



Battle of Neighborhoods Doha, Qatar

By

Muhammad Jehanzaib Rao

IN Solution Architect Ooredoo, Qatar



Problem Background



- Relocation within Doha, Qatar is needed due to change in office location.
 - To avoid being stuck in traffic jams
- Requirement is to find an area that is similar to Al Dafna area with respect to nearby venues.
- Also, new area should be closer to Ooredoo HQ2.
- Areas that are closer to Ooredoo HQ2 are
 - Al Hilal
 - Najma
 - Umm Ghuwailina

Data Acquisition and Cleaning

- Data was taken from Wikipedia [link](#). It contained list of all communities of Doha.
- Latitudes and Longitudes of all communities were added to the data by using geocode library.
- All unnecessary columns were removed along with rows that did not have Latitudes and longitudes. Final data had 51 neighborhoods of Doha.

```
print(df_doha.shape)  
df_doha.head()
```

```
(51, 3)
```

	Neighborhood	Latitude	Longitude
0	Al Bidda	25.290243	51.526697
1	Al Dafna	25.319584	51.536284
2	Al Hilal	25.260376	51.546488
3	Al Kharayej	25.209836	51.454974
4	Al Khulaifat	25.285061	51.552715

Exploring Neighborhoods

- Foursquare API was used to explore up to 100 nearby venues from each Neighborhood within the radius of 1000 meters.
- It resulted in 12398 different venues of 190 unique categories.
- A new Data Frame was then created with Neighborhood and frequency of each venue in that neighborhood as shown below;

```
print(doha_grouped.shape)  
doha_grouped.head()
```

(50, 191)

	Neighborhood	Afghan Restaurant	African Restaurant	Airport	Airport Service	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auditorium	Auto Garage	BBQ Joint	Bagel Shop
0	Al Bidda	0.0	0.00	0.0	0.0	0.000000	0.00	0.021739	0.000000	0.0	0.000000	0.000000	0.0	0.0	0.021739	0.00
1	Al Dafna	0.0	0.01	0.0	0.0	0.010000	0.01	0.000000	0.000000	0.0	0.020000	0.000000	0.0	0.0	0.000000	0.01
2	Al Hilal	0.0	0.00	0.0	0.0	0.000000	0.00	0.000000	0.000000	0.0	0.041667	0.000000	0.0	0.0	0.000000	0.00
3	Al Kharayej	0.0	0.00	0.0	0.0	0.000000	0.00	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.00
4	Al Khulaifat	0.0	0.00	0.0	0.0	0.013333	0.00	0.000000	0.013333	0.0	0.000000	0.026667	0.0	0.0	0.000000	0.00

Exploring Neighborhoods

- A new DataFrame was then created with top 5 venues of each neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Al Bidda	Café	Middle Eastern Restaurant	Hotel	Italian Restaurant	Dessert Shop
1	Al Dafna	Coffee Shop	Hotel	Café	Middle Eastern Restaurant	Italian Restaurant
2	Al Hilal	Department Store	Pizza Place	Coffee Shop	Indian Restaurant	Sandwich Place
3	Al Kharayej	Coffee Shop	Zoo	Filipino Restaurant	French Restaurant	Food Truck
4	Al Khulaifat	Hotel	Café	Indian Restaurant	Middle Eastern Restaurant	Restaurant

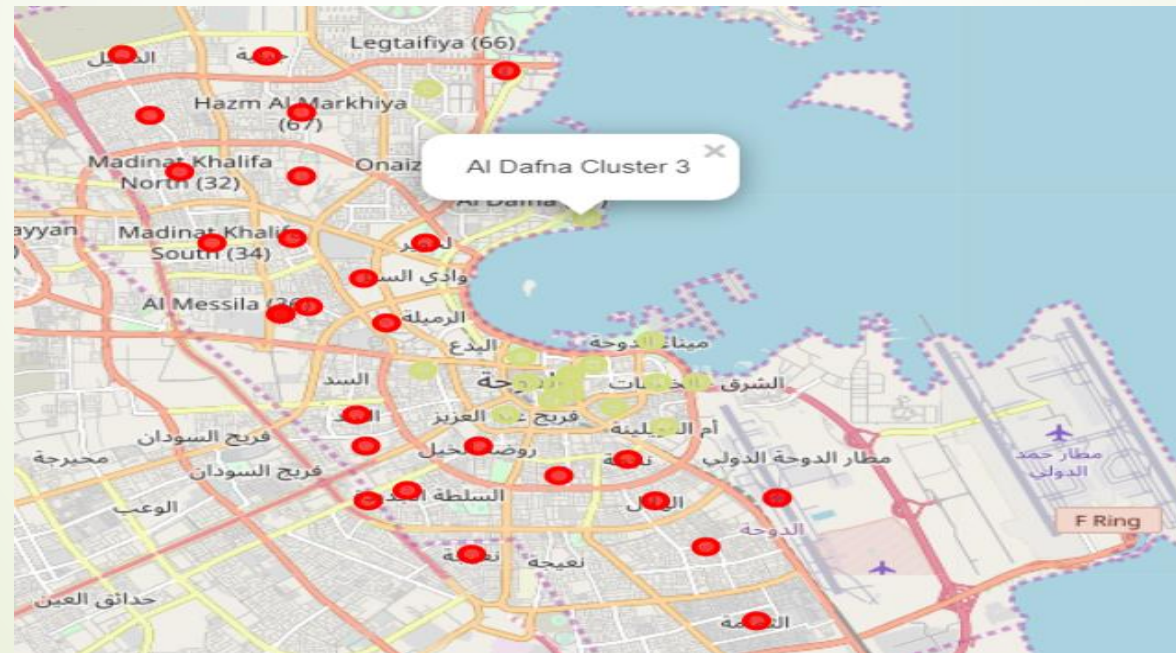
Machine Learning: K-means Clustering

- KMeans clustering algorithm was used for segmentation.
- Cluster Labels were added in Data Frame as shown below;

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Al Bidda	25.290243	51.526697	3.0	Café	Middle Eastern Restaurant	Hotel	Italian Restaurant	Dessert Shop
1	Al Dafna	25.319584	51.536284	3.0	Coffee Shop	Hotel	Café	Middle Eastern Restaurant	Italian Restaurant
2	Al Hilal	25.260376	51.546488	0.0	Department Store	Pizza Place	Coffee Shop	Indian Restaurant	Sandwich Place
3	Al Kharayej	25.209836	51.454974	2.0	Coffee Shop	Zoo	Filipino Restaurant	French Restaurant	Food Truck
4	Al Khulaifat	25.285061	51.552715	3.0	Hotel	Café	Indian Restaurant	Middle Eastern Restaurant	Restaurant

Machine Learning: K-means Clustering

- Majority of Neighborhoods were observed to be part of two clusters.
- It was verified by plotting map by Folium.



Analyzing Clusters

Cluster 1

```
doha_merged.loc[doha_merged['Cluster Labels'] == 0, doha_merged.columns[[0] + list(range(4, doha_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	Al Hilal	Department Store	Pizza Place	Coffee Shop	Indian Restaurant	Sandwich Place
5	Al Mansoura	Indian Restaurant	Hotel	Fast Food Restaurant	Café	Asian Restaurant
6	Al Markhiya	Café	Coffee Shop	Asian Restaurant	Restaurant	Pharmacy
10	Al Qassar	Beach	Hotel	Café	Nightclub	Restaurant
12	Al Sadd	Middle Eastern Restaurant	Café	Coffee Shop	Fast Food Restaurant	Italian Restaurant
15	Al Thumama	Coffee Shop	Boutique	Middle Eastern Restaurant	Convenience Store	Shopping Mall
17	Doha International Airport	Fast Food Restaurant	Indian Restaurant	Café	Restaurant	Department Store
19	Duhail	Afghan Restaurant	Spa	Flower Shop	Gym / Fitness Center	French Restaurant
21	Fereej Al Nasr	Coffee Shop	Middle Eastern Restaurant	Café	Fast Food Restaurant	Asian Restaurant
23	Fereej Bin Omran	Restaurant	Asian Restaurant	Department Store	Convenience Store	Pizza Place
24	Fereej Kulaib	Middle Eastern Restaurant	Fast Food Restaurant	Restaurant	Indian Restaurant	Cafeteria
25	Hamad Medical City	Coffee Shop	Park	Hotel	Intersection	Café
26	Hazm Al Markhiya	Cupcake Shop	Park	Food & Drink Shop	Shipping Store	Gym / Fitness Center

Analyzing Clusters

- It was observed that one cluster belongs to Industrial Area and another to Zoo Areas.

Cluster 2

```
doha_merged.loc[doha_merged['Cluster Labels'] == 1, doha_merged.columns[[0] + list(range(4, doha_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
27	Industrial Area	Business Service	Zoo	Currency Exchange	French Restaurant	Food Truck

Cluster 3

```
doha_merged.loc[doha_merged['Cluster Labels'] == 2, doha_merged.columns[[0] + list(range(4, doha_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
3	Al Kharayej	Coffee Shop	Zoo	Filipino Restaurant	French Restaurant	Food Truck
14	Al Tarfa	Coffee Shop	Food Court	Zoo	Filipino Restaurant	French Restaurant

Analyzing Clusters

- Cluster 4 has Hotel as one main Venue along with Cafes, restaurants and Coffee shops.

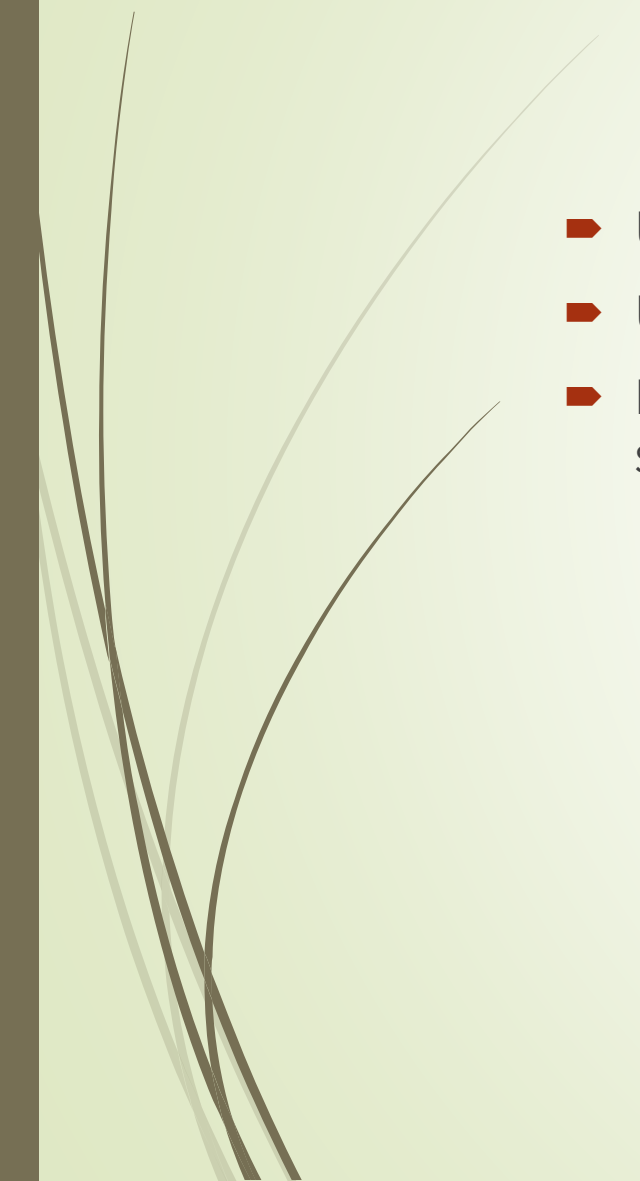
Cluster 4

```
doha_merged.loc[doha_merged['Cluster Labels'] == 3, doha_merged.columns[[0] + list(range(4, doha_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Al Bidda	Café	Middle Eastern Restaurant	Hotel	Italian Restaurant	Dessert Shop
1	Al Dafna	Coffee Shop	Hotel	Café	Middle Eastern Restaurant	Italian Restaurant
4	Al Khulaifat	Hotel	Café	Indian Restaurant	Middle Eastern Restaurant	Restaurant
7	Al Messila	Hotel	Middle Eastern Restaurant	Café	Restaurant	Coffee Shop
8	Al Mirqab	Hotel	Café	Middle Eastern Restaurant	Restaurant	Coffee Shop
9	Al Najada	Hotel	Middle Eastern Restaurant	Café	Restaurant	Coffee Shop
11	Al Rufaa	Hotel	Café	Restaurant	Middle Eastern Restaurant	Indian Restaurant
13	Al Souq	Hotel	Café	Middle Eastern Restaurant	Restaurant	Coffee Shop
16	Barahat Al Jufairi	Hotel	Café	Middle Eastern Restaurant	Restaurant	Coffee Shop
18	Doha Port	Hotel	Café	Restaurant	Harbor / Marina	Park
20	Fereej Abdul Aziz	Hotel	Indian Restaurant	Fried Chicken Joint	Fast Food Restaurant	Restaurant
22	Fereej Bin Mahmoud	Hotel	Café	Hookah Bar	Indian Restaurant	Intersection
35	Musheireb	Hotel	Middle Eastern Restaurant	Café	Restaurant	Indian Restaurant
38	New Al Mirqab	Hotel	Café	Middle Eastern Restaurant	Restaurant	Coffee Shop
42	Old Al Ghanim	Hotel	Café	Middle Eastern Restaurant	Indian Restaurant	Coffee Shop
44	Onaiza	Fruit & Vegetable Store	BBQ Joint	Hotel	Bed & Breakfast	Health & Beauty Service
47	Umm Ghuwailina	Hotel	Middle Eastern Restaurant	Pizza Place	Fast Food Restaurant	Department Store



Conclusion

- Umm Ghuwailina is in same cluster with Al Dafna Area.
 - Umm Ghuwailina is closer to Ooredoo HQ2 as well.
 - Data Analysis with and results of segmentation through K Means Clustering shows that I should move to Umm Ghuwailina area.
- 



Thank You