Capstone Project- The Battle of Neighbourhoods

Opening Asian restaurant in London

Rashmi Bhandarkar

25-04-2021

1 Introduction

1.1 Background

As increasing numbers of consumers wants to dine out or take prepared food home, the operations has skyrocketed from 155,000 about 40 years ago to nearly 960,000 today. Owning and running your own restaurant business is a dream of many people but hard reality is that many restaurants fail during their first year, frequently, due to a lack of planning. There is still a 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.

London is the capital and largest city of England and the United Kingdom. Opening restaurant in the capital city like London can be challenging as you need to make 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 ideal location for opening Asian restaurant in London through analysis of demographics of London to choose best borough and explore neighbourhoods of that borough.

1.2 Target audience

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

2 Description of the data

To solve the problem, I need data which will help us to understand demographical representation of London so below data will be used in the analysis.

2.1 Data Sources

1. Demonstrate the **Ethnic make-up** of London(2011 Census). This can also be obtained from below link:
   https://en.wikipedia.org/wiki/Demography_of_London
2. Find list of all the **boroughs of London**. This data can be obtained from link below
   https://en.wikipedia.org/wiki/London_boroughs
3. Find demography of London which will give more details about **Racial make-up** of London boroughs (2011 Census). This data can be obtained from below link : https://en.wikipedia.org/wiki/Demography_of_London
4. Find all **Neighbourhoods of Newham** which can be obtained through web scraping from link below
   https://en.wikipedia.org/wiki/London_Borough_of_Newham#Districts
5. Geographical co-ordinates of Boroughs of London and Districts of Newham was obtained with the help of Geopy Library (Geocoding Web Services).
6. I will use Foursquare location data (**Foursquare API**) to explore neighbourhoods of Newham and get number of restaurants within defined radius of each neighbourhood.

2.2 Exploratory Data Analysis

Ethnic make-up of London(2011 Census)

To demonstrate 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 was then read into pandas dataframe using “read_html” method. The table scraped from Wikipedia was multi-index table after reading it into pandas dataframe we got multi-index columns with inprecise column names (Fig.1).

```
In [3]: london_ethnic_fig=pd.DataFrame(tables[0])
        london_ethnic_fig.head()
```

Out[3]:

	Ethnic Group	1991[6]		2001[7]		2011[8]		Change 2001–2011
	Ethnic Group	Number	%	Number	%	Number	%	%
0	White: British[Note 1]	NaN	NaN	4287661.0	59.79%	3669284	44.89%	14.43%
1	White: Irish	256470.0	3.83%	220468.0	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%

```
In [4]: 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 needs to be performed to the original dataset to make it easier to create our visualization such as transformation of multi-index columns into single index columns then 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	1033961	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	59844	0.84	101500	1.24

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

Above figures are giving some insights about the “number of ethnic groups” and “percentages (%) of ethnic group” as per 1991, 2001 and 2011 Census data. Selecting only columns that are essential for better 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]:

Ethnic Group	1991 Census[%]	2001 Census[%]	2011 Census[%]
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. Percentage (%) of ethnic group

Plotted “% of ethnic group” for 1991, 2001 and 2011 Census data on the bar graph using Matboltlib visualization library (Fig.4).

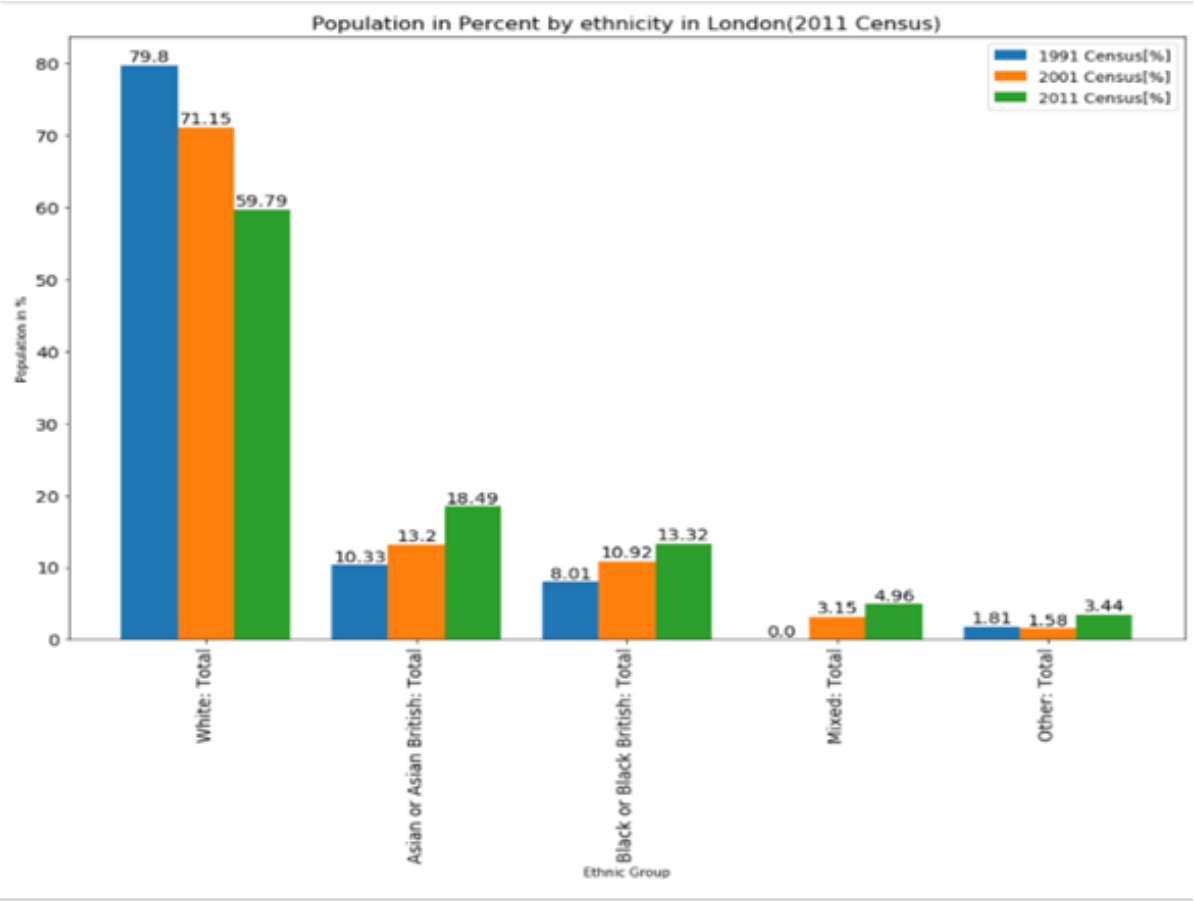

Figure 4. Population in percentages by ethnicity in London

From the above bar charts, it is clear that there is major White of population in London which is obvious but interestingly there is a sharp fall from 71.15% to 59.79% in 2011 census on the other hand Asian proportion of population increased from 13.2% to 18.49% and it is constantly expanding. Although there is huge difference between proportion of White and Asian population it is evident that 2nd largest non white groups are Asians. So, opening Asian restaurant in the London would be a great choice considering increasing number of Asian population compared to other ethnic groups. In the further analysis we will find out boroughs with highest Asian population as it will be a key factor to make more profit to the restaurant owner/investor.

To find list of all boroughs of London

As data is not readily available on the internet, to find the London borough names, I used BeautifulSoup to scrap the Wikipedia page. Some string manipulation is done using regular expression to extract exact names of the boroughs.(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 resulted data frame, the coordinates of borough “Tower Hamlet” looks unusual, so I 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. Coordinates of all boroughs of London

Racial make-up of London boroughs (2011 Census).

To analyse Racial make-up of London Wikipedia page is scraped using, BeautifulSoup library. This table in the Wikipedia page 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 in the dataframe (Fig.7).

Out[25]:

	London_Borough	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	Croydon	55.1	6.6	16.4	20.2	1.8

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

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. Merged dataframe with London borough and Asian race make-up proportion.

To obtain geographical co-ordinates of London I used 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 receiving geographical coordinates of London , I am superimposing Asian race proportion obtained from Racial-make up table on the map of London, using geospatial data and folium visualization library (Fig.10).

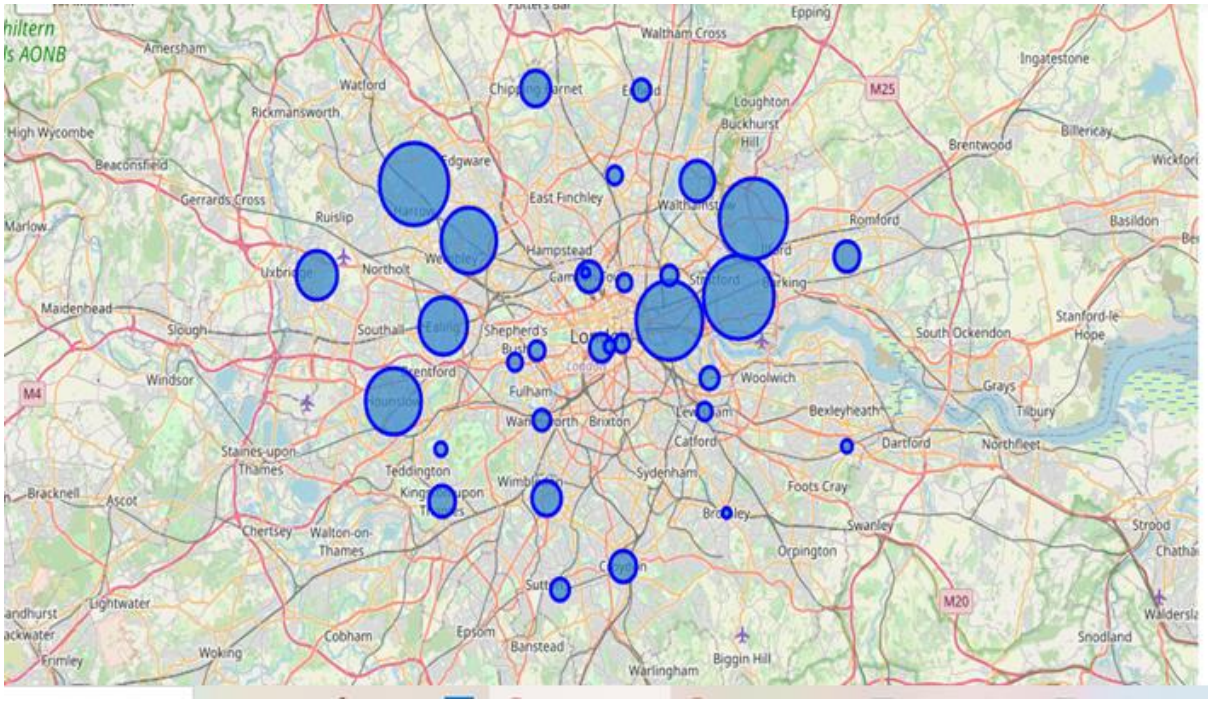

Figure 10. Asian ethnic proportion in London

The above map Indicating proportion of different races by London borough. As it can be observed from the above map there are 5-6 boroughs with quite a good number of Asian race populations in London. It is a wonderful news for someone who is looking to open restaurant in London as they have many options in terms of selecting boroughs.

Newham borough stands at 1st position with largest Asian community from many decades (43.5) among all other boroughs of London. Upon some more research I found that Newham borough is topmost borough with max. Asian race population and at 20th position for total population among all other English districts. So based on the demography of London I narrowed down my search for best borough to Newham.

Neighbourhoods of Newham

From the above analysis, we found best borough for opening Asian restaurant now it is the time to find best neighbourhood in the borough of Newham. So I extracted neighbourhoods of Newham borough by web scraping Wikipedia page which has this information. Again, I used BeautifulSoup library to extract the list of neighbourhoods and build a dataset.

The geographical coordinates of Newham obtained from geopy library. Upon checking dataset I found it gave similar coordinates for “Stratford City” and “Stratford” so I

decided to remove data for “Stratford City” . The coordinates of Manor park and Upton were also incorrect so I replaced it with correct coordinates. It is very important to cross check coordinates received from geopy as sometimes due places with similar names it can give wrong coordinates and it may affect our analysis (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. Neighbourhoods of Newham with geographical coordinates

There are final 23 neighbourhoods in Newham which I will superimpose on the map of Newham.

Fig.12. Superimposed neighbourhoods on the map of Newham

This concludes the Introduction and Data Sources and Description of data section