



Predicting best location and type of business to invest in Dubai-UAE.

Zakaria Jobran | July 07. 2021

1. Introduction

1.1 Background:

Dubai city is that the most thickly settled city within the UAE. it is multicultural. It provides ton of business opportunities and business friendly environment. it is attracted many alternative players into the market. it is a world hub of business and commerce. the town may be a major center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, ancient media, advertising, legal services, accountancy, insurance, theater, fashion, and therefore the arts in the UAE.

This also means the market is very competitive. because it is highly developed city so price of doing business is additionally one of the highest. Thus, any new business venture or growth must be analyzed carefully. The insights derived from analysis will offer sensible understanding of the business atmosphere that facilitate in strategically targeting the market. this can help in reduction of risk. and therefore, the come back on Investment are going to be reasonable.

1.2 Problem

An agent hired me to find out which type of business would be good to invest in Dubai and asked me to pick a location for that business.

To do that. Various factors need to be studied in order to decide on the Location such as:

- Neighborhoods Population.
- Dubai Demographics.
- Locations in Dubai.
- Startup costs.
- Capital and Finance.
- Competitions.

Even though well-funded agent. need to choose the correct location to start its first venture. If this is successful, they can replicate the same in other locations. First move is very important, thereby choice of location is very important.

1.3 Interest

To advocate the right location. Agent has appointed Pine Tree State to guide of the information Science team. the target is to find and recommend to the management that neighborhood of metropolis is going to be most suitable option to begin a replacement business. The Management additionally expects to know the principle of the recommendations made.

This would interest anyone who wants to start a new business in Dubai.

2. Data acquisition and cleaning

2.1 Data source

This report used data source can be found [here](#) which it has Dubai's neighborhoods.

And reading location latitude and longitude for the data from the above link from [this site](#).

2.2 Data cleaning & feature selection.

After retrieving the data from the source. I monitored the data by removing some neighborhoods that are not have coordinates on [foursquare.com](#) or not recognized.

Converting data types as required to be able to do statistical methods on them.

Then merge both data from the data frames to work in one data frame.

Do some statistical operations to find:

- a- Neighborhood with most population.
- b- Most common business invested in the area.

These operations to help to find the suitable place and type of investment to start new business.

2.3 Feature selection

After cleaning and preparing data, we select *neighborhood* and *population* as futures to select proper location to open new business.

3. Exploratory Data Analysis

We found that most population area was business bay with 191000.

Out[7]:

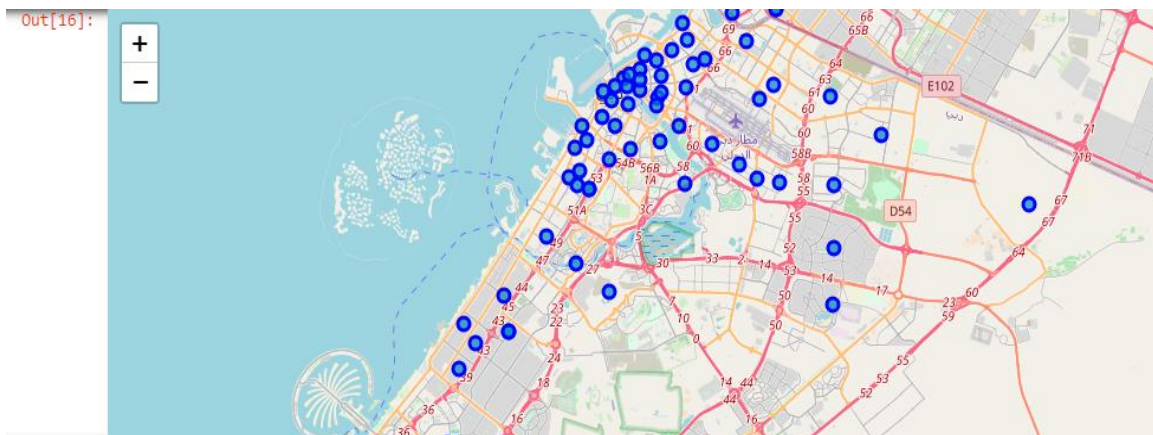
	Community_Number	Community_Name	Community_Population
76	346	Business Bay	191000

Furthermore, using geocoder library we retrieve latitude and longitude for every neighbor as follow:

Out[14]:

	Community_Number	Community_Name	Community_Population	Latitude	Longitude
0	126	Abu Hail	21414	25.286029	55.328865
1	283	Aleyas	1706	25.211788	55.536023
2	333	Al Bada'a	18816	25.224626	55.268483
3	122	Al Baraha	7823	25.281062	55.319466
4	373	Al Barsha First	1248	25.096326	55.198402
5	671	Al Barsha South First	1248	25.074716	55.232048
6	114	Al Buteen	2364	25.263057	55.320584
7	111	Al Corniche	1135	25.283817	55.312858
8	113	Al Dhagaya	10896	25.272641	55.300241
9	945	Al Fagaa	22	24.736692	55.529170
10	214	Al Garhoud	4466	25.240737	55.352039

After that using folium library, we display a map for every neighborhood.



After that finding venues for every neighborhood in the data.

Out[23]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abu Hail	25.286029	55.328865	Baithak Restaurant	25.288937	55.327372	Asian Restaurant
1	Abu Hail	25.286029	55.328865	Lively	25.285194	55.325276	Track
2	Abu Hail	25.286029	55.328865	Jannati Health Club and Spa	25.285408	55.325168	Spa
3	Abu Hail	25.286029	55.328865	Hamriya Park	25.285710	55.333000	Park
4	Abu Hail	25.286029	55.328865	Pond Park - Al Qusais	25.288060	55.332606	Park
5	Al Bada'a	25.224626	55.268483	Al Boom Diving Club	25.227329	55.266449	Pool
6	Al Bada'a	25.224626	55.268483	Al Bada'a Park	25.224104	55.268313	Park
7	Al Bada'a	25.224626	55.268483	cafe aroma	25.226160	55.270326	Café
8	Al Bada'a	25.224626	55.268483	Asma Tailor	25.226449	55.270716	Tailor Shop
9	Al Bada'a	25.224626	55.268483	Al Wasl Park (حديقة الوصل)	25.222958	55.265674	Park

With statistical operations we found that Trade center1 & 2 have more venues more than others.

	Mirrah	28	28	28	28	28	28
Muhaisnah First	1	1	1	1	1	1	1
Nad Shamma	8	8	8	8	8	8	8
Naif	13	13	13	13	13	13	13
Port Saeed	54	54	54	54	54	54	54
Rigga Al Buteen	44	44	44	44	44	44	44
Trade Centre 1	93	93	93	93	93	93	93
Trade Centre 2	93	93	93	93	93	93	93
Umm Hurair First	4	4	4	4	4	4	4
Umm Ramool	3	3	3	3	3	3	3
Umm Suqeim First	34	34	34	34	34	34	34
Warsan First	6	6	6	6	6	6	6
Za'abeel Second	4	4	4	4	4	4	4

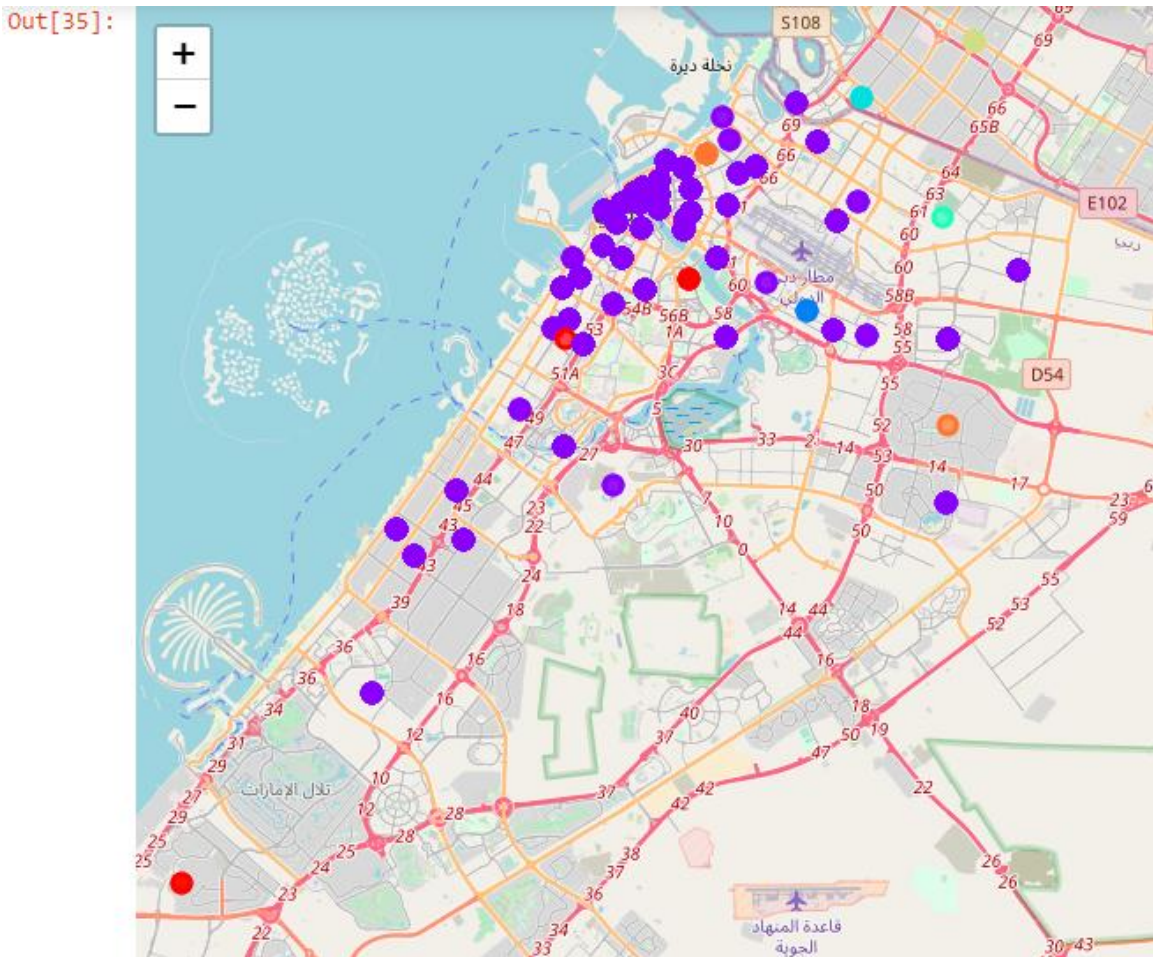
4. Predictive Modeling

Using clustering model, I used *kcluster* library with 7 clusters, with top ten venues and append them to the original dataset.

Out[34]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	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	25.286029	55.328865	Baithak Restaurant	25.288937	55.327372	Asian Restaurant	6	Park	Spa	Asian Restaurant	Track	ATM	F
1	Abu Hail	25.286029	55.328865	Lively	25.285194	55.325276	Track	6	Park	Spa	Asian Restaurant	Track	ATM	F
2	Abu Hail	25.286029	55.328865	Jannati Health Club and Spa	25.285408	55.325168	Spa	6	Park	Spa	Asian Restaurant	Track	ATM	F
3	Abu Hail	25.286029	55.328865	Hamriya Park	25.285710	55.333000	Park	6	Park	Spa	Asian Restaurant	Track	ATM	F

Finally, visualizing clusters in the map as follow.



As a result, we found that the blue cluster the most venue in Dubai which is demonstrated as cluster 2(cluster index 1).

When analyzing each cluster, as shown in the below figures

Cluster 1

```
Dubai_merged.loc[Dubai_merged['Cluster Labels'] == 0, Dubai_merged.columns[[1] + list(range(5, Dubai_merged
```

	Neighborhood Latitude	Venue Longitude	Venue Category	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
547	25.220860	55.278277	Coffee Shop	0	Coffee Shop	ATM	Performing Arts Venue	Music Venue	Nail Salon
715	25.028782	55.125027	Coffee Shop	0	Coffee Shop	ATM	Performing Arts Venue	Music Venue	Nail Salon
716	25.028782	55.125027	Coffee Shop	0	Coffee Shop	ATM	Performing Arts Venue	Music Venue	Nail Salon
1098	25.242212	55.325432	Park	0	Coffee Shop	Park	ATM	Performing Arts Venue	Music Venue
1099	25.242212	55.323613	Coffee Shop	0	Coffee Shop	Park	ATM	Performing Arts Venue	Music Venue
1100	25.242212	55.319879	Coffee Shop	0	Coffee Shop	Park	ATM	Performing Arts Venue	Music Venue
1101	25.242212	55.318589	Coffee Shop	0	Coffee Shop	Park	ATM	Performing Arts Venue	Music Venue

Cluster 2

Dubai_merged.loc[Dubai_merged['Cluster Labels'] == 1, Dubai_merged.columns[[1] + list(range(5, Dubai_me										
			Neighborhood		Restaurant	Supply Store	Restaurant		Center	Sports
1141	25.163154	55.423351	Kitchen Supply Store	1	Indian Restaurant	Kitchen Supply Store	Chinese Restaurant		Garden Center	Athletics & Sports
1142	25.163154	55.425349	Garden Center	1	Indian Restaurant	Kitchen Supply Store	Chinese Restaurant		Garden Center	Athletics & Sports
1143	25.163154	55.417187	Chinese Restaurant	1	Indian Restaurant	Kitchen Supply Store	Chinese Restaurant		Garden Center	Athletics & Sports
1144	25.163154	55.420807	Athletics & Sports	1	Indian Restaurant	Kitchen Supply Store	Chinese Restaurant		Garden Center	Athletics & Sports
1145	24.426362	54.407521	Coffee Shop	1	Café	Organic Grocery	Coffee Shop		Shoe Store	ATM
1146	24.426362	54.408767	Organic Grocery	1	Café	Organic Grocery	Coffee Shop		Shoe Store	ATM
1147	24.426362	54.407614	Café	1	Café	Organic Grocery	Coffee Shop		Shoe Store	ATM
						Organic				

Cluster 3

```
: Dubai_merged.loc[Dubai_merged['Cluster Labels'] == 2, Dubai_merged.columns[[1] + list(range(5, Dubai_merge
```

```
:
```

	Neighborhood	Latitude	Longitude	Venue Category	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1102	25.231032	55.363126	Auto Garage	2	Auto Garage	ATM	Persian Restaurant	Music Venue	Nail Salon	
1103	25.231032	55.370951	Auto Garage	2	Auto Garage	ATM	Persian Restaurant	Music Venue	Nail Salon	
1104	25.231032	55.366726	Auto Garage	2	Auto Garage	ATM	Persian Restaurant	Music Venue	Nail Salon	

Cluster 4

```
: Dubai_merged.loc[Dubai_merged['Cluster Labels'] == 3, Dubai_merged.columns[[1] + list(range(5, Dubai_merged.shape[1]))]]
```

	Neighborhood	Latitude	Longitude	Venue Category	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 Common V
398	25.305900	55.384942	Indian Restaurant	3	Indian Restaurant	ATM	Performing Arts Venue	Museum	Music Venue	Nail Salon	Nightclub	
762	25.306377	55.384942	Indian Restaurant	3	Print Shop	Indian Restaurant	Performing Arts Venue	Museum	Music Venue	Nail Salon	Nightclub	
763	25.306377	55.392912	Print Shop	3	Print Shop	Indian Restaurant	Performing Arts Venue	Museum	Music Venue	Nail Salon	Nightclub	

Cluster 5

```
: Dubai_merged.loc[Dubai_merged['Cluster Labels'] == 4, Dubai_merged.columns[[1] + list(range(5, Dubai_merged.shape[1]))]]
```

Cluster 6

```
: Dubai_merged.loc[Dubai_merged['Cluster Labels'] == 5, Dubai_merged.columns[[1] + list(range(5, Dubai_merge
```

```
:

```

	Neighborhood Latitude	Venue Longitude	Venue Category	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
410	25.326002	55.432073	Restaurant	5	Fast Food Restaurant	Restaurant	Lounge	ATM	Persian Restaurant
411	25.326002	55.435927	Fast Food Restaurant	5	Fast Food Restaurant	Restaurant	Lounge	ATM	Persian Restaurant
412	25.326002	55.435304	Lounge	5	Fast Food Restaurant	Restaurant	Lounge	ATM	Persian Restaurant
413	25.326002	55.432073	Restaurant	5	Fast Food Restaurant	Restaurant	Lounge	ATM	Persian Restaurant
414	25.326002	55.435927	Fast Food Restaurant	5	Fast Food Restaurant	Restaurant	Lounge	ATM	Persian Restaurant
415	25.326002	55.435304	Lounge	5	Fast Food Restaurant	Restaurant	Lounge	ATM	Persian Restaurant

Cluster 7

```
] : Dubai_merged.loc[Dubai_merged['Cluster Labels'] ==6, Dubai_merged.columns[[1] + list(range(5, Dubai_merged
```

```
] :

```

	Neighborhood Latitude	Venue Longitude	Venue Category	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	25.286029	55.327372	Asian Restaurant	6	Park	Spa	Asian Restaurant	Track	ATM
1	25.286029	55.325276	Track	6	Park	Spa	Asian Restaurant	Track	ATM
2	25.286029	55.325168	Spa	6	Park	Spa	Asian Restaurant	Track	ATM
3	25.286029	55.333000	Park	6	Park	Spa	Asian Restaurant	Track	ATM
4	25.286029	55.332606	Park	6	Park	Spa	Asian Restaurant	Track	ATM
621	25.190661	55.419432	Park	6	Park	ATM	Movie Theater	Museum	Music Venue

After that, I found that the most common business invested in, is Hotel, Cafés and Coffee shops respectively.

Venue Category	
Hotel	99
Café	75
Coffee Shop	70
Middle Eastern Restaurant	64
Indian Restaurant	47
Restaurant	36
Asian Restaurant	31
Fast Food Restaurant	24
Burger Joint	23
Gym	23
Chinese Restaurant	19
Clothing Store	16
Shopping Mall	16
Ice Cream Shop	16
Sandwich Place	15

5. Conclusions

As a result for this study. We found that the best choice was to invest in hotels followed by cafés or coffee shops in Business Bay neighborhood, because it has most population and venues with same investments and interests.