Capstone Project – The Battle of Neighbourhoods (Part-1)

1. Business Problem/Introduction:

Choosing Best Location to start an Indian Restaurant in London:

For the success of any business, demand plays the most pivotal role and demand relies on the needs and wants of people. They look for expediency, ease and prompt availability; hence location is of utmost importance. Like any other factor, even demography has its own pros and cons- competition being one of them. To succeed, novelty plays a significant role. Creativity, thinking out of box to attract the masses and giving them world class service is the key to success. If we can identify the target audience, define the restaurant type, we'll be well on our way to choosing a restaurant location that sets our business up for success. There's a lot of work, planning and preparation that goes into opening a restaurant.

There are some steps that we need to take in order to achieve our goal of opening a restaurant in London:

- **Target market**: In order to open a restaurant you need to be familiar with exactly who you are targeting to bring into our venue. What ethnicity people are you targeting? Also you need to aware of the age group, location, the amount of money they are willing to spend.

- **Restaurant Style:** The first thing you need to decide is what type of restaurant you want to open and what style it will hold. What type of cuisine will you specialize in? What type of service will be providing to the customers? Will it be self-service or waiter-service? Do you want to open a casual cafe, or a fine-dining restaurant?

- **Competitors:** Along with familiarising with the target customers, you must do so with your competitors. Look at the similar businesses in your area and well-renowned restaurants of the same category. Consider their marketing strategies along with the kind of services they provide. You need to be able to compete with these establishments, so take on board what they're already doing and better it. Also, be aware of similar type of restaurants in the area you are looking to open, because if there is already lot of them, chances are you're entering business in over-saturated area and the potential for you to succeed quickly may be diminished. So, consider opening in an area where there is a gap in the market and demand for your service. However you must also consider the negative like if there are no restaurants in the area similar to your theme, what's the reason? It could be either it will not work or nobody has followed that path yet. So, before choosing the location for your restaurant this is important.

- **Interiors of Restaurant and locality**: Some factors to keep in mind when finding premises for your restaurant:

    - How accessible is the location? How many competitors are there if any?
    - Make sure it is in the area of your target customer.
    - Cost is another factor. Your ideal location may prove to be little costly, so it is likely you will have to compromise on few things.

- **Create your Menu**: Menu is the core of any restaurant. It's important to get it right, as it is the deciding factors for customers voluntarily making a visit to your restaurant again. Your services could be impeccable, your premises may be exquisite, decor is amazing but without a good menu, you shall have no luck in success. Based on what ethnicity you are targeting, your menu should include those cuisines.

2. Data Section:

London is one of the most ethnically diverse cities in the world. At the 2017 census, London had a population of 8.9 million. Of this number, 44.9% are White British, 34% of the population were born outside the UK.

We will focus on the Boroughs of London, for our restaurant problem and we will work on getting the data from all the Boroughs. There are 32 London boroughs with a population around 150,000 to 300,000. To find the best location to start an Indian Restaurant we consider parameters for which we need datasets:

a) Population of the target customers of London based on their:
   - Ethnicity
   - Age
   - Employment Status
   - Gender
   - Income

b) We also need the data for existing licensed restaurants in each Borough, considering our competition factor.

c) At last, we will consider the domestic annual spend estimates and also the tourist levels.

All the above required information is available at London Data store, which is an open data-sharing portal and is free also. In this portal anyone can access the information relating to the city. The data that is available in the Data store is in CSV and XLS format.

The link for the London Data Store - https://data.london.gov.uk/

Along with the above datasets we will be also use FourSquare API to collect information on other venues/competitors in the neighbourhoods of London. You can query the FourSquare website for top locations in London. You can use the FourSquare API to get the supplemental geographical data. We can use FourSquare API to get list of similar restaurants in the location that we selected to get an overview of competition in the chosen area.

How the data will be used to solve the problem:

The data will be used as follows: Using Foursquare and geopy data to map the locations/ neighbourhoods and clustered in groups. Use foursquare and geopy data to map the location with the highest population of Asian communities and then identify the venues near that.

3. Methodology Section:

This section represents the main component of the report where the data is gathered, prepared for analysis. The tools described are used here and the Notebook cells indicate the execution of steps.

The analysis and the strategy

The strategy is to map the locations and find the best one to open a restaurant. The choice is made based on the demands imposed:

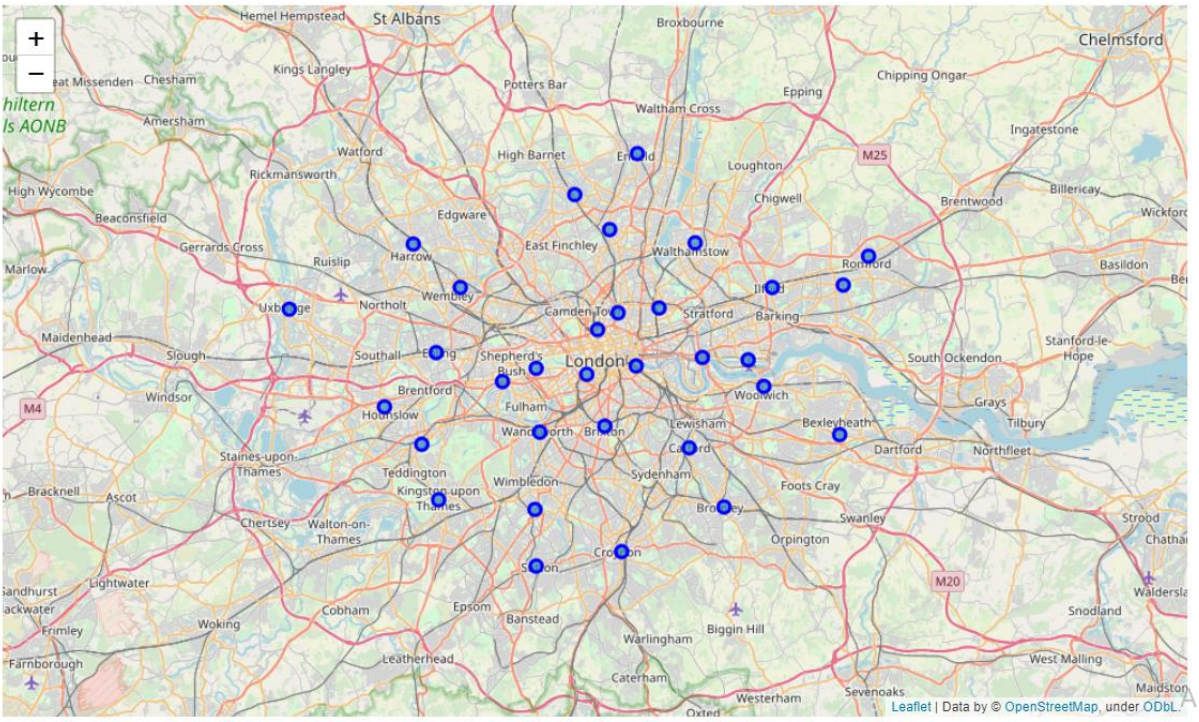
target audience, style of the restaurant, competitors, domestic annual spend estimates, tourist levels and so on. This visual approach and maps with pop ups labels allow quick identification of location, price and other features, thus making the selection very easy.

The processing of these Data and its mapping will allow answering the key questions to make a decision:

- What ethnicity people are we targeting?
- What age group people we are focusing on more?
- What type of cuisine will you specialize in?
- What type of service will be providing to the customers?
- Will it be self-service or waiter-service? Do you want to open a casual cafe, or a fine-dining restaurant?
- However you must also consider the negative like if there are no restaurants in the area similar to your theme, what's the reason?
- How accessible is the location?
- How many competitors are there if any?

We will start by mapping the data and reading the latitude and longitude coordinates of all boroughs in London from a Wikipedia link. Then we get the latitude and longitude of London city using geopy library.

The geographical coordinate of London City are 51.5073219, -0.1276474.



Preferred Location for Indian Restaurant

Based on the target audience we choose the location. For Indian Restaurant, we are focusing on ethnicities like India, Pakistan, China, Bhutan, Nepal and Srilanka. So basically we are focusing on Asian people. As the borough Newham is having high Asian population, we will consider to Segment and Cluster only the neighbourhoods of this borough. For that we have to get the latitude and longitude details of all the neighbourhoods of Newham borough.

Then we created a dataframe from the areas list for the location Newham:

	Borough	Area	Code
0	Newham	Beckton	TQ435815
1	Newham	Canning Town	TQ405815
2	Newham	Custom House	TQ408807
3	Newham	East Ham	TQ425835
4	Newham	Forest Gate	TQ405855
5	Newham	Little Ilford	TQ435855
6	Newham	Manor Park	TQ425855
7	Newham	Maryland	TQ391849
8	Newham	North Woolwich	TQ435795
9	Newham	Plaistow	TQ405825
10	Newham	Silvertown	TQ415795
11	Newham	Stratford	TQ385845
12	Newham	Upton Park	TQ405837
13	Newham	West Ham	TQ405837

Now, we create a dataframe from the list of href links , then we merge the Areas and href links dataframes and dropping the rows where data is not present. Using geo_codes we get the coordinates of all the areas in the Newham borough and then we create a dataframe of geo coordinates:

	Code	Latitude	Longitude
0	TQ435815	51.514206	0.066634
1	TQ405815	51.514959	0.023429
2	TQ408807	51.507696	0.027431
3	TQ425835	51.532430	0.053041
4	TQ405855	51.550902	0.025024
5	TQ435855	51.550148	0.068263
6	TQ425855	51.550401	0.053850
7	TQ391849	51.545857	0.004608
8	TQ435795	51.496234	0.065821
9	TQ405825	51.523945	0.023828
10	TQ415795	51.496738	0.037029
11	TQ385845	51.542410	-0.004196
12	TQ405837	51.534728	0.024306
13	TQ405837	51.534728	0.024306

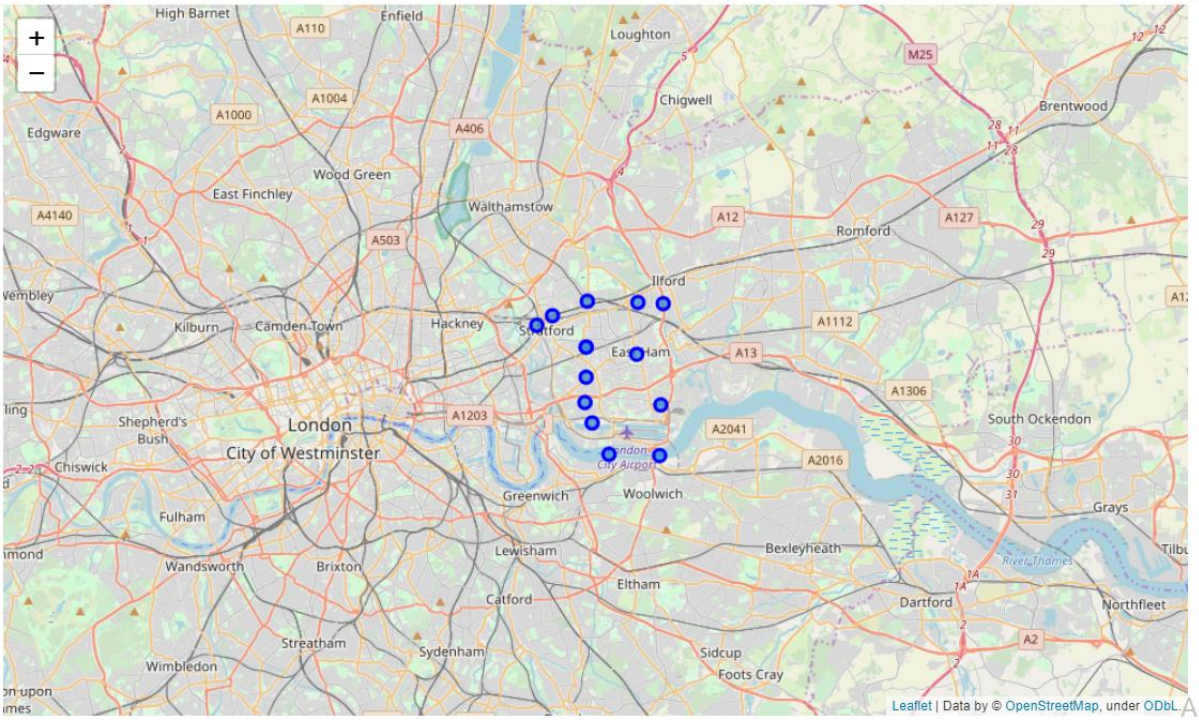
After that we merge the neighbourhoods and Geocode dataframes and then we change the name of the column ‘Area’ to ‘Neighbourhood’

	Borough	Neighborhood	Latitude	Longitude
0	Newham	Beckton	51.514206	0.066634
1	Newham	Canning Town	51.514959	0.023429
2	Newham	Custom House	51.507696	0.027431
3	Newham	East Ham	51.532430	0.053041
4	Newham	Forest Gate	51.550902	0.025024
5	Newham	Little Ilford	51.550148	0.068263
6	Newham	Manor Park	51.550401	0.053850
7	Newham	Maryland	51.545857	0.004608
8	Newham	North Woolwich	51.496234	0.065821
9	Newham	Plaistow	51.523945	0.023828
10	Newham	Silvertown	51.496738	0.037029
11	Newham	Stratford	51.542410	-0.004196
12	Newham	Upton Park	51.534728	0.024306
13	Newham	West Ham	51.534728	0.024306

Now let’s get the coordinates of Newham Borough are:

The geograpical coordinate of Newham are 51.52999955, 0.02931796029382208.

Now let’s visualize the Neighbourhoods of Newham Borough:



From the map we can see that there is cluster of selected neighbourhood from the Newham area. Using FourSquare credentials we start by exploring first neighbourhood in Newham Borough and then get the neighbourhood latitude and longitude value. Now get the top 100 places of Newham neighbourhood within 500 kms. You can check the code for the list of 100 places. Then we extract the categories of the venues, clean the json file and then structure it into a pandas a dataframe . We check for the nearby venues and we see how many venues were returned by Foursquare:

9 venues were returned by Foursquare.

We now explore the neighbourhoods in Newham Borough, by first creating the API request URL, then make the GET request and then returned the relevant information for each nearby venue. We now write the code to sun the above function on each neighbourhood and create a new dataframe called Newham venues and we check how many venues were returned for each neighbourhood:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Beckton	9	9	9	9	9	9
Canning Town	4	4	4	4	4	4
Custom House	30	30	30	30	30	30
East Ham	12	12	12	12	12	12
Forest Gate	10	10	10	10	10	10
Little Ilford	4	4	4	4	4	4
Manor Park	4	4	4	4	4	4
Maryland	24	24	24	24	24	24
North Woolwich	23	23	23	23	23	23
Plaistow	11	11	11	11	11	11
Silvertown	8	8	8	8	8	8
Stratford	72	72	72	72	72	72
Upton Park	6	6	6	6	6	6
West Ham	6	6	6	6	6	6

There are 97 unique categories that are curated from all the returned venues.

Now analyze each neighbourhood in Newham Borough based on the types of venues in those neighbourhoods. Group rows by neighbourhoods and then take mean frequency of occurrence of each category

	Neighborhood	Accessories Store	African Restaurant	Art Gallery	Bagel Shop	Bakery	Bar	Boat or Ferry	Bookstore	Boutique	Brewery	Bridal Shop	Bridge	Bubble Tea Shop	Burger Joint	Bus Line
0	Beckton	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0
1	Canning Town	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0
2	Custom House	0.0	0.000000	0.0	0.033333	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.033333	0.0	0.0	0.0
3	East Ham	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0
4	Forest Gate	0.0	0.000000	0.0	0.000000	0.1	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0
5	Little Ilford	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0
6	Manor Park	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0
7	Maryland	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0
8	North Woolwich	0.0	0.000000	0.0	0.000000	0.0	0.0	0.043478	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0
9	Plaistow	0.0	0.090909	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0

After that we print each neighbourhood along with the top 5 most common venues:

```
----Beckton----
                   venue  freq
0  Furniture / Home Store  0.11
1    Gym / Fitness Center  0.11
2                     Pub  0.11
3          Clothing Store  0.11
4          Shopping Plaza  0.11


----Canning Town----
               venue  freq
0  Convenience Store  0.25
1               Park  0.25
2   Business Service  0.25
3        Gas Station  0.25
4  Accessories Store  0.00
```

The above screenshot is the example of 5 venues from Beckton and Canning town area. We sort the venues in descending order, then create the new dataframes and display the top 10 venues for each neighbourhood.

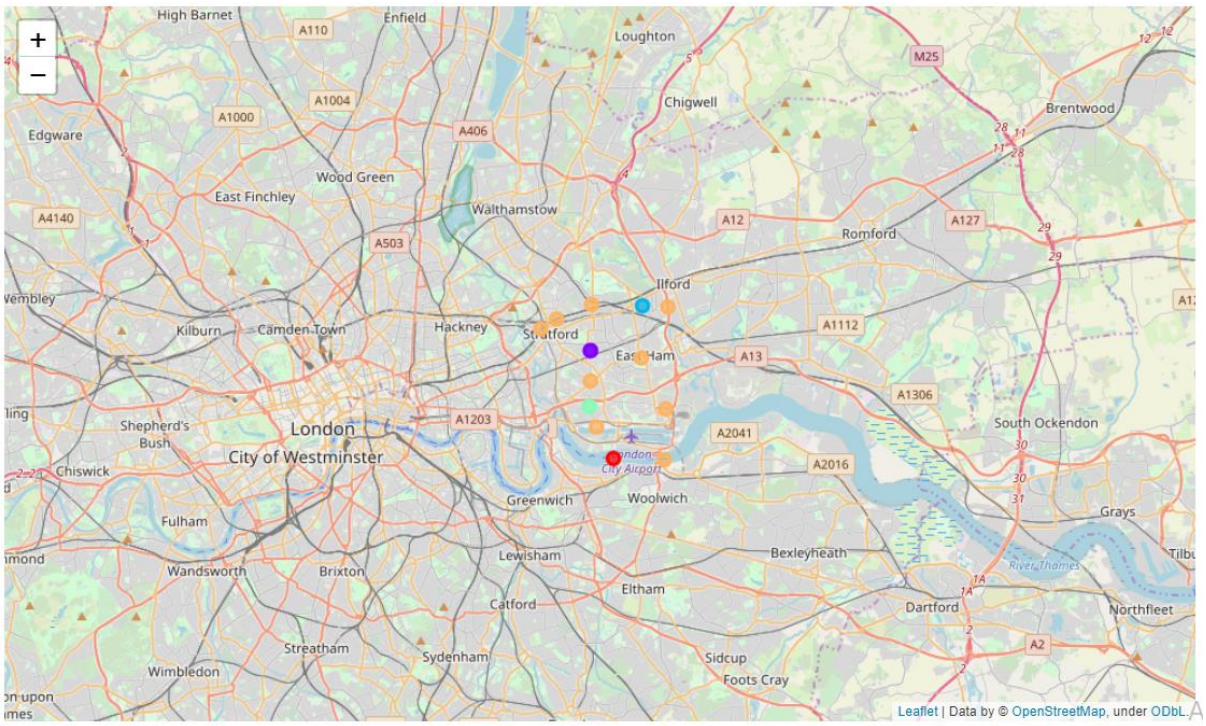
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Beckton	Shopping Plaza	Gym / Fitness Center	Hotel	Grocery Store	Clothing Store	Furniture / Home Store	Pub	Light Rail Station	Discount Store	Electronics Store
1	Canning Town	Convenience Store	Park	Gas Station	Business Service	Wine Bar	Fish & Chips Shop	Dessert Shop	Discount Store	Doner Restaurant	Donut Shop
2	Custom House	Hotel	Coffee Shop	Gym / Fitness Center	Convenience Store	Chinese Restaurant	Pub	Sandwich Place	Scenic Lookout	Gym	Hotel Bar
3	East Ham	Electronics Store	Fast Food Restaurant	Optical Shop	Gym Pool	Coffee Shop	Clothing Store	Pub	Chinese Restaurant	Sandwich Place	Park
4	Forest Gate	Grocery Store	Train Station	Bakery	Pub	Chinese Restaurant	Café	Fish & Chips Shop	Fast Food Restaurant	Wine Bar	Electronics Store
5	Little Ilford	Ice Cream Shop	Grocery Store	Fried Chicken Joint	Indian Restaurant	Gym Pool	Fish & Chips Shop	Department Store	Dessert Shop	Discount Store	Doner Restaurant
6	Manor Park	Indian Restaurant	Gym / Fitness Center	Restaurant	Gas Station	Wine Bar	Fast Food Restaurant	Dance Studio	Department Store	Dessert Shop	Discount Store
7	Maryland	Pub	Hotel	Café	Grocery Store	Supermarket	Bus Stop	Liquor Store	Eastern European Restaurant	Sculpture Garden	Dance Studio
8	North Woolwich	Pier	Hotel	Clothing Store	Warehouse Store	Pharmacy	Pub	Italian Restaurant	River	Sandwich Place	Café
9	Plaistow	Grocery Store	Park	Gym / Fitness Center	Vietnamese Restaurant	African Restaurant	Indian Restaurant	Pub	Café	Bus Stop	Electronics Store

4. Result:

We start with clustering the neighbourhoods. Now using K-means algorithm to cluster the neighbourhoods into 5 clusters. After the clustering we check cluster labels generated from each row in the dataframe .Then we create a dataframe that includes the cluster as well as the top 10 venues for each neighbourhoods.

	Borough	Neighborhood	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
0	Newham	Beckton	51.514206	0.066634	4	Shopping Plaza	Gym / Fitness Center	Hotel	Grocery Store	Clothing Store	Furniture / Home Store	Pub	Light Rail Station
1	Newham	Canning Town	51.514959	0.023429	3	Convenience Store	Park	Gas Station	Business Service	Wine Bar	Fish & Chips Shop	Dessert Shop	Discount Store
2	Newham	Custom House	51.507696	0.027431	4	Hotel	Coffee Shop	Gym / Fitness Center	Convenience Store	Chinese Restaurant	Pub	Sandwich Place	Scenic Lookout
3	Newham	East Ham	51.532430	0.053041	4	Electronics Store	Fast Food Restaurant	Optical Shop	Gym Pool	Coffee Shop	Clothing Store	Pub	Chinese Restaurant
4	Newham	Forest Gate	51.550902	0.025024	4	Grocery Store	Train Station	Bakery	Pub	Chinese Restaurant	Café	Fish & Chips Shop	Fast Food Restaurant
5	Newham	Little Ilford	51.550148	0.068263	4	Ice Cream Shop	Grocery Store	Fried Chicken Joint	Indian Restaurant	Gym Pool	Fish & Chips Shop	Department Store	Dessert Shop
6	Newham	Manor Park	51.550401	0.053850	2	Indian Restaurant	Gym / Fitness Center	Restaurant	Gas Station	Wine Bar	Fast Food Restaurant	Dance Studio	Department Store
7	Newham	Maryland	51.545857	0.004608	4	Pub	Hotel	Café	Grocery Store	Supermarket	Bus Stop	Liquor Store	Eastern European Restaurant
8	Newham	North Woolwich	51.496234	0.065821	4	Pier	Hotel	Clothing Store	Warehouse Store	Pharmacy	Pub	Italian Restaurant	River
9	Newham	Plaistow	51.523945	0.023828	4	Grocery Store	Park	Gym / Fitness Center	Vietnamese Restaurant	African Restaurant	Indian Restaurant	Pub	Café

We add clustering labels then merged the neighbourhood dataframe with Newham borough dataframe to add latitude/longitude for each neighbourhood.
Now let's visualize the Cluster:



After examining the cluster we can recommend that Beckton, Custom House, Maryland, Eastham and Manor Park are the best neighborhoods in Newham borough, to open the Indian restaurant. This is because in these areas, the most common venue visited by the public is the restaurants and as these areas are highly populated with people from Asian countries, opening an Indian Restaurant would definitely be a good idea.

5. Discussion:

I feel this capstone project presented me a good opportunity to practice and apply the Data Science tools and methodologies learned.

In general this course help me learn skills and acquire a good starting point to become a professional Data scientist, also the overall organization, content and lab works presented during the course are positively good.

6. Conclusion:

This project has shown me practical application to resolve real situation that has impacting personal and financial impact using Data Science tools.
I have created a good project that I can present an example to show my potential.

I feel rewarded with the efforts, time and money spent. This covers all the topics basics which is good as a starting point to learn big things.