Using K-Means clustering to find the best boroughs of London to open a coffee shop

Long Nguyen

Alma Mater Studiorum - Università di Bologna

minhlong.nguyen@studio.unibo.it

4th September 2020

Table of Contents

Introduction

- Ha Trang just moved to London, and she wants to open a new Café.
- However, she needs to be careful in choosing the place to open her coffee shop because of heavy competition from other coffee shops in London.

Figure 0: Computer scientists start counting from 0

Data Sources

- **Foursquare:** In this project, we use Foursquare location data to retrieve popular restaurants near each borough's geographical coordinates.
- **GeoPy:** In this project, we also use GeoPy to retrieve geographical coordinates of 32 boroughs of London, and then use this data to retrieve the list of popular venues for that area by making an API call to Foursquare.
- **London Borough Atlas:** This dataset is publicly available and can be downloaded as a xlsx file. In this project, we use this dataset to retrieve 32 London borough names.

Data Retrieval

Retrieving geographical coordinates of London boroughs

Firstly, we use GeoPy to retrieve the geographical coordinates of London.

```
from geopy.geocoders import Nominatim
geolocator = Nominatim(user_agent="open_street_map")
location = geolocator.geocode("London")
print('Latitude: ', location.latitude, ', 'Longitude: '...
,location.longitude)

>> Latitude:  51.5073219 , Longitude:  -0.1276474
```

We will use the coordinates to detect incorrect values as well as to create a map of London later using Folium

Data Retrieval

Retrieving geographical coordinates of London boroughs

Next, we retrieve borough names from London Borough Atlas.

	Borough		Borough
0	Barking and Dagenham	16	Hounslow
1	Barnet	17	Islington
2	Bexley	18	Kensington and Chelsea
3	Brent	19	Kingston upon Thames
4	Bromley	20	Lambeth
5	Camden	21	Lewisham
6	Croydon	22	Merton
7	Ealing	23	Newham
8	Enfield	24	Redbridge
9	Greenwich	25	Richmond upon Thames
10	Hackney	26	Southwark
11	Hammersmith and Fulham	27	Sutton
12	Haringey	28	Tower Hamlets
13	Harrow	29	Waltham Forest
14	Havering	30	Wandsworth
15	Hillingdon	31	Westminster

Figure 1: 32 London boroughs

Then we pass the DataFrame to GeoPy to retrieve geographical coordinates of boroughs.

Data Retrieval

Retrieving geographical coordinates of London boroughs

	Borough	Longitude	Latitude		Borough	Longitude	Latitude
0	Barking and Dagenham	0.150504	51.554117	16	Hounslow	-0.361347	51.468613
1	Barnet	-0.200226	51.653090	17	Islington	-0.099905	51.538429
2	Bexley	-82.936864	39.969238	18	Kensington and Chelsea	-0.168220	51.487542
3	Brent	-87.164718	32.937346	19	Kingston upon Thames	-0.306262	51.409627
4	Bromley	0.014814	51.402805	20	Lambeth	-0.117287	51.501301
5	Camden	-75.119891	39.944840	21	Lewisham	-0.010133	51.462432
6	Croydon	-0.101957	51.371305	22	Merton	-0.188099	51.410803
7	Ealing	-0.305195	51.512655	23	Newham	0.029318	51.530000
8	Enfield	-0.081018	51.652085	24	Redbridge	0.045410	51.576320
9	Greenwich	-0.004542	51.482084	25	Richmond upon Thames	-0.305720	51.440372
10	Hackney	-0.049362	51.543240	26	Southwark	-0.103458	51.502922
11	Hammersmith and Fulham	-0.223640	51.492038	27	Sutton	-100.643236	30.567295
12	Haringey	-0.105410	51.587930	28	Tower Hamlets	1.298669	51.128863
13	Harrow	-0.337316	51.596827	29	Waltham Forest	-71.235800	42.375640
14	Havering	-2.337475	51.004361	30	Wandsworth	-0.193261	51.457027
15	Hillingdon	-0.448335	51.542519	31	Westminster	-0.126540	51.500444

Figure 2: Geographical coordinates of London's boroughs

Looking closely at the DataFrame, we can clearly notice that there are some incorrect geographical coordinates (e.g. Bexley's coordinates)

Data Retrieval

Retrieving geographical coordinates of London boroughs

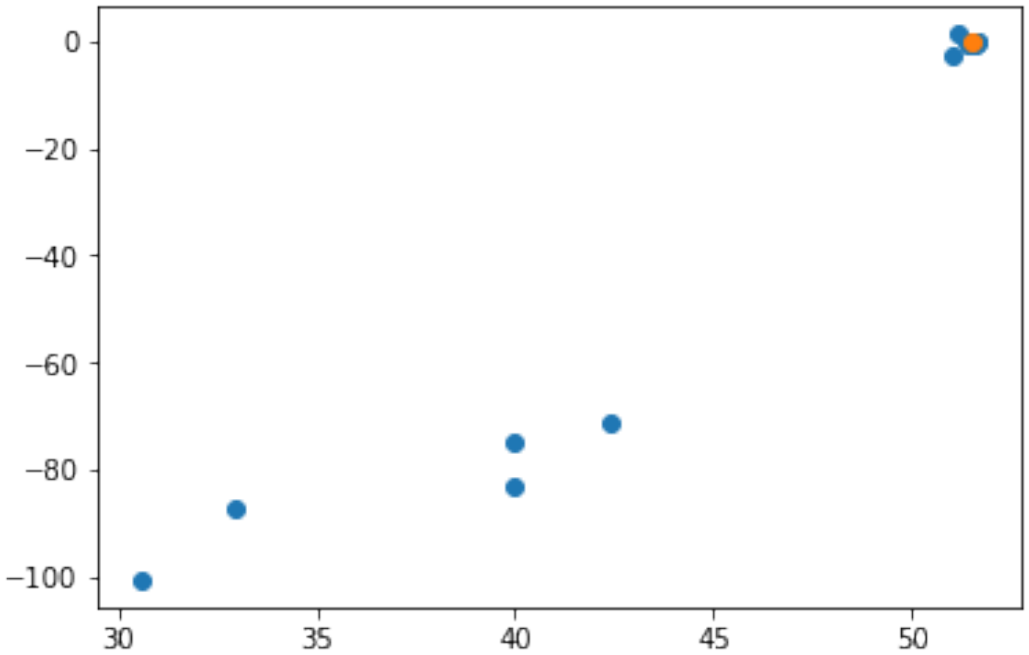

Figure 3: Scatter plot of geographical coordinates of London's boroughs

There is a dense cluster surrounding London's geographical coordinates and five outliers. It implies that the location dataset has some incorrect values. Therefore, we need to modify the input in order to retrieve the correct values.

Data Retrieval

Retrieving geographical coordinates of London boroughs

We modify the DataFrame by adding 'London', so we can retrieve the correct values.

	Borough		Borough
0	Barking and Dagenham, London	16	Hounslow, London
1	Barnet, London	17	Islington, London
2	Bexley, London	18	Kensington and Chelsea, London
3	Brent, London	19	Kingston upon Thames, London
4	Bromley, London	20	Lambeth, London
5	Camden, London	21	Lewisham, London
6	Croydon, London	22	Merton, London
7	Ealing, London	23	Newham, London
8	Enfield, London	24	Redbridge, London
9	Greenwich, London	25	Richmond upon Thames, London
10	Hackney, London	26	Southwark, London
11	Hammersmith and Fulham, London	27	Sutton, London
12	Haringey, London	28	Tower Hamlets, London
13	Harrow, London	29	Waltham Forest, London
14	Havering, London	30	Wandsworth, London
15	Hillingdon, London	31	Westminster, London

Figure 4: 32 London boroughs

Data Retrieval

Retrieving geographical coordinates of London boroughs

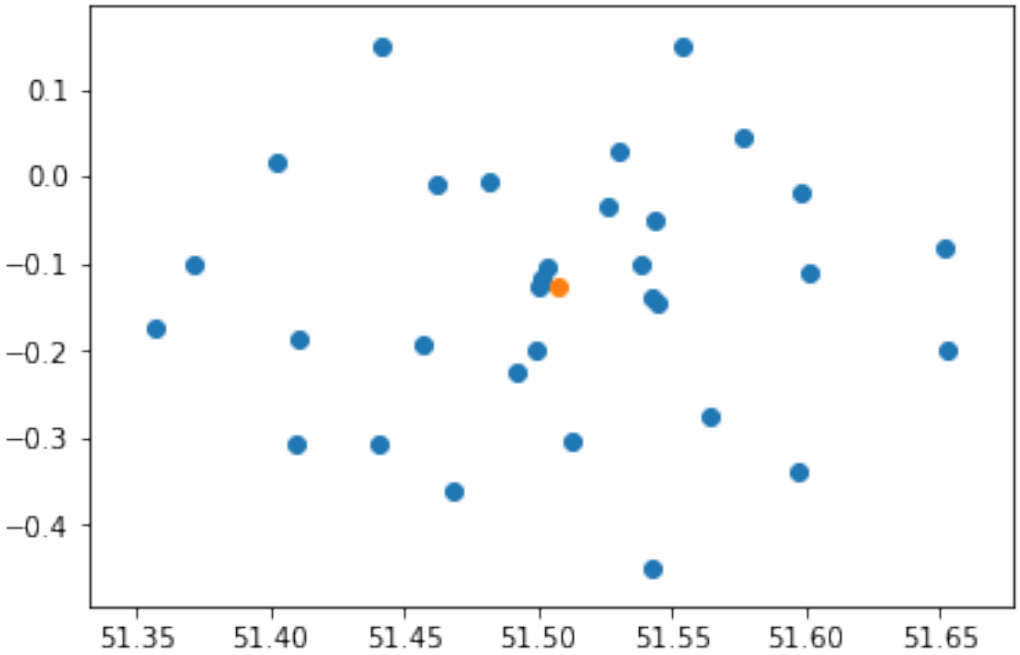

Figure 5: The scatter plot after modifying the borough names

The scatter plot matches the shape of London. We will visualize these geographical coordinates using Folium.

Data Retrieval

Retrieving geographical coordinates of London boroughs

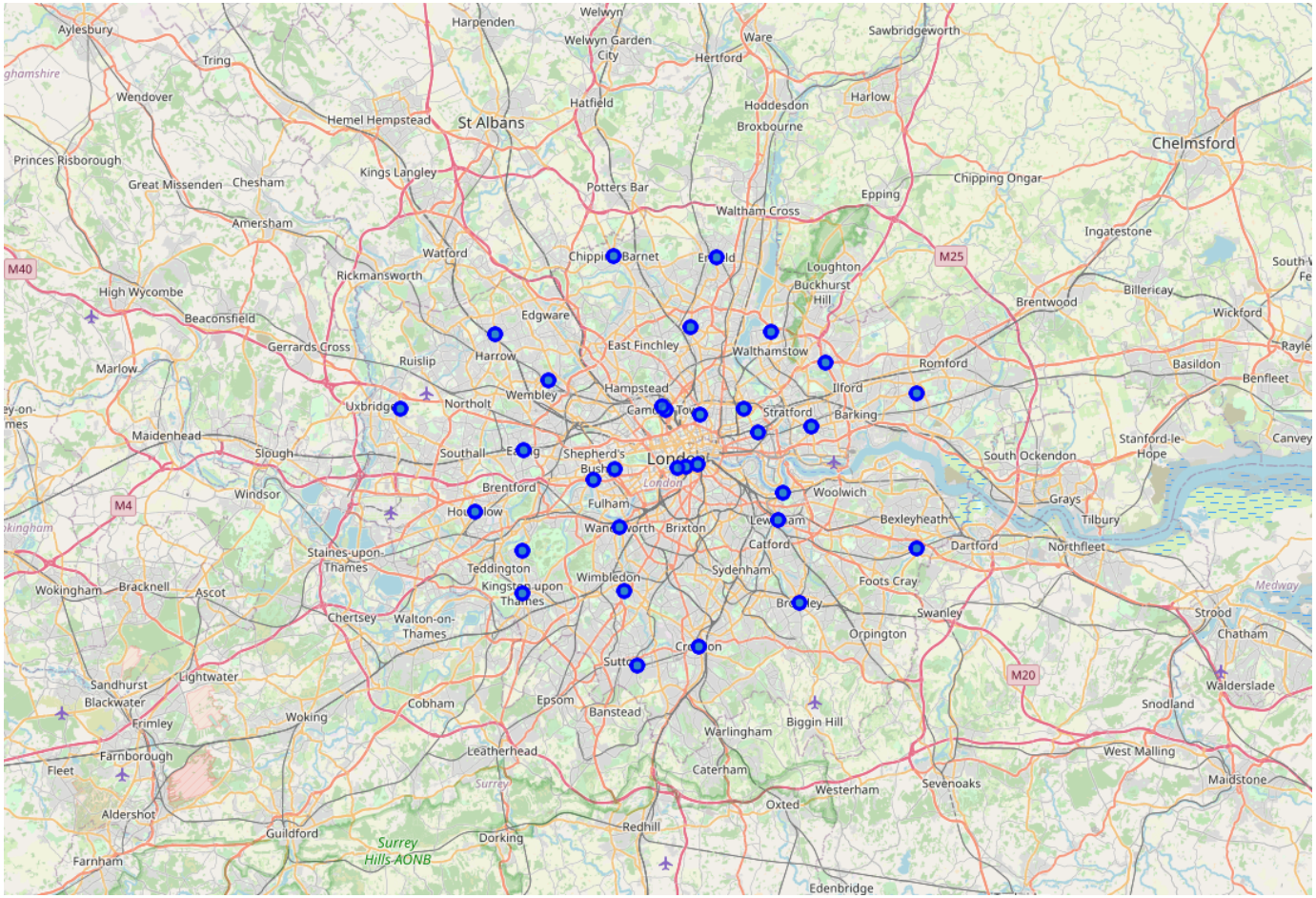

Figure 6: Visualizing geographical coordinates of London boroughs using Folium

Data Retrieval

Retrieving popular restaurants in London boroughs

- Next, we make an API call to Foursquare to retrieve popular restaurant near each London borough's geographical coordinates.
- We set *radius* = 2000, *limit* = 100 and *query* = "*restaurant*"
- After extracting features, we store the result in a DataFrame.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Barking and Dagenham, London	51.554117	0.150504	Lara Grill	51.562445	0.147178	Turkish Restaurant
1	Barking and Dagenham, London	51.554117	0.150504	The Pipe Major	51.545795	0.165834	Restaurant
2	Barking and Dagenham, London	51.554117	0.150504	Heath Pie Shop	51.560414	0.147655	Diner
3	Barking and Dagenham, London	51.554117	0.150504	New China Gold	51.561269	0.141839	Chinese Restaurant
4	Barking and Dagenham, London	51.554117	0.150504	Papa John's Pizza	51.542671	0.147628	Pizza Place
5	Barking and Dagenham, London	51.554117	0.150504	Gunay's Cafe	51.546347	0.165795	Café
6	Barking and Dagenham, London	51.554117	0.150504	McDonald's	51.565406	0.145384	Fast Food Restaurant
7	Barnet, London	51.653090	-0.200226	Joie de Vie	51.653659	-0.201288	Bakery
8	Barnet, London	51.653090	-0.200226	Spizzico	51.656600	-0.201802	Italian Restaurant
9	Barnet, London	51.653090	-0.200226	Potters Pantry	51.651802	-0.179730	Café
10	Barnet, London	51.653090	-0.200226	PizzaExpress	51.658073	-0.200862	Pizza Place
11	Barnet, London	51.653090	-0.200226	Fresh Fry Fish 'n' Chips	51.646451	-0.187072	Fish & Chips Shop
12	Barnet, London	51.653090	-0.200226	Dudley's Pancake House	51.652965	-0.199625	Restaurant
13	Barnet, London	51.653090	-0.200226	Subway	51.651235	-0.197072	Restaurant
14	Barnet, London	51.653090	-0.200226	Domino's Pizza	51.652675	-0.198837	Pizza Place

Figure 7: Popular restaurants near each borough's geographical coordinates

Exploratory Data Analysis

Visualizing the number of Café in each borough

From the last DataFrame, we extracts row in which Venue Category is Café, then count the number of Italian restaurants in each borough

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Barking and Dagenham, London	1	1	1	1	1	1
Barnet, London	2	2	2	2	2	2
Brent, London	6	6	6	6	6	6
Bromley, London	5	5	5	5	5	5
Camden, London	14	14	14	14	14	14
Croydon, London	6	6	6	6	6	6
Ealing, London	8	8	8	8	8	8
Enfield, London	5	5	5	5	5	5
Greenwich, London	20	20	20	20	20	20
Hackney, London	19	19	19	19	19	19

Figure 8: The number of Café in each borough

Exploratory Data Analysis

Visualizing the number of Café for each borough

Then, we plot a bar chart displaying the number of Italian restaurants in each borough.

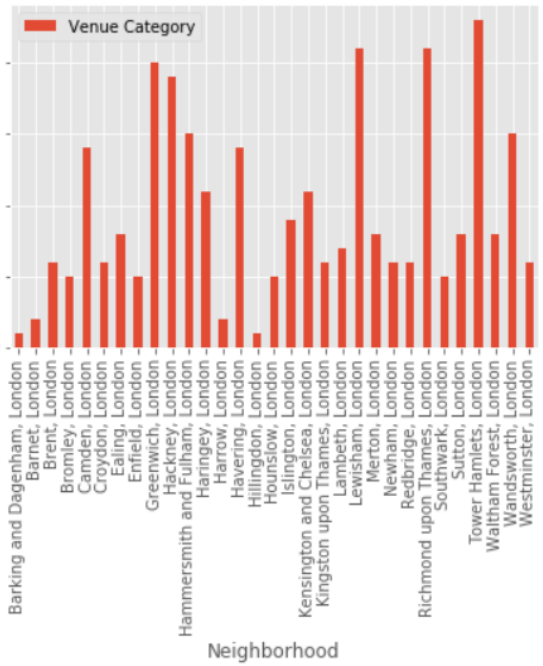

Figure 9: The number of Café in each borough

Exploratory Data Analysis

Displaying top 10 restaurants in each borough

Next, we will use one hot encoding technique to convert categorical data (Venue Category) to numerical data for further analysis.

	Borough	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	BBQ Joint	...
0	Barking and Dagenham, London	0	0	0	0	0	0	0	0	0	...
1	Barking and Dagenham, London	0	0	0	0	0	0	0	0	0	...
2	Barking and Dagenham, London	0	0	0	0	0	0	0	0	0	...
3	Barking and Dagenham, London	0	0	0	0	0	0	0	0	0	...
4	Barking and Dagenham, London	0	0	0	0	0	0	0	0	0	...

Figure 10: Converting Venue Category to numerical data

Exploratory Data Analysis

Displaying top 10 restaurants in each borough

Then we group rows by 'Borough' and calculate the frequency of each type of restaurants in each borough.

Borough	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	BBQ Joint	Bagel Shop	...
Barking and Dagenham, London	0.0	0.0	0.000000	0.0	0.00	0.000000	0.0	0.0	0.0	0.00	...
Barnet, London	0.0	0.0	0.000000	0.0	0.00	0.000000	0.0	0.0	0.0	0.00	...
Bexley, London	0.0	0.0	0.068966	0.0	0.00	0.000000	0.0	0.0	0.0	0.00	...
Brent, London	0.0	0.0	0.025641	0.0	0.00	0.025641	0.0	0.0	0.0	0.00	...
Bromley, London	0.0	0.0	0.021739	0.0	0.00	0.043478	0.0	0.0	0.0	0.00	...

Figure 11: Frequency of each type of restaurants in each borough

We will use this data to construct a table displaying top 10 restaurants in each borough and use to cluster boroughs using K-means clustering algorithm.

Exploratory Data Analysis

Displaying top 10 restaurants in each borough

Finally, we construct a table displaying top 10 restaurants in each borough.

	Neighborhood	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	Barking and Dagenham, London	Chinese Restaurant	Restaurant	Turkish Restaurant	Diner	Café	Fast Food Restaurant	Pizza Place	Eastern European Restaurant	Currywurst Joint	Deli / Bodega
1	Barnet, London	Restaurant	Pizza Place	Chinese Restaurant	Café	Indian Restaurant	Fast Food Restaurant	Italian Restaurant	Modern European Restaurant	Seafood Restaurant	Bakery
2	Bexley, London	Fast Food Restaurant	Chinese Restaurant	Italian Restaurant	American Restaurant	Greek Restaurant	Steakhouse	Restaurant	Burger Joint	Portuguese Restaurant	Mediterranean Restaurant
3	Brent, London	Indian Restaurant	Restaurant	Sandwich Place	Fast Food Restaurant	Café	Pizza Place	Bakery	Chinese Restaurant	Portuguese Restaurant	Italian Restaurant
4	Bromley, London	Indian Restaurant	Pizza Place	Café	Sandwich Place	Italian Restaurant	Fast Food Restaurant	Sushi Restaurant	Bakery	Chinese Restaurant	Asian Restaurant
5	Camden, London	Café	Burger Joint	Pizza Place	Italian Restaurant	Bakery	Deli / Bodega	Vegetarian / Vegan Restaurant	Middle Eastern Restaurant	French Restaurant	Gastropub
6	Croydon, London	Fast Food Restaurant	Sandwich Place	Indian Restaurant	Pizza Place	Café	Bakery	Portuguese Restaurant	Italian Restaurant	Sushi Restaurant	Asian Restaurant
7	Ealing, London	Café	Pizza Place	Italian Restaurant	Indian Restaurant	Bakery	Fast Food Restaurant	Restaurant	Japanese Restaurant	Burger Joint	Diner
8	Enfield, London	Pizza Place	Café	Indian Restaurant	Fast Food Restaurant	Portuguese Restaurant	Restaurant	Fish & Chips Shop	English Restaurant	Bakery	Italian Restaurant
9	Greenwich, London	Café	Indian Restaurant	Pizza Place	Vietnamese Restaurant	Japanese Restaurant	Chinese Restaurant	Burger Joint	Bakery	Restaurant	Italian Restaurant

Figure 12: Top 10 restaurants in each borough

Modeling & Result

Using K-means to cluster boroughs

Next, we will cluster London borough using the mentioned data. We set the number of clusters to 3.

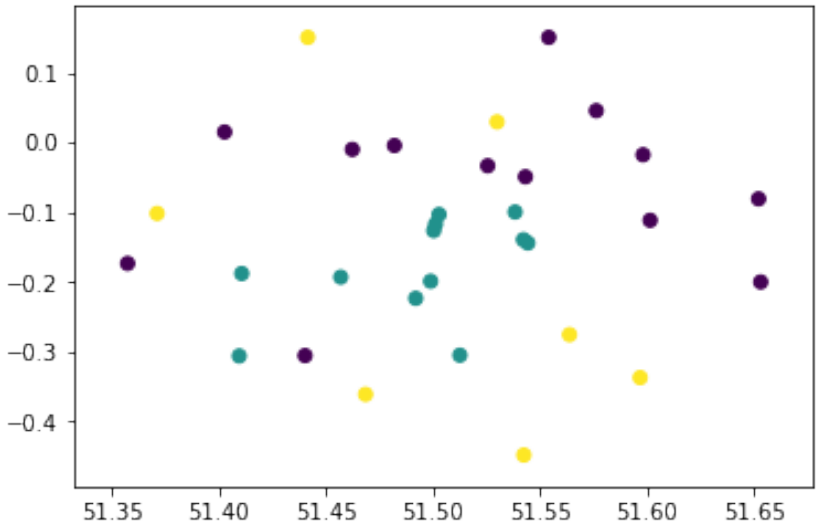

Figure 13: Scatter plot of geographical coordinates of clustered boroughs

Modeling & Result

Using K-means to cluster boroughs

Finally, we visualize geographical coordinates of clustered boroughs using folium.

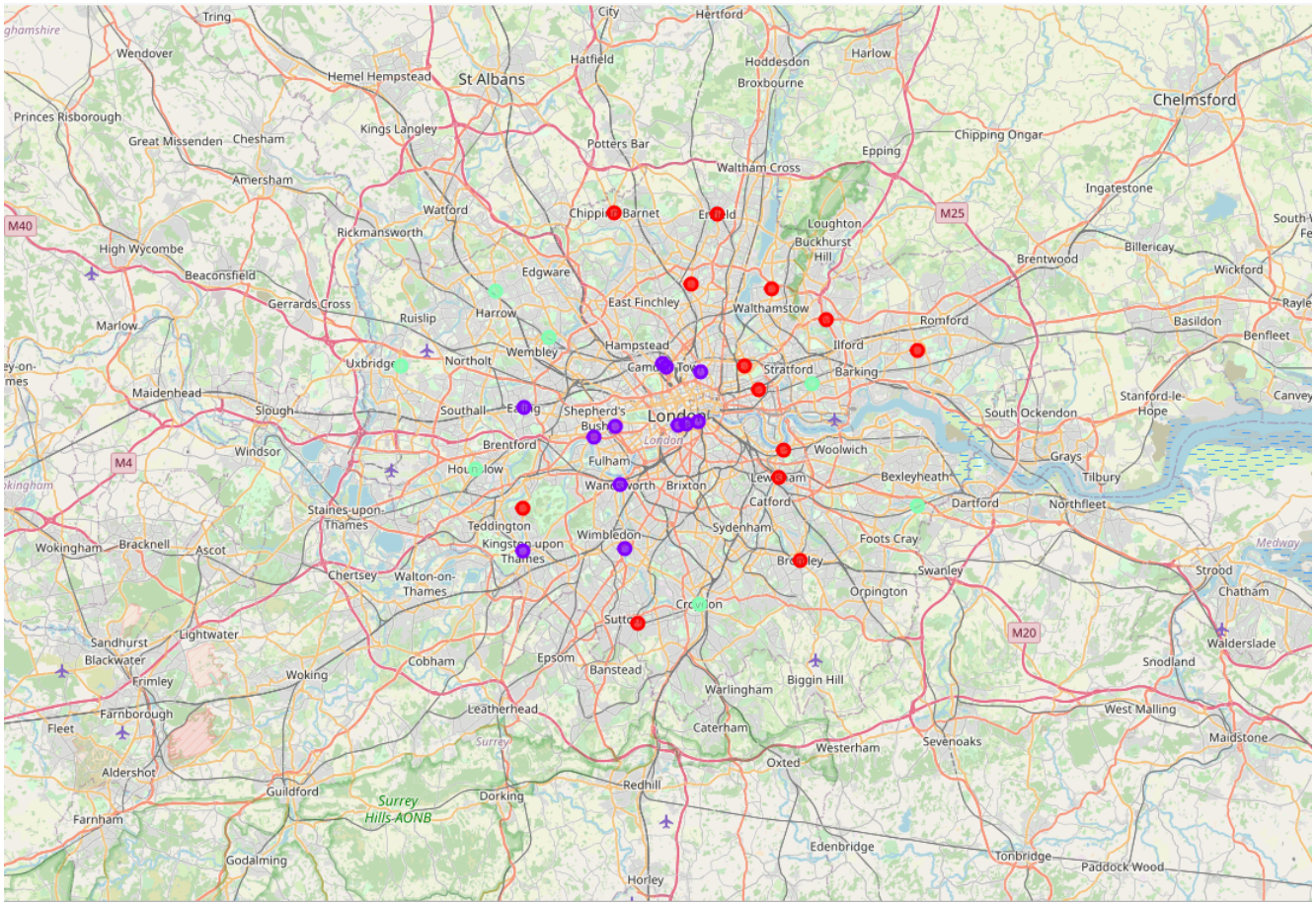

Figure 14: Visualize geographical coordinates of clustered boroughs using Folium

Cluster 0:

	Neighborhood	Cluster	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	Barking and Dagenham, London	0	Chinese Restaurant	Restaurant	Turkish Restaurant	Diner	Café	Fast Food Restaurant	Pizza Place	Eastern European Restaurant	Currywurst Joint	Deli / Bodega
1	Barnet, London	0	Restaurant	Pizza Place	Chinese Restaurant	Café	Indian Restaurant	Fast Food Restaurant	Italian Restaurant	Modern European Restaurant	Seafood Restaurant	Bakery
4	Bromley, London	0	Indian Restaurant	Pizza Place	Café	Sandwich Place	Italian Restaurant	Fast Food Restaurant	Sushi Restaurant	Bakery	Chinese Restaurant	Asian Restaurant
8	Enfield, London	0	Pizza Place	Café	Indian Restaurant	Fast Food Restaurant	Portuguese Restaurant	Restaurant	Fish & Chips Shop	English Restaurant	Bakery	Italian Restaurant
9	Greenwich, London	0	Café	Indian Restaurant	Pizza Place	Vietnamese Restaurant	Japanese Restaurant	Chinese Restaurant	Burger Joint	Bakery	Restaurant	Italian Restaurant
10	Hackney, London	0	Café	Restaurant	Bakery	Pizza Place	Italian Restaurant	Vegetarian / Vegan Restaurant	Turkish Restaurant	Gastropub	Deli / Bodega	Vietnamese Restaurant
12	Haringey, London	0	Café	Indian Restaurant	Fast Food Restaurant	Turkish Restaurant	Greek Restaurant	Mediterranean Restaurant	Sandwich Place	Italian Restaurant	Middle Eastern Restaurant	Bakery
21	Lewisham, London	0	Café	Restaurant	Indian Restaurant	Fish & Chips Shop	Gastropub	Fast Food Restaurant	Bakery	Food Truck	Pizza Place	Italian Restaurant
24	Redbridge, London	0	Pizza Place	Café	English Restaurant	Bakery	Indian Restaurant	Restaurant	Chinese Restaurant	Thai Restaurant	Fast Food Restaurant	Italian Restaurant
25	Richmond upon Thames, London	0	Café	Italian Restaurant	Indian Restaurant	Bakery	Restaurant	French Restaurant	Gastropub	Thai Restaurant	Deli / Bodega	Sandwich Place
27	Sutton, London	0	Café	Bakery	Italian Restaurant	Pizza Place	Indian Restaurant	Fast Food Restaurant	Restaurant	Irish Pub	Chinese Restaurant	Portuguese Restaurant
28	Tower Hamlets, London	0	Café	Pizza Place	Indian Restaurant	Italian Restaurant	Bakery	Fast Food Restaurant	Burger Joint	Fried Chicken Joint	Restaurant	Turkish Restaurant
29	Waltham Forest, London	0	Café	Turkish Restaurant	Pizza Place	Restaurant	Sandwich Place	Italian Restaurant	Fast Food Restaurant	Indian Restaurant	Bakery	Halal Restaurant

Figure 15: Cluster 0

Most popular venues: Cafe, Indian restaurants, Pizza places

Cluster 1:

	Neighborhood	Cluster	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
5	Camden, London	1	Café	Burger Joint	Pizza Place	Italian Restaurant	Bakery	Deli / Bodega	Vegetarian / Vegan Restaurant	Middle Eastern Restaurant	French Restaurant	Gastropub
7	Ealing, London	1	Café	Pizza Place	Italian Restaurant	Indian Restaurant	Bakery	Fast Food Restaurant	Restaurant	Japanese Restaurant	Burger Joint	Diner
11	Hammersmith and Fulham, London	1	Café	Italian Restaurant	Thai Restaurant	Pizza Place	Persian Restaurant	Sandwich Place	Gastropub	French Restaurant	Middle Eastern Restaurant	Indian Restaurant
14	Havering, London	1	Café	Italian Restaurant	Pizza Place	Burger Joint	Bakery	Deli / Bodega	Middle Eastern Restaurant	French Restaurant	Indian Restaurant	Greek Restaurant
17	Islington, London	1	Café	French Restaurant	Burger Joint	Bakery	Pizza Place	Gastropub	Middle Eastern Restaurant	Mediterranean Restaurant	Japanese Restaurant	Breakfast Spot
18	Kensington and Chelsea, London	1	Café	Italian Restaurant	Restaurant	Pizza Place	Bakery	Burger Joint	Gastropub	Japanese Restaurant	French Restaurant	Mediterranean Restaurant
19	Kingston upon Thames, London	1	Café	Thai Restaurant	Burger Joint	Italian Restaurant	Gastropub	Japanese Restaurant	Indian Restaurant	French Restaurant	Sushi Restaurant	German Restaurant
20	Lambeth, London	1	Bakery	Indian Restaurant	Café	Burger Joint	Japanese Restaurant	Steakhouse	French Restaurant	Sushi Restaurant	Seafood Restaurant	Restaurant
22	Merton, London	1	Café	Indian Restaurant	Fast Food Restaurant	Sushi Restaurant	Italian Restaurant	Thai Restaurant	Pizza Place	Burger Joint	Restaurant	Sandwich Place
26	Southwark, London	1	Restaurant	Italian Restaurant	Bakery	Steakhouse	Café	French Restaurant	Burger Joint	Pizza Place	Portuguese Restaurant	Seafood Restaurant
30	Wandsworth, London	1	Café	Italian Restaurant	Thai Restaurant	Pizza Place	Burger Joint	Bakery	Indian Restaurant	Asian Restaurant	Restaurant	Sandwich Place
31	Westminster, London	1	Indian Restaurant	Bakery	Steakhouse	Café	Burger Joint	Italian Restaurant	French Restaurant	Seafood Restaurant	Sushi Restaurant	Pizza Place

Figure 15: Cluster 1

Most popular venues: Cafe, Indian restaurants, Italian restaurants

Cluster 2:

	Neighborhood	Cluster	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
2	Bexley, London	2	Fast Food Restaurant	Chinese Restaurant	Italian Restaurant	American Restaurant	Greek Restaurant	Steakhouse	Restaurant	Burger Joint	Portuguese Restaurant	Mediterranean Restaurant
3	Brent, London	2	Indian Restaurant	Restaurant	Sandwich Place	Fast Food Restaurant	Café	Pizza Place	Bakery	Chinese Restaurant	Portuguese Restaurant	Italian Restaurant
6	Croydon, London	2	Fast Food Restaurant	Sandwich Place	Indian Restaurant	Pizza Place	Café	Bakery	Portuguese Restaurant	Italian Restaurant	Sushi Restaurant	Asian Restaurant
13	Harrow, London	2	Indian Restaurant	Fast Food Restaurant	Pizza Place	Sandwich Place	Chinese Restaurant	Fish & Chips Shop	Steakhouse	Middle Eastern Restaurant	Donut Shop	Thai Restaurant
15	Hillingdon, London	2	Fast Food Restaurant	Pizza Place	Indian Restaurant	Italian Restaurant	Chinese Restaurant	Burger Joint	Fish & Chips Shop	Mexican Restaurant	Middle Eastern Restaurant	Portuguese Restaurant
16	Hounslow, London	2	Indian Restaurant	Fast Food Restaurant	Café	Pizza Place	Restaurant	Chinese Restaurant	Asian Restaurant	Sandwich Place	Bakery	Middle Eastern Restaurant
23	Newham, London	2	Fast Food Restaurant	Indian Restaurant	Café	Bakery	Fish & Chips Shop	Sandwich Place	Pizza Place	Wings Joint	Asian Restaurant	Breakfast Spot

Figure 15: Cluster 2

Most popular venues: Fastfood restaurants, Indian restaurants

Modeling & Result

Discussion

- The most popular restaurants in boroughs in Cluster 0 are Cafe, Indian restaurants, Pizza places.
- The most popular restaurants in boroughs in Cluster 1 are Cafe, Indian restaurants, Italian restaurants.
- The most popular restaurants in boroughs in Cluster 2 are Fastfood restaurants, Indian restaurants.
- The best boroughs to open a new coffee shop are boroughs in cluster 3 due to much less competition from other coffee shops.