

Tuesday, December 4, 2018

The Battle of Neighborhoods

Coursera Capstone Project for IBM Data Science Certificate

Introduction

The United Arab Emirates (UAE; Arabic: دولة الإمارات العربية المتحدة), sometimes simply called the Emirates, is the most modern and advance countries across the Middle East region often referred to as the Business Hub of Middle East. The UAE is a federation of seven emirates consisting of Abu Dhabi, Ajman, Dubai, Fujairah, Ras al-Khaimah, Sharjah and Umm al-Quwain.

Dubai is the largest and most populous city in the United Arab Emirates (hereinafter referred to as UAE). The UAE is the most modern and advanced countries across the Middle East region and is also the central business hub location for local, regional and global businesses. Dubai is one of the 7 states in the UAE and welcomes over 30000 passengers each day through its state of the art world class airports and hosts over 60% of the entire expatriate population across the country.

Objectives

Being a very busy city, one of the most challenging aspects is to find the best neighborhood in Dubai. A neighborhood that is near the work location, within the budget, and offers quality facilities. Our project is envisaged to analyze and classify Sectors of Dubai by leveraging Foursquare APIs and Machine Learning techniques such as Segmentation and Clustering. The final objective of this project is to Segment/Classify Sectors of Dubai based on Common Places captured from Foursquare APIs.

- Similarities and Dissimilarities between the Sectors (Sector 2 and Sector 3) of Dubai
- Classification of Areas by attractions for Residents, Business Owners and Tourists

Target Audience

- Tourists
- New Migrants
 - Short Term Visitors
 - Long Term Resident Migrants
- Business Owners Looking for Locations
- Locations for New Restaurant Openings

Workflow

- Prepare and Preprocess Data
- Integration of Data with Geocoder
- Clustering of Sectors and Map Visualization
- Foursquare APIs Integration and Venue Categorization
- Analysis of Sectors
- K-Means Clustering and Common Place Categorization

Data

The Data has been acquired through Wikipedia and maintained in a CSV format at GitHub. The dataset contains 9 Sectors and 225 Communities. Click [here](#) for GitHub repository.

CommunityCode	Emirate	Sectors	Population	Area	AreaArabic
101	Dubai	Sector 1	2	NAKHLAT DEIRA	نخلة ديرة
111	Dubai	Sector 1	1550	AL CORNICHE	الكورنيش
112	Dubai	Sector 1	8821	AL RASS	الراس
113	Dubai	Sector 1	14983	AL DHAGAYA	الضفة
114	Dubai	Sector 1	2583	AL BUTEEN	البتين

Figure 1 - Data Snapshot

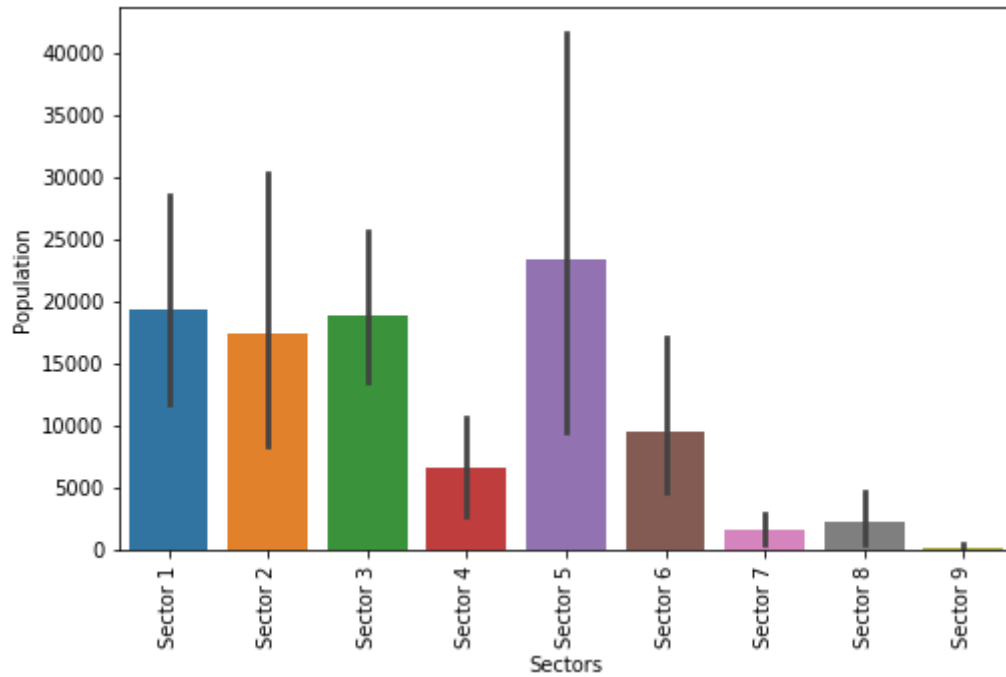


Figure 2 - Population by Sectors

Integration of Longitude and Latitude

Upon integrating the Location data with Geocoder, we obtained the Longitude and Latitude of each location in our dataset as shown in the snapshot below:

CommunityCode	Emirate	Sectors	Population	Area	AreaArabic	Latitude	Longitude
101	Dubai	Sector 1	2	NAKHLAT DEIRA	نخلة ديرة	25.38	55.39
111	Dubai	Sector 1	1550	AL CORNICHE	الكورنيش	25.28	55.33
112	Dubai	Sector 1	6621	AL RASS	الراس	25.27	55.30
113	Dubai	Sector 1	14963	AL DHAGAYA	الضفة	25.27	55.30
114	Dubai	Sector 1	2563	AL BUTEEN	الطين	25.26	55.32

Figure 3 - Longitude and Latitude

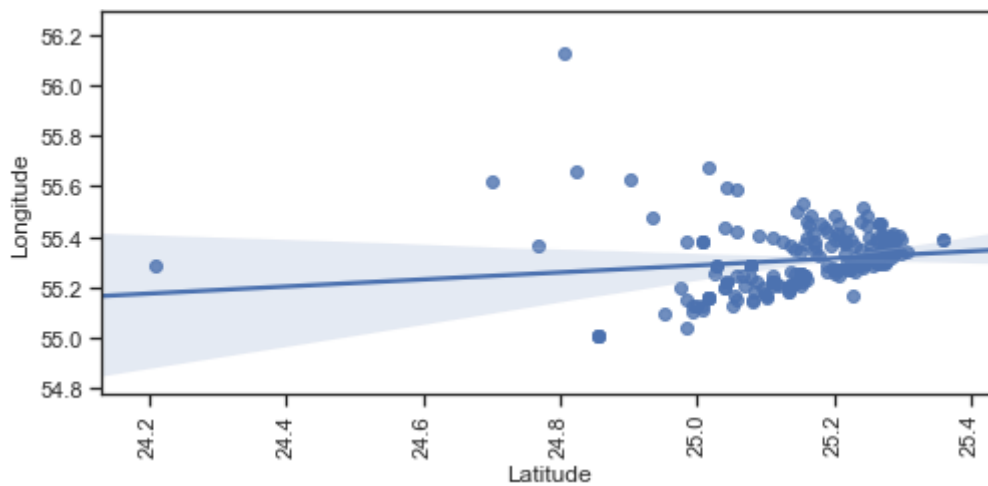


Figure 4 - By Longitude and Latitude



Figure 5 - Folium Map with Locations

Our Methodology

We will use Foursquare APIs to explore the Neighborhoods, Identify Most Common Venues and build Clusters by utilizing K-Means Clustering Algorithm. For this, we will filter Sector 2 and 3 of Dubai.



Figure 6 - Sector 2 Map

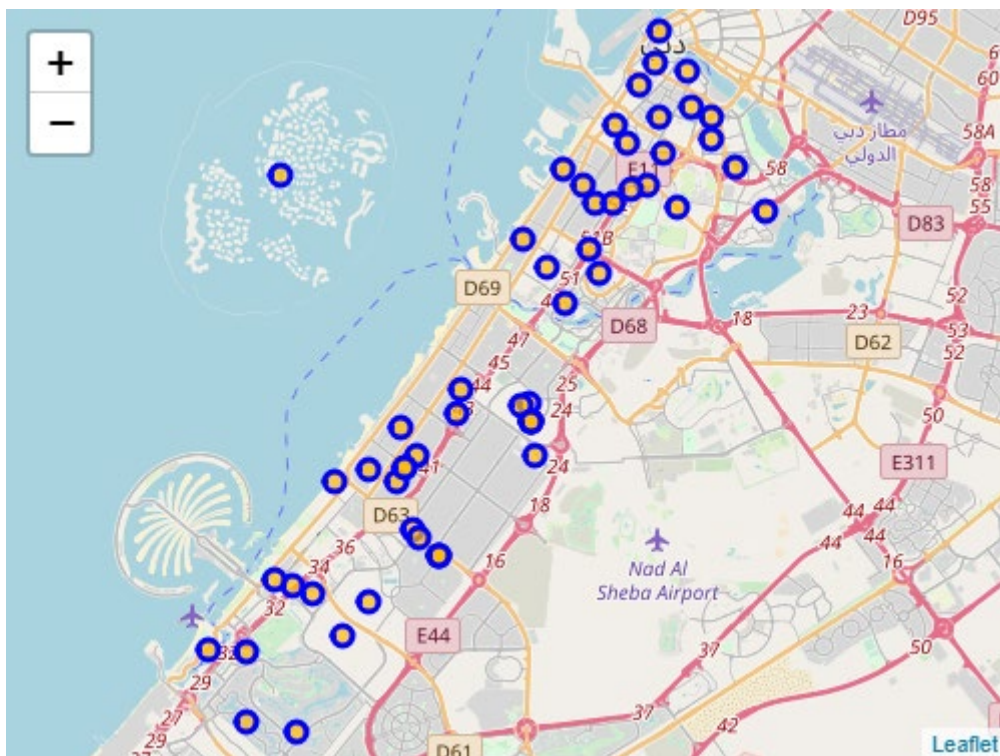


Figure 7 - Sector 3 Map

There are 77 unique categories in Sector 2.

Area	Venue	Venue Latitude	Venue Longitude	Venue Category
NAD SHAMMA	Subway	25.21	55.39	Sandwich Place
NAD SHAMMA	Al Maaref Sports Hall	25.22	55.39	Badminton Court
NAD SHAMMA	كافتريا كوخ النجاج	25.22	55.39	Burger Joint
NAD SHAMMA	Jumeriah Zabeel Saray (Dubai, United Arab Emir...	25.21	55.38	Hotel
NAD SHAMMA	Shaikh's Airport (مطار الشيوخ)	25.21	55.39	Airport

Figure 8 - Sector 2 Venues

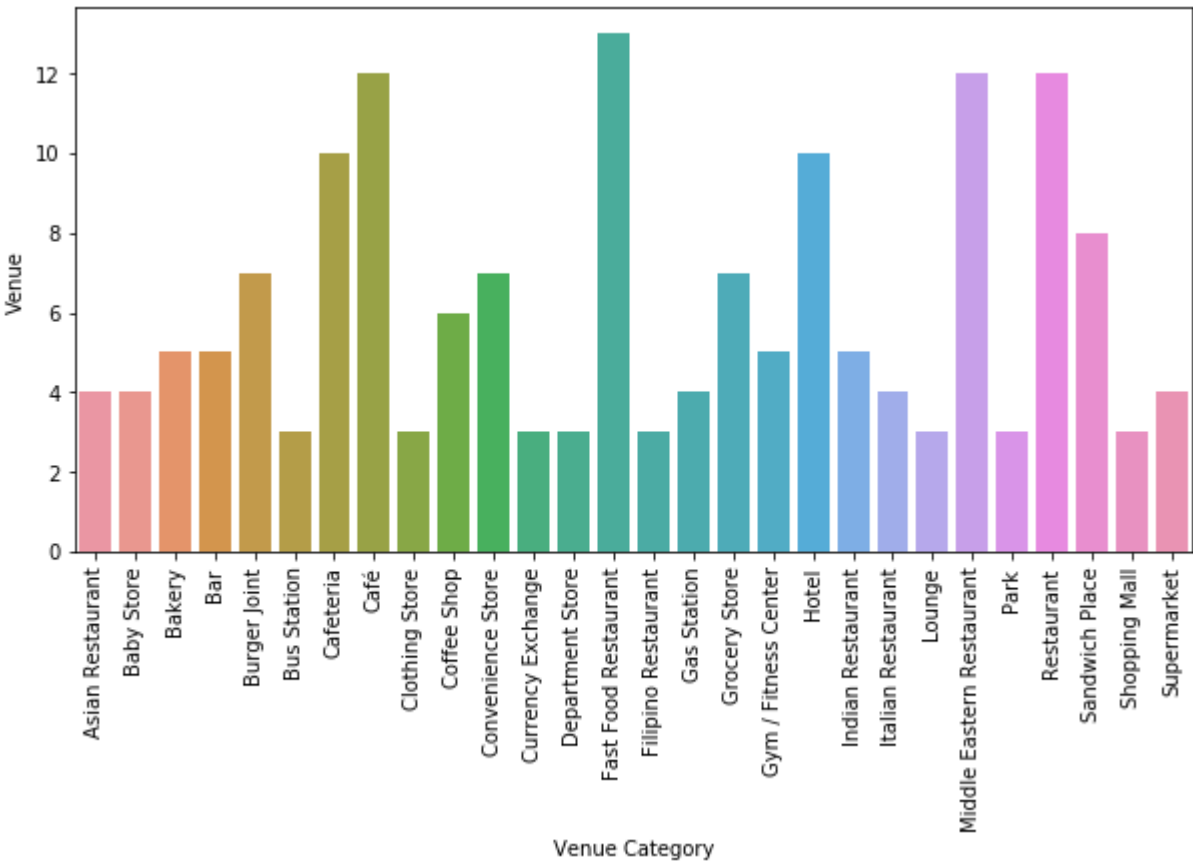
There are 204 unique categories in Sector 3.

Area	Venue	Venue Latitude	Venue Longitude	Venue Category
JUMEIRA BAY	Public Cafe	25.08	55.15	Café
JUMEIRA BAY	Pullman Dubai Jumeirah Lakes Towers-Hotel and ...	25.08	55.15	Hotel
JUMEIRA BAY	Le Petit Belge	25.08	55.15	Belgian Restaurant
JUMEIRA BAY	Friends Avenue Cafe JLT	25.08	55.15	Café
JUMEIRA BAY	The White Room Spa	25.08	55.15	Spa

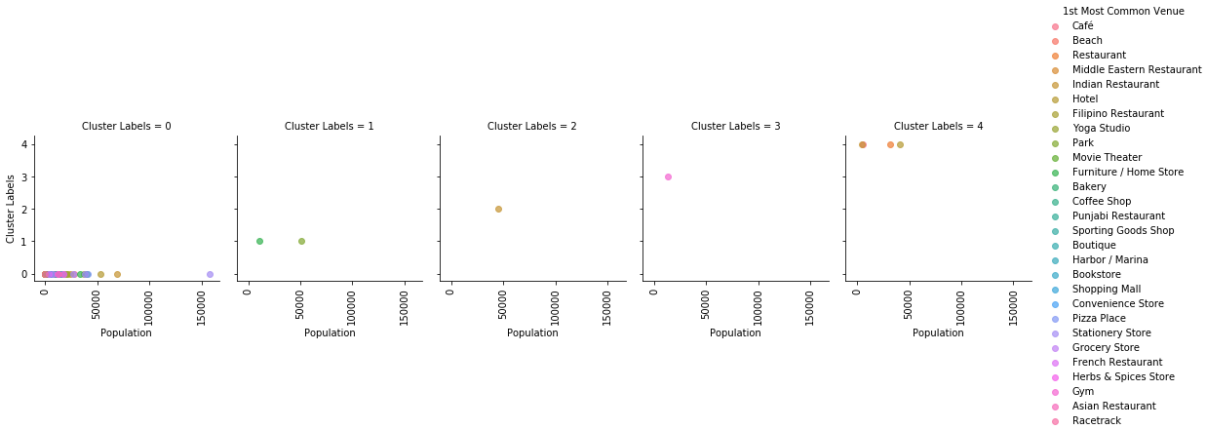
Figure 9 - Sector 3 Venues

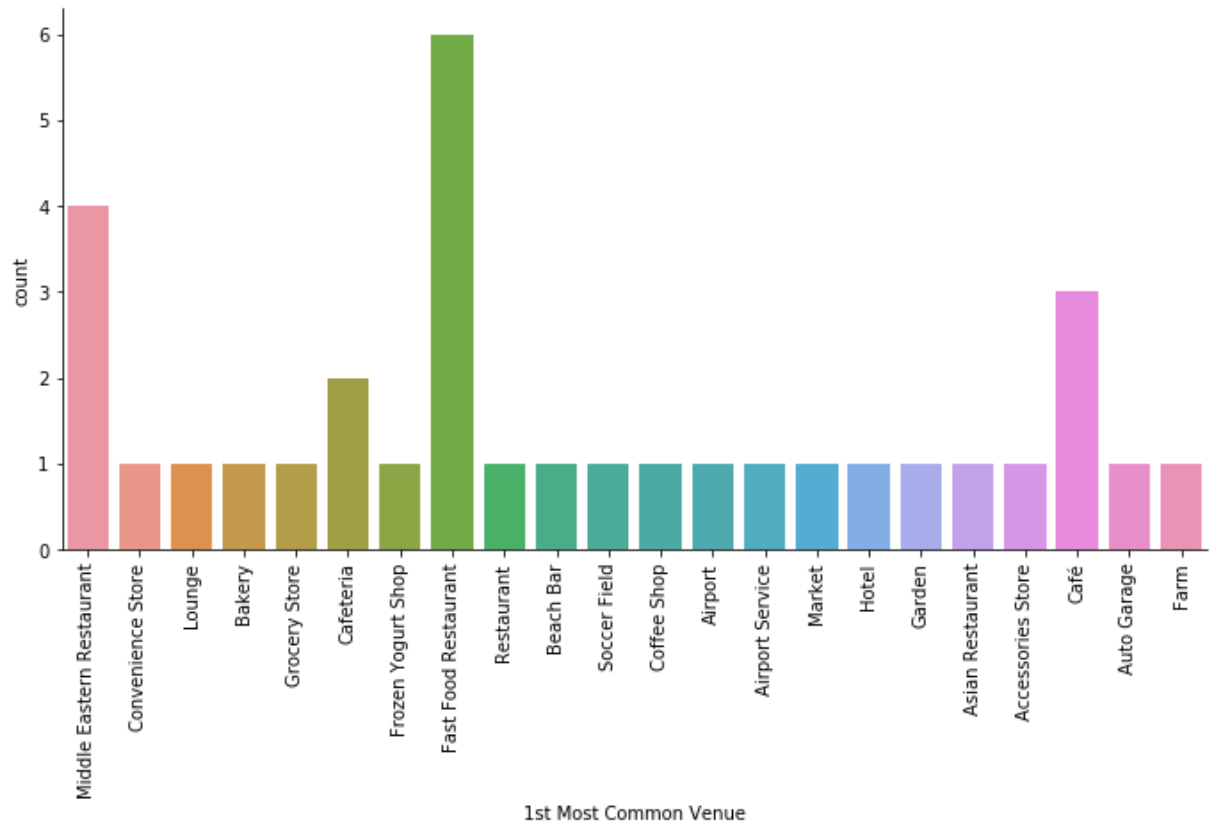
Sector Analysis

Analysis of Sector 2

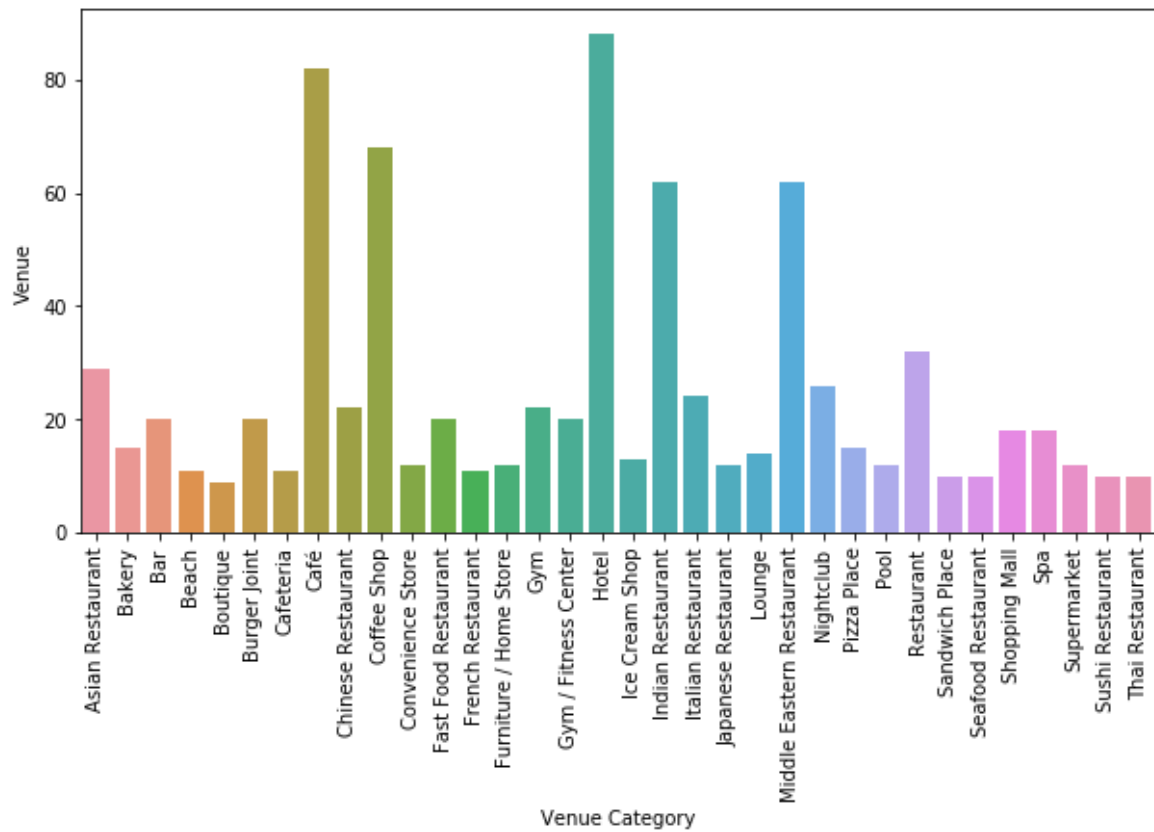


	Area	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	AL AYAS	Middle Eastern Restaurant	Café	Bar	Restaurant	Burger Joint	Italian Restaurant	Gym / Fitness Center	Hotel
1	AL GARHOUD	Convenience Store	Airport Terminal	Dessert Shop	Hotel	Vietnamese Restaurant	Cafeteria	Café	Candy Store
2	AL KHAHANEJ ONE	Lounge	Comfort Food Restaurant	Vietnamese Restaurant	Convenience Store	Cafeteria	Café	Candy Store	Clothing Store
3	AL MEZHAR FIRST	Bakery	Cafeteria	Grocery Store	Sandwich Place	Vietnamese Restaurant	Convenience Store	Café	Candy Store
4	AL MEZHAR SECOND	Grocery Store	Vietnamese Restaurant	Cosmetics Shop	Cafeteria	Café	Candy Store	Clothing Store	Coffee Shop

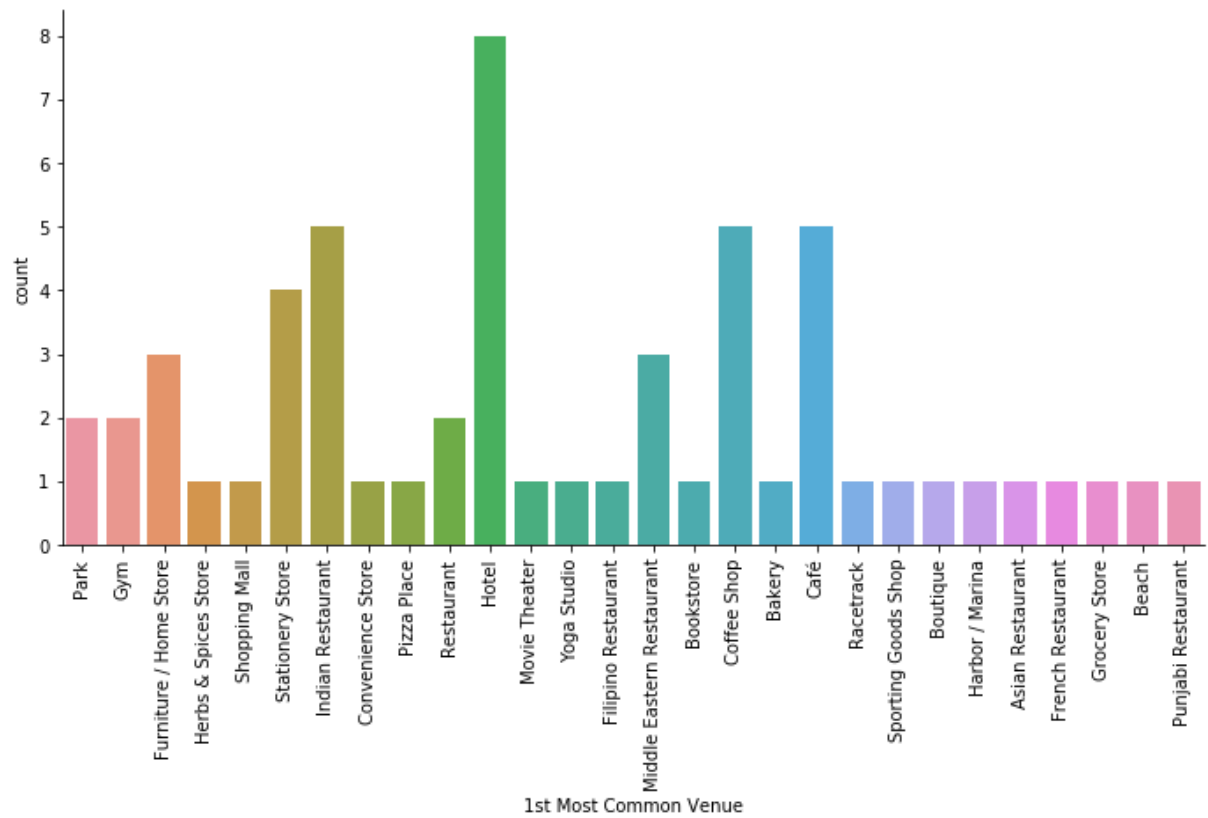
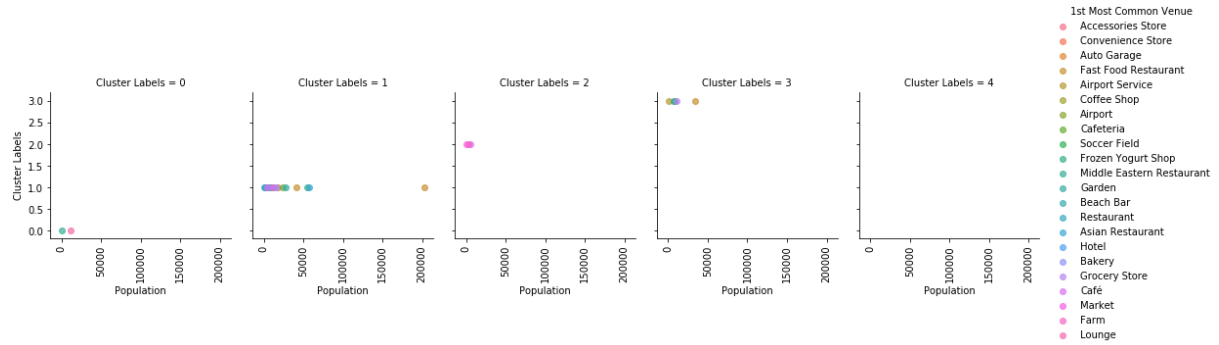




Analysis of Sector 3



	Area	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	AL BADA	Park	Indian Restaurant	Pool	Café	Middle Eastern Restaurant	Zoo	Food Court	Furniture / Home Store
1	AL BAESHAA SECOND	Gym	BBQ Joint	Café	Coffee Shop	Photography Studio	Photography Lab	Restaurant	Arts & Crafts Store
2	AL BARSHAA FIRST	Furniture / Home Store	Arts & Crafts Store	Art Gallery	Coffee Shop	Middle Eastern Restaurant	Football Stadium	Health & Beauty Service	Fast Food Restaurant
3	AL BARSHAA THIRD	Herbs & Spices Store	Cafeteria	Café	Fast Food Restaurant	Zoo	Garden	Furniture / Home Store	Frozen Yogurt Shop
4	AL GOZE FIRST	Shopping Mall	Food & Drink Shop	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Fountain



Community/Code	Emirate	Sectors	Population	Area	AreaArabic	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
302	Dubai	Sector 3	20	JUMEIRA BAY	جزيرة جميرا	25.08	55.15	0	Café	Mediterranean Restaurant	Hotel	Chinese Restaurant	Supermarket	Convenience Store	Light Rail Station	Spa
303	Dubai	Sector 3	4	WORLD ISLANDS	جزر العالم	25.23	55.17	0	Beach	Zoo	Gift Shop	General Entertainment	Gastropub	Garden	Furniture / Home Store	Frozen Yogurt Shop
304	Dubai	Sector 3	3	JUMEIRA ISLAND 2	جزيرة جميرا 2	25.08	55.15	0	Restaurant	Pool	Zoo	Food & Drink Shop	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint
311	Dubai	Sector 3	2	AL SHANDAGA	الشندغة	25.27	55.30	0	Middle Eastern Restaurant	Art Gallery	Tunnel	Flower Shop	Beach	Market	Zoo	Food Truck
312	Dubai	Sector 3	44783	AL BUQ AL KABEER	البرق الكبير	25.28	55.29	2	Indian Restaurant	Hotel	Asian Restaurant	Chinese Restaurant	Electronics Store	Pub	Tailor Shop	Flea Market

Clusters

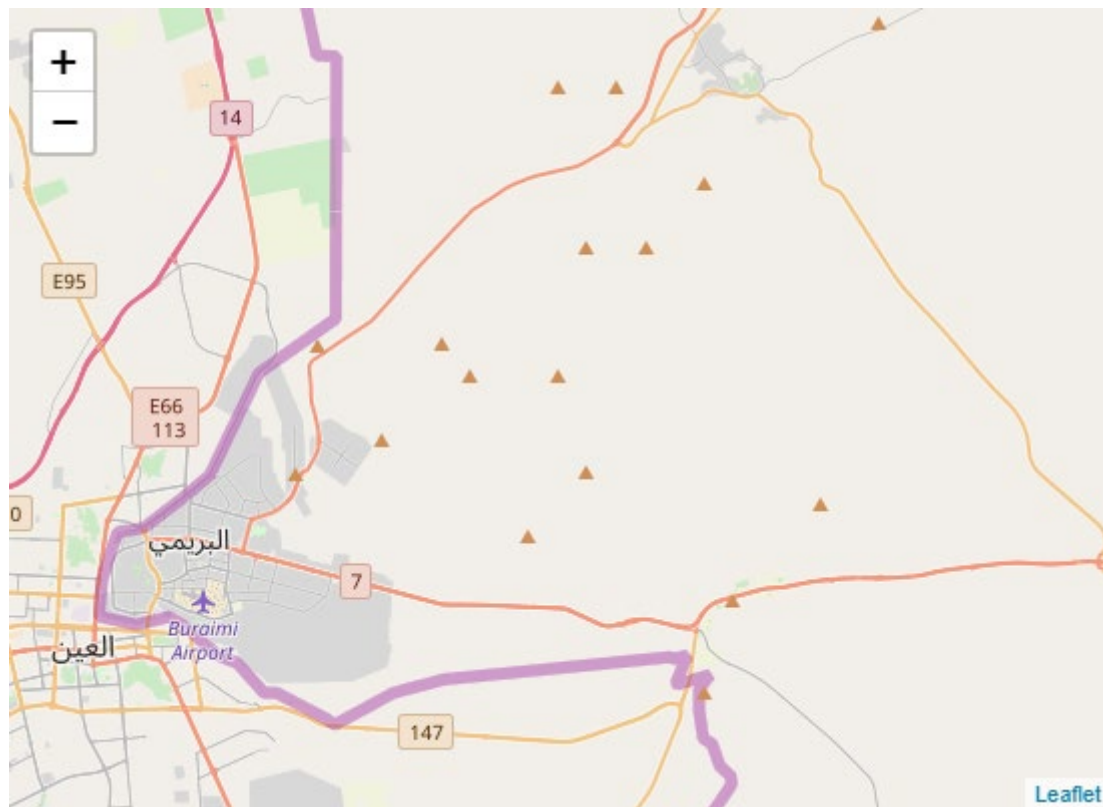


Figure 10 - Sector 2 Clusters

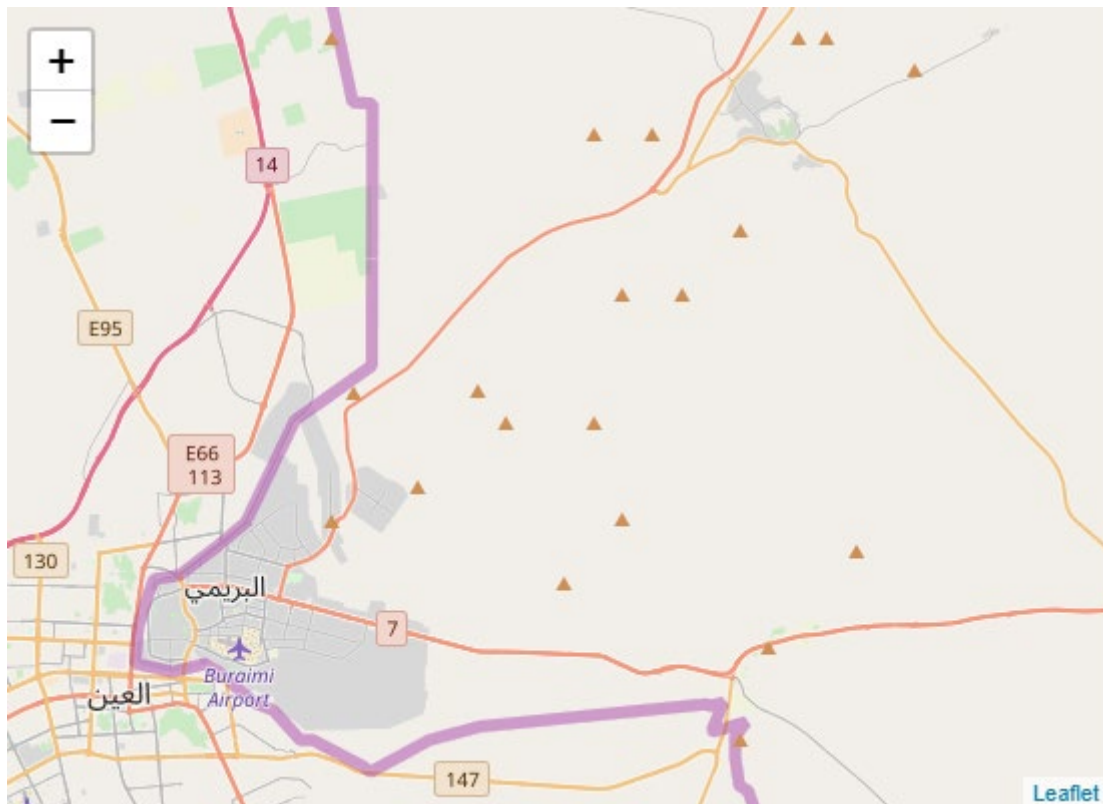


Figure 11 - Sector 3 Clusters

Results

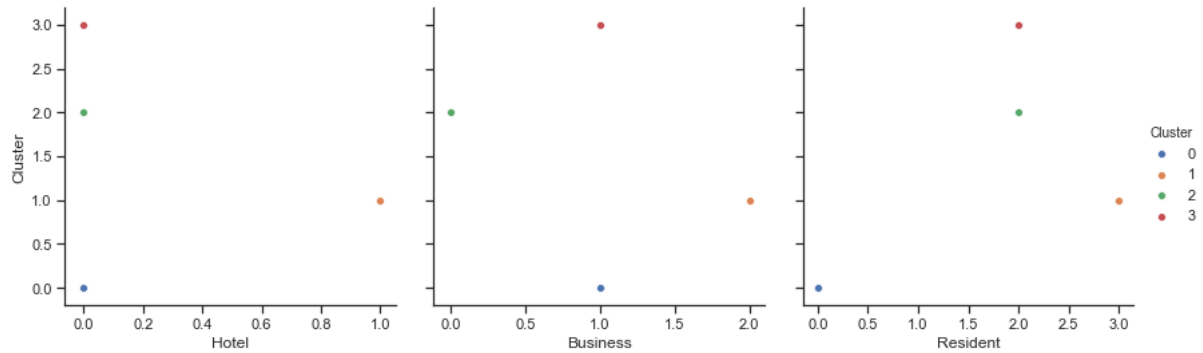


Figure 12 - Sector 2 by Cluster

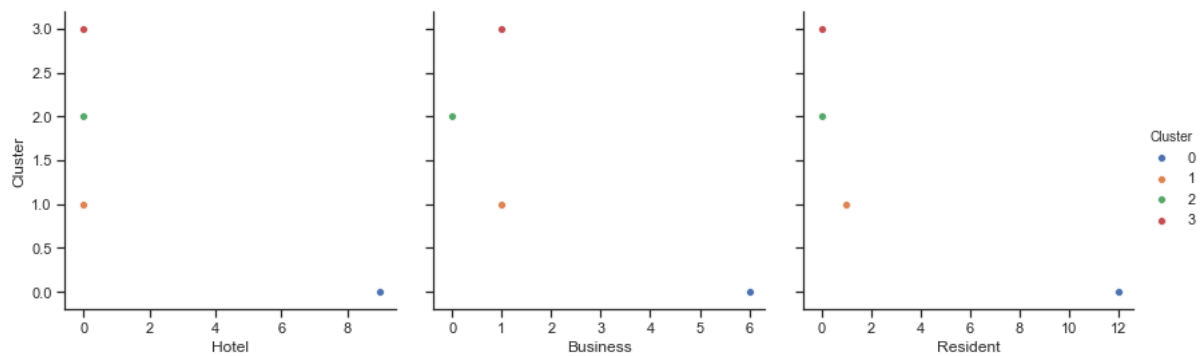


Figure 13 - Sector 3 by Cluster

Discussion

As illustrated above in the Clusters of Sector 2 and Sector 3, we learn that a more pristine Classification can be achieved through the grouping of most commonly available Venue Categories.

Sector 2 Cluster Division

- Cluster 1 Restaurants
- Cluster 2 Fast Food Restaurants
- Cluster 3 Farm

Sector 3 Cluster Division

- Cluster 1 Hotels
- Cluster 2 Café
- Cluster 3 Convenience Store

To extend this, we can further classify the Cluster Category Taxonomy so that we have a more concise view of Area Classification alongside deep Correlation Analysis. Further analysis, however is still pending due to the lack of Data Availability. The richer dataset we could obtain the better the probability of delivering new insights is.

Conclusion

Foursquare API is a powerful utility that enables us to capture detailed data about Common Places based on Geo Spatial Information of Dubai. Foursquare API has been proven valid in case of our Business Problem as defined above.

- Similarities and Dissimilarities between the Sectors of Dubai
- Classification of Areas by proximity, Residential, Tourism and Other

In the end, Dubai is a heavily populated city and it offers a lot of attraction throughout its Sectors. However, it is proven that Sector 3 offers more attraction towards Tourism whilst Sector 2 has a subtler influence of Residents.