The Battle of Neighborhoods

Dubai

1.Introduction

Dubai is one of the seven emirates that in the United Arab Emirates. Dubai is the emirate of the United Arab Emirates with the highest population and the second largest area (4.114 km²) after Abu Dhabi. Today, Dubai is a commercial and cultural hub as a globally developing city in the Middle East and Persian Gulf region. Although the economy of Dubai has historically been

founded on the oil industry, the Emirate has created important revenue items in the fields of tourism, airlines, real estate transactions and financial services, which it conducts with western-style business methods. Dubai has recently attracted the attention of the world with its large-scale construction projects and sporting events. As a result, thousands of tourists and expats go to Dubai every year.

The project's aim is to help the new expats and tourists who have difficulties to find suitable places. So, using the foursquare API, you can view the restaurant, park, beach, etc. I made various analyzes to find the best place for them. The analysis of the two most visited and settled neighborhoods is also included in this project. This can help Dubai's new guests to get an overview of the places.

2. Data Description

For this project, the Foursquare API is used. A list of neighborhoods in Dubai is downloaded and their respective location in longitude and latitude coordinates is obtained from https://en.wikipedia.org/wiki/List_of_communities_in_Dubai. The data is can be seen below.

	Community Number	Community (English)	Community (Arabic)	Area(km2)	Population(2000)	Population density(/km2)
0	126.0	Abu Hail	أبو هيل	1.27 km²	21414	16,861.4/km²
1	711.0	Al Awir 1	العوير الأولى	NaN	NaN	NaN
2	721.0	Al Awir 2	العوير الثانية	NaN	NaN	NaN
3	283.0	Al Ayas	العياص	162.4 km2	1706	162.4/km2
4	333.0	Al Bada'a	البدع	0.82 km²	18816	22946/km²

I only used Community (English) column and changed the column name to Community; others are dropped. Moreover, communities' latitude and longitude values are added. I used 'geopy' library in order to find the latitude and longitude values of each community.

	Community	Latitude	Longitude
0	Abu Hail, Dubai	25.285942	55.329444
1	Al Awir 1, Dubai	25.169420	55.541634
2	Al Awir 2, Dubai	25.169420	55.541634
3	Al Ayas, Dubai	25.229395	55.275679
4	Al Bada'a, Dubai	25.224107	55.268344

3.Methodology

Firstly, I defined the credentials to connect Foursquare API and explored the venues of 'Abu Hail' community in order to check. The result is shown below.

	name	categories	lat	lng
0	Hamriya Park	Park	25.285710	55.333000
1	Pond Park - Al Qusais	Park	25.288060	55.332606
2	Lively	Track	25.285194	55.325276
3	Jannati Health Club and Spa	Spa	25.285408	55.325168

After that I got all 1504 venues for 77 communities which coordination info is found. The data frame is as follow:

```
print(dubai_venues.shape)
dubai_venues.head()
```

```
(1504, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abu Hail, Dubai	25.285942	55.329444	Hamriya Park	25.285710	55.333000	Park
1	Abu Hail, Dubai	25.285942	55.329444	Pond Park - Al Qusais	25.288060	55.332606	Park
2	Abu Hail, Dubai	25.285942	55.329444	Lively	25.285194	55.325276	Track
3	Abu Hail, Dubai	25.285942	55.329444	Jannati Health Club and Spa	25.285408	55.325168	Spa
4	Al Ayas, Dubai	25.229395	55.275679	Regal	25.231710	55.276337	Women's Store

Then all the venues grouped by neighborhoods and the values are counted. There were 217 unique categories. For feature extraction One Hot Encoding is used. In the result, each feature is a category that belongs to a venue. All feature results are binary which means that 1 means this category is found in the venue and 0 means this category is not found in the venue. Then this result is grouped by neighborhoods which can be seen below.

	Neighborhood	Accessories Store	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Aquarium	Arcade	Art Gallery	Art Museum	...	Tunnel	Turkish Restaurant	Vegeta / Ve Restau
0	Abu Hail, Dubai	0.00	0.0	0.0	0.000000	0.0	0.00	0.0	0.0	0.0	...	0.0	0.0	0.0
1	Al Ayas, Dubai	0.00	0.0	0.0	0.000000	0.0	0.00	0.0	0.0	0.0	...	0.0	0.0	0.0
2	Al Bada'a, Dubai	0.00	0.0	0.0	0.000000	0.0	0.00	0.0	0.0	0.0	...	0.0	0.0	0.0
3	Al Baraha, Dubai	0.00	0.0	0.0	0.090909	0.0	0.00	0.0	0.0	0.0	...	0.0	0.0	0.0
4	Al Barsha 1, Dubai	0.00	0.0	0.0	0.000000	0.0	0.00	0.0	0.0	0.0	...	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
67	Umm Al Sheif, Dubai	0.00	0.0	0.0	0.000000	0.0	0.00	0.0	0.0	0.0	...	0.0	0.0	0.0
68	Umm Ramool, Dubai	0.00	0.0	0.0	0.000000	0.0	0.00	0.0	0.0	0.0	...	0.0	0.0	0.0
69	Wadi Alamardi, Dubai	0.00	0.0	0.0	0.000000	0.0	0.00	0.0	0.0	0.0	...	0.0	0.0	0.0

This followed by the finding top 5 venue categories and calculations of their frequencies for every neighborhood. For example, Downtown Dubai's top 5 venue categories are café (0.09), restaurant (0.08), boutique (0.07), Middle Eastern restaurant (0.05) and coffee shop (0.05). The total frequency of them is 0.34 so the remaining 0.66 is other venue categories. On the other hand Emirates Hill has tennis court (0.5) and bakery (0.5). This implies that Emirates Hill does not have any other venue categories.

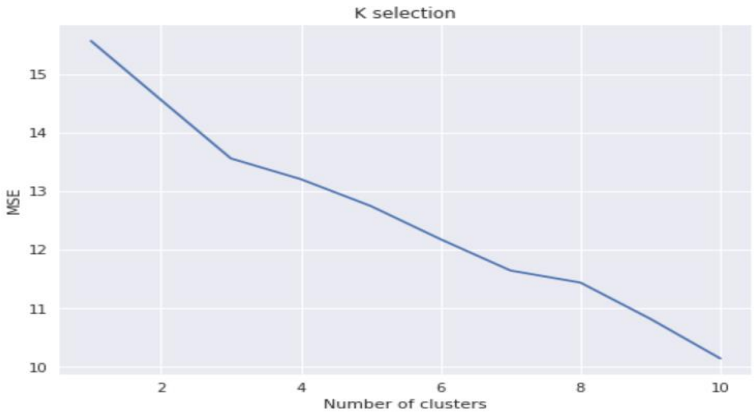
```
----Downtown Dubai, Dubai----
                       venue  freq
0                       Café  0.09
1                 Restaurant  0.08
2                   Boutique  0.07
3  Middle Eastern Restaurant  0.05
4                Coffee Shop  0.05


----Emirates Hill First, Dubai----
               venue  freq
0       Tennis Court   0.5
1             Bakery   0.5
2  Accessories Store   0.0
3      National Park   0.0
4          Nightclub   0.0
```

The top 10 venue categories of each neighborhood are found in order to use this table for future calculations. The result is displayed in below table:

	Neighborhood	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	Abu Hail, Dubai	Park	Spa	Track	Food	Gas Station	Garden	Gaming Cafe	Furniture / Home Store	Fried Chicken Joint	French Restaurant
1	Al Ayas, Dubai	Electronics Store	Grocery Store	Indian Restaurant	Middle Eastern Restaurant	Filipino Restaurant	Asian Restaurant	Restaurant	Japanese Restaurant	Women's Store	Fried Chicken Joint
2	Al Bada'a, Dubai	Park	Pool	Café	Tailor Shop	Middle Eastern Restaurant	Yoga Studio	Garden	Gaming Cafe	Furniture / Home Store	Fried Chicken Joint
3	Al Baraha, Dubai	Middle Eastern Restaurant	Hotel	Coffee Shop	Convenience Store	Mobile Phone Shop	Lounge	Spa	Café	American Restaurant	Furniture / Home Store
4	Al Barsha 1, Dubai	Hotel	Breakfast Spot	Restaurant	Bed & Breakfast	Bakery	Egyptian Restaurant	Convenience Store	Shawarma Place	Coffee Shop	Sandwich Place

We have some common venue categories in neighborhoods. For the purpose of doing unsupervised learning to found similarities between neighborhoods, a clustering algorithm is implemented. I used K-means algorithm to cluster the neighborhoods.

It is necessary to assign the number of clusters because it is an input of this algorithm. For this reason, the elbow method is considered. A chart that compares error vs number of clusters is done and the elbow is selected.



According to the chart above the number of clusters is assigned to 7 and K-means algorithm implemented.

Here is a merged table with cluster labels for each neighborhood.

Latitude	Longitude	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	Cluster Labels
25.285942	55.329444	Park	Spa	Track	Food	Gas Station	Garden	Gaming Cafe	Furniture / Home Store	Fried Chicken Joint	French Restaurant	1
25.229395	55.275679	Electronics Store	Grocery Store	Indian Restaurant	Middle Eastern Restaurant	Filipino Restaurant	Asian Restaurant	Restaurant	Japanese Restaurant	Women's Store	Fried Chicken Joint	1
25.224107	55.268344	Park	Pool	Café	Tailor Shop	Middle Eastern Restaurant	Yoga Studio	Garden	Gaming Cafe	Furniture / Home Store	Fried Chicken Joint	1
25.281368	55.319413	Middle Eastern Restaurant	Hotel	Coffee Shop	Convenience Store	Mobile Phone Shop	Lounge	Spa	Café	American Restaurant	Furniture / Home Store	5
25.111504	55.197187	Hotel	Breakfast Spot	Restaurant	Bed & Breakfast	Bakery	Egyptian Restaurant	Convenience Store	Shawarma Place	Coffee Shop	Sandwich Place	5

4.Results

The study of venues of each neighborhood gave me some results in order to help tourists and expats. Lets see what are the results.

Firstly, I investigated venues that people might interest and where them located and which neighborhoods are famous with that venue category. These venue categories are coffee shop, restaurants, beach, pub, gym and park.

	Neighborhood	1st Most Common Venue
0	Abu Hail, Dubai	Park
1	Al Bada'a, Dubai	Park

	Neighborhood	1st Most Common Venue
0	Palm Jumeira, Dubai	Beach

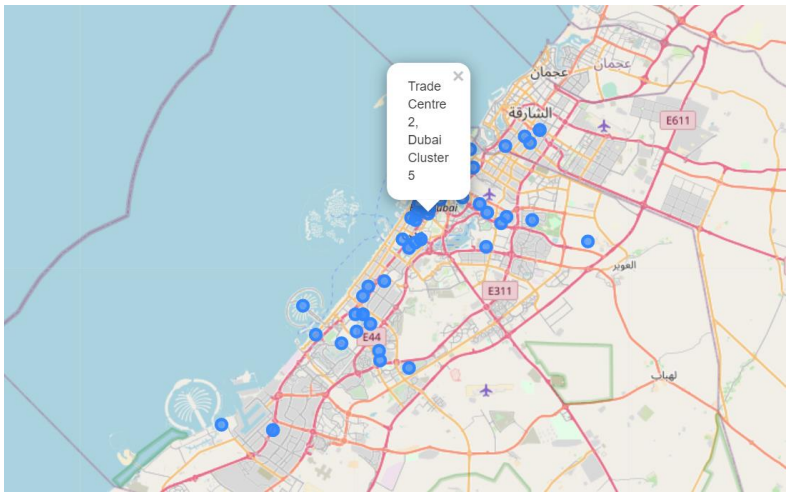
	Neighborhood	1st Most Common Venue
0	Al Quoz First, Dubai	Gym / Fitness Center

	Neighborhood	1st Most Common Venue
0	Al Qusais Second, Dubai	Pub
1	Za'abeel Second, Dubai	Pub

	Neighborhood	1st Most Common Venue
0	Al Barsha South 2, Dubai	Coffee Shop
1	Al Jafiliya, Dubai	Coffee Shop
2	Al Qusais First, Dubai	Coffee Shop
3	Al Satwa, Dubai	Coffee Shop
4	Emirates Hill Third, Dubai	Coffee Shop
5	Mirdif, Dubai	Coffee Shop
6	Ras Al Khor, Dubai	Coffee Shop
7	Rigga Al Buteen, Dubai	Coffee Shop
8	Za'abeel First, Dubai	Coffee Shop

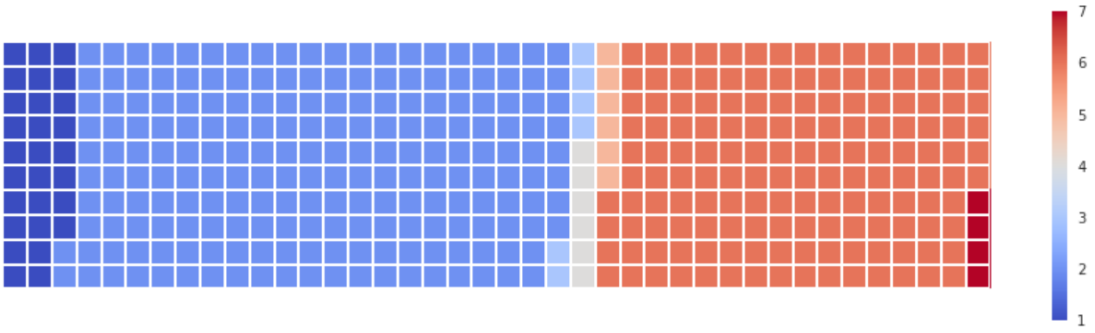
	Neighborhood	1st Most Common Venue
0	Al Baraha, Dubai	Middle Eastern Restaurant
1	Al Hudaiba, Dubai	Asian Restaurant
2	Al Jaddaf, Dubai	Seafood Restaurant
3	Al Karama, Dubai	Indian Restaurant
4	Al Mamzar, Dubai	Restaurant
5	Al Mina, Dubai	Mexican Restaurant
6	Al Muraqqabat, Dubai	Middle Eastern Restaurant
7	Al Muteena, Dubai	Chinese Restaurant
8	Al Qusais Industrial Fifth, Dubai	Fast Food Restaurant
9	Al Rigga, Dubai	Restaurant
10	Al Shindagha, Dubai	Middle Eastern Restaurant
11	Arabian Ranches, Dubai	Restaurant
12	Business Bay, Dubai	Indian Restaurant
13	Jebel Ali 1, Dubai	Mexican Restaurant
14	Jebel Ali 2, Dubai	Mexican Restaurant
15	Umm Al Sheif, Dubai	French Restaurant

After the clustering, data is plotted in a geographical map to get a notion of the world location. In the two following images are shown the neighborhoods in Dubai. Due to some codding reasons the cluster colors are the same but when you click the location you can see that the clustering is done correctly.

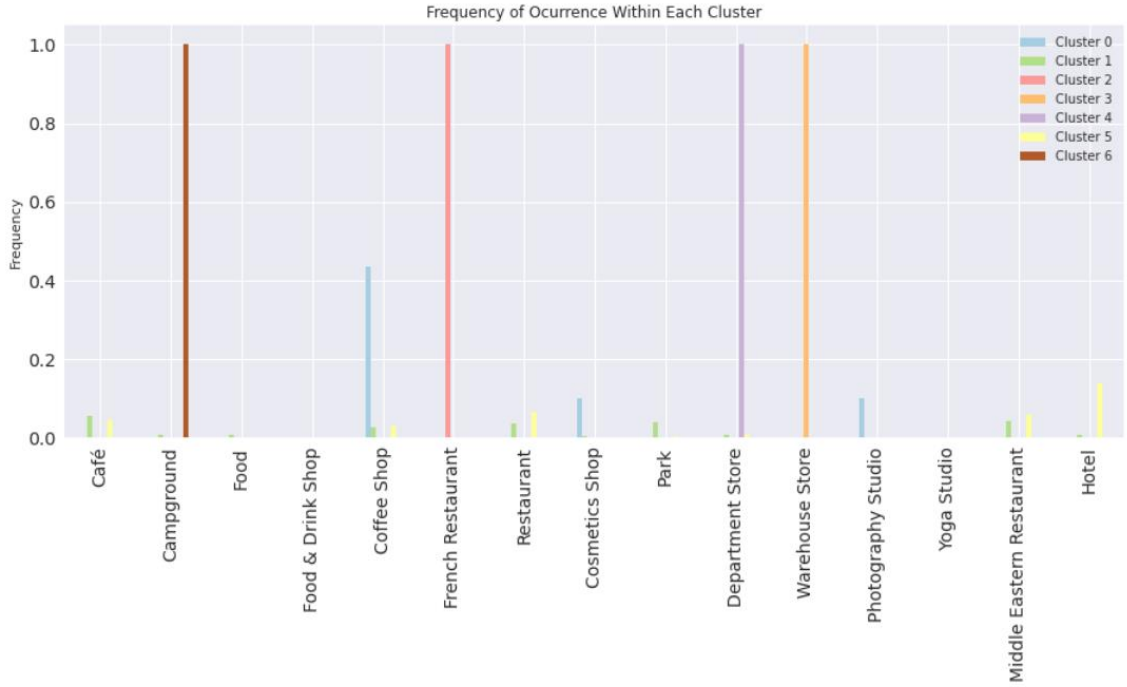


In the image below it is found the proportion of the neighborhoods assigned to each cluster. For this reason, a waffle chart is implemented. There are two major clusters which are 1 and 5.

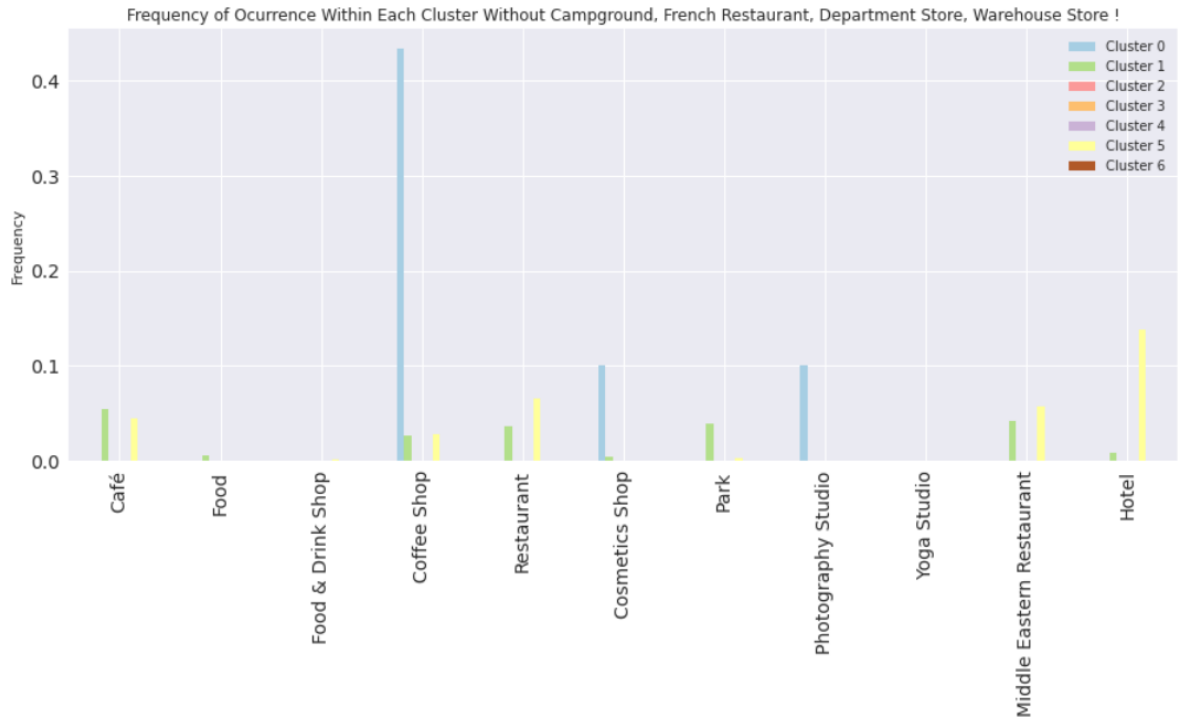
```
Total number of tiles is 400
Cluster 0: 28
Cluster 1: 200
Cluster 2: 6
Cluster 3: 6
Cluster 4: 6
Cluster 5: 150
Cluster 6: 6
```



Bar charts are implemented to find which venue categories are in which clusters and their frequencies.



The image above shows that there some categories with a high frequency of 1 or very close to 1. These categories are campground, French restaurant, department store and warehouse store. These are related to the only one cluster. So, in order to find better results, we can omit these categories.



It can be seen that Cluster 0 mostly includes coffee shops, cosmetics shop and photography studio etc.

A word cloud is generated in order to help finding similar neighborhoods easily. In this way, a person trying to locate a similar neighborhood in Dubai can locate it looking for the neighborhoods with the same color.

5. Discussion

The project is useful only for a limited area because there is a limited amount of data, we can request by using de Foursquare API. The aim of this project is to help people who are tourists visiting Dubai or expats in order to choose the neighborhoods which they want to locate based on the most common venues in it. People can choose a neighborhood by using this project's results. For example, if a person wants to go to the yoga studio and after that she wants to grab a cup of coffee then she can choose a neighborhood in cluster 1 where them have both.

Moreover, there are clusters with one neighborhood. This clusters are cluster 2, cluster 3, cluster 4 and cluster 6. When we analyze these clusters, we can see that they have a frequency of 1 in some venues. For example, cluster 4 has all the department stores. Thus, the algorithm is working well since there is no other cluster with similar venues.

6. Conclusion

As a result of the global world, people travel a lot in our age (if we ignore the coronavirus situation which is happening nowadays). I tried to help to people who wants to travel to Dubai.

In today's time of digital world, data science plays vital role. It is always good to use new tools to make life easier and became one step ahead. I have used some common libraries like geopy, folium etc. Also, I have used foursquare API to explore the venues of each neighborhoods and K-means algorithms. These form small part of data science, but It is a good start.

I wish you a healthy day.