Selecting best locations for a new café in London

Coursera Capstone Project - The Battle of Neighborhoods

Ahmed Baaziz

April 17, 2020

******

1. Introduction:

1.1 Background

Experts agree that Britain now has one of the most vibrant coffee cultures in the world, which is showcased by the constant increase year after year of the number of the London Coffee Festival attendees - more than 31000 people attended last year edition (2019). In addition to that, UK residents consume 95 million cups of coffee per day and their love for the black elixir is growing. While the coffee culture is booming in London, we also see the number of cafés and coffee shops more than doubling in UK since 2010. In fact, London's West End is now home to more Starbucks' branches than the whole of Australia.

This created a strong competition between cafés (independents and chains), but also generated opportunities for entrepreneurs and independent coffee shops to establish themselves and grow. Therefore, it is important to know which neighborhoods in London are the best to open a new cafés shops and which areas are attractive for coffee lovers.

1.2 Problem

Beside having an original menu and an attractive atmosphere, the location of a café is fundamental to its success. People pick a café for its convenience – e.g. close to their office, or in an area knowing for good cafés, so it’s important to pick the right location for your coffee. But this can be challenging in a city as big and as crowded as London.

This project aims to support coffee enthusiasts and entrepreneurs who are thinking about opening a new café and help to identify the best locations in London.

1.3 Interest

Our project supports entrepreneurs who are thinking to open a new café in London streets, but hesitate between different locations or don't know where their future coffee shop will have higher chances to grow and be profitable. Selecting the right location will be a god advantage for them and may determine how much successful their coffee business will be.

2. Data acquisition

2.1 Data sources

The data should cover the list of London boroughs and neighborhoods, with latitude and longitude coordinates and list of cafés for each neighborhood. We may also consider looking at the population density - unfortunately we couldn't find data covering this per neighborhood, so we will work with population per borough.

Such a data is not available, so we had to collect it from different sites or datasets:

- London boroughs and neighborhoods data is available in the Wikipedia page – List of areas of London.
- The Python library Geopy was used to get the latitude and longitude of each neighborhood.
- London population per borough data is available in the website gov.uk - Population borough London.csv.
- Foursquare API was used to request the data related to available cafés and coffee shops, their locations and details.

2.2 Data cleaning and wrangling

Data downloaded or scraped from multiple sources had to be grouped into one table. But before that, we had to understand the data and clean it, to generate datasets with the needed columns.

Using the BeautifulSoup library, we ended-up with duplicated rows in the dataframe we built from the data extracted from the Wikipedia page. In addition to that, we dropped the first row in the dataframe, since it's listing the columns' names, the rows without a neighborhood (stated as "not assigned" and removing the numbers in some cells, related to how they were referenced in the Wikipedia page. After renaming the column, we did a final check to confirm no neighborhoods were listed more than once, linked to different boroughs, and then reset our index. We ended-up with a dataset listing 527 unique neighborhoods, linked to 38 unique boroughs.

	Neighborhood	Borough
0	Abbey Wood	Bexley
1	Acton	Ealing
2	Addington	Croydon
3	Addiscombe	Croydon
4	Albany Park	Bexley
5	Aldborough Hatch	Redbridge
6	Aldgate	City of London
7	Aldwych	Westminster
8	Alperton	Brent
9	Anerley	Bromley

Table 1 - Extract from the dataframe of London neighborhoods

The Python library Geopy can be challenging to use and is not very reliable. We had to run it few times before getting the needed latitude and longitude coordinates. In some cases, the code returned a "service not available" error if we exceeded a certain number of requests in one time. As a workaround, we had to split the data into multiple sections, and execute the code for each one of them separately. In addition to that, some rows ended-up without coordinates, their number was very small compared to the total data and could not impact our analysis, so we just delated them from the final dataset.

	Neighborhood	Borough	Latitude	Longitude
506	Westcombe Park	Greenwich	51.4842	0.0188787
507	Westminster	Westminster	51.5014	-0.12493
508	Whetstone	Barnet	51.6302	-0.174884
509	White City	Hammersmith and Fulham	51.5119	-0.224236
510	Whitechapel	Tower Hamlets	51.5186	-0.0620807
511	Whitton	Richmond upon Thames	51.4512	-0.357976
513	Willesden	Brent	51.5466	-0.235866
514	Wimbledon	Merton	51.4215	-0.206403
515	Winchmore Hill	Enfield	51.6334	-0.103362
516	Wood Green	Haringey	51.596	-0.109147
517	Woodford	Redbridge	51.6068	0.0340119
518	Woodford Green	Redbridge	51.6118	0.0240796
519	Woodlands	Hounslow	51.4722	-0.337432
520	Woodside	Croydon	51.3871	-0.0653308

Table 2 - Extract from the dataframe of London neighborhoods with latitude and longitude coordinates

Since we decided to use the population by borough data, we downloaded the csv file, understood the data and selected the needed columns. The data was already well organized and clean, we just selected the needed rows (population for the year 2019) and saved it in a separate dataframe - to be used in the end of the project.

	Borough	Population
124	Barnet	402363
384	Croydon	396548
1268	Newham	359470
436	Ealing	354184
228	Brent	340710
488	Enfield	339480
1112	Lambeth	338028
280	Bromley	334292
1632	Wandsworth	328828
1424	Southwark	327271

Table 3 - Extract from the dataframe of London population per borough

Last step was about getting the list of cafés and coffee shop per neighborhood. As mentioned above, we did use the Foursquare API in Python – you'll need to have a developer Foursquare profile and also be aware of the limited number of request that you can execute per day, based on your profile and subscription. The list returned by the Foursquare

API listed all the venues in our neighborhoods. Since we were only interested in cafés and coffee shops, we filtered our dataset based on the venues' category.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	ID	Venue	Venue Latitude	Venue Longitude	Venue Category
Acton	51.508140	-0.273261	517d323fe4b0b4d24c007345	Costa Coffee	51.506879	-0.269368	Coffee Shop
Acton	51.508140	-0.273261	4d971048a2c654815e9dce53	Frank's Cafe	51.508083	-0.270300	Café
Addiscombe	51.379692	-0.074282	533fea17498e1f39b6eab646	The Tram Stop	51.380188	-0.073378	Café
Aldgate	51.514248	-0.075719	4f70a772e4b0f375fc669005	The Association	51.513733	-0.079132	Coffee Shop
Aldgate	51.514248	-0.075719	4cb1d779562d224b09ec2188	Fazenda	51.516169	-0.077414	Coffee Shop
Aldgate	51.514248	-0.075719	59f97d45acb00b6e13f7d430	Black Sheep Coffee	51.513990	-0.075459	Coffee Shop

Table 4 - Extract from the dataframe of London cafés

3. Exploratory data analysis

3.1 Dispersion of neighborhoods per borough

We started exploring our data checking the distribution of the neighborhoods per boroughs. This helped to understand how London neighborhoods are split and give an idea how the data will be visualized in a map.

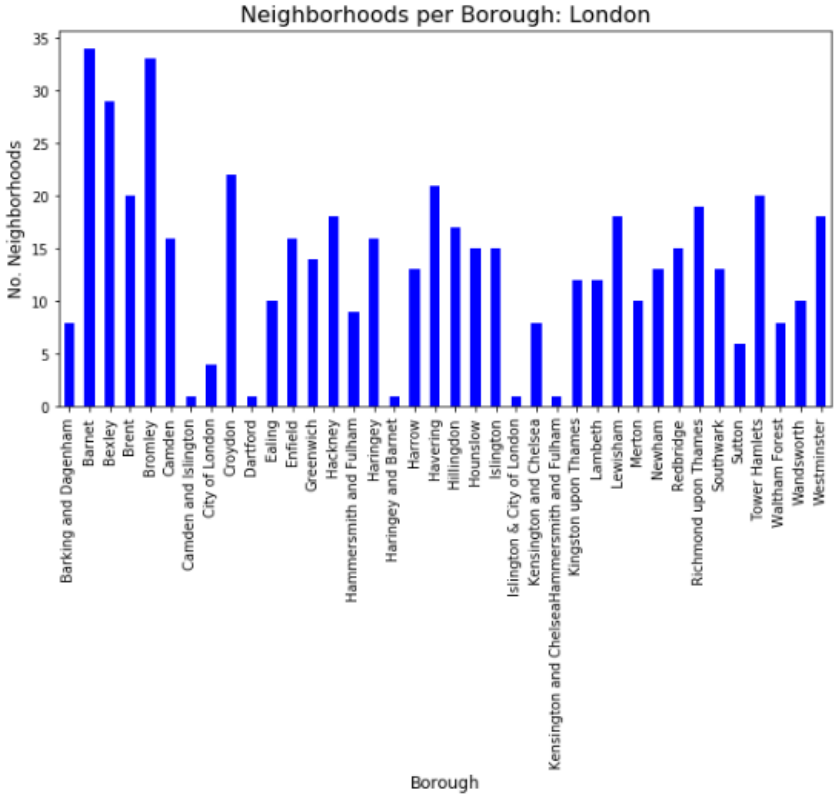

Figure 1 – Neighborhoods per borough - London

3.2 Dispersion of cafés per borough

Digging more into the data, we wanted to understand more the distribution of cafés in London, per neighborhood and per borough. Due to the large number of neighborhoods, we decided to only visualise the distribution of cafés per borough (figure 2).

It seems that the boroughs close to the centre of London, and known by tourists and locals by as good spots to hang-out and shopping, are the ones with the highest concentration of cafés. This is expected, since these areas will have a high density of visitors every day, and are situated around attraction sites. Sounds like these sites could be good candidates, but the advantage of being in Camden or Westminster means that we will face a strong competition.

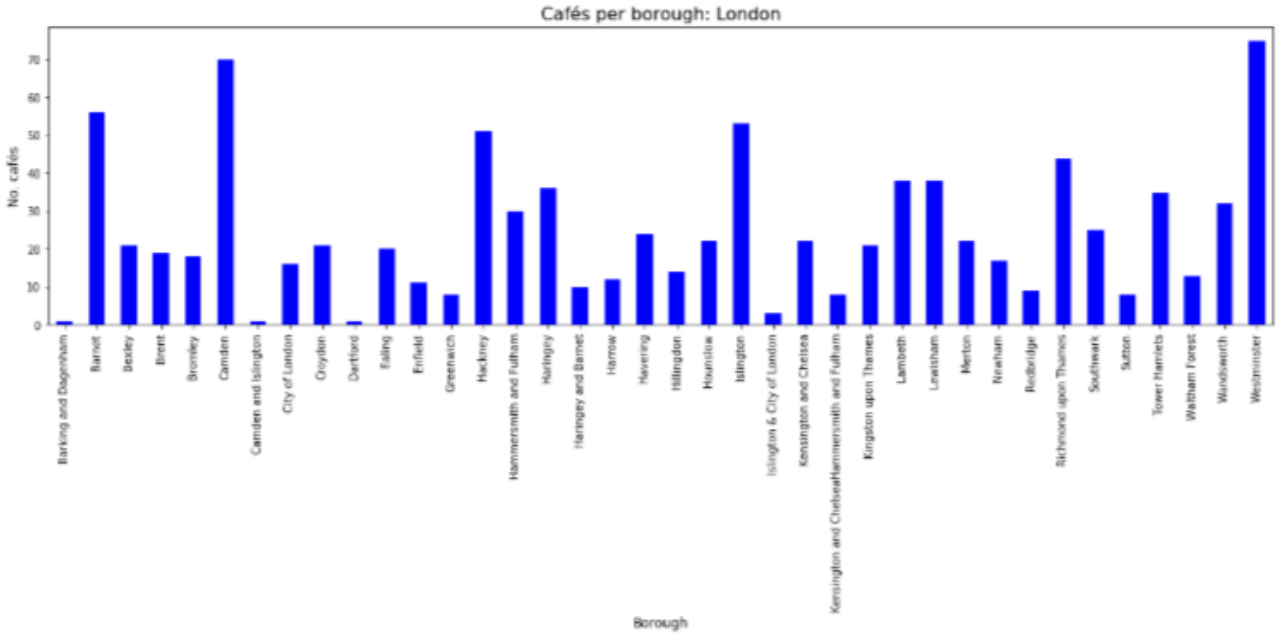

Figure 2 - Cafés per borough - London

3.3 Cafés rating per neighborhoods

We next wanted to understand in a better way the coffee culture in London, which neighborhoods are known for their good cafés and if coffee shops rating has a relationship with specific neighborhoods or boroughs.

To get the details of each café in our dataframe, we used again the Foursquare API and extracted the number of likes, tips and rating for each venue, cleaned the data, then created a new dataset where we list the average cafés rating per neighborhood.

Visualising the data in a histogram, it was clear that the majority of neighborhoods were having a good rate between 7.0 and 7.5 (on a scale between 1.0 and 10.0) – over 40 neighborhoods in total (figure 4). Since we want to identify the best locations to open a new café in London, we decided to focus on the neighborhoods having an average café rate higher than 8.0 (figure 5).

	Neighborhood	Average Rating
142	Limehouse	9.100000
25	Brixton	9.100000
124	Homerton	9.000000
67	Deptford	8.900000
154	Nag's Head	8.850000
234	West Green	8.700000
22	Bowes Park	8.700000
14	Belsize Park	8.600000
63	Dalston	8.600000

Table 5 - Extract from the dataframe of average cafés rating per neighborhood - London

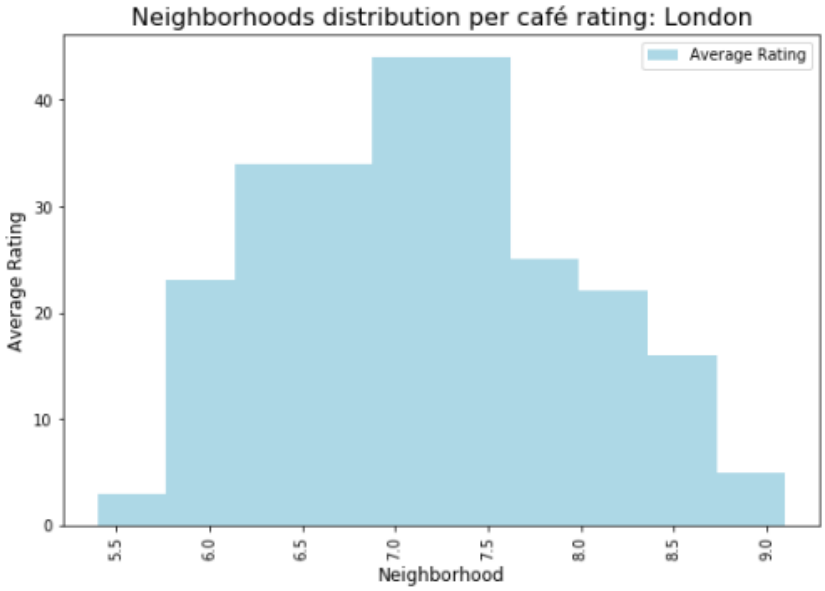

Figure 4 - Neighborhoods distribution per café rating - London

To analyse further the data, we visualized our list of neighborhoods in London map, using the Folium Python library (figure 6). Aligned with our first hypothesis, the majority of the neighborhoods with high average cafés rating are close to London city centre, around tourist areas – but we are still have a large list of neighborhoods and we will need to reduce it more.

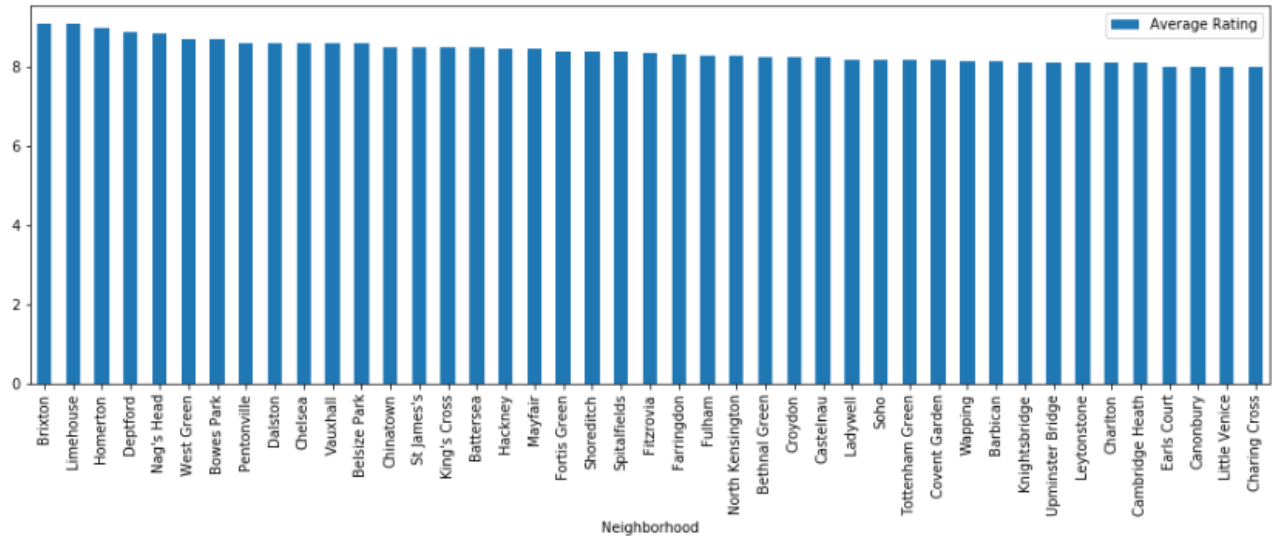

Figure 5 - Top neighborhoods with high average cafés rating

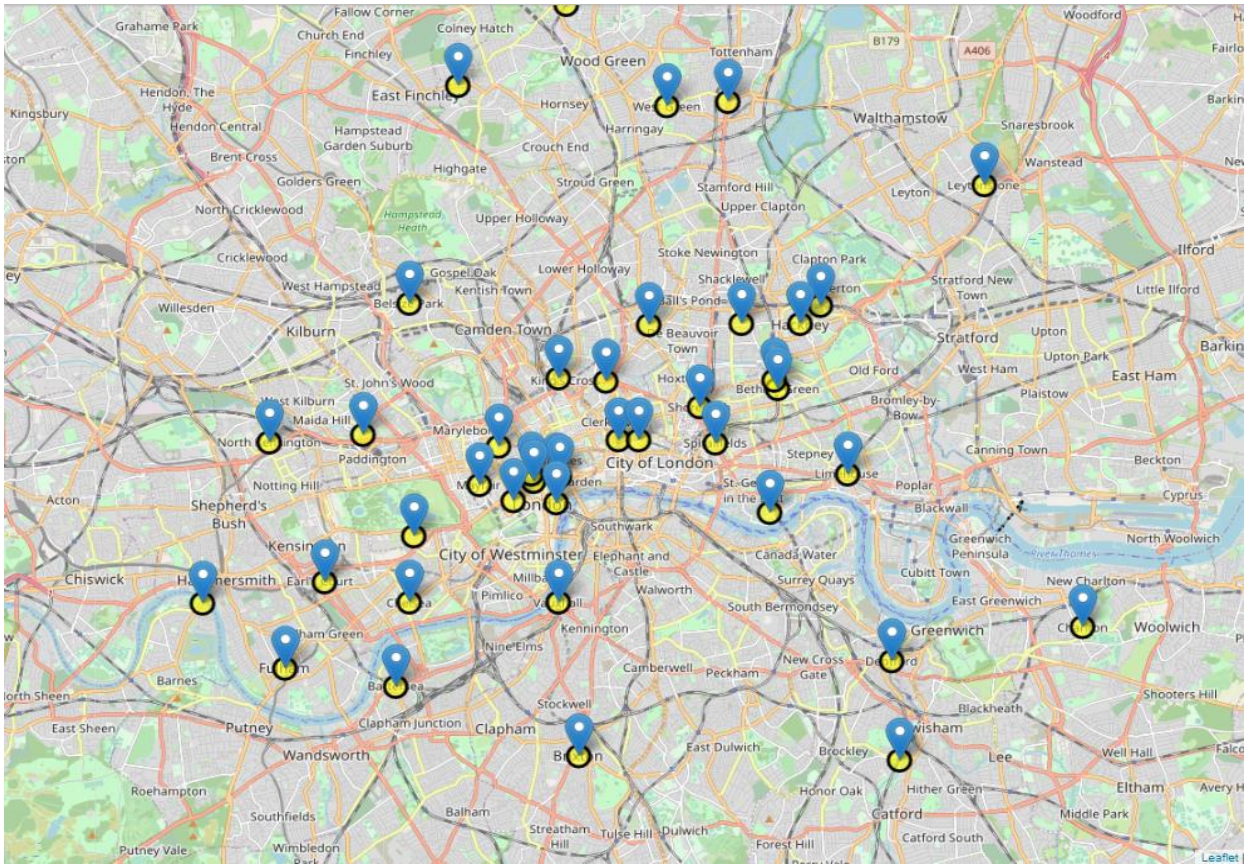

Figure 6 - Visualizing high average cafés rating neighborhoods in London map

3.4 Optimizing the list of proposed neighborhoods

In addition to the average cafés rating score, we will also check the number of cafés per neighborhood and try to identify if there is a relationship between both conditions.

We can see in figure 7 that the majority of the neighborhoods with a great average rating score have 1 or 2 cafés open in the area. While this is a positive sign regarding to potential competition in these neighborhoods, we do assume that they are not known by people as good spots for a great coffee, due to the low number of open cafés. Probably, the high rating is coming from locals who live in that neighborhoods, know the café and have been using it for years.

On the other hand, we saw that three neighborhoods have more than 7 cafés (Shoreditch, Fitzrovia and Knightsbridge). These neighborhoods are in Westminster and Camden boroughs, which we highlighted earlier in our report as being areas with high and strong competition and probably not ideal for an entrepreneur to open a new café there.

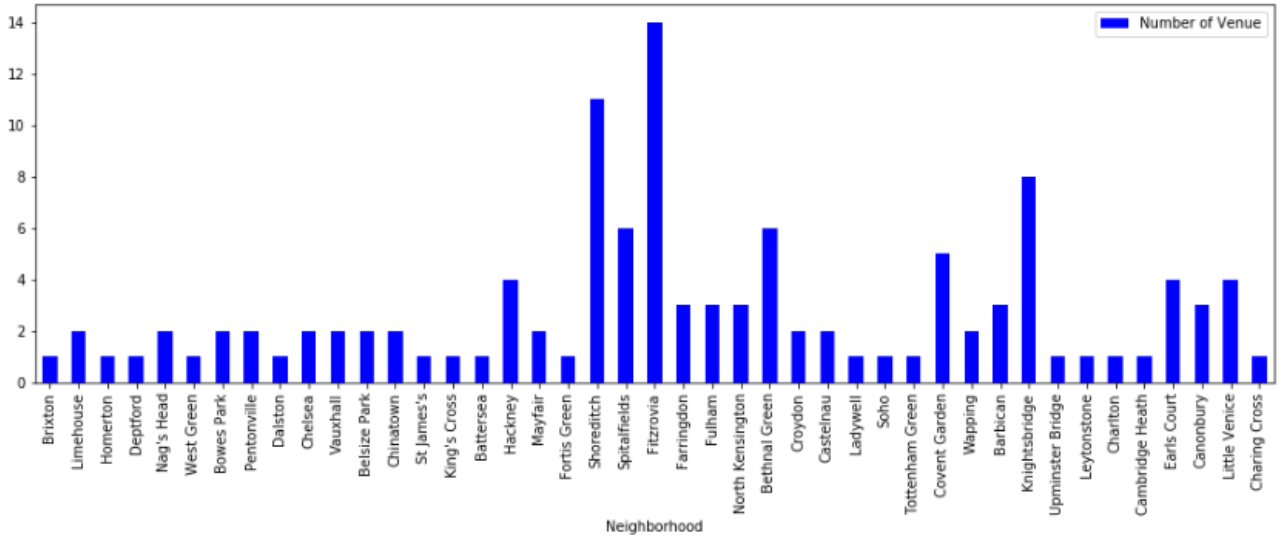

Figure 7 - Number of cafés in top neighborhoods - London

We decided to only focus on neighborhoods with an average number of cafés already open (between 3 and 6), and of course with a high cafés rating score (figure 8).

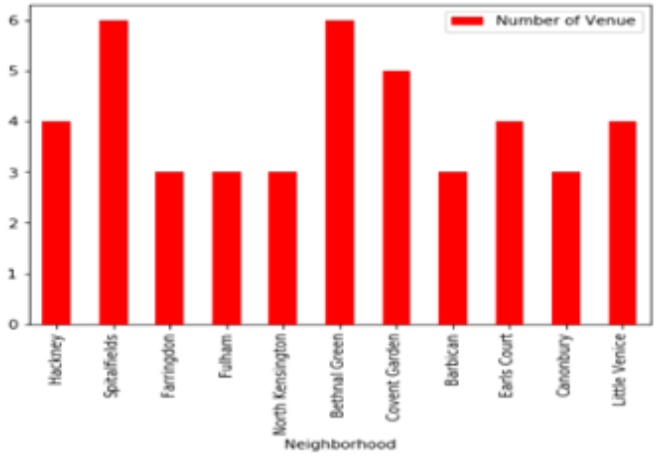

Figure 8 - List of top neighborhoods with reasonable number of cafés - London

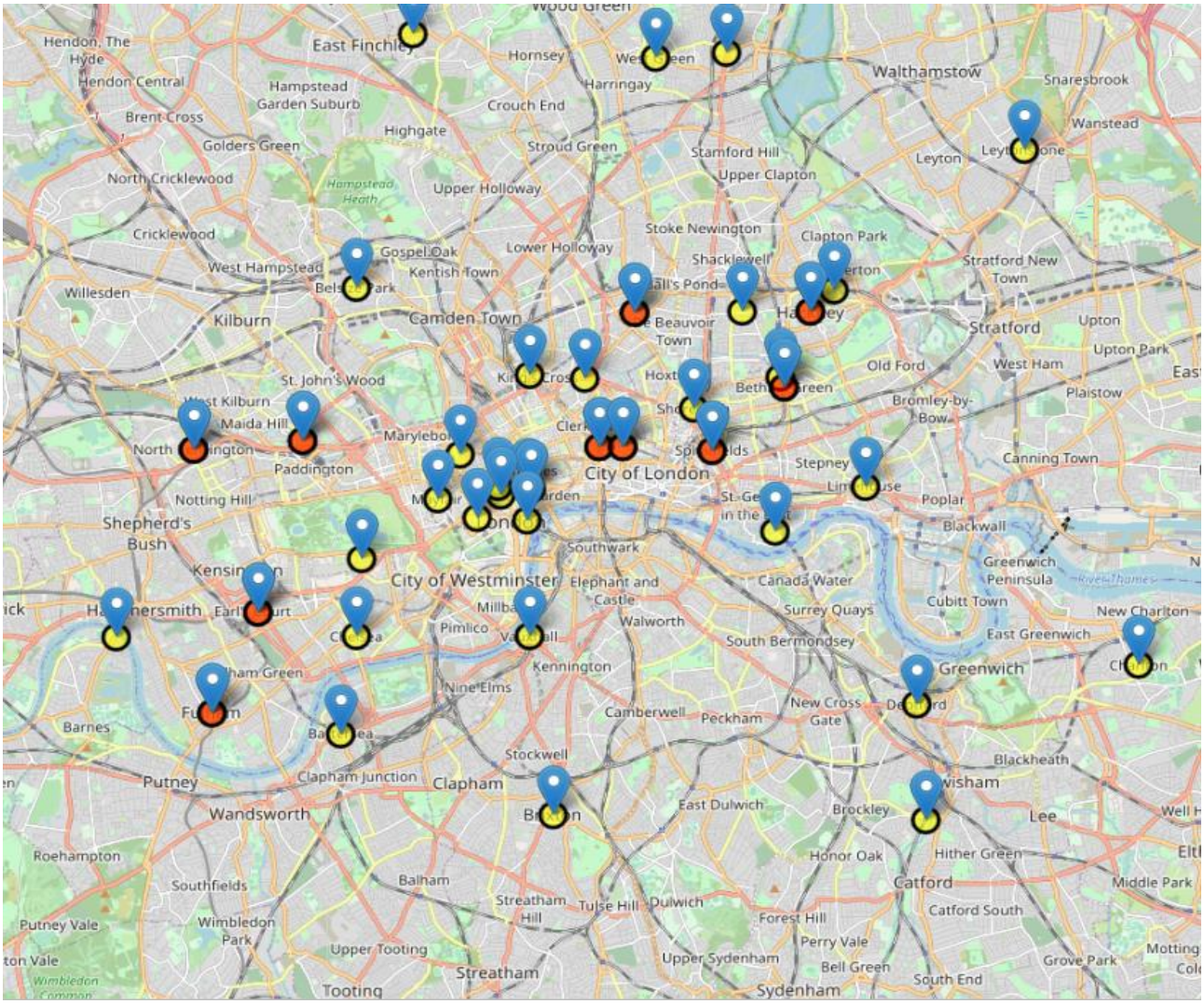

Figure 9 - Visualizing top neighborhoods in London map (in red)

We ended-up with 10 potential good neighborhoods to open a new café. The number is still high, and will not be very helpful to a new entrepreneur who is not familiar with the city.

We applied the last criteria to reduce the number of good locations and highlight the best three neighborhoods - we compared our list to the population density per borough and select the neighborhoods located in a high population borough, which are:

1) Hackney
2) Spitalfields
3) Farringdon

4. Results and discussion

Our analysis shows that although there is a great number of cafés in London (around 800 in 500 neighborhoods), there are few locations with acceptable density, close to city center and very appreciated for their coffee cultures.

The highest concentration of cafés was detected in Camden and Westminster, which are touristic areas and neighborhoods known by a high density of people. We tried to avoid these neighborhoods, since the competition is already strong there and they probably have faithful customers not willing to change their habits.

We focused on popular areas that offer a combination of a good coffee culture (knowing for their good cafés), closeness to city center and a strong socio-economic dynamic (close to business areas like Canary Wharf). After that, we had to start looking at the neighborhoods with a reasonable (fairly low) number of open cafés and part of the most popular neighborhoods.

In the end of our project, we were able to list the best three locations to open a new café in London, and where an entrepreneur will potentially have the highest chances to succeed.

5. Conclusions

The purpose of this project was to identify London neighborhoods close to center with low number of cafés, in order to aid entrepreneurs in narrowing down the search for optimal locations for their new coffee shop. By calculating the café density and average rating per neighborhood/borough using Foursquare API/data, we first identified around 50 neighborhoods that can be considered for further analysis. Then we started looking at the competition and the population (which can be translated to potential future customers).

To have good chances of success, it is recommended to open a new café in Hackney neighborhood. There are only 4 other cafés already open, with an average rating of 8.46 on a scale from 1.0 to 10.00 (considered very high comparing to other areas) and with a decent population (around 300000 people). To not limit the choice to one neighborhood, we also recommended Spitalfields and Farringdon as the 2nd and 3rd best locations to open a café. This will offer a better flexibility for the entrepreneurs knowing that it's not easy to find a shop to rent or buy in city center.

As a final note, these results can be further developed by doing similar analysis by streets (focusing on the three proposed areas) and also consider the real estate situation (the availability of a shop, how much renting will cost, etc...) which will increase the chance to build a sustainable and lucrative coffee business. This will require different datasets.