The Battle of Neighbourhoods

Introduction

In this project, I will apply some key techniques and tools learnt in the series of Data Science courses, such as Python data analysis, data visualization in graphs and maps, Foursquare API and machine learning technique specifically K-means clustering, to analyze a practical problem of picking location.

Background

London, a fascinating metropolis with the world's best attractions, is one of the most visited city all over the globe. Feeling like international travel having been restricted for like a decade, I miss my previous stays in the city so bad and can't help looking forward to the next visit. If I have one day in London, this would be it – a park in the morning, a museum or art gallery to follow, a Japanese restaurant for meal and a musical/live performance in a theater to end.

Business Problem

To better fulfill the day, I'd like to do some trip planning to figure out where to go with all the four types of interested venues nearby.

As I usually visit a city around by public transport, the target of this analysis will be a Tube station, specifically in Zone 1, where most of the attractions and sites are located.

Audience

The analysis is expected to be easily applicable to trip planning for any tourist with customized choice of city and venues of interest.

Data Preparation

1. Extract the list of London Tube stations in Zone 1 from wiki webpage using BeautifulSoup and transform to Pandas data frame https://en.wikipedia.org/wiki/List_of_stations_in_London_fare_zone_1

   There are 84 tube stations and 9 boroughs in Zone 1.

	Station	Local authority	Managed by	1981 to 1983[6]	Notes
0	Aldgate	City of London	London Underground	City	
1	Aldgate East	Tower Hamlets	London Underground	City	
2	Angel	Islington	London Underground	City	
3	Baker Street	Westminster	London Underground	West End	
4	Bank	City of London	London Underground	City	

	Boroughs
0	City of London
1	Tower Hamlets
2	Islington
3	Westminster
4	Southwark
5	Camden
6	Kensington & Chelsea
7	Hackney
8	Lambeth

2. Look for stations in the most popular boroughs to foreign and domestic tourists to trim down the scope.

   According to statistics of tourism trips by borough in Greater London from https://data.gov.uk/dataset/ee5038be-d2be-4ab6-a612-70ade60eca12/tourism-trips-borough, Westminster, Kensington & Chelsea, Camden, City of London and Tower Hamlets are the top 5 popular boroughs in Zone 1.

3. Filter out Zone-1 Tube stations managed by London Underground in the 5 most popular boroughs.

   The number of stations is down to 54.

	Station	Local authority	Managed by
0	Aldgate	City of London	London Underground
1	Aldgate East	Tower Hamlets	London Underground
2	Baker Street	Westminster	London Underground
3	Bank	City of London	London Underground
4	Barbican	City of London	London Underground

4. The list of stations is further narrowed down based on annual entries and exits by passengers of a station, importing from xls file *https://data.london.gov.uk/download/london-underground-performance-reports/b6ab04fc-9062-4291-b514-7fa218073b4c/multi-year-station-entry-and-exit-figures.xls* and converting into a data frame.

5. The least busy stations with annual entries and exits by passengers less than 5mn are dropped out.

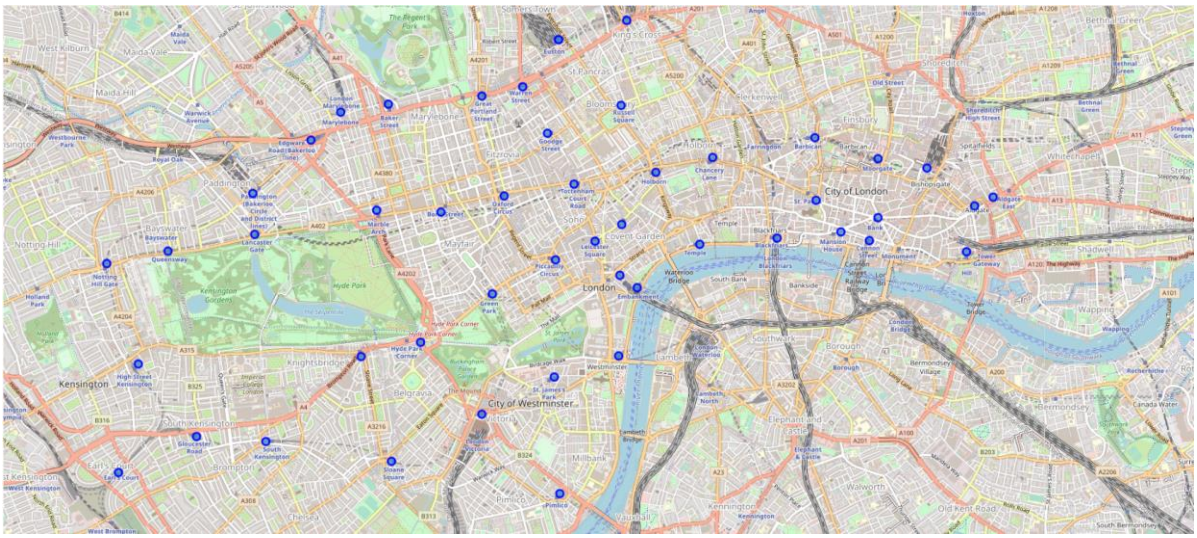
   There are 49 stations remaining.

	Station	Local authority	Managed by	PAX mn
42	Queensway	Westminster	London Underground	8.523970
43	Great Portland Street	Westminster	London Underground	8.081276
44	Goodge Street	Camden	London Underground	7.524287
45	Edgware Road (Cir)	Westminster	London Underground	7.148956
46	Mansion House	City of London	London Underground	6.405554
47	Lancaster Gate	Westminster	London Underground	6.243768
48	Hyde Park Corner	Westminster	London Underground	5.538477
49	Bayswater	Westminster	London Underground	4.714236
50	Edgware Road (Bak)	Westminster	London Underground	4.623646
51	Regent's Park	Westminster	London Underground	3.611762

6. Get geo codes of Tube stations from https://www.doogal.co.uk/LondonStationsCSV.ashx and merge with the existing data frame

	Station	Local authority	Managed by	PAX mn	Latitude	Longitude	Zone	Postcode
0	Aldgate	City of London	London Underground	8.846940	51.514342	-0.075627	1	EC3N 1AH
1	Aldgate East	Tower Hamlets	London Underground	13.998292	51.515082	-0.073001	1	E1 7PT
2	Baker Street	Westminster	London Underground	28.784591	51.523130	-0.156904	1	NW1 5LA
3	Bank	City of London	London Underground	61.796153	51.513347	-0.089000	1	EC3V 3LA
4	Barbican	City of London	London Underground	11.827212	51.520215	-0.097722	1	EC1A 4JA
5	Blackfriars	City of London	London Underground	14.825714	51.511587	-0.102995	1	EC4V 4DD

7. Now, the dataset of select Tube stations is ready to explore nearby venues via Foursquare API.



Methodology (Data Exploration and Analysis)

i. Foursquare

1. Obtain nearby venues around each select Tube station from Foursquare

```
In [34]: nearby_venues = getNearbyVenues(names=df_Z1_topborpax['Station'], latitudes=df_Z1_topborpax['Latitude'], longitudes=df_Z1_topborpax['Longitude'])
```

```
In [35]: nearby_venues.head()
```

Out[35]:

	Station	Station Latitude	Station Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Distance	Venue Category
0	Aldgate	51.514342	-0.075627	Hotel Indigo	4c1a886d8b3aa593c5c3955f	51.512740	-0.075920	179	Hotel
1	Aldgate	51.514342	-0.075627	Dorsett City London	598f92741ffed779d8055669	51.514036	-0.075812	36	Hotel
2	Aldgate	51.514342	-0.075627	Discount Suit Company	52c34344498ef50111c018ef	51.516705	-0.075506	263	Cocktail Bar
3	Aldgate	51.514342	-0.075627	The Association	4f70a772e4b0f375fc669005	51.513733	-0.079132	252	Coffee Shop
4	Aldgate	51.514342	-0.075627	1Rebel	53749f5c498e46fef6b4c193	51.515569	-0.080040	334	Gym / Fitness Center

```
In [36]: nearby_venues.shape
```

```
Out[36]: (4900, 9)
```

A total of 4900 nearby venues in total are returned.

2. Look up the venue categories that I am interested in from Foursquare website *https://developer.foursquare.com/docs/build-with-foursquare/categories/*, which include 'Park', 'Art Gallery', 'Art Museum', 'History Museum', 'Japanese Restaurant', 'Opera House' and 'Theater'.

3. Create a data frame that contains only the nearby venues of selected categories.

```
category_select = ['Park', 'Art Gallery', 'Art Museum', 'History Museum', 'Japanese Restaurant', 'Opera House', 'Theater']
nearby_venues_select = nearby_venues[nearby_venues['Venue Category'].isin(category_select)].reset_index(drop=True)
nearby_venues_select.shape
```

```
Out[37]: (324, 9)
```

There are 324 such venues.

4. Check if there are adequate venues for each category of interest by looking at the number of venues by category, and it looks so.

```
In [38]: nearby_venues_select.groupby('Venue Category').count()
         # nearby_venues_select.groupby('Station').count()
```

Out[38]:

Venue Category	Station	Station Latitude	Station Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Distance
Art Gallery	60	60	60	60	60	60	60	60
Art Museum	28	28	28	28	28	28	28	28
History Museum	28	28	28	28	28	28	28	28
Japanese Restaurant	52	52	52	52	52	52	52	52
Opera House	7	7	7	7	7	7	7	7
Park	66	66	66	66	66	66	66	66
Theater	83	83	83	83	83	83	83	83

5. Summarise venues by category for each station applying the onehot code method.

```
In [40]: nearby_venues_select_grouped = nearby_venues_select_onehot.groupby('Station').sum().reset_index()
         nearby_venues_select_grouped.head()
```

Out[40]:

	Station	Art Gallery	Art Museum	History Museum	Japanese Restaurant	Opera House	Park	Theater
0	Aldgate	2	0	0	1	0	0	0
1	Aldgate East	2	0	0	2	0	0	0
2	Baker Street	1	0	1	1	0	1	1
3	Bank	2	0	1	0	0	1	2
4	Barbican	2	0	1	0	0	1	1

6. Group (i) Art Museum and History Museum into Museum; and (ii) Museum and Art Gallery into a combined Art & Museum category.

   Create a score for each station based on the total number of venues in selected categories.

```
In [41]: nearby_venues_select_grouped['Museum'] = nearby_venues_select_grouped['Art Museum'] + nearby_venues_select_grouped['History Museum']
         nearby_venues_select_grouped['Art/Museum'] = nearby_venues_select_grouped['Art Gallery'] + nearby_venues_select_grouped['Museum']
         nearby_venues_select_grouped['Performance'] = nearby_venues_select_grouped['Opera House'] + nearby_venues_select_grouped['Theater']
         nearby_venues_select_grouped['Total'] = nearby_venues_select_grouped['Art/Museum'] + nearby_venues_select_grouped['Japanese Restaurant'] + nearby_venues_select_grouped['Park'] + nearby_venues_select_grouped['Performance']
         nearby_venues_select_grouped['Score'] = nearby_venues_select_grouped['Total']/nearby_venues_select_grouped["Total"].sum()
         nearby_venues_select_grouped.head()
```

Out[41]:

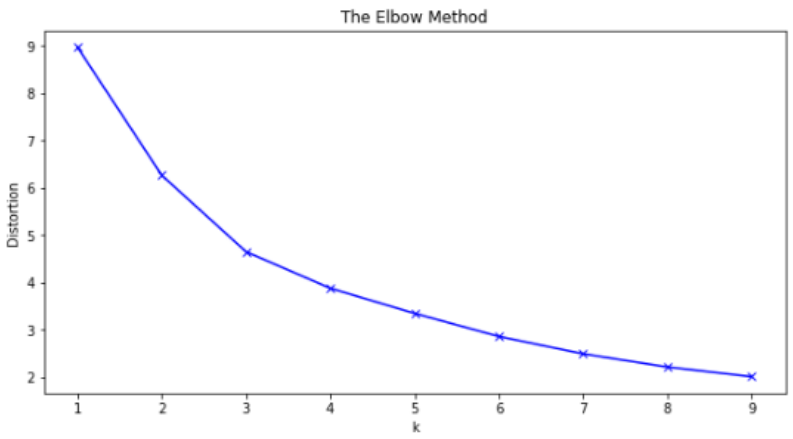
	Station	Art Gallery	Art Museum	History Museum	Japanese Restaurant	Opera House	Park	Theater	Museum	Art/Museum	Performance	Total	Score
0	Aldgate	2	0	0	1	0	0	0	0	2	0	3	0.009259
1	Aldgate East	2	0	0	2	0	0	0	0	2	0	4	0.012346
2	Baker Street	1	0	1	1	0	1	1	1	2	1	5	0.015432
3	Bank	2	0	1	0	0	1	2	1	3	2	6	0.018519
4	Barbican	2	0	1	0	0	1	1	1	3	1	5	0.015432

ii. **K-means clustering**

1. Determine the best cluster number K

   The Elbow Method suggest the best K around 5

```
In [44]: plt.figure(figsize=(10,5))
         plt.plot(K, distortions, 'bx-')
         plt.xlabel('k')
         plt.ylabel('Distortion')
         plt.title('The Elbow Method')
         plt.show()
```



2. Perform K-Means clustering and add resultant cluster labels for each station

```
In [45]: kclusters = 5

kmeans = KMeans(n_clusters=kclusters, init='k-means++', max_iter=15, random_state=8).fit(k_test)

kmeans.labels_[0:10]
```

```
Out[45]: array([0, 1, 0, 0, 0, 0, 3, 2, 0, 2], dtype=int32)
```

```
In [46]: nearby_venues_select_grouped.insert(0, 'Cluster Labels', kmeans.labels_)
nearby_venues_select_grouped = pd.merge(df_Z1_topborpax, nearby_venues_select_grouped, on='Station')
nearby_venues_select_grouped.head()
```

Out[46]:

	Station	Local authority	Managed by	PAX mn	Latitude	Longitude	Zone	Postcode	Cluster Labels	Art Gallery	Art Museum	History Museum	Japanese Restaurant	Opera House	Park	Theater	Museum	Art/Museum	Performance	Total	Score
0	Aldgate	City of London	London Underground	8.846940	51.514342	-0.075627	1	EC3N 1AH	0	2	0	0	1	0	0	0	0	2	0	3	0.009259
1	Aldgate East	Tower Hamlets	London Underground	13.998292	51.515082	-0.073001	1	E1 7PT	1	2	0	0	2	0	0	0	0	2	0	4	0.012346
2	Baker Street	Westminster	London Underground	28.784591	51.523130	-0.156904	1	NW1 5LA	0	1	0	1	1	0	1	1	1	2	1	5	0.015432
3	Bank	City of London	London Underground	61.796153	51.513347	-0.089000	1	EC3V 3LA	0	2	0	1	0	0	1	2	1	3	2	6	0.018519
4	Barbican	City of London	London Underground	11.827212	51.520215	-0.097722	1	EC1A 4JA	0	2	0	1	0	0	1	1	1	3	1	5	0.015432

3. Identify stations with at least one venue for each of the food, park, art/museum and performance category nearby, and remove stations that lack any.

```
In [47]: stations_select = nearby_venues_select_grouped.loc[(nearby_venues_select_grouped[['Japanese Restaurant', 'Park', 'Art/Museum', 'Performance']] != 0). all(axis=1)]
stations_select.sort_values(['Score'], ascending=False).reset_index(drop=True)
```

Out[47]:

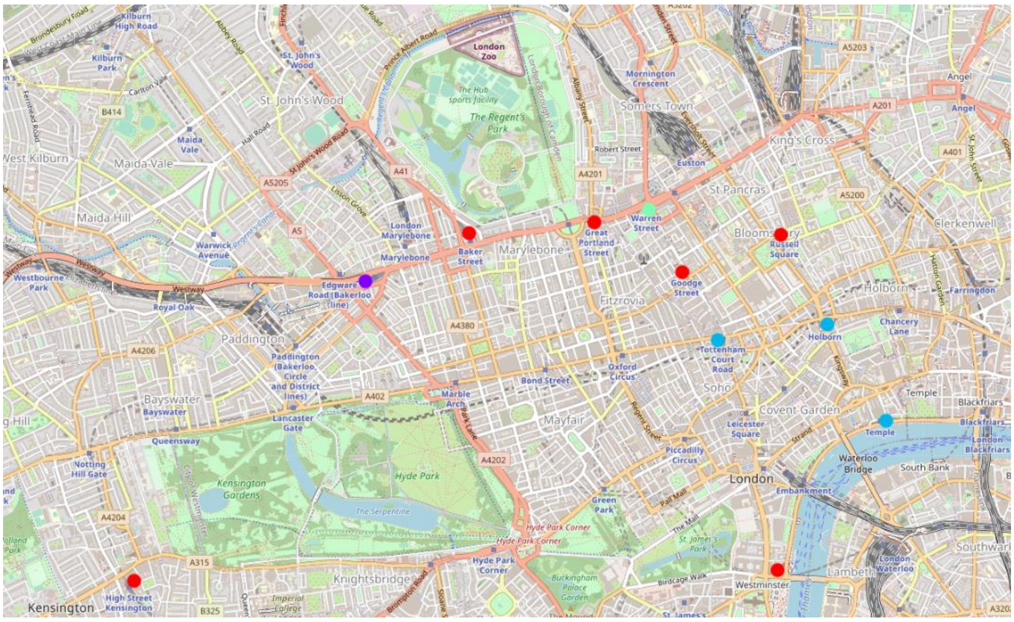
	Station	Local authority	Managed by	PAX mn	Latitude	Longitude	Zone	Postcode	Cluster Labels	Art Gallery	Art Museum	History Museum	Japanese Restaurant	Opera House	Park	Theater	Museum	Art/Museum	Performance	Total	Score
0	Temple	Westminster	London Underground	8.946533	51.511041	-0.113726	1	WC2R 2PH	2	2	0	2	2	1	3	8	2	4	9	18	0.055556
1	Russell Square	Camden	London Underground	11.449159	51.523013	-0.124590	1	WC1N 1LG	0	1	0	4	3	0	4	2	4	5	2	14	0.043210
2	Holborn	Camden	London Underground	32.538456	51.517236	-0.119841	1	WC2B 6AA	2	1	0	2	2	1	1	5	2	3	6	12	0.037037
3	Westminster	Westminster	London Underground	25.595405	51.501402	-0.125002	1	SW1A 2JR	0	3	2	0	1	0	3	2	2	5	2	11	0.033951
4	Tottenham Court Road	Westminster	London Underground	41.334148	51.516211	-0.131110	1	W1D 2DA	2	1	0	1	1	0	1	5	1	2	5	9	0.027778
5	Goodge Street	Camden	London Underground	7.524287	51.520613	-0.134808	1	W1T 2HF	0	1	0	1	1	0	1	2	1	2	2	6	0.018519
6	Baker Street	Westminster	London Underground	28.784591	51.523130	-0.156904	1	NW1 5LA	0	1	0	1	1	0	1	1	1	2	1	5	0.015432
7	Edgware Road (Cir)	Westminster	London Underground	7.148956	51.519998	-0.167668	1	NW1 5DH	1	1	0	0	2	0	1	1	0	1	1	5	0.015432
8	Great Portland Street	Westminster	London Underground	8.081276	51.523789	-0.143961	1	W1W 5PP	0	2	0	0	1	0	1	1	0	2	1	5	0.015432
9	High Street Kensington	Kensington & Chelsea	London Underground	12.762568	51.500729	-0.191566	1	W8 5SA	0	1	0	1	1	1	1	0	1	2	1	5	0.015432
10	Warren Street	Camden	London Underground	20.110864	51.524580	-0.138278	1	NW1 3AA	3	1	0	0	1	0	2	1	0	1	1	5	0.015432

As shown above, there are 10 stations that satisfy the requirement, in clusters 0, 1, 2 and 3. Not any station in cluster 4 is in the list, since not venues of all categories are available around.

Out[49]:

	Station	Local authority	Managed by	PAX mn	Latitude	Longitude	Zone	Postcode	Cluster Labels	Art Gallery	Art Museum	History Museum	Japanese Restaurant	Opera House	Park	Theater	Museum	Art/Museum	Performance	Total	Score
36	Pimlico	Westminster	London Underground	10.971039	51.489553	-0.133108	1	SW1V 2JA	4	0	1	0	0	0	6	2	1	1	2	9	0.027778
37	Queensway	Westminster	London Underground	8.523970	51.510449	-0.187420	1	W2 4SS	4	0	0	0	0	0	1	0	0	0	0	1	0.003086

4. Display the clustering results in the map



5. Look into more detailed characteristics of each cluster by summary tables and charts

```
In [51]: stations_select.loc[stations_select['Cluster Labels'] == 0].sort_values(['Score'], ascending=False)
```

Out[51]:

	Station	Local authority	Managed by	PAX mn	Latitude	Longitude	Zone	Postcode	Cluster Labels	Art Gallery	Art Museum	History Museum	Japanese Restaurant	Opera House	Park	Theater	Museum	Art/Museum	Performance	Total	Score
38	Russell Square	Camden	London Underground	11.449159	51.523013	-0.124590	1	WC1N 1LG	0	1	0	4	3	0	4	2	4	5	2	14	0.043210
48	Westminster	Westminster	London Underground	25.595405	51.501402	-0.125002	1	SW1A 2JR	0	3	2	0	1	0	3	2	2	5	2	11	0.033951
17	Goodge Street	Camden	London Underground	7.524287	51.520613	-0.134808	1	W1T 2HF	0	1	0	1	1	0	1	2	1	2	2	6	0.018519
2	Baker Street	Westminster	London Underground	28.784591	51.523130	-0.156904	1	NW1 5LA	0	1	0	1	1	0	1	1	1	2	1	5	0.015432
18	Great Portland Street	Westminster	London Underground	8.081276	51.523789	-0.143961	1	W1W 5PP	0	2	0	0	1	0	1	1	0	2	1	5	0.015432
20	High Street Kensington	Kensington & Chelsea	London Underground	12.762568	51.500729	-0.191566	1	W8 5SA	0	1	0	1	1	1	1	0	1	2	1	5	0.015432

```
In [52]: stations_select.loc[stations_select['Cluster Labels'] == 1].sort_values(['Score'], ascending=False)
```

Out[52]:

	Station	Local authority	Managed by	PAX mn	Latitude	Longitude	Zone	Postcode	Cluster Labels	Art Gallery	Art Museum	History Museum	Japanese Restaurant	Opera House	Park	Theater	Museum	Art/Museum	Performance	Total	Score
12	Edgware Road (Cir)	Westminster	London Underground	7.148956	51.519998	-0.167668	1	NW1 5DH	1	1	0	0	2	0	1	1	0	1	1	5	0.015432

```
In [53]: stations_select.loc[stations_select['Cluster Labels'] == 2].sort_values(['Score'], ascending=False)
```

Out[53]:

	Station	Local authority	Managed by	PAX mn	Latitude	Longitude	Zone	Postcode	Cluster Labels	Art Gallery	Art Museum	History Museum	Japanese Restaurant	Opera House	Park	Theater	Museum	Art/Museum	Performance	Total	Score
43	Temple	Westminster	London Underground	8.946533	51.511041	-0.113726	1	WC2R 2PH	2	2	0	2	2	1	3	8	2	4	9	18	0.055556
21	Holborn	Camden	London Underground	32.538456	51.517236	-0.119841	1	WC2B 6AA	2	1	0	2	2	1	1	5	2	3	6	12	0.037037
44	Tottenham Court Road	Westminster	London Underground	41.334148	51.516211	-0.131110	1	W1D 2DA	2	1	0	1	1	0	1	5	1	2	5	9	0.027778

```
In [54]: stations_select.loc[stations_select['Cluster Labels'] == 3].sort_values(['Score'], ascending=False)
```

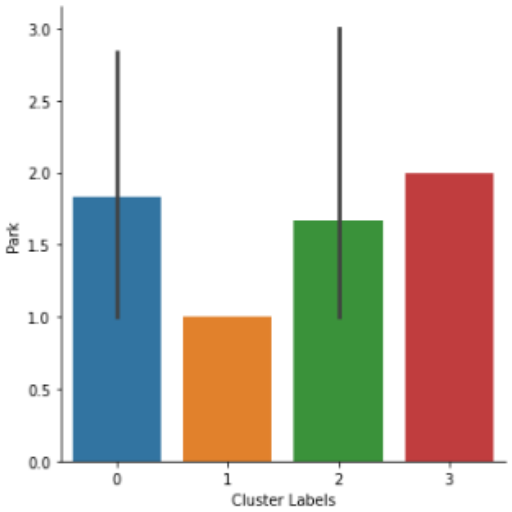
Out[54]:

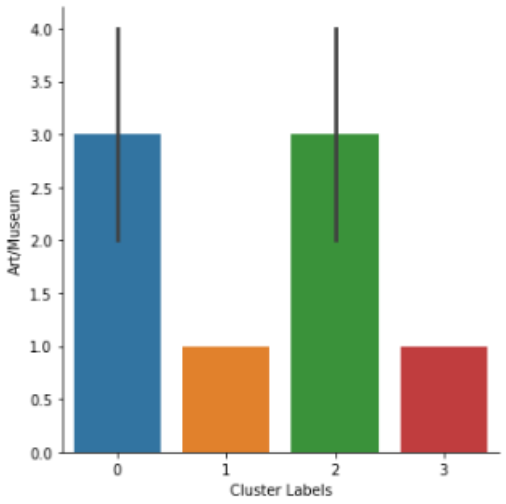
	Station	Local authority	Managed by	PAX mn	Latitude	Longitude	Zone	Postcode	Cluster Labels	Art Gallery	Art Museum	History Museum	Japanese Restaurant	Opera House	Park	Theater	Museum	Art/Museum	Performance	Total	Score
47	Warren Street	Camden	London Underground	20.110864	51.52458	-0.138278	1	NW1 3AA	3	1	0	0	1	0	2	1	0	1	1	5	0.015432

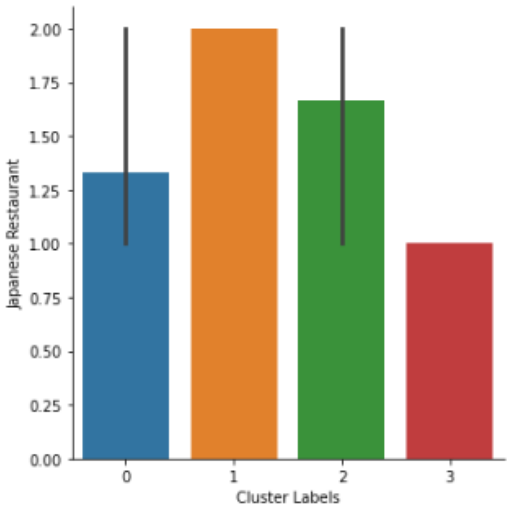
```
In [55]: stations_select.groupby('Cluster Labels').mean()
```

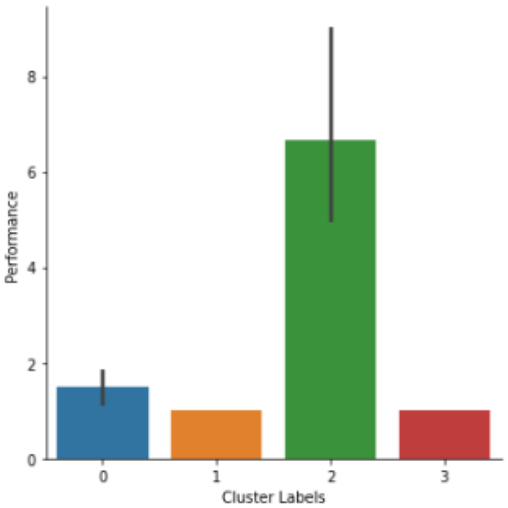
Out[55]:

Cluster Labels	PAX mn	Latitude	Longitude	Art Gallery	Art Museum	History Museum	Japanese Restaurant	Opera House	Park	Theater	Museum	Art/Museum	Performance	Total	Score
0	15.699548	51.515446	-0.146139	1.500000	0.333333	1.166667	1.333333	0.166667	1.833333	1.333333	1.500000	3.0	1.500000	7.666667	0.023663
1	7.148956	51.519998	-0.167668	1.000000	0.000000	0.000000	2.000000	0.000000	1.000000	1.000000	0.000000	1.0	1.000000	5.000000	0.015432
2	27.606379	51.514829	-0.121559	1.333333	0.000000	1.666667	1.666667	0.666667	1.666667	6.000000	1.666667	3.0	6.666667	13.000000	0.040123
3	20.110864	51.524580	-0.138278	1.000000	0.000000	0.000000	1.000000	0.000000	2.000000	1.000000	0.000000	1.0	1.000000	5.000000	0.015432

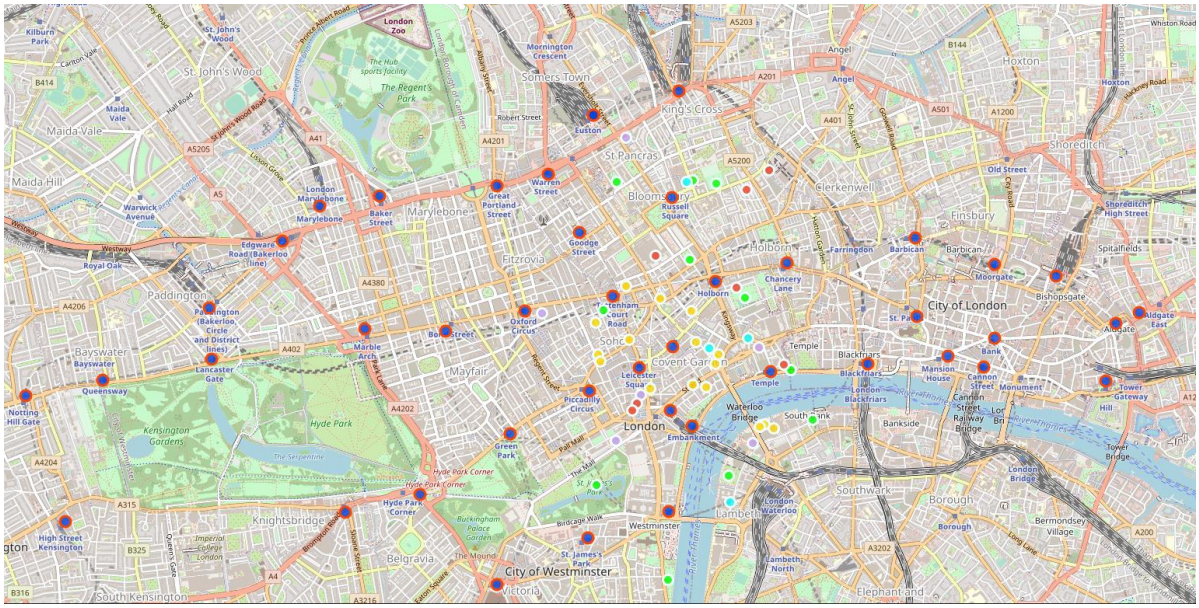








6. Venues around the top 5 stations are listed out and shown on map.



Tube station: blue Park: light green Museum: red

Art gallery: lavender Japanese restaurant: light blue Theater: yellow

	Station	Station Latitude	Station Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Distance	Venue Category
0	Holborn	51.517236	-0.119841	Sir John Soane's Museum	4ac518d3f964a5204fa720e3	51.516833	-0.117540	165	History Museum
1	Holborn	51.517236	-0.119841	Lincoln's Inn Fields	4ad862b4f964a520291121e3	51.516114	-0.116558	259	Park
2	Holborn	51.517236	-0.119841	Gillian Lynne Theatre	4ac518e9f964a520ccab20e3	51.515231	-0.122492	289	Theater
3	Holborn	51.517236	-0.119841	Royal Opera House	4ac518e8f964a5207fab20e3	51.513083	-0.121968	485	Opera House
4	Holborn	51.517236	-0.119841	British Museum	4ac518d2f964a5203da720e3	51.519009	-0.126437	497	History Museum
5	Holborn	51.517236	-0.119841	Shaftesbury Theatre	4ac518e7f964a5205cab20e3	51.516065	-0.125886	438	Theater
6	Holborn	51.517236	-0.119841	Roka	545505a2498e550f5aefc0af	51.513312	-0.116194	504	Japanese Restaurant
7	Holborn	51.517236	-0.119841	Lyceum Theatre	4bad083df964a52078263be3	51.511598	-0.119785	627	Theater
8	Holborn	51.517236	-0.119841	180 The Strand	52fe72e0498e6484322373d8	51.512671	-0.115009	608	Art Gallery
9	Holborn	51.517236	-0.119841	eat TOKYO	56e9c481498e37117cecacf4	51.512637	-0.120488	513	Japanese Restaurant
10	Holborn	51.517236	-0.119841	Dominion Theatre	4ac518e7f964a52044ab20e3	51.516932	-0.129649	680	Theater
11	Holborn	51.517236	-0.119841	Duchess Theatre	4ac518e7f964a52047ab20e3	51.512239	-0.119460	556	Theater
12	Russell Square	51.523013	-0.124590	British Museum	4ac518d2f964a5203da720e3	51.519009	-0.126437	463	History Museum
13	Russell Square	51.523013	-0.124590	Hare & Tortoise	4b411423f964a520aac025e3	51.524086	-0.123015	161	Japanese Restaurant
14	Russell Square	51.523013	-0.124590	Gordon Square	4ada5af6f964a520c22121e3	51.524098	-0.130763	444	Park
15	Russell Square	51.523013	-0.124590	Charles Dickens Museum	4be69ded2468c92835880143	51.523523	-0.116409	569	History Museum
16	Russell Square	51.523013	-0.124590	Brunswick Square Gardens	4c28812a97d00f479ee13fea	51.524177	-0.122322	203	Park
17	Russell Square	51.523013	-0.124590	Bloomsbury Square	4ac518d3f964a52082a720e3	51.518740	-0.122622	494	Park
18	Russell Square	51.523013	-0.124590	Dominion Theatre	4ac518e7f964a52044ab20e3	51.516932	-0.129649	762	Theater
19	Russell Square	51.523013	-0.124590	eat TOKYO	505a370de4b034c534cd8d67	51.519849	-0.117410	609	Japanese Restaurant
20	Russell Square	51.523013	-0.124590	Coram's Fields	4bca013ecc8cd13afcabbccf	51.523979	-0.119736	352	Park
21	Russell Square	51.523013	-0.124590	Crypt Gallery	4c9269f87088a1433141b943	51.527110	-0.129690	576	Art Gallery
22	Russell Square	51.523013	-0.124590	Sir John Soane's Museum	4ac518d3f964a5204fa720e3	51.516833	-0.117540	843	History Museum
23	Russell Square	51.523013	-0.124590	The Postal Museum	4dcc1a2afa76d745f064241e	51.524853	-0.113885	769	History Museum
24	Russell Square	51.523013	-0.124590	Shaftesbury Theatre	4ac518e7f964a5205cab20e3	51.516065	-0.125886	778	Theater
25	Russell Square	51.523013	-0.124590	Roka	4b2fd935f964a52023f124e3	51.518992	-0.135308	866	Japanese Restaurant

26	Temple	51.511041	-0.113726	Two Temple Place	4ee1d846a17c79c0e9344b09	51.511523	-0.112236	116	History Museum
27	Temple	51.511041	-0.113726	180 The Strand	52fe72e0498e6484322373d8	51.512671	-0.115009	202	Art Gallery
28	Temple	51.511041	-0.113726	Temple Gardens	4ac518d1f964a520d1a620e3	51.511154	-0.111472	156	Park
29	Temple	51.511041	-0.113726	National Theatre	4ac7a2caf964a520adb820e3	51.507376	-0.114793	414	Theater
30	Temple	51.511041	-0.113726	Lyceum Theatre	4bad083df964a52078263be3	51.511598	-0.119785	424	Theater
31	Temple	51.511041	-0.113726	Roka	545505a2498e550f5aefc0af	51.513312	-0.116194	305	Japanese Restaurant
32	Temple	51.511041	-0.113726	Dorfman Theatre	542eec4e498ee31baa076ef7	51.507167	-0.113350	432	Theater
33	Temple	51.511041	-0.113726	Olivier Theatre	5022b0c1e4b0464790dcff76	51.507414	-0.114397	406	Theater
34	Temple	51.511041	-0.113726	Lyttelton Theatre	4b7fe34ef964a5209c4230e3	51.507249	-0.114926	430	Theater
35	Temple	51.511041	-0.113726	Savoy Theatre	4ac518e7f964a5205bab20e3	51.509944	-0.120864	509	Theater
36	Temple	51.511041	-0.113726	Hayward Gallery	4ac518d2f964a5203ba720e3	51.506136	-0.115709	563	Art Gallery
37	Temple	51.511041	-0.113726	Royal Opera House	4ac518e8f964a5207fab20e3	51.513083	-0.121968	614	Opera House
38	Temple	51.511041	-0.113726	Duchess Theatre	4ac518e7f964a52047ab20e3	51.512239	-0.119460	419	Theater
39	Temple	51.511041	-0.113726	Lincoln's Inn Fields	4ad862b4f964a520291121e3	51.516114	-0.116558	597	Park
40	Temple	51.511041	-0.113726	eat TOKYO	56e9c481498e37117cecacf4	51.512637	-0.120488	501	Japanese Restaurant
41	Temple	51.511041	-0.113726	Bernie Spain Gardens	4c21fbbc11de20a13c1084ce	51.507717	-0.109118	488	Park
42	Temple	51.511041	-0.113726	Sir John Soane's Museum	4ac518d3f964a5204fa720e3	51.516833	-0.117540	696	History Museum
43	Temple	51.511041	-0.113726	Vaudeville Theatre	4ac518e7f964a5205fab20e3	51.510129	-0.122269	600	Theater
44	Tottenham Court Road	51.516211	-0.131110	Dominion Theatre	4ac518e7f964a52044ab20e3	51.516932	-0.129649	129	Theater
45	Tottenham Court Road	51.516211	-0.131110	Soho Square	4ac518cff964a5206ba620e3	51.515286	-0.132209	128	Park
46	Tottenham Court Road	51.516211	-0.131110	Soho Theatre	4ac518e8f964a5209fab20e3	51.514384	-0.133025	242	Theater
47	Tottenham Court Road	51.516211	-0.131110	British Museum	4ac518d2f964a5203da720e3	51.519009	-0.126437	449	History Museum
48	Tottenham Court Road	51.516211	-0.131110	Palace Theatre	4ac518e7f964a52053ab20e3	51.513211	-0.129365	355	Theater
49	Tottenham Court Road	51.516211	-0.131110	Roka	4b2fd935f964a52023f124e3	51.518992	-0.135308	424	Japanese Restaurant
50	Tottenham Court Road	51.516211	-0.131110	Les Miserables Show	527be8b8498e02ffd3fb693d	51.512190	-0.132807	462	Theater
51	Tottenham Court Road	51.516211	-0.131110	Queen's Theatre	4ac518e7f964a52059ab20e3	51.511745	-0.132562	507	Theater
52	Tottenham Court Road	51.516211	-0.131110	The Photographers' Gallery	4ac518cef964a52009a620e3	51.515070	-0.138969	559	Art Gallery
53	Westminster	51.501402	-0.125002	Victoria Tower Gardens	4b9f9704f964a5205d2d37e3	51.496729	-0.125016	520	Park
54	Westminster	51.501402	-0.125002	St James's Park	4ac518cdf964a520f2a520e3	51.503253	-0.132995	590	Park
55	Westminster	51.501402	-0.125002	Okan	5a99bc4f95a7222a9d15d51e	51.502040	-0.118192	477	Japanese Restaurant
56	Westminster	51.501402	-0.125002	Jubilee Gardens	4bdec9150d69b713580266d3	51.503901	-0.118339	539	Park
57	Westminster	51.501402	-0.125002	Institute of Contemporary Arts (ICA)	4ac518b9f964a520b6a120e3	51.506265	-0.130817	674	Art Gallery
58	Westminster	51.501402	-0.125002	National Gallery	4ac518cdf964a520e6a520e3	51.508876	-0.128478	866	Art Museum
59	Westminster	51.501402	-0.125002	Hayward Gallery	4ac518d2f964a5203ba720e3	51.506136	-0.115709	832	Art Gallery
60	Westminster	51.501402	-0.125002	Sainsbury Wing National Gallery	5320a0a6498e8dd49f5aa981	51.508384	-0.129001	825	Art Museum
61	Westminster	51.501402	-0.125002	National Portrait Gallery	4ac518d2f964a52047a720e3	51.509438	-0.128032	918	Art Gallery
62	Westminster	51.501402	-0.125002	National Theatre	4ac7a2caf964a520adb820e3	51.507376	-0.114793	970	Theater
63	Westminster	51.501402	-0.125002	The London Coliseum	567ff1aa498e8f5ed88b0c9d	51.509846	-0.127005	950	Theater

Results

- Cluster 0 has the most stations (6 stations), while Cluster 2 has the most number of venues around each station on average (13 venues).

- In terms of each venue category, Cluster 1, Cluster 3, Cluster 0/2, Cluster 2 outperform in the Japanese Restaurant, Park, Art/Museum, Theater category respectively.

- Cluster 1 and Cluster 3 are not bad in terms of food and green spaces, but they lag in Art/Museum and Theater.

- Cluster 0 is similar to Cluster 2 in regards to Food, Park and Art/Museum, but weaker in Theater.

- As a whole, Cluster 2 appears to be the better choice, with balanced ranking in all categories. In addition, as seen from the map, the stations in Cluster 2 (light blue circles) are close to each other, and thus flexible to pass from one to the other.

- The top 5 stations with the highest score (representing the total number of venue of all categories) are Temple (Cluster 2), Russell Square (Cluster 0), Holborn (Cluster 2), Tottenham Court Road (Cluster 0) and Westminster (Cluster 2).

Conclusion

This project leverages techniques and tools learnt in the Data Science courses, including Python data analysis, data visualization in graphs and maps, Foursquare API and machine learning technique specifically K-means clustering, to facilitate individuals' trip planning. Specifically, the problem is to pick out Zone 1 tube stations in London that have parks, museums/art galleries, Japanese restaurants and theaters around. The analysis is easily adaptable to any customized choice of city and venue categories of interest.

For future enhancement, attributes related to price and popularity (e.g. count of likes and ratings) of venues from Foursquare could be added in the analysis for more comprehensive assessment.