IBM Data Science Professional Certificate

Battle of the Neighbourhoods: Where in South West London should a restaurant owner open a new healthy eating restaurant?

Rebecca Williams

Business Problem

A healthy eating restaurant chain owner has proposed a plan to open a new restaurant in South West London.

The company saw an opportunity for the restaurant in South West London as it has so far been successful in various other areas in London.

The task the company has put forward is to find out which Neighbourhood in South West London would be the best place to open the restaurant.

The **objective** of this project is to find the best area in South West London for the restaurant to be located.

Data

| Data Source 1 | Data Source 2 | Data Source 3 |

London Location Data

https://en.wikipedia.org/wiki/List_of_areas_of_London

Extracted using the BeautifulSoup package

Latitude and Longitude Data

Generated using Python Geocoder

Needed for finding nearby venues and plotting maps

London Venue Data

The Foursquare API will be used to get the geographical data of the venues in South West London

Resulting Data Frame

	Neighbourhood	Borough	PostTown	PostCode
0	Abbey Wood	Bexley, Greenwich	LONDON	SE2
1	Acton	Ealing, Hammersmith and Fulham	LONDON	W3
1	Acton	Ealing, Hammersmith and Fulham	LONDON	W4
2	Addington	Croydon	CROYDON	CR0
3	Addiscombe	Croydon	CROYDON	CR0

Data Cleaning

- Drop columns 'DialCode' and 'GridRef'
- Split Neighbourhoods with 2 post code values into separate rows
- Keep only South West postcodes
- The resulting Data Frame is shown on the right

	Neighbourhood	Borough	PostTown	PostCode
0	Balham	Wandsworth	LONDON	SW12
1	Barnes	Richmond upon Thames	LONDON	SW13
2	Battersea	Wandsworth	LONDON	SW11
3	Belgravia	Westminster	LONDON	SW1
4	Brixton	Lambeth	LONDON	SW2

Latitude and Longitude

```python
def get_latlong(arcgis_geocoder):

    lat_long = None

    while(lat_long is None):
        g = geocoder.arcgis('{}, London, United Kingdom'.format(arcgis_geocoder))
        lat_long = g.latlng
    return lat_long
```
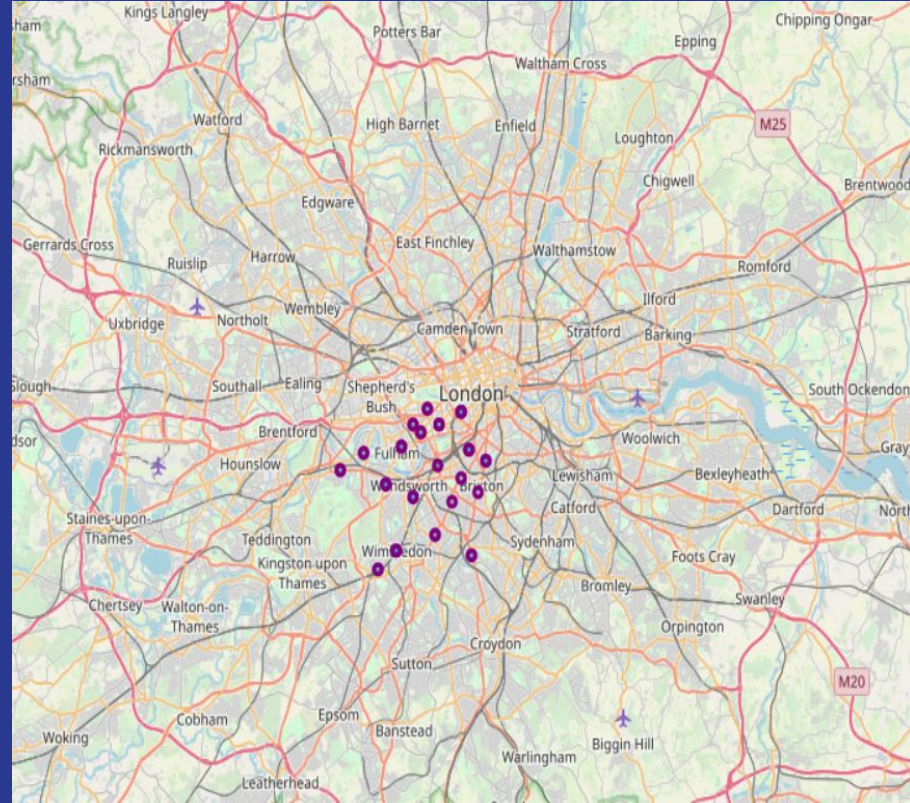
- Python Geocoder package with arcgis_geocoder was used to get the lat/long of each Neighbourhood

- A function, shown above right, was defined and applied to each Neighbourhood

- The resulting Data Frame is shown right

	Neighbourhood	Borough	PostTown	PostCode	Latitude	Longitude
0	Balham	Wandsworth	LONDON	SW12	51.44822	-0.14839
1	Barnes	Richmond upon Thames	LONDON	SW13	51.47457	-0.24212
2	Battersea	Wandsworth	LONDON	SW11	51.46760	-0.16290
3	Belgravia	Westminster	LONDON	SW1	51.49713	-0.13829
4	Brixton	Lambeth	LONDON	SW2	51.45295	-0.12083

Folium Map

A map of the Neighbourhoods in South West London generated using the Foium package

Methodology

Foursquare Location Data

- The Foursquare API will be used to get the geographical data of the venues in South West London.
- Foursquare holds a database 105 million places, and provides various different categories of venue data.
- A function is generated to extract the category types of each venue and the newly obtained data is put into a data frame.

The venues data set contains 2539 venues.

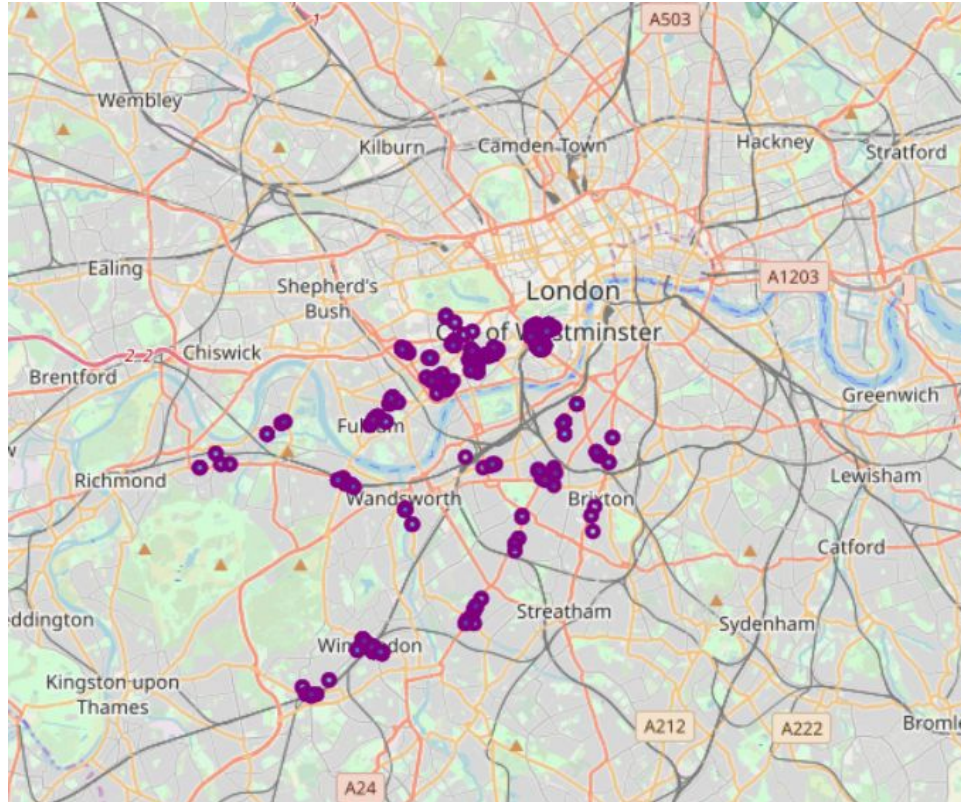|	Neighborhood	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
0	Balham	51.44822	-0.14839	Ciullosteria	51.447144	-0.148981	Italian Restaurant
1	Balham	51.44822	-0.14839	M1LK	51.444450	-0.150913	Coffee Shop
2	Balham	51.44822	-0.14839	We Brought Beer	51.444324	-0.150656	Beer Store
3	Balham	51.44822	-0.14839	Brickwood Coffee & Bread	51.444509	-0.151127	Coffee Shop
4	Balham	51.44822	-0.14839	The Grove	51.448286	-0.150964	Pub

Unique Categories

- Coffee Shop is the most common venue in South West London with 174

- Italian and Indian restaurants have a high count so this needs to be considered when deciding on a location for the healthy eating restaurant

	Count
Coffee Shop	174
Pub	153
Hotel	124
Café	112
Italian Restaurant	85
Sandwich Place	67
Indian Restaurant	67
Bar	62
Gym / Fitness Center	62
Bakery	53

Restaurant Venue Data Frame

	Neighborhood	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
0	Balham	51.44822	-0.14839	Ciullosteria	51.447144	-0.148981	Italian Restaurant
9	Balham	51.44822	-0.14839	The Georgian	51.452242	-0.147677	Caucasian Restaurant
16	Balham	51.44822	-0.14839	Holy Cow	51.447371	-0.149012	Indian Restaurant
17	Balham	51.44822	-0.14839	The Kebab Company	51.445773	-0.150407	Fast Food Restaurant
20	Balham	51.44822	-0.14839	Nando's	51.445837	-0.150603	Portuguese Restaurant
29	Balham	51.44822	-0.14839	Megan's On The Hill	51.443974	-0.150208	Restaurant
31	Balham	51.44822	-0.14839	Chatkhara	51.452637	-0.147465	Indian Restaurant
33	Balham	51.44822	-0.14839	Chicken Cottage	51.452127	-0.147835	Fast Food Restaurant
37	Barnes	51.47457	-0.24212	Awesome Thai Cuisine	51.474905	-0.240909	Thai Restaurant
45	Barnes	51.47457	-0.24212	Côte Brasserie	51.472424	-0.246774	French Restaurant

Folium Map of Restaurant Venues in SW London

One Hot Encoding

- Used to turn the restaurant categories into numerical data
- In the extract below, none of the locations have any of the shown venues

	Neighborhood	American Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant
0	Balham	0	0	0	0	0	0
1	Balham	0	0	0	0	0	0
2	Balham	0	0	0	0	0	0
3	Balham	0	0	0	0	0	0
4	Balham	0	0	0	0	0	0

Extract of Mean of Each Venue

	Neighborhood	American Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Australian Restaurant	Auto Garage	BBQ Joint	Bagel Shop	Bakery
0	Balham	0.0	0.0	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.057143
1	Barnes	0.0	0.0	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.045455
2	Battersea	0.0	0.0	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.000000
3	Belgravia	0.0	0.0	0.010101	0.0	0.010101	0.0	0.010101	0.0	0.0	0.0	0.010101
4	Brixton	0.0	0.0	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.038462

Top 4 Most Common Venue

The top 4 most common venues of Chelsea, Clapham, Barnes and Battersea are shown on the right

```
----Chelsea----
                   venue  freq
0                   Café  0.06
1                    Pub  0.05
2                  Hotel  0.05
3   Japanese Restaurant   0.05


----Clapham----
             venue  freq
0              Pub  0.08
1             Café  0.06
2     Burger Joint  0.05
3       Restaurant  0.04
```

```
----Barnes----
                    venue  freq
0             Pizza Place  0.09
1          Farmers Market  0.09
2                    Park  0.09
3       Food & Drink Shop  0.09


----Battersea----
                    venue  freq
0                    Café  0.12
1                     Pub  0.08
2       Indian Restaurant  0.08
3             Supermarket  0.08
```

Top 10 Most Common Venues in Each Neighbourhood

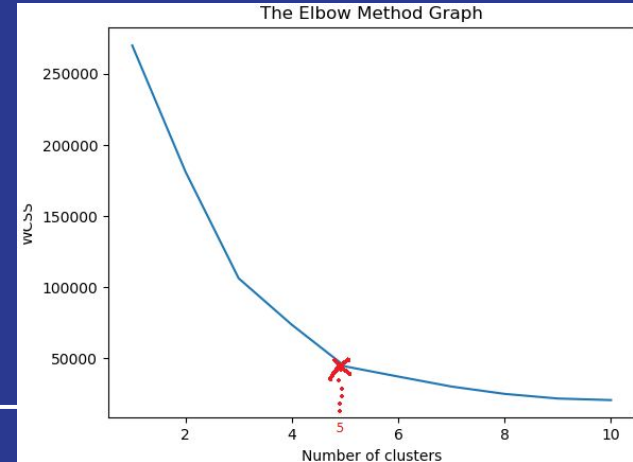|	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Balham	Coffee Shop	Grocery Store	Pub	Bakery	Indian Restaurant
1	Barnes	Food & Drink Shop	Farmers Market	Pub	Pizza Place	Park
2	Battersea	Café	Bar	Supermarket	Pub	Indian Restaurant
3	Belgravia	Coffee Shop	Hotel	Sandwich Place	Theater	Restaurant
4	Brixton	Pub	Coffee Shop	Indian Restaurant	Caribbean Restaurant	Portuguese Restaurant

K-Means Clustering

- A task of identifying subgroups in the data
- Data points in the same subgroup (cluster) are very similar
- Data points in different clusters are very different
- An unsupervised method that identifies k number of centroids
- Centroids are kept as small as possible
- The graph shows K=5

The Elbow Method

Used to determine the optimal number of clusters into which the data may be clustered.

To determine the optimal number of clusters, we have to select the value of k at the "elbow" ie the point after which the distortion/inertia start decreasing in a linear fashion.
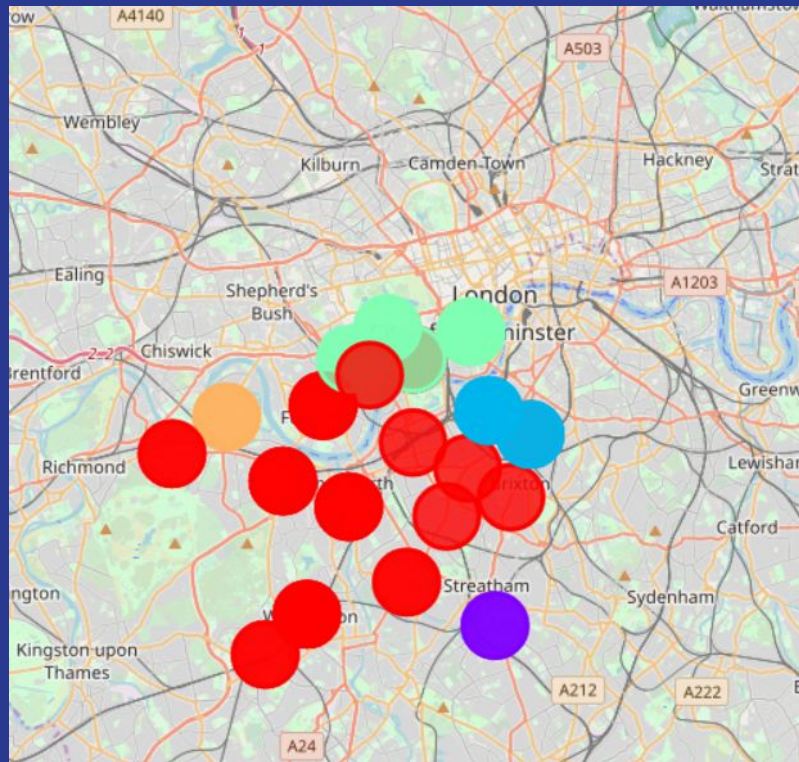

The Elbow Method Graph

Cluster Labels

- The number of clusters is set to 5
- New data frame of top 10 venues and cluster labels is created
- Array of K-Means labels:

```
array([0, 4, 0, 3, 0, 0, 4, 0, 0, 0, 3, 0, 0, 0, 3, 0, 3, 0, 3, 0, 2, 1,
       2, 0, 3, 0, 0, 0, 0, 3, 0, 0, 3, 2, 1, 0, 0, 2, 0, 0, 3, 0],
      dtype=int32)
```

Results

Cluster 1

- Most common venues are Coffee Shop, Pub, Cafe, Bar
- Restaurants appear in the 3rd column
- Indian restaurants are 3rd most common, followed by Caribbean and Portuguese
- Kensington and Chelsea have Japanese as 3rd most common

Cropped extract of Cluster 1

	Borough	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Cluster Labels	4th Most Common Venue	
0	Wandsworth	-0.14839	Coffee Shop	Grocery Store	Pub	0	Bakery	
2	Wandsworth	-0.16290	Café	Bar	Supermarket	0	Pub	
4	Lambeth	-0.12083	Pub	Coffee Shop	Indian Restaurant	0	Caribbean Restaurant	
5	Lambeth	-0.11249	Pub	Coffee Shop	Indian Restaurant	0	Caribbean Restaurant	
6	Kensington and Chelsea	-0.16248	Café	Pub	Hotel	0	Japanese Restaurant	C
8	Kensington and Chelsea	-0.16248	Café	Pub	Hotel	0	Japanese Restaurant	C
9	Lambeth, Wandsworth	-0.13922	Pub	Café	Burger Joint	0	Bar	G
10	Merton	-0.20796	Bar	Coffee Shop	Pub	0	Sushi Restaurant	C
13	Wandsworth	-0.18997	Coffee Shop	Pub	Café	0	Supermarket	
14	Richmond upon Thames	-0.26591	Coffee Shop	Grocery Store	Pub	0	Pizza Place	

Results

Cluster 2

- Most common venue is Pub
- Followed by Bar and Pizza Place
- This cluster also includes a Burger Joint and Grilled Meat Restaurant

	Borough	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Cluster Labels	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
23	Croydon	-0.12753	Pub	Bar	Pizza Place	1	Park	Diner	Burger Joint	Supermarket	Playground	Video Game Store	Grilled Meat Restaurant
41	Lambeth	-0.12753	Pub	Bar	Pizza Place	1	Park	Diner	Burger Joint	Supermarket	Playground	Video Game Store	Grilled Meat Restaurant

Results

Cluster 3

- Pub is msot common venue
- Portuguese restaurant is the 2nd most common
- Gym and Indian restaurant is 3rd most common
- Cluster also includes Park in the top ten, and gyms in different Neighbourhoods in the top 10

	Borough	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Cluster Labels	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
22	Wandsworth	-0.13056	Pub	Portuguese Restaurant	Gym / Fitness Center	2	Fish & Chips Shop	Indian Restaurant	Pizza Place	Performing Arts Venue	Bakery	Coffee Shop	Bus Stop
24	Lambeth	-0.13056	Pub	Portuguese Restaurant	Indian Restaurant	2	Bakery	Park	Gym / Fitness Center	Pizza Place	Performing Arts Venue	Restaurant	Caribbean Restaurant
25	Lambeth	-0.11249	Pub	Portuguese Restaurant	Indian Restaurant	2	Bakery	Park	Gym / Fitness Center	Pizza Place	Performing Arts Venue	Restaurant	Caribbean Restaurant
39	Lambeth	-0.13056	Pub	Portuguese Restaurant	Indian Restaurant	2	Bakery	Park	Gym / Fitness Center	Pizza Place	Performing Arts Venue	Restaurant	Caribbean Restaurant
40	Lambeth	-0.11249	Pub	Portuguese Restaurant	Indian Restaurant	2	Bakery	Park	Gym / Fitness Center	Pizza Place	Performing Arts Venue	Restaurant	Caribbean Restaurant
44	Lambeth	-0.13056	Pub	Portuguese Restaurant	Gym / Fitness Center	2	Fish & Chips Shop	Indian Restaurant	Pizza Place	Performing Arts Venue	Bakery	Coffee Shop	Bus Stop

Results

Cluster 4

- Most common venues are Coffee Shops, Hotels and Cafes
- Followed by Hotels, Cafes an Exhibit and Italian Restaurant
- Also Sandwich Places, Gyms, Juice Bars
- Top restaurants are Italian, Indian, Japanese and Chinese

Cropped Extract of Cluster 4

	Borough	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Cluster Labels	4th Most Common Venue
3	Westminster	-0.13829	Coffee Shop	Hotel	Sandwich Place	3	Theater
11	Kensington and Chelsea	-0.18971	Hotel	Italian Restaurant	Garden	3	Pub
12	Kensington and Chelsea	-0.18144	Hotel	Italian Restaurant	Garden	3	Pub
16	Kensington and Chelsea	-0.17404	Café	Exhibit	Science Museum	3	Hotel
18	Westminster	-0.13829	Coffee Shop	Hotel	Sandwich Place	3	Theater
20	Westminster	-0.13829	Coffee Shop	Hotel	Sandwich Place	3	Theater
27	Westminster	-0.13829	Coffee Shop	Hotel	Sandwich Place	3	Theater
32	Kensington and Chelsea	-0.17404	Hotel	Café	Exhibit	3	Pub
33	Kensington and Chelsea	-0.16248	Hotel	Café	Exhibit	3	Pub

Results

Cluster 5
- Food and Drink Shop is most common venue
- Followed by Farmers Market
- Most common restaurants are Pizza Place, French Restaurant and Thai Restaurant
- This cluster also has Movie Theater as the 9th most common venue

	Borough	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Cluster Labels	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	Richmond upon Thames	-0.24212	Food & Drink Shop	Farmers Market	Pub	4	Pizza Place	Park	Bakery	French Restaurant	Thai Restaurant	Movie Theater	Gastropub
7	Richmond upon Thames	-0.24212	Food & Drink Shop	Farmers Market	Pub	4	Pizza Place	Park	Bakery	French Restaurant	Thai Restaurant	Movie Theater	Gastropub

Conclusion

- The objective of this project was to find the best area in South West London to open a new healthy eating restaurant
- Cluster 3 has been decided for the location of the new restaurant
- There are a variety of venues promoting a healthier lifestyle
- Different types of restaurant venues, healthy eating establishment will make a nice addition
- Restaurant to be located in Lambeth in proximity to the gym/fitness centres and parks
- This will hopefully attract those who are looking to make healthier lifestyle choices