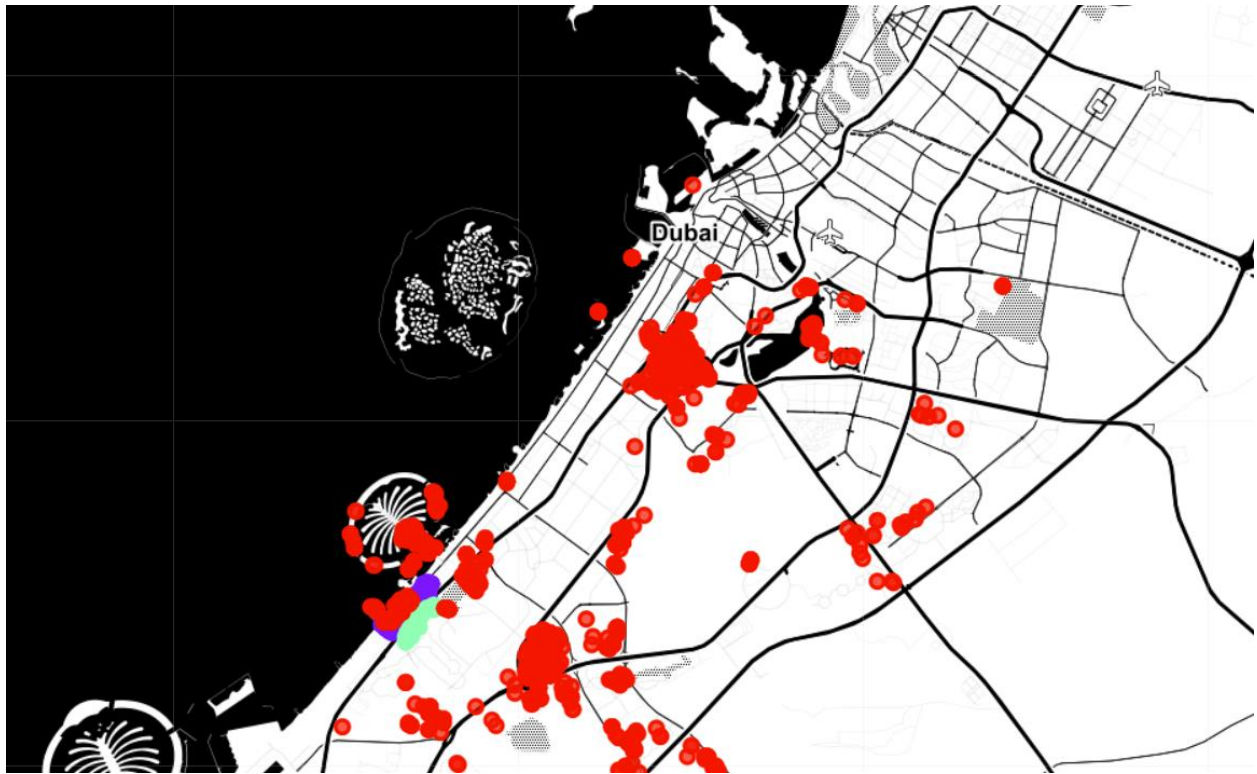# Capstone Project
The Battle of Neighbourhoods (Week 2)

Date of submission: 20th February 2021

Background

Dubai is a real estate hub, probably the only place in the word where you can get a studio apartment that comes with its own swimming pool and tennis court. In short, properties of all sizes and facilities are available.

Business Problem

And having such large options makes its really difficult to choose the right ones that fits your residential needs or investment plans. A solution is required that could help stakeholders in pinning down neighbourhoods based over the facilities and amenities they provide and also understand any possible correlation between certain amenities and property value.

Target Audience

1. People and investors planning to rent or buy properties
2. Real estate dealers

Description of The Data

To resolve our problem, we need data of different properties across Dubai, their prices and amenities and facilities near to it.

Source of The Data

1. List of properties:

The data is scraped from the real estate portal and it is anonymized. It consists of more than 1900+ properties containing 38 features. Its usage is licensed under CC0: Public Domain.

The data contains over 1900 records along with their geolocation and their prices. Some other information is also part of the dataset but not applicable to our use.

Link to dataset: https://www.kaggle.com/dataregress/dubai-properties-dataset

2. Amenities and finicalities next to properties:

We will utilize Foursquare API for scrap list of facilities and venues next to our properties, using their location part of our dataset.

Methodology

First, we will import our dataset in a CSV format, since the dataset contains a lot of other information that isn't of our use so we would need to filter out the unnecessary information and focus on features which can be helpful to address our problems.

Than we have sorted the data according to the prices and assigned a rank to each property to represent its value, lower the rank higher that value of property in price. Then these properties are mapped on Dubai's map using the geo coordinates of the dataset so better visualize what we are working with.
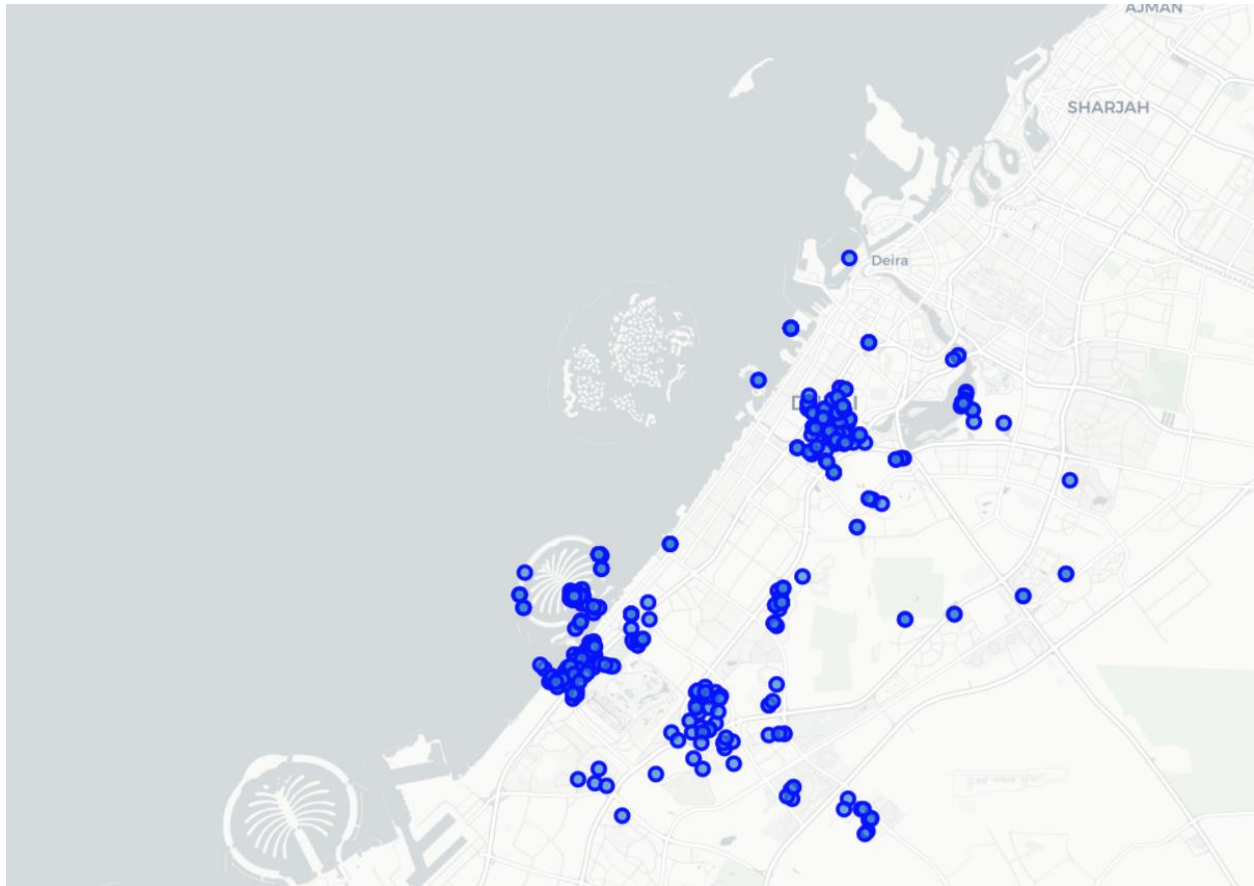
Next, we will use utilizes Foursquare's API venue endpoint to retrieve a list of functionalities surrounding our properties.

Applying one hot encoding to all retrieved property and take mean for each respective to each neighbourhood.

Next, we will sort the resulting dataset according to viability of a particular facility that the user wants to target and is interested in.

In the end we will apply K-means clustering to divide the data (our neighbourhoods) in the chunks of neighbourhoods having high concentration of our category to neighbourhoods with the lowest. Visualizing each cluster on the map in the end.

Since, we will be using k=3 for this example we will get three resulting clusters, mapped in different colours over the map.

Observations

Over individually analysing each cluster. We can see that Cluster -2 contains the neighbourhoods with highest concentrations of the target category which in our case are 'Gym'(s), the neighbourhood on top in this category is 'Al Barari', similarly in opposite Cluster -0 contains the neighbourhood with least gyms available in the surrounding of a property such as 'Al Sufouh'.

However, no specific correlation between the existence of gyms nearby and prices of a property were observed.
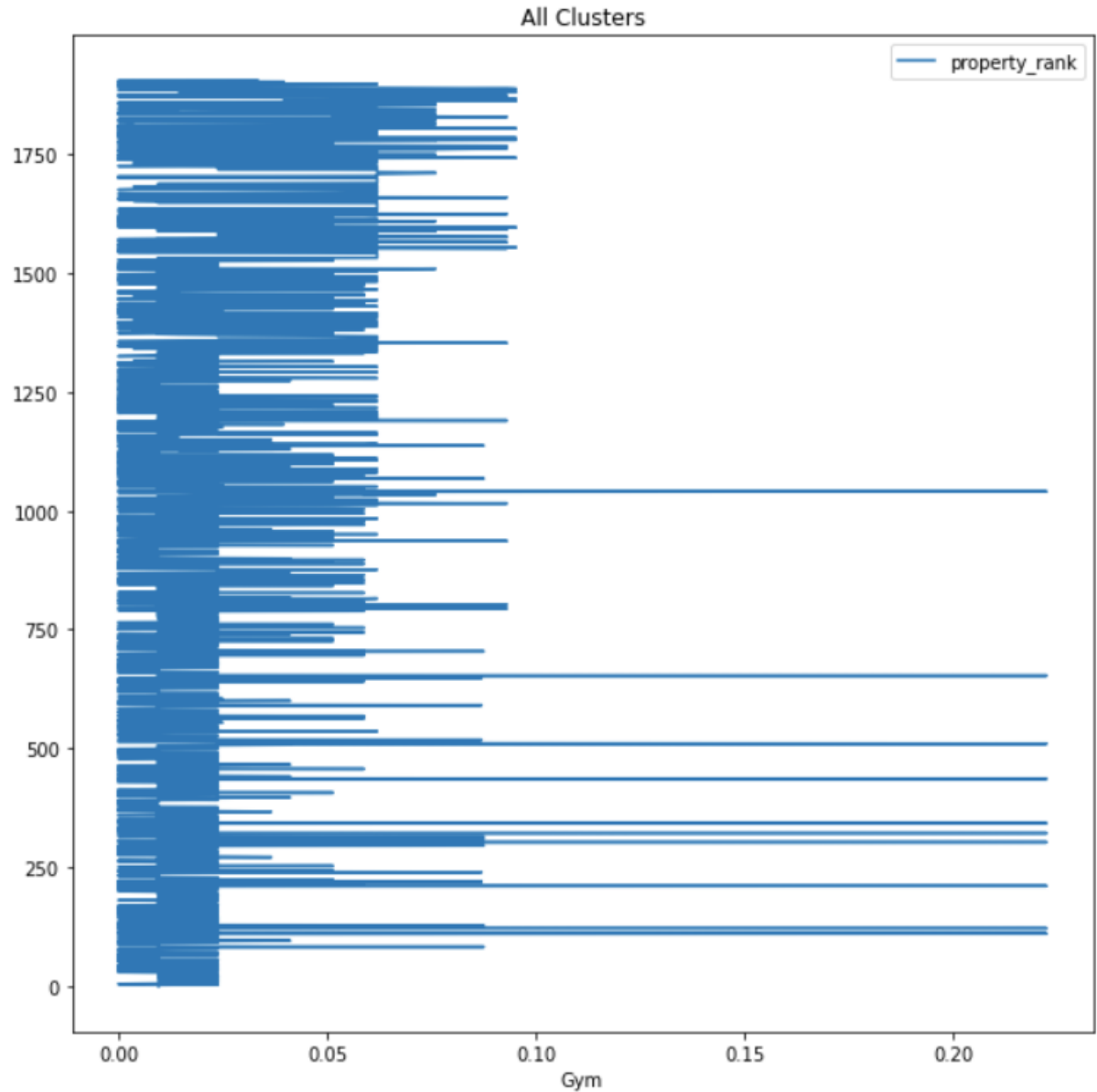
Results

It has been found and analysed that in order to own or rent a property that has feasible gym accessibility, the right neighbourhood to live in is *Al Barari*.

Conclusion

During this project, we worked through a already existing dataset to work our way through the solution of categorising the properties and neighbourhood as per the preferred target category, which in our case was 'Gym'. We did geolocation and clustering to divided out dataset and information accordingly to nearby facilities and their concentrations and finally found results of most ideal and worst neighbourhood for particular facility.