Problem Identification

Many Countries are populated with variety of people encompassing to different faith and different cultures. This leads to the introduction of different kinds of food cuisine in the country. People will like to find and check out a particular cuisine eg: Italian, Chinese etc and would end up recommending it or not for others. They would like to attempt this in a particular manner without missing any cuisine at all and would like to keep a track of it as well. If all the different kinds of restaurants in a particular city can be clustered and arranged in a manner with ratings will give an idea for the customer to devour the cuisine which has appealed the customers attention. The tourist who visit these countries would to like to experience the local cuisine in the area. This machine learning analyses will be able to give them a idea and what areas would be the best to visit. This data can also be explored in a such manner to find the most sought out cuisine and what would be the appropriate location to open a restaurant as well.

Data Description

Data Processing is the core of machine learning. It is the most time-consuming process of all when compared with other components of machine learning. Thereby requiring scrupulous effort in processing. Therefore, if performed well and properly documented, this would result in an output which would have a high quality in terms of its insight it can provide to stakeholders.

The Data which will be used for the analysis are presented below.

1. Web Scrape the Wikipedia Table To Obtain The areas around London and its postcodes from https://en.wikipedia.org/wiki/List_of_areas_of_London
2. Use geocoder to find the latitude and longitude for each area and update the table.
3. Use Foursquare API to obtain venues around each location for the analysis

Data Wrangling

This is main part of the analysis. The data obtained through Web Scraping is uncleaned . Therefore this needs to arranged in a format which can be read by the machine learning algorithm and also be able to interpreted to provide insights to stakeholders. Beautifulsoup API was used to obtain the table providing the areas around London from https://en.wikipedia.org/wiki/List_of_areas_of_London. After the cleaning the table is organised as below.

	Location	Borough	Post Town	Postcode
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

Obtaining Longitudes and Latitudes

The longitudes and Latitudes for each Postcode is obtained using the Geocoder API. The code for this is presented below.

```
def lat_lan_finder(postCode):
    lat_lng_coords = None
    while(lat_lng_coords is None):
        g = geocoder.arcgis('{}, London, United
Kingdom'.format(postCode))
        lat_lng_coords = g.latlng
    return lat_lng_coords
```

Using This function the longitudes and latitudes for every post code was found the table was updated. The updated table is presented below.

INDEX	Location	Borough	Postcode	Latitude	Longitude
0	Upton Park	Newham	E13	51.52653	0.02876
1	West Ham	Newham	E15	51.54014	0.00278
3	Woodford	Redbridge	E18	51.58977	0.03052
4	Hackney	Hackney	E8	51.54505	-0.05532
5	Angel	Islington	N1	51.52969	-0.08697

Assumptions

Certain Assumptions were made before attempting the Foursquare API call to prevent exceeding premium call which are listed below.

1. Unique Post codes were only analysed. As duplicate post codes still returned the same results thus this increased time and memory consumption. Eliminating this decreased the time consuming in obtaining results.
2. The Post town of London was only analysed thereby removing all Post towns apart from London.

Therefore , The Foursquare API was used to analyse 151 Unique post codes after simplifying the database.

Foursquare API

The Foursquare API was used to obtain the venues around the areas for clustering the city for providing a restaurant list what to visit in each city for stakeholders without neglecting the restaurants the Locality is famous for.

The Foursquare id was the food category which Is **'4d4b7105d754a06374d81259'** . The limit was placed for 200 in a radius of 1000m. This was as it is convenient for a person to cover this distance by walk .

The Foursquare API returned 104 Unique restaurants for all the areas. The majority of restaurant categories are presented below image.

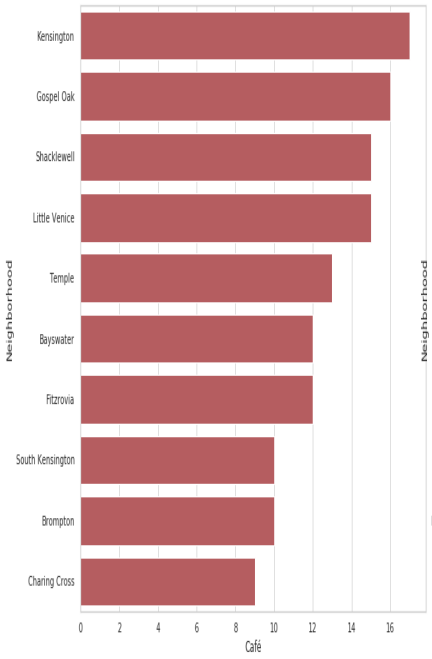
	Count
Café	537
Italian Restaurant	241
Sandwich Place	193
Bakery	181
Pizza Place	173

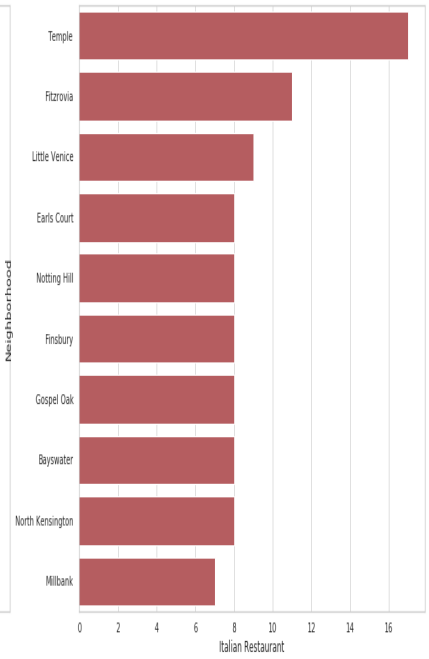
Neighborhood	Count
Temple	181
Angel	122
Little Venice	104
Fitzrovia	100
Tower Hill	89

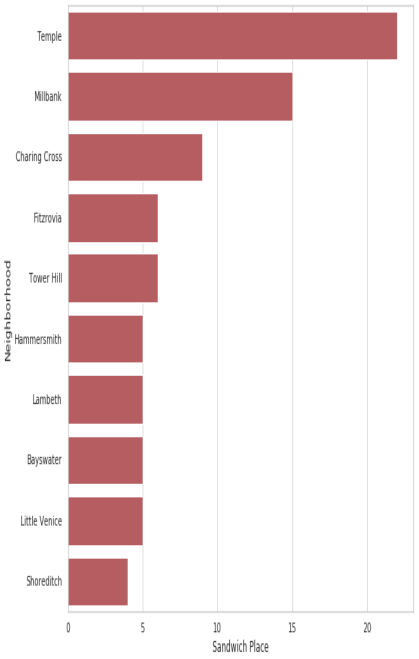
It can be seen that Café occupies the most and records of 537 venues in more than 150 venues around the city of London. This is not to be amazed as the culture of United Kingdom is brewed upon this.

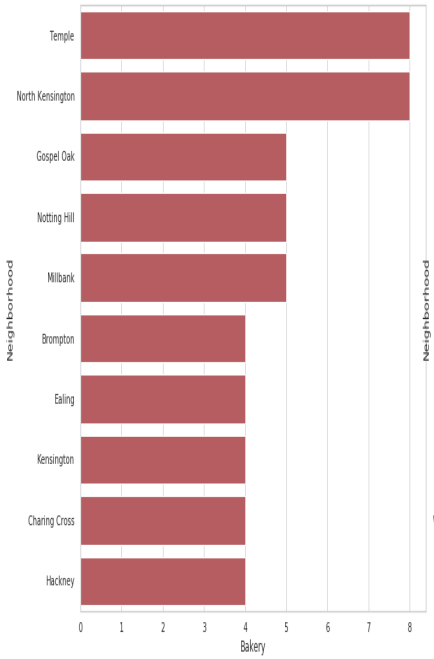
Furthermore, it was analysed to find out which localities recorded and provided highest venues and it was found out that Temple locality was the highest providing 181 venues. This is a legal district of London. This is also provides that there are more trending venues around this locality as higher food category venues depicting the ability to cater to a large group of people in the area.

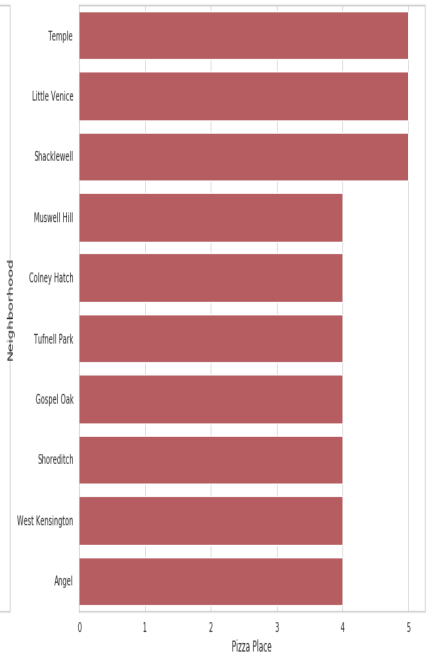
The below image presents the localities which present the highest food venues for the categories presented in above table. It is also visualised that Temple locality has the highest venues for most categories as well.


Neighborhood
Kensington
Gospel Oak
Shacklewell
Little Venice
Temple
Bayswater
Fitzrovia
South Kensington
Brompton
Charing Cross
0 2 4 6 8 10 12 14 16
Café


Neighborhood
Temple
Fitzrovia
Little Venice
Earls Court
Notting Hill
Finsbury
Gospel Oak
Bayswater
North Kensington
Millbank
0 2 4 6 8 10 12 14 16
Italian Restaurant


Neighborhood
Temple
Millbank
Charing Cross
Fitzrovia
Tower Hill
Hammersmith
Lambeth
Bayswater
Little Venice
Shoreditch
0 5 10 15 20
Sandwich Place


Neighborhood
Temple
North Kensington
Gospel Oak
Notting Hill
Millbank
Brompton
Ealing
Kensington
Charing Cross
Hackney
0 1 2 3 4 5 6 7 8
Bakery


Neighborhood
Temple
Little Venice
Shacklewell
Muswell Hill
Colney Hatch
Tufnell Park
Gospel Oak
Shoreditch
West Kensington
Angel
0 1 2 3 4 5
Pizza Place

The venues were simplified and grouped according to the mean of occurrence in each neighbourhood in order to begin the machine learning procedure. After the one hot encoding and mean grouping the following table is obtained.

	Neighborhood	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	BBQ Joint	Bagel Shop	Bakery
0	Acton	0.0	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.120000
1	Aldwych	0.0	0.0	0.0	0.012346	0.000000	0.0	0.0	0.012346	0.000000	0.049383
2	Angel	0.0	0.0	0.0	0.000000	0.016393	0.0	0.0	0.032787	0.008197	0.000000
3	Arnos Grove	0.0	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.000000
4	Balham	0.0	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.076923

Furthermore, for each locality the top 10 most trended venues were also tabulated to understand what is famous for each venues.

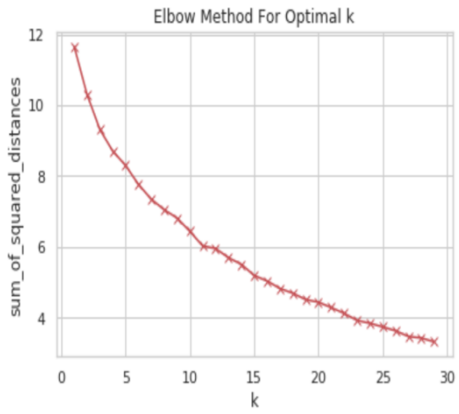
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Acton	Café	Bakery	Italian Restaurant	French Restaurant	Breakfast Spot	Restaurant	Burger Joint	English Restaurant	Portuguese Restaurant	Chinese Restaurant
1	Aldwych	Sandwich Place	Café	Korean Restaurant	Restaurant	Japanese Restaurant	Chinese Restaurant	Bakery	Italian Restaurant	Burger Joint	Fish & Chips Shop
2	Angel	Café	Italian Restaurant	Food Truck	Breakfast Spot	Sandwich Place	Vietnamese Restaurant	Pizza Place	Sushi Restaurant	BBQ Joint	Restaurant
3	Arnos Grove	Fast Food Restaurant	Pizza Place	Café	Spanish Restaurant	Italian Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	Chinese Restaurant	Portuguese Restaurant	Sandwich Place
4	Balham	Fish & Chips Shop	Indian Restaurant	Café	Fast Food Restaurant	Bakery	Breakfast Spot	Pizza Place	Chinese Restaurant	Portuguese Restaurant	Caucasian Restaurant

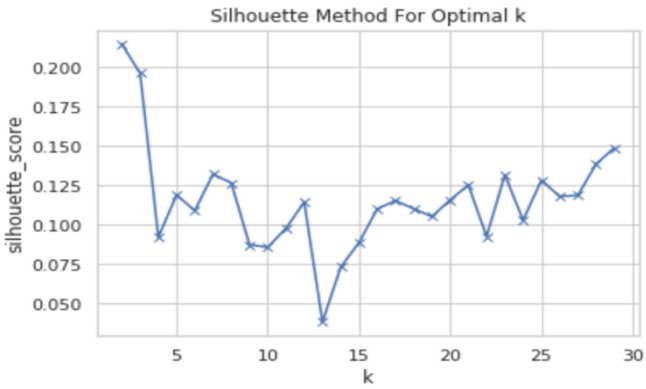
Machine Learning

The algorithm used in performing the Machine Learning is KMeans Clustering algorithm. This is an hierarchical algorithm, which groups similar items in the same cluster by minimising the inter cluster distance and maximising the intra cluster distances for each data point.

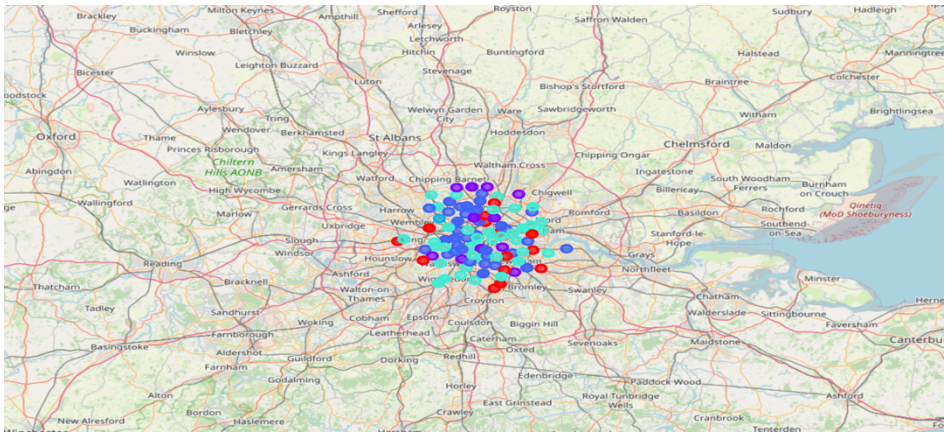
```
# set number of clusters
kclusters = 5

London_grouped_clustering =
London_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters,
random_state=0).fit(London_grouped_clustering)

# check cluster labels generated for each row in the
dataframe
kmeans.labels [0:10]
```

Initially K of 5 was chosen to understand the visibility but the most feasible k was found to be 9 from the Silhouette Score and Minimum Sqaured Sum of Distances from the following graphs





It can be seen elbow begins at 8 and there is low Silhouette score at 9 thus this was chosen. After the performing the clustering for K=9 the following map is obtained.



It can be seen that only 5 cluster were utilised which are the dominated clusters. Following depicts the most frequented venues for each cluster.

```
Cluster 0
Café                        13
Fast Food Restaurant         2
Fried Chicken Joint          2
Name: 1st Most Common Venue, dtype: int64
```

```
Cluster 1
Café                        7
Pizza Place                 4
Fast Food Restaurant        3
Turkish Restaurant          1
Italian Restaurant          1
Name: 1st Most Common Venue, dtype: int64
```

```
Cluster 2
Café                        39
Indian Restaurant            3
Italian Restaurant           2
Vietnamese Restaurant        1
Bakery                       1
Asian Restaurant             1
Pizza Place                  1
Name: 1st Most Common Venue, dtype: int64
----------------------------------------------
```

```
Cluster 3
Bakery    1
Name: 1st Most Common Venue, dtype: int64
----------------------------------------------
```

```
Cluster 4
Café                        11
Indian Restaurant           10
Fast Food Restaurant         8
Sandwich Place               5
Chinese Restaurant           3
Pizza Place                  2
Turkish Restaurant           2
Fish & Chips Shop            2
Restaurant                   2
Portuguese Restaurant        1
Japanese Restaurant          1
Italian Restaurant           1
French Restaurant            1
Fried Chicken Joint          1
Burger Joint                 1
Name: 1st Most Common Venue, dtype: int64
----------------------------------------------
```

Discussion and Future Works

From the Clustering provided it can be seen certain defects still exists due to less data regarding the demographics of the locality. If information related to demographical features such as Population, Income and house prices this can be used to cluster in an better and more convincing clusters for the audiences even though the current solution also does its purpose by the giving the stakeholders which locality to visit in order to have an awareness of its popularity.

The Limitation of calls for the Foursquare API limits the analyse without giving the ability to obtain vital information such as ratings of each venues providing a high excellence of visions to the concerned parties allowing them to make improved decisions.