Capstone Project - The Battle of the Neighborhoods (Week 1)

IBM Data Science Professional

Opening a new Business Park in Dubai, UAE

By: Tushar Tyagi

June 2020

Table of Contents

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 Bangalore, 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 km²	21,414	16,861.4/km²
711	Al Awir First	العوير الأولى			
721	Al Awir Second	العوير الثانية			
333	Al Bada	البدع	0.82 km²	18,816	22946/km²
122	Al Baraha	البراحة	1.104 km²	7,823	7,086/km²
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 km²	2,364	33,771/km²
113	Al Dhagaya	الضغاية	0.125 km²	10,896	21,451/km²
214	Al Garhoud	القرهود	4 km²	4,466	1,116.5/km²

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***.