Hotel Recommendation system

By: Noor Qarabash

Introduction

The Excel event center in London is a great venue for conventions and exhibitions, so it is important for visitors to find good hotels to stay in when visiting, in this report we will develop a simple rating score for hotels surrounding the Excel center based on user ratings, distance from the venue and the number of nearby restaurants.

This system can be used by individuals who want to attend events at the excel center as well as organizations aiming to organize events and would need ideal accommodation for staff or attendees.

Data Description

All data used in this project is from the Foursequre API but in 3 steps:

1. First, we will use the search venue option to find hotels around the Excel center
2. We will use the Id for each hotel to find the rating of said hotel
3. We will use the coordinates of each hotel to search for nearby restaurants.

All data will be appended into one data frame that we will use to compute a single score for each hotel

Methodology

This project uses content-based filtering techniques by identifying features of the hotels and using them to produce a recommendation. The process was conducted in 4 main steps, Data acquisition, Data wrangling, Score calculation, Visualization.

Data acquisition

We start the process with establishing Foursequere credentials and determining the latitude and longitude of our center location, i.e the Excel center in London,

then we search for keyword “hotel” in a radius of “10 Km”, We store the results in a dataframe that contain information about the hotels in that area.

This data frame contains details about the hotel including: ***Name, Id, Coordinates, Distance*** and many more that will be filtered because we don’t need them. What we do need is the rating and the number of nearby restaurants, and we will need to acquire them separately.

The rating for each venue acquired through foursequre API by using the id of each restaurant in the data frame as a query and storing the ratings in a list that is later appended to our main dataframe.

The number of nearby restaurants is not directly provided by foursequre but we can find it by using the search query ‘restaurants’ and a radius of 500 m we then calculate the number of returned results and store it in a list that is later appended to our main dataframe.

Data wrangling

Data cleanup and optimization was performed along with data acquisition, and can be summarized in:

1. The Hotel search results were converted to a dataframe and filtered for relevant data only, any unnecessary columns were dropped.

The resulting data frame is contains information for 30 hotels and is shown below

:

	name	lat	lng	distance	id
0	Hotels.com	51.531290	-0.106847	9738	4f869499e4b02b999b92c5b1
1	Moxy Hotel London Excel	51.508510	0.041466	883	58c3eb9d424f933529f89a16
2	Marriott Hotels International	51.517167	-0.109152	9605	4ff19e2fe4b02f36dc3fd764
3	The Tower Hotel	51.506392	-0.073223	7066	4b27f875f964a520098d24e3
4	Lucky 8 Hotels	51.559290	0.068484	6323	569cc2f5498e5a25c48f4794

2. The ratings was appended into the dataframe in column ‘Rating’. Any restaurant that has not been rated yet was dropped from the dataframe.
3. The number of nearby restaurants was appended into the dataframe in column ‘Nearby’.

4. We will also have to reset the index twice to account for the dropped fields and the sorted list.

The resulting is about 10 hotels and is shown below:

5]:

	name	lat	lng	distance	id	rating	Nearby
1	Moxy Hotel London Excel	51.508510	0.041466	883	58c3eb9d424f933529f89a16	7.2	3
3	The Tower Hotel	51.506392	-0.073223	7066	4b27f875f964a520098d24e3	7.4	9
18	Point A Hotel Shoreditch	51.522623	-0.084403	8001	58fbbe8dc21cb16c170a4cf4	7.1	19
19	Imperial Hotel	51.521697	-0.123935	10683	4b839ecdf964a5206e0b31e3	4.7	30
20	Ace Hotel Shoreditch Gym	51.525626	-0.077550	7615	52f339e8498ea23d1c2d31d8	6.8	20
22	Courthouse Hotel Shoreditch	51.527278	-0.079561	7797	5739dbf2498ede53d70af880	7.4	20
24	Club Quarters Hotel, St. Paul's	51.513882	-0.101095	9017	4b58e941f964a5205f7228e3	6.2	25
25	The Z Hotel Shoreditch	51.526960	-0.088204	8366	55061264498e0ee5ed850341	7.4	13
26	The Bridge Hotel Southwark	51.499040	-0.101780	9100	4b6c8b0ff964a520eb412ce3	5.4	8
29	The Rathbone Hotel	51.518707	-0.135557	11442	4be52dd5d4f7c9b6b8232520	5.4	30

5. Before visualization we also remove any special characters from the hotel names so it does not interfere with our code.

Score calculation

To evaluate every hotel this project establishes a rating score that combines distance, rating and restaurants nearby, the higher the score the more suitable the hotel is for prospective event attendees.

To calculate the score we first normalize the values of the 3 columns, this will eliminate any bias for example the distance has quite high values ranging from few hundreds to thousands of meters while the rating is from 1 to 10, using a Sklearn library we normalize the values that they range from 0-1

For simplicity we will assume that all 3 values have equal importance for our scoring system. Ideally, we want the highest rating and nearby values, with the lowest distance. So our equation will be:

Score = (Rating+ Nearby restaurants-distance)/2*10

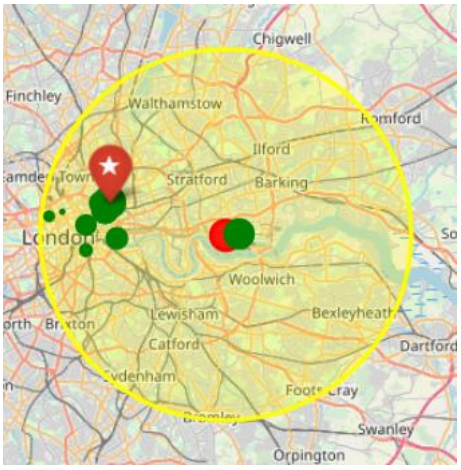
Note that we divided by 2 (being the highest possible value for the score) to normalize it then multiply by 10 to make the score out of 10 that it becomes more relevant.

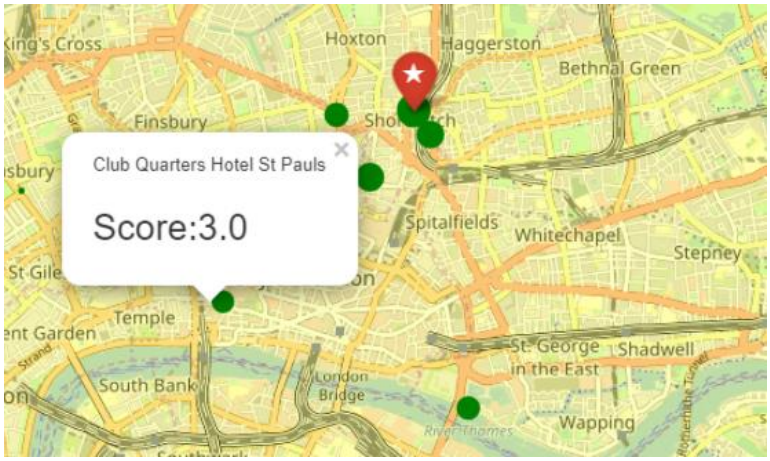
Now that we have calculated he score we can easily print out the most recommended hotels.

Visualization

Finally, we can present our results on a folium map that highlights:

1. The excel Event center in the center of the map represented by a red circle
2. The area around the excel center slightly larger than the initial search radius, represented by a transparent yellow circle.
3. The Hotels that have been scored represented by green circles with size varying based on their respective scores.
4. A Star icon representing the recommended best hotel.
5. When clicking on each hotel the pop up displays the name and score of the hotel.





Results and discussion

The result of this project is a list of hotels around the event center with scores for evaluation and recommendation

	name	lat	lng	score
0	Courthouse Hotel Shoreditch	51.527278	-0.079561	4.87
1	Moxy Hotel London Excel	51.508510	0.041466	4.63
2	Point A Hotel Shoreditch	51.522623	-0.084403	4.04
3	Ace Hotel Shoreditch Gym	51.525626	-0.077550	3.85
4	The Z Hotel Shoreditch	51.526960	-0.088204	3.31
5	The Tower Hotel	51.506392	-0.073223	3.18
6	Club Quarters Hotel St Pauls	51.513882	-0.101095	3.00
7	The Rathbone Hotel	51.518707	-0.135557	1.30
8	Imperial Hotel	51.521697	-0.123935	0.36
9	The Bridge Hotel Southwark	51.499040	-0.101780	-1.67

The results were also displayed on a map were any person interested can easily identify what the best hotels in the area are.

This project is limited by the fact that pricing for hotel accommodation changes due to peak time and availability, and had to be disregarded from the recommendation process, if it was possible to retrieve recent pricing information from foursequre it can improve on the results.

It is also possible to make the scoring equation flexible depending on the preference of the user, for example they might care more for the distance than they do the number of nearby restaurants in this case we can have variable coefficients that give different weight to each of our 3 variables.

Conclusion:

This project presented a simplified recommendation system that can offer event attendees a list of suggested hotels evaluated based on features that would make their experience of an even better. The result is a list of 10 hotels ranked based on a combination of the distance, rating and nearby restaurants attributes. The results were visualized on an easily readable map that any event attendee can use to identify the ideal hotel for them.