Opening a Restaurant within Croydon

Introduction/Business Problem

London[1] is the capital of the United Kingdom. It is one of the largest financial hubs in the world. With a population of over 9 million people and being a very diverse city, large and small businesses thrive in different parts of the city. London is divided into 32 London boroughs and the City of London.

In this study we would be exploring areas within one of the London Borough of Croydon, which is the largest London borough by population.

As the London Borough of Croydon has an extensive shopping district and night-time economy, restaurant owners and investors can benefit from this analysis to help determine what kind of restaurant to open to maximise profit.

It is important for restaurant owners to find a location that has a continuous stream of traffic, convenient parking, and is in proximity to other businesses.

A map will be created to show what kinds of restaurants are popular in the areas within Croydon. Based on our analysis we can determine the restaurant type and style of service that would be best suited for each area so that investors can make targeted decisions.

Data

The data below would be used for our analysis

- **List of London Boroughs (Wikipedia)[2] -** This holds the coordinates of all the boroughs and local authorities within London
- **List of areas of London (Wikipedia[3] and Wikimedia[4])** - This holds a list of areas in London with their coordinates; It would be cleaned and reduced to areas within the London Borough of Croydon and shown on a map
- **Foursquare API[5]** will be used to get the most common venues within all the areas of interest.

Each area will be analyzed for unique venue categories.
Venues will be clustered and displayed on a map. From our analysis we can determine what kinds of restaurants would thrive in different areas of Croydon

---

[1] https://en.wikipedia.org/wiki/London

[2] https://en.wikipedia.org/wiki/London_Borough_of_Croydon

[3] https://en.wikipedia.org/wiki/List_of_areas_of_London

[4] https://tools.wmflabs.org/kmlexport?article=Category%3AAreas+of+London

[5] https://developer.foursquare.com

Methodology

Data Cleaning

Data from the List of London Boroughs was used to determine the central location of London borough of Croydon.

Using the Beautiful Soup Library, data from the List of areas of London was cleaned and reduced to areas within Croydon.

	Location	London_borough	Post_town	Postcode_district	Dial_code	OS_grid_ref
0	Addington	Croydon	CROYDON	CR0	020	TQ375645
1	Addiscombe	Croydon	CROYDON	CR0	020	TQ345665
2	Coombe	Croydon	CROYDON	CR0	020	TQ342647
3	Coulsdon	Croydon	COULSDON	CR5	020, 01737	TQ298596
4	Croydon	Croydon	CROYDON	CR0	020	TQ335655

Location Data[6] for these areas where derived from saved GPS coordinates on Wikimedia.
The Location data was saved to Github. Some areas' coordinates (South Norwood, Woodside, Coombe) had to be manually sourced from Google Maps.

	Location	Latitude	Longtitude
0	Abbey Wood	51.4864	0.1109
1	Acton	51.513519	-0.270661
2	Acton Green	51.510515	-0.262668
3	Acton Vale	51.511	-0.258
4	Addington	51.3583	-0.0305

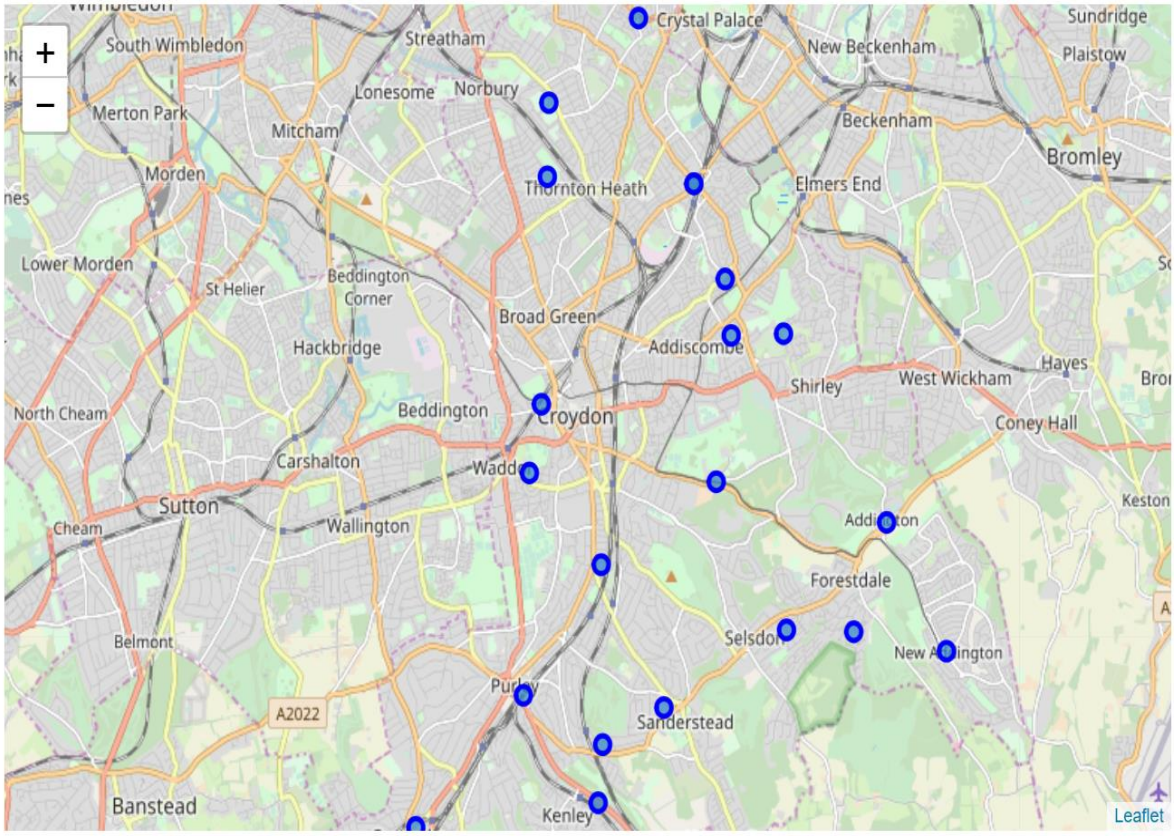
Beautiful Soup was also used to clean the data and retrieve coordinates for all the areas. The location data and the areas data were merged

	Location	London_borough	Post_town	Postcode_district	Dial_code	OS_grid_ref	Latitude	Longtitude
0	Addington	Croydon	CROYDON	CR0	020	TQ375645	51.3583	-0.0305
1	Addiscombe	Croydon	CROYDON	CR0	020	TQ345665	51.381	-0.0663
2	Coombe	Croydon	CROYDON	CR0	020	TQ342647	NaN	NaN
3	Coulsdon	Croydon	COULSDON	CR5	020, 01737	TQ298596	51.3211	-0.1386
4	Croydon	Croydon	CROYDON	CR0	020	TQ335655	51.3727	-0.1099

[6] https://raw.githubusercontent.com/aniebiet/Coursera_Capstone/master/Capstone_Project/areas.gpx

Visualizing the areas

Using Folium library, I visualized geographic details of the areas within the London Borough of Croydon by super imposing the coordinates on a map.



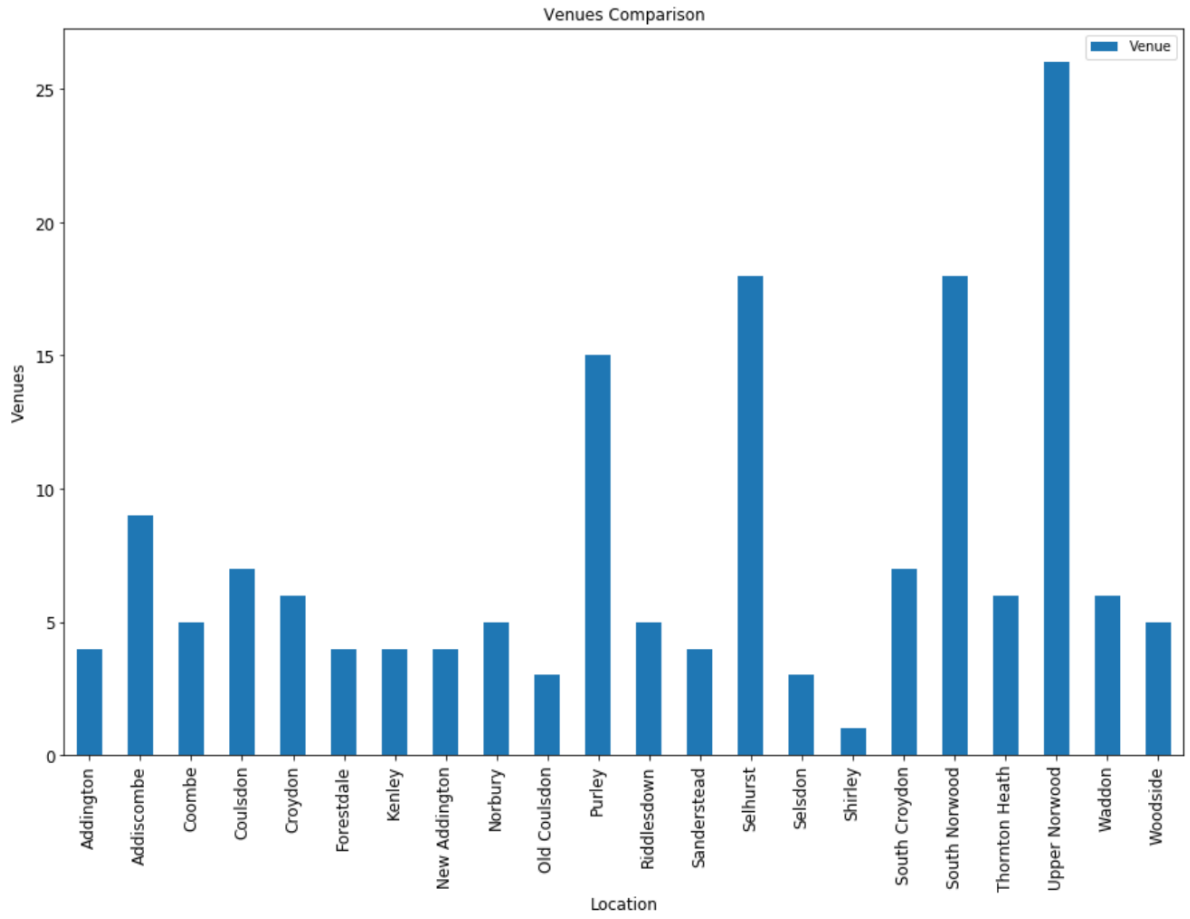
Venue Data

Foursquare API was used to explore the areas and segment them. I set the limit as 100 venue and the radius 500 meter for each area from their given latitude and longitude coordinate. 172 results were returned. Part of the result is as follows.

	Location	Latitude	Longtitude	Venue	Venue Latitude	Venue Longtitude	Venue Category
0	Addington	51.3583	-0.0305	Shell	51.357047	-0.033468	Gas Station
1	Addington	51.3583	-0.0305	The Cricketers (Harvester)	51.357833	-0.032844	English Restaurant
2	Addington	51.3583	-0.0305	Addington Village Interchange	51.356374	-0.032680	Bus Station
3	Addington	51.3583	-0.0305	Addington Village London Tramlink Stop	51.356276	-0.032923	Tram Station
4	Addiscombe	51.3810	-0.0663	Co-op Food	51.381969	-0.069717	Grocery Store

Exploratory Data Analysis

Venues were compared to have an idea of what area had the most venues. The bar chart comparison is as follows. Upper Norwood had the most venues whist Shirley had the least.



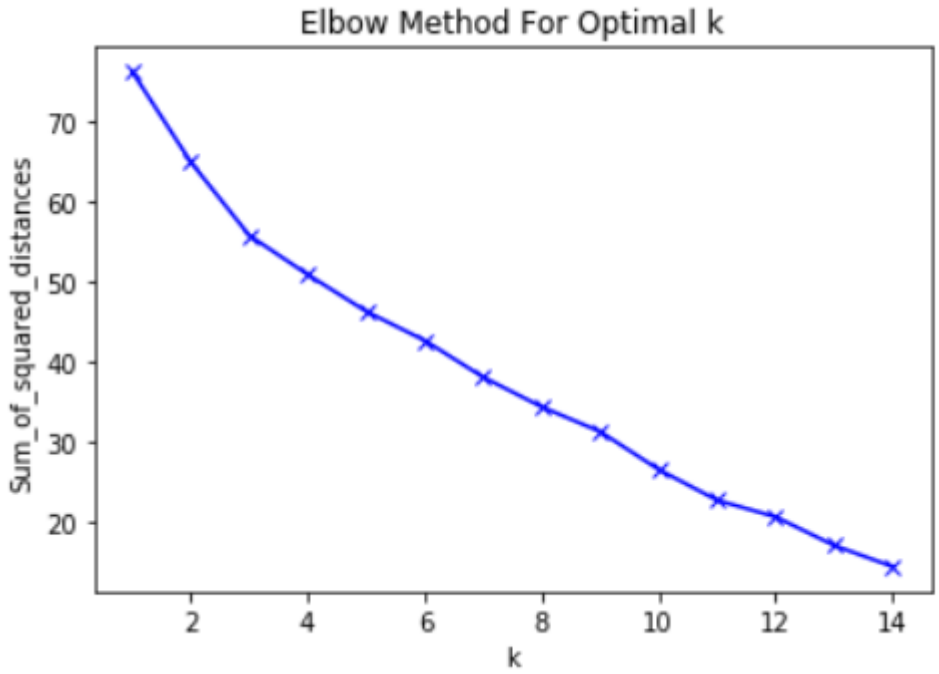
61 unique categories were returned by Foursquare. Having normalized the data, a dataframe with each areas' top 10 venue category was created.

	Location	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	Addington	Tram Station	Gas Station	English Restaurant	Bus Station	Waste Facility	Coffee Shop	Gastropub	Garden Center	Furniture / Home Store	French Restaurant
1	Addiscombe	Grocery Store	Park	Café	Bakery	Waste Facility	Fish & Chips Shop	Diner	English Restaurant	Fast Food Restaurant	Flea Market
2	Coombe	Café	Park	Hotel	French Restaurant	Restaurant	Waste Facility	English Restaurant	Convenience Store	Diner	Fish & Chips Shop
3	Coulsdon	Supermarket	Coffee Shop	Pub	Martial Arts Dojo	Pharmacy	Grocery Store	French Restaurant	Flea Market	Fish & Chips Shop	Furniture / Home Store
4	Croydon	Supermarket	Grocery Store	Tram Station	Warehouse Store	Park	Fast Food Restaurant	Fish & Chips Shop	Convenience Store	Diner	English Restaurant

Unsupervised Learning

Using the k-Means algorithm, the venues in the areas would be clustered into 3 clusters.

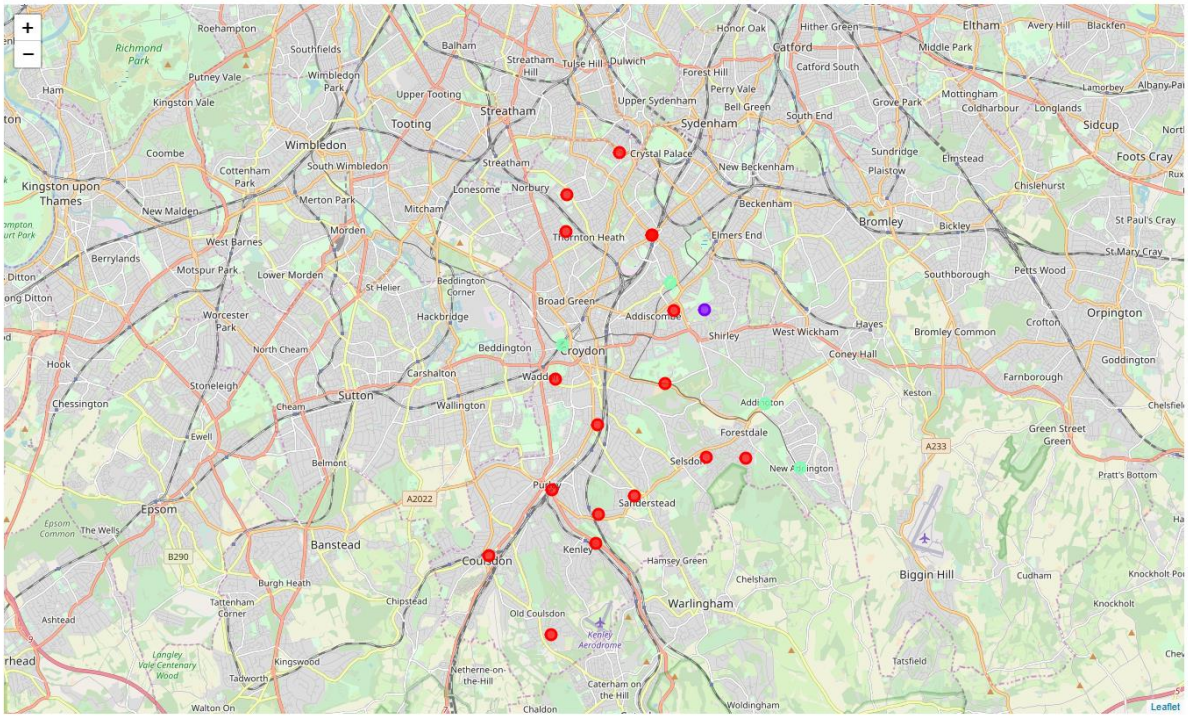
The optimal k was derived using the MinMaxScaler to rescale the data and plotting the values of k in a graph. The elbow method was then used to determine the optimum k as 3.



Results

By gathering data from different sources and transforming the data, the k-means algorithm was used to cluster the areas. The map with all the areas being represented in their respective clusters was generated using the folium library.

	Location	London_borough	Post_town	Postcode_district	Dial_code	OS_grid_ref	Latitude	Longtitude	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	7t Co
0	Addington	Croydon	CROYDON	CR0	020	TQ375645	51.3583	-0.0305	2	Tram Station	Gas Station	Bus Station	English Restaurant	Waste Facility	Convenience Store	
1	Addiscombe	Croydon	CROYDON	CR0	020	TQ345665	51.3810	-0.0663	0	Park	Grocery Store	Café	Bakery	Food Truck	English Restaurant	Fa Res
2	Coombe	Croydon	CROYDON	CR0	020	TQ342647	51.3633	-0.0696	0	Park	Restaurant	Hotel	Café	French Restaurant	Diner	Res
3	Coulsdon	Croydon	COULSDON	CR5	020, 01737	TQ298596	51.3211	-0.1386	0	Supermarket	Coffee Shop	Convenience Store	Pub	Martial Arts Dojo	Pharmacy	
4	Croydon	Croydon	CROYDON	CR0	020	TQ335655	51.3727	-0.1099	2	Tram Station	Grocery Store	Warehouse Store	Chinese Restaurant	Supermarket	Park	Fa Res



Discussions

As the value of k was determined using the elbow method which is not always very reliable, other values of k were used to observe how the clusters would change. Setting the value of k to 5 did not give any better modelling of the data.

	Location	OS_grid_ref	Latitude	Longtitude	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	10th Most Common Venue
1	Addiscombe	TQ345665	51.3810	-0.0663	0	Park	Grocery Store	Café	Bakery	Food Truck	English Restaurant	Fast Food Restaurant	Fish & Chips Shop	Flea Market	Waste Facility
2	Coombe	TQ342647	51.3633	-0.0696	0	Park	Restaurant	Hotel	Café	French Restaurant	Diner	English Restaurant	Fast Food Restaurant	Fish & Chips Shop	Waste Facility
3	Coulsdon	TQ298596	51.3211	-0.1386	0	Supermarket	Coffee Shop	Convenience Store	Pub	Martial Arts Dojo	Pharmacy	Grocery Store	French Restaurant	Furniture / Home Store	Food Truck
5	Forestdale	TQ366625	51.3450	-0.0380	0	Golf Course	Home Service	Campground	Waste Facility	Convenience Store	Gastropub	Gas Station	Garden Center	Furniture / Home Store	French Restaurant
6	Kenley	TQ327600	51.3242	-0.0969	0	Grocery Store	Platform	Train Station	Pub	Flea Market	Diner	English Restaurant	Fast Food Restaurant	Fish & Chips Shop	Food Truck
8	Norbury	TQ315695	51.4092	-0.1083	0	Convenience Store	Gym / Fitness Center	Bar	Mediterranean Restaurant	Café	Flea Market	English Restaurant	Fast Food Restaurant	Fish & Chips Shop	Waste Facility
9	Old Coulsdon	TQ315575	51.3018	-0.1143	0	Construction & Landscaping	Home Service	Café	Middle Eastern Restaurant	Chinese Restaurant	Diner	Gastropub	Gas Station	Garden Center	Furniture / Home Store
10	Purley	TQ313615	51.3373	-0.1141	0	Pizza Place	Platform	Pub	Gym	Sandwich Place	Fast Food Restaurant	Convenience Store	Coffee Shop	Café	Pharmacy
11	Riddlesdown	TQ327608	51.3313	-0.0957	0	Fish & Chips Shop	Train Station	Indian Restaurant	Waste Facility	Construction & Landscaping	Gastropub	Gas Station	Garden Center	Furniture / Home Store	French Restaurant
12	Sanderstead	TQ337613	51.3358	-0.0818	0	Gourmet Shop	Chinese Restaurant	Coffee Shop	Park	Art Gallery	Gastropub	Gas Station	Garden Center	Furniture / Home Store	Diner
13	Selhurst	TQ340684	51.3995	-0.0747	0	Platform	Café	Coffee Shop	Gas Station	Park	Asian Restaurant	Supermarket	Sports Bar	Sandwich Place	Indian Restaurant

From the clustered map we can see that the areas in the red cluster would be the best areas to establish a restaurant. The areas within the London Borough of Croydon that can are in this cluster are as follows

- Addiscombe
- Coombe
- Coulsdon
- Forestdale
- Kenley
- Norbury
- Old Coulsdon
- Purley
- Riddlesdown
- Sanderstead
- Selhurst
- Selsdon
- South Croydon
- South Norwood
- Thornton Heath
- Upper Norwood
- Waddon

These areas have proximity to other businesses. We can also see amongst the three most common venues in these areas we mostly have Cafes, Coffee Shops and Ethnic Restaurants. The style of service that will be most suited in these areas would be fast casual as this is the style most suited for cafes and coffee shops.

Conclusion

Having analyzed the data and applied machine learning techniques I was able to determine what areas were suitable for opening a restaurant. This kind of analysis can be performed on other locations to determine what areas would be suitable for any kind of business.