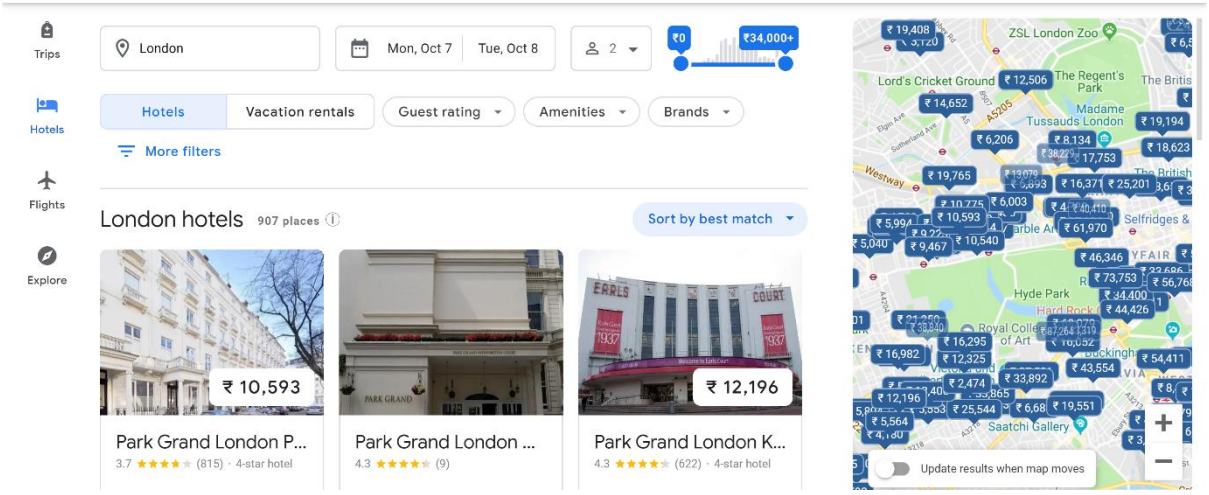
IBM Data Science Capstone Project

Finding an optimum location for a hotel in London

1. Introduction and Background

London is the most popular tourist destination in the whole world, attracting around 30 million tourists every year.[1] This is a great market for the tourism and hospitality industry with around 1500 hotels in the city.[2]

Let's say a worldwide hotel chain wants to open a new hotel in London, but is not sure of which neighbourhood it should choose. Before beginning this project, I looked up London hotels on Google Maps and found out that most popular luxury hotels are clustered around parks like Hyde Park and Regent's Park. While it is obvious why park hotels are so popular in London – the great view and open spaces for a family to wander, areas like Hyde Park and Regent's Park are already full of hotels.



Thus, I thought of analyzing which other London neighbourhoods have a variety of parks and open spaces, low competition in terms of other hotels, as well as surrounding venues which match the target demographic of the hotel chain, i.e. tourist families. As a luxury hotel chain, they would be able to afford to open at popular parkside venues as well.

Thus, the question:

In the city of London, if a worldwide hotel chain targeting tourist families is planning to open a new hotel, where would you recommend they open it?

This data may prove useful to hotel chains who are looking to expand into London's constantly expanding tourism market but also look at viable but not overcrowded locations for their property.

---

[1] "London's Tourism Industry." London's Tourism Industry, www.uncsbrp.org/tourism.htm.
[2] "Number of Hotels in London by Room Amount 2015." Statista, 2015, www.statista.com/statistics/487966/number-hotels-london-by-room-amount-united-kingdom/.

2. Data

To solve this problem, I plan to use the following data:

1. List of neighbourhoods in London, which all serve as possible options for the hotel chain - https://en.wikipedia.org/wiki/List_of_places_in_London
2. Latitude and Longitude coordinates of each of these neighbourhoods, which would help visualize them and their surroundings on a map.
3. Venue data in order to locate parks, other hotels and surrounding venues that would help in the analysis and clustering of neighbourhoods.

3. Methodology

- I scraped data about the 32 neighbourhoods in London from the Wikipedia page https://en.wikipedia.org/wiki/List_of_places_in_London with the help of Python requests and the beautifulsoup package.
- I then used the Geocoder package to get the geographical coordinates for each of the neighbourhoods.
- After that, I used the Foursquare API to get venue data for each of these neighbourhoods, i.e. surrounding landmarks, attractions and places to visit. Based on that data, I gained insights to the number of parks present in each neighbourhood.
- Using machine learning techniques like K-means clustering, I calculated and clustered neighbourhoods which have a high number of parks and open spaces and then visualized them using Folium, a map visualization library.
- By then, I had narrowed down to a cluster of 7 neighbourhoods with a high number of parks. I then checked for hotels in these neighbourhoods using the venue data, and narrowed down the list to 4 neighbourhoods with low hotel competition.
- For each of the remaining neighbourhoods, I calculated the 10 most common venues and judged which of them fit the target demographic of the hotel chain better, thus concluding the optimum neighbourhood for opening a new hotel in London.

4. Results

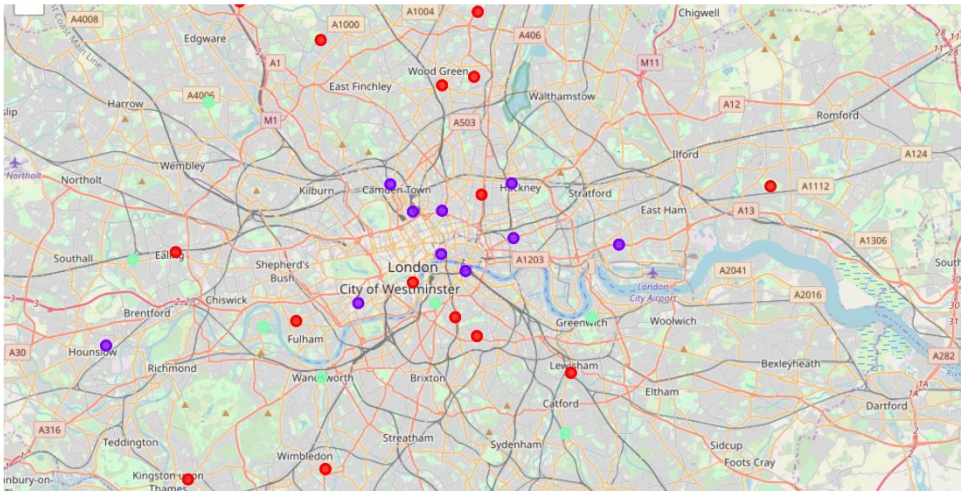


Figure 1: Neighbourhoods clustered around parks and open spaces.

Purple - low frequency, Red - medium frequency, Green - high frequency.

	Neighbourhood	Park	Cluster Labels	Latitude	Longitude
5	Camden	0.000000	1	51.532360	-0.127960
10	Hackney	0.020000	1	51.545050	-0.055320
13	Harrow	0.010000	1	51.513180	-0.106980
14	Havering	0.010000	1	51.544605	-0.144105
16	Hounslow	0.011364	1	51.471391	-0.351375
17	Islington	0.000000	1	51.532790	-0.106140
23	Newham	0.010989	1	51.517368	0.022979
26	Southwark	0.000000	1	51.505410	-0.089210
27	Sutton	0.010000	1	51.490987	-0.167417
28	Tower Hamlets	0.000000	1	51.520220	-0.054310

Figure 2: Low Frequency Cluster

	Neighbourhood	Park	Cluster Labels	Latitude	Longitude
0	Barking and Dagenham	0.054054	0	51.543932	0.133157
1	Barnet	0.045455	0	51.627294	-0.253759
2	Bexley	0.038961	0	51.622832	-0.080656
3	Brent	0.034884	0	51.609768	-0.194688
6	Croydon	0.050000	0	51.593209	-0.083390
7	Ealing	0.050000	0	51.514060	-0.300730
8	Enfield	0.030000	0	51.540021	-0.077501
11	Hammersmith and Fulham	0.040000	0	51.482690	-0.212910
12	Haringey	0.040000	0	51.589264	-0.106405
15	Hillingdon	0.040000	0	51.484225	-0.096480
19	Kingston upon Thames	0.030000	0	51.410874	-0.291946
21	Lewisham	0.050000	0	51.459160	-0.012130
22	Merton	0.040000	0	51.415640	-0.191420
24	Redbridge	0.050000	0	51.475778	-0.080698
31	Westminster	0.030000	0	51.500080	-0.128020

Figure 3: Medium Frequency Cluster

:

	Neighbourhood	Park	Cluster Labels	Latitude	Longitude
4	Bromley	0.086207	2	51.431820	-0.016566
9	Greenwich	0.060000	2	51.484540	0.002750
18	Kensington and Chelsea	0.080000	2	51.510380	-0.331470
20	Lambeth	0.070000	2	51.490840	-0.111080
25	Richmond upon Thames	0.080000	2	51.480210	-0.237180
29	Waltham Forest	0.079365	2	51.581761	-0.276969
30	Wandsworth	0.060000	2	51.456820	-0.194520

Figure 4: High Frequency Cluster

	Neighbourhoods	Hotel
4	Bromley	0.00
9	Greenwich	0.00
18	Kensington and Chelsea	0.04
20	Lambeth	0.07
25	Richmond upon Thames	0.01
29	Waltham Forest	0.00
30	Wandsworth	0.00

Figure 5: Frequency of hotels in neighbourhoods from Cluster 3

	Neighbourhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	Bromley	Grocery Store	Park	Supermarket	Fast Food Restaurant	Pub	Bus Stop	Italian Restaurant	Coffee Shop	Train Station	Platform
9	Greenwich	Pub	Park	Garden	Grocery Store	Café	Turkish Restaurant	Scenic Lookout	Furniture / Home Store	French Restaurant	History Museum
29	Waltham Forest	Indian Restaurant	Supermarket	Grocery Store	Park	Pub	Sandwich Place	Fast Food Restaurant	Coffee Shop	Gym / Fitness Center	Café
30	Wandsworth	Pub	Coffee Shop	Café	Park	Pizza Place	Gym / Fitness Center	Supermarket	Thai Restaurant	Bakery	Gym

Figure 6: Top 10 venues in the neighbourhoods with no hotels

5. Discussion

From these observations, I thought that given the target demographic of the hotel is tourist families, Bromley matches this demographic more than the other neighbourhoods. The presence of a train station and a bus stop in the neighbourhood would be an important factor for tourists. Furthermore, grocery stores and supermarkets are also more common in this neighbourhood than others. The restaurants in Bromley like Fast Food Restaurants and Italian Restaurants also cater more to the family demographic, compared to pubs and Turkish/French restaurants. Waltham Forest would be a close second in terms of shopping and restaurants, but Bromley has an added advantage of the train station and bus stop which would make travel easier for tourists and having a hotel close to both a station and a park would be a huge plus for the company.

However, this data depends on the venues and data obtained from the Foursquare API. On a sandbox API, the number of calls is limited. Furthermore, additional data should have been obtained on other factors like safety and accessibility of the neighbourhood. It also discards neighbourhoods with a lot of parks and already existing hotels. It may not be mandatory for the neighbourhood to have no nearby hotels at all for the new hotel to do well there. It also assumes that the hotel chain would want to open a hotel near a park, and not near, say the River Thames or any other popular spot. I would recommend these abovementioned areas as future possibilities to explore.

Also, if the hotel chain chooses to target a demographic like working adults and open a business hotel, the choice of neighbourhood would definitely be different and would be one located close to downtown London and its office buildings.

6. Conclusion

However, assuming the hotel is targeting tourist families with children, opening a parkside hotel would be highly recommended and Bromley would be an ideal neighbourhood to start a new hotel, given that other factors like safety and accessibility check out.