Applied Data Science Capstone

Opening a Billiards Café

Rayyan Abid Ali

Introduction

Due to the current COVID-19 pandemic there have been restrictions that result in less people going outside. Although this is disadvantageous for most businesses one can hypothesise that when the COVID-19 issue is resolved there will most probably be and upsurge of people leaving their homes to go outside. A billiards cafe is a great way for friends and family to interact while having fun. The aim of this project is to be able to find a suitable location to open a billiards cafe in Riyadh, Saudi Arabia to take advantage of the aforementioned speculation.

Data

The data for this project will be retrieved from the Foursquare API. Using the number of visitors and check-ins for current billiards cafés with the stats endpoint, as well as nearby venues with the explore endpoint, enough information can be gathered to cluster similar locations and see which billiard cafes have the most people gone to and why that particular café has more people than others. Using this information, we can then find areas to open our billiards café.

NOTE: Upon starting the project it was found that in order to get the number of visitors and check-ins for a venue you must be the owner of the venue and so instead of the number of people the deciding feature had to be changed to the ratings of the venues. Although ratings are not really the way to go there was no choice as there is a large lack of data in Riyadh, Saudi Arabia for billiards venues.

Methodology

The first step is to get the location of the existing billiards café's in Riyadh. This was done by calling the following Foursquare API URL:

```
https://api.foursquare.com/v2/venues/search?&client_id={CLIENT_ID}&client_secret={CLIENT_SECRET}&v={VERSION}&ll=24.7136,46.6753&categoryId=4bf58dd8d48988d1e3931735&limit=100
```

Where the CLIENT_ID and CLIENT_SECRET are the credentials of the Foursquare API application created in week 4 of the course. The version is set to represent the data up till the 15th of November 2020. The category parameter was set to be the category number for pool table venues.

After getting the JSON file we extracted it and made a dataframe consisting of the venues' id and name as well their latitude and longitude coordinates. Some of the venues name were in Arabic so they were converted to English with the help of the *translate* module. The dataframe is named billiards_data and the first five rows are shown in Figure 1.

	id	name	location.lat	location.lng
0	500094bfe4b042c51bee94fd	Al Samar Billiards and Playstation	24.708333	46.661671
1	503f6ec9e4b08d67c33747c2	meeting time	24.688202	46.638902
2	56807a07498e89ea767ab3f8	Hour Cafe	24.631776	46.611611
3	52a4944611d2abd78d2538e0	Al Wizarat	24.594221	46.693558
4	5bc10279c0cacb0039d3182e	The Black Eight Poll Hall	24.751186	46.781500

Figure 1 billiards_data

These venue were then visualized on a map using the *Folium* module as shown in Figure 2.

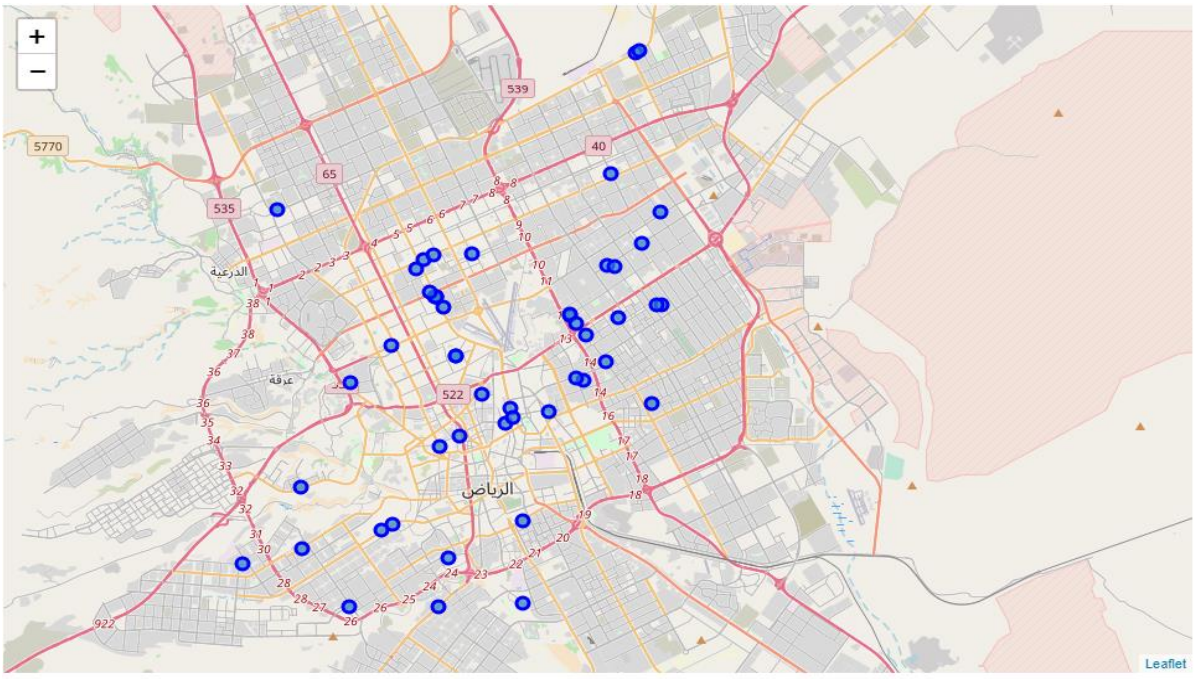

Figure 2 Locations of billiards venues

Now a URL needs to be made that can call forward the Foursquare API to give the details of the venue (which contains the ratings). Since a sandbox account is being used and this is a premium call, we can get a max of 50 venues. Luckily for us the number of venues is actually exactly 50 already.

The URL is as follows:

```
https://api.foursquare.com/v2/venues/{VENUE_ID[i]}/?&client_id={CLIENT_ID}&client_secret={CLIENT_SECRET}&v={VERSION}&limit=50
```

[i] here represents the numbers 0-49 as this URL has to loop through the 50 different venues.

After getting the respective json files we then get the ratings and append it to the existing billiards_data dataframe as shown in Figure 3.

	id	name	location.lat	location.lng	Ratings
0	500094bfe4b042c51bee94fd	Al Samar Billiards and Playstation	24.708333	46.661671	7.6
1	503f6ec9e4b08d67c33747c2	meeting time	24.688202	46.638902	0.0
2	56807a07498e89ea767ab3f8	Hour Cafe	24.631776	46.611611	0.0
3	52a4944611d2abd78d2538e0	Al Wizarat	24.594221	46.693558	0.0
4	5bc10279c0cacb0039d3182e	The Black Eight Poll Hall	24.751186	46.781500	0.0

Figure 3 Updated billiards_data

Here is where we see the lack of data for billiards venues in Riyadh. There are actually 38 venues that have not been rated. To carry on further analysis we assume that these places have not been rated due to less people visiting it. This assumption can be helpful as it can tell us places to not build our billiards café.

The next step was to explore areas around each of the venues. What were the other type of venues around these billiards cafes and do they make a difference to the ratings?

The URL for this search is:

```
https://api.foursquare.com/v2/venues/explore?&client_id={CLIENT_ID}&client_secret={CLIENT_SECRET}&v={VERSION}&ll={lat[i]},{lng[i]}
```

Where lat and lng are the latitudes and longitudes of the different venues and [i] once again ranges from 0-49 for the 50 different venues.

From getting the respective JSON files it is seen that there are a total of 157 different venues across the billiard cafes. A table was made which represented how many of these nearby venues are next to our billiards cafes. The table is named df_nearbyVenues and the first five rows are shown in Figure 4.

	Ratings	Donut Shop	Spa	Pharmacy	Gym	Chinese Restaurant	Juice Bar	Ice Cream Shop	Jewelry Store	Lounge	...	Street Food Gathering	Theater	Theme Restaurant	Multiplex	Optical Shop
name																
Al Samar Billiards and Playstation	7.6	2.0	1.0	2.0	2.0	1.0	2.0	1.0	1.0	1.0	...	0.0	0.0	0.0	0.0	0.0
meeting time	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
Hour Cafe	0.0	1.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
Al Wizarat	0.0	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
The Black Eight Poll Hall	0.0	1.0	1.0	3.0	0.0	0.0	1.0	2.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0

Figure 4 df_nearbyVenues

Now that the number of venues next to the different billiards cafes are known, we can make a KMeans cluster algorithm to cluster different venues and try to interpret the resulting data.

Before we can do that however, we need to find an optimum number of clusters to pick. This is done using the elbow method. With the *scikitlearn* module a test was run for the above dataset (dropping the ratings column) for clusters ranging from 2 to 9. A graph of SSE (Sum of squared errors) against number of clusters was plotted (Figure 5) and the optimum value was said to be found at number of clusters=6.

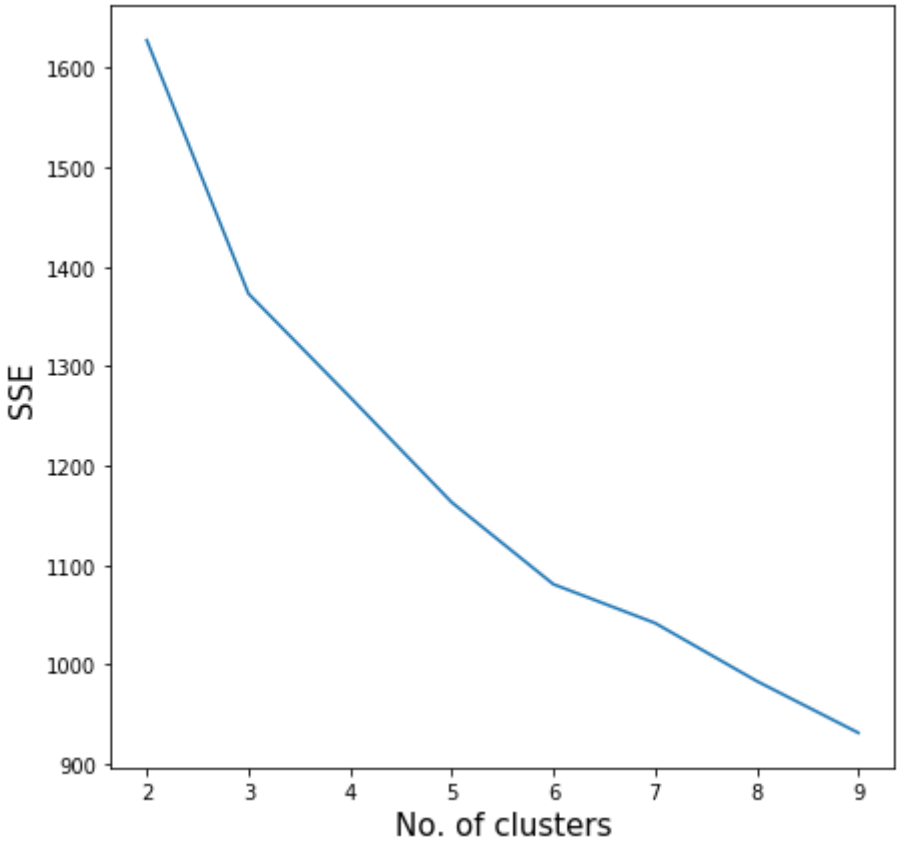

Figure 5 SSE - number of clusters

After finding the optimum number of clusters a KMeans pipeline was run and the clusters plotted on a map (Figure 6).

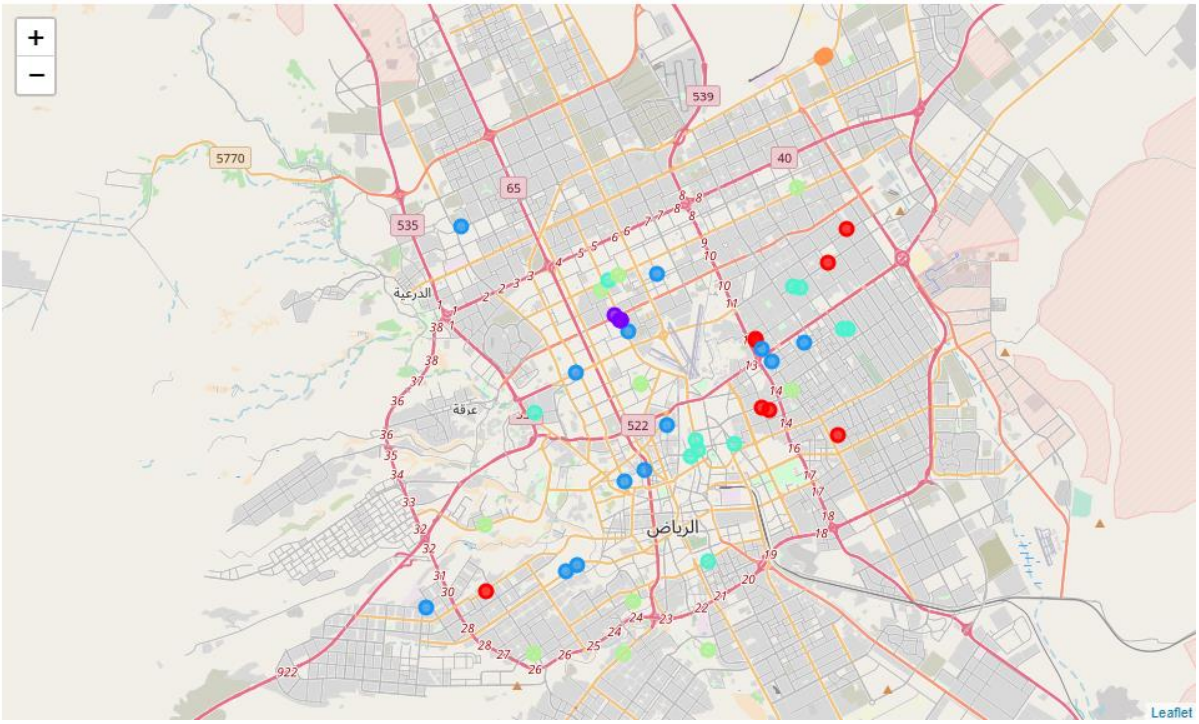

Figure 6 Clusters Visualized

Tables were then made to examine the relation between these clusters. The first few rows of these tables are shown in Figures 7 to 14 below.

	Cluster Labels	Ratings	Donut Shop	Spa	Pharmacy	Gym	Chinese Restaurant	Juice Bar	Ice Cream Shop	Jewelry Store	...	Street Food Gathering	Theater	Theme Restaurant	Multiplex	Optical Shop
name																
Blessed Shed	0	0.0	1.0	0.0	0.0	0.0	0.0	3.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0
Super touch (billiards) and (cafe)	0	0.0	1.0	0.0	0.0	0.0	0.0	2.0	2.0	0.0	...	0.0	0.0	0.0	0.0	0.0
Black Ball Champions Pool Hall	0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0
Celebrity Billiards	0	0.0	3.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
billards rawdah	0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0

Figure 7 Venues in Cluster 0

	Cluster Labels	Ratings	Donut Shop	Spa	Pharmacy	Gym	Chinese Restaurant	Juice Bar	Ice Cream Shop	Jewelry Store	...	Street Food Gathering	Theater	Theme Restaurant	Multiplex	Optical Shop	Auto Worksho
name																	
Elite Time Billiards	1	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	2.0	...	0.0	0.0	0.0	0.0	0.0	0.0
Elite Center for Billiards and Snooker	1	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	2.0	...	0.0	0.0	0.0	0.0	0.0	0.0
North Billiards	1	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	2.0	...	0.0	0.0	0.0	0.0	0.0	0.0
North Center for Snooker and Billiards	1	6.7	1.0	0.0	0.0	0.0	0.0	0.0	1.0	2.0	...	0.0	0.0	0.0	0.0	0.0	0.0
fitness time	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	3.0	...	0.0	0.0	0.0	0.0	0.0	0.0

Figure 8 Venues in Cluster 1

	Cluster Labels	Ratings	Donut Shop	Spa	Pharmacy	Gym	Chinese Restaurant	Juice Bar	Ice Cream Shop	Jewelry Store	...	Street Food Gathering	Theater	Theme Restaurant	Multiplex	Optical Shop
name																
Al Samar Billiards and Playstation	2	7.6	2.0	1.0	2.0	2.0	1.0	2.0	1.0	1.0	...	0.0	0.0	0.0	0.0	0.0
Young imagination	2	0.0	2.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0
Snooker - Billiards Billiards	2	7.2	3.0	0.0	1.0	2.0	0.0	1.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0
Fal - Fall Sports Center	2	5.3	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0
Final Shoot (Final Shoot for Bowling and Billiards)	2	7.4	2.0	1.0	1.0	1.0	0.0	2.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0

Figure 9 Venues in Cluster 2

	Cluster Labels	Ratings	Donut Shop	Spa	Pharmacy	Gym	Chinese Restaurant	Juice Bar	Ice Cream Shop	Jewelry Store	...	Street Food Gathering	Theater	Theme Restaurant	Multiplex
name															
meeting time	3	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
The Black Eight Poll Hall	3	0.0	1.0	1.0	3.0	0.0	0.0	1.0	2.0	1.0	...	0.0	0.0	0.0	0.0
Omar's Pool	3	0.0	2.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	...	0.0	0.0	0.0	0.0
A basket of billiard stars	3	0.0	4.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	...	0.0	0.0	0.0	0.0
BeautyLoungeGym	3	0.0	1.0	0.0	2.0	0.0	0.0	1.0	4.0	0.0	...	0.0	0.0	0.0	0.0

Figure 10 Venues in Cluster 3

	Cluster Labels	Ratings	Donut Shop	Spa	Pharmacy	Gym	Chinese Restaurant	Juice Bar	Ice Cream Shop	Jewelry Store	...	Street Food Gathering	Theater	Theme Restaurant	Multiplex	Optical Shop
name																
Hour Cafe	4	0.0	1.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0
Al Wizarat	4	0.0	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	...	0.0	0.0	0.0	0.0	0.0
Baraish Billard	4	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
Coffee Planet	4	6.3	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
Billiardo & Snooker Club	4	0.0	1.0	0.0	0.0	0.0	0.0	2.0	2.0	0.0	...	0.0	0.0	0.0	0.0	0.0

Figure 11 Venues in Cluster 4

	Cluster Labels	Ratings	Donut Shop	Spa	Pharmacy	Gym	Chinese Restaurant	Juice Bar	Ice Cream Shop	Jewelry Store	...	Street Food Gathering	Theater	Theme Restaurant	Multiplex	Optical Shop	Auto Workshop
name																	
Sands Leisure Lounge	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
al rimal billiards	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0

Figure 12 Venues in Cluster 5

	Ratings	Donut Shop	Spa	Pharmacy	Gym	Chinese Restaurant	Juice Bar	Ice Cream Shop	Jewelry Store	Lounge	...	Street Food Gathering	Theater	Theme Restaurant
Cluster Labels														
0	0.000000	1.375000	0.000000	0.500000	0.250000	0.000000	1.375000	1.125000	0.000000	0.000000	...	0.000000	0.000000	0.000000
1	1.340000	0.800000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	2.200000	0.800000	...	0.000000	0.000000	0.000000
2	3.215385	1.846154	0.230769	0.538462	0.846154	0.076923	0.769231	1.076923	0.076923	0.230769	...	0.076923	0.076923	0.076923
3	1.309091	2.090909	0.181818	0.818182	0.272727	0.000000	0.818182	1.363636	0.363636	0.000000	...	0.000000	0.000000	0.000000
4	1.890909	1.090909	0.090909	0.181818	0.181818	0.000000	1.000000	1.090909	0.000000	0.000000	...	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	...	0.000000	0.000000	0.000000

Figure 13 Mean Rating per Cluster

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Coffee Shop	15.0	27.0	77.0	33.0	43.0	0.0
Donut Shop	11.0	4.0	24.0	23.0	12.0	0.0
Dessert Shop	20.0	0.0	17.0	9.0	15.0	0.0
Ice Cream Shop	9.0	5.0	14.0	15.0	12.0	0.0
Breakfast Spot	3.0	4.0	14.0	22.0	7.0	0.0
...	...	...	...	...	...	...
Theme Park	2.0	0.0	0.0	0.0	0.0	0.0
Pie Shop	4.0	0.0	0.0	2.0	2.0	0.0
American Restaurant	1.0	0.0	0.0	0.0	0.0	0.0
Electronics Store	1.0	5.0	0.0	3.0	0.0	0.0
Egyptian Restaurant	0.0	1.0	0.0	1.0	0.0	0.0

Figure 14 Number of nearby venues for each cluster

Results and Discussion

Figure 13 shows us the mean rating for each cluster group. We can see that the cluster with the highest mean rating is cluster 2 followed by cluster 4. We then turn to Figure 14 to analyse why this is so. From Figure 14 we see that the clusters 2 and 4 have more confectionary shops close to the billiards cafes than most of the other clusters. They also have little to no restaurants, electronic stores etc. What we can gather from this is that when people go to play billiards they usually go after or before having some coffee or sweets and so an optimum position to open one's own billiards café would be next to these stores. Of course, we must also make sure that the location has less competition and is close to the city so it can be easily accessed. We can then have a look at the cluster map to see which location best suits these conditions. Looking at the map for light blue (Cluster 2) and light green (Cluster 4) circles which are close to the city, we see that the area around Al Samar Billiards and Playstation, Snooker Billiards and Snooker Cafe is an ideal area to build the billiards cafe as there is not much competition there. This are is shown in Figure 15.

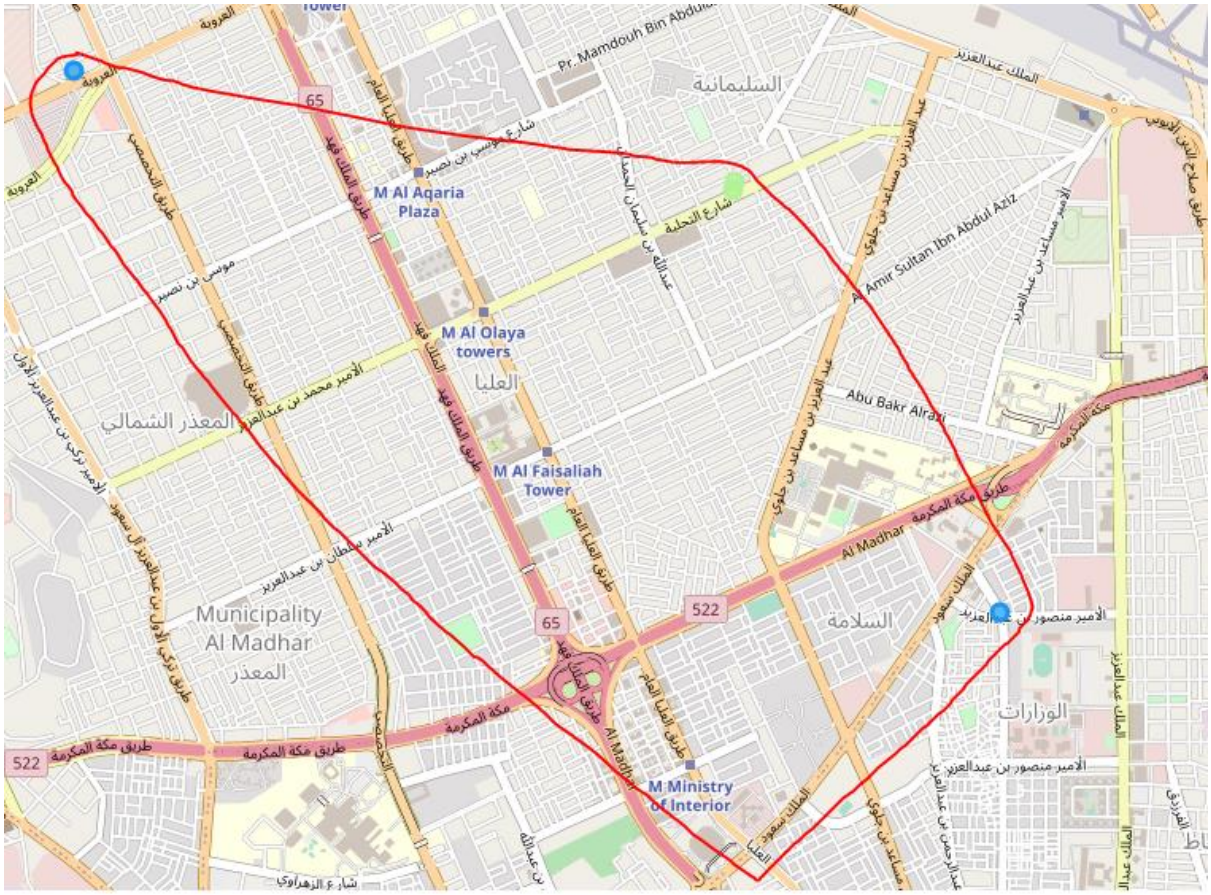

Figure 15 Optimum area to make a billiards cafe in Riyadh

Conclusion

The aim of this project was to build an optimum billiards café in Riyadh, Saudi Arabia. Given the limited data in Riyadh an optimum location was still found. This was done by leveraging the Foursquare API to find locations of billiards cafes in Riyadh and cluster them into groups using KMeans clustering.