Analyzing the Neighborhoods in Dubai for setting up a New Fast Food Restaurant

IBM Applied Data Science Capstone Project

By: Ahmad Kanbari

2021-6-12

Since the first of January 2021, I have been working to achieve an IBM Data Science Professional certification, It took me 6 months to reach this level. The course teaches what it means to be a data scientist and the tools that you need to succeed in this field. The classes included Python programming, SQL databases, machine learning, statistics, and more. As a Capstone for the certification, students are required to define a business problem and pull data from the web and location data from Foursquare to solve it. This report details my process for determining the best spot in Dubai to open a Fast Food Restaurant or a food truck, from problem statement to data preparation to analysis to recommendations. Detailed code for this project is located on GitHub and can be found at the end of this post.

1- Introduction

1.1- Business Problem

Food is a lucrative business industry and the Middle East is the hub of delicious culinary delights from all over the world. Nowadays with covid-19 presence, Fast-food restaurants have proven to provide safe and quick meals to people on the go and offer an alternative to traditional, sit-down restaurants. Imagine you have opened a fast-food restaurant or a food truck whether its Western fast-food style where you serve burgers and drinks or it's a middle eastern style where you sell shawarma, falafel, and other sandwiches. By following the food code of UAE and choosing the right place, This business will be so profitable as it provides tasty food in a quick, low-cost manner.

Before opening a fast-food restaurant, however, researches the city and determine if your fast-food business has enough potential demand from the population. So what factors should you analyze to determine the best place in

the city to start your fast food restaurant, and how do you do it? This report outlines some basic assumptions, data sets, and analyses that can inform your decision when selecting the optimal neighborhood in Dubai for opening a Fast Food restaurant or your food truck.

1.2- Target Audience

The target audience of this report would be anyone who wants to buy or build a fast food restaurant or fast food truck in Dubai.

2- Data

The dataset that I used in this project is of Dubai communities scrapped from Wikipedia. This dataset contains a list of 153 communities of Dubai.

Data source:

https://en.wikipedia.org/wiki/List_of_communities_in_Dubai

I scrapped the data from the table of Wikipedia using a python library called 'Beautiful soup'. We will use only 4 columns of the dataset i.e. Community Number, Community (English) and Community (Arabic) for venue analysis of Dubai and the population density column. Following data sources will be needed to extract/generate the required information:

- I Will use the GEOPY python library to generate Latitude and Longitude for the Abu Dhabi Island
- I Will use the GEOCODER python library to geocode coordinates (Latitudes and Longitude for schools in Abu Dhabi.
- number of venues and their type and location in every neighborhood will be obtained using FOURSQUARE API

3 – Methodology

3.1- The first step in this project is to scrap the data from Wikipedia page then clean it up by dropping unnecessary columns, renaming the remaining columns, and getting rid of NaN values.

We can grab this data and place it into a pandas data frame using python. This is going to help us clean and analyze it to generate insights.

As we see the data contains 153 records and 7 columns

```
df_Dubai.shape
```

```
(153, 7)
```

Let’s clean up the data and delete unnecessary columns and Nan values and makes it ready for analysis. The result table :

```
[14] df_Dubai.head(7)
```

	Neighborhood_Number	Neighborhoods	Neighborhood_Arabic	Population_density
0	126.0	Abu Hail	ابو هيل	16,861.4/km2
1	711.0	Al Awir First	العوير الأولى	NaN
2	721.0	Al Awir Second	العوير الثانية	NaN
3	283.0	Aleyas	العياص	162.4/km2
4	333.0	Al Bada'a	البدع	22946/km2
5	122.0	Al Baraha	البراحة	7,086/km2
6	373.0	Al Barsha First	البرشاء الأولى	33/km2

Now after cleaning it has 153 records and 4 columns

```
[15] df_Dubai.shape
```

```
(153, 4)
```

3.2- The second step is to get the longitude and latitude of Dubai city and each neighborhood in Dubai by using geopy python libraries.

Let’s show the top five Neighborhoods with their longitude and latitude data.

```
[25] Dubai_data.head(5)
```

	Neighborhoods	Latitude	Longitude
0	Abu Hail	25.286029	55.328865
1	Al Awir First	25.185184	55.565170
2	Aleyas	25.211788	55.536023
3	Al Bada'a	25.224626	55.268483
4	Al Baraha	25.281062	55.319466

3.2.1-Then we can create a map that shows all our Neighborhoods in Dubai

Folium library was used to visualize a map of Dubai and neighborhoods were superimposed on top of it as in the picture below:

Then I used Foursquare API to explore areas around center of each neighborhood and find the 10 most common venues in each neighborhood. I've set the radius of search to 500 meters around the center of each neighborhood and limit to 100 venues.

Foursquare data is robust and provides location data for Apple and Uber. Foursquare API will allow us to retrieve information about the most popular spots in each neighborhood in San Francisco. This will be an insightful indication of foot traffic for different venue types. Calling the Foursquare API returns a JSON file, which can be turned into a data frame for analysis in a python notebook.

But Let me first explore AlAwir first neighborhood in Dubai:

```
[ ] #exploring the venues of 'Al Awir First, Dubai'
    Dubai_data.loc[1,'Neighborhoods']
```

```
'Al Awir First'
```

```
[ ] neighborhood_latitude = Dubai_data.loc[0, 'Latitude'] # neighborhood latitude value
    neighborhood_longitude = Dubai_data.loc[0, 'Longitude'] # neighborhood longitude value

    neighborhood_name = Dubai_data.loc[1, 'Neighborhoods'] # neighborhood name

    print('Latitude and longitude values of {} are {}, {}.'.format(neighborhood_name,
                                                                   neighborhood_latitude,
                                                                   neighborhood_longitude))
```

```
Latitude and longitude values of Al Awir First are 25.2860207, 55.328065.
```

By using FourSquar API and getting The results, clean it and categorize it wi will get a result table of all venues in Al Awir first neighborhood and their categories:

	name	categories	lat	lng
0	Al Douri Mart Supermarket & Roastery	Supermarket	25.285869	55.328174
1	Baithak Restaurant	Asian Restaurant	25.288937	55.327372
2	Lively	Track	25.285194	55.325276
3	Jannati Health Club and Spa	Spa	25.285408	55.325168
4	Hamriya Park	Park	25.285710	55.333000

4- This will lead us to the 4th step which is doing the same process for each neighborhood in Dubai :

This is an example of the data resulting from the analysis using foursquare API:

```
print(Dubai_venues.shape)
Dubai_venues.head()
```

```
(544, 7)
```

	Neighborhoods	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abu Hail	25.286029	55.328865	Al Douri Mart Supermarket & Roastery	25.285869	55.328174	Supermarket
1	Abu Hail	25.286029	55.328865	Baithak Restaurant	25.288937	55.327372	Asian Restaurant
2	Abu Hail	25.286029	55.328865	Lively	25.285194	55.325276	Track
3	Abu Hail	25.286029	55.328865	Jannati Health Club and Spa	25.285408	55.325168	Spa
4	Abu Hail	25.286029	55.328865	Hamriya Park	25.285710	55.333000	Park

The result was 544 venues returned by Foursquare in Dubai. The next step was to create a table, which shows a list of the top 10 venues for each Borough:

5- In the fifth step we will analyze the data, by digging further into each of the neighborhoods to see the most popular types of venues for each neighborhood. To do this, we will take the following steps:

- Create a data frame of venue categories with pandas one-hot encoding

- Use pandas group by to get the mean of the one-hot encoded venue categories
- Transpose the data frame and arrange in descending order

The result table will contain 80 records(Neighborhoods) and 147 Category type:

```
[ ]  dubai_grouped.shape

     (80, 147)
```

The table will look like that :

```
dubai_grouped = dubai_onehot.groupby('Neighborhoods').mean().reset_index()
dubai_grouped
```

	Neighborhoods	Afghan Restaurant	African Restaurant	Airport Terminal	American Restaurant	Arcade	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Garage	Auto Workshop	BBQ Joint	Baby Store
0	Abu Hail	0.0	0.0	0.000	0.0	0.0	0.000000	0.000000	0.166667	0.0	0.000000	0.0	0.000000	0.0
1	Al Awir First	0.0	0.0	0.000	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.0
2	Al Bada'a	0.0	0.0	0.000	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.0
3	Al Baraha	0.0	0.0	0.000	0.1	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.0
4	Al Barsha First	0.0	0.0	0.000	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.0
5	Al Buteen	0.0	0.0	0.000	0.0	0.0	0.000000	0.000000	0.100000	0.0	0.000000	0.0	0.000000	0.0
6	Al Corniche	0.0	0.0	0.000	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.0
7	Al Dhagaya	0.1	0.0	0.000	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.0
8	Al Garhoud	0.0	0.0	0.000	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.0
9	Al Hamriya Port	0.0	0.0	0.000	0.0	0.5	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.0
10	Al Hamriya, Dubai	0.0	0.0	0.000	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.0

```
[ ]  num_top_venues = 5
     i =1
     for hood in dubai_grouped['Neighborhoods']:
         print("----"+hood+"----",i)
         i = i+i
         temp = dubai_grouped[dubai_grouped['Neighborhoods'] == hood].T.reset_index()
         temp.columns = ['venue','freq']
         temp = temp.iloc[1:]
         temp['freq'] = temp['freq'].astype(float)
         temp = temp.round({'freq': 2})

         print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
         print('\n')
```

The code provides us with the top 5 venues for each neighborhood:

```
----Abu Hail---- 1
              venue  freq
0              Park  0.33
1               Spa  0.17
2  Asian Restaurant  0.17
3             Track  0.17
4       Supermarket  0.17


----Al Awir First---- 2
                 venue  freq
0    Convenience Store  0.25
1  Fast Food Restaurant  0.25
2   Seafood Restaurant  0.25
3          Coffee Shop  0.25
4  Pakistani Restaurant  0.00


----Al Bada'a---- 3
                      venue  freq
0                      Park   0.4
1                      Pool   0.2
2  Middle Eastern Restaurant   0.2
3                      Café   0.2
4     Outdoors & Recreation   0.0
```

This data is important because it is giving us an idea of the atmosphere of each of these neighborhoods. That helps us to know whether any location is already a hot spot for other fast-food restaurants and food trucks. So, I'm going to create several data frames from this information and use them to examine the atmosphere of our potential neighborhoods.

Let's try to know more about Neighborhoods by Knowing the 10 most common venues categories in each Neighborhood.

```
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhoods']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhoods'] = dubai_grouped['Neighborhoods']

for ind in np.arange(dubai_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(dubai_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

The result will be looks like that :

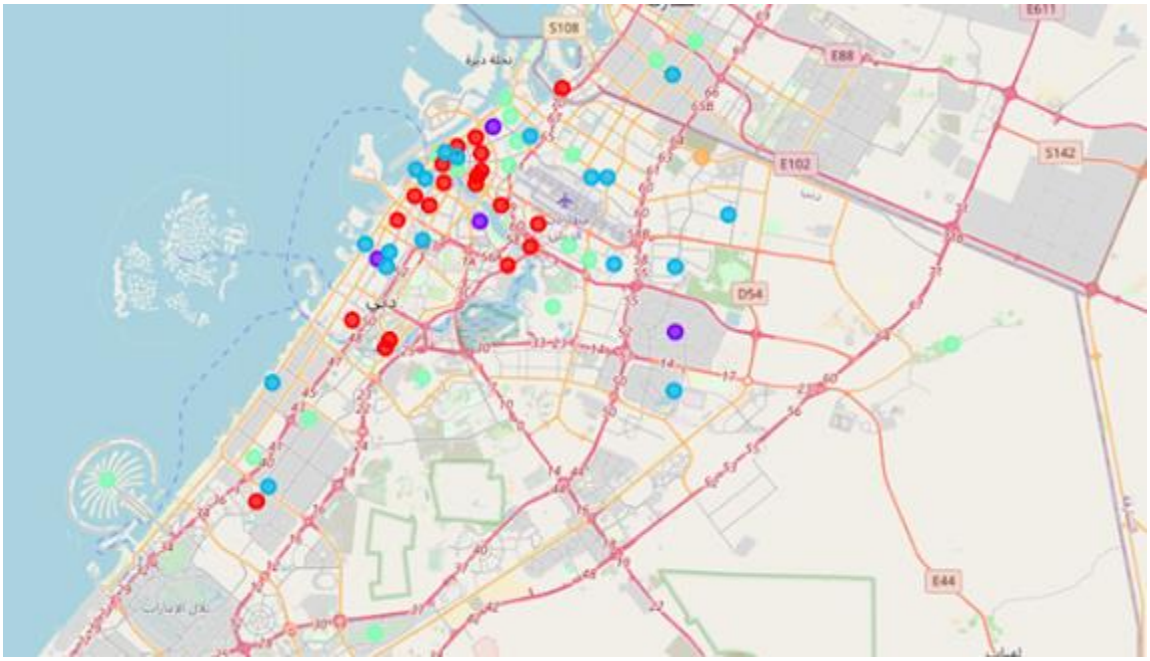
	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Abu Hail	Park	Track	Spa	Asian Restaurant	Supermarket	Yoga Studio	Flower Shop	Fried Chicken Joint	French Restaurant	Food Truck
1	Al Awir First	Coffee Shop	Fast Food Restaurant	Convenience Store	Seafood Restaurant	Flower Shop	Fried Chicken Joint	French Restaurant	Food Truck	Food & Drink Shop	Food
2	Al Bada'a	Park	Middle Eastern Restaurant	Pool	Café	Flower Shop	Fried Chicken Joint	French Restaurant	Food Truck	Food & Drink Shop	Food
3	Al Baraha	Hotel	Middle Eastern Restaurant	Spa	Lounge	Convenience Store	Café	IT Services	American Restaurant	Arts & Crafts Store	Food Truck
4	Al Barsha First	Ice Cream Shop	Hotel	Juice Bar	Café	Burrito Place	Bakery	Egyptian Restaurant	Turkish Restaurant	Indian Restaurant	Fruit & Vegetable Store

6- In the sixth step we will use machine learning and use K-means clustering algorithm to group neighborhoods into 5 distinct groups, based on the top 10 venues in each neighborhood. By clustering our neighborhoods based on their popular venue categories This will help us get a feel for which neighborhoods are like each other based on the venues people like to visit in each one.

The results of the K-Means algorithm will result in the following table:

	Neighborhoods	Latitude	Longitude	Cluster Labels	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
0	Abu Hail	25.286029	55.328865	1	Park	Track	Spa	Asian Restaurant	Supermarket	Yoga Studio	Flower Shop	Fried Chicken Joint
1	Al Awir First	25.185184	55.585170	3	Coffee Shop	Fast Food Restaurant	Convenience Store	Seafood Restaurant	Flower Shop	Fried Chicken Joint	French Restaurant	Food Truck
3	Al Bada'a	25.224626	55.268483	1	Park	Middle Eastern Restaurant	Pool	Café	Flower Shop	Fried Chicken Joint	French Restaurant	Food Truck
4	Al Baraha	25.281082	55.319466	0	Hotel	Middle Eastern Restaurant	Spa	Lounge	Convenience Store	Café	IT Services	American Restaurant
5	Al Barsha First	25.111724	55.206427	0	Ice Cream Shop	Hotel	Juice Bar	Café	Burrito Place	Bakery	Egyptian Restaurant	Turkish Restaurant
7	Al Buteen	25.263057	55.320584	0	Hotel	Ice Cream Shop	Restaurant	Asian Restaurant	Fried Chicken Joint	Coffee Shop	Middle Eastern Restaurant	Fish Market

Let's visualize the clusters using folium python library:



7- After clustering, we will select neighborhoods where parks are one of the most first 3 venue categories and try to find the ideal neighborhood for our business.

The best Neighborhood to open a fast food restaurant will be a place where many people can gather from different places in Dubai, so neighborhoods where parks and shopping centers are located will be better for a fast food restaurant or food truck, for being able to reach the maximum number of possible customers. also within those neighborhoods the ideal neighborhood will have the highest population density.

Let's find out the best place for opening a fast food restaurant in Dubai

```
#there are 7 neighborhood will a good optimal place to open a neighborhood

#select full all data for those neighborhoods
df_neighborhoods_withparks = dubai_parks_combined['Neighborhoods']

df_neighborhoods_withparks
```

```
0    Al Warqa'a Third
1          Al Bada'a
2           Abu Hail
3     Al Safa Second
4             Mirdif
5    Umm Hurair First
6       Al Safa First
Name: Neighborhoods, dtype: object
```

Let’s merge our table with the df_Dubai table to add population density columns to our existing table. the result will be : (some columns are not shown)

	Neighborhoods	Latitude	Longitude	Cluster Labels	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	10t Mos Comm Venu
0	Abu Hail	25.288029	55.328865	1	Park	Track	Spa	Asian Restaurant	Supermarket	Yoga Studio	Flower Shop	Fried Chicken Joint	French Restaurant	Foo Truc
1	Al Bada'a	25.224626	55.269483	1	Park	Middle Eastern Restaurant	Pool	Café	Flower Shop	Fried Chicken Joint	French Restaurant	Food Truck	Food & Drink Shop	Foo
2	Al Safa First	15.604586	32.557628	1	Garden	Soccer Stadium	Park	Concert Hall	BBQ Joint	Food	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Foo Truc
3	Al Safa Second	29.324196	30.830316	1	Café	Park	Yoga Studio	Flower Shop	Fried Chicken Joint	French Restaurant	Food Truck	Food & Drink Shop	Food	Fle Mark
4	Al Warqa'a Third	25.190861	55.422755	1	Park	Yoga Studio	Flower Shop	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Food Truck	Food & Drink Shop	Food	Fle Mark
5	Mirdif	25.221171	55.422420	2	Coffee Shop	Park	Seafood Restaurant	Café	Spa	Nail Salon	Gym	Grocery Store	Yoga Studio	Flow Sho
6	Umm Hurair First	25.242212	55.321784	0	Pharmacy	Park	Café	Coffee Shop	Yoga Studio	Flower Shop	Fried Chicken Joint	French Restaurant	Food Truck	Food Dri Sho

Solution :

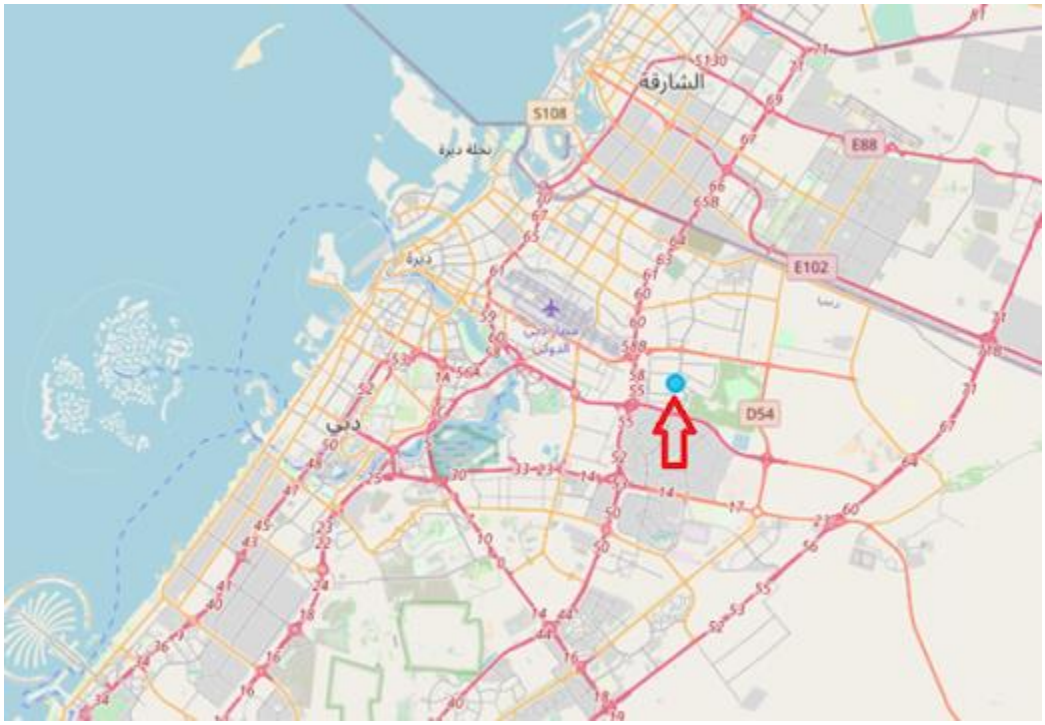
Option 1: Choosing the ideal place based on less competition.

for this option, we are going to use cluster 0 to choose the ideal neighborhood where food trucks and fast food restaurants are not among the 10 most common venues category.

As The above table provided the Mirdif neighborhood is the only place where there are no food trucks or fast-food restaurants as the ideal neighborhood.

Selecting this neighborhood may lead to the rapid growth of your business due to lack of competition and due to be near a park.

Let's show the location of the Mirdif neighborhood :



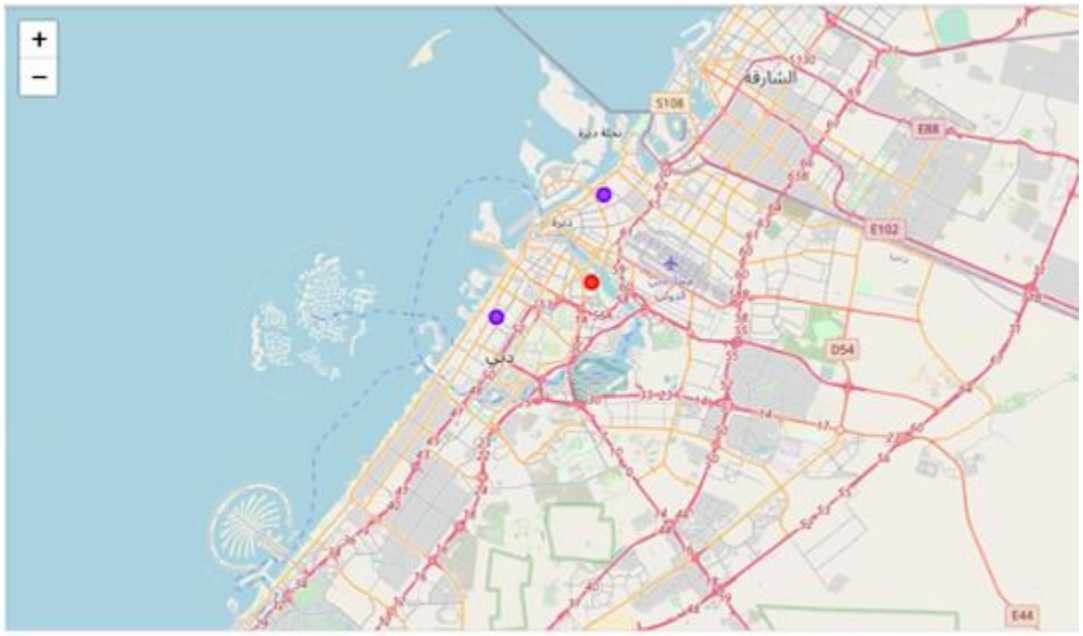
Option2: choosing ideal place only based on population density

let's sort our neighborhoods on population densities (population per area) in descending order and select the top 3 neighborhoods that contain park among 1st,2nd, and 3rd most common venues and have the highest population density.

```
#let's choose top 3 neighborhoods with top populatin density
top3_neighborhoods= top_neighborhoodss.iloc[:3,:]
top3_neighborhoods
```

Population_density	Neighborhoods	Latitude	Longitude	Cluster Labels	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
22946.0	Al Bada'a	25.224626	55.268483	1	Park	Middle Eastern Restaurant	Pool	Café	Flower Shop	Fried Chicken Joint	French Restaurant	Food Truck
16861.4	Abu Hail	25.286029	55.328865	1	Park	Track	Spa	Asian Restaurant	Supermarket	Yoga Studio	Flower Shop	Fried Chicken Joint
1926.0	Umm Hurair First	25.242212	55.321784	0	Pharmacy	Park	Café	Coffee Shop	Yoga Studio	Flower Shop	Fried Chicken Joint	French Restaurant

Let’s show their locations :



Discussion Section

The major purpose of this project is to suggest the ideal neighborhood in Dubai for who wants to open a fast food restaurant or a food truck. being near popular venues where people get together, checking for population density and competitiveness are among important factors to choose the ideal neighborhood.

Results

If you want to choose the ideal neighborhood where the competition is low you should go with Mirdif neighborhood. If you care more about population density you can choose among top neighborhoods where population density is high such as Al Babaji, Abu Hail or Umm Hurair First.

Conclusion

Foursquare data and K-means algorithm, alongside folium mapping tools, can be powerful tools helping people decide the best places for doing such a business. where Data is the driver of the decision. This approach can be replicated for various other tasks, e.g. comparing two different cities or choosing the best neighborhood to live in.