IBM DATA SCIENCE – CAPSTONE

Battle Of The Neighborhoods

Choosing The Best Vacation Destination Between Dubai, Paris, New York and Kuala Lumpur

Muhammad Umar Salman

Introduction and Problem Description

In this Capstone Project, I will use my learnings and understandings from IBM's Professional Data Science Certificate to solve this never-ending problem of which place is it better to go for a vacation between these four famous cities. Many foreign tourists are trying to find a vacation spot which gives the whole package and is worth them taking their time out of their lives to come and enjoy. Tourists want to choose a place that has a variety of activities to do and a variety of places to go. In this project I will try to explore the two cities to see which city offers the best vacation destination in terms of hotels, restaurants, malls, museums, sight-seeing venues and fun outdoor activities.

This data will be able to solve this problem as we can see what the reviews, ratings and tips for places in the city are. How many 5 – Star hotels does the city have. How many different categories of places are available and how far are their distances from the places where the tourists plan to live. Furthermore, we can see what Tourists look for in ideal vacation venues and cross check them with cities to see whether those cities provide those places or not.

Data Section

1. Requirements

   The data required in this is all types of categories for vocational purposes such as Hotels, Restaurants, Shopping Malls, Museums, Sight-Seeing Venues and Fun-Outdoor Activities. For that purpose, I will use the Foursquare API to 'Search' and 'Explore' all these Vacation venues to observe and do exploratory data analysis on the following cities.

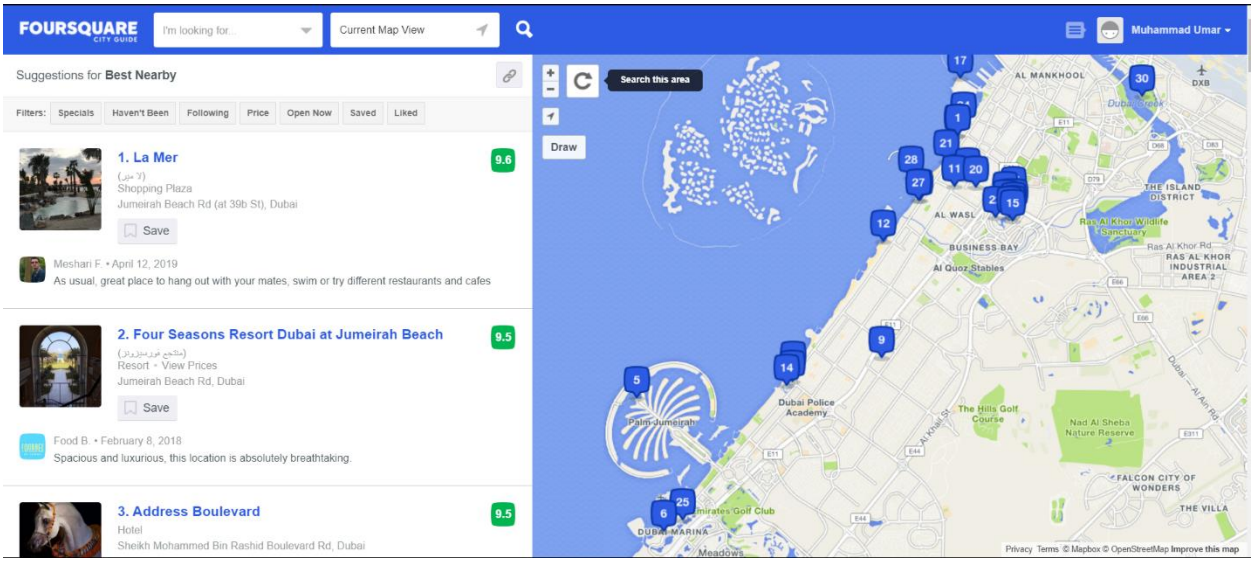
2. Source

   The data for this project will be collected using Foursquare API using the following link: https://foursquare.com/developers. In order to extract our data, I will use my Client ID and Client Secret to create a query which will fetch the data I need from the website.

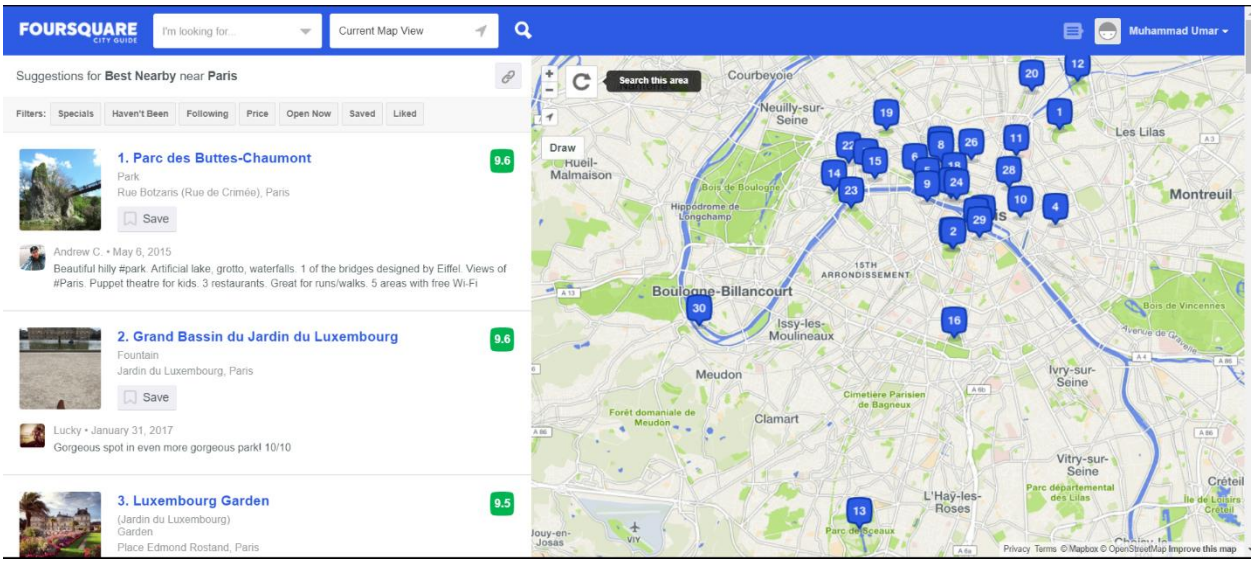
```
CLIENT_ID = '******' # your Foursquare ID
CLIENT_SECRET = '******' # your Foursquare Secret
VERSION = '******' # Foursquare API version
```

```
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
        CLIENT_ID,
        CLIENT_SECRET,
        VERSION,
        lat,
        long,
        RADIUS,
        LIMIT)
```

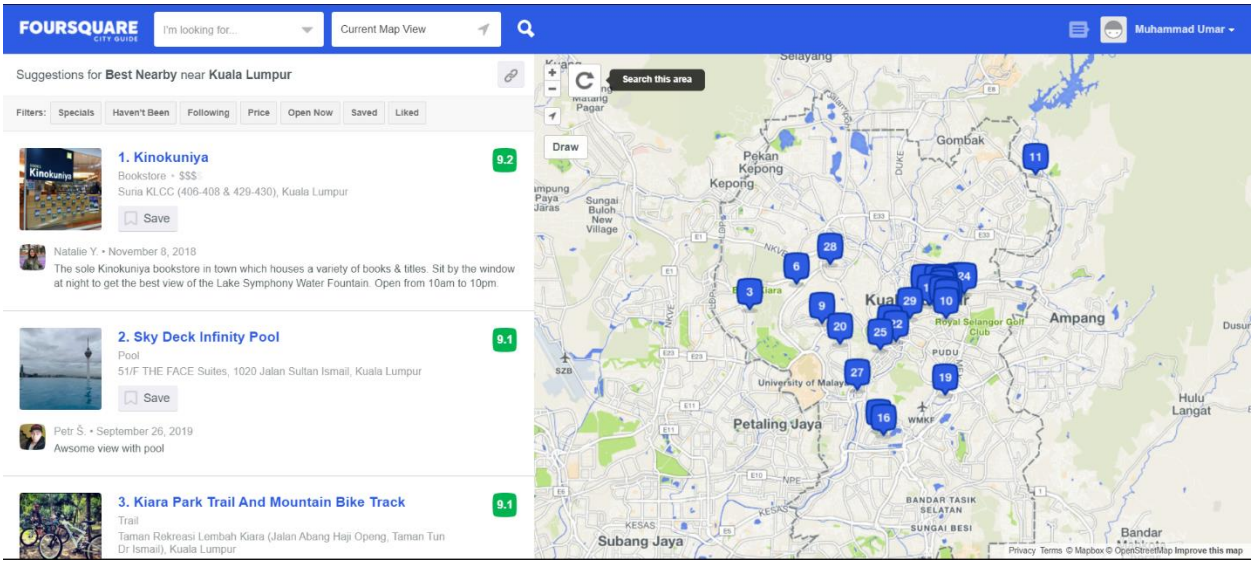
Dubai


FOURSQUARE
CITY GUIDE
I'm looking for...
Current Map View
Muhammad Umar
Suggestions for Best Nearby
Filters: Specials Haven't Been Following Price Open Now Saved Liked
1. La Mer
9.6
Shopping Plaza
Jumeirah Beach Rd (at 39b St), Dubai
Save
Meshari F. • April 12, 2019
As usual, great place to hang out with your mates, swim or try different restaurants and cafes
2. Four Seasons Resort Dubai at Jumeirah Beach
9.5
Resort • View Prices
Jumeirah Beach Rd, Dubai
Save
Food B. • February 8, 2018
Spacious and luxurious, this location is absolutely breathtaking.
3. Address Boulevard
9.5
Hotel
Sheikh Mohammed Bin Rashid Boulevard Rd, Dubai
Search this area
Draw
AL MANKHOOL
DXB
THE ISLAND DISTRICT
AL WASL
BUSINESS BAY
Al Quoz Stables
Ras Al Khor Wildlife Sanctuary
Ras Al Khor Rd
RAS AL KHOR INDUSTRIAL AREA 2
Dubai Police Academy
The Hills Golf Course
Nad Al Sheba Nature Reserve
FALCON CITY OF WONDERS
THE VILLA
Palm Jumeirah
Emirates Golf Club
DUBAI MARINA
Meadows
Privacy Terms © Mapbox © OpenStreetMap Improve this map

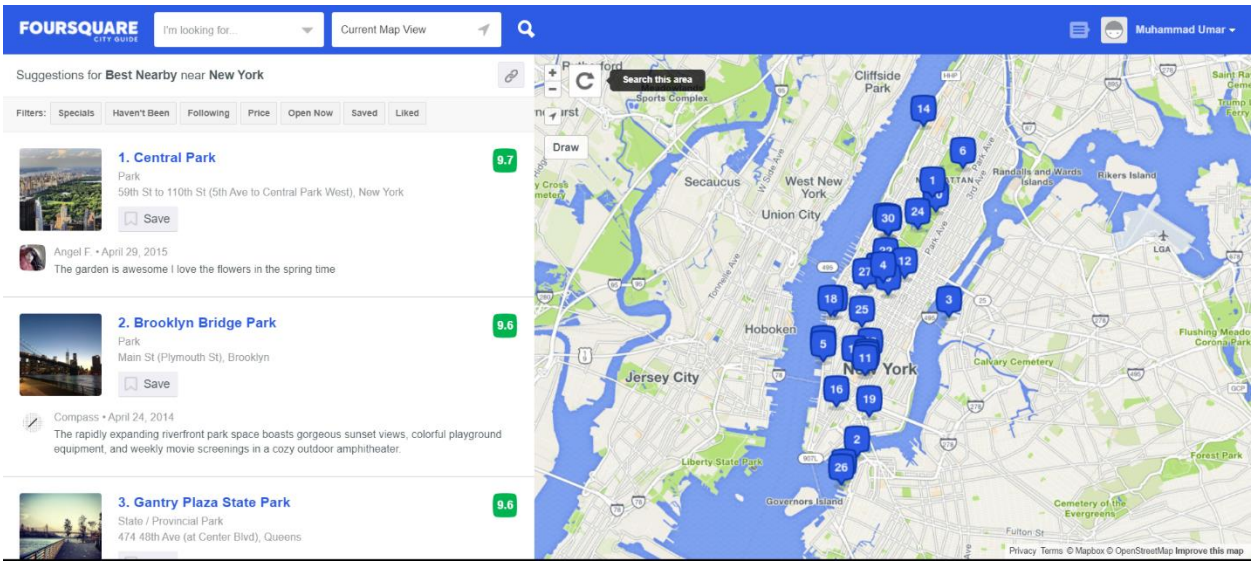
Paris


FOURSQUARE
CITY GUIDE
I'm looking for...
Current Map View
Muhammad Umar
Suggestions for Best Nearby near Paris
Filters: Specials Haven't Been Following Price Open Now Saved Liked
1. Parc des Buttes-Chaumont
9.6
Park
Rue Botzaris (Rue de Crimée), Paris
Save
Andrew C. • May 6, 2015
Beautiful hilly #park. Artificial lake, grotto, waterfalls. 1 of the bridges designed by Eiffel. Views of #Paris. Puppet theatre for kids. 3 restaurants. Great for runs/walks. 5 areas with free Wi-Fi
2. Grand Bassin du Jardin du Luxembourg
9.6
Fountain
Jardin du Luxembourg, Paris
Save
Lucky • January 31, 2017
Gorgeous spot in even more gorgeous park! 10/10
3. Luxembourg Garden
9.5
(Jardin du Luxembourg)
Garden
Place Edmond Rostand, Paris
Search this area
Draw
Courbevoie
Neuilly-sur-Seine
Rueil-Malmaison
Bois de Boulogne
Hippodrome de Longchamp
Les Lilas
Montreuil
15TH ARRONDISSEMENT
Boulogne-Billancourt
Issy-les-Moulineaux
Meudon
Forêt domaniale de Meudon
Clamart
Cimetière Parisien de Bagneux
Bois de Vincennes
Ivry-sur-Seine
Vitry-sur-Seine
Parc départemental des Lilas
Créteil
L'Haÿ-les-Roses
Parc de Sceaux
Privacy Terms © Mapbox © OpenStreetMap Improve this map

Kuala Lumpur



New York



3. Audience

The audience for this report will be the foreign tourists and travelers looking for their perfect vacation destination. It will not only benefit these travelers but will also give insight to the tourist industry as to what things make a place a better spot and more likely place for people to go for their holidays. **Tourism** industry is important for the benefits it brings and due to its role as a commercial activity that creates demand and growth for many more industries.

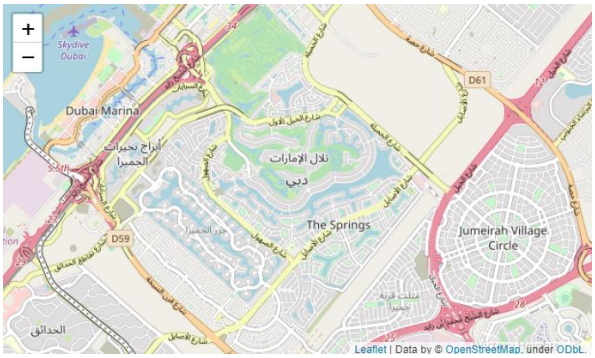
Methodology Section

1. Feature Engineering

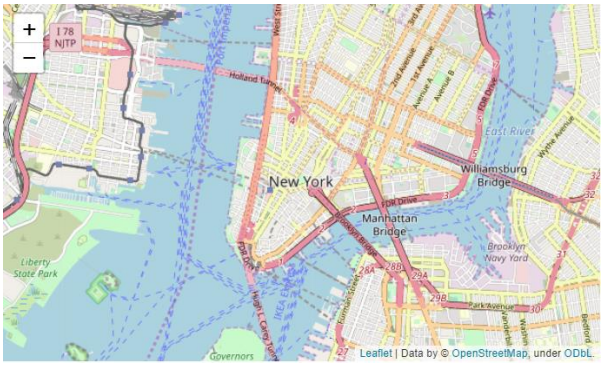
   First of all, we requested venues from the Foursquare API as shown above. For each of the 4 cities we looked at all the with a RADIUS of 10000. From the Geolocator we found out the location of all the cities and mapped them out as shown below.

Paris

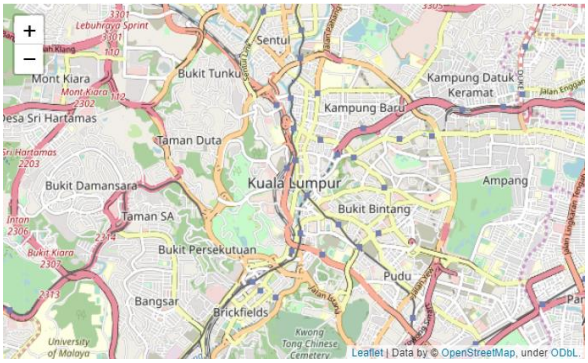
Dubai



New York



Kuala Lumpur



We then converted the raw JSON to a DataFrame.

	id	name	categories	referralId	hasPerk	location.address	location.lat	location.lng	location.labeledLatLng
0	4c73295213228cfa74312d65	Dataran DBKL	[{'id': '4bf58dd8d48988d163941735', 'name': 'P...	v-1585689957	False	Menara DBKL	3.151427	101.694462	[{'label': 'display', 'la 3.151427172961877
1	4c75ac31ff1fb60ca9faf6a7	Menara DBKL 1	[{'id': '4bf58dd8d48988d129941735', 'name': 'C...	v-1585689957	False	Jalan Raja Laut	3.152314	101.694662	[{'label': 'display', 'la 3.152314298073572
2	4ec4f9b9e300e6894f206ae2	Dewan Bandaraya Kuala Lumpur Menara 2	[{'id': '4bf58dd8d48988d129941735', 'name': 'C...	v-1585689957	False	Bangunan DBKL Menara 2, Tingkat 6	3.151609	101.694551	[{'label': 'display', 'la 3.15160866153506,
3	4d65cd0a56746dcbae433fff	Cafeteria DBKL	[{'id': '4bf58dd8d48988d142941735', 'name': 'A...	v-1585689957	False	Tkt 2, Menara DBKL 1	3.152154	101.694922	[{'label': 'display', 'la 3.152153558414952

However, a lot of feature engineering had to be done. So firstly, we dropped all the columns which we were not going to need, and which

didn’t give us any meaning. Then we looked into categories and extracted out the venue. We divided all the venues for all the city Data Frames into 6 very generic categories all which would be essential in a vacation trip. The following are:

	Category	Rank
0	Hotels	1
1	Activities	2
2	Food	3
3	Sight Seeing	4
4	Travel	5
5	Shopping	6

We assigned a rank to each of the categories based on the importance it had for tourists which on popular opinion was in this order. After all the Feature Engineering we managed to shape our Data Frame to be exactly as we had hoped.

	id	name	categories	latitude	longitude	distance
0	58f0cd8db1ec13241266420a	New York	Pizza Place	40.712784	-74.005940	8
1	4a676321f964a52051c91fe3	New York City Hall	City Hall	40.712659	-74.005880	13
2	4b475390f964a520f12e26e3	Mary's Coffee Shop	Coffee Shop	40.712786	-74.005944	8
3	4b57b0dff964a520293c28e3	MTA Subway - City Hall (R/W)	Metro Station	40.713394	-74.006934	107
4	4b79a5e8f964a52037082fe3	NY Gift Shop	Gift Shop	40.712733	-74.005978	3
5	3fd66200f964a520d8f11ee3	City Hall Park	Park	40.712415	-74.006724	69
6	51a4bc7c498e469047be66d6	City Hall Council Chambers	City Hall	40.712736	-74.005472	45
7	5ae3d3db3ba767002cf97a85	El Mexicano Restaurante & Cafe	Mexican Restaurant	40.675417	-74.000486	4179
8	4dd77d22b3adc64ae076925d	St comunity beach	Beach	40.712781	-74.005944	8
9	51b3b7da498efda7cefb2660	Dahlia's Mexican Restaurant	Mexican Restaurant	40.718754	-74.010337	763
10	4e31861d7d8b9b256bcc43a8	June's Fruit Cart	Food Truck	40.712786	-74.005944	8

2. Data Analysis

We then created for all cities a single Data Frame which we were going to use as a metric to discover which city gave the best results. From all the categories returned for each city we first took a count of the categories which actually fit for our problem. Then we took counts for all the individual broad categories for every city. Since we ranked the categories in terms of importance, we then calculated the weighted count for each individual category. The constructed data frame is shown below:

	City	Category Count	Hotels	Activities	Food	Sight Seeing	Travel	Shopping	Hotels Weighted	Activities Weighted	Food Weighted	Sight Seeing Weighted	Travel Weighted	Shopping Weighted	Weighted Category Count
0	Paris	26	1	2	11	2	4	6	6	10	44	6	8	6	80
1	Kuala Lumpur	38	2	7	18	6	0	5	12	35	72	18	0	5	142
2	New York	44	4	5	21	4	6	4	24	25	84	12	12	4	161
3	Dubai	33	1	18	10	2	0	2	6	90	40	6	0	2	144

We then marked the clusters for each category on the map. The diagrams are shown in the result section. We added up all the values for each city and got a total score. To normalize the score based on the number of categories that were actually present in that vicinity of the city we divided the Total by the Category count to get the normalized scores for each city. We sorted them by these values which led us to discover which city was the best in terms of the facilities they provide for tourists on vacations.

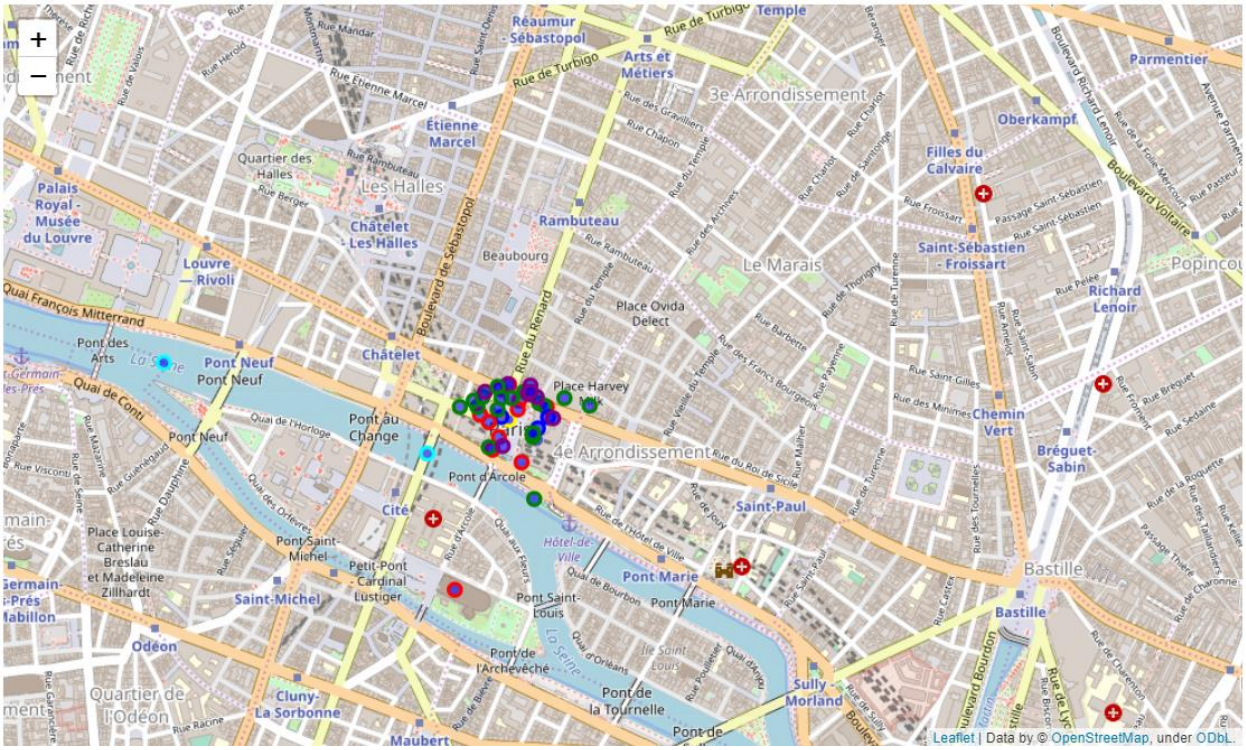
Results

First, we will see the clusters formed from the mapping of the categories in each city as show below:

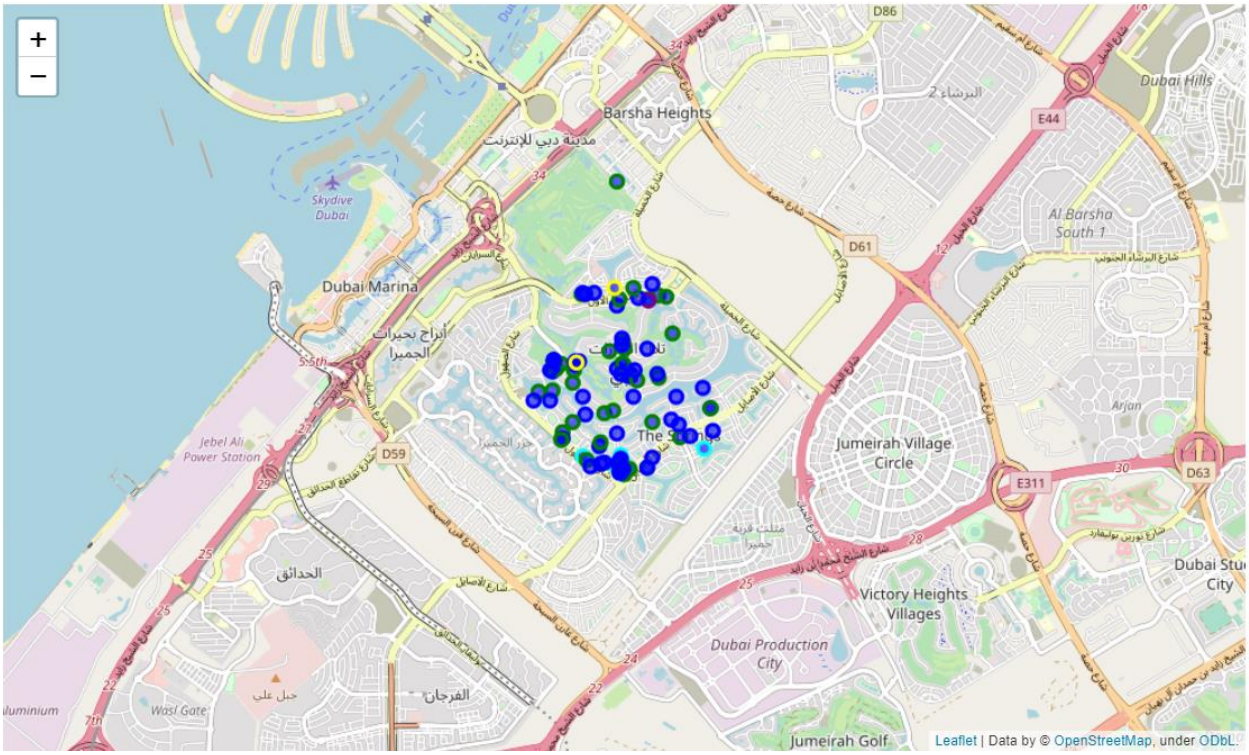
Color of the clusters

- Hotels - Yellow
- Activities - Blue
- Food - Green
- Sight Seeing - Cyan
- Travel - Red
- Shopping - Purple

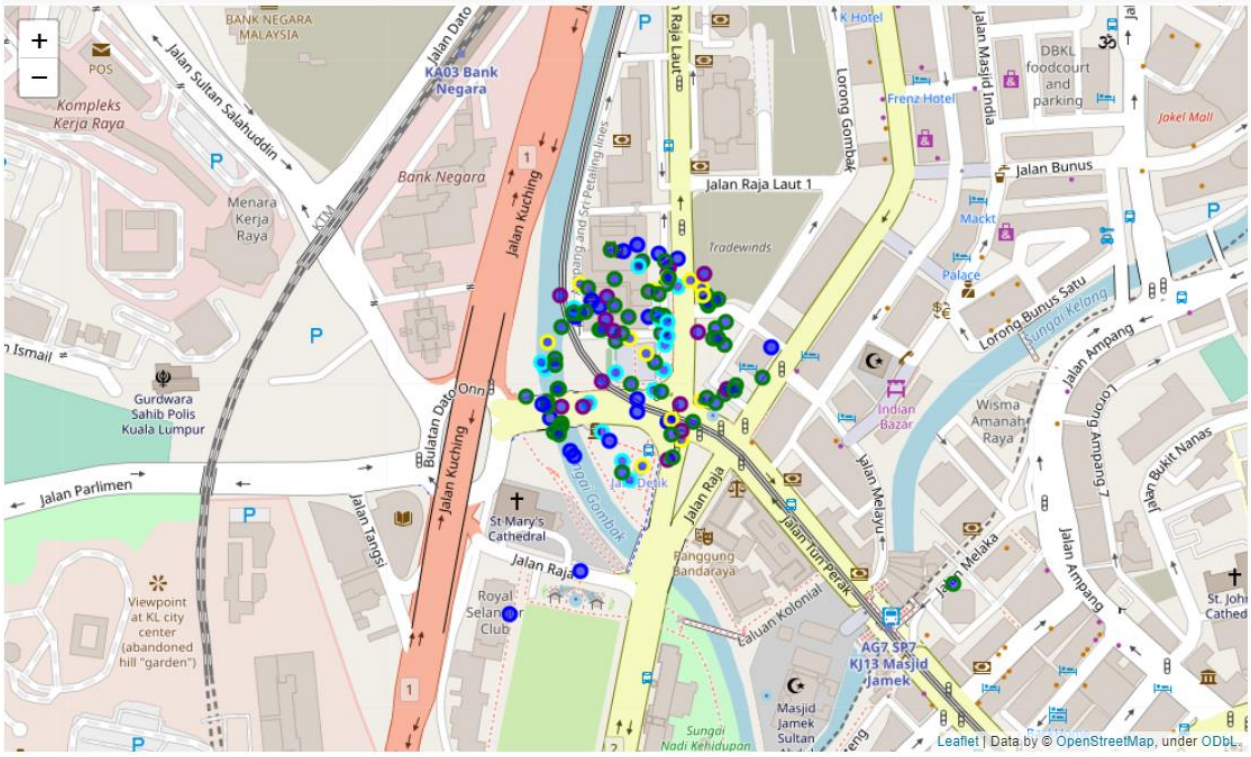
Paris


Réaumur - Sébastopol
Temple
Arts et Métiers
3e Arrondissement
Parmentier
Étienne Marcel
Oberkampf
Filles du Calvaire
Quartier des Halles
Les Halles
Palais Royal - Musée du Louvre
Châtelet - Les Halles
Rambuteau
Beaubourg
Saint-Sébastien - Froissart
Le Marais
Louvre — Rivoli
Richard Lenoir
Quai François Mitterrand
Place Ovida Delect
Pont des Arts
Pont Neuf
Châtelet
Place Harvey
Pont au Change
4e Arrondissement
Chemin Vert
Bréguet-Sabin
Pont d'Arcole
Cité
Saint-Paul
Place Louise-Catherine Breslau et Madeleine Zillhardt
Pont Saint-Michel
Petit-Pont - Cardinal Lustiger
Hôtel-de-Ville
Pont Marie
Bastille
Saint-Michel
Pont Saint-Louis
Odéon
Pont de l'Archevêché
Quartier de l'Odéon
Cluny-La Sorbonne
Pont de la Tournelle
Sully Morland
Maubert
Leaflet | Data by © OpenStreetMap, under ODbL.

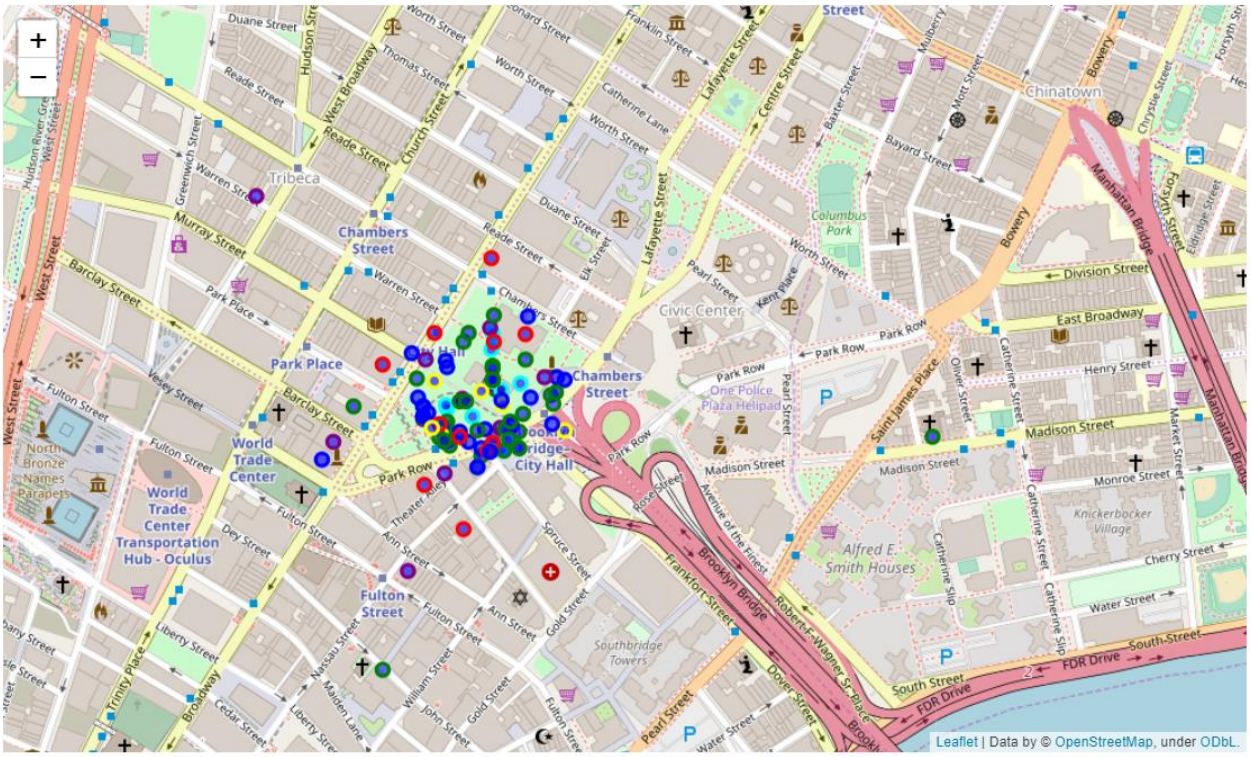
Dubai


D86
Dubai Hills
Barsha Heights
مدينة دبي للإنترنت
E44
Skydive Dubai
Al Barsha South 1
D61
Dubai Marina
أبراج بحيرات الجميرا
Jebel Ali Power Station
D59
The Springs
Jumeirah Village Circle
E311
D63
Arjan
الحدائق
Victory Heights Villages
Dubai Production City
Wasl Gate
الفرجان
Jumeirah Golf
Leaflet | Data by © OpenStreetMap, under ODbL.

Kuala Lumpur



New York



Now looking the total scores of each of the cities

	City	Category Count	Hotels	Activities	Food	Sight Seeing	Travel	Shopping	Hotels Weighted	Activities Weighted	Food Weighted	Sight Seeing Weighted	Travel Weighted	Shopping Weighted	Weighted Category Count	Total
0	Paris	26	1	2	11	2	4	6	6	10	44	6	8	6	80	212
1	Kuala Lumpur	38	2	7	18	6	0	5	12	35	72	18	0	5	142	360
2	New York	44	4	5	21	4	6	4	24	25	84	12	12	4	161	410
3	Dubai	33	1	18	10	2	0	2	6	90	40	6	0	2	144	354

We then normalized these results and got the following

City	Score
Dubai	10.727273
Kuala Lumpur	9.473684
New York	9.318182
Paris	8.153846

From the following results we will now draw our conclusion discuss further observation that could be done.

Discussion

From the results above and the exploratory data analysis we can see that further analysis can be done by considering tips and ratings of all the venues. We could also look at successful countries in vacations and do a cluster analysis and see as to which place resembles the target place which is known to be a good vacation spot. We could also read reviews and use Natural language processing to see whether the reviews are positive or negative sentiment

Conclusion

From the following report and analysis that we have conducted it is the results show that Dubai among Paris, New York and Kuala Lumpur is the vacation destination as it scored most in all relevant categories and had the highest score. Due to Dubai's many outdoor activities and 5 star hotels and amazing Metro travel station it is clear that it deserves first place.