

Best London neighbourhood for a coffee shop

Alberto Lapedriza, September 2019

Table of Contents

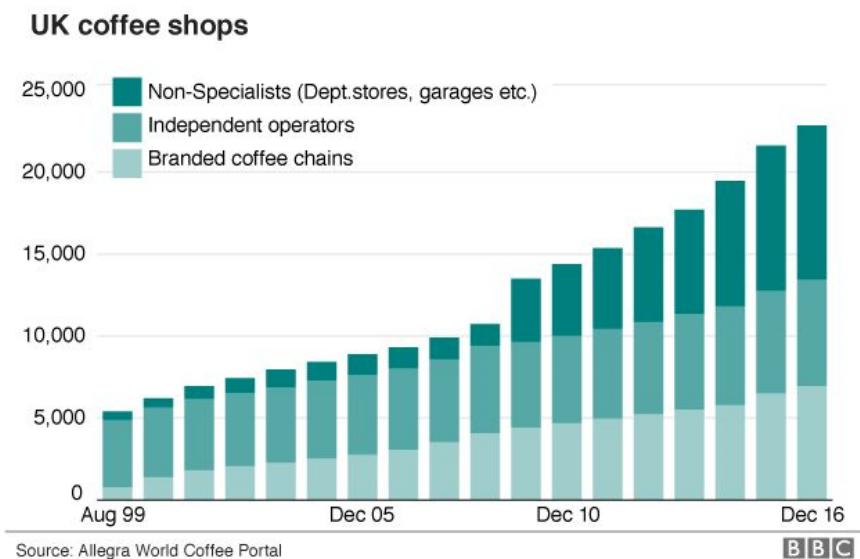
- 1) Introduction & Business Problem
- 2) Data
- 3) Methodology
- 4) Analysis
- 5) Results and Discussion
- 6) Conclusion

Introduction & Business Problem

UK's coffee consumption has soared to 95 million cups a day in 2018, up from 70 million in 2008. That's an increase of 25 million over the last 10 years. This coffee popularity has translated in a big increase in coffee shops or cafés in London, the UK capital and one of the world's global cities. Most of these popular cafés are big chains such as the USA's Starbucks or the UK's own chains Costa Coffee or Café Nero.

But Londoners are starting to get tired of having the same chain coffee every day and are starting to look into more independent and speciality coffee shops, where they take great care of each coffee cup and they use single origin, ethically sourced and organic coffee beans that are roasted locally by artisan roasters.

London's obsession with coffee is showing no signs of slowing. Across the city, cafés are constantly popping up, serving up perfectly executed flat whites, espressos and cold-drip Americanos to the masses. As the following graph from the BBC demonstrates, the number of cafés and other hospitality venues selling coffee has increased significantly since the early 2000s.



Aim

The aim of this project is to find out which neighbourhood in London would be the best to open a new café.

Target audience

The target audience would be an entrepreneur or group of entrepreneurs that are looking to set up their new independent café in London. This project would help them to find out which are the neighbourhoods with more and less cafés so they can open their café in the less saturated neighbourhood.

Data

This section will describe the data that will be used to solve the problem.

Dataset 1

London is divided in 32 London boroughs and "the City of London" which is the central part of London or downtown. Each borough then has several neighbourhoods (although there are some neighbourhoods that may expand several boroughs).

Our London neighbourhood and borough data comes from the Wikipedia page "List of areas of London": https://en.wikipedia.org/wiki/List_of_areas_of_London

The data is presented in a Wikipedia table and we have transformed it to a pandas data frame for our analysis. In order to narrow down the neighbourhoods to those that are actually part of London and not the nearby counties we removed all those neighbourhoods where the 'post town' was not London. The following table shows the top 5 rows of our data frame.

	Neighbourhood	Borough
0	Abbey Wood	Bexley, Greenwich
1	Acton	Ealing, Hammersmith and Fulham
2	Aldgate	City
3	Aldwych	Westminster
4	Anerley	Bromley

Dataset 2

To obtain the coordinate data of the London neighbourhoods, the Geocoder package is used to get the latitude and longitude for each neighbourhood. Both of these are needed for the Foursquare API.

The Geocoder location data will be used to enrich the data frame of London neighbourhoods obtained from Wikipedia above.

	Neighbourhood	Borough	Latitude	Longitude
0	Abbey Wood	Bexley, Greenwich	51.492450	0.121270
1	Acton	Ealing, Hammersmith and Fulham	51.513240	-0.267460
2	Aldgate	City	51.513308	-0.077762
3	Aldwych	Westminster	51.513307	-0.117092
4	Anerley	Bromley	51.412330	-0.065390

Dataset 3

The Foursquare API will be used to search for a specific venue category (in our case cafés or coffee shops) for the geographical location data for each London neighbourhood.

We will use the explore Foursquare API and the parameter 'section'=coffee, which can have one of the following values: food, drinks, coffee, shops, arts, outdoors, sights, trending, nextVenues , to limit the venues that we found to those that serve coffee (these may include cafés and coffee shops but also restaurants or ice cream parlours).

Methodology

This section describes all the analysis that was performed in the project.

First of all, we created a function to explore the venues of all the neighbourhoods in London using the Foursquare API. The following table shows the first 5 rows of the resulting dataframe:

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbey Wood	51.492450	0.121270	Bean @ Work	51.491172	0.120649	Coffee Shop
1	Abbey Wood	51.492450	0.121270	Abbey Cafe	51.489754	0.120822	Café
2	Aldgate	51.513308	-0.077762	The Association	51.513733	-0.079132	Coffee Shop
3	Aldgate	51.513308	-0.077762	Kahaila Aldgate	51.514046	-0.077001	Coffee Shop
4	Aldgate	51.513308	-0.077762	Benk + Bo	51.515731	-0.075875	Bakery

This dataframe contains all venues that serve coffee (including bakeries or restaurants), but we only want to focus on Cafés or Coffee Shops for the scope of this project. Hence we removed from our table all the venues where the category is not 'Café' or 'Coffee Shop'. These are the first 5 rows of the resulting dataframe:

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbey Wood	51.492450	0.121270	Bean @ Work	51.491172	0.120649	Coffee Shop
1	Abbey Wood	51.492450	0.121270	Abbey Cafe	51.489754	0.120822	Café
2	Aldgate	51.513308	-0.077762	The Association	51.513733	-0.079132	Coffee Shop
3	Aldgate	51.513308	-0.077762	Kahaila Aldgate	51.514046	-0.077001	Coffee Shop
4	Aldgate	51.513308	-0.077762	Notes Coffee Roaster & Wine Bar	51.514643	-0.080671	Coffee Shop

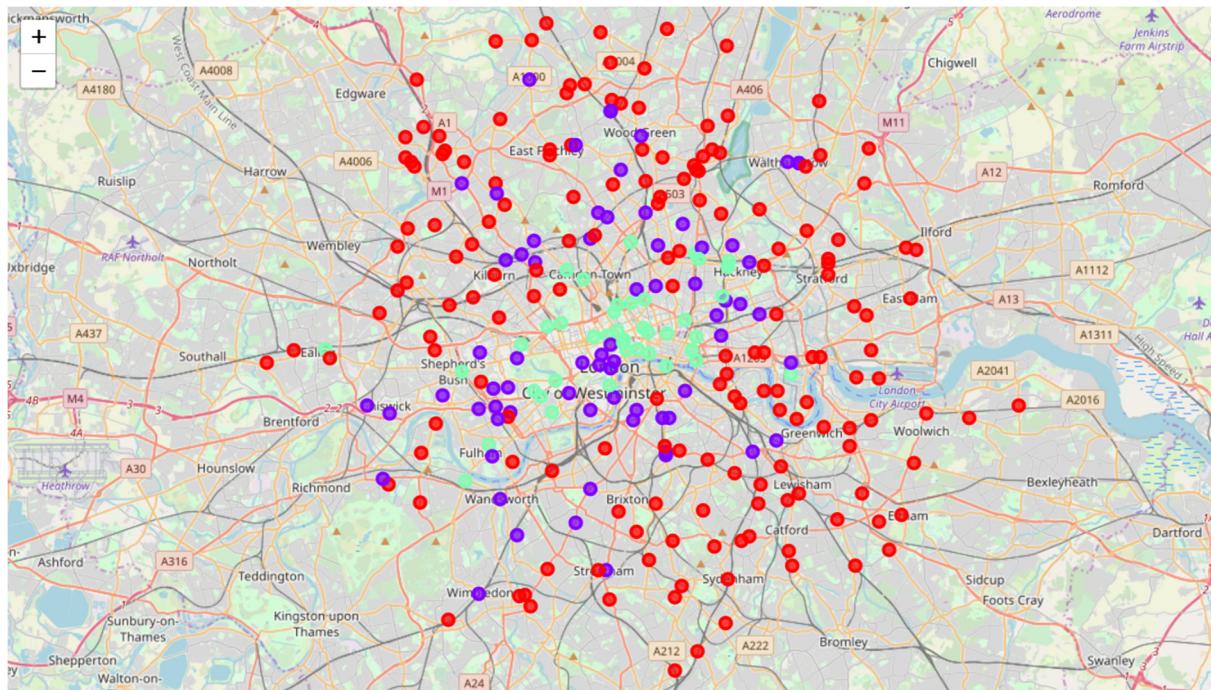
Then we created a table to show how many cafés are in each neighbourhood. The following table shows the top 10 neighbourhoods with more coffee shops:

Neighbourhood	Count
Pentonville	29
Lisson Grove	26
Paddington	26
Aldgate	24
Oval	23
Barbican	23
Islington	22
Shepherd's Bush	22
Bloomsbury	22
Bedford Park	22

The neighbourhoods were clustered together based on the number of cafes using the k-means algorithm. This algorithm grouped the neighbourhoods in three different clusters (0, 1 and 2). The following table shows the first 5 rows of the resulting data frame with each neighbourhood assigned to a cluster (see 'Cluster Labels' column):

Cluster Labels	Neighbourhood	Count	Borough	Latitude	Longitude
0	Abbey Wood	2	Bexley, Greenwich	51.492450	0.121270
1	Aldgate	24	City	51.513308	-0.077762
2	Aldwych	20	Westminster	51.513307	-0.117092
3	Angel	5	Islington	51.500500	-0.060510
4	Archway	12	Islington	51.565746	-0.134917

Finally the neighbourhoods were plotted on a London map and coloured based on their cluster: red for cluster 0, purple for cluster 1 and light green for cluster 3:



Results

Throughout our analysis we have selected 299 neighbourhoods in London belonging to one of each of the 32 London boroughs. We have removed those neighbourhoods that are on the outskirts to reduce the number of neighbourhoods to explore through the Foursquare API, as it has limited number of free requests.

The Foursquare API was used to explore venues that belong to the 'coffee' section parameter. This helped us to limit the venues retrieved from the API to those that sell coffee. The number of venues obtained was: 2833.

Those venues included cafés and coffee shops but also restaurants, ice cream parlours or any other venues serving coffee. Hence we restricted our results to those with venue category of 'café' and 'coffee shop'. The number of venues remaining were: 2215.

We obtained the number of cafés per neighbourhood. The top 10 neighbourhoods with more cafés ranged from Pentonville with 29 to Bedford Park with 22. Pentonville is a neighbourhood on the northern fringe of Central London which is located near the international transportation hub of Kings Cross-St Pancras train station. Pentonville is also very close to the Angel area famous for its restaurants and shops. Both of these factors explain the high number of coffee shops in Pentonville.

	Neighbourhood	Count
184	Pentonville	29
144	Lisson Grove	26
178	Paddington	26
1	Aldgate	24
177	Oval	23
8	Barbican	23
126	Islington	22
200	Shepherd's Bush	22
22	Bloomsbury	22
13	Bedford Park	22

On the other hand, all the bottom 10 neighbourhoods with less number of cafes in London have 1 cafe. These neighbourhoods such as Hackney Marshes, Totteridge or Neasden are all residential neighbourhoods in the outer rim of London, where people that work in central London live; hence this explains the lack of coffee shops.

	Neighbourhood	Count
174	Oakleigh Park	1
145	Little Ilford	1
43	Catford	1
103	Hackney Marshes	1
240	Totteridge	1
155	Middle Park	1
66	Dartford	1
234	The Hyde	1
164	Neasden	1
122	Honor Oak	1

Our analysis used Machine Learning to cluster the different London neighbourhoods based on the number of cafes in each one, in order to find those neighbourhoods with more and less cafes.

The K-means algorithm grouped the neighbourhoods in 3 clusters:

- Cluster 0: those neighbourhoods with a minimum number of cafes of 1 and a maximum of 7.

count	168.000000
mean	3.904762
std	1.720737
min	1.000000
25%	3.000000
50%	4.000000
75%	5.000000
max	7.000000

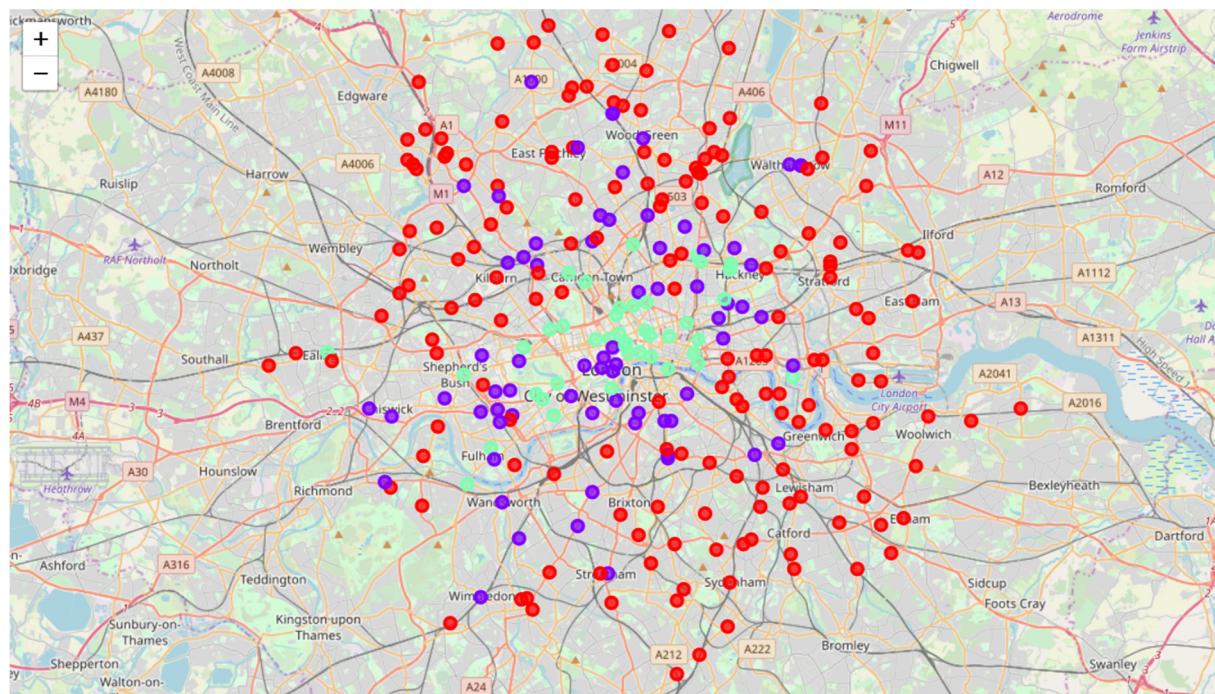
- Cluster 1: those neighbourhoods with a minimum number of cafes of 8 and a maximum of 15.

count	70.000000
mean	10.457143
std	2.012178
min	8.000000
25%	9.000000
50%	10.000000
75%	12.000000
max	15.000000

- Cluster 2: those neighbourhoods with a minimum number of cafes of 16 and a maximum of 29.

count	42.000000
mean	20.119048
std	2.864436
min	16.000000
25%	18.000000
50%	20.000000
75%	21.750000
max	29.000000

We plotted the location of the three clusters on the map and that helped us to understand the meaning of each of the clusters.



Cluster 0 includes all those neighbourhoods with a low number of coffee shops. These tend to be in the outskirts of London, and are mainly residential areas. People living here commute into London for both work and leisure.

Cluster 1 includes all those neighbourhoods with a medium number of coffee shops. These tend to be closer to central London, and are a mixture of residential, work and leisure areas.

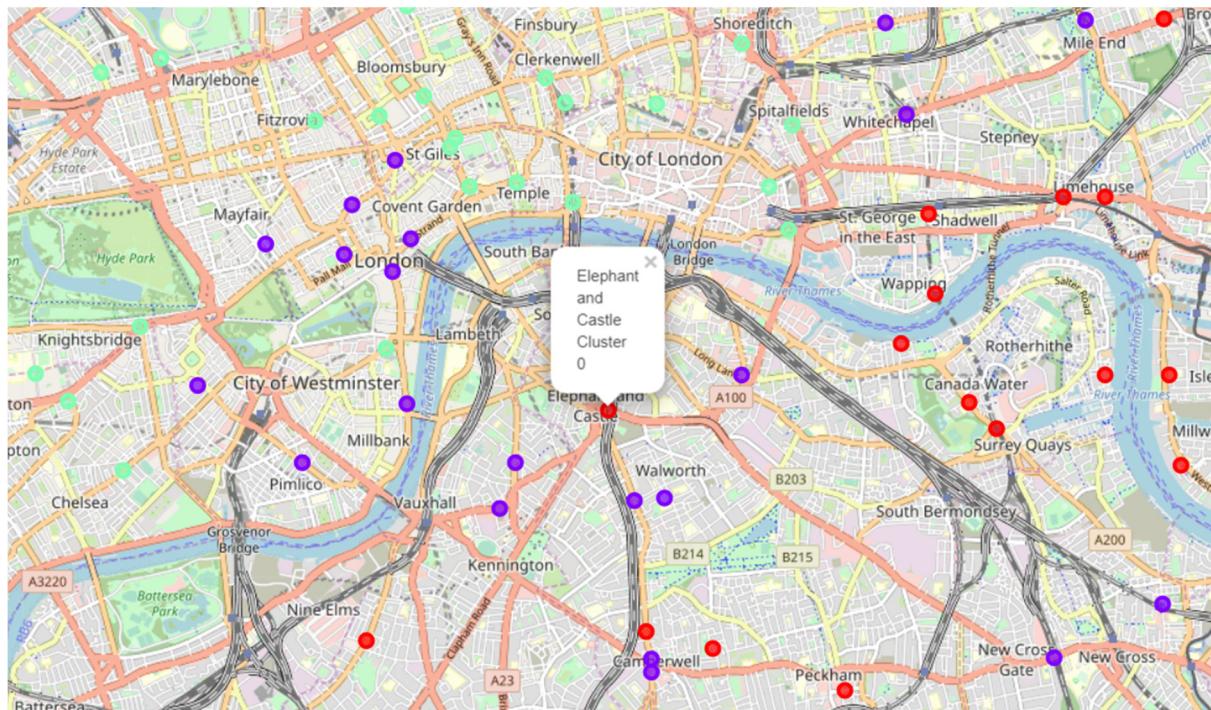
Cluster 2 includes all those neighbourhoods with a high number of coffee shops. These neighbourhoods are part of central London, and their main use is leisure (shopping and restaurant districts) and work (office buildings). These are also the most touristic neighbourhoods.

Discussion

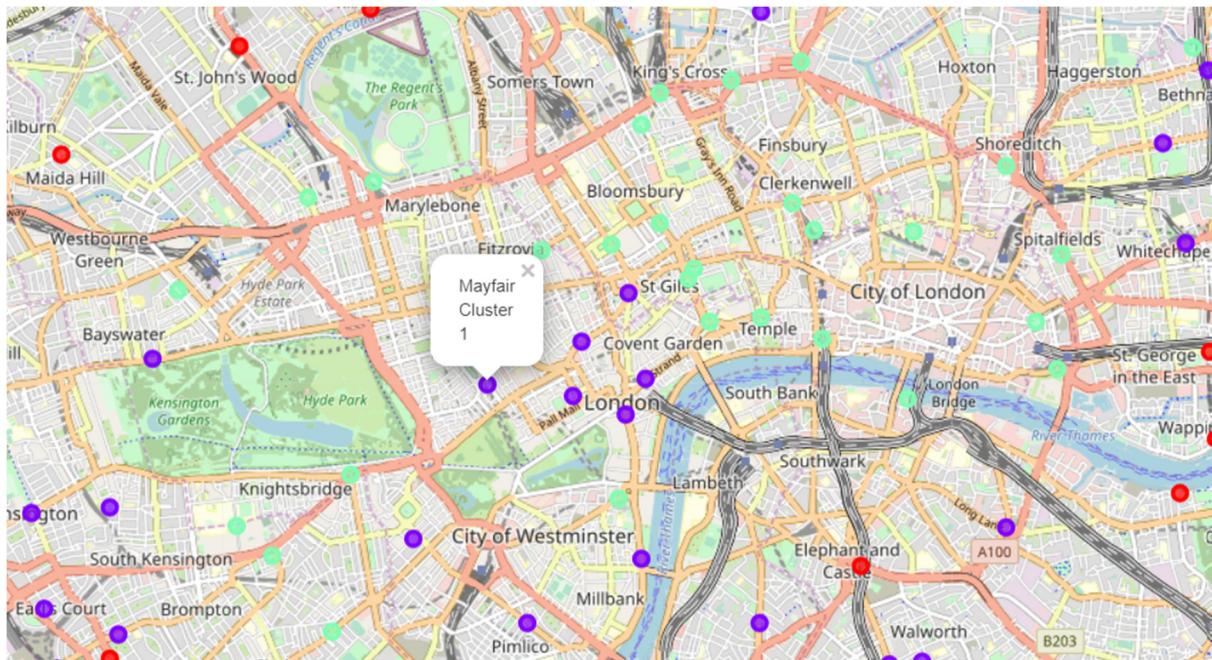
Our Machine Learning analysis has helped us identify three clusters of neighbourhoods based on their number of coffee shops. This, together with the location of these clusters on the map, has helped us understand which would be the best neighbourhood London to open a new coffee shop.

These are the neighbourhoods we would recommend for a potential entrepreneur looking to open a coffee shop:

- One of the best neighbourhoods is Elephant and Castle. This neighbourhood is part of Cluster 0 (low number of cafes) and according to our analysis has 6 cafes. In addition, this neighbourhood is close to central London and is undergoing urban regeneration with a lot of investment. It is one of the up and coming areas of London.

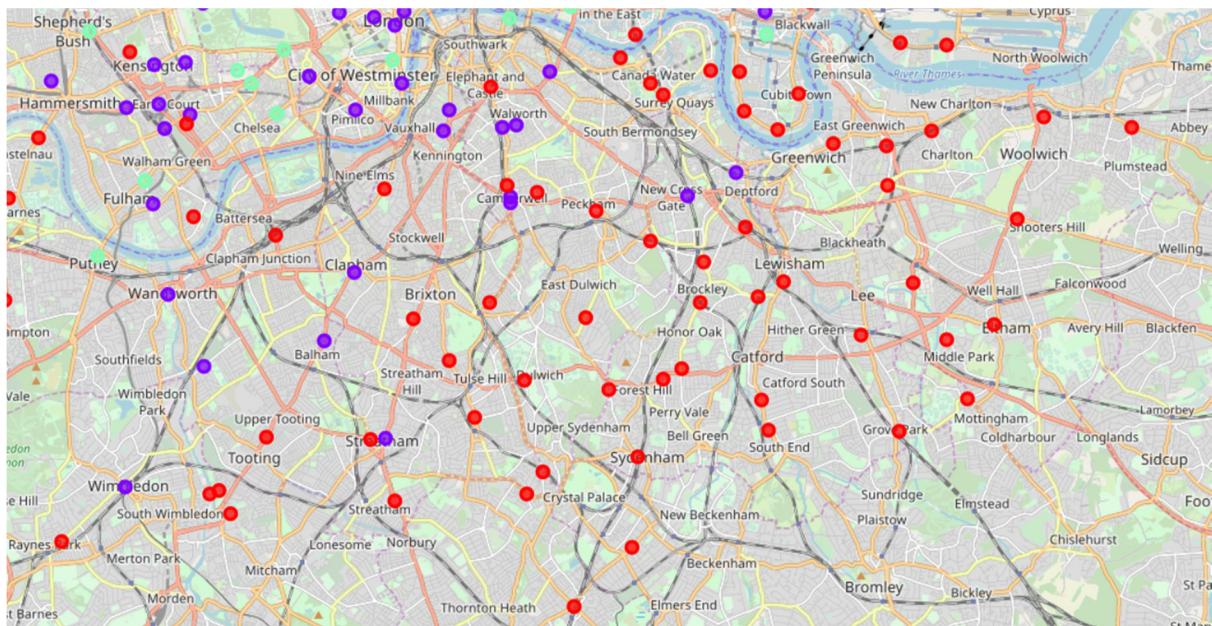


- Another great neighbourhood would be Mayfair. This neighbourhood is part of Cluster 1 (medium number of cafés) and according to our analysis has 13 cafes. Mayfair is an area famous for its high-end restaurants and hotels, and luxury shops. This may make this location very good for an upper-end speciality café.

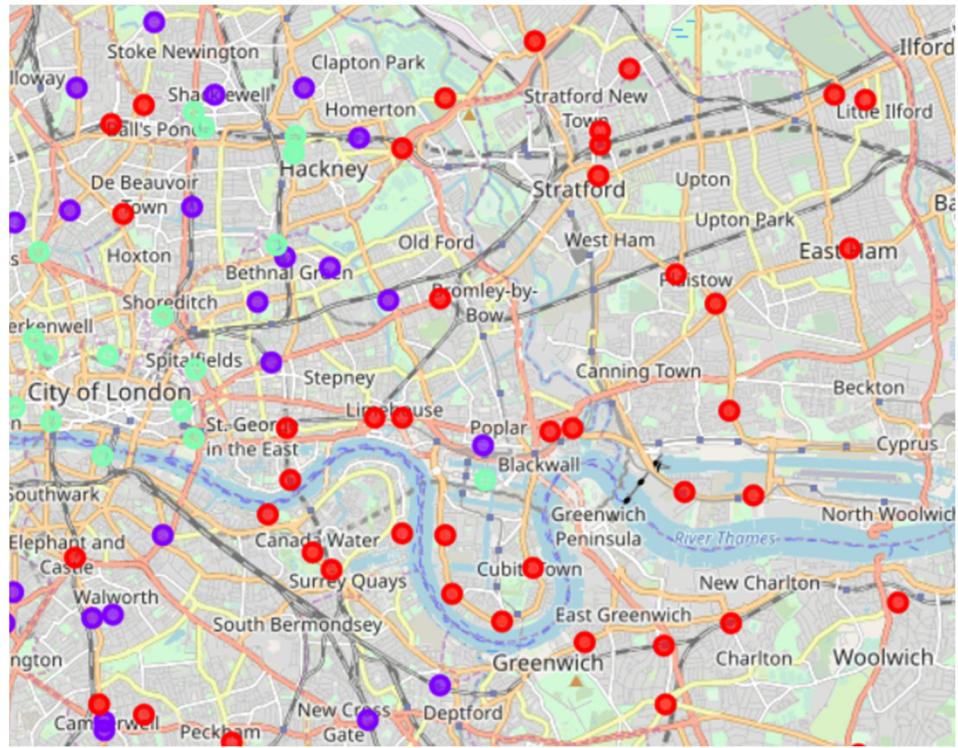


In addition to these two neighbourhoods we can also recommend areas of London based on the location of the cluster 0 neighbourhoods (with less than 7 cafes) on the London map. We have identified two areas where there is a low number of cafés in general:

- South London:



- East London:



Both of these areas of London have been traditionally less touristic and have less leisure and work districts, however they are becoming trendier and undergoing significant regeneration. A clear example would be Stratford (which right now only has 4 cafés). This neighbourhood was majorly regenerated and developed for the London 2012 Olympic Games and has become a shopping and leisure destination. Hence these would be areas with potential.

Conclusion

The purpose of this project was to identify London neighbourhoods with low number of coffee shops in order to aid stakeholders in narrowing down the search for optimal location for a new coffee shop.

By calculating coffee shop density from Foursquare data we have first identified neighbourhoods with high and low number of cafés. Clustering of those locations was then performed in order to group the neighbourhoods into 3 clusters based on their number of cafés (high, medium and low). Finally the neighbourhoods and their clusters were plotted in a map to find the geographical distribution of the three clusters. This allowed us to identify that in general, neighbourhoods in central London have a higher number of cafes than those in the outskirts. However, some interesting outliers were found (such as the neighbourhoods of Elephant and Castle and Mayfair), which would be good neighbourhoods for a new coffee shop.

The final decision on the optimal coffee shop location will be made by stakeholders based on the number of cafes already in the neighbourhood (data presented in this project) but also taking into consideration additional factors such as attractiveness of each location (proximity to parks or shopping centre), proximity to a tube station, levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighbourhood etc.