurant	Chinese Restaurant	Restaurant	Bistro	Italian Restaurant	Lebanese Restaurant	Middle Eastern Restaurant
5	East Village	Newham	51.548108	-0.009177	0	Café	Italian Restaurant	Vegetarian / Vegan Restaurant	Eastern European Restaurant	Ice Cream Shop	Dessert Shop	Mexican Restaurant	Modern European Restaurant	Coffee Shop
9	Mill Meads	Newham	51.530370	-0.003497	0	Café	Food & Drink Shop	Thai Restaurant	Fish & Chips Shop	Street Food Gathering	Comfort Food Restaurant	Fast Food Restaurant	English Restaurant	Eastern European Restaurant
10	North Woolwich	Newham	51.500407	0.064154	0	Coffee Shop	Breakfast Spot	Sandwich Place	Italian Restaurant	Chinese Restaurant	Dessert Shop	Fish & Chips Shop	Fast Food Restaurant	English Restaurant
11	Plaistow	Newham	51.531154	0.016683	0	Coffee Shop	Bulgarian Restaurant	Food & Drink Shop	Fish & Chips Shop	Café	Fried Chicken Joint	Breakfast Spot	Doner Restaurant	Asian Restaurant
12	Plashet	Newham	51.540008	0.039274	0	Indian Restaurant	Fast Food Restaurant	Sandwich Place	Asian Restaurant	Ice Cream Shop	Vegetarian / Vegan Restaurant	Comfort Food Restaurant	English Restaurant	Eastern European Restaurant
14	Stratford	Newham	51.541289	-0.003547	0	Café	Coffee Shop	Italian Restaurant	Burger Joint	Sandwich Place	Pizza Place	Ice Cream Shop	Dessert Shop	Donut Shop
15	Stratford Marsh	Newham	51.539325	-0.009594	0	Café	Restaurant	Burger Joint	English Restaurant	Sandwich Place	Coffee Shop	Pizza Place	Ice Cream Shop	Italian Restaurant
16	Stratford New Town	Newham	51.550678	0.002977	0	Restaurant	Pizza Place	Café	Indian Restaurant	Italian Restaurant	Coffee Shop	Food Truck	English Restaurant	Eastern European Restaurant
17	Temple Mills	Newham	51.550617	-0.007472	0	Coffee Shop	Café	Burger Joint	Italian Restaurant	Pizza Place	Ice Cream Shop	Latin American Restaurant	Dessert Shop	English Restaurant
18	Upton Park	Newham	51.535106	0.033984	0	Fish & Chips Shop	Asian Restaurant	Fast Food Restaurant	Sandwich Place	Ice Cream Shop	Indian Restaurant	Pizza Place	Comfort Food Restaurant	English Restaurant
22	Upton	Newham	51.542278	0.026435	0	Fast Food Restaurant	Asian Restaurant	Sandwich Place	Ice Cream Shop	Indian Restaurant	Café	Comfort Food Restaurant	Vegetarian / Vegan Restaurant	Dessert Shop

In Cluster 0, most common venues are café/Coffee shops, Sandwich place, Italian restaurant, Asian restaurant, Indian restaurant, fast food restaurant are quite eminent.

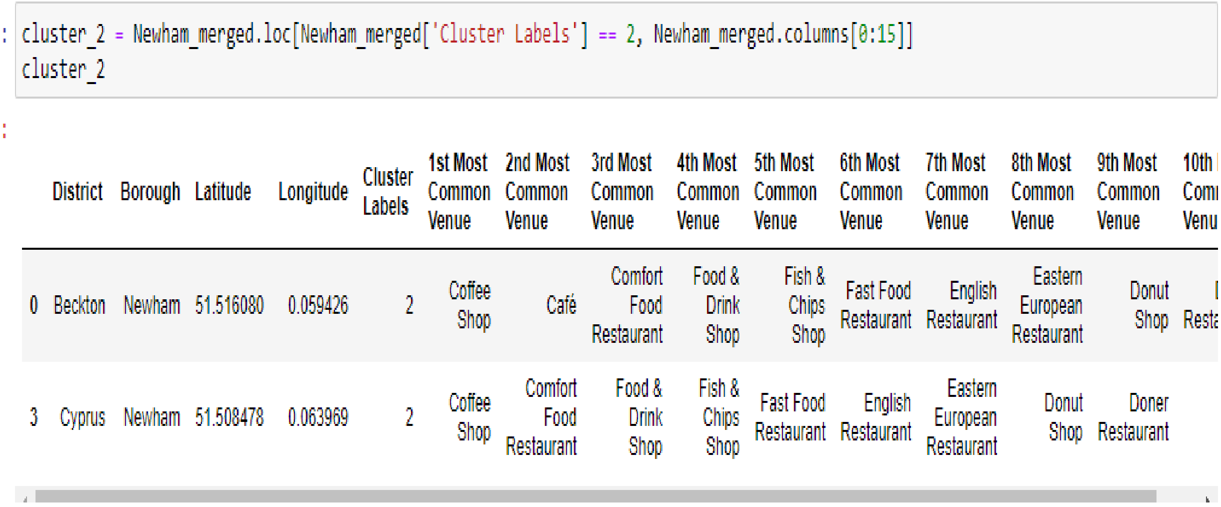
Cluster1:



	District	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 Most Common Venue	8th Most Common Venue	9th Most Commoı Venue
4	East Ham	Newham	51.532963	0.055320	1	Indian Restaurant	Fast Food Restaurant	Coffee Shop	Sandwich Place	Pizza Place	Vegetarian / Vegan Restaurant	English Restaurant	Eastern European Restaurant	Don Sh
7	Little Ilford	Newham	51.550298	0.062522	1	Indian Restaurant	Fast Food Restaurant	Ice Cream Shop	Restaurant	Vegetarian / Vegan Restaurant	Comfort Food Restaurant	English Restaurant	Eastern European Restaurant	Don Sh
8	Maryland	Newham	51.546053	0.005922	1	Pizza Place	Café	Coffee Shop	Indian Restaurant	Burger Joint	Mediterranean Restaurant	English Restaurant	Eastern European Restaurant	Don Sh
13	Silvertown	Newham	51.501363	0.038518	1	Coffee Shop	Sandwich Place	Restaurant	Vegetarian / Vegan Restaurant	Mexican Restaurant	Asian Restaurant	Bistro	Café	Chines Restaura
21	Manor Park	Newham	51.550330	0.048580	1	Indian Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Comfort Food Restaurant	Fish & Chips Shop	Fast Food Restaurant	English Restaurant	Eastern European Restaurant	Don Sh

In cluster 1, Indian restaurants, Fast food restaurants and Vegetarian / Vegan Restaurant are most common.

Cluster 2:



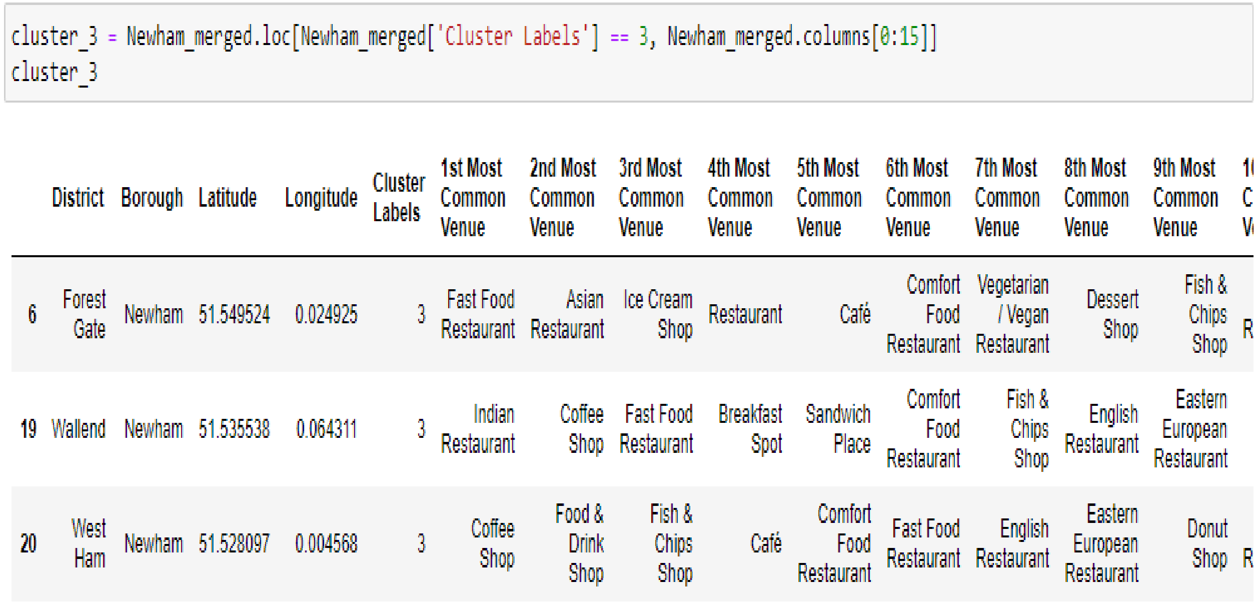
```
: cluster_2 = Newham_merged.loc[Newham_merged['Cluster Labels'] == 2, Newham_merged.columns[0:15]]
  cluster_2
```

:

	District	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Comı Venu
0	Beckton	Newham	51.516080	0.059426	2	Coffee Shop	Café	Comfort Food Restaurant	Food & Drink Shop	Fish & Chips Shop	Fast Food Restaurant	English Restaurant	Eastern European Restaurant	Donut Shop	Resta
3	Cyprus	Newham	51.508478	0.063969	2	Coffee Shop	Comfort Food Restaurant	Food & Drink Shop	Fish & Chips Shop	Fast Food Restaurant	English Restaurant	Eastern European Restaurant	Donut Shop	Doner Restaurant	

In cluster2, Coffee shop and Vegetarian / Vegan Restaurant are most common venues.

Cluster 3:



```
cluster_3 = Newham_merged.loc[Newham_merged['Cluster Labels'] == 3, Newham_merged.columns[0:15]]
cluster_3
```

	District	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	1 C V
6	Forest Gate	Newham	51.549524	0.024925	3	Fast Food Restaurant	Asian Restaurant	Ice Cream Shop	Restaurant	Café	Comfort Food Restaurant	Vegetarian / Vegan Restaurant	Dessert Shop	Fish & Chips Shop	R
19	Wallend	Newham	51.535538	0.064311	3	Indian Restaurant	Coffee Shop	Fast Food Restaurant	Breakfast Spot	Sandwich Place	Comfort Food Restaurant	Fish & Chips Shop	English Restaurant	Eastern European Restaurant	
20	West Ham	Newham	51.528097	0.004568	3	Coffee Shop	Food & Drink Shop	Fish & Chips Shop	Café	Comfort Food Restaurant	Fast Food Restaurant	English Restaurant	Eastern European Restaurant	Donut Shop	R

In cluster3, Indian restaurant, Fast Food Restaurant, restaurant, Vegetarian / Vegan Restaurant and are most common venues.

5 Discussion

Due to the diversity of the Newham in each neighborhood, there is an assortment of most common venues and there are numerous ethnic restaurants as well. Our analysis is focused on finding optimal neighborhood for opening Asian restaurant so to understand the clusters let us find out which neighborhood has the most common venues related to Asian ethnicity. From cluster 0, Custom House, Plashet, Upton Park, Upton, Silvertown are the neighborhoods with the highest number of Asian restaurants. In cluster 1, Indian Restaurant is most common across all the neighborhoods and these are not crowded with other Asian cuisines. Cluster 2 is not famous for Asian cuisine hence opening an Asian restaurant in these neighborhoods will not be profitable. In Cluster 3 Forest Gate and Wallend has Asian restaurant in top 2 most common venue.

6 Conclusion

One application of Clustering Algorithm, k-Means or others, to a multi-dimensional dataset, a very inquisitive result can be curated which helps to understand and visualize the data. The neighborhoods of Newham borough are very briefly segmented into four clusters based on the most common venue hence when looking for a restaurant location, one must consider who else is doing business in the neighborhood. If there are already many restaurants with the same concept of ethnic cooking, then it will not be a profitable deal to choose that location such neighborhoods are mostly appearing in cluster 0. While neighborhoods in cluster 1 are most common for Asian ethnic venue but at the same time, these are less crowded with Asian restaurants. To enjoy maximum patrons in the restaurant, the neighborhoods from cluster 1 are assumed the best choice to open Asian restaurant. The results of this project can be improved and made more inquisitive by considering neighborhoods of other boroughs which have high proportion of Asian population. The scope of this project can be expanded further to choose best borough for opening Asian or other ethnic concept restaurants and suggest a new vendor a profitable location in a diverse city like London.