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 located at: 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 neighbourhood 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 neighbourhood 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