The Battle of Neighbourhoods

Coursera Capstone Project

Introduction

- This project leverages the techniques and tools learnt in the Data Science courses to analyse a practical problem in regards to picking location
  - Python data analysis
  - Data visualization in graphs and maps
  - Foursquare API
  - Machine learning technique, specifically K-means clustering

Business problem and Audience

- I am a tourist in London, planning my one-day activity: a park in the morning, a museum or art gallery to follow, a Japanese restaurant for meal and a musical/live performance in a theatre to end
- Which tube stations in Zone 1 should I go with all these four types of venues nearby?
- The analysis is aimed to be adaptable for trip planning by any tourist with customised choice of city and personalised interests.

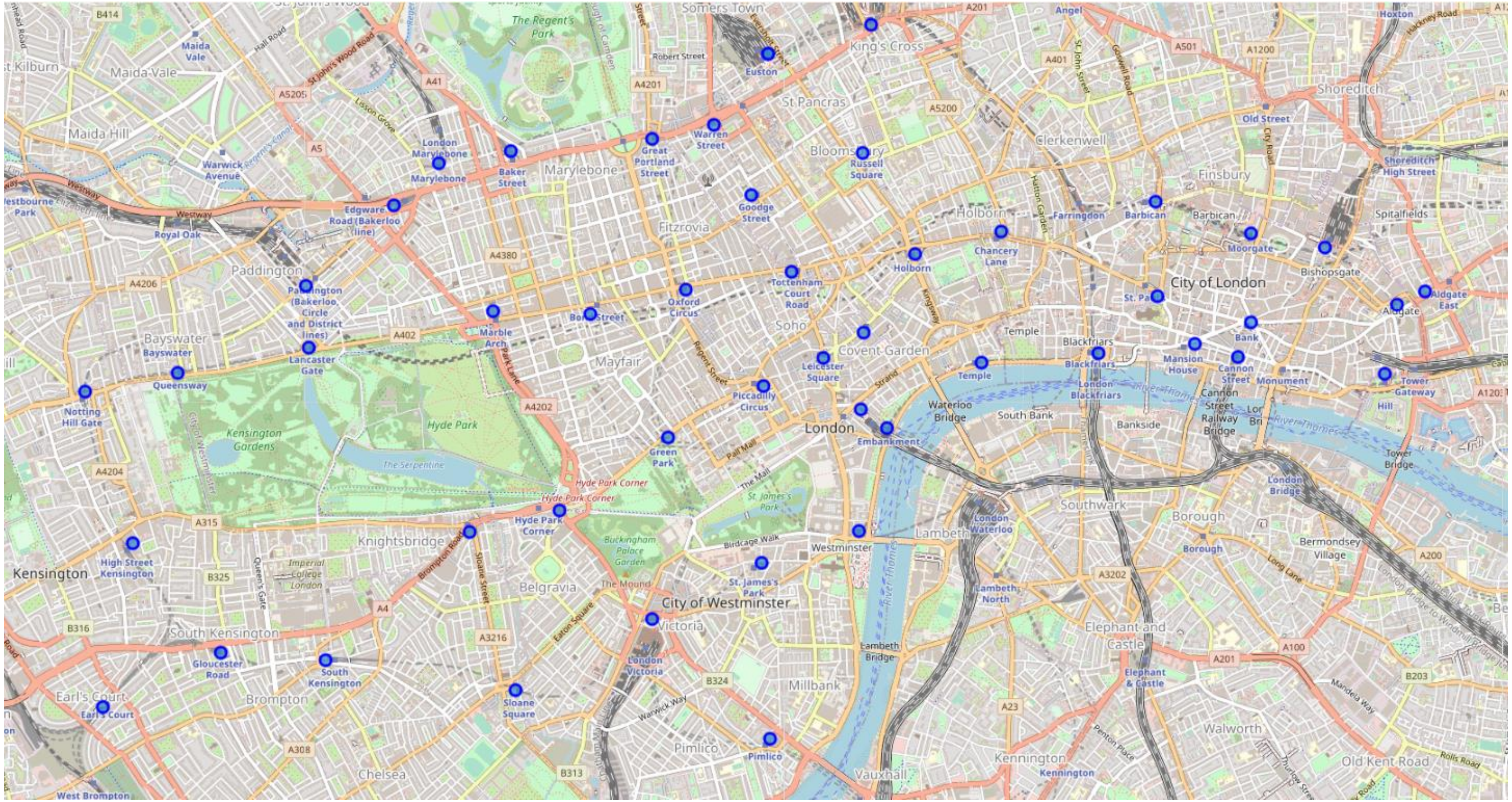
Data

- List of London tube stations in Zone 1 (https://en.wikipedia.org/wiki/List_of_stations_in_London_fare_zone_1)
- Borough-level statistics of tourism trips by foreign and domestic tourists (https://data.gov.uk/dataset/ee5038be-d2be-4ab6-a612-70ade60eca12/tourism-trips-borough)
- Annual entries and exits by passengers of a station (https://data.london.gov.uk/download/london-underground-performance-reports/b6ab04fc-9062-4291-b514-7fa218073b4c/multi-year-station-entry-and-exit-figures.xls)
- Geo codes of tube stations (https://www.doogal.co.uk/LondonStationsCSV.ashx)

Data Preparation

- Selection criteria out of the Zone-1 tube stations (84 in total):
  - in the top 5 boroughs with the most foreign and domestic tourism trips: Westminster, Kensington & Chelsea, Camden, City of London and Tower Hamlets
  - managed by London Underground
  - annual entries and exits by passengers >= 5 million
- Technique used in data processing:
  - Web scrapping by BeautifulSoup
  - Pandas data frame methods (read excel, drop, merge, replace etc.)
- 49 candidate tube stations in Zone-1 selected, together with their geo codes.

Candidate stations on map



Methodology for Data Analysis

- Foursquare: to obtain nearby venues around each select station and get venue categories
  - Categories selected: Park, Art Gallery, Art Museum, History Museum, Japanese Restaurant, Opera House and Theater
- Pandas data analysis
- Onehot dummy variables
- K-means clustering
  - Determine the best number of clusters by Elbow Method (the best K = 5)

Results (i) – 324 venues, 11 stations and 4 clusters

- There are 324 venues of selected categories in total surrounding the 49 candidate tube stations

Venue Category	
Art Gallery	60
Art Museum	28
History Museum	28
Japanese Restaurant	52
Opera House	7
Park	66
Theater	83

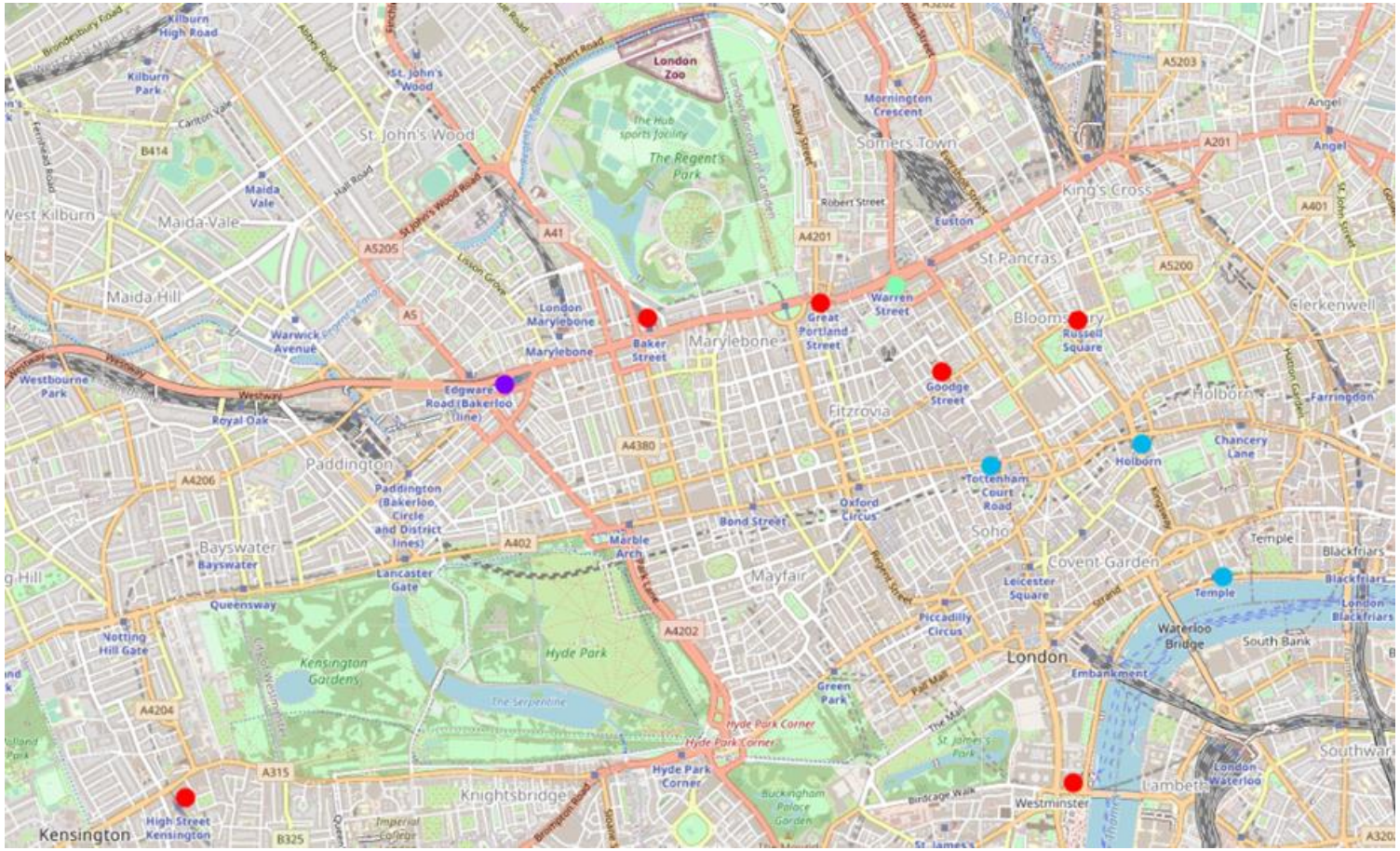
- Only stations with at least one venue for each category (i.e. park, art/museum, food and performance) nearby are kept, and stations that lack venues of any one category are removed

- 11 stations are shortlisted, are grouped into 4 clusters

	Station	Local authority	Managed by	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	1	WC2R 2PH	2	2	0	2	2	1	3	8	2	4	9	18	0.055556
1	Russell Square	Camden	London Underground	1	WC1N 1LG	0	1	0	4	3	0	4	2	4	5	2	14	0.043210
2	Holborn	Camden	London Underground	1	WC2B 6AA	2	1	0	2	2	1	1	5	2	3	6	12	0.037037
3	Westminster	Westminster	London Underground	1	SW1A 2JR	0	3	2	0	1	0	3	2	2	5	2	11	0.033951
4	Tottenham Court Road	Westminster	London Underground	1	W1D 2DA	2	1	0	1	1	0	1	5	1	2	5	9	0.027778
5	Goodge Street	Camden	London Underground	1	W1T 2HF	0	1	0	1	1	0	1	2	1	2	2	6	0.018519
6	Baker Street	Westminster	London Underground	1	NW1 5LA	0	1	0	1	1	0	1	1	1	2	1	5	0.015432
7	Edgware Road (Cir)	Westminster	London Underground	1	NW1 5DH	1	1	0	0	2	0	1	1	0	1	1	5	0.015432
8	Great Portland Street	Westminster	London Underground	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	1	W8 5SA	0	1	0	1	1	1	1	0	1	2	1	5	0.015432
10	Warren Street	Camden	London Underground	1	NW1 3AA	3	1	0	0	1	0	2	1	0	1	1	5	0.015432

Not any station in cluster 4 is in the list, since not venues of all categories are available around.

Results (ii) – visualize clusters on map



Red: Cluster 0 (6 stations)

- Baker Street
- Goodge Street
- Great Portland Street
- High Street Kensington
- Russel Square
- Westminster

Purple: Cluster 1 (1 station)

- Edgware Road

Blue: Cluster 2 (3 stations)

- Holborn
- Temple
- Tottenham Court Road

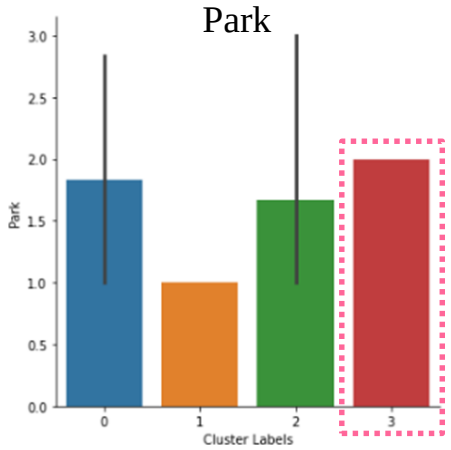
Green: Cluster 3 (1 station)

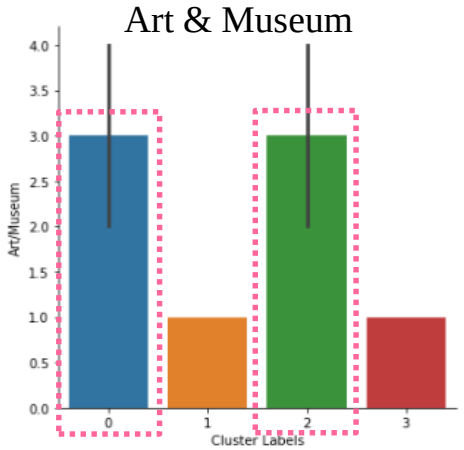
- Warren Street

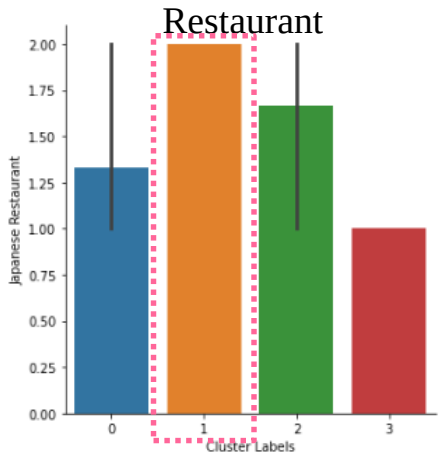
Results (iii) – cluster comparison

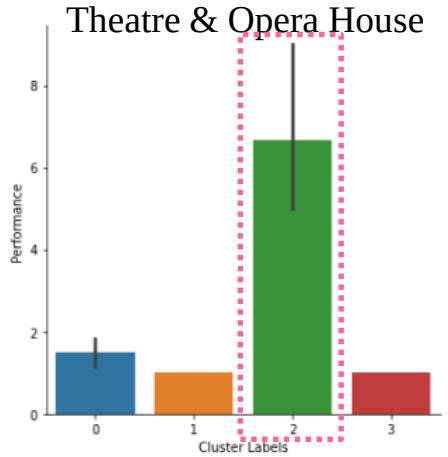
Mean description of clusters and average number of venues by category

Cluster Labels	Art Gallery	Art Museum	History Museum	Japanese Restaurant	Opera House	Park	Theater	Museum	Art/Museum	Performance	Total	Score
0	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	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	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	1.000000	0.000000	0.000000	1.000000	0.000000	2.000000	1.000000	0.000000	1.0	1.000000	5.000000	0.015432









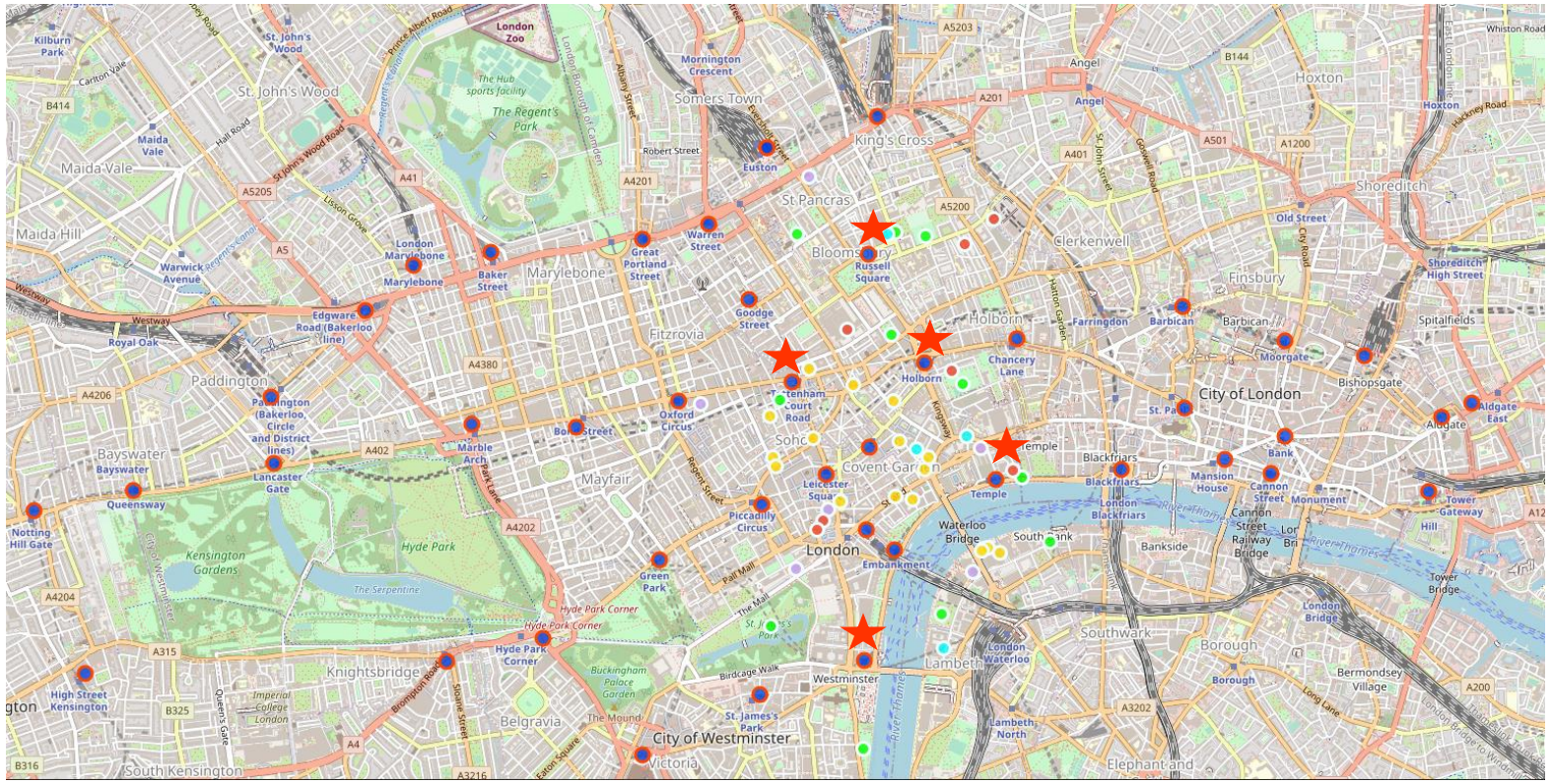
Observations

- While Cluster 0 has the most stations (6), Cluster 2 has the most number of venues (13) around each station on average.
- Cluster 3, 0/2, 1 and 2 are the best in Park, Art & Museum, Japanese Restaurant and Theater category respectively.
- Clusters 1 and 3 are not bad in terms of food and green spaces, but lag in Art & Museum and Theater.
- Cluster 0 is similar to Cluster 2 in Park, Art/Museum and Restaurant categories, but weaker in the Theater category.
- As a whole, **Cluster 2** appears to be the better choice, with balanced ranking in all categories. In addition, as seen from the map, stations in Cluster 2 (light blue circles) are close to each other, and thus flexible to pass from one to the other.

Results (iv) – top 5 stations and venues nearby on map

- The top 5 stations are Temple, Russell Square, Holborn, Tottenham Court Road and Westminster, ranked by the total number of venues nearby.

	Station	Local authority	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	1	WC2R 2PH	2	2	0	2	2	1	3	8	2	4	9	18	0.055556
1	Russell Square	Camden	1	WC1N 1LG	0	1	0	4	3	0	4	2	4	5	2	14	0.043210
2	Holborn	Camden	1	WC2B 6AA	2	1	0	2	2	1	1	5	2	3	6	12	0.037037
3	Westminster	Westminster	1	SW1A 2JR	0	3	2	0	1	0	3	2	2	5	2	11	0.033951
4	Tottenham Court Road	Westminster	1	W1D 2DA	2	1	0	1	1	0	1	5	1	2	5	9	0.027778



Park: light green

Museum: red

Art gallery: lavender

Japanese restaurant: light blue

Theatre and Opera House: yellow

Conclusions

- This project leverages key techniques and tools learnt in the series of Data Science courses (e.g. Foursquare API and K-Means clustering) to facilitate individuals' trip planning.
- Specifically, the problem to solve is to pick out Zone-1 tube stations in London that has parks, museums/art galleries, Japanese restaurants and theaters around.
- The analysis is expected to be easily adaptable to any customised 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.