Association between area-level deprivation, demographic trends and restaurant locations in London

Adrian Turculet

Contents

1.0 Introduction

1.1 Background

London is the capital of England and the United Kingdom and one of the largest and most important cities in the world. It has a diverse range of people and cultures that currently make the city one of the most populate cities in the world with 9,787,426 inhabitants registered at the 2011 census.[1] With such a large number of inhabitants, the deprivation levels and population density can vary immensely between the 32 local authorities, also known as the London boroughs.

1.2 Business problem

London is not only known for its population size and diversity but also for the restaurant industry which plays an important role in the city's economy. For this reason, in this project I will be comparing and analysing the different deprivation levels and population demographics across London's boroughs and investigate the association between deprivation levels and restaurant locations, specifically, what types of restaurants and their prices can be found in different boroughs.

1.3 Interest

The findings of this research can help potential/existing restaurateurs looking to open a new restaurant or other business stakeholders, by identifying new suitable locations for their business, based on various characteristics such as:

- Restaurant price range
- Number of competing restaurants in the area
- Population size/density
- Area-level deprivation
- Region size

2.0 Data

To achieve the goals set out in this project, various types of data will be explored and analysed. This includes real-time location data using the Foursquare API[2], geospatial data using Python Geocoder library[3] and GADM (Database of Global Administrative Areas) data, Indices of Deprivation data[4] and London Borough Profiles data using the online London Datastore[5].

2.1 Foursquare API

Foursquare is a social networking service that provides independent location data. The purpose of the platform is to provide users with information about businesses and attractions around them using real-time location data.[6] This API will be instrumental in identifying detailed information about restaurants across London.

2.2 Python Geocoder library

Geocoder is a simple and consistent geocoding library written in Python. It helps developers to locate the coordinates of addreses, cities, countries and landmarks across the globe. The Geocoder will help get the coordinates for every London borough.

2.3 GADM (Database of Global Administrative Areas) data

GADM is a high-resolution database of country administrative areas, with a goal of "all countries, at all levels, at any time period. The database includes shapefiles that are used in Geographic Information System (GIS) applications.

2.4 Indices of Deprivation data and London Borough Profiles data

The Index of Multiple Deprivation 2019 combines a number of indicators, chosen to cover a range of economic, social and housing issues, into a single deprivation score for each small area in England. This allows each area to be ranked relative to one another according to their level of deprivation.[4]
Index of deprivation will be used to assign a deprivation score to each London borough.

The London Borough Profiles help paint a general picture of an area by presenting a range of headline indicator data in both spreadsheet and map form to help show statistics covering demographic, economic, social and environmental datasets for each borough, alongside relevant comparator areas. This information will be used to understand each borough’s demographic trends.

2.5 London Borough Profiles

The London Borough Profiles help paint a general picture of an area by presenting a range of headline indicator data in both spreadsheet and map form to help show statistics covering demographic, economic, social and environmental datasets for each borough, alongside relevant comparator areas. [5]

2.6 Data sources

[1] "2011 Census – Built-up areas" ONS. Retrieved 08 January 2020
[2] “Foursquare API”
[3] “Python Geocoder”
[4] “Indices of Multiple Deprivation 2019, Borough”
[5] “London Borough Profiles data 2017”
[6] “Foursquare”

2.7 Shape files for the creation of the London boroughs map

https://data.gov.uk/dataset/6cdebf5d-c69b-4480-8c9c-53ab8a816b9d/statistical-gis-boundary-files-for-london

2.8 Converting the shapefiles into GeoJSON format for rendering in Folium:

https://mapshaper.org/

2.9 Data cleansing and Feature selection

Data downloaded from the sources mentioned above were combined into one table ready for analysis. The first step was to clean the **London Boroughs Profiles** data. We know that London is made of 33 individual boroughs but the dataset included 38 records. The 5 additional records contained values at bigger geographical levels than borough-level so I excluded them from the data.

Next, selecting the features required for analysis. There were 84 features in total from which I kept 7 features that will be linked to the additional datasets:

- Borough code
- Borough name
- Population size
- Region size
- Population density
- Average Age
- Unemployment rates

The official **Index of Multiple Deprivation 2019 in England and Wales** was used to calculate the deprivation score at borough level.

The feature values were then transformed into numeric values for analysis.

After gathering the relevant features about each London borough I then used the **Geocoder** library in Python to fetch each borough’s geographical coordinates and link them to the Boroughs profiles dataset.

Finally, using the **FourSquare API**, I fetched all venues for each borough (with a limit of 100 per borough due to account limitations). The data was then filtered to keep only the venues that were actual restaurants.

Using the restaurant dataset created above, I looped through the FourSquare API database to fetch the restaurant price Tier (from 1 to 4 as Cheap to Very Expensive).

3.0 Methodology

In order to explore the whole array of datasets in a fast and efficient way, I used various Python libraries used in data manipulation like **Pandas** and **Numpy**

When exploring the datasets, some issues have been identified that needed to be sorted before any inferential statistical testing would be performed.

3.1 Data aggregation

The dataset containing the Deprivation index was at a smaller geographical level (Super Lower Output Area) compared to the London Borough Profiles.

	LSOA code (2011)	LSOA name (2011)	Local Authority District code (2019)	Local Authority District name (2019)	Index of Multiple Deprivation (IMD) Rank	Index of Multiple Deprivation (IMD) Decile
0	E01000001	City of London 001A	E09000001	City of London	29,199	9
1	E01000002	City of London 001B	E09000001	City of London	30,379	10
2	E01000003	City of London 001C	E09000001	City of London	14,915	5
3	E01000005	City of London 001E	E09000001	City of London	8,678	3
4	E01000006	Barking and Dagenham 016A	E09000002	Barking and Dagenham	14,486	5

To get a deprivation score at borough level, I calculated the median **Index of Multiple Deprivation** at Local Authority District lever for each borough:

	Local Authority District name (2019)	Index of Multiple Deprivation (IMD) Decile
21	Blackpool	2.41
159	Manchester	2.54
145	Knowsley	2.56
8	Barking and Dagenham	2.68
112	Hackney	2.74
154	Liverpool	2.75
220	Sandwell	2.81
18	Birmingham	2.89
142	Kingston upon Hull, City of	2.95
190	Nottingham	2.96

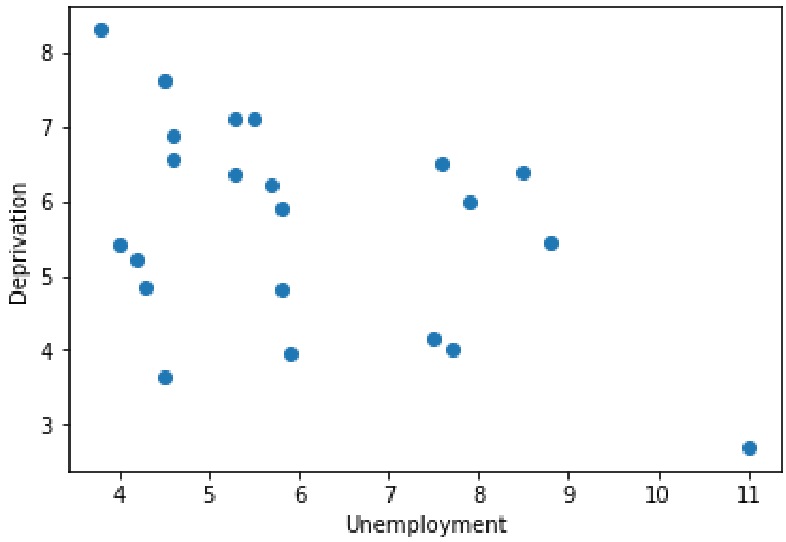
3.2 Exploratory Data Analysis

3.2.1 Missing Data

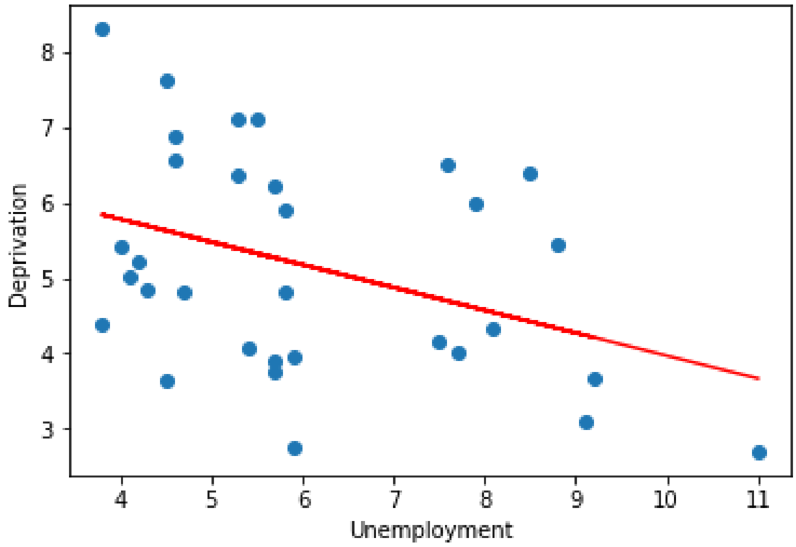
The London Borough Profiles dataset had a missing value in one borough for the 'Unemployment' rate. It is important that there are no missing values if I want the results to be complete and accurate.

To address the missing value issue, I used a **classification model** using the machine learning algorithm Linear Regression to predict the missing Unemployment rate for the borough "City of London" from the **sci-kit learn** Python library.

I used the Deprivation scores and Unemployment rates for each borough and visualised the relationship between them using **pyplot** from **matplotlib**.



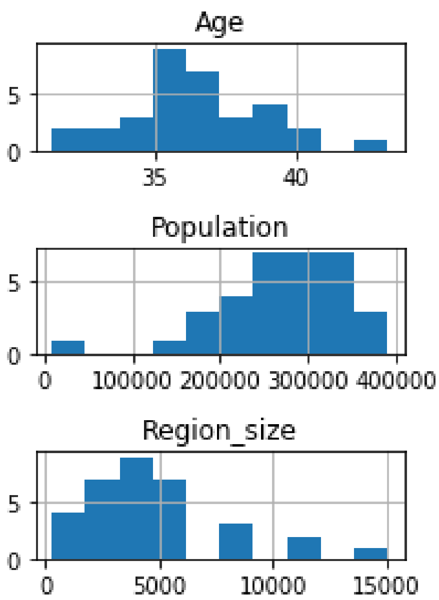
The hypothesis here is that if the deprivation score (1 most deprived – 8 least deprived) for a borough is low then the Unemployment rate is high.



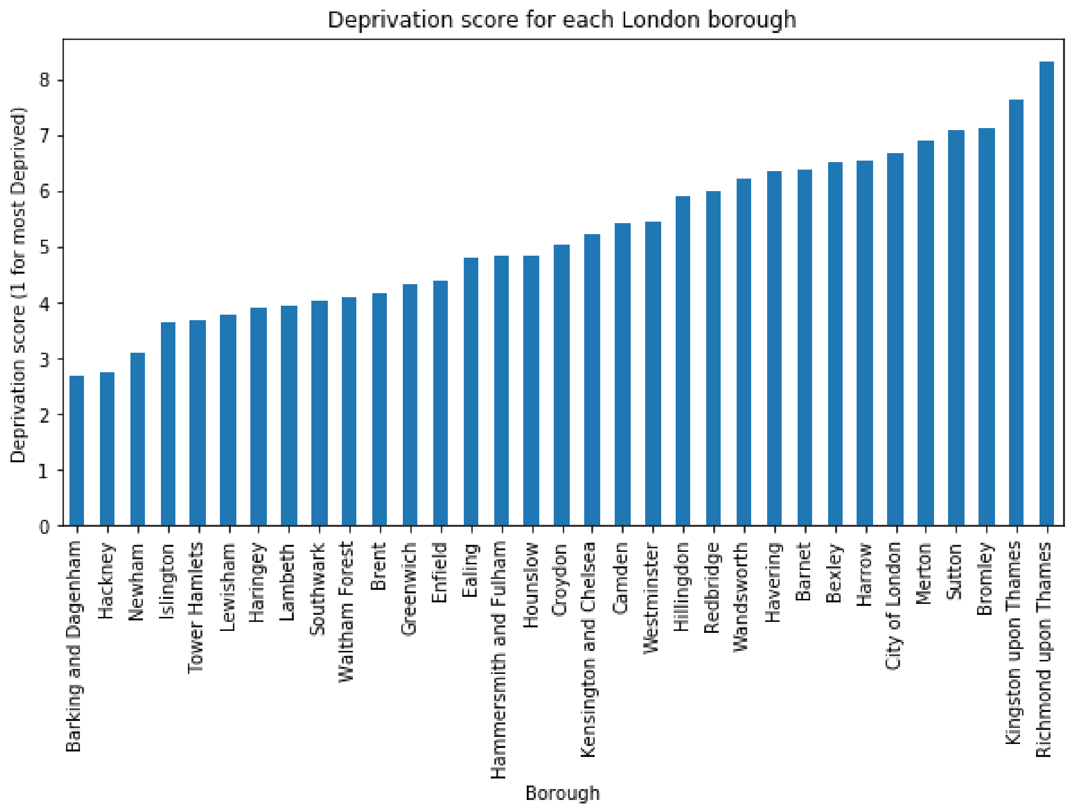
After fitting the prediction line over the existing data, the Unemployment rate for "City of London" was predicted to be 4.97

3.2.2 Analyse data distribution

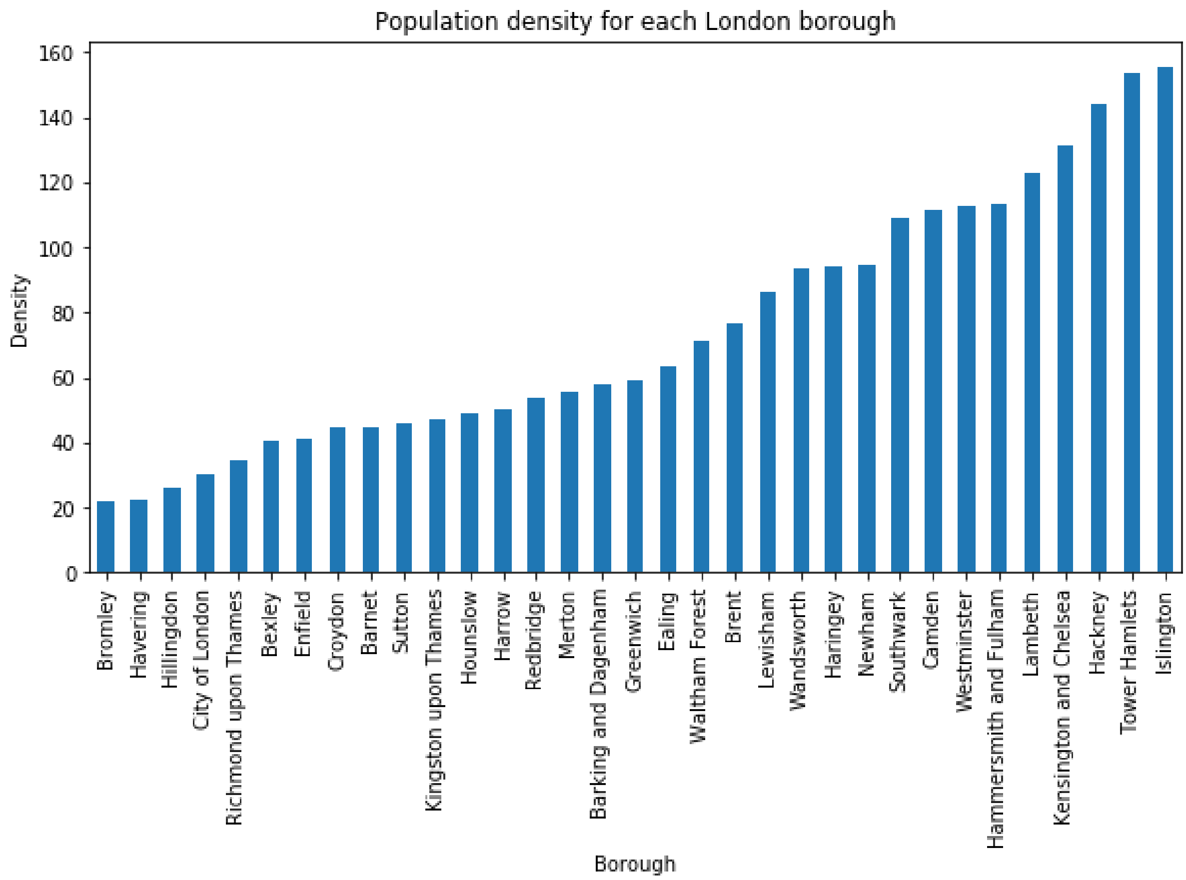
Based on the histograms below, we can see that the data is not evenly distributed except the values from the “Age“ feature.



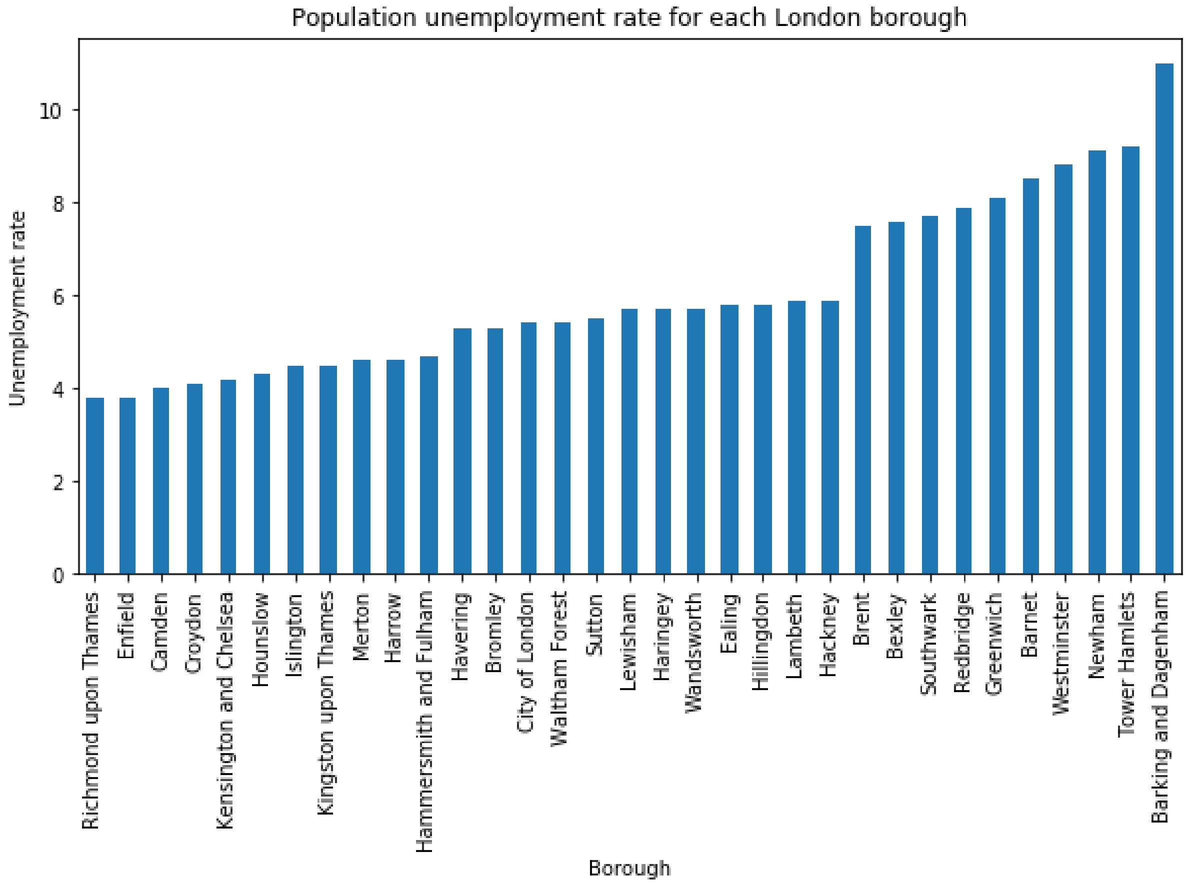
Below is a representation of the deprivation scores for each borough. It’s important to note the most deprived boroughs – “Barking and Dagenham”, “Hackney” and “Newham” and the least deprived boroughs – “Kingston upon Thames” and “Richmond upon Thames”



Deprivation score for each London borough
Deprivation score (1 for most Deprived)
0
1
2
3
4
5
6
7
8
Barking and Dagenham
Hackney
Newham
Islington
Tower Hamlets
Lewisham
Haringey
Lambeth
Southwark
Waltham Forest
Brent
Greenwich
Enfield
Ealing
Hammersmith and Fulham
Hounslow
Croydon
Kensington and Chelsea
Camden
Westminster
Hillingdon
Redbridge
Wandsworth
Havering
Barnet
Bexley
Harrow
City of London
Merton
Sutton
Bromley
Kingston upon Thames
Richmond upon Thames
Borough



Population density for each London borough
Density
0
20
40
60
80
100
120
140
160
Bromley
Havering
Hillingdon
City of London
Richmond upon Thames
Bexley
Enfield
Croydon
Barnet
Sutton
Kingston upon Thames
Hounslow
Harrow
Redbridge
Merton
Barking and Dagenham
Greenwich
Ealing
Waltham Forest
Brent
Lewisham
Wandsworth
Haringey
Newham
Southwark
Camden
Westminster
Hammersmith and Fulham
Lambeth
Kensington and Chelsea
Hackney
Tower Hamlets
Islington
Borough



The “London Boroughs” dataframe is now covers demographic, economic and social information for each borough and is ready for the next step in our analysis.

	Code	Borough	Population	Population_density	Age	Unemployment	Region_size	Deprivation
0	E09000002	Barking and Dagenham	209000	57.9	32.9	11.0	3611	2.68
1	E09000012	Hackney	274300	144.0	33.1	5.9	1905	2.74
2	E09000025	Newham	342900	94.7	32.1	9.1	3620	3.09
3	E09000019	Islington	231200	155.6	34.8	4.5	1486	3.64
4	E09000030	Tower Hamlets	304000	153.7	31.4	9.2	1978	3.68

3.3 Fetching geospatial information for each London borough and their venues

Using the Geocoder library in Python and connecting to the FourSquare API, I extracted the borough and venue geographical information.

The dataset contained 3269 venues from which there were 664 restaurants.

	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue id	Tier
0	Barking and Dagenham	51.554117	0.150504	Nando's	51.567729	0.116807	Portuguese Restaurant	5d72e633b40a620008bf3518	2
1	Barking and Dagenham	51.554117	0.150504	Ciao Bella	51.576103	0.182819	Italian Restaurant	4bf70cfdc07c9c74b690bbef	2
2	Barking and Dagenham	51.554117	0.150504	The Greyhound (Harvester)	51.568429	0.119456	English Restaurant	4dea493c18386283a3e063f6	0
3	Barking and Dagenham	51.554117	0.150504	The Pipe Major	51.545800	0.165860	Restaurant	4e7a1f5414954a343fb58258	2
4	Barking and Dagenham	51.554117	0.150504	Burger King	51.565342	0.193468	Fast Food Restaurant	4adcbef1f964a5205a2f21e3	1

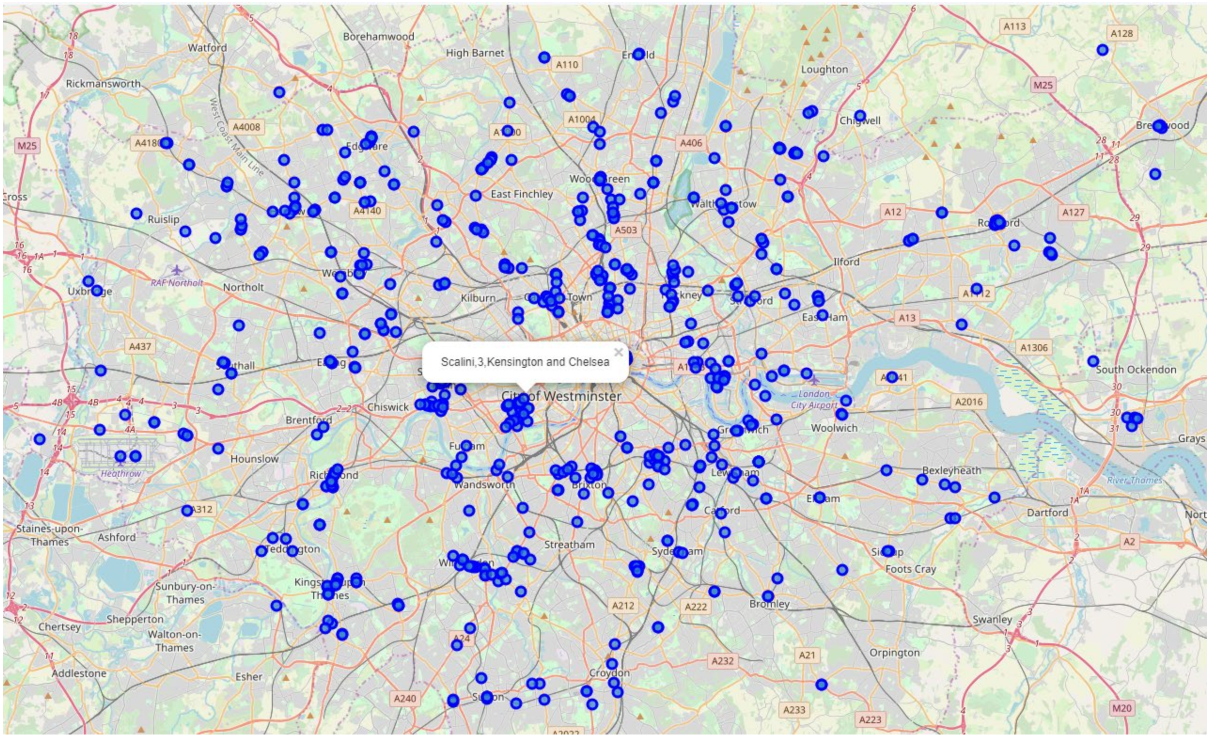
For each restaurant, I fetched the **Venue name**, **id**, **category** and **price tier** through the FourSquare API.

For the purpose of this study I removed any restaurant venues that no Price Tier value available on FourSquare (i.e. “Tier”=0)

3.4 Visualise data using Folium

Folium makes it easy to visualize data that’s been manipulated in Python on an interactive leaflet map. We use the dataset generated in the previous steps to create interactive maps containing all restaurant venues in London fetched using the FourSquare API and each Borough.

Below is a geographical representation of all restaurants fetched from FourSquare.



3.4 One-hot encoding process

Calculate the most common restaurant types for each borough and separately the number of restaurants for each different price tie

	Borough	Afghan Restaurant	African Restaurant	American Restaurant	Argentinian Restaurant	Asian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Caribbean Restaurant	Chinese Restaurant	...	Southern / Soul Food Restaurant	Spanish Restaurant	Sri Res
0	Barking and Dagenham	0.000000	0.000000	0.125000	0.000000	0.000000	0.062500	0.000000	0.000000	0.000000	...	0.000000	0.000000	0
1	Barnet	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.038462	0.076923	...	0.000000	0.000000	0
2	Bexley	0.000000	0.000000	0.100000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0
3	Brent	0.000000	0.000000	0.000000	0.000000	0.037037	0.000000	0.000000	0.000000	0.074074	...	0.000000	0.000000	0
4	Bromley	0.000000	0.076923	0.000000	0.076923	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0
5	Camden	0.000000	0.000000	0.000000	0.052632	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0
6	City of London	0.000000	0.000000	0.000000	0.000000	0.038462	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0
7	Croydon	0.000000	0.052632	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.052632	...	0.000000	0.000000	0
8	Ealing	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.071429	0.000000	...	0.071429	0.000000	0
9	Enfield	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0
10	Greenwich	0.000000	0.000000	0.000000	0.071429	0.142857	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0
11	Hackney	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0
12	Hammersmith and Fulham	0.000000	0.000000	0.000000	0.000000	0.035714	0.000000	0.000000	0.000000	0.142857	...	0.000000	0.000000	0
13	Haringey	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.034483	0.034483	...	0.000000	0.034483	0
14	Harrow	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.055556	...	0.000000	0.000000	0
15	Havering	0.000000	0.000000	0.058824	0.000000	0.058824	0.000000	0.000000	0.000000	0.058824	...	0.000000	0.000000	0
16	Hillingdon	0.041667	0.000000	0.000000	0.041667	0.000000	0.000000	0.000000	0.000000	0.041667	...	0.000000	0.000000	0
17	Hounslow	0.000000	0.000000	0.000000	0.100000	0.050000	0.000000	0.000000	0.000000	0.050000	...	0.000000	0.000000	0
18	Islington	0.041667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0
19	Kensington and Chelsea	0.000000	0.000000	0.037037	0.037037	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0
20	Kingston upon Thames	0.000000	0.000000	0.000000	0.000000	0.038462	0.000000	0.038462	0.000000	0.038462	...	0.000000	0.000000	0
21	Lambeth	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.136364	0.000000	...	0.000000	0.000000	0
22	Lewisham	0.000000	0.000000	0.000000	0.058824	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0
23	Merton	0.000000	0.000000	0.000000	0.035714	0.071429	0.035714	0.000000	0.000000	0.000000	...	0.000000	0.000000	0
24	Newham	0.000000	0.000000	0.000000	0.055556	0.000000	0.000000	0.000000	0.000000	0.055556	...	0.000000	0.000000	0
25	Redbridge	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.050000	...	0.000000	0.000000	0
26	Richmond upon Thames	0.000000	0.000000	0.000000	0.117647	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0
27	Southwark	0.000000	0.052632	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.052632	...	0.000000	0.000000	0
28	Sutton	0.000000	0.000000	0.062500	0.000000	0.000000	0.000000	0.000000	0.000000	0.125000	...	0.000000	0.000000	0
29	Tower Hamlets	0.000000	0.000000	0.000000	0.050000	0.050000	0.000000	0.000000	0.000000	0.050000	...	0.000000	0.050000	0
30	Waltham Forest	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.052632	...	0.000000	0.000000	0
31	Wandsworth	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0
32	Westminster	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.090909	0

33 rows × 58 columns

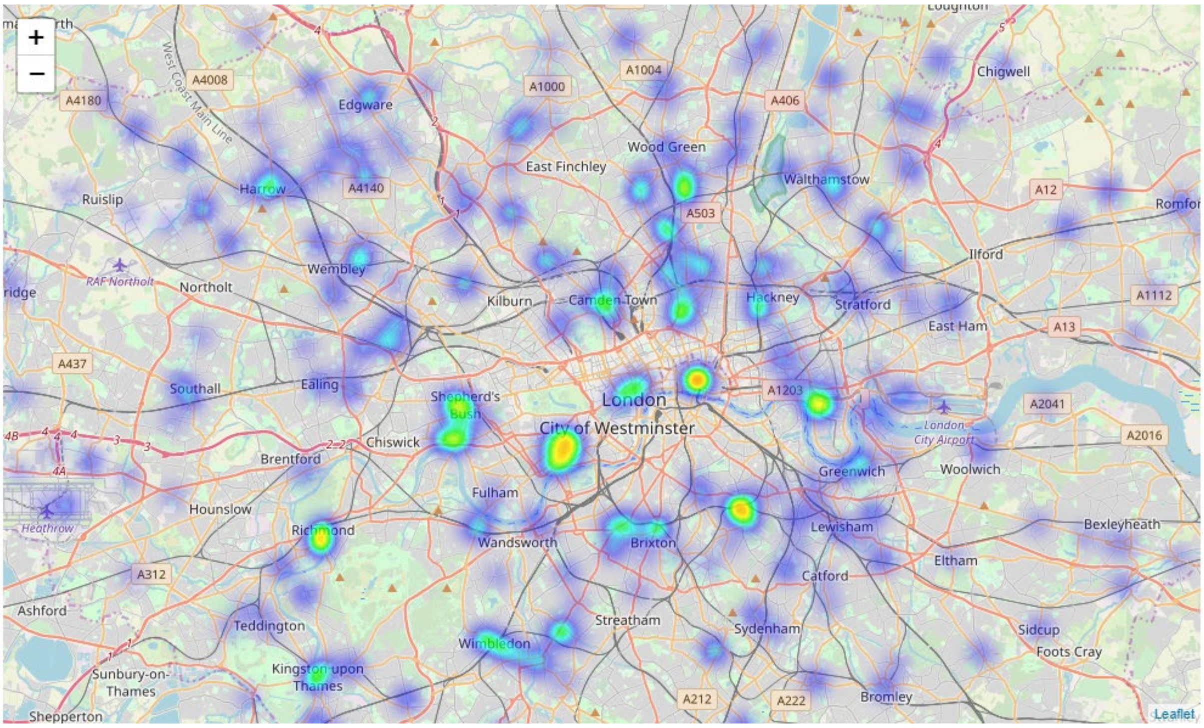
	Borough	Population	Population_density	Age	Unemployment	Region_size	Deprivation	Lat	Long	£	££	£££	££££
0	Barking and Dagenham	209000	57.9	32.9	11.0	3611	2.68	51.554117	0.150504	5	9	0	0
1	Hackney	274300	144.0	33.1	5.9	1905	2.74	51.549049	-0.047801	2	9	0	0
2	Newham	342900	94.7	32.1	9.1	3620	3.09	51.530000	0.029318	2	12	0	1
3	Islington	231200	155.6	34.8	4.5	1486	3.64	51.547156	-0.101694	2	17	2	0
4	Tower Hamlets	304000	153.7	31.4	9.2	1978	3.68	51.514562	-0.035012	2	12	2	1

To better understand how the restaurants are distributed across London and how the restaurant pricing differs, I created a heatmap showing the density and restaurant price range in each borough.

The zones showcasing a bright yellow and orange core represent the boroughs with a high number of restaurants and high price rates in a small radius.

These zones are most apparent in the following borrows:

- Kensington and Chelsea
- City of London
- Richmond
- Kingston upon Thames

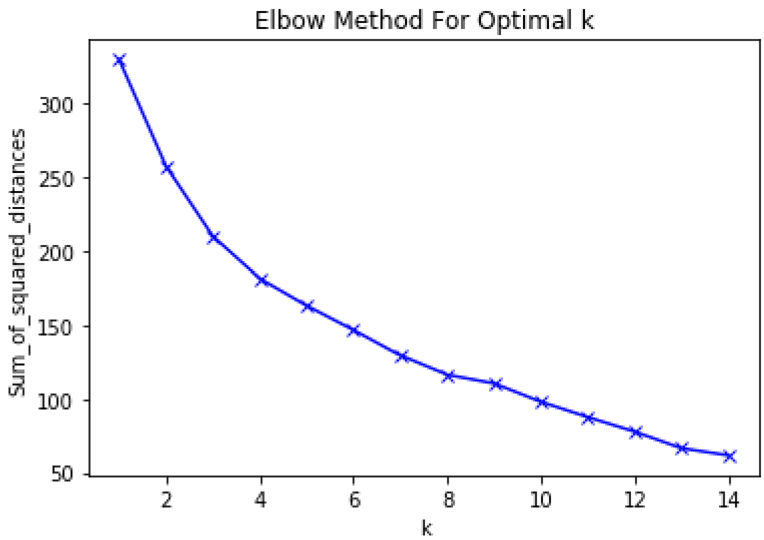


3.5 Unsupervised ML algorythm - Clustering (KMeans)

Now that I have detailed information about each borough, including restaurant counts and prices, I applied the Kmeans Clustering algorythm to create distinct zones of interest

Before running the clustering algorithm, I normalized the data over the standard deviation using StandardScaler() from the sci-kit learn Python library.

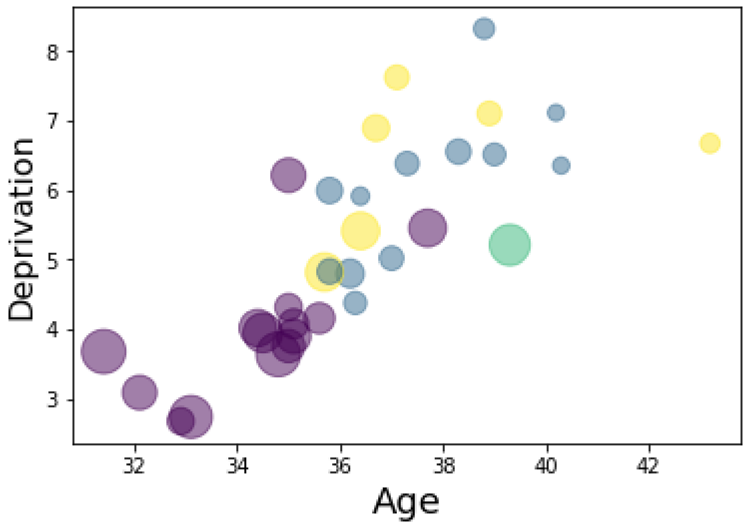
Now in order to determine the optimum number of clusters used to categorise the data, I used the elbow method for optimal K.



Once the optimal K was found (k=4), I applied the Kmeans clustering algorithm and started analysing it using various visualisation tools.

4.0 Results and Discussion

In the graph below each colour represents a cluster (4 clusters) and each bubble is a borough (32 boroughs). The size of the bubble represents the Population density in a borough.

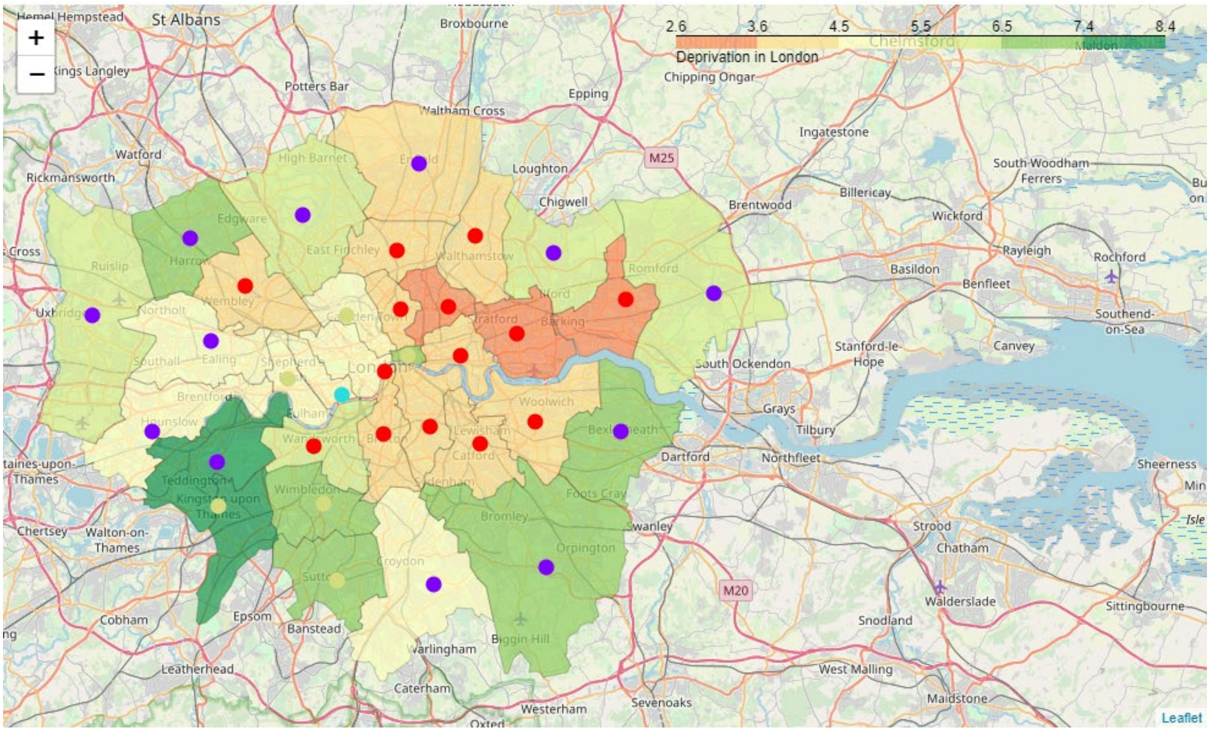


We can clearly conclude that the purple cluster includes boroughs that have a very low deprivation score, are densely populated where residents are on average of young age.

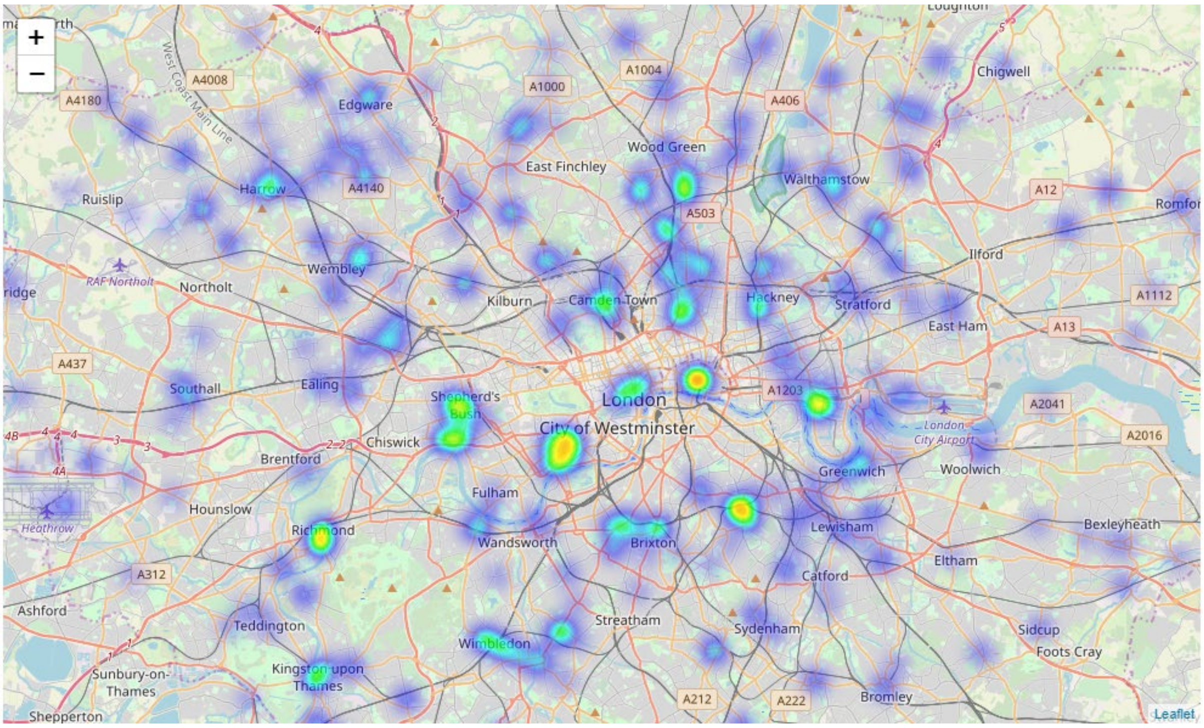
To further understand each cluster, it is crucial to look at them from different perspectives.

4.1 Map of London with superimposed layer of deprivation and clusters

Now that we can see the clusters created by our model on top of the deprivation layers, we can start to understand how boroughs are clustered.



The areas in orange represent the most deprived areas in London while the dark green represents the least deprived.To further understand restaurant placement, we recall the restaurant heatmap showing the density and highest price tiers.



Analysing the most common restaurant types by borough has helped discover some interesting insights in how restaurants are located:

- Barking and Dagenham, the most deprived borough has the Fast food restaurants as most frequent in the borough. This is not surprising considering this borough has the highest unemployment rate and poorest deprivation score in London.

- Newham, the third most deprived borough has the highest frequency of Indian restaurants. It is not unusual to see this as Newham has a very large Indian community.
- Two of the most up and coming and trendiest boroughs in London, Camden and Hackney, have the Vegan/Vegetarian restaurants as the most common type of restaurant.

In contrast, the least deprived boroughs with the highest number of £££ and ££££ ratings, have the highest number of Italian restaurants:

- City of London
- Kensington and Chelsea
- Kingston upon Thames

Based on the types of restaurants discovered in each borough, we can also predict that these boroughs can host a large number of people from countries with a specific cuisine. For example:

- Barnet and Enfield have the highest number of Turkish restaurants which can also mean these boroughs have a large number or turkish nationals living there
- Richmond upon Thames is the only borough with the highest frequency of German restaurants. These could show that there might be a large German community living here.

These hypotheses can be further analysed to best explain the reason behind these restaurant placements.

5.0 Conclusion

After clustering was performed, we can infer on the differences between each cluster and the borough it is in:

- **Cluster 0** represents the most deprived areas with the highest unemployment rate and youngest population. Restaurants in these areas range between cheap and moderate prices.
- **Cluster 1** includes the largest boroughs in terms of region size with the smallest population density. Unemployment rate and deprivation score are below average and span a majority of moderate priced restaurants.
- **Cluster 2** is unique and it includes only one borough. This is a rich area hosting the most expensive restaurants in London. It is a densely -populated area with the lowest Unemployment rate and oldest population - averaging at 39.3
- **Cluster 3** can be distinguished on the map as the most affluent boroughs located in the west and south-west part of London. These are not dense populated areas with a population age averaging 38 years.

In conclusion, the purpose of this study was to identify an association between area-level deprivation, demographic trends and restaurant locations in London. We can see the data analysis performed, reveals some interesting ties between the different types of boroughs and restaurant location.

We also discovered some insights by looking at the most frequent restaurant type in each borough. This prompts further investigation regarding each borough's major communities and the most frequent restaurant type.