**The battle of neighborhoods**

**Business problem**

Qatar is a country in the middle east where there are a lot of investors thinking of investing in new venues, deciding what project to invest in and in which location can be challenging since there are already many existing venues all around the country. the location and the category of the project are key factors when deciding for an investment. In this project, the neighborhoods in Doha, Qatar are clustered based on these factors to help investors make a decision.

**Data**

The data needed for this model is the names and locations of the neighborhoods in Doha, the names and categories of venues in each neighborhood.

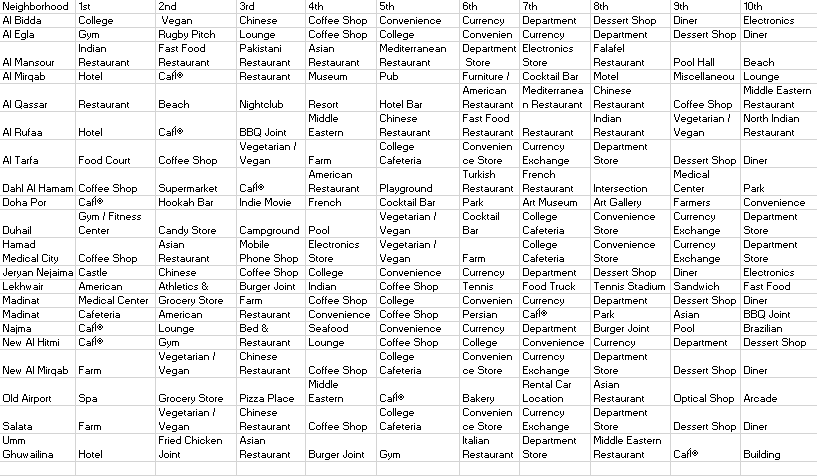
Data about neighborhoods location was collected manually from Wikipedia as it did not readily exist. The data about venues was taken from Foursquare API using the explore endpoint.

**Mythology**

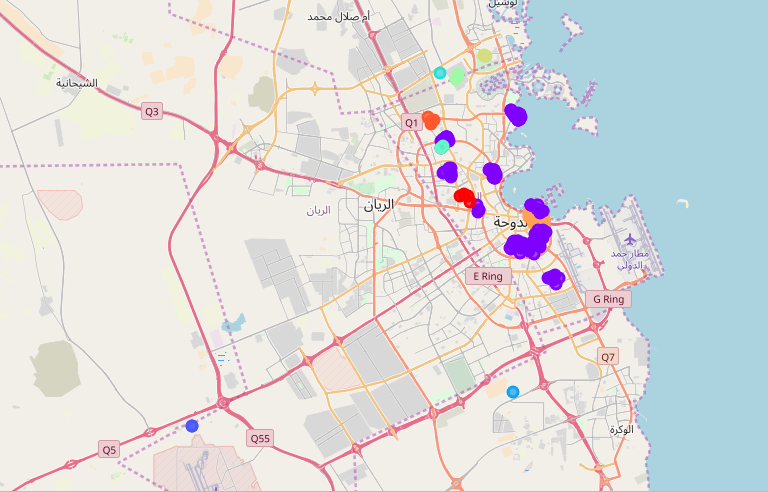
Neighborhoods locations was first used to get nearby venues from the Foursquare API, the data extracted was each venue name, category, id and location and the data was appended to a new Data frame next to the neighborhood name and location

Next, categories were one hot encoded for the ease of modeling, the data was grouped by the neighborhood and the occurrence for each category in neighborhood was calculated. The model used for clustering was kmeans with 10 clusters.

**Results** :the venues were grouped by the neighborhood and the following graph shows the number of different venues each neighborhood have :It can be seen from the graph that al rufaa had the highest number of venues(44)

The graph below shows the 10 most common venues for each neighborhood :

K means clustering was applied to the data set with number of clusters =10,the map below shows the different colors of clusters on the map of doha .



**Discussion**

Although location and category are important factors when it comes to deciding about investment, but the decision should not be limited to only these .as there are many other factors that could be added to this model to make the classification more useful and accurate. studying the nature of the economy and the population in the neighborhood are some examples of data that can be added to the model.

**Conclusion**

After clustering the data into 10 clusters , two were found to be quite larger than the other 8 clusters as was shown in the jupyter notebook, this suggests that these two clusters has the most venues in doha that are similar to each other and that the rest are few with great dissimilarities.