February 28, 2021

IBM'S DATA SCIENCE PROFESSIONAL CERTIFICATE

RESTAURANTS IN CENTRAL LONDON

AMUTHAN THIRUVALLUVAN

Table of Contents

1. Introduction

Keen interest and wish to develop my skills in Data Science had led me to pursue IBM's Data Science Professional Certification on Coursera: https://www.coursera.org/professional-certificates/ibm-data-science. During this course I was able to learn how to use Data Science tools such as Jupyter Notebook, GitHub and IBM Watson Studio. The main programming language used in this course was Python, which is packed with powerful libraries that can be utilised for Data Science such as Pandas, Numpy, Matplotlib, Seaborn, Folium, Scikit-learn and SciPy.

The final assignment in this course was called the "Capstone Project" where I was required to use the various tools and methodologies learned throughout this course to solve a real-life business problem. This business problem had to involve the use of location data derived from Foursquare (https://foursquare.com) using API.

The conditions set out to pass this assignment were to:

- Create a Jupyter notebook which contains the code used to conduct my analysis
- Submission of the Jupyter notebook on my Github repository
- A detailed report consisting of the business problem, methodology and findings.
- A blogpost or presentation of the analysis.

This is my full report, hope you enjoy it!

2. Business Problem

Central London is one of the world's most popular tourist destinations, attracting visitors from all over the world. It is a vibrant part of the city, packed with entertainment centres and great food. Furthermore, it is one of the most influential business centres in the world. This makes it an attractive location for businesses, especially restaurants.

The theoretical problem owner for this project is a well-established multiple restaurant chain owner from the suburbs of London who is looking to replicate his success by opening up a restaurant in Central London. The restaurant owner has a variety of restaurants covering different cuisines.

There are many things to consider for the restaurant owner before proceeding with this business venture e.g. costs, availability, supplies, staff etc. Let's assume the restaurant owner is happy with all the other aspects of this venture and is now left with the final problem which is to find the optimal location to open a restaurant in Central London.

2.1 Audience

This business problem is targeted at a group consisting of successful business owners (specifically restaurant owners) who wish to open a restaurant in Central London. Although, it can also be targeted at new business owners as long as there is enough capital available to open up in Central London, due to costs being at the premium side of the scale. The beauty of Data Science is that once a methodology is developed it can be applied to different variables of the same scenario quite easily. Therefore, this could be targeted at any business owner looking to open a restaurant almost anywhere in the world.

3. Data

The following list of data will be used to conduct this analysis:

- List of Post Districts in Central London
- Geo-coordinates of the districts in Central London
- Popular restaurants by categories in these districts

The list of Post Districts will be obtained from the following Wikipedia page: https://en.wikipedia.org/wiki/EC_postcode_area. Here, the data required is stored in a table called "List of postcode districts".

The Geo-coordinates will be calculated using the geocoder package within Python.

The popular restaurants will be gathered from Foursquare using API.

4. Methodology

4.1 High level summary

I will be using the post district data to find the geolocation for each district, then make API calls to Foursquare in order to get the surrounding venue details for each district. After cleaning this data into a usable format, I will run an unsupervised machine learning algorithm called *k*-means clustering to group the districts based on the restaurant types in these districts.

4.2 Post Districts Data

First step was to web scrape the Post Districts data from the Wikipedia page https://en.wikipedia.org/wiki/EC_postcode_area. To do this I used the Pandas library in Python which has a read html function. This function can be used to read in the data from a webpage and place it into a pandas dataframe. Once read in, I cleaned the data by removing unnecessary columns, relabelled the existing columns and removed any rows which did not have a District name associated with it.

The result is the following dataframe:

	Postal Code	City	District
0	EC1A	LONDON	St Bartholomew's Hospital
1	EC1M	LONDON	Clerkenwell, Farringdon
2	EC1N	LONDON	Hatton Garden
3	EC1R	LONDON	Finsbury, Finsbury Estate (west)
4	EC1V	LONDON	Finsbury (east), Moorfields Eye Hospital
5	EC1Y	LONDON	St Luke's, Bunhill Fields
6	EC2A	LONDON	Shoreditch
7	EC2M	LONDON	Broadgate, Liverpool Street
8	EC2N	LONDON	Old Broad Street, Tower 42
9	EC2R	LONDON	Bank of England

Only showing top 10 rows from the dataframe

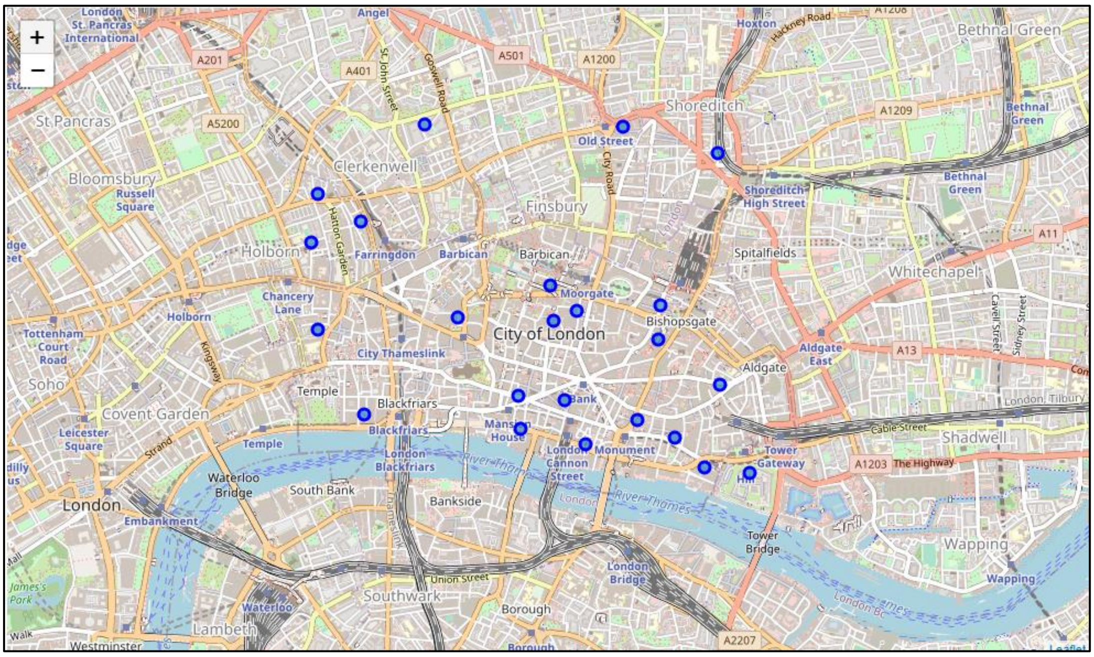
Once I had a clean dataframe, I was then able to use the geocoders (Nominatim) function which is a geolocation service from the geopy package (installation required) to loop through the Postal Codes to identify the Latitude and Longitude of each District. This information was then placed back into the dataframe.

The result is the following table:

	Postal Code	City	District	Latitude	Longitude
0	EC1A	LONDON	St Bartholomew's Hospital	51.516355	-0.099137
1	EC1M	LONDON	Clerkenwell, Farringdon	51.521011	-0.106675
2	EC1N	LONDON	Hatton Garden	51.520027	-0.110511
3	EC1R	LONDON	Finsbury, Finsbury Estate (west)	51.522350	-0.110057
4	EC1V	LONDON	Finsbury (east), Moorfields Eye Hospital	51.525715	-0.101704

Only showing top 5 rows from the table

Once I had the geolocation of each district, I could then plot these areas onto a map to confirm they are accurate. I used the Folium library to perform this visualisation:



4.3 Foursquare Data

4.3.1 Testing API call for one district

Once I had the geolocations of the districts, I used the data from Foursquare to view the venues situated within a close radius of each district. I tested the function for the first district on the table which is “St Bartholomew’s Hospital”. I chose 500 meters as the radius to perform the search as the districts in Central London are quite close to each other. The result from the Foursqare API call was a JSON file, which was inspected to create a function in Python to pull out the relevant information and place into a pandas dataframe.

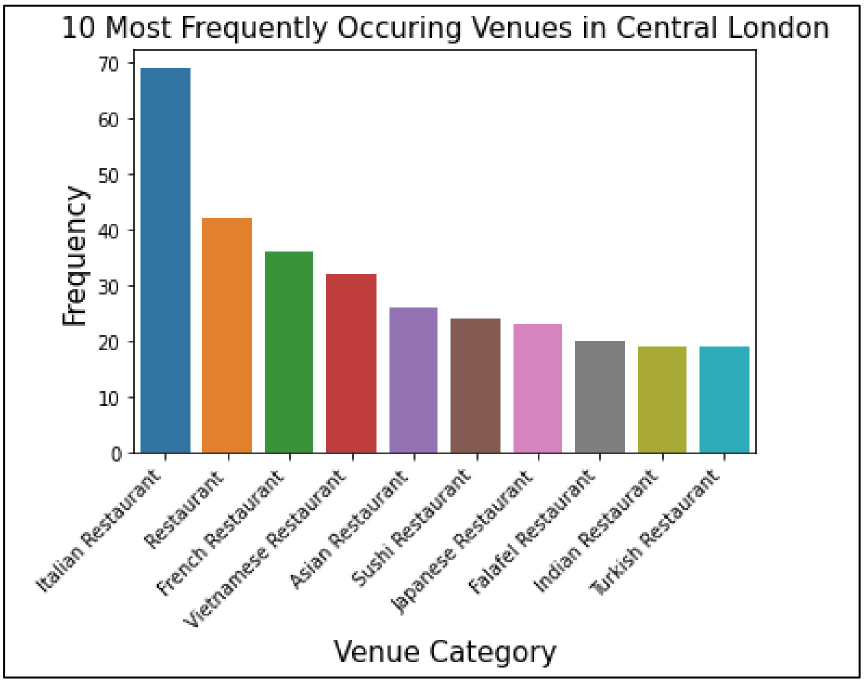
The result is the following dataframe which contains the details of the venues situated within a 500-meter radius of the St Bartholomew's Hospital district.

	name	categories	lat	lng
0	Pilpel	Falafel Restaurant	51.515195	-0.098462
1	Virgin Active	Gym / Fitness Center	51.518047	-0.097661
2	Postman's Park	Park	51.516860	-0.097643
3	Christ Church Greyfriars Garden	Garden	51.515670	-0.098760
4	St Bartholomew the Great (St Bartholomew-the-G...	Church	51.518631	-0.099890
5	Paternoster Square	Plaza	51.514572	-0.099226
6	Stationers' Hall	Event Space	51.514292	-0.101487
7	Museum of London	History Museum	51.518019	-0.096060
8	One New Change Rooftop	Roof Deck	51.513912	-0.095775
9	Paul	Bakery	51.514130	-0.099306

4.2.2 Extracting information for all districts

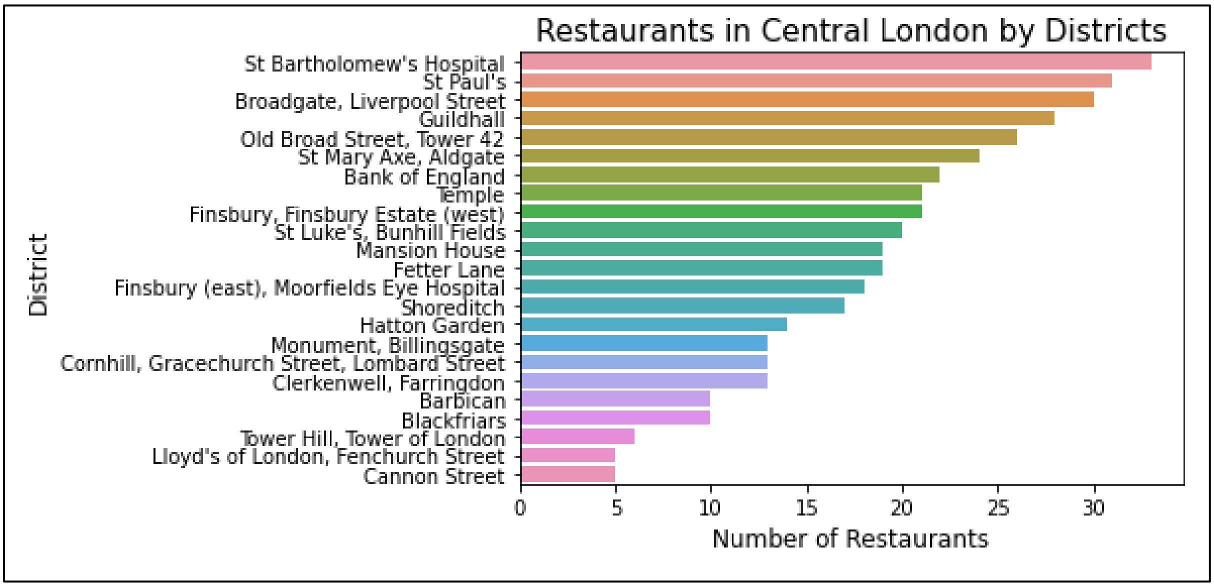
Once I was comfortable with the information being returned and I was able to define a function to extract the required details, and apply the same logic to perform the API call for each district. This brought back information on 1,723 venues in Central London. As I am only interested in restaurants, I extracted only the information related to restaurants, this refined the list to 418 venues of which there were 42 unique categories (restaurant type).

Analysis of the top 10 restaurant types in these districts using the seaborn and matplotlib libraries produced the following graph:



If we exclude “Restaurant” from the dataset as this could include any cuisine type, of the top 10 restaurant types in Central London, Italian restaurants equate to 24%, which is nearly twice as much as the 2nd most popular restaurant being French with 13%.

Once I removed “Restaurants” from the venue categories, I performed an overall frequency analysis using seaborn and matplotlib to compare the number of restaurants in each district.



4.2.3 Analysing each district

With the extracted information now in a pandas dataframe, I performed one-hot encoding method to convert the categorical values to binary vectors which is required for many machine learning models. Simply put, this transforms the dataframe by placing the unique restaurant types as column headers and the values as either 0 or 1 where 1 is Yes and 0 is No.

	District	Argentinian Restaurant	Asian Restaurant	Cantonese Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	English Restaurant	Falafel Restaurant	Fast Food Restaurant	French Restaurant	German Restaurant	Gı Restau
1	St Bartholomew's Hospital	0	0	0	0	0	0	0	1	0	0	0	
2	St Bartholomew's Hospital	0	0	0	0	0	0	0	0	0	1	0	
3	St Bartholomew's Hospital	0	0	0	0	0	0	0	0	0	0	0	
4	St Bartholomew's Hospital	0	0	0	0	0	0	0	1	0	0	0	
5	St Bartholomew's Hospital	0	0	0	0	0	0	0	0	0	0	0	

Only showing top 5 rows

After the dataframe had been one-hot encoded, the next step was to group the rows by districts in terms of the means of the frequency for each restaurant.

	District	Argentinian Restaurant	Asian Restaurant	Cantonese Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	English Restaurant	Falafel Restaurant	Fast Food Restaurant	French Restaurant	German Restaurant	G Restau
0	Bank of England	0.000000	0.090909	0.000000	0.000000	0.000000	0.045455	0.000000	0.000000	0.000000	0.090909	0.045455	0.00
1	Barbican	0.000000	0.100000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.100000	0.00
2	Blackfriars	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.100000	0.000000	0.00
3	Broadgate, Liverpool Street	0.033333	0.033333	0.033333	0.033333	0.000000	0.000000	0.066667	0.033333	0.000000	0.066667	0.033333	0.00
4	Cannon Street	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.20
5	Clerkenwell, Farringdon	0.000000	0.076923	0.000000	0.000000	0.000000	0.000000	0.076923	0.153846	0.000000	0.076923	0.000000	0.00
6	Cornhill, Gracechurch Street, Lombard Street	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.076923	0.000000	0.076923	0.153846	0.000000	0.00
7	Fetter Lane	0.052632	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.105263	0.052632	0.210526	0.000000	0.00
8	Finsbury (east), Moorfields Eye Hospital	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
9	Finsbury, Finsbury Estate (west)	0.000000	0.047619	0.000000	0.000000	0.000000	0.000000	0.000000	0.095238	0.047619	0.047619	0.000000	0.04

Only showing the top 10 rows

With this information I then created a dataframe consisting of the top 10 restaurants for each district:

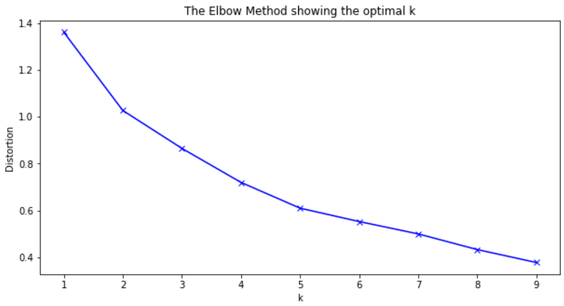
	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bank of England	Italian Restaurant	Seafood Restaurant	Japanese Restaurant	Asian Restaurant	Indian Restaurant	French Restaurant	Vietnamese Restaurant	German Restaurant	New American Restaurant	Scandinavian Restaurant
1	Barbican	Italian Restaurant	Vietnamese Restaurant	German Restaurant	Asian Restaurant	Turkish Restaurant	Vegetarian / Vegan Restaurant	Sushi Restaurant	Indian Restaurant	Korean Restaurant	Kebab Restaurant
2	Blackfriars	Italian Restaurant	Modern European Restaurant	Vietnamese Restaurant	French Restaurant	Turkish Restaurant	Japanese Restaurant	Seafood Restaurant	Korean Restaurant	Kebab Restaurant	Indian Restaurant
3	Broadgate, Liverpool Street	Indian Restaurant	Mediterranean Restaurant	Italian Restaurant	Middle Eastern Restaurant	English Restaurant	Sushi Restaurant	French Restaurant	Japanese Restaurant	Asian Restaurant	Cantonese Restaurant
4	Cannon Street	Italian Restaurant	Vietnamese Restaurant	Japanese Restaurant	Greek Restaurant	Malay Restaurant	Latin American Restaurant	Korean Restaurant	Kebab Restaurant	Indian Restaurant	German Restaurant

Only showing top 5 rows

4.2.4 K-means Clustering

I then ran an unsupervised machine learning algorithm known as the *k*-means clustering from the scikit-learn package to cluster the districts based on the most common restaurants.

First, I used the elbow method to calculate the optimal number of clusters to use.



From the graph, 5 was deemed as an appropriate number of clusters to group the districts by.

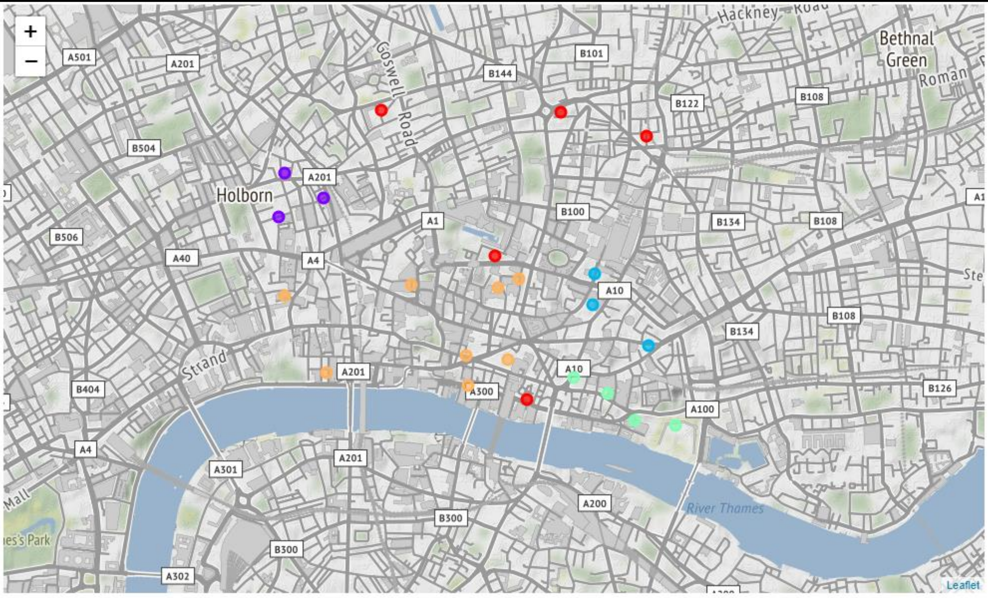
5. Results

Running this data through the *k*-means clustering algorithm with the k-value being 5 produces the following results. Each district had been classified with a particular cluster. Since there are 5 clusters, the Cluster Labels range is between 0 and 4.

	District	Latitude	Longitude	Cluster Labels	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
0	St Bartholomew's Hospital	51.516355	-0.099137	4	Italian Restaurant	Asian Restaurant	Modern European Restaurant	French Restaurant	English Restaurant	Japanese Restaurant	Sushi Restaurant	Falafel Restaurant	Korean Restaurant
1	Clerkenwell, Farringdon	51.521011	-0.106675	1	Vietnamese Restaurant	Falafel Restaurant	Asian Restaurant	Italian Restaurant	Middle Eastern Restaurant	Sushi Restaurant	Spanish Restaurant	English Restaurant	Modern European Restaurant
2	Hatton Garden	51.520027	-0.110511	1	Vietnamese Restaurant	Falafel Restaurant	Sushi Restaurant	Spanish Restaurant	Korean Restaurant	Middle Eastern Restaurant	Portuguese Restaurant	French Restaurant	Lebanese Restaurant
3	Finsbury, Finsbury Estate (west)	51.522350	-0.110057	1	Sushi Restaurant	Vietnamese Restaurant	Middle Eastern Restaurant	Spanish Restaurant	Falafel Restaurant	Portuguese Restaurant	Lebanese Restaurant	Japanese Restaurant	Italian Restaurant
4	Finsbury (east), Moorfields Eye Hospital	51.525715	-0.101704	0	Italian Restaurant	Vietnamese Restaurant	Sushi Restaurant	Japanese Restaurant	Mexican Restaurant	Middle Eastern Restaurant	Moroccan Restaurant	Seafood Restaurant	German Restaurant
5	St Luke's, Bunhill Fields	51.525630	-0.086289	0	Italian Restaurant	Vietnamese Restaurant	Turkish Restaurant	Japanese Restaurant	Ramen Restaurant	Scandinavian Restaurant	Asian Restaurant	Korean Restaurant	Middle Eastern Restaurant
6	Shoreditch	51.524365	-0.078885	0	Italian Restaurant	Vietnamese Restaurant	Japanese Restaurant	Peruvian Restaurant	Kebab Restaurant	Vegetarian / Vegan Restaurant	Middle Eastern Restaurant	Indian Restaurant	New American Restaurant
7	Broadgate, Liverpool Street	51.516950	-0.083340	2	Indian Restaurant	Mediterranean Restaurant	Italian Restaurant	Middle Eastern Restaurant	English Restaurant	Sushi Restaurant	French Restaurant	Japanese Restaurant	Asian Restaurant
8	Old Broad Street, Tower 42	51.515305	-0.083495	2	Turkish Restaurant	Italian Restaurant	Indian Restaurant	Sushi Restaurant	French Restaurant	English Restaurant	Argentinian Restaurant	Portuguese Restaurant	Latin American Restaurant
9	Bank of England	51.516675	-0.089874	4	Italian Restaurant	Seafood Restaurant	Japanese Restaurant	Asian Restaurant	Indian Restaurant	French Restaurant	Vietnamese Restaurant	German Restaurant	New American Restaurant

Only showing top 10 rows

This information can now be displayed on a folium map as the following:



We can see 23 markers on the map, one for each district and 5 different colours each reflecting a different cluster.

The following are the clusters and their 10 most common restaurants. Based on the top restaurants in each cluster, we can classify the type of cluster. This is noted under each cluster.

Cluster 1

```
CL_merged.loc[CL_merged['Cluster Labels'] == 0, CL_merged.columns[[0] + list(range(4, CL_merged.shape[1]))]]
```

	District	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
4	Finsbury (east), Moorfields Eye Hospital	Italian Restaurant	Vietnamese Restaurant	Sushi Restaurant	Japanese Restaurant	Mexican Restaurant	Middle Eastern Restaurant	Moroccan Restaurant	Seafood Restaurant	German Restaurant	Kebab Restaurant
5	St Luke's, Bunhill Fields	Italian Restaurant	Vietnamese Restaurant	Turkish Restaurant	Japanese Restaurant	Ramen Restaurant	Scandinavian Restaurant	Asian Restaurant	Korean Restaurant	Middle Eastern Restaurant	Peruvian Restaurant
6	Shoreditch	Italian Restaurant	Vietnamese Restaurant	Japanese Restaurant	Peruvian Restaurant	Kebab Restaurant	Vegetarian / Vegan Restaurant	Middle Eastern Restaurant	Indian Restaurant	New American Restaurant	Korean Restaurant
11	Barbican	Italian Restaurant	Vietnamese Restaurant	German Restaurant	Asian Restaurant	Turkish Restaurant	Vegetarian / Vegan Restaurant	Sushi Restaurant	Indian Restaurant	Korean Restaurant	Kebab Restaurant
20	Cannon Street	Italian Restaurant	Vietnamese Restaurant	Japanese Restaurant	Greek Restaurant	Malay Restaurant	Latin American Restaurant	Korean Restaurant	Kebab Restaurant	Indian Restaurant	German Restaurant

Italian and Thai cuisine dominant cluster

Cluster 2

```
CL_merged.loc[CL_merged['Cluster Labels'] == 1, CL_merged.columns[[0] + list(range(4, CL_merged.shape[1]))]]
```

	District	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
1	Clerkenwell, Farringdon	Vietnamese Restaurant	Falafel Restaurant	Asian Restaurant	Italian Restaurant	Middle Eastern Restaurant	Sushi Restaurant	Spanish Restaurant	English Restaurant	Modern European Restaurant	French Restaurant
2	Hatton Garden	Vietnamese Restaurant	Falafel Restaurant	Sushi Restaurant	Spanish Restaurant	Korean Restaurant	Middle Eastern Restaurant	Portuguese Restaurant	French Restaurant	Lebanese Restaurant	Cuban Restaurant
3	Finsbury, Finsbury Estate (west)	Sushi Restaurant	Vietnamese Restaurant	Middle Eastern Restaurant	Spanish Restaurant	Falafel Restaurant	Portuguese Restaurant	Lebanese Restaurant	Japanese Restaurant	Italian Restaurant	Mexican Restaurant

Thai and Middle Eastern

Cluster 3

```
CL_merged.loc[CL_merged['Cluster Labels'] == 2, CL_merged.columns[[0] + list(range(4, CL_merged.shape[1]))]]
```

	District	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
7	Broadgate, Liverpool Street	Indian Restaurant	Mediterranean Restaurant	Italian Restaurant	Middle Eastern Restaurant	English Restaurant	Sushi Restaurant	French Restaurant	Japanese Restaurant	Asian Restaurant	Cantonese Restaurant
8	Old Broad Street, Tower 42	Turkish Restaurant	Italian Restaurant	Indian Restaurant	Sushi Restaurant	French Restaurant	English Restaurant	Argentinian Restaurant	Portuguese Restaurant	Latin American Restaurant	Malay Restaurant
12	St Mary Axe, Aldgate	Argentinian Restaurant	Asian Restaurant	Turkish Restaurant	English Restaurant	French Restaurant	South American Restaurant	Italian Restaurant	Portuguese Restaurant	Cantonese Restaurant	Falafel Restaurant

Diverse cuisine cluster

Cluster 4

```
CL_merged.loc[CL_merged['Cluster Labels'] == 3, CL_merged.columns[[0] + list(range(4, CL_merged.shape[1]))]]
```

	District	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
13	Lloyd's of London, Fenchurch Street	French Restaurant	Asian Restaurant	Turkish Restaurant	Japanese Restaurant	South American Restaurant	Vietnamese Restaurant	Korean Restaurant	Kebab Restaurant	Italian Restaurant	Indian Restaurant
14	Tower Hill, Tower of London	French Restaurant	Asian Restaurant	Turkish Restaurant	Tapas Restaurant	Italian Restaurant	South American Restaurant	Vietnamese Restaurant	Korean Restaurant	Kebab Restaurant	Japanese Restaurant
15	Monument, Billingsgate	French Restaurant	Falafel Restaurant	Asian Restaurant	Turkish Restaurant	Italian Restaurant	English Restaurant	South American Restaurant	Indian Restaurant	Portuguese Restaurant	Vietnamese Restaurant
16	Cornhill, Gracechurch Street, Lombard Street	Turkish Restaurant	Italian Restaurant	French Restaurant	Latin American Restaurant	Tapas Restaurant	Japanese Restaurant	Indian Restaurant	South American Restaurant	English Restaurant	Fast Food Restaurant

French, Asian and Turkish cuisine dominant cluster

Cluster 5

```
CL_merged.loc[CL_merged['Cluster Labels'] == 4, CL_merged.columns[[0] + list(range(4, CL_merged.shape[1]))]]
```

	District	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	St Bartholomew's Hospital	Italian Restaurant	Asian Restaurant	Modern European Restaurant	French Restaurant	English Restaurant	Japanese Restaurant	Sushi Restaurant	Falafel Restaurant	Korean Restaurant	Indian Restaurant
9	Bank of England	Italian Restaurant	Seafood Restaurant	Japanese Restaurant	Asian Restaurant	Indian Restaurant	French Restaurant	Vietnamese Restaurant	German Restaurant	New American Restaurant	Scandinavian Restaurant
10	Guildhall	Vietnamese Restaurant	French Restaurant	Asian Restaurant	Italian Restaurant	Seafood Restaurant	Indian Restaurant	Modern European Restaurant	Sushi Restaurant	Japanese Restaurant	German Restaurant
17	Fetter Lane	Italian Restaurant	French Restaurant	Falafel Restaurant	Sushi Restaurant	Vietnamese Restaurant	Japanese Restaurant	Fast Food Restaurant	Indian Restaurant	Argentinian Restaurant	Turkish Restaurant
18	St Paul's	Italian Restaurant	Vietnamese Restaurant	Asian Restaurant	French Restaurant	Modern European Restaurant	Japanese Restaurant	Seafood Restaurant	Falafel Restaurant	Chinese Restaurant	Sushi Restaurant
19	Mansion House	Vietnamese Restaurant	Italian Restaurant	Sushi Restaurant	French Restaurant	Asian Restaurant	Turkish Restaurant	Japanese Restaurant	Dim Sum Restaurant	English Restaurant	Seafood Restaurant
21	Blackfriars	Italian Restaurant	Modern European Restaurant	Vietnamese Restaurant	French Restaurant	Turkish Restaurant	Japanese Restaurant	Seafood Restaurant	Korean Restaurant	Kebab Restaurant	Indian Restaurant
22	Temple	Italian Restaurant	Modern European Restaurant	Asian Restaurant	Falafel Restaurant	Vietnamese Restaurant	Fast Food Restaurant	French Restaurant	Indian Restaurant	Korean Restaurant	Argentinian Restaurant

Italian and diverse cuisine cluster

6. Discussion

6.1 Interpretations

This unsupervised machine learning model highlights how dominant the Italian and Thai cuisine restaurants are in Central London. Those who are familiar with Central London would agree with these results. The districts where Italian restaurants aren't dominant tends to be on the outer circle.

The purpose of this analysis was to find the best location for the restaurant owner to open in Central London. From the analysis conducted, we can say that the large proportion of Italian and Thai restaurant would suggest this being a popular choice for customers in Central London. However, if the restaurant owner was to open up an Italian restaurant in the Italian dominant districts, this would result in heavy competition. It would have to be an extremely well-established chain to beat off the competition. What we could advise from this analysis is potentially exploring the option to open in a district from one of the clusters where Italian restaurants are not as common. Likewise, the same logic can be applied when considering other cuisine types.

6.2 Limitations

We have only considered the frequency of restaurants within districts for our analysis to avoid opening in a highly competitive area. However, for a wider analysis we need to consider why certain cuisine types are so dominant in certain districts? The location data we have used is from Foursquare, although a widely used source it could be incomplete e.g. we had exclude 42 restaurants from the analysis due to this not being labelled with a cuisine type. Therefore, it would be prudent to consider using other location data in conjunction for a more complete analysis. In addition, the account we used to make the API calls to Foursquare was a basic free membership account, which limits the number of API calls we can make. A paid subscription can provide a more detailed result.

7. Conclusion

Taking the limitations into consideration we have achieved the goal set out to advise on the optimal location to open a restaurant in Central London for a certain cuisine type. We have done this using readily available open-source information and tools as highlighted in the Introduction section.

What is amazing about the field of data science is the amount of free support that is available from the data science community. Because of the sheer volume of tools and libraries available for data science, it is almost impossible to fully understand how to use them all. If you are curious about data science and know how to apply the core library packages (most of those listed in this report), there are plenty of comprehensive documents and support available to help.