Capstone Project — The Battle of Communities in Dubai: Restaurants

Dara Sakhnini

Wednesday, July 8, 2020

1. Introduction

Dubai is a growing city with many opportunities for investors and working professionals. The city, once a desert, grows and develops into one of the richest cities in the world. Therefore, it becomes increasingly important to examine and understand its communities quantitatively.

Entrepreneurs, investors, and city planners have an interest in identifying opportunities and growing urban footprints in specific communities. Therefore, after the top food venues for each community are presented, this information could be of interest to potential investors looking to open a new restaurant in Dubai.

Also, for tourists visiting Dubai, it would be of interest to cluster similar communities according to their top food venues. Therefore, when visiting Dubai, for the utmost experience, it would be preferable to visit a community from each cluster in order to insure variety.

2. Data requirements

For this project, we need following data:

a. List of communities in Dubai from Wikipedia https://en.wikipedia.org/wiki/List_of_communities_in_Dubai
b. Latitude and Longitude of each communities from Geocoder from Geopy Library of Python.
c. Different types of venues in each community from Foursquare API.

3. Methodology

3.1 Data Preparation

3.1.1 Scrapping data from Wikipedia

I first make use of 'List of communities in Dubai' page from Wikipedia to scrap the table to create a data frame. For this, I've used pandas to transform the data in the table on the Wikipedia page into a data frame containing name of the 131 communities, Area, Population, and Population density, as shown in figure 1.

	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 First	العوير الأولى	NaN	NaN	NaN
2	721.0	Al Awir Second	العوير الثانية	NaN	NaN	NaN
3	333.0	Al Bada	البدع	0.82 km²	18816	22946/km²
4	122.0	Al Baraha	البراحة	1.104 km²	7823	7,086/km²

Figure 1 Scrapping data from Wikipedia

3.1.2 Getting Coordinates

Next objective is to get the coordinates of these 131 communities using geocoder class of Geopy client, as shown in figure 2.

	Community Number	Community (English)	Community (Arabic)	Area(km2)	Population(2000)	Population density(/km2)	Latitude	Longitude
0	126.0	Abu Hail	أبو هيل	1.27 km²	21414	16,861.4/km²	25.2859	55.3294
1	711.0	Al Awir First	العوير الأولى	NaN	NaN	NaN	null	null
2	721.0	Al Awir Second	العوير الثانية	NaN	NaN	NaN	null	null
3	333.0	Al Bada	البدع	0.82 km²	18816	22946/km²	24.7542	46.8218
4	122.0	Al Baraha	البراحة	1.104 km²	7823	7,086/km²	25.2814	55.3194

Figure 2 Getting Coordinates

3.1.3 Data Cleaning

Next comes cleaning the data by deleting null values, converting all values into floats, removing units, and deleting unrequired columns. The cleaned data is shown in figure 3.

	Neighborhood	Area(km2)	Population(2000)	Population density(/km2)	Latitude	Longitude
0	Abu Hail	1.270	21414.0	16861.4	25.2859	55.3294
3	Al Bada	0.820	18816.0	22946.0	24.7542	46.8218
4	Al Baraha	1.104	7823.0	7086.0	25.2814	55.3194
5	Al Barsha First	38.100	1248.0	33.0	25.1083	55.2048
11	Al Buteen	0.070	2364.0	33771.0	25.2631	55.3206

Figure 3 Data Cleaning

3.1.4 Map

I used python folium library to visualize geographic details of Dubai and its 131 major communities. I used latitude and longitude values to get the visual shown in figure 4.

```
The geograpical coordinate of Dubai are 25.0750095, 55.18876088183319.
```

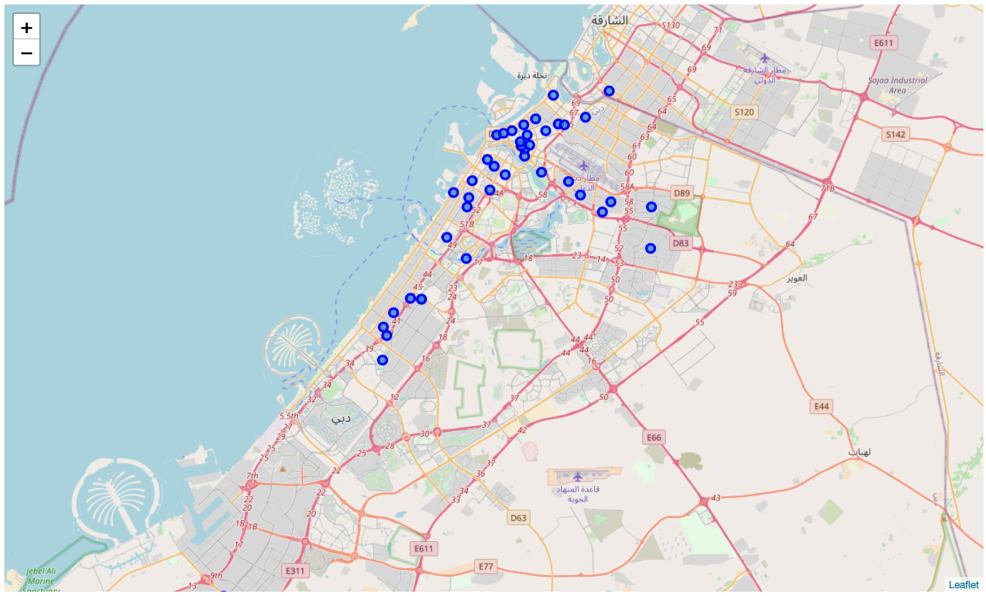

Figure 4 Map

3.2 Data Analysis

3.2.1 Fetching Foursquare Location Data

Finally, let's make use of Foursquare API and get the top 100 venues that are within a radius of 500 meters of each community in Dubai. First four venues are shown in the figure 5.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abu Hail	25.285942	55.329444	Hamriya Park	25.285710	55.333000	Park
1	Abu Hail	25.285942	55.329444	Pond Park - Al Qusais	25.288060	55.332606	Park
2	Abu Hail	25.285942	55.329444	Al Farooj	25.288186	55.326324	Restaurant
3	Abu Hail	25.285942	55.329444	Zahr El Laymoun	25.289486	55.330753	Middle Eastern Restaurant
4	Al Bada	24.754227	46.821832	Tamimi Markets Express (التميمي إكسبرس)	24.754251	46.824108	Supermarket

Figure 5 List of Venues

3.2.2 Analyzing Venue Data

In the figure below, it can be seen that in these Dubai communities, there is a total of 163 unique categories of venues. Also, coffee shops are the most common, as it is at the top of the list in the figure below.

```
print('There are {} uniques categories.'.format(len(dubai_venues['Venue Category'].unique())))
```

```
There are 163 uniques categories.
```

```
dubai_venues_top10 = dubai_venues['Venue Category'].value_counts().rename_axis('unique_values').reset_index(name='counts').head(10)
dubai_venues_top10
```

	unique_values	counts
0	Café	84
1	Hotel	81
2	Coffee Shop	59
3	Middle Eastern Restaurant	59
4	Restaurant	38
5	Indian Restaurant	37
6	Asian Restaurant	26
7	Fast Food Restaurant	22
8	Ice Cream Shop	21
9	Bakery	14

Figure 6 Analyzing Venue Data

3.2.3 Filtering Restaurant Data

Next, I will concentrate on Restaurant category only. I have filtered the restaurant venues as shown in the sample data frame.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
2	Abu Hail	25.285942	55.329444	Al Farooj	25.288186	55.326324	Restaurant
3	Abu Hail	25.285942	55.329444	Zahr El Laymoun	25.289486	55.330753	Middle Eastern Restaurant
7	Al Bada	24.754227	46.821832	La Casa Pasta	24.753953	46.824110	Italian Restaurant
23	Al Bada	24.754227	46.821832	McDonald's (ماكدونالدز)	24.753915	46.824254	Fast Food Restaurant
25	Al Bada	24.754227	46.821832	McDonald's ماكدونالدز	24.750435	46.822568	Fast Food Restaurant

Figure 7 Filtering Restaurant Data

Middle Eastern restaurants top the chart as we can see in the figure below, figure 8.

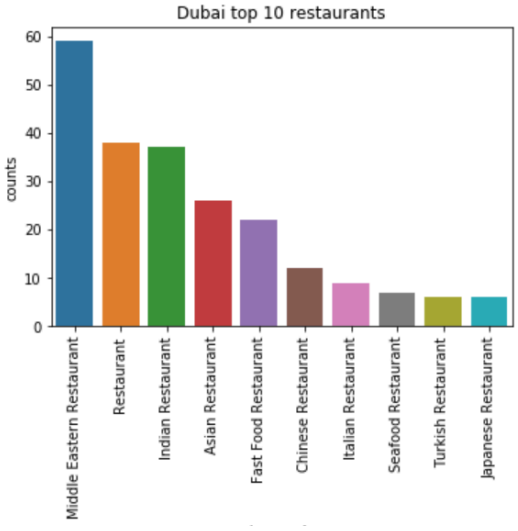

Figure 8 Top 10 restaurants in Dubai

Al Karma has the highest number of restaurants, as seen below.

	unique_values	counts
0	Al Karama	29
1	Al Hudaiba	18
2	Al Muraqqabat	17
3	Umm Hurair First	16
4	Al Buteen	15
5	Nadd Al Shiba Third	13
6	Al Rigga	12
7	Al Raffa	12
8	Jumeira First	12
9	Umm Hurair Second	12

Figure 9 Communities with highest number of restaurants

3.2.4 Top Restaurants in each community

Let's analyze each neighborhood to know about the top 5 venues of each one, as shown in figure a.

So we create a data-frame with pandas one hot encoding for the venue categories, use pandas group by on neighborhood column and calculate the mean of the frequency of occurrence of each venue category, and finally output each neighborhood along with the top 5 most common venues.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Abu Hail	Middle Eastern Restaurant	Restaurant	Yemeni Restaurant	Greek Restaurant	Kebab Restaurant
1	Al Bada	Fast Food Restaurant	Middle Eastern Restaurant	Italian Restaurant	Indian Restaurant	Yemeni Restaurant
2	Al Baraha	Middle Eastern Restaurant	American Restaurant	Yemeni Restaurant	Greek Restaurant	Kebab Restaurant
3	Al Barsha First	Indian Restaurant	Seafood Restaurant	Middle Eastern Restaurant	Chinese Restaurant	Restaurant
4	Al Buteen	Restaurant	Middle Eastern Restaurant	Mediterranean Restaurant	Asian Restaurant	Turkish Restaurant
5	Al Garhoud	Restaurant	Yemeni Restaurant	Lebanese Restaurant	Kebab Restaurant	Japanese Restaurant
6	Al Hudaiba	Middle Eastern Restaurant	Asian Restaurant	Chinese Restaurant	Seafood Restaurant	Filipino Restaurant
7	Al Jaddaf	Indian Restaurant	Yemeni Restaurant	Vegetarian / Vegan Restaurant	Korean Restaurant	Kebab Restaurant
8	Al Jafiliya	Fast Food Restaurant	Restaurant	Yemeni Restaurant	Greek Restaurant	Kebab Restaurant
9	Al Karama	Indian Restaurant	Asian Restaurant	African Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant

Figure 10 Top Restaurants in each community

3.2.5 Clustering

Finally, we try to cluster these the communities based on the venue categories using K-Means clustering. So, we expect the communities to be clustered based on the similarities of venue categories.

We finally form a comprehensive data frame with cluster labels, latitude and longitude, most common venues, and population and area data.

	Neighborhood	Area(km2)	Population(2000)	Population density(/km2)	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Abu Hail	1.270	21414.0	16861.4	25.2859	55.3294	0.0	Middle Eastern Restaurant	Restaurant	Yemeni Restaurant	Greek Restaurant	Kebab Restaurant
3	Al Bada	0.820	18816.0	22946.0	24.7542	46.8218	2.0	Fast Food Restaurant	Middle Eastern Restaurant	Italian Restaurant	Indian Restaurant	Yemeni Restaurant
4	Al Baraha	1.104	7823.0	7086.0	25.2814	55.3194	0.0	Middle Eastern Restaurant	American Restaurant	Yemeni Restaurant	Greek Restaurant	Kebab Restaurant
5	Al Barsha First	38.100	1248.0	33.0	25.1083	55.2048	0.0	Indian Restaurant	Seafood Restaurant	Middle Eastern Restaurant	Chinese Restaurant	Restaurant
11	Al Buteen	0.070	2364.0	33771.0	25.2631	55.3206	2.0	Restaurant	Middle Eastern Restaurant	Mediterranean Restaurant	Asian Restaurant	Turkish Restaurant

Figure 11 Full table with cluster labels

We can represent these 3 clusters in a leaflet map using Folium library as shown in the following figure. Burj Khalifa Downtown area, which is represented by a purple dot in figure a, has a variety of restaurants different than other communities.

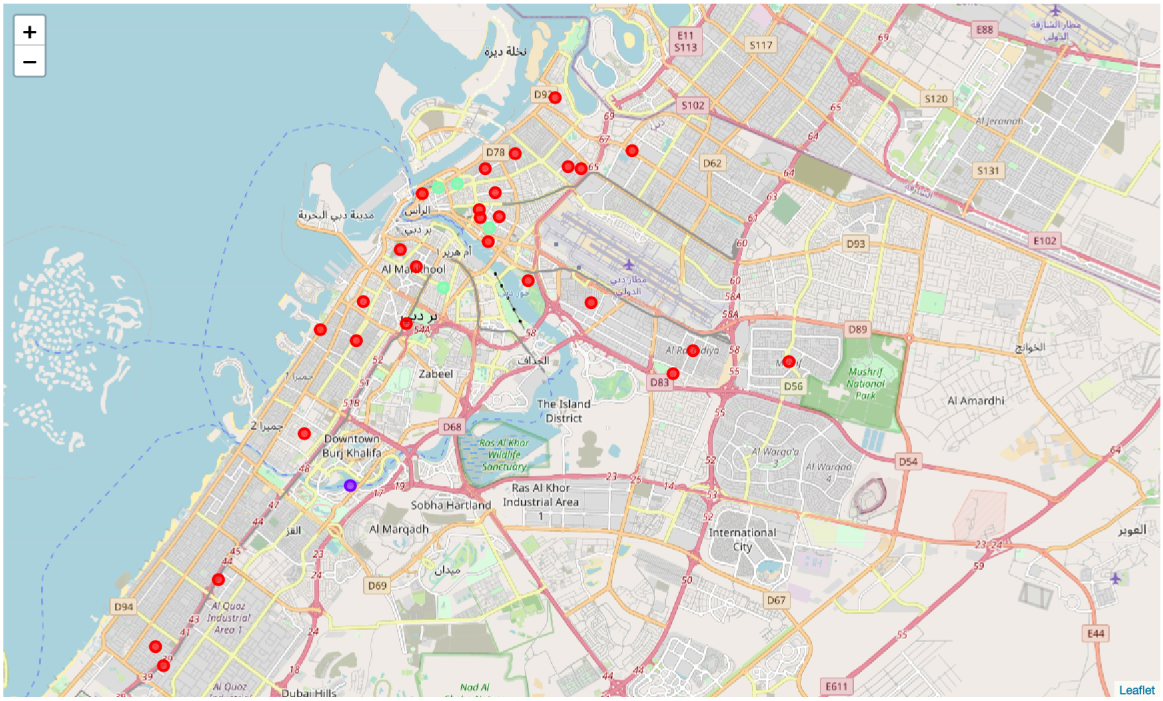

Figure 12 Map with cluster representations

4. Results and discussions

We got a glimpse of the Restaurants in Dubai and were able to find out some interesting insights which might be useful to travelers as well as people with business interests. Let's summarize our findings:

· Middle Eastern restaurants top the charts of most common venues in Dubai communities.

· Al Karama community has maximum number of restaurants.

· Since the clustering was based only on the category of restaurants on each district, Burj Khalifa Downtown area has a variety of restaurants different than other communities.

· Al Mamzar, Za'abeel First, and Al Rashidiya has the least number of restaurants.

The clustering is based on the most common venues obtained from Foursquare data.

However, in this analysis, we have ignored other factors like distance of the venues from closest stations, range of prices of restaurants, and so on, since we don't have such data and it would be difficult to obtain it for a small exploratory study like ours. Hence, our analysis only helps to get a rough overview of Restaurants distribution by categories in the 131 communities of Dubai.

Also, these results also could potentially vary if we use other clustering techniques.

5. Conclusion

There are many real-life scenarios where data can be extremely useful for a wide range of people, from tourists to people with business interests. Like seen in the example above, data was used to cluster neighborhoods in Dubai based on the most common restaurants in its 131 communities. The results can help a traveler to decide about the district that fit the most his needs.

I have made use of an API library to scrap data from Wikipedia, used Foursquare API to explore the major communities of Dubai, and visualized the results of clustering the communities using Folium leaflet map. Finally, some of the drawbacks and chance for improvements to represent even more realistic pictures are mentioned.