Capstone project: The Battle of the Neighbourhoods.

An analysis of London's Borough of Merton to determine where's best to situate a new Chinese restaurant.

Author: Colin Kagame, June 2020.

Summary.

This assignment attempts to analyse and explore the different neighbourhoods of a borough in the city of London, UK called Merton. The sole purpose for this is to find the most suitable neighbourhood in which to open up a Chinese restaurant within Merton. To perform this analysis, we will use K-means clustering, an unsupervised machine learning algorithm.

1. Introduction.

Background: London is England's capital and the largest city of England and the United Kingdom. It is considered one of the world's most visited cities and boasts a very vibrant economy, attracting multitudes of people from all over the world for tourism, education, work and settlement purposes. London has a diverse range of peoples and cultures, and more than 300 languages are spoken within the region. The city's estimated mid-2018 municipal population (corresponding to Greater London) was close to 9 million making it the third most populous of any city in Europe.

> According to Trustforlondon, London's population is estimated to reach 10 million.

For these reasons it is clear that London has a large demand relating to food/catering to accommodate not only its large population, but also its culturally diverse population. For this assignment we will explore, cluster, and perform an in-depth analysis on one of the boroughs of London to determine where is best to open up a Chinese restaurant.

Target audience: The target audience for this is someone who is seeking to open up a restaurant in the residential boroughs of Southwest London, particularly Merton. One can benefit by finding out which are the most popular restaurant types in the neighbourhoods/towns of Merton as per how many of them are n a borough and customer ratings. For this assignment we chose to focus on a businessperson who wants to open a Chinese restaurant, but the analysis results could aid other individuals finding what restaurant types/cuisines are popular in particular neighbourhoods.

2. Data.

Two sets of data were utilised in this assignment. The first dataset required was that constituting the city and corresponding neighbourhood data needed to explore Merton, London, UK which is our city of choice for this analysis. For this I used data from Wikipedia, similar to what we did in Week 3's Toronto analysis where data was scraped from a Wikipedia page using beautifulsoup etc. The Wikipedia page I have used for this assignment is titled, 'List of Areas of London.' This constitutes the data we need such as Boroughs, Districts, Postcodes and more. I aim to scrape this Wikipedia page in similar fashion to our Toronto task, and then store this data in data frames.

Then for the venues around the locations obtained above, including their coordinate details, we obtained the data from Foursquare's venue/search API that allowed us to pull this data from Foursquare's database.

Data sources used for this assignment:

a) **Wikipedia:** London location data, such as towns, boroughs and postcodes etc. For this we used Wikipedia: https://en.wikipedia.org/wiki/List_of_areas_of_London

b) **Foursquare:** For data relating to venue categories, venues and their respective neighbourhoods, including their latitude and longitudes values we utilised the Foursquare API.

- Foursquare Search/Places API: https://developer.foursquare.com/docs/api-reference/venues/search/
- Foursquare Venue Categories. https://developer.foursquare.com/docs/build-with-foursquare/categories/

The data obtained from Foursquare was then used in conjunction to that which we scraped off the Wikipedia page to create the necessary data frames, maps etc.

Libraries to be used.

Pandas: For creating and manipulating data frames.

- Folium: Python visualization library to visualize the neighbourhoods cluster distribution of using interactive leaflet map.
- Scikit Learn: To import k-means clustering.
- JSON: Library to handle JSON files.
- XML: To separate data from presentation XML i.e. stores data in plain text format.
- Geocoder: For retrieving Location Data.
- Beautiful Soup and Requests: To scrape Wikipedia page and to handle http requests.
- Matplotlib: Python Plotting Module.

3. Methodology.

3.1: Data preparation, cleaning and further pre-processing.

I started off by **collecting the first set of data.** I scraped the table containing a list of the areas of London consisting essential information such as Locations, Boroughs, Post Towns and Districts. To do this I utilised the beautifulsoup library which is responsible for pulling HTML and XML files.

Original Wikipedia page.

Location	London borough	Post town	Postcode district	Dial code	OS grid ref
Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728
Aldborough Hatch	Redbridge[9]	ILFORD	IG2	020	TQ455895
Aldgate	City[10]	LONDON	EC3	020	TQ334813
Aldwych	Westminster[10]	LONDON	WC2	020	TQ307810
Alperton	Brent[11]	WEMBLEY	HA0	020	TQ185835
Anerley	Bromley[11]	LONDON	SE20	020	TQ345695
Angel	Islington[8]	LONDON	EC1, N1	020	TQ345665
Aperfield	Bromley[11]	WESTERHAM	TN16	01959	TQ425585
Archway	Islington[12]	LONDON	N19	020	TQ285875
Ardleigh Green	Havering[12]	HORNCHURCH	RM11	01708	TQ535895
Arkley	Barnet[12]	BARNET, LONDON	EN5, NW7	020	TQ225955
Arnos Grove	Enfield[12]	LONDON	N11, N14	020	TQ295925
Balham	Wandsworth[13]	LONDON	SW12	020	TQ285735
Bankside	Southwark[14]	LONDON	SE1	020	TQ325795
Barbican	City[14]	LONDON	EC1	020	TQ322818
Barking	Barking and Dagenham[14]	BARKING	IG11	020	TQ440840
Barkingside	Redbridge[15]	ILFORD	IG6	020	TQ445895

Scraped data frame. (Figures below show head () i.e. first 5 rows.)

[2]:

	Location	London borough	Post town	Postcode district	Dial code	OS grid ref
0	Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
2	Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728

I then proceeded to do the following with regards to further **cleansing of the data:**

- Removed unnecessary columns i.e. 'Dial code' and 'OS grid ref.'
- Removed hyperlinks/references
- Removed values that contain more than one result.
- Renamed the column names to single word strings to eliminate spaces.
- Changed the results of Post town to Sentence case.

Result:

[4]:

	Location	Borough	Town	Postcode
0	Abbey Wood	Bexley	London	SE2
1	Acton	Ealing	London	W3
2	Addington	Croydon	Croydon	CR0
3	Addiscombe	Croydon	Croydon	CR0
4	Albany Park	Bexley	Bexley	DA5

- I then filtered the data to show only results belonging to London's borough of Merton. In addition, I reset the index values of our new data frame.

From this:

[5]:

	Location	Borough	Town	Postcode
107	Colliers Wood	Merton	London	SW19
294	Lower Morden	Merton	Morden	SM4
305	Merton Park	Merton	London	SW19
311	Mitcham	Merton	Mitcham	CR4
313	Morden	Merton	Morden	SM4
314	Morden Park	Merton	Morden	SM4
383	Raynes Park	Merton	London	SW20
419	South Wimbledon	Merton	London	SW19
428	St Helier	Merton	Morden	SM4
520	Wimbledon	Merton	London	SW19

To this:

[7]:

	Location	Borough	Town	Postcode
0	Colliers Wood	Merton	London	SW19
1	Lower Morden	Merton	Morden	SM4
2	Merton Park	Merton	London	SW19
3	Mitcham	Merton	Mitcham	CR4
4	Morden	Merton	Morden	SM4
5	Morden Park	Merton	Morden	SM4
6	Raynes Park	Merton	London	SW20
7	South Wimbledon	Merton	London	SW19
8	St Helier	Merton	Morden	SM4
9	Wimbledon	Merton	London	SW19

Using geocoder to retrieve geolocation data points:

After successfully achieving our first clean dataset as illustrated previously, we then proceeded utilise the geocoder package to retrieve geographical location data. I installed and imported the necessary geocoder libraries and then set the location boundaries to 'Merton, UK'. I then saved the 'Location' column values of our first dataframe into a list and then for each of these locations/neighbourhoods, I requested their respective latitude and longitude values. These latitudes and longitudes values were then saved to a separate pandas dataframe. The final step here was to merge the coordinates dataframe with our first locations dataframe as shown below.

[10]:

	Location	Borough	Town	Postcode	Latitude	Longitude
0	Colliers Wood	Merton	London	SW19	51.420690	-0.175670
1	Lower Morden	Merton	Morden	SM4	51.388005	-0.209364
2	Merton Park	Merton	London	SW19	51.411418	-0.208302
3	Mitcham	Merton	Mitcham	CR4	53.796853	-3.006069
4	Morden	Merton	Morden	SM4	51.411810	-0.192434
5	Morden Park	Merton	Morden	SM4	51.411810	-0.192434
6	Raynes Park	Merton	London	SW20	51.403040	-0.207220
7	South Wimbledon	Merton	London	SW19	51.415640	-0.191420
8	St Helier	Merton	Morden	SM4	51.398485	-0.187018
9	Wimbledon	Merton	London	SW19	51.415640	-0.191420

On running the .shape() method, we saw the dataframe contains one borough (Merton) and 10 neighbourhoods/locations.

[12]:
```
print('This area has {} boroughs and contains {} locations.'.format(
        len(df2 ['Borough'].unique()),
        df2 .shape[0]
    )
)
```

```
This area has 1 boroughs and contains 10 locations.
```

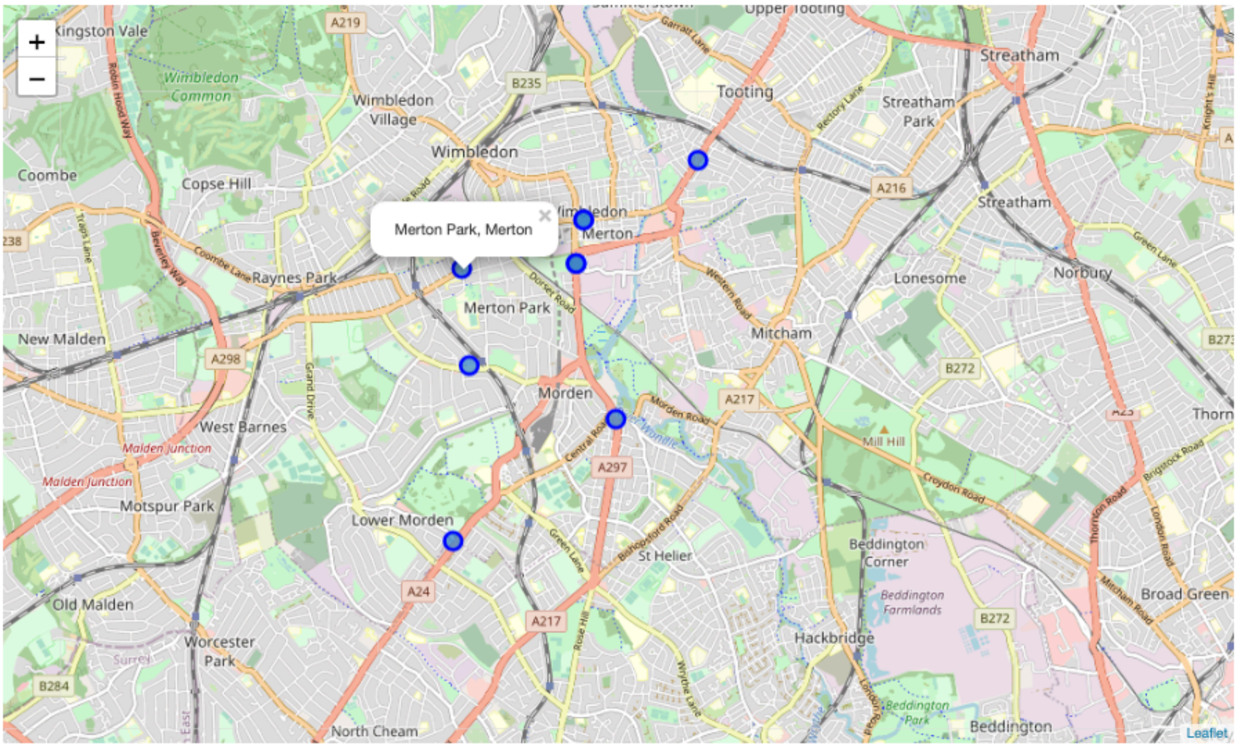
3.2: Exploratory data analysis.

- **Visualising the borough of Merton, London, UK.**

Below is a map visualising Merton and its corresponding neighbourhoods from the previously created data frame above. I utilised Folium's python library to achieve this.

A map of the London borough of Merton and its neighbourhoods.

[14]:



- **Utilising Foursquare to explore Merton's venues.**

In this section we utilise the Foursquare API to explore and segment Merton's neighbourhoods.

For each neighbourhood/location, we are gathering a list of all the different restaurants, their categories (cuisines) and their corresponding venue latitude and longitude values thus creating an even more informative data frame to work with. The data frame image below only shows the first 10 rows. We also learn that there are 217 venues across Merton that are restaurants.

In this section we explore and analyse multiple facets with regard to categories, locations, quantities and spread of the restaurants across the neighbourhoods of Merton.

A data frame containing all restaurants in Merton, their latitude, longitude values and their category.

[18]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Colliers Wood	51.42069	-0.17567	Thai Town	51.421436	-0.175059	Thai Restaurant
1	Colliers Wood	51.42069	-0.17567	Little Vietnam	51.415625	-0.192852	Vietnamese Restaurant
2	Colliers Wood	51.42069	-0.17567	Rosa's Thai Cafe	51.427388	-0.168848	Thai Restaurant
3	Colliers Wood	51.42069	-0.17567	Confucius	51.420605	-0.163035	Chinese Restaurant
4	Colliers Wood	51.42069	-0.17567	Golden House	51.420179	-0.187175	Chinese Restaurant
5	Colliers Wood	51.42069	-0.17567	Daddy Bao	51.425509	-0.164612	Asian Restaurant
6	Colliers Wood	51.42069	-0.17567	Dinner King	51.426281	-0.165999	Chinese Restaurant
7	Colliers Wood	51.42069	-0.17567	Koi Ramen Bar	51.429048	-0.166357	Ramen Restaurant
8	Colliers Wood	51.42069	-0.17567	Maki Yaki	51.418269	-0.193211	Sushi Restaurant
9	Colliers Wood	51.42069	-0.17567	Takahashi	51.416424	-0.192304	Sushi Restaurant

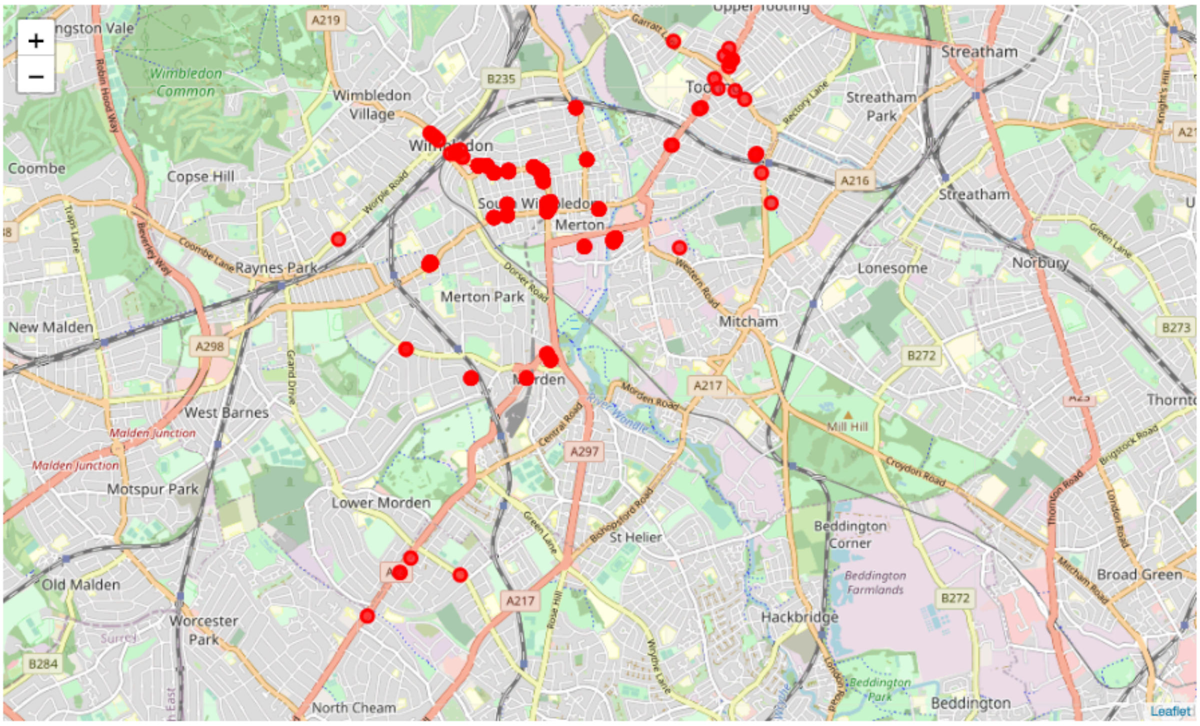
[19]:

```
merton_venues_asian.shape
```

[19]:

```
(217, 7)
```

Visualising **ALL restaurants across Merton**, London, UK.



Visualising only **Chinese restaurants across Merton**, London, UK.

I then filtered the previous data frame containing all restaurants to show only Merton's Chinese restaurants. We can observe that there are about 80 Chinese restaurants.

[70]:

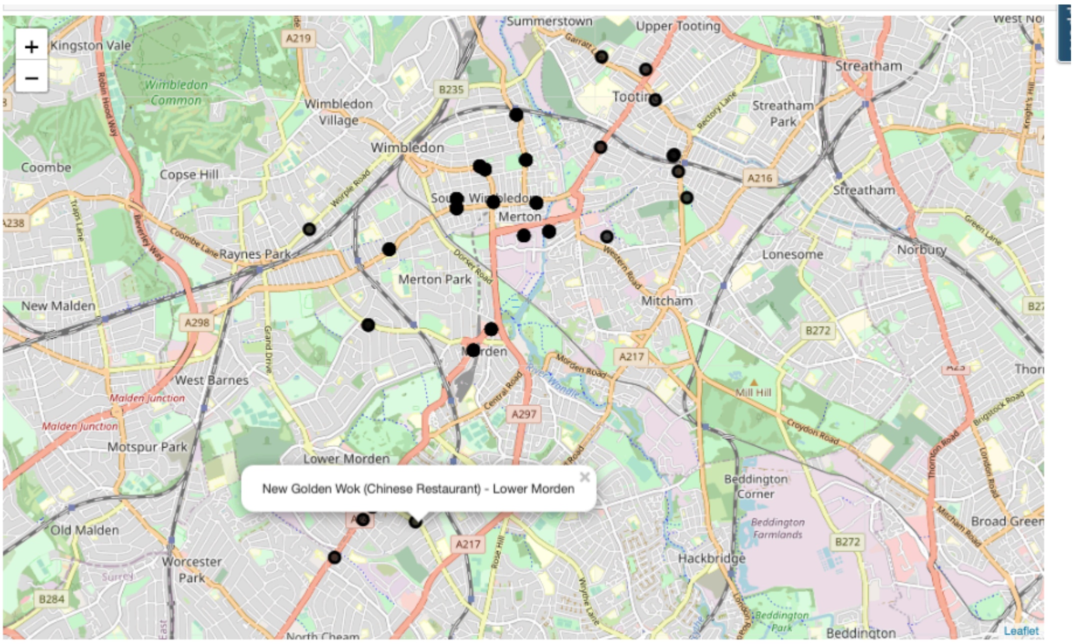
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
2	Colliers Wood	51.42069	-0.17567	Dinner King	51.426281	-0.165999	Chinese Restaurant
6	Colliers Wood	51.42069	-0.17567	The Kung Fu	51.415888	-0.192432	Chinese Restaurant
7	Colliers Wood	51.42069	-0.17567	Chinese Food @ Tooting Market	51.429357	-0.167543	Chinese Restaurant
11	Colliers Wood	51.42069	-0.17567	Peter's Chinese	51.416333	-0.160918	Chinese Restaurant
14	Colliers Wood	51.42069	-0.17567	China Dragon	51.430656	-0.174815	Chinese Restaurant

[71]:

```
print(Chinese_Rest.shape)
```

```
(80, 7)
```

[72]:



Number of Chinese restaurants per neighbourhood.

Neighborhood	Chinese Restaurants
Colliers Wood	14
Lower Morden	4
Merton Park	8
Mitcham	3
Morden	14
Morden Park	14
Raynes Park	4
South Wimbledon	11
St Helier	2
Wimbledon	11

Number of restaurants (ALL) per borough.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Colliers Wood	33	33	33	33	33	33
Lower Morden	5	5	5	5	5	5
Merton Park	30	30	30	30	30	30
Mitcham	3	3	3	3	3	3
Morden	36	36	36	36	36	36
Morden Park	36	36	36	36	36	36
Raynes Park	8	8	8	8	8	8
South Wimbledon	31	31	31	31	31	31
St Helier	4	4	4	4	4	4
Wimbledon	31	31	31	31	31	31

We also learn as shown below that there are only 11 restaurant categories within Merton. This indicates that as a Chinese restaurant in the borough, the competition would be only 10 other cuisines/restaurant categories which is a positive factor.

```
[23]: print('There are {} unique categories.'.format(len(merton_venues_asian['Venue Category'].unique())))
There are 11 unique categories.
```

3.3: Machine Learning.

In order to analyse the data for machine learning purposes, we used the **One hot encoding** technique which transforms categorical data into numerical data. The 'venue category' occurrence frequencies for each neighbourhood were obtained. In this case the venue category is the restaurant cuisine category e.g. Thai Restaurant, Chinese Restaurant etc.

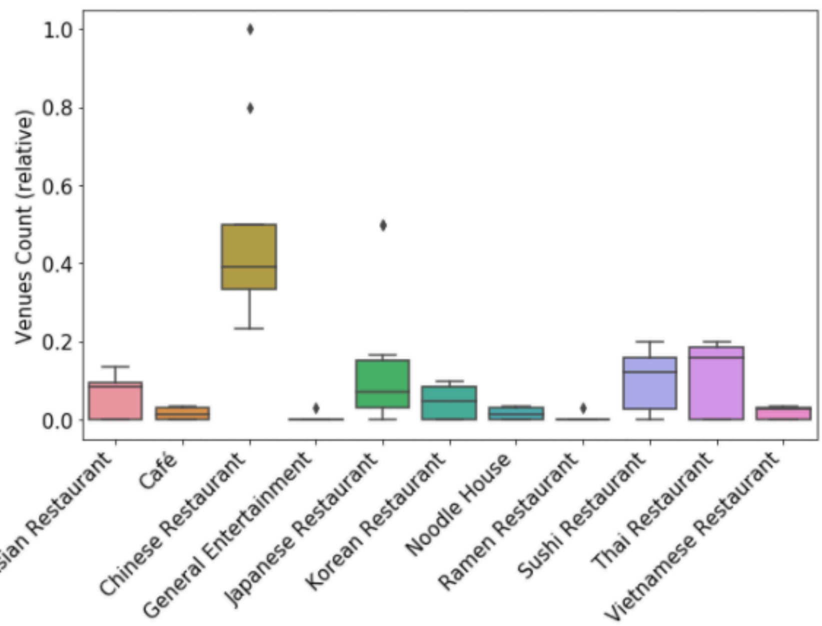
	Neighborhood	Asian Restaurant	Café	Chinese Restaurant	General Entertainment	Japanese Restaurant	Korean Restaurant	Noodle House	Ramen Restaurant	Sushi Restaurant	Thai Restaurant	Vietnamese Restaurant
0	Colliers Wood	0	0	0	0	0	0	0	0	0	1	0
1	Colliers Wood	0	0	0	0	0	0	0	0	0	0	1
2	Colliers Wood	0	0	0	0	0	0	0	0	0	1	0
3	Colliers Wood	0	0	1	0	0	0	0	0	0	0	0
4	Colliers Wood	0	0	1	0	0	0	0	0	0	0	0

We then proceed to group the rows in the dataframe above by 'Neighbourhood' and then calculate a mean/average of the occurrence frequency for each venue category. Utilising the seaborn library, a box plot representing these values was then plotted as is shown below.

	Neighborhood	Asian Restaurant	Café	Chinese Restaurant	General Entertainment	Japanese Restaurant	Korean Restaurant	Noodle House	Ramen Restaurant	Sushi Restaurant	Thai Restaurant	Vietnamese Restaurant
0	Colliers Wood	0.090909	0.000000	0.424242	0.030303	0.030303	0.060606	0.000000	0.030303	0.151515	0.151515	0.030303
1	Lower Morden	0.000000	0.000000	0.800000	0.000000	0.000000	0.000000	0.000000	0.000000	0.200000	0.000000	0.000000
2	Merton Park	0.133333	0.033333	0.233333	0.000000	0.166667	0.033333	0.033333	0.000000	0.133333	0.200000	0.033333
3	Mitcham	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	Morden	0.083333	0.027778	0.361111	0.000000	0.111111	0.083333	0.027778	0.000000	0.111111	0.166667	0.027778
5	Morden Park	0.083333	0.027778	0.361111	0.000000	0.111111	0.083333	0.027778	0.000000	0.111111	0.166667	0.027778
6	Raynes Park	0.000000	0.000000	0.500000	0.000000	0.500000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	South Wimbledon	0.096774	0.032258	0.322581	0.000000	0.032258	0.096774	0.032258	0.000000	0.161290	0.193548	0.032258
8	St Helier	0.000000	0.000000	0.500000	0.000000	0.500000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	Wimbledon	0.096774	0.032258	0.322581	0.000000	0.032258	0.096774	0.032258	0.000000	0.161290	0.193548	0.032258

Box plot illustrating frequency of occurrence for restaurant categories across Merton, London, UK

We can observe that Chinese restaurants are very popular across Merton due to the highest Frequency. The chart has few outliers indicating that the data is not highly skewed.



I also created a data frame that tells us the popularity ratings (top 5) of 'All' restaurants across Merton, where we can observe as we did with the box plot that Chinese restaurants are the most popular overall.

We see that Chinese restaurants were the most popular areas except in St Helier and Raynes Park. In these 2 neighbourhoods, Chinese restaurants came second.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Colliers Wood	Chinese Restaurant	Thai Restaurant	Sushi Restaurant	Asian Restaurant	Korean Restaurant
1	Lower Morden	Chinese Restaurant	Sushi Restaurant	Vietnamese Restaurant	Thai Restaurant	Ramen Restaurant
2	Merton Park	Chinese Restaurant	Thai Restaurant	Japanese Restaurant	Sushi Restaurant	Asian Restaurant
3	Mitcham	Chinese Restaurant	Vietnamese Restaurant	Thai Restaurant	Sushi Restaurant	Ramen Restaurant
4	Morden	Chinese Restaurant	Thai Restaurant	Sushi Restaurant	Japanese Restaurant	Korean Restaurant
5	Morden Park	Chinese Restaurant	Thai Restaurant	Sushi Restaurant	Japanese Restaurant	Korean Restaurant
6	Raynes Park	Japanese Restaurant	Chinese Restaurant	Vietnamese Restaurant	Thai Restaurant	Sushi Restaurant
7	South Wimbledon	Chinese Restaurant	Thai Restaurant	Sushi Restaurant	Korean Restaurant	Asian Restaurant
8	St Helier	Japanese Restaurant	Chinese Restaurant	Vietnamese Restaurant	Thai Restaurant	Sushi Restaurant
9	Wimbledon	Chinese Restaurant	Thai Restaurant	Sushi Restaurant	Korean Restaurant	Asian Restaurant

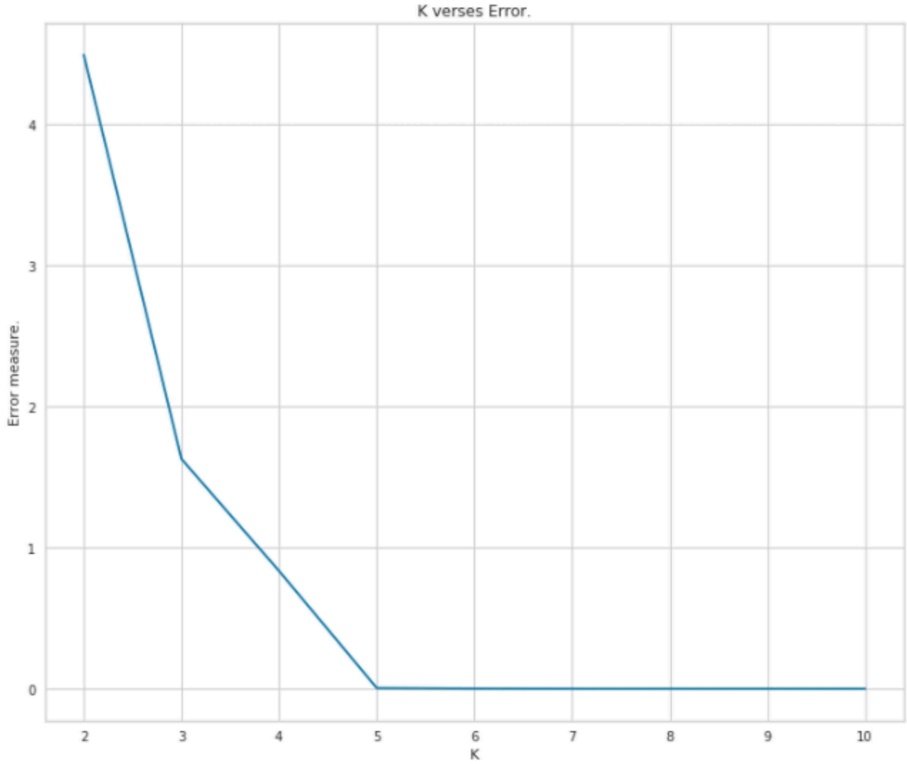
Afterwards, I created a data frame consisting the neighbourhood names and the occurrence frequency of Chinese restaurants across Merton.

	Neighborhood	Chinese Restaurant
0	Colliers Wood	0.424242
1	Lower Morden	0.800000
2	Merton Park	0.233333
3	Mitcham	1.000000
4	Morden	0.361111
5	Morden Park	0.361111
6	Raynes Park	0.500000
7	South Wimbledon	0.322581
8	St Helier	0.500000
9	Wimbledon	0.322581

CLUSTERING.

I used K-Means clustering as our method of choice for unsupervised machine learning. Here I clustered the neighbourhoods based on those that had similar averages of Chinese restaurants. In order to obtain an appropriate K value that had a better suitability fit for the model, I decided to use the **Elbow point** method. To do this, I ran a test with different numbers of K values and measured their accuracy, then chose the better K value. The better K value is indicated on the Elbow point chart at the point where there is the sharpest turn.

According to our chart this point is 5.



Now that we have determined our optimum value K as 5, I proceeded to install the Yellowbrick package and then imported the KElbowVisualizer. I then fit the K-Means model with the Elbow visualizer as illustrated below.

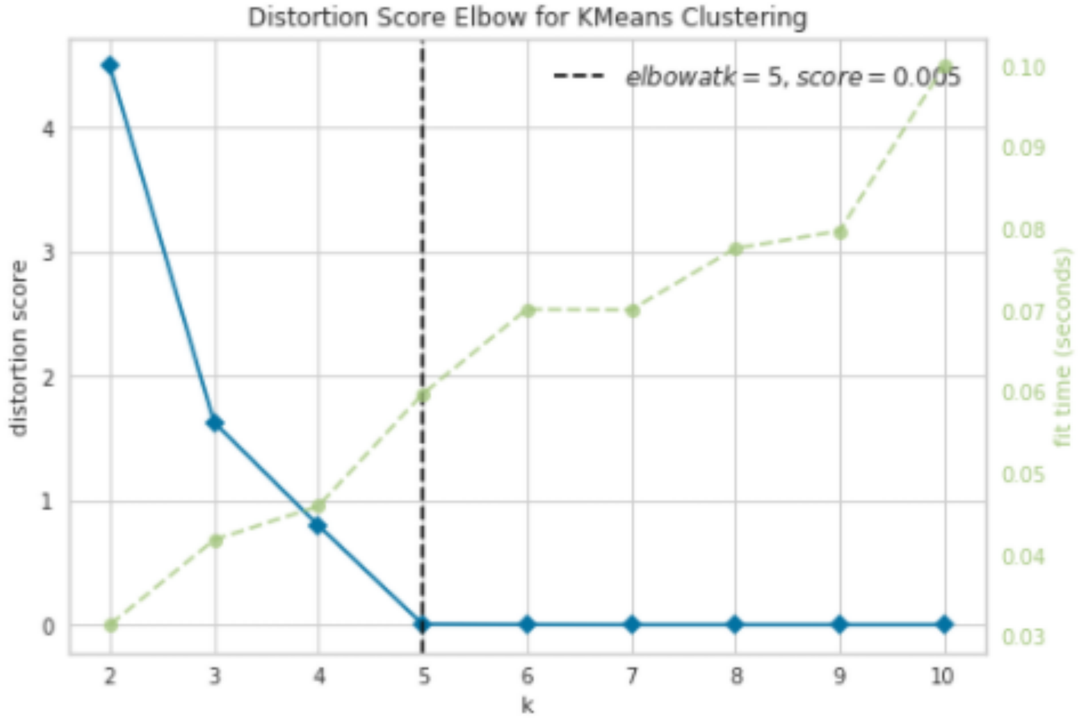
```
!pip install yellowbrick
from yellowbrick.cluster import KElbowVisualizer

# Instantiate the clustering model and visualizer

model = KMeans()
visualizer = KElbowVisualizer(model, k=(2,11))

visualizer.fit(X)
visualizer.show()
```

We then obtained the distortion score results as indicated in the diagram below. On the dotted line, we can see that K = 5 and as with clustering, items with similar features are grouped into similar clusters so locations/neighbourhoods with similar frequencies of Chinese restaurants within them were grouped into 5 clusters, starting from 0 to 4.



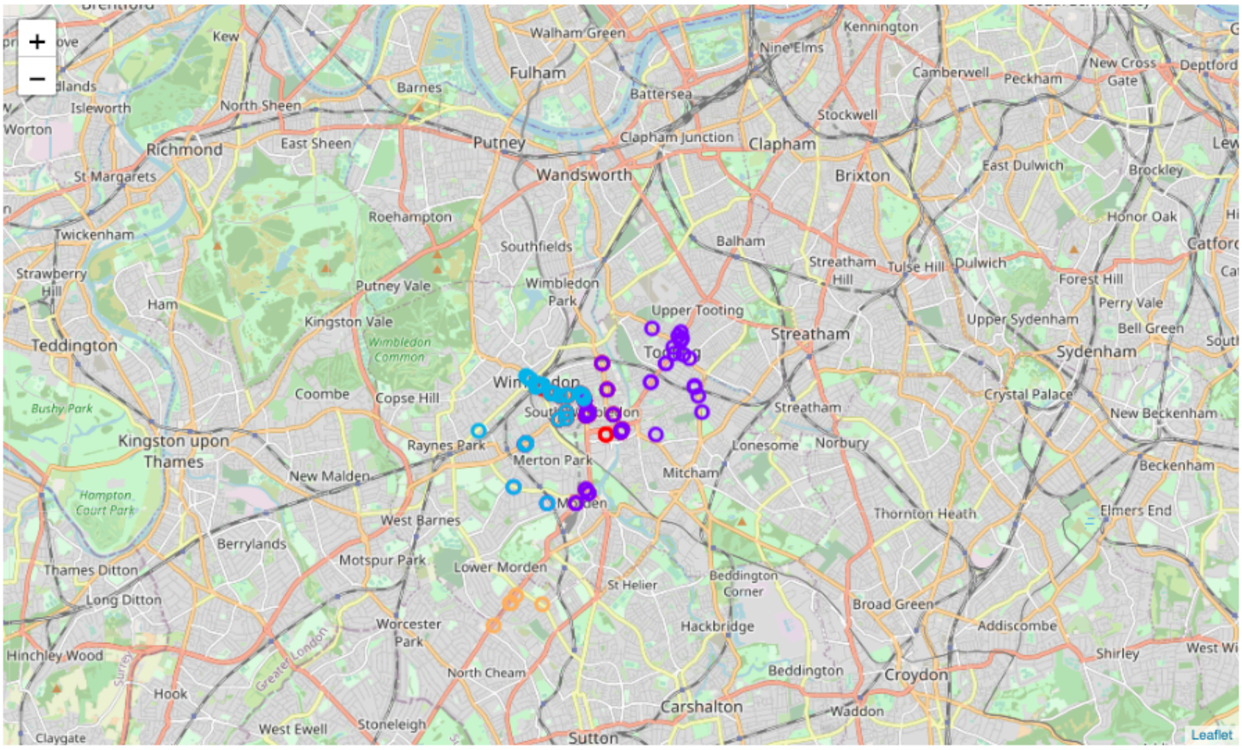
Hereafter, I proceeded to apply K-Means clustering on the Chinese restaurant frequencies data frame. I used the optimum value K of 5.

	Neighborhood	Chinese Restaurant	Cluster Labels
0	Colliers Wood	0.424242	2
1	Lower Morden	0.800000	1
2	Merton Park	0.233333	4
3	Mitcham	1.000000	3
4	Morden	0.361111	0
5	Morden Park	0.361111	0
6	Raynes Park	0.500000	2
7	South Wimbledon	0.322581	0
8	St Helier	0.500000	2
9	Wimbledon	0.322581	0

We then merged the above data frame with the venue data frame consisting all venues for all restaurant category types across Merton.

Map of each restaurant coloured based on the cluster label i.e. 5 clusters hence 5 colours.

It shows the clusters which had a similar mean/average frequency of Chinese restaurants.

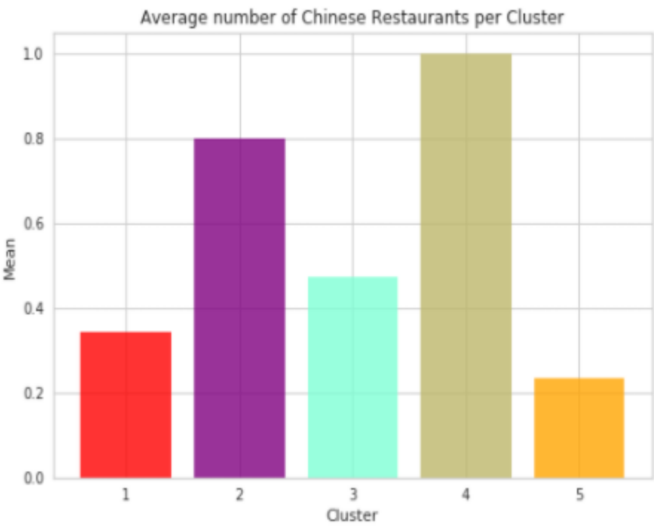


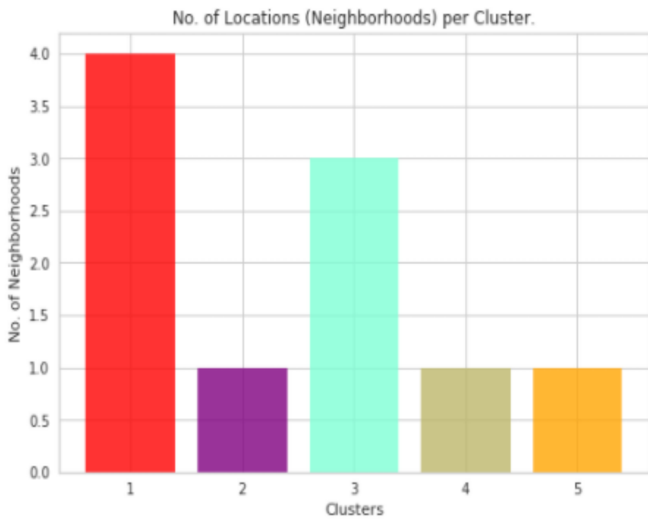
Clusters Analysis.

There are 5 clusters ranging 0,1,2,3 and 4.

Below we visualise how many Chinese restaurants are in each cluster and also how many neighbourhoods are within the clusters. Clusters 0, 1, 2, 3 and 4 are coloured red, purple, aquamarine, dark khaki and orange respectively. Clusters 1, 3 and 5 have the least neighbourhoods while 0 has the most. Also, Cluster 4 has the least number of Chinese restaurants while 3 and 1 have the most.

The information in the bar charts is very informative and crucial to business decisions as we can see that though clusters 1 and 3 have very few neighbourhoods within them, they have the highest number of Chinese restaurants. And though cluster 0 has the highest number of neighbourhoods overall, it has the second lowest number of Chinese restaurants within it.





Individual cluster analysis.

Cluster 1 has the neighbourhoods, Morden, Morden Park, South Wimbledon and Wimbledon.
These areas contain 50 Chinese restaurants. Morden and Merton Park has a higher number than the Wimbledon areas. These areas show promise for more Chinese restaurants.

Cluster 1

```
cluster1 = Chinese_merged.loc[Chinese_merged['Cluster Labels'] == 0]
cluster1_df = pd.merge(df_geo, cluster1, on='Neighborhood')
cluster1_df
```

	Neighborhood	Borough	Chinese Restaurant	Cluster Labels
0	Morden	Merton	0.361111	0
1	Morden Park	Merton	0.361111	0
2	South Wimbledon	Merton	0.322581	0
3	Wimbledon	Merton	0.322581	0

Cluster 2 contains only one neighbourhood i.e. Lower Morden. The frequency is also high at 0.8 meaning it has the highest concentration of Chinese restaurants. This indicates high demand but also high competition.

Cluster 2

```
cluster2 = Chinese_merged.loc[Chinese_merged['Cluster Labels'] == 1]
cluster2_df = pd.merge(df_geo, cluster2, on='Neighborhood')
cluster2_df
```

	Neighborhood	Borough	Chinese Restaurant	Cluster Labels
0	Lower Morden	Merton	0.8	1

Cluster 3 has 3 neighbourhoods within it with a moderate concentration of Chinese restaurants ranging 0.42 – 0.50. In this cluster Colliers Wood has the least number of Chinese restaurants while Raynes Park and St Helier feature a higher number of restaurants on average.

Cluster 3

```
cluster3 = Chinese_merged.loc[Chinese_merged['Cluster Labels'] == 2]
cluster3_df = pd.merge(df_geo, cluster3, on='Neighborhood')
cluster3_df
```

	Neighborhood	Borough	Chinese Restaurant	Cluster Labels
0	Colliers Wood	Merton	0.424242	2
1	Raynes Park	Merton	0.500000	2
2	St Helier	Merton	0.500000	2

Cluster 4 only has one neighbourhood, Mitcham. Within the clusters, it has the highest number of Chinese restaurants given its frequency score of 1. This indicates high demand in the area.

Cluster 4

```
cluster4 = Chinese_merged.loc[Chinese_merged['Cluster Labels'] == 3]
cluster4_df = pd.merge(df_geo, cluster4, on='Neighborhood')
cluster4_df
```

	Neighborhood	Borough	Chinese Restaurant	Cluster Labels
0	Mitcham	Merton	1.0	3

Cluster 5 also consists only one neighbourhood i.e. Merton Perk. This area has few Chinese restaurants which might indicate less demand.

Cluster 5

```
cluster5 = Chinese_merged.loc[Chinese_merged['Cluster Labels'] == 4]
cluster5_df = pd.merge(df_geo, cluster5, on='Neighborhood')
cluster5_df
```

	Neighborhood	Borough	Chinese Restaurant	Cluster Labels
0	Merton Park	Merton	0.233333	4

4: Discussion.

With the findings, we have found that the borough of Merton is indeed a good place of choice to open up restaurants, in fact 8 out of 10 of the areas have Chinese restaurants as the most popular within them, followed my Thai and Sushi restaurants. Merton boasts 10 neighbourhoods which consist a vast number of restaurant category types/cuisines. We found Mitcham has the highest concentration of Chinese restaurants in it with a frequency of 1 which is promising as Mitcham has a high population of about 70,000 inhabitants and a diverse group of people. The frequency score thus backs the population and diversity number. However, this can cause a problem as there might be a higher level of competition which would mean more effort needed to stand out and attract customers. Nonetheless, it is still a promising neighbourhood, but I would not recommend it unless a businessperson had a strong competitive advantage. The next promising neighbourhoods are in Morden i.e. Morden park and Lower Morden which had a frequency of 0.8. I would highly recommend Lower Morden as there are only about 4 Chinese restaurants meaning it is easier to get clients due to the less competition given the high demand in the area. A drawback however of these results is that they are entirely dependent on Foursquare's API which can lead to limitations. Another is that the analysis doesn't take into account certain factors such as demographics i.e. the Chinese population present within the neighbourhoods etc and also proximity to transport links like train stations where people prefer to get take-out meals after work which can raise competitive advantage in a restaurant's location placement. This concludes the analysis and recommendation of where's best in Merton, London to open a Chinese restaurant.

5: Conclusion.

Merton, London is a very popular residential borough in the United Kingdom and as seen with the results of our findings, it is an excellent place of choice to open up a Chinese restaurant and other restaurant types too. Present in Merton are a variety of cuisines in demand as reflected by the different types of restaurants which makes it an appealing place of business.

As with any solution there are always areas of improvement and with this case we could benefit further from more variety and recent information/data, such as that which could help us identify people's eating habits in an area as reflected by their credit card purchase history. Payment companies or banks, social networks such as Facebook and Instagram aggregate people's specific likes and shopping habits which would allow us to have a better pinpoint as to the demand level in an area. As more people are resorting to online shopping there are fewer 'physical' stores in an area so the count of restaurants in today's world might not accurately represent the demand level in an area. Also, there are newer machine learning algorithms and systems in place that can improve upon such an analysis.

Hopefully, with this report a businessperson can now have a good head start as to which areas are the most hopeful with regards to the business of successfully opening up a Chinese restaurant in Merton, London.