**IDENTIFYING BUSINESS OPPORTUNITIES IN THE CITY OF LONDON**

1. Introduction

Since I have now been equipped with the skills and the tools to use location data to explore a geographical location, I have come up with an idea to leverage the Foursquare location data to explore neighbourhoods in a specific city of my choice by coming up with a problem that you I will use the Foursquare location data to solve

* 1. Description and Discussion of the Business Problem

An entrepreneur has recently relocated to London from New York. He is looking to start a new venture but he is facing problems since

* He is new to the city
* He wants to learn about promising businesses and high demands.

As a business consultant, I have been assigned the duty of identifying promising business opportunity for this entrepreneur in the city of London.

The approach I have undertaken is to dive deep into the various boroughs and neighbourhoods in London and identify

* Various neighbourhoods
* Profile of people and their preferences
* Current existing businesses and their popularity
* Promising business opportunity and location
* After an in-depth analysis of the various boroughs in London ,

I have identified Camden as a focus borough for a variety of reasons.

* Close proximity of this borough to the entrepreneur's location ,
* Camden is well known for its markets.
* The markets are a major tourist attraction at weekends and is popular with young people.

The approach is to drill down further into the various neighbourhoods of Camden and segregate neighbourhoods by venues, their popularity.

After performing a cluster analysis and evaluation has helped narrow down my recommendation to a few neighbourhoods and a promising business opportunity for this entrepreneur.

* 1. Target Reader

The target reader for this report is any individual/ businesses who are interesting in learning about the boroughs in London and one is particular which is Camden. It explores the various venues and places of interest so the data can be useful to tourists or younger population that may be keen to learn more about this location and its hotspots.

1. Data – Source & Methodology

The data sources that I am using as a basis for my analysis are as below:

1. A list of London boroughs from the Wikipedia page. This data has been scraped and the link for the same is here

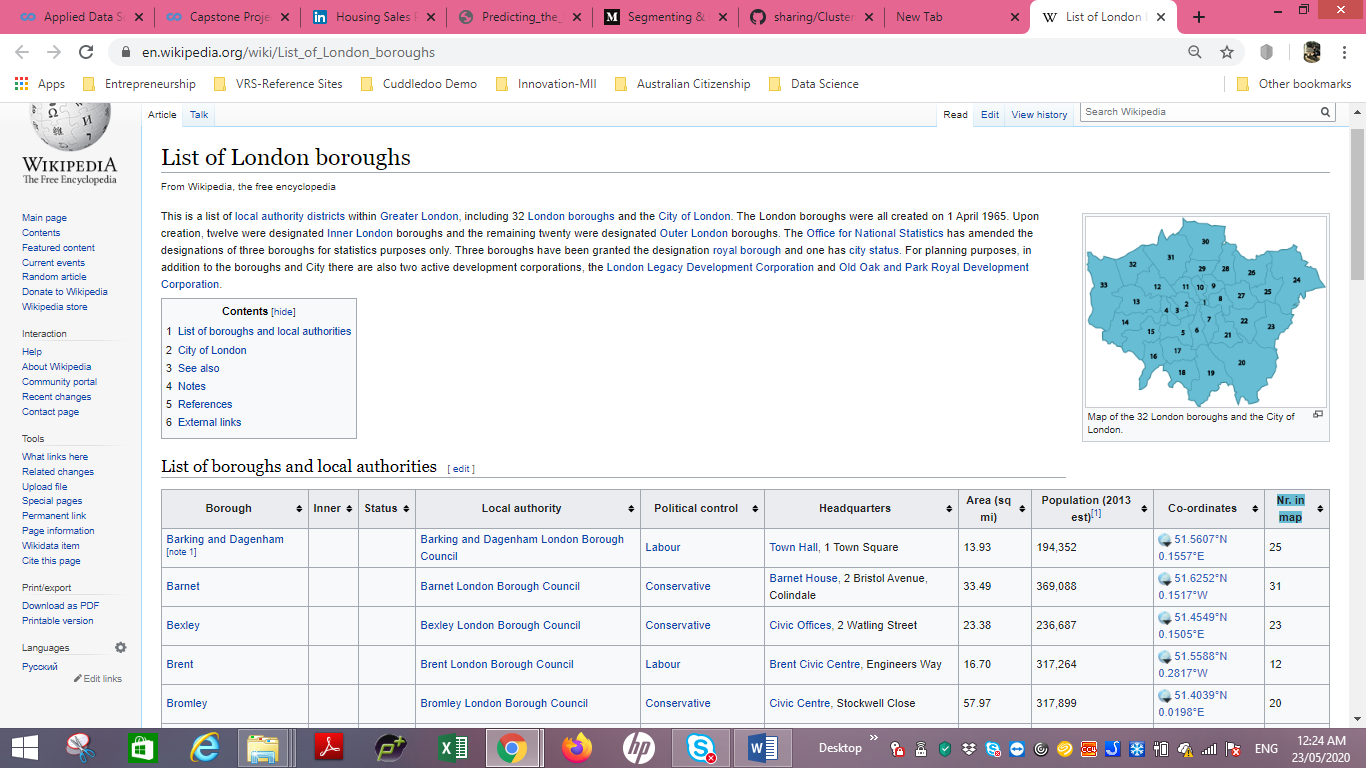
<https://en.wikipedia.org/wiki/List_of_London_boroughs>

1. Data fetched by sending messages (both regular and premium ) to the Foursquare API by using my client ID and client secret along with other mandatory fields like version and query type.

The following operations have been performed on the data:

Data Import

**Fetching and importing London borough's data from the web**. This data comprises of the borough information for all of the city of London. Below is a screenshot of the same.



**Fetching latitude and longitude details of all the boroughs.**

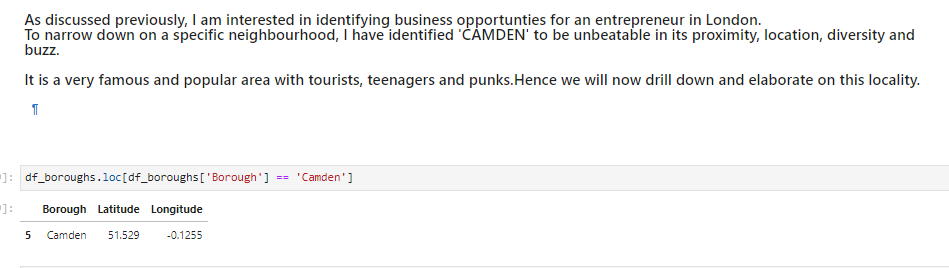


This will be needed once we try to identify venues in the different neighbourhoods around London by making use of the Foursquare API.

Data Cleansing

**Filtering the data for the selected borough = 'Camden'**

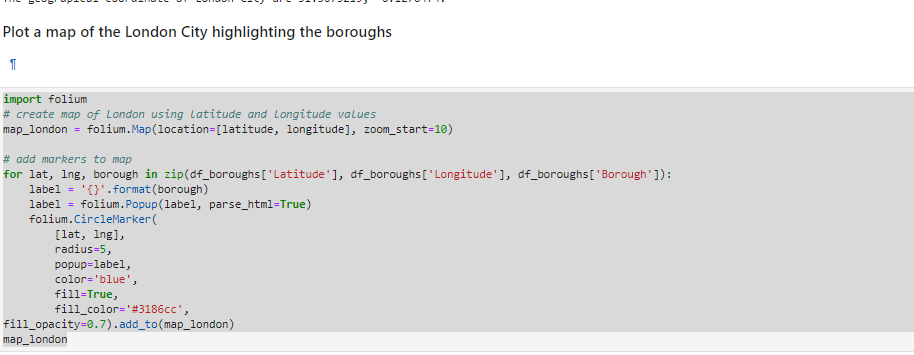
This is the preferred choice of Borough due to a detailed explanation provided in the first part of the paper.

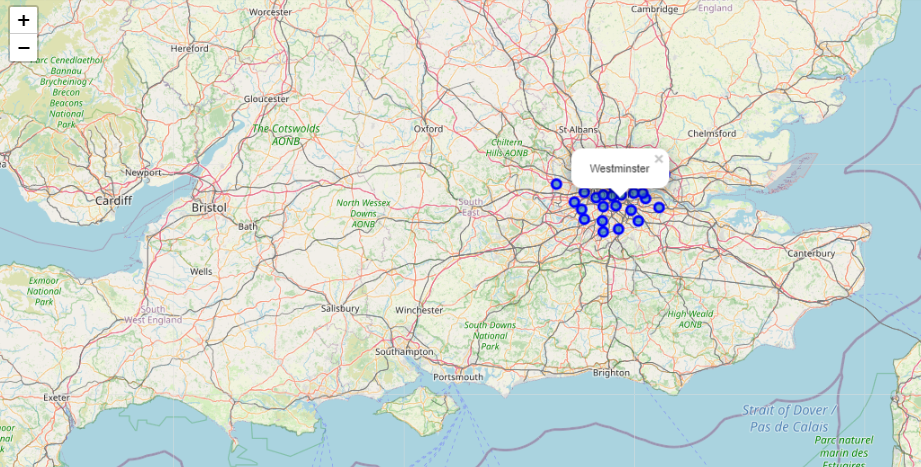


**Using Geopy library to fetch co-ordinates for London city.**



**Plot a map of the London City highlighting the boroughs.**





The map below gives us a visual feel and indication of the boroughs, how they are located to help add inputs to our judgement.

**Fetching Latitude and longitude of all the neighbourhoods of Camden**



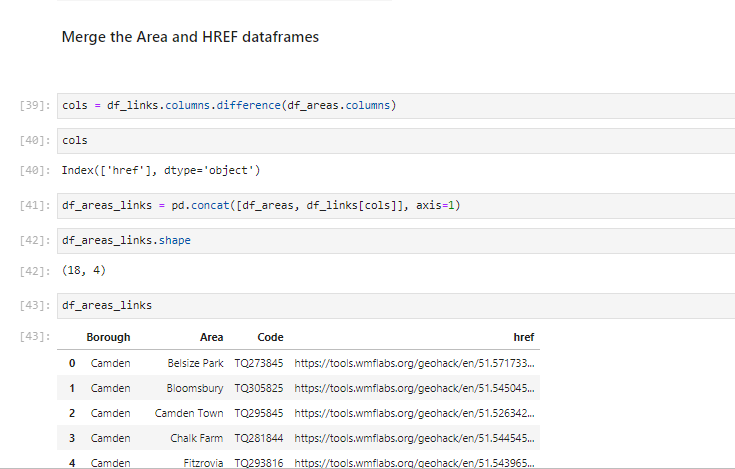
**Converting the list to a data frame**



**Building a data frame from list of her links**

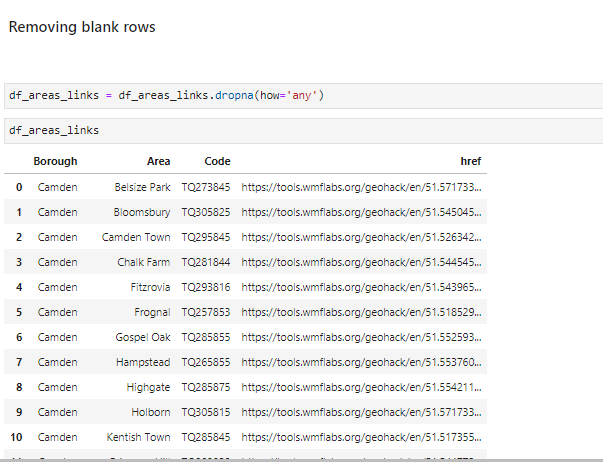


**Merge the Area and href data frames**

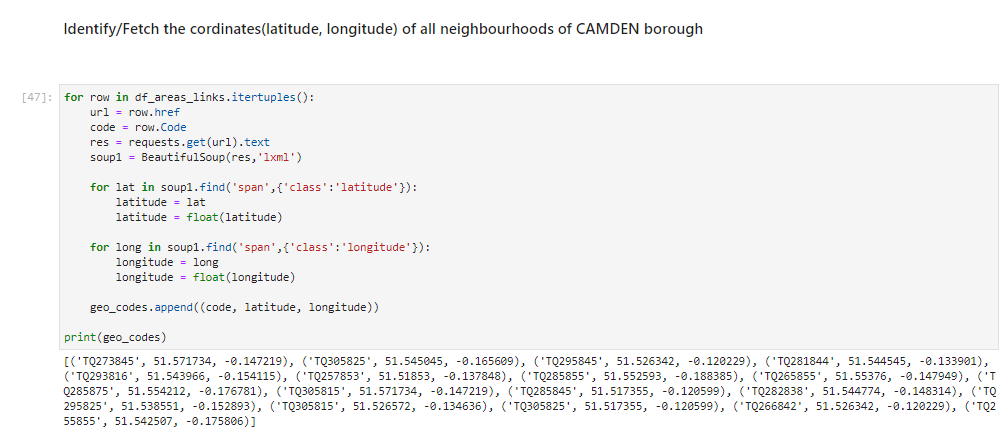


**Remove Blank Rows**

As can be seen above, there are some rows that have no data or NAN. These are adding no value so will be deleted.



**Fetch the geo co-ordinates of all the areas of CAMDEN borough.**



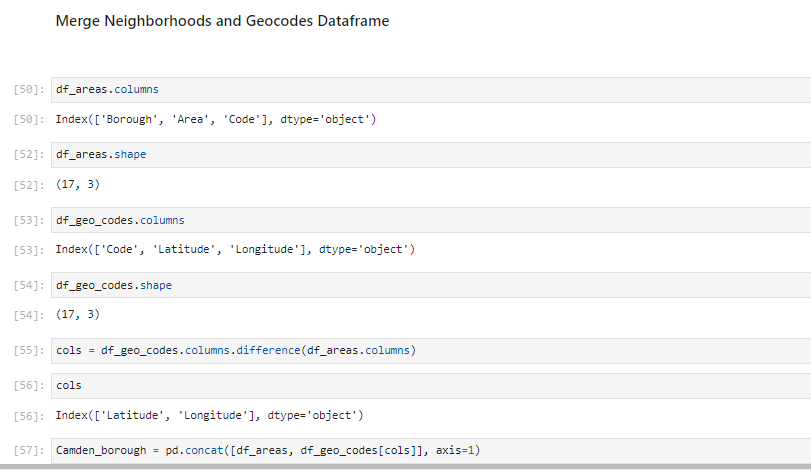
This step will aid in preparing the data to be moving to the next step. It will also help us identify venue by co-ordinates and will assist in our decision making capability.

Data Manipulation

**Build a geo-code data frame**

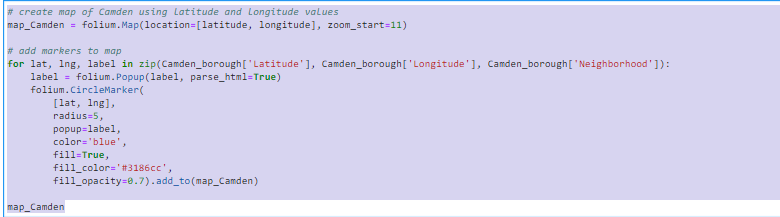


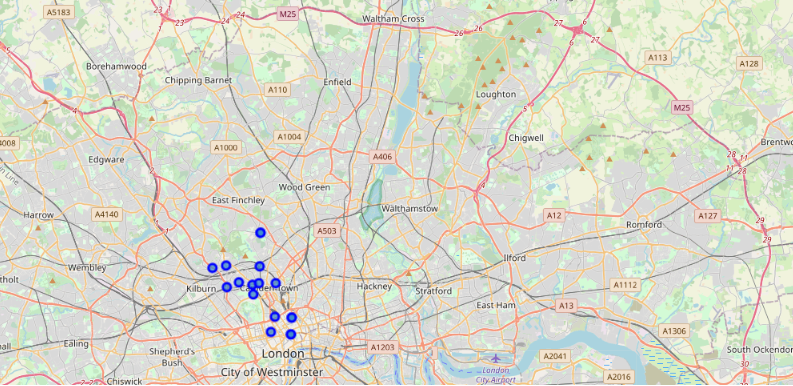
**Merge Neighbourhood and Geo-codes Data frame and Rename Column 'Area' To 'Neighbourhood' & Dropping the 'Code' column**





**Plotting a map for all the neighbourhoods of Camden.**

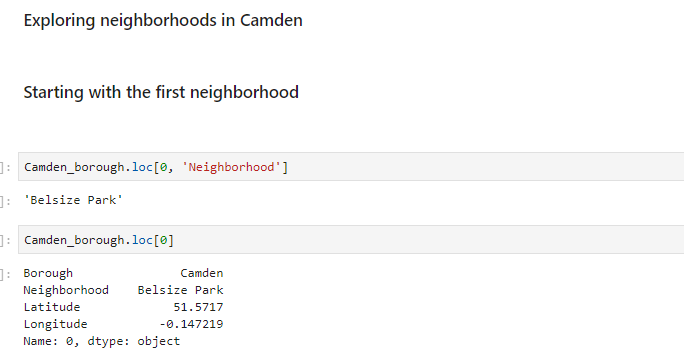




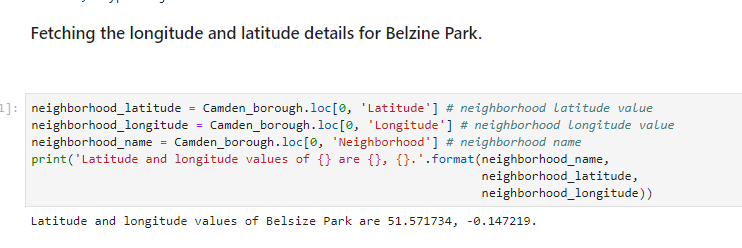
.

This visual representation gives us a good feel of the locations of the different neighbourhoods and shows us how they are placed.

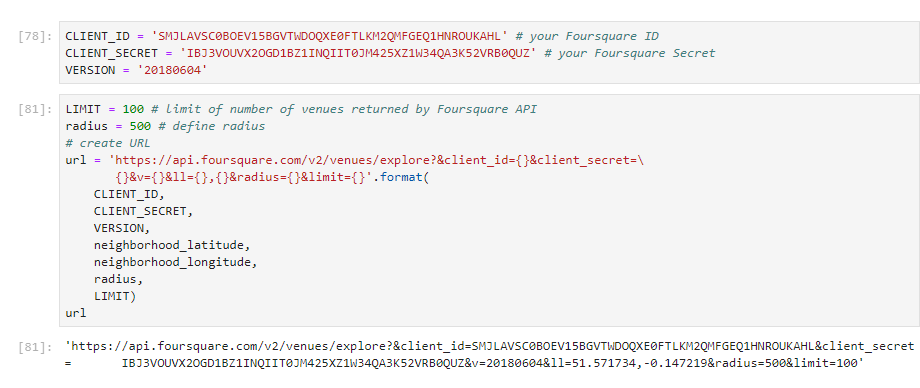
**Exploring neighbourhoods in Camden, starting with the first neighbourhood.**



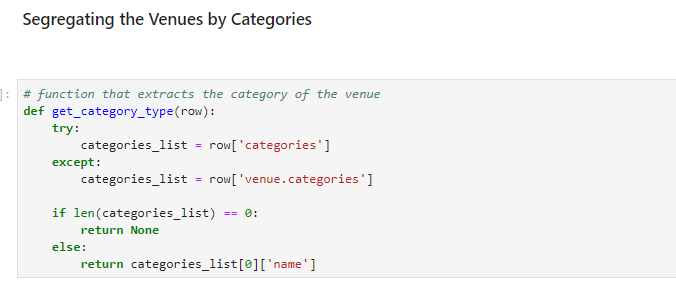
**Fetching co-ordinates for first neighbourhood, Belsize Park.**



**Fetching the top 100 places of Camden neighbourhoods within a 500 km radius by leveraging the Foursquare API.**



**Segregating venues by categories.**



**JSON Data Cleansing and formatting to data frame.**

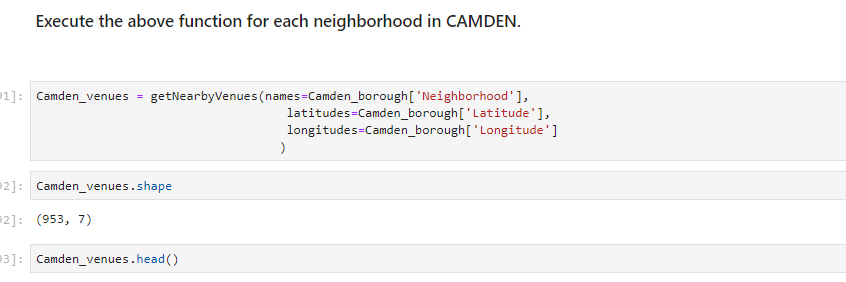




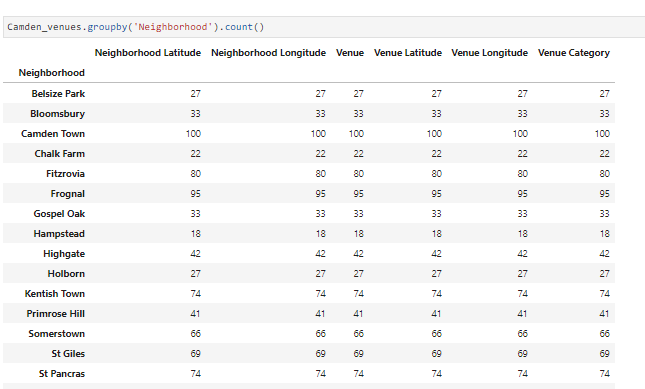
1. Results& Analysis

**Exploring the Camden neighbourhood.**

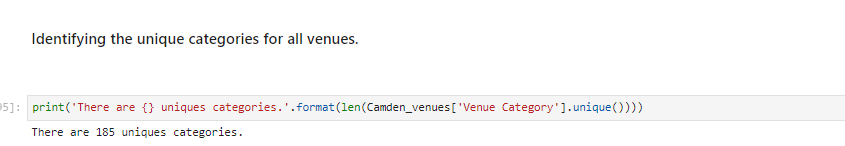




**Display venues of all neighbourhood.**

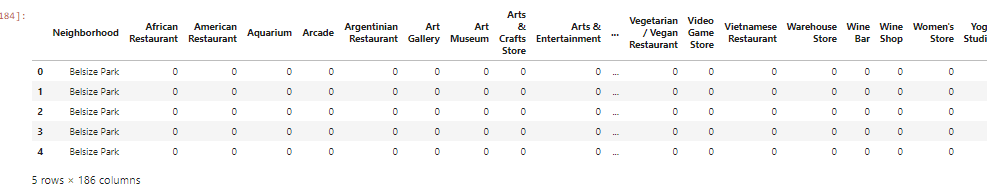


**Identify unique categories for all neighbourhoods.**



**Analysing all neighbourhoods.**

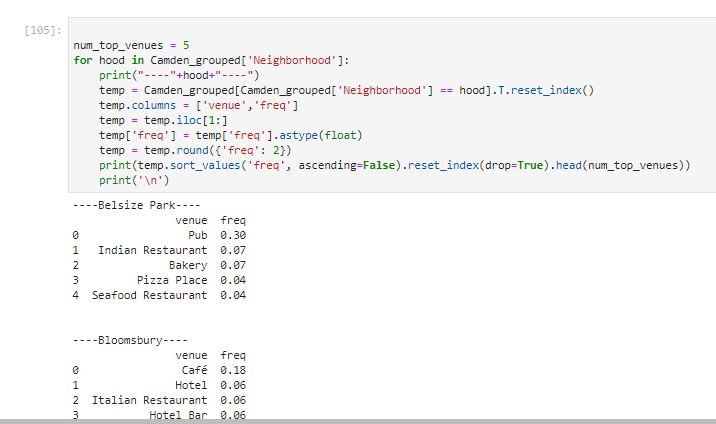




**Grouping rows by neighbourhoods**



**Printing each neighbourhood with top 5 common venues.**



………………(many more)

**Sorting venues in descending order**

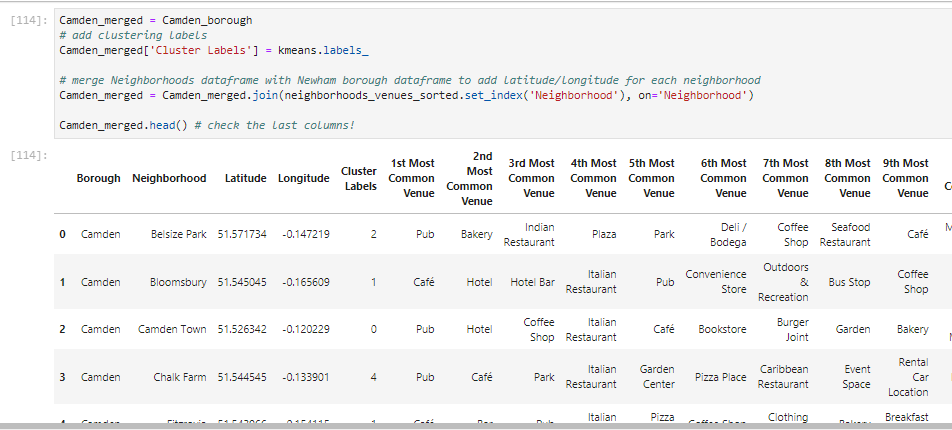


**Run K-means algorithm to cluster the neighbourhood into 5 clusters.**



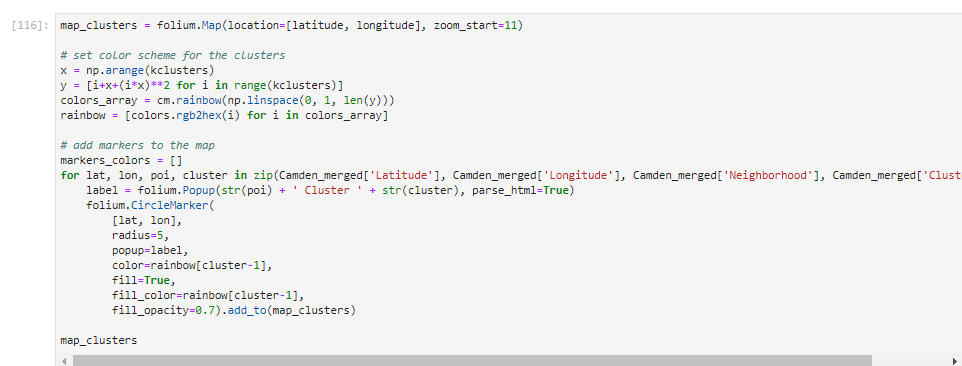
1. **Discussions**

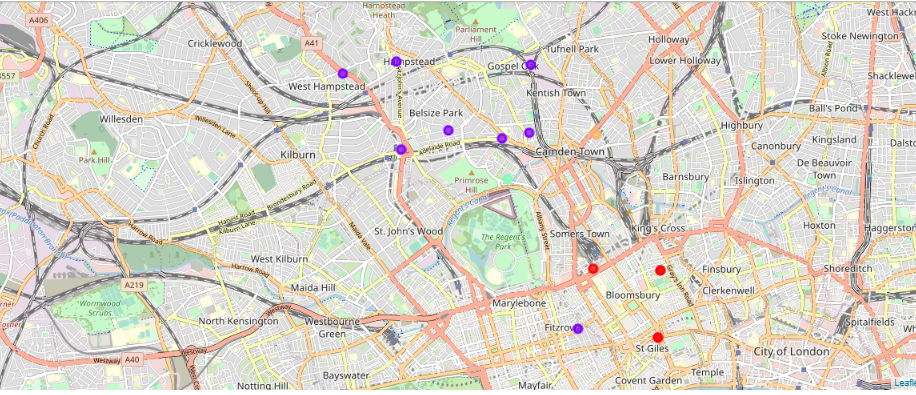
**Build a new data frame to include the cluster and top 10 venues for each neighbourhood.**



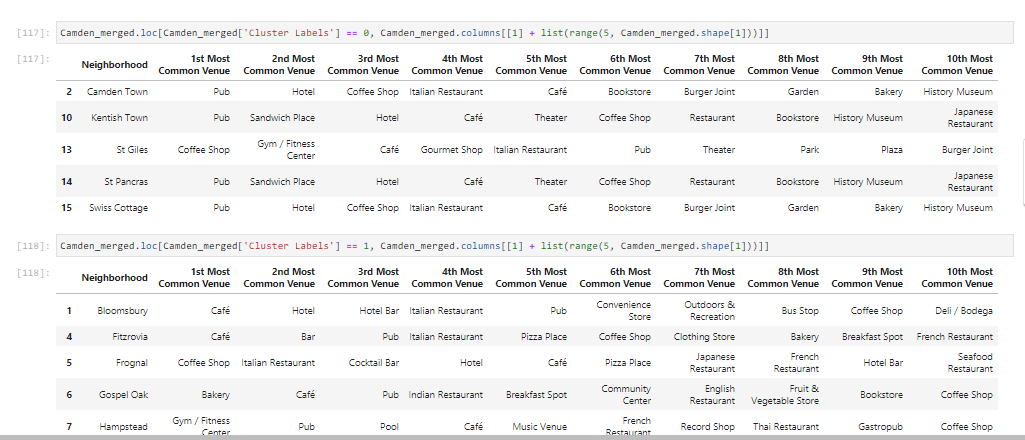
As you can see, the above data frame gives us a summarised view of the borough, neighbourhood and places of interest. This visual representation of data helps us narrow down further on the places of interest, and the profile if people who visit the

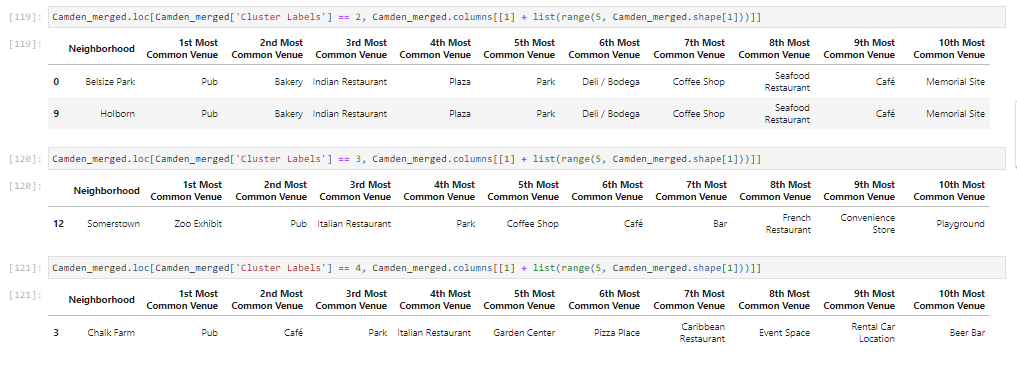
**Visualize the clusters by plotting a map.**





**Examining the Clusters**





We examine the various neighbourhoods, the cluster they belong and the popularity of venues,

For clusters 0,1,2,3 and 4. The neighbourhoods are

**6. Cluster Analysis and Conclusion**

After careful and detailed analysis of the data from the neighbourhoods, it has been identified that the neighbourhoods of Camden Town, Kentish Town, St. Pancras, Swiss Cottage, Belsize Park, Holborn and Chalk Farm have 'Pubs' as the most visited venue.

Since Camden borough is very popular among the youngsters and promotes punk culture, I have identified and recommended a new business opportunity, i.e. opening of a PUB.

The neighbourhood I have narrowed down to and identified is Camden Park since it is one of the most popular neighbourhoods of Camden and of London city.

Camden has the seventh largest economy in the UK. The area's industrial economic base has been replaced by service industries such as retail, tourism and entertainment. The area now hosts street markets and music venues that are strongly associated with alternative culture.