1.Introduction:

1.1 Background:

London is the capital and the largest city of the England and the UK. It is one of the most ethnically diverse cities in the world. For this reason, It is also seen as a world city. According to the 2011 Census, London has a total population of 8 million (approximately) of which 20% belong to Asian ethnic group which is 1.5 million approximately. Even though the Asian community is massive, there is a lack of an high-end Asian Restaurant with multiple cuisines that not only provides food but also provides service with the ambience. The restaurant industry in London is growing exponentially with the increasing demand. This demand has spurred the competition to open restaurants in a nice area of the city.

Data Science helps in identifying the appropriate market trends and evolving consumer preferences so that restaurants can better address them. Using various data analysis techniques, the London areas are explored through segmenting and clustering, to identify a good location to open an Asian restaurant.

1.2 Business Problem:

A successful Asian restaurant chain is looking to expand its operations through London. We were asked to identify and recommend the neighborhoods in London that will be good choice to start an Asian restaurant.

1.3 Target Audience:

This project will primarily help the following categories:

1. Companies that are looking to invest in food service industry of London.
2. Individuals looking to relocate neighborhoods in London with particular venues.

1.4 Data:

Data sources:

For this project, we will make use of the following data.

1.London Neighborhood's: I have used web scraping techniques to get the list of areas and boroughs in the London. I've extracted the Location, borough, post town and post codes of the areas in London.

 Data source: https://en.wikipedia.org/wiki/List_of_areas_of_London

2. London Demographics: From the following Wikipedia page, I have extracted the demographics of each Borough in London through web scraping.

Data source: https://en.wikipedia.org/wiki/Demography_of_London

3. Geopy library: To get the latitude and longitude of each neighborhood.

4. Foursquare API: I've used the foursquare API to locate various venues in each of the London neighborhoods.

Data Collection and Cleaning:

The BeautifulSoup package is used to scrape the needed data from Wikipedia. The following data frame was obtained by scraping the list of London areas from Wikipedia page:

```
Out[67]:
```

	Neighborhood	Borough	Post_town	Post_code
0	Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2
1	Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4
2	Addington	Croydon[8]	CROYDON	CR0
3	Addiscombe	Croydon[8]	CROYDON	CR0
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14
...	...	...	...	...
528	Woolwich	Greenwich	LONDON	SE18
529	Worcester Park	Sutton, Kingston upon Thames	WORCESTER PARK	KT4
530	Wormwood Scrubs	Hammersmith and Fulham	LONDON	W12
531	Yeading	Hillingdon	HAYES	UB4
532	Yiewsley	Hillingdon	WEST DRAYTON	UB7

533 rows × 4 columns

The data frame needs to be cleaned. The borough column has numbers attached to it's values that should be stripped. After stripping the numbers, the following data frame is obtained:

```
In [68]: df_1['Borough'] = df_1['Borough'].map(lambda x: x.rstrip(']').rstrip('0123456789').rstrip('['))
         df_1
```

```
Out[68]:
```

	Neighborhood	Borough	Post_town	Post_code
0	Abbey Wood	Bexley, Greenwich	LONDON	SE2
1	Acton	Ealing, Hammersmith and Fulham	LONDON	W3, W4
2	Addington	Croydon	CROYDON	CR0
3	Addiscombe	Croydon	CROYDON	CR0
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14
...	...	...	...	...
528	Woolwich	Greenwich	LONDON	SE18
529	Worcester Park	Sutton, Kingston upon Thames	WORCESTER PARK	KT4
530	Wormwood Scrubs	Hammersmith and Fulham	LONDON	W12
531	Yeading	Hillingdon	HAYES	UB4
532	Yiewsley	Hillingdon	WEST DRAYTON	UB7

533 rows × 4 columns

The demographics of the all the London boroughs is obtained from the 'Demography of London' Wikipedia page. After cleaning and parsing the html accordingly, the resulting data frame is below:

```
In [74]: df_demographics = pd.DataFrame({'Local_authority':local_authority,'White': white,'Mixed': mixed,'Asian':asian,'Blac
         df_demographics
```
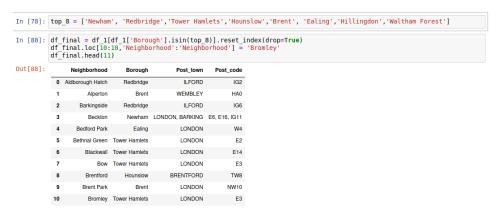
Out[74]:

	Local_authority	White	Mixed	Asian	Black	Others
0	Barnet	64.1	4.8	18.5	7.7	4.8
1	Barking and Dagenham	58.3	4.2	15.9	20	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	0.9
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	4.5	29.7	10.9	6
9	Enfield	61	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
13	Harrow	42.2	4	42.6	8.2	2.9

Since, the business focusses on Asian market the data frame is sorted in descending order of the column 'Asian' that represents the percentage of Asian population. Sorted data frame can be observed below:

```
In [13]: df_demographics.sort_values(by='Asian', inplace=True, ascending=False)
         df_demographics.reset_index(inplace=True)
         df_demographics.drop('index', axis=1, inplace=True)
         df_demographics.head()
```

Out[13]:

	Local_authority	White	Mixed	Asian	Black	Others
0	Newham	29.0	4.5	43.5	19.6	3.5
1	Harrow	42.2	4.0	42.6	8.2	2.9
2	Redbridge	42.5	4.1	41.8	8.9	2.7
3	Tower Hamlets	45.2	4.1	41.1	7.3	2.3
4	Hounslow	51.4	4.1	34.4	6.6	3.6

The top eight boroughs with highest Asian population are observed. We limit the London neighborhoods that we initially obtained to these eight boroughs. The part of the data frame can be observed below:

```
In [78]: top_8 = ['Newham', 'Redbridge','Tower Hamlets','Hounslow','Brent', 'Ealing','Hillingdon','Waltham Forest']

In [88]: df_final = df_1[df_1['Borough'].isin(top_8)].reset_index(drop=True)
         df_final.loc[10:10,'Neighborhood':'Neighborhood'] = 'Bromley'
         df_final.head(11)
```

Out[88]:

	Neighborhood	Borough	Post_town	Post_code
0	Aldborough Hatch	Redbridge	ILFORD	IG2
1	Alperton	Brent	WEMBLEY	HA0
2	Barkingside	Redbridge	ILFORD	IG6
3	Beckton	Newham	LONDON, BARKING	E6, E16, IG11
4	Bedford Park	Ealing	LONDON	W4
5	Bethnal Green	Tower Hamlets	LONDON	E2
6	Blackwall	Tower Hamlets	LONDON	E14
7	Bow	Tower Hamlets	LONDON	E3
8	Brentford	Hounslow	BRENTFORD	TW8
9	Brent Park	Brent	LONDON	NW10
10	Bromley	Tower Hamlets	LONDON	E3

The Geopy library is used to get the latitude and longitude of each neighborhood and are added as columns to the data frame. The head of the final data frame is as follows:

3

Out[95]:

	Neighborhood	Borough	Post_town	Post_code	Latitude	Longitude
0	Aldborough Hatch	Redbridge	ILFORD	IG2	51.585590	0.098750
1	Alperton	Brent	WEMBLEY	HA0	51.540804	-0.300096
2	Barkingside	Redbridge	ILFORD	IG6	51.585818	0.088624
3	Beckton	Newham	LONDON, BARKING	E6, E16, IG11	51.516080	0.059426
4	Bedford Park	Ealing	LONDON	W4	51.498020	-0.255647

4