Battle Of The Neighborhoods

CHOOSING THE BEST VACATION DESTINATION AMONG PARIS, KUALA LUMPUR, NEW YORK AND DUBAI

MUHAMMAD UMAR SALMAN

Introduction & Business Problem

- Foreign Tourists need a vacation spot which is the complete package
- They need a place where they have all sort of facilities and activities so they can make their vacation worthwhile
- I will try and attempt to find out from the following 4 cities which should be the next best vacation destination for you
- The following cities are:

1. Paris
2. Kuala Lumpur
3. Dubai
4. New York

Data Section

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.

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
```

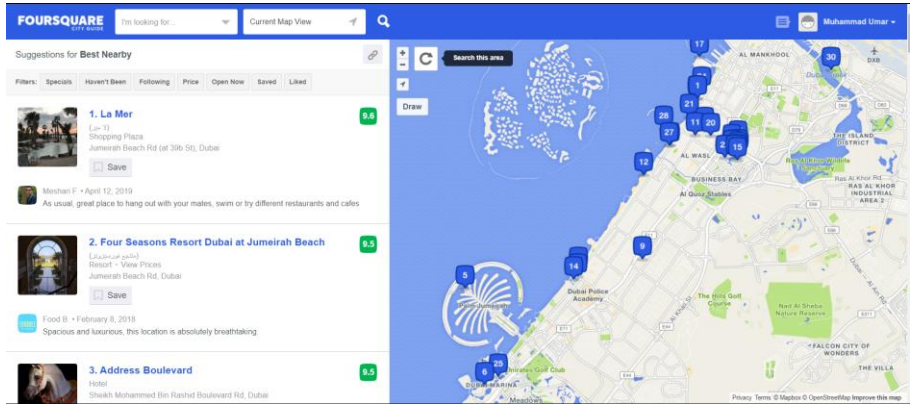
Data Section

Four Square API

- Foursquare is a social location service that allows users to explore the world around them. Users can download the Foursquare application and sign up for free, then connect their Foursquare accounts to their other social media accounts

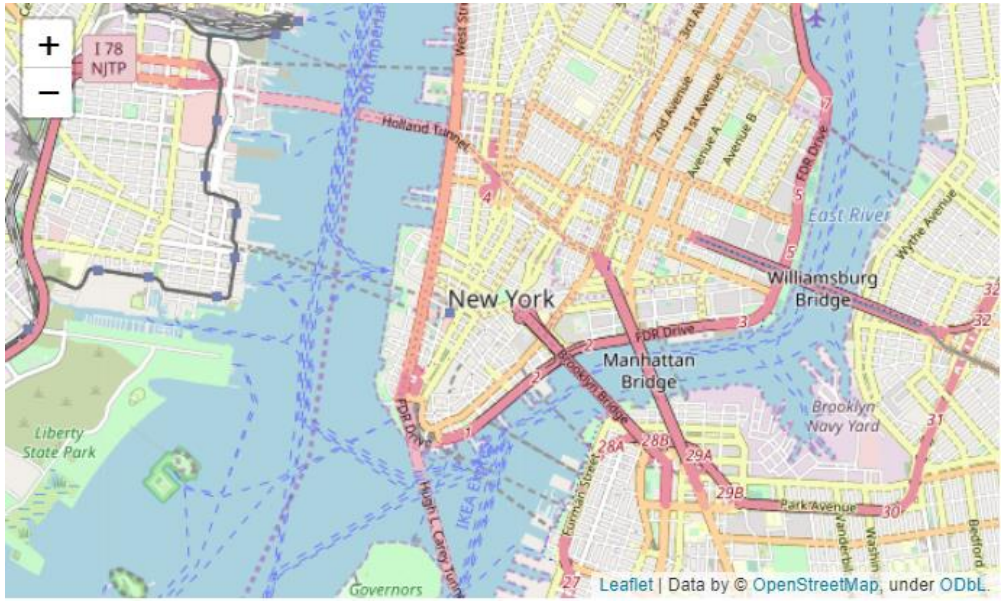
- Here we will use the Four Square API for development purposes to search and extract venues within a given radius and observe the categories and other features of that place

- Here is an example of a Four Square City Guide of Paris



Methodology Section

- 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, here is an example of two. (Kuala Lumpur and New York)



Methodology Section

- After calling the API, we received a JSON file with all the venues of that area. We then converted that JSON to a Data Frame which looks like the image below.
- After changing it to a Data Frame, we had to transform, clean and wrangle the data so that we could put it to proper use
- After that we divided the venues into 6 generic categories and ranked them according to the importance they had for vacation spots. The figure is shown below

	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

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

Methodology Section

- 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

Methodology Section

- We then marked the clusters for each category on the map.
- 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

	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

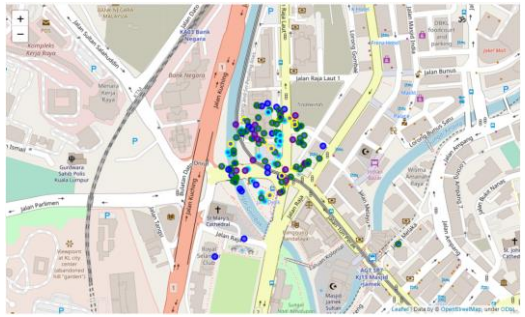
Result Section

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

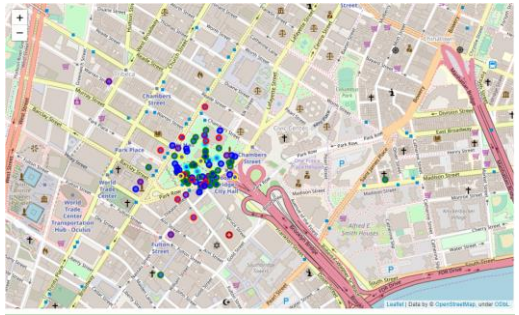
1. Hotels – Yellow
2. Activities – Blue
3. Food - Green
4. Sight Seeing – Cyan
5. Travel - Red
6. Shopping - Purple

Paris

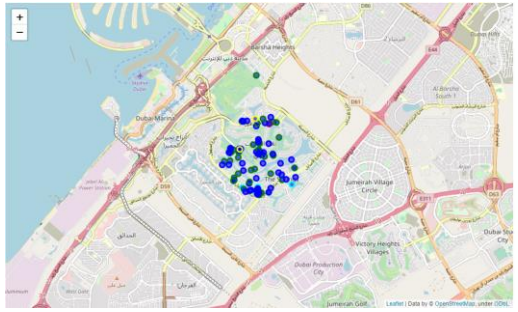
Kuala Lumpur



New York



Dubai



Result Section

- After normalizing the Total by dividing it be the Category Count, which was the total categories we picked up for that certain city.
- We sorted the scores for each city and the city with the higher score was considered the best place to go for your vacation destination.
- Here are the following results:

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

Discussion Section

- 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 Section

- 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 and abundant food options it is clear that it deserves first place.

Thank You