Recommendation of Hotel Location Selection in Inner London

Zhen Wang

1. Introduction

1.1. Background

London, is a world-famous tourist city. It attracted 20.42 million international visitors in 2018 (Tourism in London, 2019). Tourism plays an important role in UK, tourism has been the fastest growing sector in employment terms in UK since 2010 (Britain's visitor economy facts, 2019). The fast-growing tourism market means more tourists, which means more places are required for tourists to stay over the night.

1.2. Problem

This project will focus on finding an optimal location for a hotel. Specifically, since London is a large city, this report is targeted to stakeholders who are interested in opening a hotel in Inner London, UK.

2. Data Description and Cleansing

2.1. Data Components

Lee et al (2010) pointed out that in individual evaluation of hotel location model, three indexes mainly matter: safety, ease of access to transportation portals and close connection to area attractions (historic, businesses, and pleasure). Yang et al (2012) pointed out that public service infrastructure, road accessibility, subway accessibility, hotel agglomeration effect and accessibility to tourism sites were important determinants of hotel location choice. Hotel agglomeration effect means that hotel can receive benefits when they cluster together (Yang et al, 2012). Aksoy and Ozbuk (2017) mentioned some location-related attributes which influence travelers' impressions of hotel, such as traffic, security, general convenience, tourism attractions and parking.

After reviewing several literatures, **Hotel agglomeration effect**, **Accessibility to transports**, **Tourist Attractions** were selected as the key factors for selecting a hotel location in this project.

2.2. Data Sources

To be able to get three key analyzing factors, the websites were scraped as follows:

2.2.1. London Geography Related

1) London areas & postal code from **Wikipedia**
2) London areas & postal code location data from **Geocoder**
3) Inner London boroughs from **Wikipedia**
4) Download London shapefile from **London Datastore**. By using QGIS (a desktop geographic information system application), London shapefile can be transformed to GeoJson data.

2.2.2. Hotel Agglomeration Effect

Hotel location data will be obtained by using **Foursquare API**

2.2.3. Accessibility to Transports

Transports location (bus stop, bus station, light rail station, metro station, train station) data will be obtained by using **Foursquare API**

2.2.4. Tourist Attractions

Tourist attractions data will be obtained by using **TripAdvisor**

2.3. Data Cleansing

2.3.1. Foursquare Data Cleansing

After combining the London Geography data from Wikipedia and location data from Geocoder, data was scraped from Foursquare. The Foursquare data was divided into two datasets: Hotel location data and Transports location data. Yet, there were some problems with the data, lots of data points were missing neighborhood and postal code (See **Figure 1**).

Figure 1

formattedAddress	labeledLatLngs	lat	lng	neighborhood	postalCode	state
[19 Amhurst Road,, Hackney, Greater London, E8...	[{'label': 'display', 'lat': 51.54730441740626...	51.547304	-0.056107	NaN	E8 1LK	Greater London
[United Kingdom]	[{'label': 'display', 'lat': 51.53875339057779...	51.538753	-0.043925	NaN	NaN	NaN
[24-28 Charnley Road, Blackpool, FY1 4PF, Unit...	[{'label': 'display', 'lat': 51.533615, 'lng':...	51.533615	-0.039024	NaN	FY1 4PF	NaN
[106 Wick Road, Hackney, Greater London, Unite...	[{'label': 'display', 'lat': 51.53875853593553...	51.538759	-0.042126	NaN	NaN	Greater London

To make the best use of data points and make this project more accurate, QGIS was used to get this missing information. First, load the London shapefile and data points from Foursquare into QGIS, then use "Join Attributes by Location"

method (Performing Spatial Joins, 2018). Finally, neighborhood (area) and district were obtained (See **Figure 2**).

Figure 2

Foursquare Data:

formattedAddress	labeledLatLngs	lat	lng	neighborhood	postalCode	state
[19 Amhurst Road,, Hackney, Greater London, E8...	[{'label': 'display', 'lat': 51.54730441740626...	51.547304	-0.056107	NaN	E8 1LK	Greater London
[United Kingdom]	[{'label': 'display', 'lat': 51.53875339057779...	51.538753	-0.043925	NaN	NaN	NaN

Foursquare Data Join QGIS :

	area	district	hotels	categories	lat	lng
0	Hackney Central	Hackney	KIP Hotel	Hotel	51.547304	-0.056107
1	St. Mary's	Islington	The Hotel Alternative Ltd	Hotel	51.541891	-0.100508

2.3.2. TripAdvisor Data Cleansing

After scraping the Top 300 Tourist Attractions in London from TripAdvisor, Geocoder was used to get these attractions' coordinates. Also, QGIS was used to obtain the attractions' neighborhood (area) and district (See **Figure 3**). Finally, attractions located in Inner London were selected as this report targets to Inner London area.

Figure 3

TripAdvisor Data:

	Attractions
0	Tower of London
1	Churchill War Rooms
2	Tower Bridge

Geocoder

	name	lat	lng	categories
0	Tower of London	51.509740	-0.075230	Attractions
1	Churchill War Rooms	51.502250	-0.129030	Attractions
2	Tower Bridge	51.505480	-0.075380	Attractions

TripAdvisor Data Join QGIS:

	area	district	attractions	categories	lat	lng
0	St. Katharine's & Wapping	Tower Hamlets	Tower of London	Attractions	51.509740	-0.075230
1	St. James's	Westminster	Churchill War Rooms	Attractions	51.502250	-0.129030
2	North Bermondsey	Southwark	Tower Bridge	Attractions	51.505480	-0.075380

2.3.3. Data Finalize

After dropping the duplicates in these datasets, the total number for hotel is 342, transport is 574, tourist attraction is 263 in Inner London area, datasets details see Figure 4.

Figure 4

borough_hotel

	area	district	hotels	categories	lat	lng
0	Hackney Central	Hackney	KIP Hotel	Hotel	51.547304	-0.056107
1	St. Mary's	Islington	The Hotel Alternative Ltd	Hotel	51.541891	-0.100508
2	St. Mary's	Islington	Hilton London Angel Islington	Hotel	51.535605	-0.104781

borough_trans

	area	district	transports	categories	lat	lng
0	Hackney Central	Hackney	Hackney Central Station Bus Stop (H)	Bus Stop	51.546913	-0.055127
1	Homerton	Hackney	Bus Stop A	Bus Stop	51.545139	-0.055224
2	Hackney Central	Hackney	Hackney Central London Overground Station	Train Station	51.547068	-0.057052

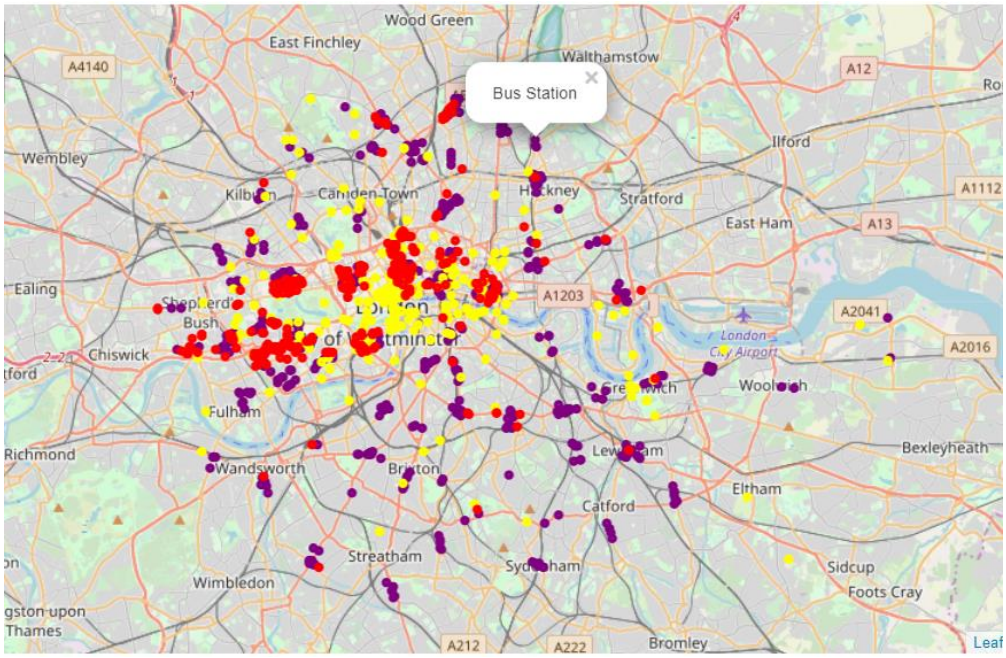
borough_attra

	area	district	attractions	categories	lat	lng
0	St. Katharine's & Wapping	Tower Hamlets	Tower of London	Attractions	51.509740	-0.075230
1	St. James's	Westminster	Churchill War Rooms	Attractions	51.502250	-0.129030
2	North Bermondsey	Southwark	Tower Bridge	Attractions	51.505480	-0.075380

3. Methodology

3.1. Data Overview – Folium Map

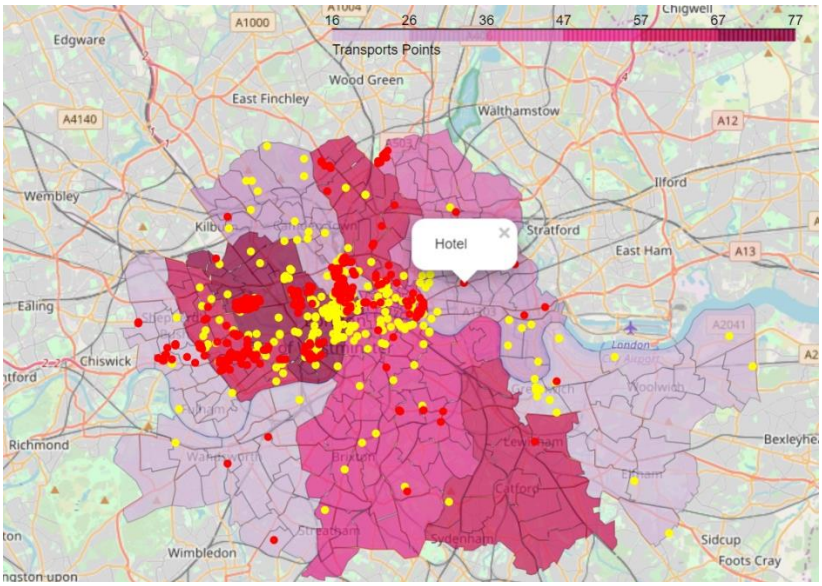
Figure 5



Folium library was used to have an overview of the geographic details of hotels', transports' and attractions' location in Inner London (see **Figure 5**). The purple points represent Transport data, the red points represent Hotel data, the yellow points represent Tourist Attraction data.

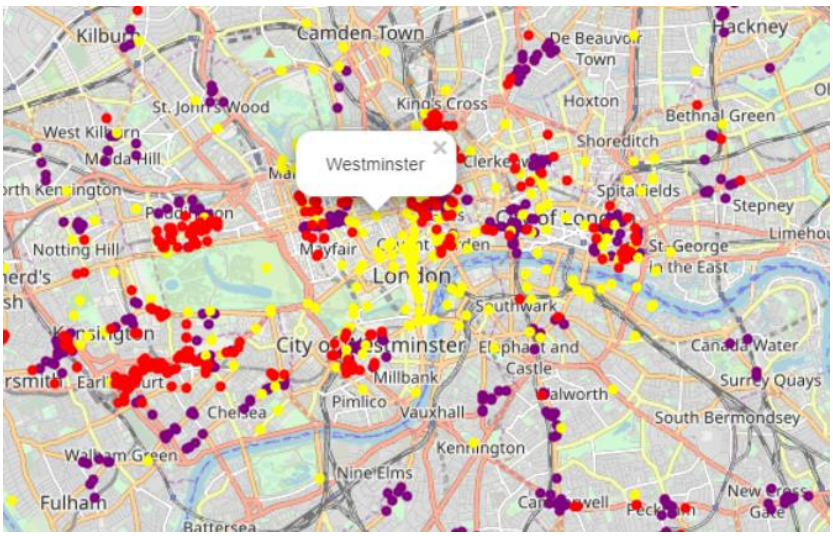
To make Inner London area edge clearer, choropleth map was superimposed onto the map by using Inner London GeoJson (See **Figure 6**). The density of Transports data is shown in Figure 6. Again, the red points represent Hotel locations, the yellow points represent Tourist Attractions in Figure 6.

Figure 6



Set "district" column as the arguments for yellow points label (See **Figure 7**), so it enables to we know which borough this point belongs to.

Figure 7



3.2. Data Summary

Merge three datasets by districts and neighborhood (area) separately.

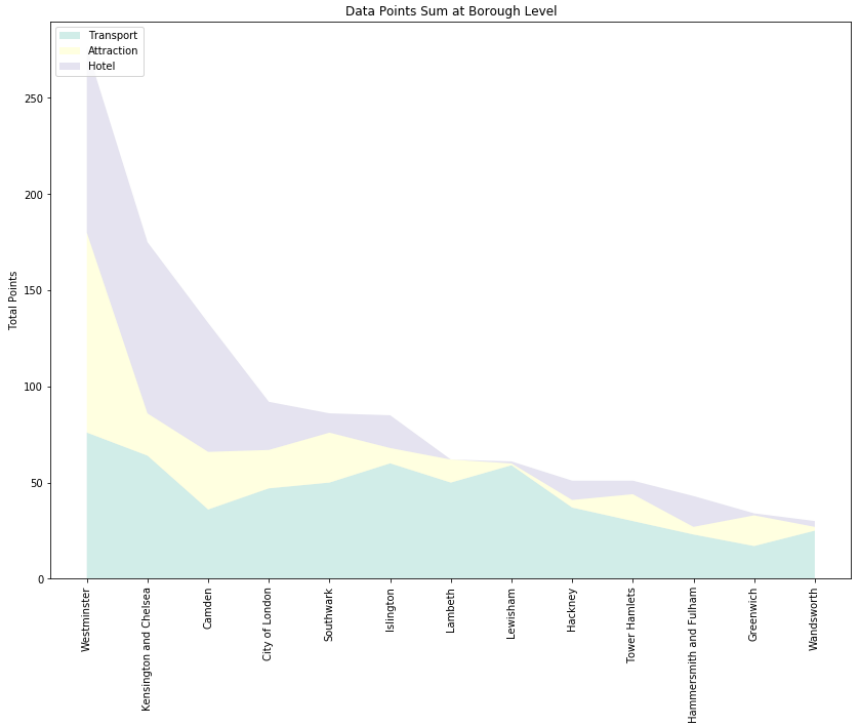
3.2.1. Area Plot - Borough Level

Sum up three data types by borough and rank the total number from maximum to minimum (See **Figure 8**). Westminster, Kensington and Chelsea, Camden, City of London and Southwark are the top 5 boroughs among the result. Area Plot was used to visualize the table in Figure 8 since it can represent cumulated totals by using numbers (See **Figure 9**). It enables segments in each borough can be clear visualized in the graph.

Figure 8

	District	Transport	Hotel	Attraction	Sum
0	Westminster	76	96.0	104	276.0
1	Kensington and Chelsea	64	89.0	22	175.0
2	Camden	36	67.0	30	133.0
3	City of London	47	25.0	20	92.0
4	Southwark	50	10.0	26	86.0
5	Islington	60	17.0	8	85.0
6	Lambeth	50	0.0	12	62.0
7	Lewisham	59	1.0	1	61.0
8	Hackney	37	10.0	4	51.0
9	Tower Hamlets	30	7.0	14	51.0
10	Hammersmith and Fulham	23	16.0	4	43.0
11	Greenwich	17	1.0	16	34.0
12	Wandsworth	25	3.0	2	30.0

Figure 9



3.2.2. Bubble Map - Neighborhood Level

Sum up three data types by neighborhood and the top 20 areas were selected for this project (See **Figure 10**).

Figure 10

Area	Transport	Hotel	Attraction	row_sum
St. James's	13	14.0	65.0	92.0
Holborn and Covent Garden	20	22.0	8.0	50.0
West End	19	16.0	14.0	49.0
Hyde Park	11	19.0	1.0	31.0
Bloomsbury	4	22.0	3.0	29.0
Brompton & Hans Town	13	9.0	6.0	28.0
Earl's Court	3	23.0	1.0	27.0
Lancaster Gate	3	21.0	1.0	25.0
Tower	8	12.0	2.0	22.0
Warwick	7	9.0	4.0	20.0
Royal Hospital	11	4.0	5.0	20.0
Junction	14	5.0	0.0	19.0
Marylebone High Street	4	13.0	2.0	19.0
Queen's Gate	3	13.0	0.0	16.0
Lewisham Central	14	1.0	0.0	15.0
Holland	3	7.0	5.0	15.0
Borough & Bankside	3	2.0	9.0	14.0
Peninsula	10	1.0	3.0	14.0
Castle Baynard	11	1.0	2.0	14.0
St. Peter's	11	2.0	0.0	13.0

Geocoder Python package was used to get these top 20 areas' coordinates. Create an Area-Districts mapping list, merge the table from figure 10 and the Area-Districts mapping list on "Area" column. At last, merge "Area" and "District" column into one column called "Area_Dis" (See **Figure 11**).

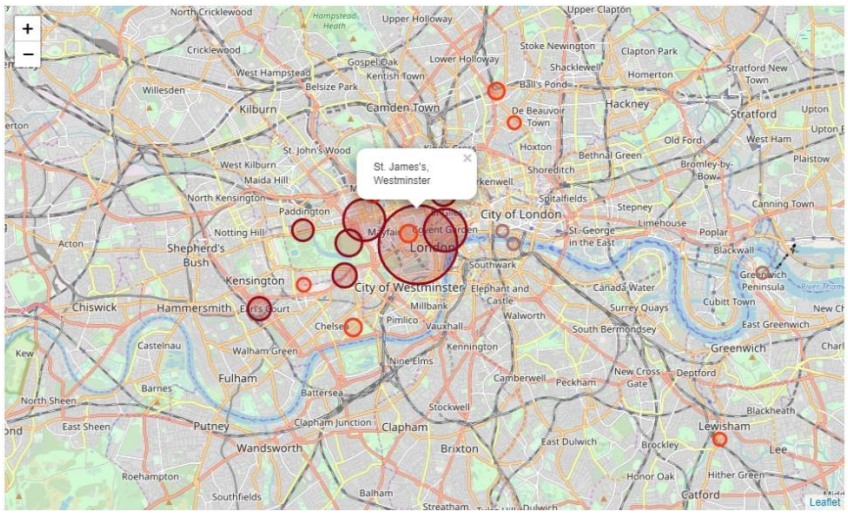
Figure 11

	Area	Transport	Hotel	Attraction	Sum	Latitude	Longitude	District	Area_Dis
0	St. James's	13	14.0	65.0	92.0	51.507820	-0.133560	Westminster	St. James's, Westminster
1	Holborn and Covent Garden	20	22.0	8.0	50.0	51.511622	-0.122485	Camden	Holborn and Covent Garden, Camden
2	West End	19	16.0	14.0	49.0	51.514140	-0.155100	Westminster	West End, Westminster
3	Hyde Park	11	19.0	1.0	31.0	51.508280	-0.161390	Westminster	Hyde Park, Westminster
4	Bloomsbury	4	22.0	3.0	29.0	51.520740	-0.123100	Camden	Bloomsbury, Camden
5	Brompton & Hans Town	13	9.0	6.0	28.0	51.500097	-0.162909	Kensington and Chelsea	Brompton & Hans Town, Kensington and Chelsea
6	Earl's Court	3	23.0	1.0	27.0	51.491870	-0.197280	Kensington and Chelsea	Earl's Court, Kensington and Chelsea
7	Lancaster Gate	3	21.0	1.0	25.0	51.511455	-0.179867	Westminster	Lancaster Gate, Westminster
8	Tower	8	12.0	2.0	22.0	51.611040	-0.109520	City of London	Tower, City of London
9	Warwick	7	9.0	4.0	20.0	51.510650	-0.137500	Westminster	Warwick, Westminster

Add bubbles to folium map. Set the "Area_Dis" column as the arguments for bubble label. Set the "Sum" column as the arguments for bubble size. Also, use the "Sum"

column to set threshold. If sum<15, set the bubble color in grey; if 15≤sum<25, set the bubble color in orange; if sum≥25, set the bubble color in red (See **Figure 12**).

Figure 12



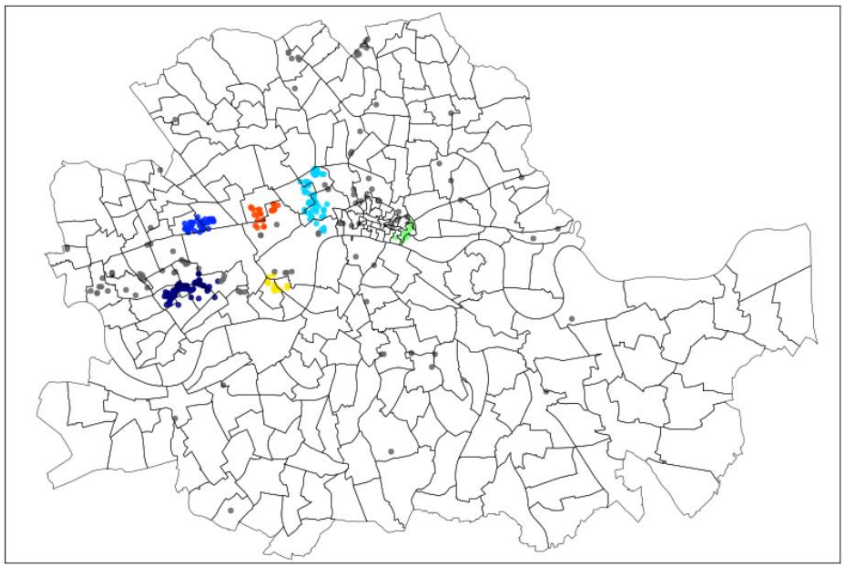
3.3. Data Clustering - DBSCAN Machine Learning

Density-Based Clustering (DBSCAN) algorithm was used here since it is especially good when dealing with spatial clusters or when there is noise in the dataset. It means that all points don't have to be assigned to a cluster if they do not belong in any. DBSCAN algorithm was used to cluster hotel, attraction and transports data separately.

3.3.1. Hotel

Hotel data points were clustered in Figure 13. Six clusters were marked.

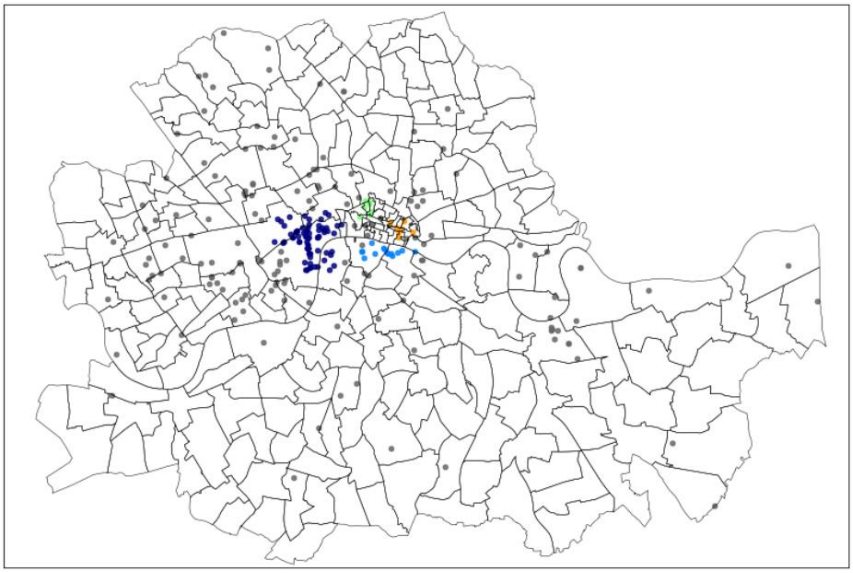
Figure 13



3.3.2. Attractions

Attractions data points were clustered in Figure 14. Four clusters were marked in the figure.

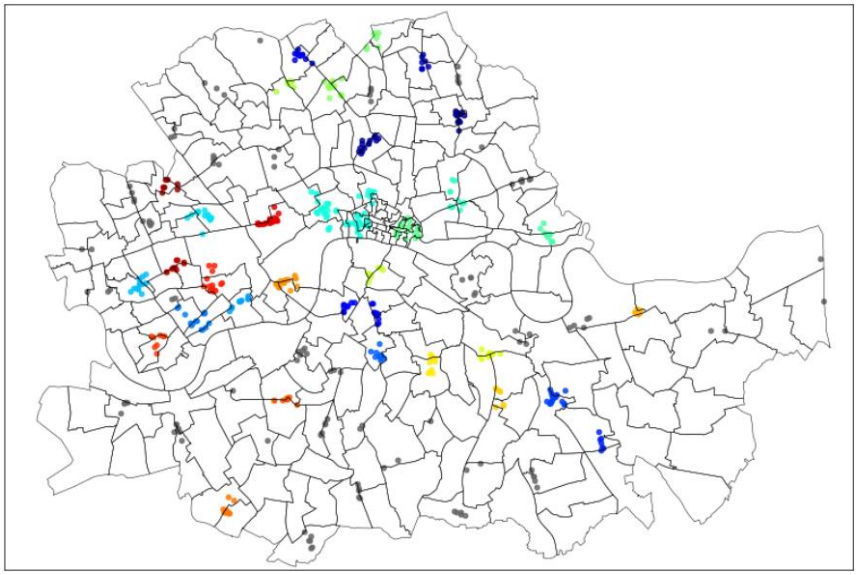
Figure 14



3.3.3. Transports

Transports data points were clustered in Figure 15. More than 20 clusters were marked in the figure.

Figure 15



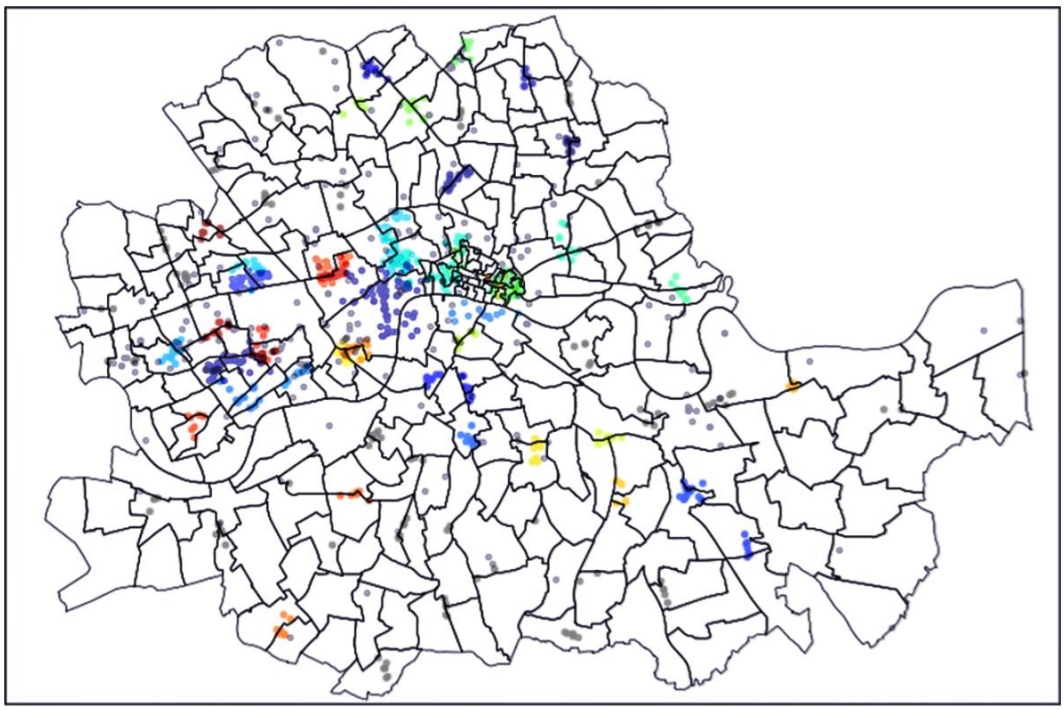
4. Results

As the results from Methodology section, Westminster, Kensington and Chelsea, Camden, City of London and Southwark are the top 5 boroughs when it comes to the sum of hotels, attractions and transports data points. St. James's, Holborn and Covent Garden, West End,

Hyde Park and Bloomsbury are the top 5 neighborhoods when it comes to the sum of hotels, attractions and transports data points.

Hotels are usually highly clustered in urban (Yang et al, 2012). As it shown in Figure 13, most of hotels are clustered together. In Figure 14 and 15, by using DBSCAN algorithm, some attractions and transports locations are also found clustered together. Figure 16 is presented after putting Figure 13, 14 and 15 into one figure.

Figure 16

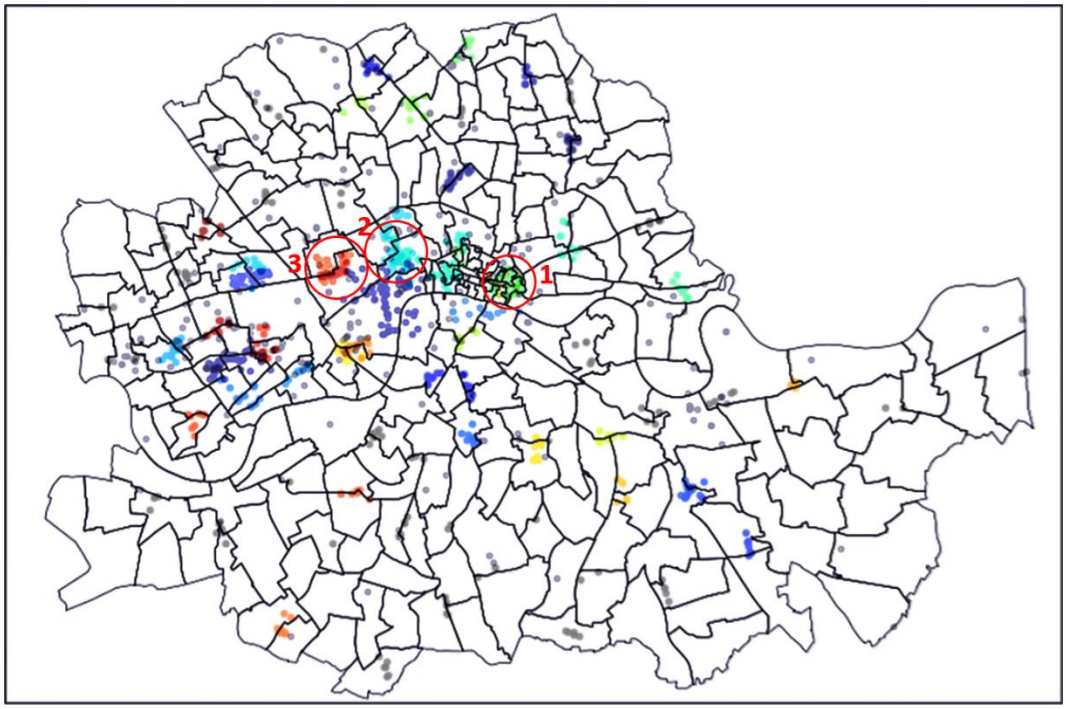


5. Discussion

5.1. Three Types of Data Overlapping

After comparing the DBSCAN results, three areas are found superposed together (See **Figure 17**). In Figure 17, all of area 1, 2 and 3 have hotel, attraction and transports cluster inside. Compare these three areas with Figure 7, it is turned out that area 1 is in City of London Borough; area 2 is in Camden Borough; area 3 is in Westminster Borough. Compare this result with Figure 8 and 9, City of London, Camden and Westminster are also listed in the top 5 boroughs.

Figure 17



More precisely, the area 1 is around the London Fenchurch St, the area 2 is around Bloomsbury neighborhood, the area 3 is around Lancaster Gate neighborhood. Both area 2 and 3 can be found in Figure 11 and 12. Area 1 cannot be found in top 20 data points sum at neighborhood level, the reason could be that even though the hotels, attractions and transports are located closely to each other, but they belong to different small neighborhoods. And because data points sum at neighborhood level was calculated by neighborhood, so the total number for one neighborhood was relatively small in area 1.

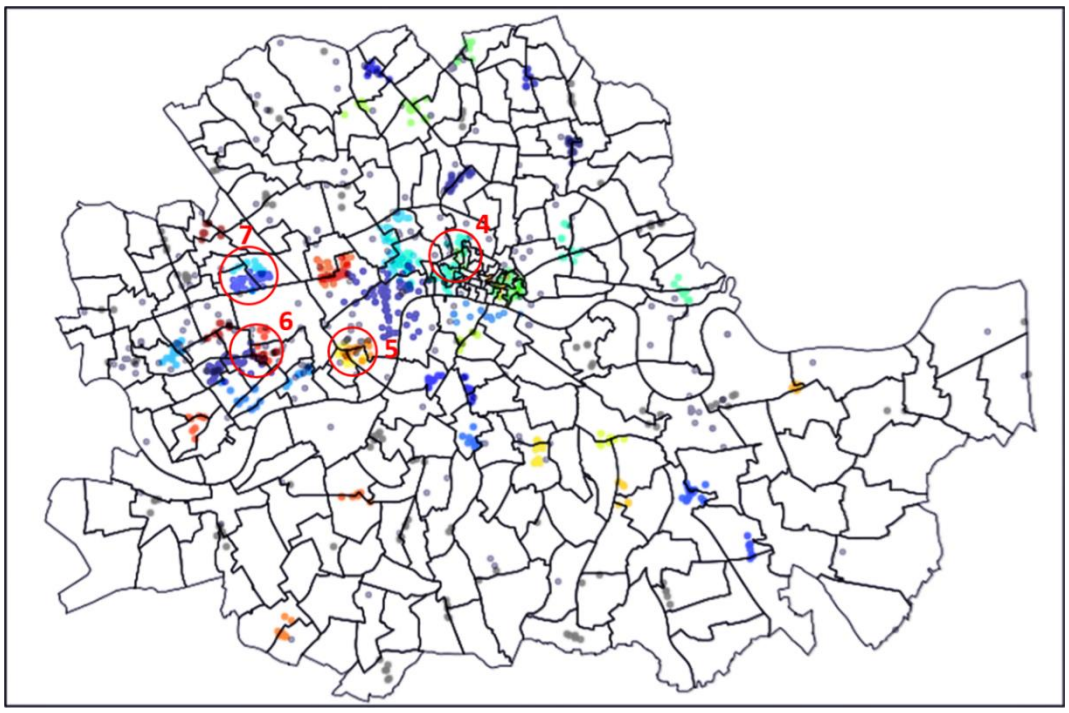
5.2. Two Types of Data Overlapping

Besides three datasets overlapping, it can be found that some areas are overlapped by two types of data, see the area 4, 5, 6 and 7 in Figure 18. Again, compare these four areas with Figure 7, it is turned out that for area 4, half of it is in City of London Borough, half of it is in Islington Borough; area 5 is in Westminster Borough; for area 6, half of it is in Westminster Borough, half of it is in Kensington and Chelsea Borough; area 7 is in Westminster Borough. Again, compare this result with Figure 8 and 9, all the boroughs mentioned just now are in the top 5 data points sum at borough level.

More precisely, area 4 is overlapped by attraction and transport data points. It is around City Thameslink. Area 5 is overlapped by hotel and transport data points. It is around Victoria St. Notice that in Figure 18, even though area 5 just has hotel and transports cluster inside, but it is very close to one of attractions clusters. Area 6 is overlapped by

hotel and transport data points. It is around Earl's Court neighborhood. Area 7 is overlapped by hotel and transport data points. It is around Lancaster Gate.

Figure 18



To summarize, area 1, 2 and 3 are in the first recommendation list to stakeholders who are interested in opening a hotel in Inner London, since all of them have three datasets superposed together. Area 5 is in the second recommendation list to stakeholders since it has hotel and transport overlapped inside and is very close to attractions. Area 4, 6 and 7 are in the third recommendation list to stakeholders since all of them just have two types datasets overlap inside. The reason of this recommendation is due to this project uses Hotel agglomeration effect, Accessibility to transports, Tourist Attractions as the key factors for selecting a hotel location. Note that stakeholder probably also needs to consider other conditions, such as real estate price and availability, security and so on. Therefore, the proposed area should only be considered as a starting point for a more detailed analysis.

6. Conclusion

This project is aim to find an optimal location for a hotel in Inner London. Foursquare data is used to have an overview of how the hotel, transport and attraction locate in Inner London first. Then, DBSCAN algorithm is used to cluster the location data from

Foursquare. Superpose three DBSCAN cluster results together to see if there are any overlapping areas. The overlapping areas were marked as the main area of interest.

This is a starting point for stakeholders, they can ultimately determine the best location for the hotel based on the hotel its own characteristics, for instance, hotel style, hotel target consumers; also the neighborhoods characteristics, for instance, the real estate price and availability, the safety situation around, public service infrastructure and so on.

References

Aksoy, S., & Ozbuk, M. Y. (2017). Multiple criteria decision making in hotel location: Does it relate to postpurchase consumer evaluations?. *Tourism management perspectives*, *22*, 73-81.

Britain's visitor economy facts. (2019). Retrieved from https://www.visitbritain.org/visitor-economy-facts.

Lee, K. W., Kim, H. B., Kim, H. S., & Lee, D. S. (2010). The determinants of factors in FIT guests' perception of hotel location. *Journal of Hospitality and Tourism Management*, *17*(1), 167-174.

Performing Spatial Joins. (2018). Retrieved from https://www.qgistutorials.com/en/docs/performing_spatial_joins.html

Tourism in London. (2019). Retrieve d from https://en.wikipedia.org/wiki/Tourism_in_London.

Yang, Y., Wong, K. K., & Wang, T. (2012). How do hotels choose their location? Evidence from hotels in Beijing. *International Journal of Hospitality Management*, *31*(3), 675-685.