Alexander Nicolussi 21.11.2020
Capstone Project – The Battle of Neighbourhoods

Nightlife-activity in London

Table of contents:

1. Introduction

This report intends to examine the distribution of nightlife-activity in London, UK. The aim of the project is to find regions in the Greater London Area that have significant nightlife-activity (Pubs, Restaurants, Theatres…). Furthermore, we will try to analyse if a cluster with high activity exists by using a **DBSCAN** cluster machine learning algorithm.
One big problem when moving to a new city is to identify a suitable area where we want to live. And a lot of studies have been done on various parameters that characterise each area, for example how pricy and international each area is. Nonetheless, it has been found that some aspects are missing in the aforementioned reports. Especially many young people have the urgent desire to live in a community with a vibrant nightlife. This report can be useful for different people. People moving to London who are seeking the nightlife can use this as a guide to find those areas. Also, investors who are planning on investing in student housing, or looking to open a new venue can use this as a reference for the different areas in the Greater London Area.
For this purpose, we analyse the centres of activity in the Greater London Area, not just Inner London, to get a full picture of the city. This can be useful because somebody might be living outside of London and the travel to the nearest area of interest could be shorter, and also because investors might be interested in not so developed parts of London.
This is just a preliminary study to determine which areas are to be considered further, which is up to the reader and their kind of interest in the data.
The problem can be solved by using the Foursquare location data. We can use the area locations and find nearby venues. These venues can then be classified. We can filter the venue-types that we are interested in and look at the frequency of occurrence in the neighbourhood of the chosen area. This frequency analysis is a good preliminary model for the evaluation of nightlife-activity.

2. Data

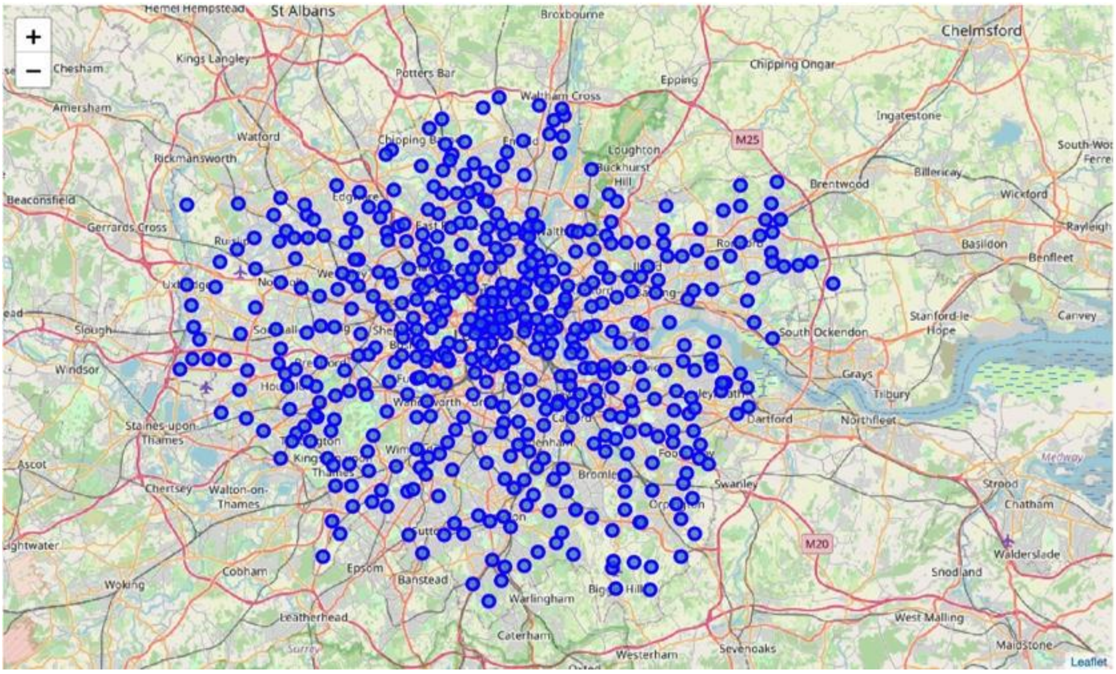
To carry out the aforementioned analysis we need data. Firstly, we need the names and locations of the various areas in the Greater London Area. We can get these from Wikipedia.

Wikipedia London Areas link: https://en.wikipedia.org/wiki/List_of_areas_of_London

We can scrape these without the use of any additional library. We can use **pandas.read_html** to extract information from the **Wikipedia** page. We can drop everything except the names. Then, we proceed by looking for the locations of these areas. We can use **geopy** to obtain the locations of the areas, meaning the latitude and longitude of each point. We use a loop to do this for every area on our list.
To obtain the venues that are present we use **Foursquare**. This allowes us to obtain data about surrounding venues. Again, we use this for every area and are able to obtain the nearest 70 venues in a radius of 500 metres. These are a statistical representation of the areas. From this we can filter venues that contain indicator names such as **Restaurant, Pub, Theatre, Bar** and so on. We can compute the frequency of these signal words and obtain a rate for nightlife-activities.

3. Methodology

We use the data that we import from Wikipedia and construct a table of the area names. With the names of the areas we obtain their latitude and longitude by using **geopy**. This table can be visualised by using the **Folium** library.



We append these in our table and have obtained a table that we can use for **Foursquare**. We create a loop to obtain the 70 nearest venues within a radius of 500 meters. We transform the obtained dataset to show only the number of venues included in our selection. The selection includes:

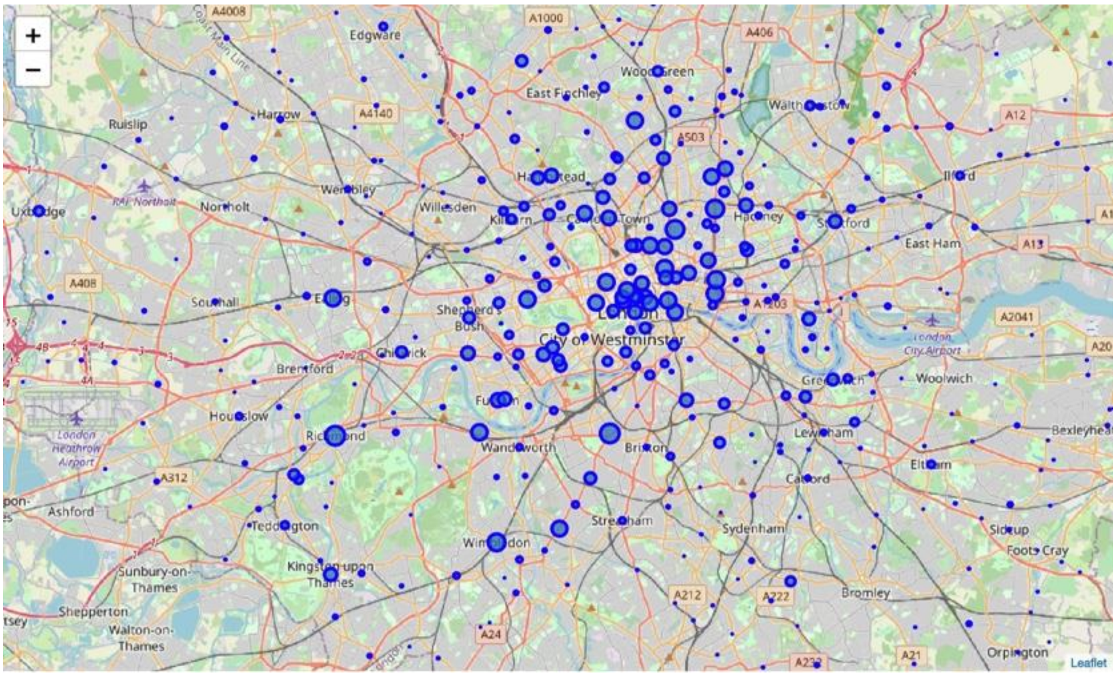
- Restaurant
- Bar
- Pub
- Theater
- Concert
- Strip Club
- Burger
- Lounge

- Diner
- Pizza

Therefore, we can use this data to plot the concentration for each of the areas. After this, we can look for a cluster. We use the **DBSCAN** algorithm from the **scikit-learn** library. We use specific values, adapted to the dataset to compute the cluster. We find a cluster in Central London and this cluster can be plotted using **Folium** and we can then extrapolate the areas in this cluster.

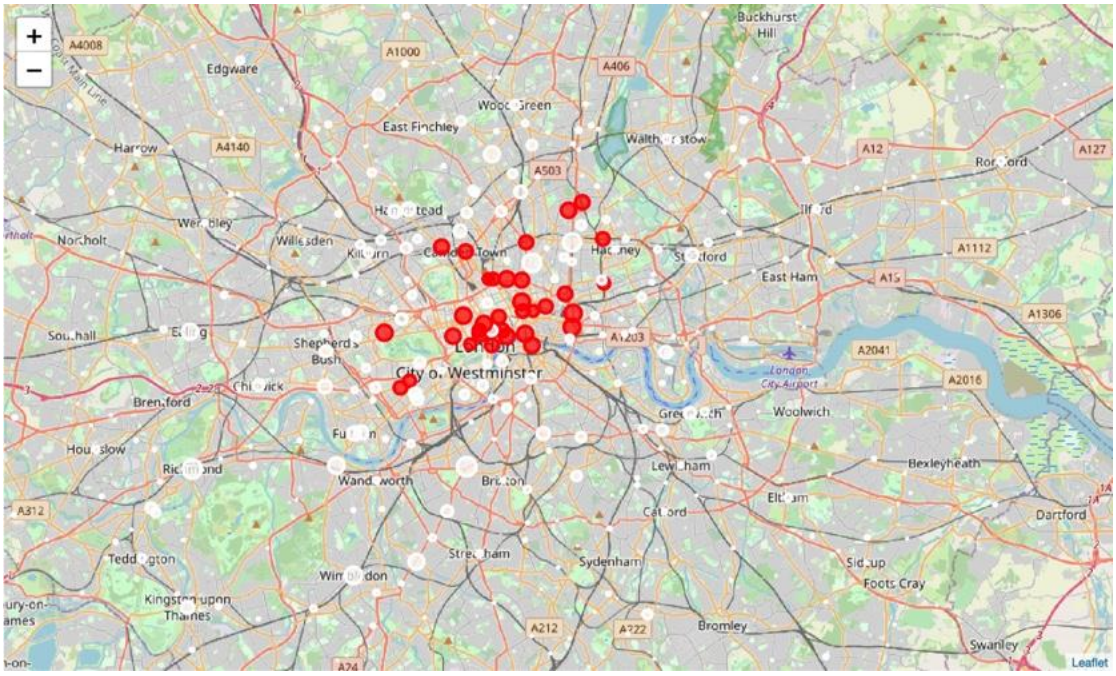
4. Results and Discussion

We wanted to find a parameter that could be used to determine how active different areas are when it comes to nightlife in London. We have used **Foursquare** to successfully obtain this parameter. Therefore, we are left with a dataset that contains a parameter for each area. A plot of this dataset can be seen in the figure below, where bigger dots represent a higher concentration.



From the image above it can be seen that Central London has a higher concentration of nightlife-activity venues. There are some regions outside the Centre that have high concentrations. These are the outskirt cities of London. They can be very attractive for investors and can easily be filtered from the calculated table. After that, we have found a cluster in Central London by using a **DBSCAN** clustering method. We can see that this cluster is formed of several areas, which have similar concentrations, around 40-50%. We have therefore found that the areas that can be found in the appendix are the areas in Central London with the highest concentration of nightlife-activities. Therefore, these are the ones that can be analysed further, depending on the sought-after goal.

The author knows that the model is not perfect. Some areas in Central London with higher concentrations are left out because the **DBSCAN** algorithm does not consider them. One could say that this analysis was obvious, "of course nightlife is better in the centre!", but the author thinks that everything needs to be verified even if obvious. Furthermore, this analysis can be used to determine which areas in Central London are better suited, as they still have some variation among them.



It can be seen from the above image, where dots in red are part of the cluster, that the cluster is formed in central London. This cluster includes areas such as Camden Town and Chalk Farm which are known to be active at night. The full list of areas in the cluster can be found in the appendix. We have found the areas of interest even though the model is not perfect.

5. Conclusion

We have successfully found the different concentrations of nightlife-activities in the Greater London Area. Furthermore, we were able to find a cluster in Central London. This means that Central London is the most consistent greater area of London. The concentration is higher in Central London than Outer London even if some high concentration centres exist.

6. Appendix

	Area_name	frequency	latitude	longitude	Clus_Db
4	Aldgate	35	51.514248	-0.075719	0
5	Aldwych	30	51.513131	-0.117593	0
8	Angel	31	51.531842	-0.105714	0
13	Bankside	32	51.507499	-0.099302	0
14	Barbican	24	51.520150	-0.098683	0
21	Bayswater	34	51.512276	-0.188385	0
33	Bethnal Green	25	51.530346	-0.056163	0
35	Blackfriars	34	51.511585	-0.103767	0
53	Brompton	28	51.491822	-0.178326	0
60	Camden Town	30	51.542305	-0.139560	0
70	Chalk Farm	31	51.544114	-0.153481	0
71	Charing Cross	29	51.507497	-0.123689	0
77	Chinatown	32	51.511680	-0.130496	0
83	Clerkenwell	35	51.523727	-0.105555	0
92	Covent Garden	25	51.512874	-0.122544	0
138	Farringdon	28	51.520124	-0.104793	0
141	Finsbury	29	51.521798	-0.091425	0
143	Fitzrovia	34	51.518764	-0.141002	0
167	Hackney Central	28	51.547061	-0.056875	0
197	Highbury	29	51.545800	-0.102713	0
201	Holborn	29	51.517934	-0.119528	0
224	King's Cross	26	51.532395	-0.123022	0
255	Mayfair	32	51.511087	-0.147058	0
303	Pentonville	33	51.532105	-0.114894	0
341	Shoreditch	30	51.526669	-0.079893	0
347	Soho	28	51.513163	-0.131175	0
352	South Kensington	26	51.494049	-0.173044	0
362	Spitalfields	36	51.519527	-0.075170	0
363	St Giles	33	51.515472	-0.128418	0
364	St James's	24	51.507908	-0.136573	0
370	St Pancras	24	51.532121	-0.125863	0
376	Stoke Newington	32	51.557697	-0.077282	0
390	Temple	33	51.510966	-0.114335	0
438	West Hackney	30	51.560859	-0.069132	0