

# “End to end journey of Dubai data coming from source systems to predictive modelled output”

Dubai has a diverse infrastructure, diverse society which directly decides the quality of living. Which is what makes Dubai a preferred city for people to live in. When I decided to move to Dubai, purpose was to explore and find a good area with better facilities around neighborhood. Questions I aimed to answer through this project are -

1. List and visualize all major parts of Dubai city with Top infrastructure
2. What are best locations in Dubai as per infrastructure?
3. Which areas have the potential for development of infrastructure of different kinds?
4. Which areas lack infrastructure facilities?
5. What is the best place to stay in Dubai?

Divided the document in steps to show the flow of analytical process.

**Step 1.** Scrapped data from webpage using various python libraries and acquired list of various communities of Dubai.

	Community Number\n	Community (English)	Community (Arabic)	Area(km2)	Population(2000)	Population density(/km2)\n
0	126\n	Abu Hail	أبو هيل	1.27 km2	21,414	16,861.4/km2\n
1	711\n	Al Awir First	العوير الأولى			\n
2	721\n	Al Awir Second	العوير الثانية			\n
3	283\n	Aleyas	العياص	162.4 km2	1,706	162.4/km2\n
4	333\n	Al Bada'a	البدع	0.82 km2	18,816	22946/km2\n
...	...	...	...	...	...	...
147	913\n	Umm Nahad Third				\n
148	914\n	Umm Nahad Fourth				\n
149	971\n	Saih Al-Dahal				\n
150	951\n	Saih Al Salam				\n
151	931\n	Al Lisaili				\n

152 rows x 6 columns

**Step 2.** Found respective latitude and longitude of each community while dropping those whose addresses were not found.

	Community Number	Community (English)	Latitude	Longitude	City
0	126	Abu Hail	25.286029	55.3289	Dubai
1	711	Al Awir First	25.185184	55.5652	Dubai
3	283	Aleyas	25.211788	55.536	Dubai
4	333	Al Bada'a	25.224626	55.2685	Dubai
5	122	Al Baraha	25.281062	55.3195	Dubai
...	...	...	...	...	...
139	621	Warsan First	25.163154	55.4221	Dubai
141	861	Yaraah	14.424065	44.0112	Dubai
142	325	Za'abeel First	25.227589	55.2888	Dubai
143	337	Za'abeel Second	-6.177421	35.75	Dubai
144	841	Margham	24.899518	55.6255	Dubai

90 rows x 5 columns

**Step 3.** Defined Foursquare [network service] credentials, version and searched for 200 top venues within 600 meters of each community

	Community (English)	Community Number	Latitude	Longitude	City	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Abu Hail	126	25.286029	55.328865	Dubai	Al Douri Mart Supermarket & Roastery	25.285869	55.328174	Supermarket
1	Abu Hail	126	25.286029	55.328865	Dubai	Balthak Restaurant	25.288937	55.327372	Asian Restaurant
2	Abu Hail	126	25.286029	55.328865	Dubai	Lively	25.285194	55.325276	Track
3	Abu Hail	126	25.286029	55.328865	Dubai	Jannati Health Club and Spa	25.285408	55.325168	Spa
4	Abu Hail	126	25.286029	55.328865	Dubai	Hamriya Park	25.285710	55.333000	Park
...	...	...	...	...	...	...	...	...	...
1531	Za'abeel Second	337	-6.177421	35.749954	Dubai	Younique Pub Dodoma	-6.180089	35.749450	Pub
1532	Za'abeel Second	337	-6.177421	35.749954	Dubai	Nyerere Square	-6.179483	35.752127	Park
1533	Za'abeel Second	337	-6.177421	35.749954	Dubai	Masterpub	-6.178891	35.753326	Pub
1534	Za'abeel Second	337	-6.177421	35.749954	Dubai	Double M	-6.178769	35.753961	Electronics Store
1537	Margham	841	24.899518	55.625454	Dubai	Dubai Desert Conservation Reserve	24.903710	55.627335	Nature Preserve

368 rows × 9 columns

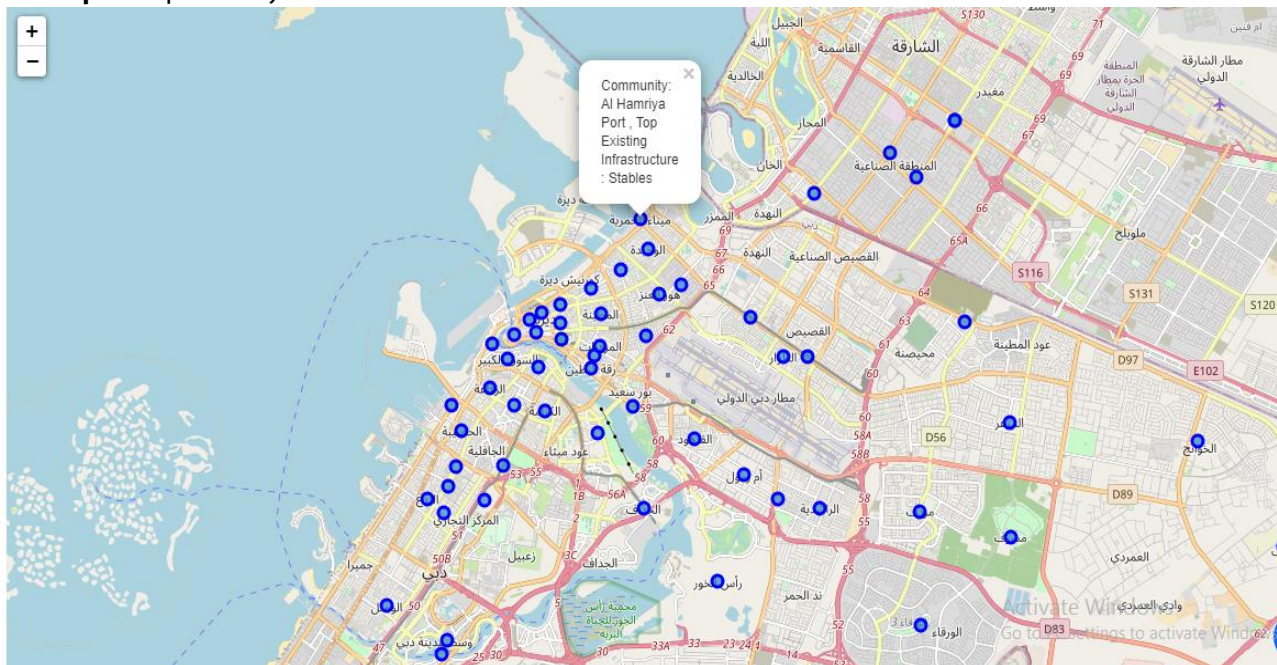
**Step 4.** Grouped the data by Community's name and its number to find unique venue categories.

**Step 5.** Analyzed each community for venue category and displayed top 5 existing facility for each community.

0		Downtown Dubai	Dubai	Hotel	Coffee Shop	Café	Middle Eastern Restaurant	French Restaurant
1	111	Al Corniche	Dubai	French Restaurant	Bakery	Pizza Place	Flower Shop	Food Court
2	112	Al Ras	Dubai	History Museum	Jewelry Store	Middle Eastern Restaurant	Farmers Market	Café
3	113	Al Dhagaya	Dubai	Afghan Restaurant	Jewelry Store	Hotel	Men's Store	Fast Food Restaurant
4	114	Al Buteen	Dubai	Hotel	Restaurant	Middle Eastern Restaurant	Coffee Shop	Ice Cream Shop
...	...	...	...	...	...	...	...	...
75	600	Arabian Ranches	Dubai	Restaurant	Sports Bar	Golf Course	Pool	Yoga Studio
76	616	Nadd Al Shiba Third	Dubai	Café	Coffee Shop	Middle Eastern Restaurant	Turkish Restaurant	Gym
77	621	Warsan First	Dubai	Chinese Restaurant	Fast Food Restaurant	Indian Restaurant	Moroccan Restaurant	Gym
78	711	Al Awir First	Dubai	Fast Food Restaurant	Seafood Restaurant	Coffee Shop	Convenience Store	Yoga Studio
79	841	Margham	Dubai	Nature Preserve	Yoga Studio	Egyptian Restaurant	Gastropub	Gas Station

80 rows × 8 columns

## Step 6. Exploratory Visualization



# Feature Business Problem

**Step 1:** Defined 'Quality Infrastructure' like school, Hospitals, theater, ATM, Office, Bus Station, Bank, Market, Business Service, Monument / Landmark,Resort,Hospital,Police Station, School, College, Café, Park, Playground, Convention Center, College Auditorium, Government Building, Airport Terminal etc.

**Step 2:** Queried the 'Community with top 5 facility' dataset with respect to these 'Quality Infrastructure' parameters.

	Community (English)	Community Number	Latitude	Longitude	City	VenueName	VenueLatitude	VenueLongitude	VenueCategory
4	Abu Hail	126	25.286029	55.328865	Dubai	Hamriya Park	25.285710	55.333000	Park
5	Abu Hail	126	25.286029	55.328865	Dubai	Pond Park - Al Qusais	25.288060	55.332606	Park
14	Al Bada'a	333	25.224626	55.268483	Dubai	Al Bada'a Park	25.224104	55.268313	Park
15	Al Bada'a	333	25.224626	55.268483	Dubai	cafe aroma	25.226160	55.270326	Café
17	Al Bada'a	333	25.224626	55.268483	Dubai	Al Wasl Park (حديقة الوصل)	25.222958	55.265674	Park
...	...	...	...	...	...	...	...	...	...
1525	Za'abeel First	325	25.224255	55.286425	Dubai	Novotel World Trade Centre Dubai	25.221415	55.288294	Hotel
1528	Za'abeel First	325	25.224255	55.286425	Dubai	Entre-Nous	25.221485	55.288592	Restaurant
1532	Za'abeel Second	337	-6.177421	35.749954	Dubai	Nyerere Square	-6.179483	35.752127	Park
1534	Za'abeel Second	337	-6.177421	35.749954	Dubai	Double M	-6.178769	35.753961	Electronics Store
1535	Za'abeel Second	337	-6.177421	35.749954	Dubai	Jamatini	-6.181216	35.748118	Bus Station

396 rows x 9 columns

Step 3: Found out communities with total number of infrastructure and discovered the area which is best according to infrastructure.

59	
Community (English)	Rigga Al Buteen
Community Number	125
City	Dubai
Bank	1
Bus Station	0
Business Service	0
Café	3
Electronics Store	0
Farmers Market	0
Gym / Fitness Center	0
Hotel	19
Indie Movie Theater	0
Light Rail Station	0
Market	0
Metro Station	0
Office	0
Park	0
Pharmacy	0
Playground	0
Resort	0
Restaurant	6

## Potential area for the development of infrastructure of different kinds

### Looking for area in development of infrastructure in Community

Step 1. *My choice of infrastructure for which Community has highest potential: 'Restaurant'*

0	Abu Hail
2	Al Baraha
5	Al Dhagaya
8	Al Hudaiba
9	Al Jaddaf
12	Al Khabisi
13	Al Kifaf
14	Al Mamzar
15	Al Manara
17	Al Merkad
18	Al Mina
19	Al Mizhar First
21	Al Murar
22	Al Mushrif
29	Al Raffa
30	Al Ras
31	Al Rashidiya
33	Al Sabkha
34	Al Safa First
35	Al Satwa
37	Al Souq Al Kabeer
39	Al Warqa'a Third
43	Downtown Dubai
45	Emirates Hill First
46	Emirates Hill Second
49	Jebel Ali 1
50	Jebel Ali 2

Step 2. *My choice of area for which one infrastructure has highest potential: 'Al Karama'*

These are infrastructures with highest potential in Al Karama area :

Bank

Bus Station

Business Service

Electronics Store

Farmers Market

Indie Movie Theater

Light Rail Station

Market

Metro Station

Office

Park

Pharmacy

Playground

Resort

Shopping Mall



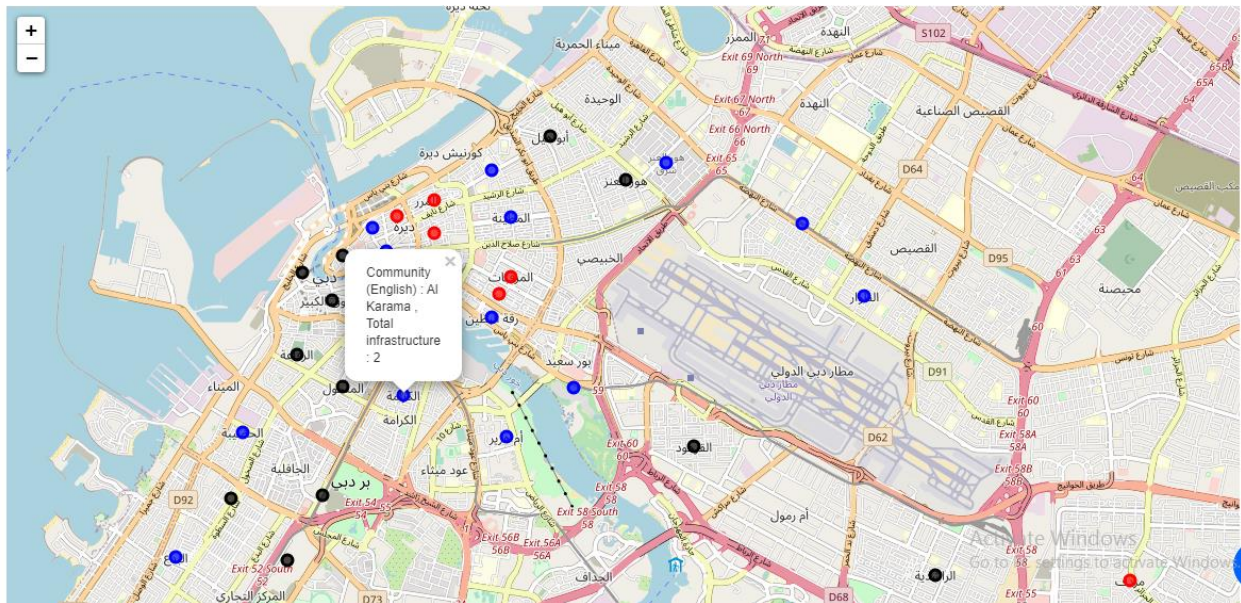
## Best place to stay within a city for vital infrastructure facilities

**Step 1.** Defined 'Vital infrastructure' like hospital, food, park etc. and displayed communities that had max number of vital infrastructures

Community (English)		Total infrastructure
3	Al Buteen	3
13	Al Muraqqabat	3
14	Al Murar	3
29	Ayal Nasir	3
38	Mirdif	3
40	Naif	3

Clustered the data set to find out which place would be the best to stay in considering all the above observations

**Step 1:** Set number of clusters for 'Total Infrastructure' column and perform kmeans Clustering on them and visualize the dataset.



## Observations

1. Cluster red shows the community with the greatest number of infrastructures
2. Cluster blue shows communities with medium number of infrastructures
3. Cluster black shows communities with low number of total infrastructures

Most of the infrastructures are concentrated in the western areas of Dubai city, with the highest number in cluster 1 and moderate number in cluster 3. On the other hand, cluster 2 has a very low number of infrastructures in the community. This represents a great opportunity and high potential areas to open new infrastructures as it is very little to no competition from existing varied infrastructures. Meanwhile, one can specifically check the infrastructure of choice against the Community area of our choice.

