# Introduction/Business Problem

London is a diverse city with a history stretching back to the days of the roman empire. As such, the city has been home to a number of different cultures and ethnicities. In present day London, certain boroughs and regions of the city are concentrated to have speciality foods and services, e.g. Chinatown for Asian food and Soho for fusion. The purpose of this project is to identify the most common borough location for a given speciality food. Such a list will allow prospective businesses to identify an ideal location to establish a new restaurant either within a known community or to help in diversifying the offerings within the borough.

# Data

The project will first compile a list of london boroughs and associated coordinates from the following wikitable link: https://en.wikipedia.org/wiki/List_of_London_boroughs (https://en.wikipedia.org/wiki/List_of_London_boroughs)

A venue list will then be compiled for each location using the Foursquare API and any food venues with the "restaurant" or "cafe" identifier will be filtered out and processed to identify food speciality. As the boroughs vary in size from 4.6--34 mi^2, the venue list will be compiled based on 60% of a respective borough size to minimize duplicate venue generation between neighboring boroughs. This data would then be assessed by cluster analysis to identify the primary speciality. The presence of the style compared to other venues within in the borough will also be assessed to help compile a ranked list of boroughs whereby a particular food speciality is most likely found.

# Results

The documented code for this project is given in 'LondonClusterSegmentationFood.ipynb.' The wikitable within the url, https://en.wikipedia.org/wiki/List_of_London_boroughs (https://en.wikipedia.org/wiki/List_of_London_boroughs), was scraped for the borough name, geographic location, and the borough size in square kilometres. The venue list within each borough was then acquired over a radius encompassing 60% of the size of the borough (Fig 1) and filtered for food venues containing "Restaurant" or "Cafe" in the "Venue Category" column.

| | Borough | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Barking and Dagenham | 51.5607 | 0.1557 | Central Park | 51.559560 | 0.161981 | Park |
| 1 | Barking and Dagenham | 51.5607 | 0.1557 | Lara Grill | 51.562445 | 0.147178 | Turkish Restaurant |
| 2 | Barking and Dagenham | 51.5607 | 0.1557 | Costa Coffee | 51.576890 | 0.179497 | Coffee Shop |
| 3 | Barking and Dagenham | 51.5607 | 0.1557 | The Range | 51.575550 | 0.180254 | Furniture / Home Store |
| 4 | Barking and Dagenham | 51.5607 | 0.1557 | Asda | 51.565770 | 0.143393 | Supermarket |

## Fig 1: Representative dataframe output for listing venues for a given borough.

This filtered list was then further processed to identify the most commonly found food venues for each borough (Fig 2) and processed through k-means cluster analysis to categorize the boroughs by their common food venues.

| | Borough | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| 0 | Barking and Dagenham | Turkish Restaurant | American Restaurant | Italian Restaurant | Brazilian Restaurant | Restaurant |
| 1 | Barnet | Italian Restaurant | Turkish Restaurant | Indian Restaurant | Sushi Restaurant | Restaurant |
| 2 | Bexley | Italian Restaurant | English Restaurant | American Restaurant | Mediterranean Restaurant | Greek Restaurant |
| 3 | Brent | Indian Restaurant | Latin American Restaurant | Restaurant | English Restaurant | Italian Restaurant |
| 4 | Bromley | Turkish Restaurant | American Restaurant | Italian Restaurant | Asian Restaurant | Indian Restaurant |

## Fig 2: Representative dataframe output to identify the most common food venues for a given borough
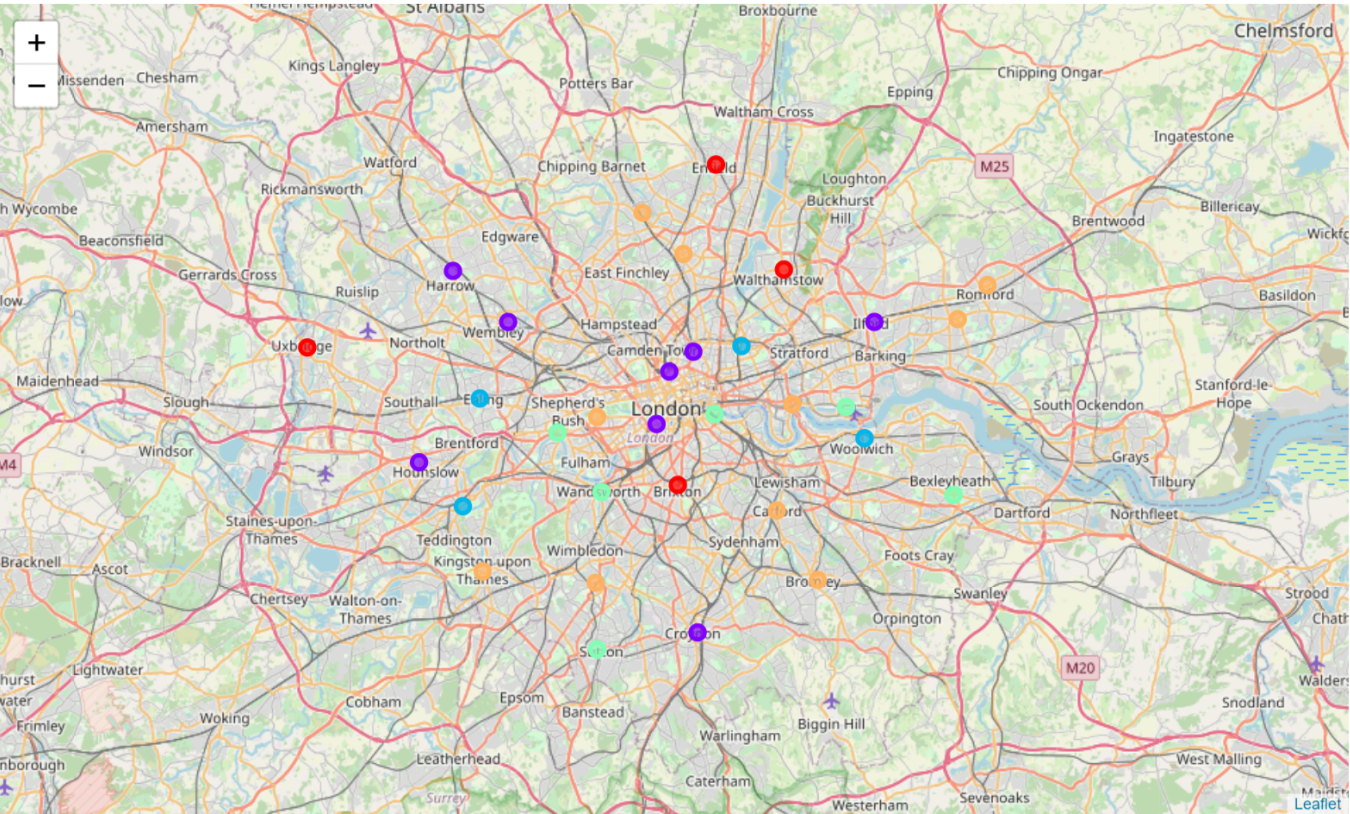


## Fig 3: Map to visualize cluster analysis readout. Cluster locations are centralized at each borough

Given that multiple boroughs can be contained within a single cluster, the venue distribution in each cluster was further identified (Fig 4).

| Cluster Labels | Boroughs in Cluster | 1st Most Speciality | 1st Fraction | 2nd Most Speciality | 2nd Fraction | 3rd Most Speciality | 3rd Fraction | 4th Most Speciality | 4th Fraction |
|---|---|---|---|---|---|---|---|---|---|
| 0 | [Enfield, Hillingdon, Lambeth, Waltham Forest] | Portuguese Restaurant | 0.250 | Turkish Restaurant | 0.500000 | Middle Eastern Restaurant | 0.250000 | Vietnamese Restaurant | 0.250000 |
| 1 | [Brent, Camden, Croydon, Harrow, Hounslow, Isl... | Indian Restaurant | 0.625 | Portuguese Restaurant | 0.250000 | Middle Eastern Restaurant | 0.250000 | English Restaurant | 0.250000 |
| 2 | [Ealing, Greenwich, Hackney, Richmond upon Tha... | Vietnamese Restaurant | 0.750 | Caribbean Restaurant | 0.250000 | Fast Food Restaurant | 0.250000 | Japanese Restaurant | 0.250000 |
| 3 | [Bexley, Hammersmith and Fulham, Newham, South... | Italian Restaurant | 0.500 | English Restaurant | 0.166667 | Italian Restaurant | 0.333333 | Mediterranean Restaurant | 0.166667 |
| 4 | [Barking and Dagenham, Barnet, Bromley, Haring... | Turkish Restaurant | 0.300 | American Restaurant | 0.300000 | Italian Restaurant | 0.300000 | Indian Restaurant | 0.200000 |

## Fig 4: Most commonly found specialities in each cluster based on total venues within the boroughs.

Here, it can be observed that while Portuguese restaurants comprise of 25% of the most common venues within the boroughs in the first cluster, Turkish restaurants are more commonly found within those boroughs at 50%. A list of the least common venues within the boroughs was also identified, where more similarity with the venue specialities in the clusters was observed.

| Cluster Labels | Boroughs in Cluster | 1st Least Common Speciality | 1st Fraction | 2nd Least Common Speciality | 2nd Fraction | 3rd Least Common Speciality | 3rd Fraction | 4th Least Common Speciality | 4th Fraction |
|---|---|---|---|---|---|---|---|---|---|
| 0 | [Enfield, Hillingdon, Lambeth, Waltham Forest] | Modern European Restaurant | 0.75 | Mediterranean Restaurant | 0.75 | Malay Restaurant | 0.750 | Lebanese Restaurant | 0.750000 |
| 1 | [Brent, Camden, Croydon, Harrow, Hounslow, Isl... | Okonomiyaki Restaurant | 0.50 | Modern European Restaurant | 0.50 | Mediterranean Restaurant | 0.375 | Malay Restaurant | 0.375000 |
| 2 | [Ealing, Greenwich, Hackney, Richmond upon Tha... | Okonomiyaki Restaurant | 0.50 | Middle Eastern Restaurant | 0.50 | Mediterranean Restaurant | 0.500 | Malay Restaurant | 0.500000 |
| 3 | [Bexley, Hammersmith and Fulham, Newham, South... | Modern European Restaurant | 0.50 | Middle Eastern Restaurant | 0.50 | Mediterranean Restaurant | 0.500 | Malay Restaurant | 0.666667 |
| 4 | [Barking and Dagenham, Barnet, Bromley, Haring... | Modern European Restaurant | 0.40 | Middle Eastern Restaurant | 0.50 | Mediterranean Restaurant | 0.400 | Malay Restaurant | 0.500000 |

## Fig 5: Least commonly found specialities in each cluster based on total venues within the boroughs.

# Discussion and Conclusions

In this study, the food venue data within the city of London, UK was segmented and classified into various clusters to determine the common food venues in the London boroughs. This information is designed to indicate to restaurateurs the approximate make up of each borough and how to approach a new business venture specializing in a particular cuisine. From the data acquired, certain similarities were observed in the venue specialities when classifying botht he most common and most esoteric venues within London. While all of the clusters expressed unique values for the first and second most common specialities, the subsequent values were found to be mroe uniform across the clusters. This may be a result of the search parameters being too relaxed, thereby permitting venue overlap within different boroughs.

Furthermore, the current data does not provide insight as to venue popularity in the designated area, such as population demographics within the borough; however, future work can further refine the classification algorithm to better identify these differences to help businesses decide on their venture.