# **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.

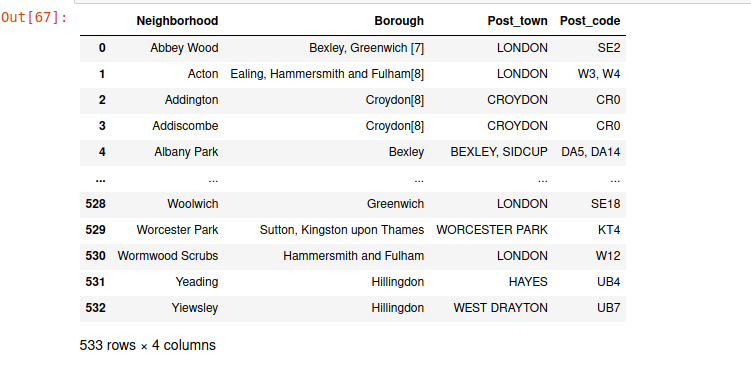
Data source: <https://en.wikipedia.org/wiki/Demography_of_London>

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

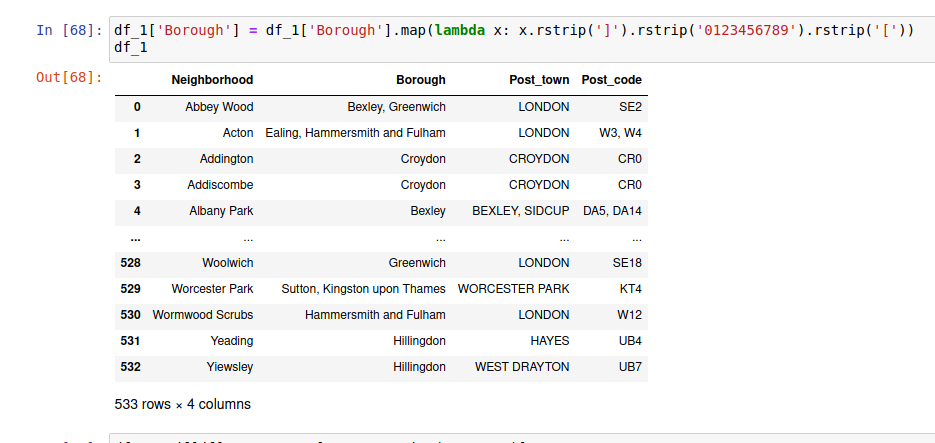
4. Foursquare API: I’ve used the foursquare API to locate various venues in each of the London neighborhoods.

**Data Collection and Cleaning:**

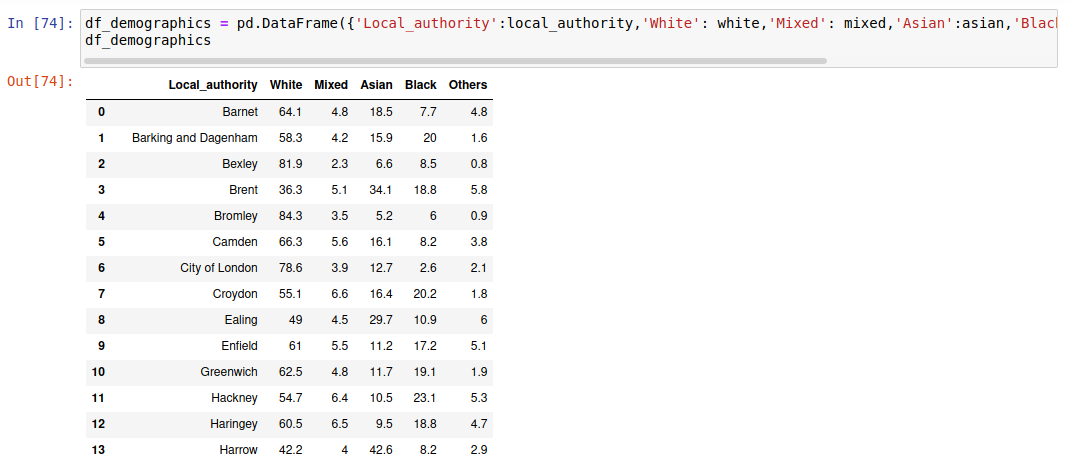
The BeautifulSoup package is used to scrape the needed data from Wikipedia. The following data frame was obtained by scraping the list of London areas from Wikipedia page:



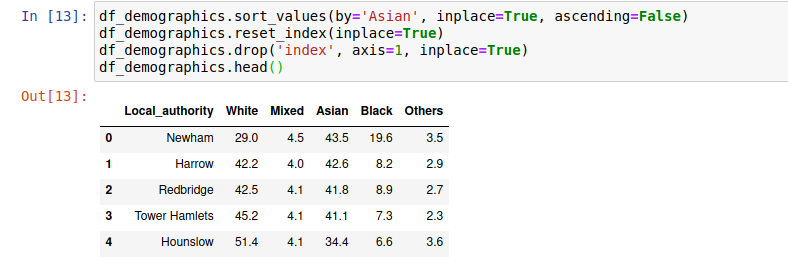
The data frame needs to be cleaned. The borough column has numbers attached to it’s values that should be stripped. After stripping the numbers, the following data frame is obtained:



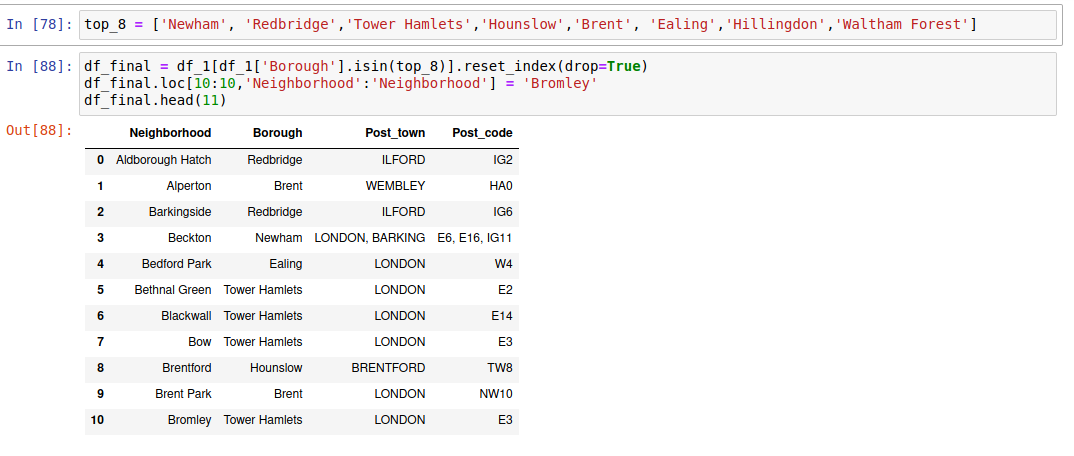
The demographics of the all the London boroughs is obtained from the ‘[Demography of London’](https://en.wikipedia.org/wiki/Demography_of_London) Wikipedia page. After cleaning and parsing the html accordingly, the resulting data frame is below:



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:



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:



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:

