Determining the “Green” Neighborhoods of London

Author: Aleeza Ladha

Date: 31/01/2020

1. Introduction: Business Problem

In this project we will try to help a real estate client that is about to move to London find a neighborhood to live in. As her family are outdoor people, her main criteria is that there is greenery, i.e. parks with in close proximity

The aim of this project will be to provide a list of locations within London/Greater London in which our real estate client will have close access to greenery. We will provide a list and map of these “green” neighborhoods in which the client can use as a baseline to then further refine to satisfy her family’s other living criteria.

2. Data Acquisition and Cleaning

Based on definition of our problem, the main factors that will influence our decision are:

- The density of "green" areas within in a 200m radius of each neighborhood.
- Distance of "green" neighborhood from city center

Note that "green" areas will be defined as parks, gardens, botanical gardens.

2.1 Data Sources:

The following data sources were used to extract/generate the required information:

- A list of London neighborhoods is extracted from Wikipedia page: https://en.wikipedia.org/wiki/List_of_areas_of_London.
- Coordinates of each London neighborhood will extracted from csv file extracted from: https://www.doogal.co.uk/london_postcodes.php
- The number of "green" areas and their type and location in every neighborhood will be obtained using **Foursquare API**
- The coordinate of London center will be obtained using **Google Maps API geocoding**

2.2 Data Correlation and Cleaning:

London neighborhood data:

London neighborhood data was scraped from the Wikipedia page defined above in order to isolate the table below:

	Location	London borough	Post town	Postcode district	Dial code	OS grid ref
0	Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
2	Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728
...	...	...	...	...	...	...
528	Woolwich	Greenwich	LONDON	SE18	020	TQ435795
529	Worcester Park	Sutton, Kingston upon Thames	WORCESTER PARK	KT4	020	TQ225655
530	Wormwood Scrubs	Hammersmith and Fulham	LONDON	W12	020	TQ225815
531	Yeading	Hillingdon	HAYES	UB4	020	TQ115825
532	Yiewsley	Hillingdon	WEST DRAYTON	UB7	020	TQ063804

Only the OS grid reference and location were needed for our analysis, and therefore, all other columns were deleted and all rows with NAN were dropped.

London neighborhood coordinate data:

London coordinate data was extracted from a csv file present online. The csv file was loaded directly into Juptyer notebook in order to facilitate access. The OS grid data and corresponding longitude and latitude coordinates were isolated from this rather dense file.

It was found that multiple coordinates pairs existed for each grid coordinate, and therefore the mean of the latitude coordinate and the mean of the longitude coordinate were taken as the grid ref coordinate, and stored in a dataframe.

The two dataframes were combined in order to obtain the coordinates for each neighborhood defined in the Wikipedia table. All NAN rows were dropped.

e.g.

	OS_grid_ref	Location	Latitude	Longitude
0	TQ345665	Addiscombe	51.382106	-0.068576
1	TQ334813	Aldgate	51.515328	-0.078387
2	TQ307810	Aldwych	51.513626	-0.117027
3	TQ185835	Alperton	51.538395	-0.292108
4	TQ345695	Anerley	51.408792	-0.066646

Foursquare API:

Foursquare was then used in order to gain information about the venue types present within a 200m radius in each neighborhood.

The “green” venues defined as parks, gardens and botanical gardens, were then isolated for each neighbourhood (within a 200m radius) and the total green areas determined and presented in a dataframe:

	Location	Botanical_Garden	Garden	Park	Total
0	Addiscombe	0	0	0	0
1	Aldgate	0	0	0	0
2	Aldwych	0	0	0	0
3	Alperton	0	0	0	0
4	Anerley	0	0	1	1
5	Angel	0	0	0	0
6	Balham	0	0	0	0
7	Bankside	0	1	1	2
8	Barbican	1	1	0	2
9	Barkingside	0	0	0	0

Once data is compiled and cleaned for NAN values, we are left with **302** London neighborhoods in which we are able to analyze.

3. Machine learning Analysis:

In this project, based on the client's preference to have close access to a greenery, we will direct our efforts to creating a list of "green" neighborhoods across London, and displaying these location on a map in aim to help the client decide where in London to move to.

In this analysis section we will use the machine learning technique - k-means clustering, in aim to form various clusters of neighborhoods ranging from locations with zero green areas within a 200m radius, to locations with multiple green areas within a 200m radius. The aim will be to present these clusters on a map in order to display the "green" neighborhoods in relation to the center of London. The stakeholder can then use this map as a baseline to determine which areas he/she would consider to live in.

K-means clustering

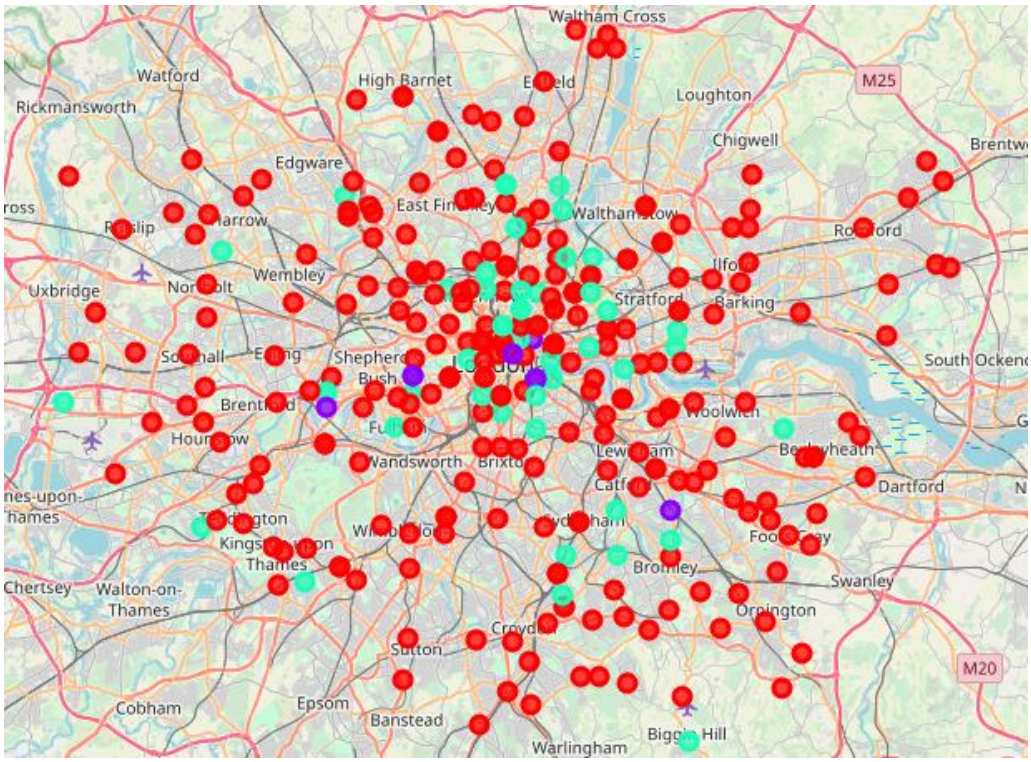
Clusters will be created in aim to separate the neighborhoods relative to their access to green areas.

After some trial and error with the number of clusters to be used, it was decided that 3 neighborhood clusters was sufficient in aim to provide a list of appropriate neighborhood candidates to the client.

Our K-means clustering analysis highlighted the following:

- Cluster 1 (red): 0 "green" areas within a 200m radius of the London neighborhood
- Cluster 2 (green): 1 "green" area within a 200m radius of the London neighborhood
- Cluster 3 (blue): 2 or more "green" areas within a 200m radius of the London neighborhood

These neighborhoods were color-coded accordingly and displayed on the London map (See Map 1 below).

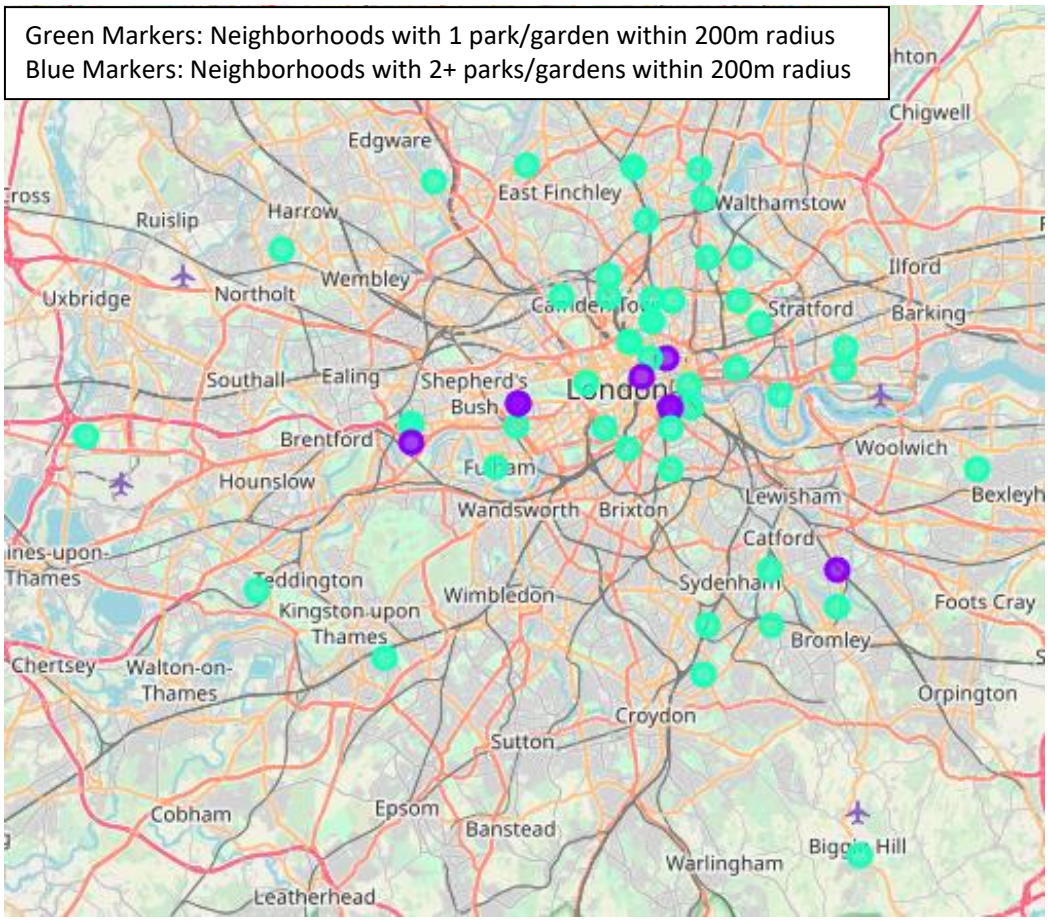


Map 1: All London Neighborhoods.

4. Results and Discussion

Our K means cluster analysis shows that 53 of the 302 London neighborhoods have at least one park or garden within a vicinity of 200m, these will be referred to as "green neighborhoods". Most London neighborhoods (249 of 302), according to this analysis, are classed as "non-green" neighborhoods, meaning that they do not have a park or garden within a 200m radius.

In reality an interactive map of all "green" neighborhoods, color coded (with markers specifying all neighborhoods) due to their "green status" along with a list of these neighborhoods is to be provided to the client concerned.



Map 2: London "Green" Neighborhoods.

index	Location	No. Green areas	index	Location	No. Green areas	index	Location	No. Green areas
0	Bankside	2	**18**	Canonbury	1	**36**	Pentonville	1
1	Barbican	2	**19**	Chiswick	1	**37**	Pimlico	1
2	Kensington	2	**20**	Church End	1	**38**	Plaistow	1
3	Newington	2	**21**	Church End	1	**39**	Plaistow	1
4	North Kensington	2	**22**	Colindale	1	**40**	South Hackney	1
5	Stepney	2	**23**	Earls Court	1	**41**	South Harrow	1
6	Temple	3	**24**	East Wickham	1	**42**	Southend	1
7	Anerley	1	**25**	Farringdon	1	**43**	St Pancras	1
8	Beckenham	1	**26**	Fulham	1	**44**	Stroud Green	1
9	Belmont	1	**27**	Grove Park	1	**45**	Tottenham	1
10	Belmont	1	**28**	Grove Park	1	**46**	Tower Hill	1
11	Belsize Park	1	**29**	Hampton	1	**47**	Tufnell Park	1
12	Berrylands	1	**30**	Harmondsworth	1	**48**	Turnpike Lane	1
13	Bloomsbury	1	**31**	Islington	1	**49**	Walworth	1
14	Camberwell	1	**32**	Kennington	1	**50**	West Hackney	1
15	Camden Town	1	**33**	King's Cross	1	**51**	Wood Green	1
16	Canary Wharf	1	**34**	Lea Bridge	1	**52**	Woodside	1
17	Canning Town	1	**35**	Mayfair	1			

Table 1: List of "Green" Neighborhoods.

The result of this analysis is that we were able to narrow down the neighborhood options for a client that intends to move to London and wants to move to an area with close access to a park and/or garden. This information will serve as a baseline or recommendation in order to help the client start further researching these neighborhoods and adding further criteria in aim to narrow down her options, e.g. distance from city center, reputation of the neighborhood etc.

5. Conclusion

The purpose of this project was to identify the London neighborhoods which had close access to greenery in order to aid a client narrow down her London neighborhood options by taking into account that she wanted to live close to a park/garden. By calculating the greenery (parks and gardens) density distribution from Foursquare data and then clustering these locations, we were able to create a short list of "green" neighborhoods to which the client could considered moving to.

The final decision on the optimal neighborhood will be made by the stakeholder based on further living criteria she deems important. The information provided to the client (Maps and tables) will serve as a solid foundation in which she can begin her decision making process.