

Capstone Project - The Battle of the Neighborhoods (Week 2)

IBM Data Science Professional

Opening a new Business Park in Dubai, UAE

By: Tushar Tyagi

June 2020



Table of Contents

1.	Introduction.....	3
2.	Business Problem.....	3
3.	Target Audience of this Project	3
4.	Data	4
5.	Methodology	6
6.	Results.....	10
7.	Discussion Section	13
8.	Conclusion	13

1. Introduction

Dubai's real estate market started in 2002 when the government of Dubai declared freehold rights for owning property for non-nationals. At the time according to the Dubai Statistics Centre the population of Dubai was 1,010,751, 75% of which were ex-pat/non-nationals.

From 2003 Dubai's property supply accelerated and spread across the Emirate but with the onset of the financial crisis of 2007-2010, Dubai's real estate market declined, after enjoying an uninterrupted boom since 2002.

After a remarkable recovery in 2012, the real estate of Dubai is set for bigger and better performance in 2013. With many huge projects planned for the present year and large supply of residential and corporate units to serve the increasing interests of investors, Dubai has already started showing signs of healthy growth.

At the present time, the UAE implements a legal framework of free zones which foster an attractive environment for businesses by offering companies primarily 100% foreign-owned companies incentives such as zero tax rates on their income and exemption from foreign exchange controls. These incentives have encouraged foreign companies to set up their business in Dubai, leading to an ever-growing demand of Business Parks and new offices and this demand has attracted a lot of real-estate developer and investors to invent in new Business Park project to meet those demands. For examples, The Business Park at Dubai South recorded a 400% growth in companies operating in the park since its inception in early 2015 with the number of companies increasing from 900 in 2015 to 4,500 in 2018. An average of 1,500 new companies has joined the park each year.

2. Business Problem

The objective of this capstone project is to analyze and find the best locations to open a Business Park in the city of Dubai, UAE. By applying Data Science methodology and Machine Learning Techniques like k-means clustering, this project aims to provide an answer to the following question:

In the city of Dubai, UAE, if a real estate developer or an investor is looking to open a new Business Park, where would you recommend that they open it?

3. Target Audience of this Project

This project is specifically targeted towards real estate developers and investors, looking for an opportunity to start and invest in new Business Park Project in the city of Dubai, UAE. Real estate is one of the biggest industries in Dubai, with a huge potential for more development in an ever-growing city of Dubai. This project serves to recommend a location or a community for a new Business Park Project, with a lot of companies and businesses moving to this part of the Middle East. In recent years, Dubai has been a major attraction to new business and there is an increasing demand for offices and workplace in Dubai, due to major benefits in terms of economic growth,

low taxes, huge investment from local government, proximity to international markets etc. Real estate industry has surged year after year in Dubai and in 2018, real estate contributed nearly 13.6% to the total GDP. Considering the aforementioned benefits for the businesses and companies in the area, Dubai seems to be a hotbed for huge investment in the field of real estate, and new real estate developers and investors can leverage this project to find a potential investing opportunity.

4. Data

To solve the problem at hand, we will need to following data:

- List of communities in Dubai, UAE. This data defines and focuses the scope of this project to the city of Dubai, UAE .
- Coordinates (Latitude and Longitude) of the communities in Dubai, UAE. This information is required to create the map to plot communities in Dubai and to obtain the venue data for a particular community.
- Venue data of all communities in Dubai, obtained by using Foursquare API will be used to perform clustering on the communities.

Data sources and Data extraction methods:

This Wikipedia page for list of communities in Dubai, United Arab Emirates (https://en.wikipedia.org/wiki/List_of_communities_in_Dubai) contains a list of all communities in the city of Dubai, with a total of 131 communities. Web scraping technique will be used to extract the data from the Wikipedia page and import the same into a *pandas* DataFrame.

Community Number	Community (English)	Community (Arabic)	Area(km ²)	Population(2000)	Population density(/km ²)
126	Abu Hail	أبو هيل	1.27 km ²	21,414	16,861.4/km ²
711	Al Awir First	العوير الأولى			
721	Al Awir Second	العوير الثانية			
333	Al Bada	البدع	0.82 km ²	18,816	22946/km ²
122	Al Baraha	البراحة	1.104 km ²	7,823	7,086/km ²
373	Al Barsha First	البرشاء الأولى			
376	Al Barsha Second	البرشاء الثانية			
671	Al Barsha South First	البرشاء جنوب الاولى			
672	Al Barsha South Second	البرشاء جنوب الثانية			
673	Al Barsha South Third	البرشاء جنوب الثالثة			
375	Al Barsha Third	البرشاء الثالثة			
114	Al Buteen	البطين	0.07 km ²	2,364	33,771/km ²
113	Al Dhagaya	الضغاية	0.125 km ²	10,896	21,451/km ²
214	Al Garhoud	القرهود	4 km ²	4,466	1,116.5/km ²

Then, we will extract the geographical coordinates of the all the communities of Dubai using Python *geocoder* package in *geopy* library. Coordinates (latitude and longitude) of all the communities will be extracted and stored in a separate DataFrame.

	Latitude	Longitude
0	25.28308	55.33435
1	25.18593	55.54126
2	25.16792	55.54331
3	25.21861	55.26406
4	25.28280	55.31678

Later, both the DataFrames will be consolidated to make a single DataFrame that stores both community names and coordinates.

	Community	Latitude	Longitude
0	Abu Hail	25.283080	55.334350
1	Al Awir First	25.185930	55.541260
2	Al Awir Second	25.167920	55.543310
3	Al Bada	25.218610	55.264060
4	Al Baraha	25.282800	55.316780
5	Al Barsha First	25.114830	55.191360
6	Al Barsha Second	25.107230	55.204850
7	Al Barsha South First	25.089580	55.234240
8	Al Barsha South Second	25.077390	55.242670
9	Al Barsha South Third	25.062290	55.239950
10	Al Barsha Third	25.093420	55.190440

Post that, *Foursquare API* will be used to extract the venue data for those communities. Foursquare API will provide many categories of the venue data, we are particularly interested in the Business Parks and overall profile of a community to help in identifying a cluster of communities, that will attract potential businesses, real estate developers and investors to start a project in the identified communities.

	Community	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Abu Hail	25.28308	55.33435	Habib Bakery	25.281124	55.332774	Bakery
1	Abu Hail	25.28308	55.33435	Mubasher Cafeteria	25.282622	55.336252	Cafeteria
2	Abu Hail	25.28308	55.33435	Pizza Hut	25.282107	55.335684	Pizza Place
3	Abu Hail	25.28308	55.33435	Bait Al Jinnie Junction	25.280546	55.330471	Scenic Lookout
4	Abu Hail	25.28308	55.33435	Hamriya Park	25.285710	55.333000	Park

This project will make use of many data science skills and methodologies such as web scraping from Wikipedia, geocoding to obtain coordinates, Foursquare API, data cleaning and data wrangling, machine learning techniques (K-means clustering) and data visualization techniques to create map using *Folium*.

5. Methodology

- a) First, we need to obtain the list of communities in the city of Dubai, UAE. The list of communities in Dubai, United Arab Emirates is available on Wikipedia page (https://en.wikipedia.org/wiki/List_of_communities_in_Dubai).

Community Number	Community (English)	Community (Arabic)	Area(km ²)	Population(2000)	Population density(/km ²)
126	Abu Hail	أبو هيل	1.27 km ²	21,414	16,861.4/km ²
711	Al Awir First	العوير الأولى			
721	Al Awir Second	العوير الثانية			
333	Al Bada	البدع	0.82 km ²	18,816	22946/km ²
122	Al Baraha	البراحة	1.104 km ²	7,823	7,086/km ²
373	Al Barsha First	البرشاء الأولى			
376	Al Barsha Second	البرشاء الثانية			
671	Al Barsha South First	البرشاء جنوب الاولى			
672	Al Barsha South Second	البرشاء جنوب الثانية			
673	Al Barsha South Third	البرشاء جنوب الثالثة			
375	Al Barsha Third	البرشاء الثالثة			
114	Al Buteen	البطين	0.07 km ²	2,364	33,771/km ²
113	Al Dhagaya	الذغاية	0.125 km ²	10,896	21,451/km ²
214	Al Garhoud	القرهود	4 km ²	4,466	1,116.5/km ²

- b) We performed Web scraping technique to extract the data from the Wikipedia page and import the same into a pandas DataFrame. The DataFrame consisted of a total of 131 communities.

	Community_no	Community	Community_arabic	Area	Population	Density
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
5	373.0	Al Barsha First	البرشاء الأولى	NaN	NaN	NaN
6	376.0	Al Barsha Second	البرشاء الثانية	NaN	NaN	NaN
7	671.0	Al Barsha South First	البرشاء جنوب الأولى	NaN	NaN	NaN

- c) The original table of Wikipedia's page had many columns describing such as Community_no, Community_arabic, Area, Population, Density etc. These columns are neither required for our analysis, nor for the clustering model, and are out of the scope of study of this project. Therefore, these columns were ignored, and we only stayed with the Community columns, given that we only needed to know the names of communities in Dubai, UAE.

	Community
0	Abu Hail
1	Al Awir First
2	Al Awir Second
3	Al Bada
4	Al Baraha
5	Al Barsha First
6	Al Barsha Second
7	Al Barsha South First

- d) Later, we extracted the geographical coordinates of these communities in the form of latitude and longitude, to be able to perform further analysis of the venues of these communities using Foursquare API. The latitudes and longitudes of the communities were extracted in a separate DataFrame.

To do so, we used Python *geocoder* package in geopy library. Coordinates (latitude and longitude) of all the communities will be extracted and stored in a separate DataFrame. The geocoder package allows us to convert addresses into geographical coordinates, in the form of latitudes and longitudes. After extracting the data, we populate the data into a pandas DataFrame.

	Latitude	Longitude
0	25.28308	55.33435
1	25.18593	55.54126
2	25.16792	55.54331
3	25.21861	55.26406
4	25.28280	55.31678

- e) Then we consolidated the two DataFrames to make a single DataFrame that stored both community names and coordinates.

	Community	Latitude	Longitude
0	Abu Hail	25.283080	55.334350
1	Al Awir First	25.185930	55.541260
2	Al Awir Second	25.167920	55.543310
3	Al Bada	25.218610	55.264060
4	Al Baraha	25.282800	55.316780
5	Al Barsha First	25.114830	55.191360
6	Al Barsha Second	25.107230	55.204850
7	Al Barsha South First	25.089580	55.234240
8	Al Barsha South Second	25.077390	55.242670
9	Al Barsha South Third	25.062290	55.239950
10	Al Barsha Third	25.093420	55.190440

- f) Then we visualized the communities of Dubai by creating a map using Folium library in python. This allowed us to verify that the geographical coordinates data returned by geocoder were correctly plotted for the communities in the city of Dubai, UAE.



- g) Next, we used Foursquare API to get the top 100 venues that are within a radius of 500 meters for each community of Dubai. We made API calls to Foursquare, by passing in the geographical coordinates of the communities in a loop. Foursquare API returned the top

100 venues data for each community in JSON format and later, we extracted the venue name, venue category, venue latitude and longitude data into a pandas DataFrame.

	Community	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Abu Hail	25.28308	55.33435	Habib Bakery	25.281124	55.332774	Bakery
1	Abu Hail	25.28308	55.33435	Mubasher Cafeteria	25.282622	55.336252	Cafeteria
2	Abu Hail	25.28308	55.33435	Pizza Hut	25.282107	55.335684	Pizza Place
3	Abu Hail	25.28308	55.33435	Bait Al Jinnie Junction	25.280546	55.330471	Scenic Lookout
4	Abu Hail	25.28308	55.33435	Hamriya Park	25.285710	55.333000	Park

No data was returned for 12 communities as there were no venues in those communities, due to their topography and geography. Cluster for these venues was handled separately later.

- h) With the data, we checked how many venues were returned for each community and examined how many unique categories existed from all the returned venues.

Community	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
Abu Hail	6	6	6	6	6	6
Al Awir First	1	1	1	1	1	1
Al Bada	11	11	11	11	11	11
Al Baraha	9	9	9	9	9	9
Al Barsha First	57	57	57	57	57	57
Al Barsha Second	24	24	24	24	24	24
Total Unique Categories in Dubai communities						

```
In [25]: print('There are total {} uniques categories.'.format(len(dubai_venues['VenueCategory'].unique())))
```

There are total 241 uniques categories.

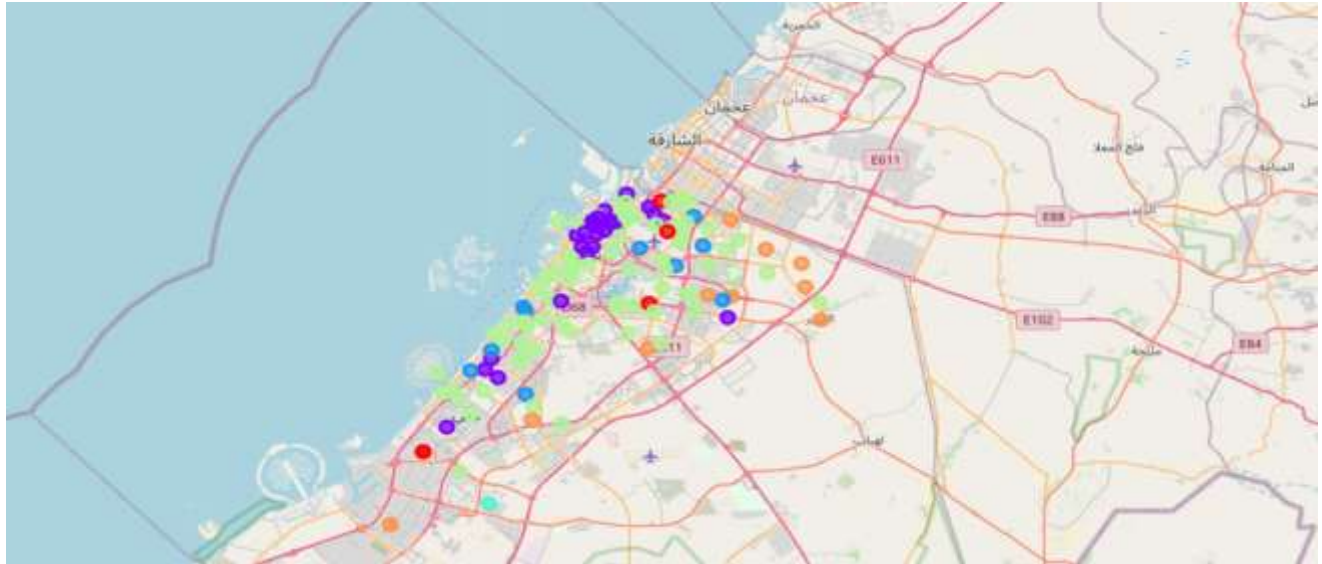
- i) Then, we analyzed each community by grouping the rows by community and taking the mean of the frequency of occurrence of each venue category. By doing so, we prepared the data to be used later in clustering.

Community	ATM	Accessories Store	Afghan Restaurant	African Restaurant	Airport Terminal	American Restaurant	Amphitheater	Antique Shop	Arcade	Art Gallery	Arts & Crafts Store	Asian Restaurant	Astrologer	Athletics & Sports	Australian Restaurant	Auto Garage	Auto Workshop	Automotive
0 Abu Hail	0.0000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000	0.00	0.00	0.00
1 Al Awir First	0.0000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000	0.00	0.00	0.00
2 Al Bada	0.0000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000	0.00	0.00	0.00
3 Al Baraha	0.0000	0.000000	0.000000	0.000000	0.00	0.111111	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000	0.00	0.00	0.00
4 Al Barsha First	0.0000	0.000000	0.000000	0.000000	0.00	0.017544	0.000000	0.0000	0.000000	0.000000	0.000000	0.017544	0.000000	0.017544	0.0000	0.00	0.00	0.00
5 Al Barsha Second	0.0000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.041667	0.0000	0.00	0.00	0.00

- j) At lastly, we performed clustering on the data by using K-means clustering. K-means clustering algorithm identified k number of centroids, and then allocated every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the

simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project.

We clustered the communities into 5 clusters and assigned a 6th cluster to the communities that returned no venues. Based on the categories of venues and the profile of each cluster, we identified the cluster of communities, more suitable for opening new Business Parks.



6. Results

The results from the k-means clustering shows that we can categorize the communities into 6 clusters based on the frequency of occurrence of various venues in a community:

- **Cluster 0:** Cluster 0 consists of communities with outdoor venues such as Football Stadiums and small-scale food joints such as Fast Food Restaurants, Fried Chicken Joints, Food Courts and Food & Drink Shops. Despite of having no Business Park and centers in this clusters, these communities do not have much to offer to the new businesses in terms of commercial value. Also, real estate developers will need to invest in enhancing infrastructure in these communities, which will be an additional overhead and cost.

Cluster 1

```

In [47]: dubai_merged.loc[dubai_merged['Cluster_Labels'] == 0, dubai_merged.columns[[0] + list(range(1, dubai_merged.shape[1]))]]
Out[47]:

```

	Community	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
29	Al Nahla Second	Restaurant	Pakistani Restaurant	Market	Yoga Studio	Fast Food Restaurant	French Restaurant	Football Stadium	Food Court	Food & Drink Shop
67	Al Tair Second	Basketball Court	Yoga Studio	Fast Food Restaurant	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Football Stadium	Food Court	Food & Drink Shop
88	Jebel Ali 1	Yoga Studio	Fast Food Restaurant	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Football Stadium	Food Court	Food & Drink Shop	Food
90	Jebel Ali 2	Yoga Studio	Fast Food Restaurant	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Football Stadium	Food Court	Food & Drink Shop	Food
114	Ras Al Khair Industrial Second	Coffee Shop	Fast Food Restaurant	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Football Stadium	Food Court	Food & Drink Shop	Food

Community Characteristics:
Most common venues in this cluster are Football Stadiums and small scale food joints such as Fast Food Restaurants, Fried Chicken Joints, Food Courts and Food & Drink Shops.

- **Cluster 1:** Cluster 1 is very commercial and mostly concentrated towards North Dubai. It mostly consists of a huge variety of Restaurants, Hotels, Pubs, shops, cafes, restaurants, bars and pubs. Cluster 1 is very commercial, with a lot of hotels, cafés and restaurants in the communities to attract businesses.

Cluster 1

```
In [48]: dubai_merged.loc[dubai_merged['cluster_labels'] == 1, dubai_merged.columns[[9] + list(range(9, dubai_merged.shape[1]))]]
```

Out[48]:

	Community	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
4	Al Barsha	Indian Restaurant	Turkish Restaurant	Cafe	Middle Eastern Restaurant	Lounge	Fast Food Restaurant	American Restaurant	Fast Court	Football Stadium
6	Al Barsha First	Food & Breakfast	Pakistani Restaurant	Park	Indian Restaurant	Middle Eastern Restaurant	Restaurant	Grocery Store	Desert Shop	Turkish Restaurant
8	Al Barsha Second	Indian Restaurant	Coffee Shop	Ice Cream Shop	Park	Chinese Restaurant	Bakery	Tennis Court	Seafood Restaurant	Dyn
14	Al Gurf Fourth	Middle Eastern Restaurant	Fast Food Restaurant	Asian Restaurant	Restaurant	Persian Restaurant	Plaza	Chinese Restaurant	Flea Market	Don Bury Restaurant
20	Al Karama	Restaurant	Pakistani Restaurant	Asian Restaurant	Fast Food Restaurant	Bakery	Chinese Restaurant	Ice Cream Shop	Middle Eastern Restaurant	European Restaurant
24	Al Kobar	Middle Eastern Restaurant	Hotel	Track	Department Store	North Indian Restaurant	Tee Room	Clothing Store	Asian Restaurant	Park
25	Al Mankar	Restaurant	Farmers Market	Fried Chicken Joint	French Restaurant	Football Stadium	Food Court	Food & Drink Shop	Food	Pondue Restaurant
27	Al Mankool	Park	Department Store	Filipino Restaurant	Chinese Restaurant	Fast Food Restaurant	Fondue Restaurant	Flea Market	Flea Market	Flower Shop
32	Al Mankool	Middle Eastern Restaurant	Chinese Restaurant	Indian Restaurant	Filipino Restaurant	Grocery Store	Asian Restaurant	Cafe	Hookah Bar	Spa
34	Al Mankool	Hotel	Fried Chicken Joint	Restaurant	Asian Restaurant	Ice Cream Shop	Middle Eastern Restaurant	Chinese Restaurant	Pakistani Restaurant	Marital Arts Gym
35	Al Mankool	Cafe	Asian Restaurant	Chinese Restaurant	Bar	Hookah Bar	Ice Cream Shop	Filipino Restaurant	Hong Kong Restaurant	Persian Restaurant
36	Al Nakha First	Women's Store	Hotel	BBQ Joint	Hookah Bar	Food Court	Football Stadium	Food & Drink Shop	Fast Food Restaurant	Food
38	Al Nakha Dubai	Pizza Place	Cafe	Women's Store	Asian Restaurant	Middle Eastern Restaurant	Intersection	Smoke Shop	Mobile Phone Shop	Filipino Restaurant

- **Cluster 2:** Cluster 2 is slightly away from the coast and a bit in the interior of Dubai. It mostly consists of small-scale businesses such as Grocery Stores, Flea Markets, Fish Markets, Small Restaurants. Such communities might seem unattractive to new businesses and thus real estate developers will not be interested in starting new Business Park in these areas.

Cluster 2

```
In [49]: dubai_merged.loc[dubai_merged['cluster_labels'] == 2, dubai_merged.columns[[9] + list(range(9, dubai_merged.shape[1]))]]
```

Out[49]:

	Community	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
7	Al Barsha South East	Pharmacy	Grocery Store	Yoga Studio	Fast Food Restaurant	Fried Chicken Joint	French Restaurant	Football Stadium	Food Court	Food & Drink Shop
13	Al Garhoud	Park	Grocery Store	Carney Exchange	African Restaurant	Filipino Restaurant	Food & Vegetable Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant
56	Al Rashidiya	Fried Chicken Joint	Burger Joint	Bus Station	Asian Restaurant	Cake/Bakery	Cafe	Yoga Studio	Flea Market	French Restaurant
89	Al Safa First	Ice Cream Shop	Tennis Court	Video Store	Italian Restaurant	Grocery Store	Flower Shop	Filipino Restaurant	Flea Market	Flea Market
85	Al Safa First	Grocery Store	Playground	Playground	Yoga Studio	Filipino Restaurant	Fish Market	Flower Shop	Flower Shop	Fondue Restaurant
71	Al Warqaa Fourth	Grocery Store	Yoga Studio	Filipino Restaurant	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Football Stadium	Food Court	Food & Drink Shop
86	Jumeira Third	Cafe	Ice Cream Shop	Tennis Court	Video Store	Italian Restaurant	Flower Shop	Filipino Restaurant	Flea Market	Flea Market
89	Mohammed Fourth	Grocery Store	Dyn	Yoga Studio	Fast Food Restaurant	Fried Chicken Joint	French Restaurant	Football Stadium	Food Court	Food & Drink Shop
102	Mohammed First	Hotel	Grocery Store	Yoga Studio	Farmers Market	French Restaurant	Football Stadium	Food Court	Food & Drink Shop	Food
125	Umm Suqeim Third	Fast Food Restaurant	Spa	Breakfast Spot	Diner	Racetrack	Pharmacy	French Restaurant	Sandwich Place	Bakery

- **Cluster 3:** Cluster 3 is very far from other clusters and lies far inland from a coast or border of Dubai. It consists of just 1 community which is an Investment Park with a few restaurants. This community being similar to the Business Park can provide competition to new projects in attracting new businesses. Also, this cluster being far inland from a coast or border of Dubai, wouldn't offer better connectivity in terms of transportation.

Cluster 3

```
In [58]: dubai_merged.loc[dubai_merged['Cluster_Labels'] == 3, dubai_merged.columns[[0] + list(range(5, dubai_merged.shape[1]))]]
```

```
Out[58]:
```

	Community	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
82	Dubai Investment Park Second	Fruit & Vegetable Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Football Stadium	Food Court	Food & Drink Shop	Food	Fondue Restaurant

Community Characteristics:

Cluster 3 is very far from other cluster and lies far inland from a coast or border of Dubai. It consists of an investment Park with a few restaurants.

- Cluster 4:** Cluster 4 is the biggest, very urban and the most commercial area of Dubai with high end Restaurants, Shopping Malls, Hotels, Yoga Studios, Bars, Cafes. Cluster 4 is very commercial, with a lot of attractive venues and will be an ideal location for potential projects for Business Park development. Also, this cluster being in the heart of the city will provide high commercial value to the real estate developers and investors.

Cluster 4

```
dubai_merged.loc[dubai_merged['Cluster_Labels'] == 4, dubai_merged.columns[[0] + list(range(5, dubai_merged.shape[1]))]]
```

	Community	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	Salad Lookout	Cafeteria	Bakery	Pizza Place	Hookah Bar	Flower Shop	Fish Market	Flea Market	Food
1	Al Ayn First	Yoga Studio	Dance Studio	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Football Stadium	Food Court	Food & Drink Shop	Food
3	Al Bada	Coffee Shop	Gym / Fitness Center	Restaurant	Beach	Shopping Plaza	Bakery	Boulogne	Electronics Store	English Restaurant
8	Al Barsha South Second	Supermarket	Office	Gym	Yoga Studio	French Restaurant	Football Stadium	Food Court	Food & Drink Shop	Food
10	Al Barsha Third	Art Gallery	Bakery	Convenience Store	Restaurant	Flea Market	Flower Shop	Fondue Restaurant	Food	
11	Al Bustan	Shopping Mall	Jewelry Store	Art Gallery	Fast Food Restaurant	Pakistani Restaurant	Farmer's Market	Flower Shop	Electronics Store	Juice Bar
12	Al Dhagaya	Jewelry Store	Afghan Restaurant	Shopping Mall	Currency Exchange	Electronics Store	Pakistani Restaurant	Clothing Store	Fast Food Restaurant	Flea Market
15	Al Hamriya, Dubai	Coffee Shop	Cafe	Indian Restaurant	Asian Restaurant	Middle Eastern Restaurant	Fast Food Restaurant	Pizza Place	Shopping Mall	Persian Restaurant
16	Al Hamriya Port	Arcade	Cafeteria	Yoga Studio	Filipino Restaurant	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Football Stadium	Food Court

Community Characteristics:

Cluster 4 is the biggest, very urban and the most commercial area of Dubai with high end Restaurants, Shopping Malls, Yoga Studios, Bars, Cafes.

- Cluster 5:** Cluster 5 consists of remote areas of Dubai with no venues or attractions and mostly consists of areas in desert. This cluster being remote and isolated from the main city has almost nothing to offer to potential Business Park projects. Also, establishing Business Park in such communities will result in heavy investments in infrastructure, which could repel few real estate developers from these communities. Moreover, connectivity to the communities is also very poor.

Cluster 5

```
dubai_merged.loc[dubai_merged['Cluster_Labels'] == 5, dubai_merged.columns[[0] + list(range(5, dubai_merged.shape[1]))]]
```

```
Out[ ]:
```

	Community	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
2	Al Ayn Second	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	Al Barsha South Third	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22	Al Khumsan First	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
23	Al Khumsan Second	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
44	Al Warqaa Fifth	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
73	Al Warqaa Third	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
77	Alayna	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
81	Jebel Ali Industrial	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
82	Jebel Ali Palm	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
104	Nad Al Sheba Fourth	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
111	Thal Al Muharikah Third	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Community Characteristics:

Cluster 5 consists of remote areas of Dubai with no venues or attractions and mostly consists of areas in desert.

7. Discussion Section

We have thoroughly analyzed the communities of Dubai using the data, we are now able to tell which areas are more suitable for establishment of a new Businesses Park project.

We can study the profile of each cluster and based on the value those communities offer to new businesses, we can decide the potential locations and rule out certain communities for building new Business Parks.

As observed from the clusters and the map in results section, based on commercial values, Cluster 4 and Cluster 1 (in order) are most suitable for new Business Park projects as they have a lot to offer to new businesses in terms of infrastructure, venues and commercial value. Therefore, this project recommends real estate developers to capitalize on these findings to open new Business Park malls in communities in Cluster 4 and Cluster 1.

Property developers with heavy financial backing can explore the options of developing Business Park in communities under Cluster 0 and Cluster 2. These communities will be cheaper in terms of property and offer little of no competition but will incur decent investment in infrastructure.

Lastly, real estate developers are advised to avoid communities in Cluster 5 as these communities not only have nothing to offer to new Businesses, but also will lead to a heavy investment in infrastructure as these communities lie in remote deserts of Dubai.

8. Conclusion

In this project, we went through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 6 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. real estate developers and investors regarding the best locations to open a new Business Park. To answer the business question that was raised in the introduction section, the answer proposed by this project is: The communities in Cluster 4 and Cluster1 are the most preferred locations to open a new Business Park. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding the areas with little commercial value, while making decisions to open a new Business Park.

Thank You!