Capstone project - The battle of neighborhoods in Dubai

1. Introduction/ Business understanding

1.1 Description of problem

Recently, the 13th edition of IPL (Indian Premier League) has been announced amid coronavirus pandemic and UAE has been chosen as the host country. The league is slated to commence from 19th of September, 2020. There has been ongoing discussion regarding the entry of audience in the stadium. The dataset of Dubai has been used to help the visitors of Dubai find places suitable for restaurant, hotel and so on during the IPL season.

1.2 Background of problem

Indian Premier League (IPL) is one of the most popular and highly valued league across the world particularly within cricket playing nation like India, Australia, England and so on. It is an India's version of T20 cricket league tournament. It gathers large audience in stadium and has huge viewership across cable TV and digital platform. Since, it is India's tournament it is mostly played in India. But in some extra ordinary condition, it is played in some other countries. This time it is UAE. UAE is also a cricket playing nation and has similar time zone as of India. Many games are slotted to be played in the stadium of Dubai as well. Dubai is located in the eastern part of the Arabian Peninsula on the coast of the Persian Gulf. Dubai aims to be the business hub of Western Asia. It is also a major global transport hub for passengers and cargo.

It is difficult for new traveler to find best places suited for them. So, using the foursquare API, I have performed various analysis on the data set of Dubai to find the best place for restaurant, hotels, parks and so on. This could help the new visitors of Dubai to get the overview of the places.

2. Data description

The dataset that used in this project is of Dubai scrapped from Wikipedia. This dataset contains the list of 131 communities of Dubai.

Data source:

https://en.wikipedia.org/wiki/List_of_communities_in_Dubai

We scrapped the data from the table of Wikipedia using a python library called 'Beautiful soup'. We will use only 3 columns of the dataset i.e. Community Number, Community (English) and Community (Arabic).

Example of dataset:

Community Number	Community (English)	Community (Arabic)	Area(km²)	Population(2000)	Population density(/km²)
891	Hatta	حتا			
721	Al Awir Second	العوير الثانية			
711	Al Awir First	العوير الأولى			
673	Al Barsha South Third	البرشاء جنوب الثالثة	38.1 km²	1248	33/km²
672	Al Barsha South Second	البرشاء جنوب الثانية	38.1 km²	1248	33/km²
671	Al Barsha South First	البرشاء جنوب الأولى	38.1 km²	1248	33/km²
622	Warsan Second	ورسان الثانية	17.1 km²	1,421	83/km²
621	Warsan First	ورسان الأولى	17.1 km²	1,421	83/km²
617	Nadd Al Shiba Fourth	ند الشبا الرابعة	8.4 km²	2,563	6.16/km²

I used 'geopy' library to find the latitude and longitude of each community. And then using foursquare API I found the venues in each community and what is each community famous for.

3. Methodology

3.1 Scrapping table of list of communities of Dubai from Wikipedia

I first read the table of Wikipedia and then iterating through data of each rows, a new data frame was created with 6 columns.

```
#Assigning the names of the column
column_names = ['CommunityNumber','Community_english', 'Community_arabic','Area', 'Population','Population_density']
df = pd.DataFrame(columns = column_names)
```

```
#searching through each row in the table of wikipedia and assigning the data of each cell to newly formed dataframe i.e. df
rows=mytable.find_all('tr')

for tr in rows:
    row_data=[]
    for td in tr.find_all('td'):
        row_data.append(td.text.strip())
    if(len(row_data)==6):
        df.loc[len(df)]=row_data
```

```
df.head()
```

Only 3 columns were kept and the rest were dropped. The columns were renamed. The new data frame looked in this way:

3.2 Adding geospatial data

Using geopy library location i.e. latitude and longitude of each communities were retrieved. The communities whose location could not be found were left out. Hence, I was left with 65 communities out of 131.

The outlook of data after adding location:

	Community	Latitude	Longitude
0	Abu Hail, Dubai	25.285942	55.329444
1	Al Baraha, Dubai	25.281368	55.319413
2	Al Buteen, Dubai	25.263057	55.320584
3	Al Garhoud, Dubai	25.239831	55.355668
4	Al Hudaiba, Dubai	25.240050	55.277459

3.3 Finding the venues of neighborhood within a radius of 500 meters using Foursquare API:

Defining the credentials to connect to Foursquare API:

```
CLIENT_ID = 'XHED4EHOOSFZ41A1QJXZ21TLW2KGUUUEH45ETS5QPVJDTXTU' # your Foursquare ID
CLIENT_SECRET = 'H53SGUBWTK44GUDTUXOI242XB5DVFTISNO4X0MGNVQKL4FEL' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version
```

First exploring the venues of neighborhood ‘Abu Hail’:

	name	categories	lat	lng
0	Pond Park - Al Qusais	Park	25.288060	55.332606
1	Zahr El Laymoun	Middle Eastern Restaurant	25.289486	55.330753
2	Lively	Track	25.285194	55.325276
3	Jannati Health Club and Spa	Spa	25.285408	55.325168

```
#we can see we have got only four venues in Abu Hail within 500m radius
print("{} venues were returned by foursquare".format(nearby_venues.shape[0]))
```

```
4 venues were returned by foursquare
```

we can see a total of four venues were returned by foursquare.

Exploring the venues of all 65 neighborhood of Dubai:

I was having problem when trying to explore the venues of all 65 neighborhoods at one go. So, what I did was I divided the 65 neighborhoods into 4 groups and then explored the venues of each neighborhood separately. When the venues of all four groups were returned, they were concatenated together.

The data frame after concatenation along with venues of each neighborhood is as follow:

```
The shape of dubai_venues_combined is:  (437, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abu Hail, Dubai	25.285942	55.329444	Pond Park - Al Qusais	25.288060	55.332606	Park
1	Abu Hail, Dubai	25.285942	55.329444	Zahr El Laymoun	25.289486	55.330753	Middle Eastern Restaurant
2	Abu Hail, Dubai	25.285942	55.329444	Lively	25.285194	55.325276	Track
3	Abu Hail, Dubai	25.285942	55.329444	Jannati Health Club and Spa	25.285408	55.325168	Spa
4	Al Baraha, Dubai	25.281368	55.319413	TGI Friday's	25.279242	55.316245	American Restaurant

The number of venues in each neighborhood returned by foursquare can be viewed as:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Abu Hail, Dubai	4	4	4	4	4	4
Al Baraha, Dubai	10	10	10	10	10	10
Al Buteen, Dubai	10	10	10	10	10	10
Al Garhoud, Dubai	10	10	10	10	10	10
Al Hudaiba, Dubai	10	10	10	10	10	10
...	...	...	...	...	...	...
Umm Al Sheif, Dubai	2	2	2	2	2	2
Umm Ramool, Dubai	5	5	5	5	5	5
Wadi Alamardi, Dubai	1	1	1	1	1	1
Za'abeel First, Dubai	10	10	10	10	10	10
Za'abeel Second, Dubai	4	4	4	4	4	4

From above table we can see Abu Hail has 4 venues, Al Baraha, Al Buteen, Al Garhoud has 40 venues and so on.

Analyzing each neighborhood using one hot encoding:

	Neighborhood	African Restaurant	American Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Garage	BBQ Joint	...	Tennis Court	Thai Restaurant	Track	Train Station	Tunnel	I
0	Abu Hail, Dubai	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
1	Abu Hail, Dubai	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
2	Abu Hail, Dubai	0	0	0	0	0	0	0	0	0	...	0	0	1	0	0	
3	Abu Hail, Dubai	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
4	Al Baraha, Dubai	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0	

5 rows × 128 columns

```
# one hot encoding
dubai_onehot = pd.get_dummies(dubai_venues_combined[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
dubai_onehot['Neighborhood'] = dubai_venues_combined['Neighborhood']

dubai_onehot.drop(['Neighborhood'],axis=1,inplace=True)
dubai_onehot.insert(loc=0, column='Neighborhood', value=dubai_venues_combined['Neighborhood'] )

dubai_onehot.head()
```

Displaying each neighborhood with top 5 most common venues:

```
----Abu Hail, Dubai----
                      venue  freq
0                     Track  0.25
1                      Park  0.25
2  Middle Eastern Restaurant  0.25
3                       Spa  0.25
4      Performing Arts Venue  0.00


----Al Baraha, Dubai----
                      venue  freq
0                     Hotel   0.2
1  Middle Eastern Restaurant   0.2
2                      Café   0.1
3        American Restaurant   0.1
4                       Spa   0.1
```

From above figure, we can see the top 5 venues of Al Baraha are Hotel, Middle Eastern Restaurant, cafe, American Restaurant and Spa. The frequency above represents that among 100% venues in Al Baraha, 20% are Hotel, 20% are Middle Eastern Restaurant, 10% are Café, 10% are American restaurant, 10% are Spa and the remaining 30% are venues other than these.

The top 10 venues of each neighborhood are displayed in below table:

	Neighborhood	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	9th Most Common Venue	10th Most Common Venue
0	Abu Hail, Dubai	Middle Eastern Restaurant	Park	Track	Spa	Filipino Restaurant	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish Market
1	Al Baraha, Dubai	Hotel	Middle Eastern Restaurant	Lounge	Coffee Shop	Café	Mobile Phone Shop	Spa	American Restaurant	Gym	Food & Drink Shop
2	Al Buteen, Dubai	Restaurant	Sandwich Place	Ice Cream Shop	Asian Restaurant	Coffee Shop	Fast Food Restaurant	Fried Chicken Joint	Hotel	Historic Site	Flower Shop
3	Al Garhoud, Dubai	Hotel	Post Office	Gastropub	Dessert Shop	Coffee Shop	Restaurant	Shipping Store	Basketball Court	Yemeni Restaurant	Flea Market
4	Al Hudaiba, Dubai	Convenience Store	Ice Cream Shop	Cafeteria	Café	Korean Restaurant	Filipino Restaurant	Juice Bar	Asian Restaurant	Hotel	Fast Food Restaurant
...	...	...	...	...	...	...	...	...	...	...	...
57	Umm Al Sheif, Dubai	Pool	French Restaurant	Ice Cream Shop	Department Store	Dessert Shop	Dim Sum Restaurant	Donut Shop	Electronics Store	Farmers Market	Fast Food Restaurant
58	Umm Ramool, Dubai	Auto Garage	Gym	Food & Drink Shop	Cafeteria	Food	Flower Shop	Flea Market	Fish Market	Yemeni Restaurant	Fried Chicken Joint
59	Wadi Alamardi, Dubai	Campground	Yemeni Restaurant	Deli / Bodega	Dessert Shop	Dim Sum Restaurant	Donut Shop	Electronics Store	Farmers Market	Fast Food Restaurant	Filipino Restaurant
60	Za'abeel First, Dubai	Coffee Shop	Middle Eastern Restaurant	Japanese Restaurant	Building	Steakhouse	Lebanese Restaurant	Indian Restaurant	Gym	Food & Drink Shop	Dim Sum Restaurant

Clustering the neighborhoods i.e. communities of Dubai based on the similarities of their venues using K- Means algorithm:

The neighborhoods have been grouped into five clusters.

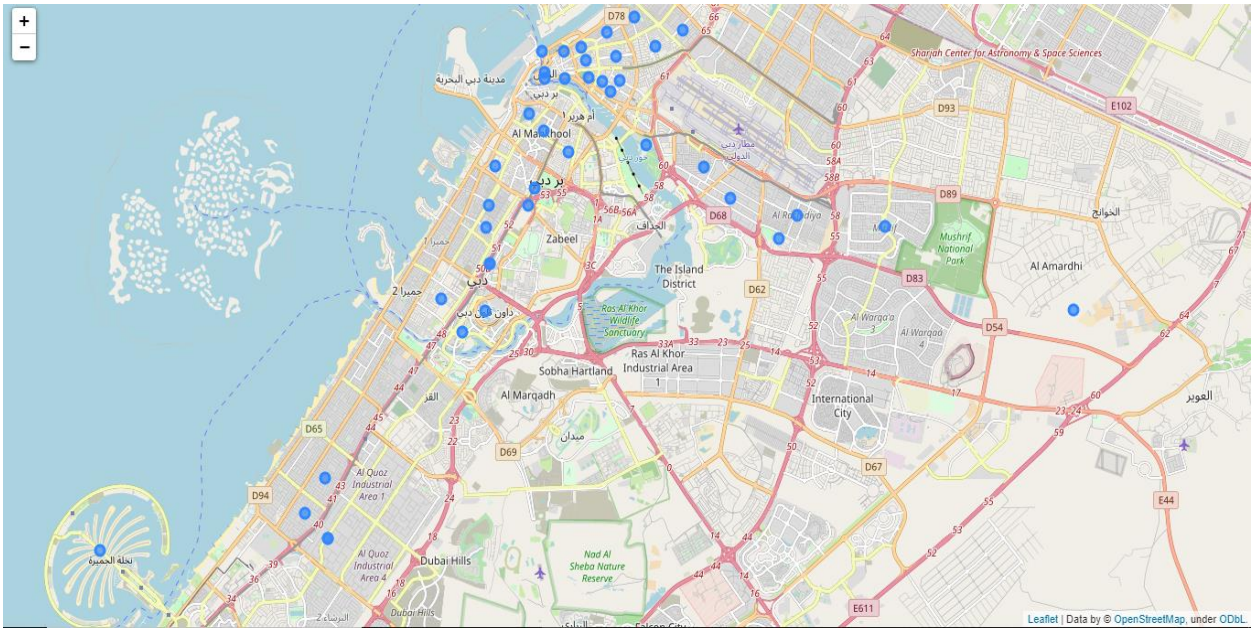
The K-Means label for each neighborhood:

```
# set number of clusters
kclusters = 5

dubai_grouped_clustering = dubai_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(dubai_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_
```

```
array([1, 1, 1, 1, 1, 1, 3, 1, 1, 1, 3, 1, 1, 1, 1, 1, 1, 0, 0, 3, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 3, 1, 1, 1, 1, 1, 1, 4, 1, 3, 0, 1,
       1, 1, 3, 1, 1, 1, 3, 1, 1, 1, 1, 3, 1, 1, 1, 2, 3, 1])
```

Now, plotting each neighborhood into map using folium library:

Folium is an essential library to visualize locations on a map. It also allows to zoom in and zoom out the map. With very lines of code, it, does amazing piece of work for visualization of data.



4. Results/ Discussions

From the study of venues of each neighborhood we got some results. Lets discuss those results here:

Finding 1:

Restaurant	19
Middle Eastern Restaurant	19
Indian Restaurant	18
Asian Restaurant	14
Japanese Restaurant	7
Mexican Restaurant	6
Fast Food Restaurant	6
Chinese Restaurant	5
Italian Restaurant	5
Korean Restaurant	4
Filipino Restaurant	4
Seafood Restaurant	3
Greek Restaurant	2
Lebanese Restaurant	2
Thai Restaurant	2

As we are discussing about IPL going to be held in UAE and being an Indian league tournament more Indians are expected to visit this place. From above data we see plenty of Indian restaurant available here. So, people from India will probably face no problem finding the restaurant of their kind. Also, cricket is mainly considered an Asian game. So, for people from across the Asia visiting the place can also find plenty of Asian restaurant.

The places where one can find Indian restaurant easily are:

	Neighborhood	1st Most Common Venue
0	Al Muteena, Dubai	African Restaurant
1	Al Qusais Industrial Fifth, Dubai	Filipino Restaurant
2	Emirates Hill Third, Dubai	Indian Restaurant
3	Marsa Dubai, Dubai	Indian Restaurant
4	Al Raffa, Dubai	Indian Restaurant
5	Al Karama, Dubai	Indian Restaurant
6	Al Jaddaf, Dubai	Italian Restaurant
7	Al Mina, Dubai	Mexican Restaurant
8	Jebel Ali 1, Dubai	Mexican Restaurant
9	Al Muraqqabat, Dubai	Middle Eastern Restaurant
10	Hor Al Anz East, Dubai	Middle Eastern Restaurant

From above data, we can see Emirates Hill Third, Marsa Dubai, Al Raffa, Al Karama are more famous for Indian restaurant.

Finding 2:

Places where hotel can be found easily are:

	Neighborhood	1st Most Common Venue
0	Al Qusais Industrial First, Dubai	Hotel
1	Naif, Dubai	Hotel
2	Nad Shamma, Dubai	Hotel
3	Al Murar, Dubai	Hotel
4	Al Sabkha, Dubai	Hotel
5	Al Mankhool, Dubai	Hotel
6	Jebel Ali Palm, Dubai	Hotel
7	Al Garhoud, Dubai	Hotel
8	Al Baraha, Dubai	Hotel
9	Jumeira Second, Dubai	Hotel
10	Trade Centre 2, Dubai	Hotel

So, if someone in Dubai is looking for place with more options available for hotel, they can choose from the places above.

Finding 3:

Places with most parks are given below:

	Neighborhood	1st Most Common Venue
0	Za'abeel Second, Dubai	Park
1	Al Qusais Second, Dubai	Park

So, people fond of parks can choose to stay in the communities/ neighborhoods mentioned above.

Finding 4:

Someone fond of beach can choose to stay in the given below:

	Neighborhood	1st Most Common Venue
0	Palm Jumeira, Dubai	Beach

Finding 5:

Many people love to have coffee frequently and it becomes for them when they don't find a coffee shop easily. So, here are the list of places more famous for having coffee shops.

	Neighborhood	1st Most Common Venue
0	Rigga Al Buteen, Dubai	Coffee Shop
1	Al Qusais First, Dubai	Coffee Shop
2	Mirdif, Dubai	Coffee Shop
3	Za'abeel First, Dubai	Coffee Shop
4	Al Satwa, Dubai	Coffee Shop
5	Jebel Ali 2, Dubai	Coffee Shop
6	Al Jafiliya, Dubai	Coffee Shop

So, these were some findings which I felt were more necessary to be known to people traveling to Dubai.

5. Conclusion

In todays time of digital world, data science plays vital role. It increases the capability of the businesses, medical instruments. It helps the businesses to analyze the behavior of their customers, and also compete with their counterpart in a fast-changing world. With an exponential increase in the use of digital instruments in various sectors, lots of data are being generated and stored every day. Hence, it becomes quite instrumental and essential to analyze

those data to gain information which could help in the improvement of various sectors by taking right decision at right time.

With this project I have made an effort to help the first time travelers to Dubai especially during the season of IPL. I have used some common libraries like geopy, folium to find the location and plot those locations on map respectively. Also, I have made use of foursquare API to explore the venues of each neighborhoods. Despite all these efforts, there are still some areas of improvements which could help in providing even more essential and realistic information from the data.