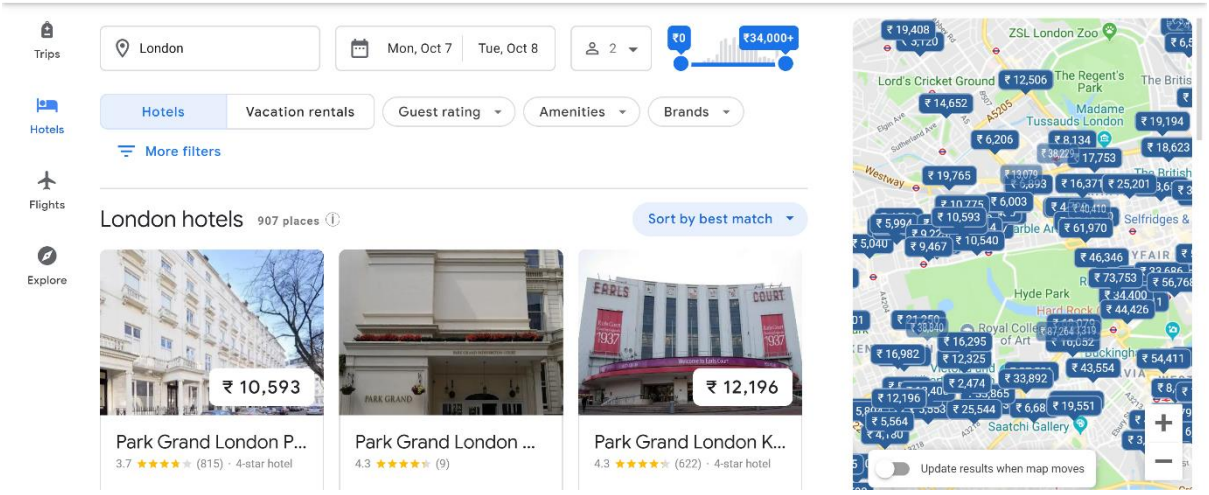
IBM Data Science Capstone Project

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.
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 plan to scrape 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 will then use the Geocoder package to get the geographical coordinates for each of the neighbourhoods. After that, I will use 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 will gain insights to the number of parks present in each neighbourhood. Using machine learning techniques like K-means clustering, I will calculate and cluster neighbourhoods which have a high number of parks and open spaces and then visualize them using Folium, a map visualization library. I will then check for hotels in these neighbourhoods using the venue data, and narrow down the list to neighbourhoods with low hotel competition. For each of the remaining neighbourhoods, I will calculate the 10 most common venues and judge which of them fit the target demographic of the hotel chain better, thus concluding the optimum neighbourhood for opening a new hotel in London.