The Battle of the Neighborhoods - Report

1.Introduction & Business Problem:

Problem Background:

London, one of the top tourism destinations in the world and home of many famous tourist attractions. London welcomed 20.42 million international tourists in 2018, a number that increases every year. London is also one of the biggest cities in the world with a total area of 1,572 km^2 (607 sqm). As a tourist, we usually travel for 4-5 days and we have the tendency of trying to see everything during that time. But in a city that bog with so many famous attractions how can someone plan their days?

Problem Description:

In this project I tried to identify the most important attractions in London and group them into 5 groups based on their distance. Attractions that might have been further away from the city center have not been taken into account because as mentioned before this project is referred to tourists that do not have many days in London and want to visit as many places as they can. Briefly my project will answer the questions below:

- Which attractions should I visit when in London?
- How should I plan my days (which attractions to see) according to the distance of the places?
- What kind of venues are nearby so I can visit? (for example, restaurants, hotels, spas etc)

Target Audience:

This project would be most successfully used by people who visit London for the first time and do not know of the attractions or the neighbourhoods of London. Also, it targets visitor that visit London for a short amount of time and want to visit mostly the city centre.

2. Data:

One city will be analysed in this project: London

We use data from https://www.alva.org.uk/details.cfm?p=423 to create an excel file (Figure 1) that includes all the major attractions in London, the category of the attraction (for example, museum, park), the number of visitors in 2018 and if it has free admission or the visitor need to pay a ticket.

Name of Attraction	Category	2018 Visitors	Type
Tower of London	Historic Properties	2.855.438	Paid
Royal Botanic Gardens, Kew*	Gardens	1.858.513	Paid
St Paul's Cathedral	Places of Worship	1657446	Paid
Royal Academy of Arts	Museums & Art Galleries	1.594.140	Paid
Westminster Abbey	Places of Worship	1.546.017	Paid
ZSL London Zoo	Wildlife	1.133.952	Paid
Houses of Parliament	Historic Properties	1.075.550	Paid
Hampton Court Palace	Historic Properties	902.584	Paid

Figure 1: Dataset used - London attractions

These data were modified (more information in the methodology section) to fit the needs of the project. Coordinates were added and attractions were erased that might have been away from the city centre.
We also used data from Foursquare that I used to find venues near each attraction (Figure 2).

	Name of Attraction	Attraction's Latitude	Attraction's Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	British Library	51.529877	-0.12772	The Sir John Ritblat Gallery: Treasures of the...	51.529666	-0.127541	Museum
1	British Library	51.529877	-0.12772	M&S Simply Food	51.530964	-0.125591	Grocery Store
2	British Library	51.529877	-0.12772	London St Pancras International Railway Statio...	51.531371	-0.126022	Train Station
3	British Library	51.529877	-0.12772	Pullman London St Pancras	51.528521	-0.128161	Hotel
4	British Library	51.529877	-0.12772	St. Pancras Renaissance Hotel London	51.529699	-0.125887	Hotel

Figure 2. Foursquare data

3. Methodology:

The project can be divided in 3 sections:

- Processing and visualising the dataset
- Importing data for venues near the attractions from Foursquare
- Clustering

Processing and visualising the dataset:

The first step of the project was to import the excel file with the data and arrange it into alphabetical order according to the attraction (Figure 3)

	Name of Attraction	Category	2018 Visitors	Type
0	British Library	Museums & Art Galleries	1437839	Free
1	British Museum	Museums & Art\nGalleries	5828552	Free
2	Chelsea Physic Garden	Gardens	66914	Paid
3	Churchill War Rooms	Visitor Centre	579612	Paid
4	Down House - Home of Charles Darwin	Historic Properties	50232	Paid

Figure 3. Data frame

After that we use geopy geocoders to find the coordinates of each attraction and update the data frame (Figure 4).

	Name of Attraction	Category	2018 Visitors	Type	latitude	longitude
0	British Library	Museums & Art Galleries	1437839	Free	51.529877	-0.127720
1	British Museum	Museums & Art\nGalleries	5828552	Free	51.519294	-0.128018
2	Chelsea Physic Garden	Gardens	66914	Paid	51.484585	-0.162179
3	Churchill War Rooms	Visitor Centre	579612	Paid	51.502202	-0.129349
4	Down House - Home of Charles Darwin	Historic Properties	50232	Paid	51.331449	0.053387

Figure 4

Using Folium we can visualise the dataset to understand the position of the attractions and to identify any outliers (Figure 5). In this case we define as outlier any attraction that is far away from the city centre.

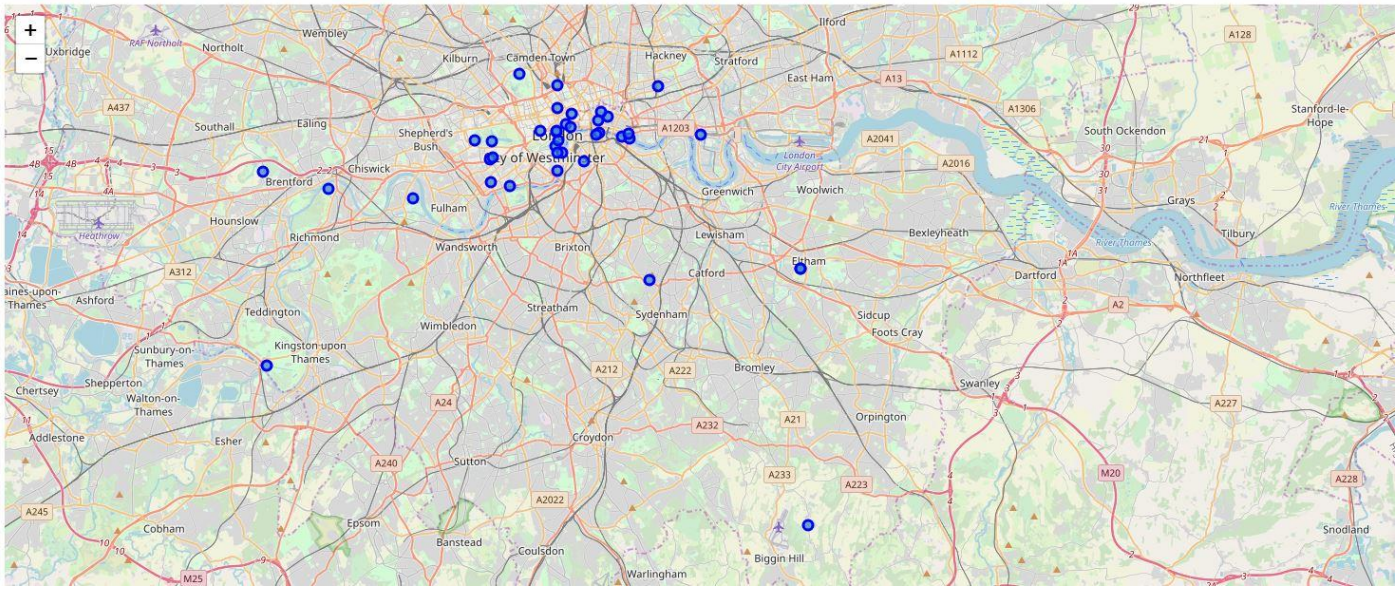

Figure 5

As we can see there are several attractions that are far from the city centre, so we modified the data frame to exclude these points. Our new map looks like this (Figure 6):

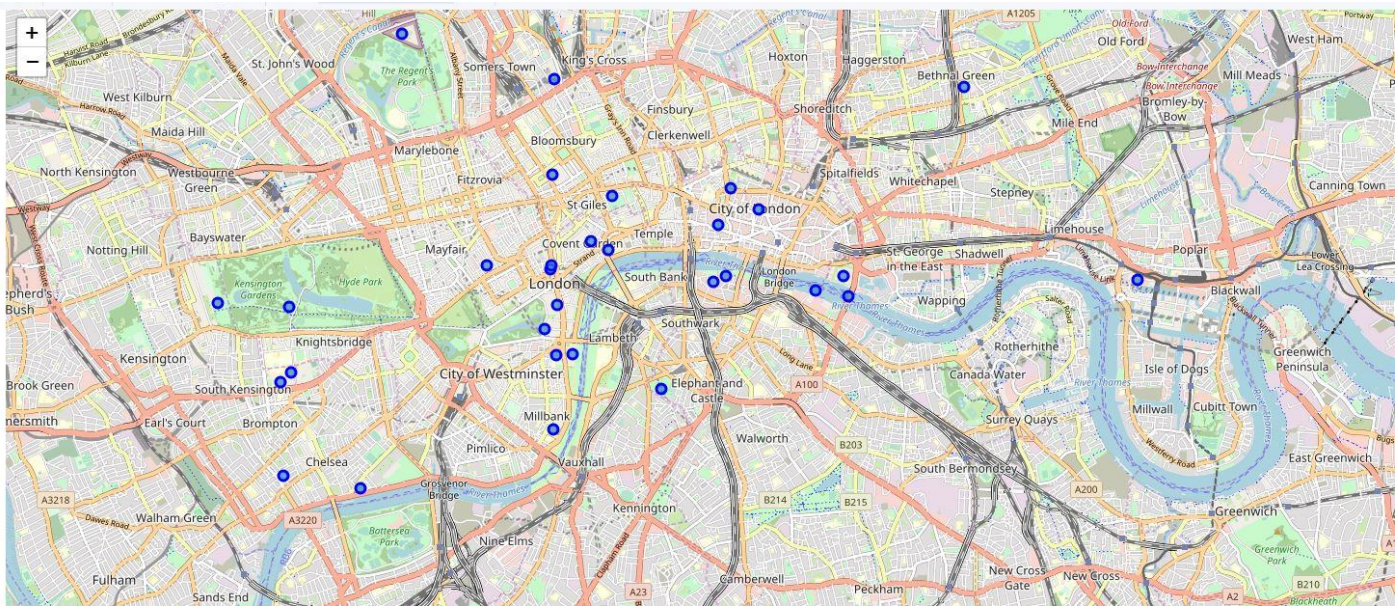

Figure 6

Importing data for venues near the attractions from Foursquare:

Using Foursquare's API we found the 100 most common venues around each attraction (Figure 7). However, categories such as museum, park were dropped because we already have them as attractions and attractions that are not so important might have been suggested.

	Name of Attraction	Attraction's Latitude	Attraction's Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	British Library	51.529877	-0.12772	The Sir John Ritblat Gallery: Treasures of the...	51.529666	-0.127541	Museum
1	British Library	51.529877	-0.12772	M&S Simply Food	51.530964	-0.125591	Grocery Store
2	British Library	51.529877	-0.12772	London St Pancras International Railway Statio...	51.531371	-0.126022	Train Station
3	British Library	51.529877	-0.12772	Pullman London St Pancras	51.528521	-0.128161	Hotel
4	British Library	51.529877	-0.12772	St. Pancras Renaissance Hotel London	51.529699	-0.125887	Hotel

Figure 7

In the bar chart below (Figure 8) we can see the number of venues according to the attraction:

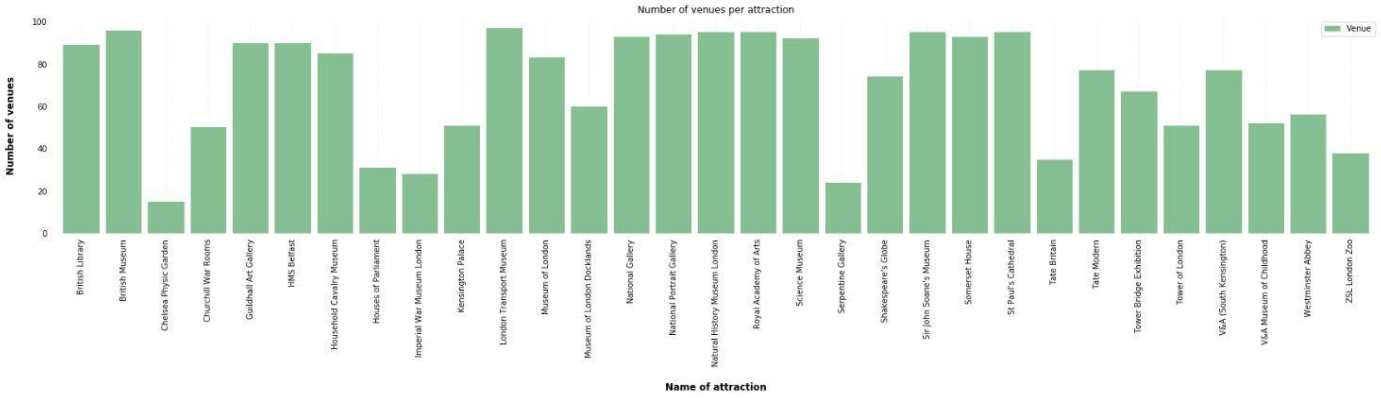

Figure 8

We also found according to each attraction what is the 5 most common venues (Figure 9)

```
----British Library----
                venue  freq
0                 Pub  0.09
1         Coffee Shop  0.09
2                Café  0.08
3               Hotel  0.07
4  Italian Restaurant  0.03


----British Museum----
          venue  freq
0   Coffee Shop  0.09
1       Exhibit  0.07
2         Hotel  0.05
3     Bookstore  0.04
4  Cocktail Bar  0.03


----Chelsea Physic Garden----
                venue  freq
0                 Pub  0.20
1       Grocery Store  0.13
2  Italian Restaurant  0.07
3     Harbor / Marina  0.07
4         Pizza Place  0.07


----Churchill War Rooms----
              venue  freq
0               Pub  0.18
1              Café  0.12
2  Outdoor Sculpture  0.10
3       Coffee Shop  0.10
4             Plaza  0.08
```

Figure 9

Afterwards, we created a data frame that shows the 10 most common venues near each attraction (Figure 10)

	Name of Attraction	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
0	British Library	Pub	Coffee Shop	Café	Hotel	Restaurant	Italian Restaurant	Sushi Restaurant	Train Station	Hotel Bar	Sandwich Place
1	British Museum	Coffee Shop	Exhibit	Hotel	Bookstore	Cocktail Bar	Bar	Japanese Restaurant	Korean Restaurant	Greek Restaurant	Gift Shop
2	Chelsea Physic Garden	Pub	Grocery Store	Pizza Place	Gym / Fitness Center	Steakhouse	French Restaurant	Italian Restaurant	Bar	Burger Joint	Harbor / Marina
3	Churchill War Rooms	Pub	Café	Coffee Shop	Outdoor Sculpture	Plaza	Hotel	Sandwich Place	Bookstore	Gift Shop	Indie Movie Theater
4	Guildhall Art Gallery	Coffee Shop	Italian Restaurant	Café	Seafood Restaurant	Gym / Fitness Center	Sushi Restaurant	Restaurant	Hotel	French Restaurant	Steakhouse
5	HMS Belfast	Coffee Shop	Hotel	Bar	Restaurant	Cocktail Bar	Pub	English Restaurant	French Restaurant	Scenic Lookout	Gym / Fitness Center
6	Household Cavalry Museum	Pub	Outdoor Sculpture	Hotel	Plaza	Coffee Shop	Café	Theater	French Restaurant	Sandwich Place	Bookstore
7	Houses of Parliament	Pub	Plaza	Café	Outdoor Sculpture	Sandwich Place	Coffee Shop	Hotel	Indian Restaurant	Bar	Concert Hall
8	Imperial War Museum London	Pub	Coffee Shop	Sandwich Place	Plaza	Hotel	Café	Bar	Dance Studio	Food & Drink Shop	Gastropub
9	Kensington Palace	Hotel	Clothing Store	Restaurant	Café	Juice Bar	French Restaurant	Grocery Store	Pub	Spa	Chinese Restaurant
10	London Transport Museum	Theater	Coffee Shop	Burger Joint	Clothing Store	Hotel	Cocktail Bar	Ice Cream Shop	Restaurant	Bakery	Dessert Shop

Figure 10

Clustering:

The final part of our code grouped attractions according to the distance between them and we used Folium to visualise these groups. To cluster the attractions into five clusters we used the K-Means clustering Algorithm. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. It uses iterative refinement approach.

Finally we can print out each group separately including the most common venues nearby, so the tourist will know which attraction is in which group and what are the most common venues.(Figures 12,13,14,15,16)

	Name of Attraction	Category	2018 Visitors	Type	Cluster Labels	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
12	Museum of London Docklands	Museums & Art\nGalleries	290488	Free	0	Coffee Shop	Italian Restaurant	Plaza	Burger Joint	Gym / Fitness Center	Indian Restaurant	Sandwich Place	Restaurant	Pizza Place	Hotel
28	V&A Museum of Childhood	Museums & Art\nGalleries	418271	Free	0	Coffee Shop	Pub	Café	Cocktail Bar	Beer Bar	Hotel	Pizza Place	Fast Food Restaurant	Grocery Store	Church

Figure 11. Group 1 - Red dots

	Name of Attraction	Category	2018 Visitors	Type	Cluster Labels	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
2	Chelsea Physic Garden	Gardens	66914	Paid	1	Pub	Grocery Store	Pizza Place	Gym / Fitness Center	Steakhouse	French Restaurant	Italian Restaurant	Bar	Burger Joint	Harbor / Marina
9	Kensington Palace	Historic Properties	550185	Paid	1	Hotel	Clothing Store	Restaurant	Café	Juice Bar	French Restaurant	Grocery Store	Pub	Spa	Chinese Restaurant
15	Natural History Museum London	Museums & Art\nGalleries	5226320	Free	1	Hotel	Exhibit	Science Museum	Café	Bakery	Ice Cream Shop	Italian Restaurant	Pizza Place	Sandwich Place	Burger Joint
17	Science Museum	Museums & Art\nGalleries	3174963	Free	1	Hotel	Café	Exhibit	Science Museum	Ice Cream Shop	Bakery	Gift Shop	Burger Joint	Italian Restaurant	Sandwich Place
18	Serpentine Gallery	Museums & Art\nGalleries	1208531	Free	1	Outdoor Sculpture	Café	Fountain	Lake	Champagne Bar	Tennis Court	Bus Stop	Bar	Restaurant	Scenic Lookout
27	V&A (South Kensington)	Museums & Art\nGalleries	3967566	Free	1	Italian Restaurant	Bakery	Juice Bar	Pizza Place	Sandwich Place	Steakhouse	Coffee Shop	Creperie	Burger Joint	Japanese Restaurant

Figure 12. Group 2 – Purple dots

	Name of Attraction	Category	2018 Visitors	Type	Cluster Labels	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
0	British Library	Museums & Art Galleries	1437839	Free	2	Pub	Coffee Shop	Café	Hotel	Restaurant	Italian Restaurant	Sushi Restaurant	Train Station	Hotel Bar	Sandwich Place
1	British Museum	Museums & Art\nGalleries	5828552	Free	2	Coffee Shop	Exhibit	Hotel	Bookstore	Cocktail Bar	Bar	Japanese Restaurant	Korean Restaurant	Greek Restaurant	Gift Shop
30	ZSL London Zoo	Wildlife	1133952	Paid	2	Zoo Exhibit	Café	Playground	Athletics & Sports	Harbor / Marina	Food & Drink Shop	Chinese Restaurant	Diner	Fountain	Gift Shop

Figure 13. Group 3 - Blue dots

	Name of Attraction	Category	2018 Visitors	Type	Cluster Labels	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	Guildhall Art Gallery	Museums & Art Galleries	116616	Free	3	Coffee Shop	Italian Restaurant	Café	Seafood Restaurant	Gym / Fitness Center	Sushi Restaurant	Restaurant	Hotel	French Restaurant	Steakhouse
5	HMS Belfast	Historic Properties	297941	Paid	3	Coffee Shop	Hotel	Bar	Restaurant	Cocktail Bar	Pub	English Restaurant	French Restaurant	Scenic Lookout	Gym / Fitness Center
11	Museum of London	Museums & Art\nGalleries	694596	Free	3	Coffee Shop	Italian Restaurant	Sandwich Place	Plaza	Sushi Restaurant	Gym / Fitness Center	French Restaurant	Hotel	Modern European Restaurant	Clothing Store
19	Shakespeare's Globe	Visitor Centre	819000	Paid	3	Coffee Shop	Gym / Fitness Center	Italian Restaurant	Pub	Hotel	Café	Asian Restaurant	Portuguese Restaurant	Spanish Restaurant	Bakery
22	St Paul's Cathedral	Places of Worship	1657446	Paid	3	Coffee Shop	Italian Restaurant	Pub	Sandwich Place	Plaza	Gym / Fitness Center	Wine Bar	Restaurant	Vietnamese Restaurant	Bakery
24	Tate Modern	Museums & Art\nGalleries	5868562	Free	3	Hotel	Gym / Fitness Center	Pub	Italian Restaurant	Coffee Shop	Performing Arts Venue	Wine Bar	Cocktail Bar	Pedestrian Plaza	Portuguese Restaurant
25	Tower Bridge Exhibition	Historic Properties	836654	Paid	3	Pub	Hotel	Coffee Shop	Italian Restaurant	Bar	English Restaurant	Scenic Lookout	Cocktail Bar	Theater	Hotel Bar
26	Tower of London	Historic Properties	2855438	Paid	3	Hotel	Gym / Fitness Center	Scenic Lookout	Coffee Shop	Tapas Restaurant	Cocktail Bar	French Restaurant	Hotel Bar	Modern European Restaurant	Beer Bar

Figure 114. Group 4 – Light blue dots

	Name of Attraction	Category	2018 Visitors	Type	Cluster Labels	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
3	Churchill War Rooms	Visitor Centre	579612	Paid	4	Pub	Café	Coffee Shop	Outdoor Sculpture	Plaza	Hotel	Sandwich Place	Bookstore	Gift Shop	Indie Movie Theater
6	Household Cavalry Museum	Museums & Art Galleries	54155	Paid	4	Pub	Outdoor Sculpture	Hotel	Plaza	Coffee Shop	Café	Theater	French Restaurant	Sandwich Place	Bookstore
7	Houses of Parliament	Historic Properties	1075550	Paid	4	Pub	Plaza	Café	Outdoor Sculpture	Sandwich Place	Coffee Shop	Hotel	Indian Restaurant	Bar	Concert Hall
8	Imperial War Museum London	Museums & Art\nGalleries	1061798	Free	4	Pub	Coffee Shop	Sandwich Place	Plaza	Hotel	Café	Bar	Dance Studio	Food & Drink Shop	Gastropub
10	London Transport Museum	Museums & Art Galleries	353921	Paid	4	Theater	Coffee Shop	Burger Joint	Clothing Store	Hotel	Cocktail Bar	Ice Cream Shop	Restaurant	Bakery	Dessert Shop
13	National Gallery	Museums & Art\nGalleries	5735831	Free	4	Theater	Hotel	Bakery	Ice Cream Shop	Pub	Wine Bar	Burger Joint	Italian Restaurant	Speakeasy	Japanese Restaurant
14	National Portrait Gallery	Museums & Art\nGalleries	1586451	Free	4	Theater	Bakery	Pub	Hotel	Ice Cream Shop	Wine Bar	Pizza Place	French Restaurant	Italian Restaurant	Speakeasy
16	Royal Academy of Arts	Museums & Art Galleries	1594140	Paid	4	Boutique	Clothing Store	Cocktail Bar	Men's Store	Women's Store	Bookstore	Dessert Shop	Hotel	Jewelry Store	Restaurant
20	Sir John Soane's Museum	Museums & Art\nGalleries	133407	Free	4	Pub	Coffee Shop	Restaurant	Sandwich Place	Hotel	Theater	Korean Restaurant	Japanese Restaurant	Italian Restaurant	Sushi Restaurant
21	Somerset House	Historic Properties	3143626	Free	4	Theater	Coffee Shop	Burger Joint	French Restaurant	Hotel	Restaurant	Cocktail Bar	Dessert Shop	Tea Room	Ice Cream Shop
23	Tate Britain	Museums & Art Galleries	1272523	Free	4	Pub	Hotel	Sandwich Place	Café	Bar	Restaurant	Coffee Shop	Exhibit	Canal Lock	Dry Cleaner
29	Westminster Abbey	Places of Worship	1546017	Paid	4	Sandwich Place	Pub	Coffee Shop	Café	Sushi Restaurant	Outdoor Sculpture	Plaza	Hotel	Italian Restaurant	Hotel Bar

Figure 125. Group 5 - Yellow dots

4. Results:

As we can see in Figure 11, we have 5 different groups (purple, red, blue, light blue and yellow dots)

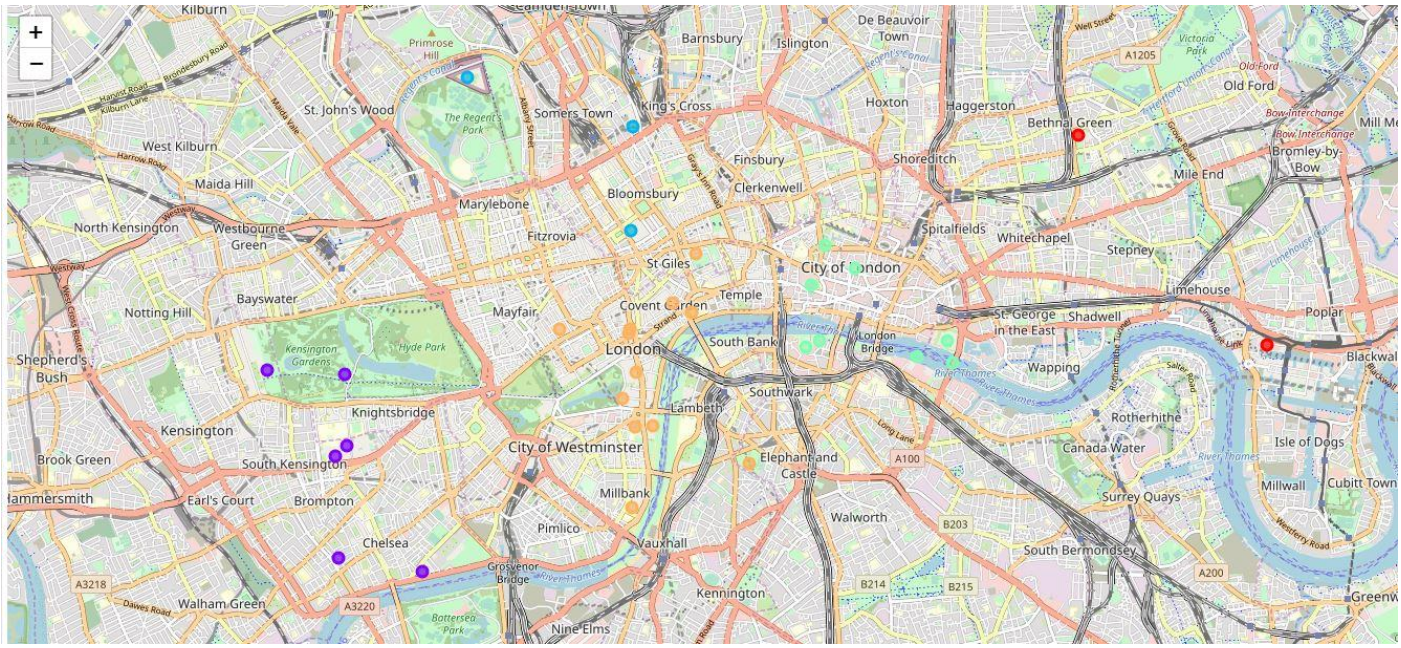

Figure 136. Map with the attractions according to groups.

Let's examine each cluster separately:

Cluster 0 or group 1 (red dots) has only 2 attractions, the two attractions that are east from Central London.
Cluster 1 or group 2 (purple dots) has 6 attractions, the ones west from Central London.
Cluster 2 or group 3 (blue dots) has 3 attractions, the ones a bit north from Central London.

We can see that Central London has too many attractions so these were divided into two groups or clusters:

Cluster 3 or group 4 (light blue dots) that has 8 attractions
Cluster 4 or group 5 (yellow dots) that has 12 attractions.

5. Discussion:

According to our analysis we can see that 5 days are quite enough to visit London – mostly Central London. If we need to check for less or more days we can just change the number of the clusters into 4 and 6 accordingly. We can see that the clusters are not equal, some clusters have 2 attractions and some clusters have 12. So that may cause a problem. However, it is obvious from Figure 16 that the cluster with the two attractions is the one that is the furthest away from the centre and also the two attractions are far away from each other, so it will take more time to commute.

Our project also shows us that the most common venues near the centre are theatres, coffee shops and pubs, making it the perfect destination for an afternoon activity or outing. On the other hand we can see that in West London the most common venues are hotels, so maybe we could propose for the tourist to stay there.

Limitations:

This project lacks to plan days to visit attractions much further away from the centre. These attractions were removed as outliers. However, if we wanted to include them, maybe for visitors that will stay more days k-means would probably not work. In order to include them we could use DBSCAN, a data clustering algorithm that groups together points that are closely packed together, marking as outliers points that lie alone in low-density regions.

Further projects:

In a following project we can also use Foursquare to identify the top rated venues near the attractions and propose them to the visitor.

6. Conclusion:

This project is meant to be used by first-time tourists in London that are supposed to travel for few days. It illustrates in a map all the major attractions and the user can also use the data to identify how many visitors visited each attraction in 2018 and if the attraction is free or not. Also, it groups attractions that are near and thus proposing ways to organise 5 days. Which attractions to see in one day. Finally, it shows the most common venues near each attraction allowing the person using the program to identify what is nearby and plan their entertainment, staying or eating.