

Capstone Project



Location Selection Restaurant
COURSERA IBM Data Science

July 2020

TABLE OF CONTENTS

| | | |
|-----|---|----|
| 1. | INTRODUCTION | 3 |
| 2. | PROBLEM DESCRIPTION | 4 |
| 3. | DATA | 5 |
| 3.1 | Toronto Geolocatoin:..... | 5 |
| 3.2 | Dubai Geolocatoin:..... | 6 |
| 3.3 | Categories of Toronto Venues..... | 6 |
| 3.4 | Categories of Dubai Venues | 7 |
| 3.5 | Compare Subcategories..... | 7 |
| 3.6 | Venues of Dubai..... | 7 |
| 4. | METHODOLOGY | 10 |
| 4.1 | Heat map to define the locations of resturants in Dubai | 11 |
| 4.2 | Heat map to define the locations of Cafés in Dubai..... | 11 |
| 4.3 | Define the coordinate of the area with the most number of resturants using the Machine Learning clustering | 12 |
| 4.4 | Define the coordinate of the area with the less number of Café resturants using the Machine Learning clustering | 13 |
| 5. | RESULT..... | 14 |
| 5.1 | Display both locations on the map: | 14 |
| 6. | DISCUSSIONS | 15 |
| 7. | CONCLUSION..... | 16 |
| | ANNEX 1 – REFERENCES | 17 |

1. Introduction

This project will be used to evaluate and compare Toronto with Dubai cities to check the diversity between both eastern and western culture in terms of venues categories. We aim to suggest a venue categories and the most suitable for this location.

2. Problem Description

Most of the individuals or entities who aim to launch a new business will face a lot of challenges to gain revenue at start up. There are many reasons behind this problem as there will be many choices that must be made between more alternatives ideas that seem to be equally and at the same time they will need to invest and look for a return (ROI) in a short period of time. The problem occurs when the business is selected without a proper market survey to check the effectiveness of the ideas and whether it will lead to getting a good market share in the desired time based on the feasibility study.

Methodology

We will use both Toronto and Dubai cities to list the categories that can help new business owners to take a decision as follows:

1. Download and Explore both Toronto & Dubai Locations Datasets
2. Explore Neighborhoods in both cities
3. Cluster the venues in Dubai and suggest new business based on the bottom 5 categories
4. Check missing category in Dubai and suggest top 5 based on Toronto trending venues dataset

3. Data

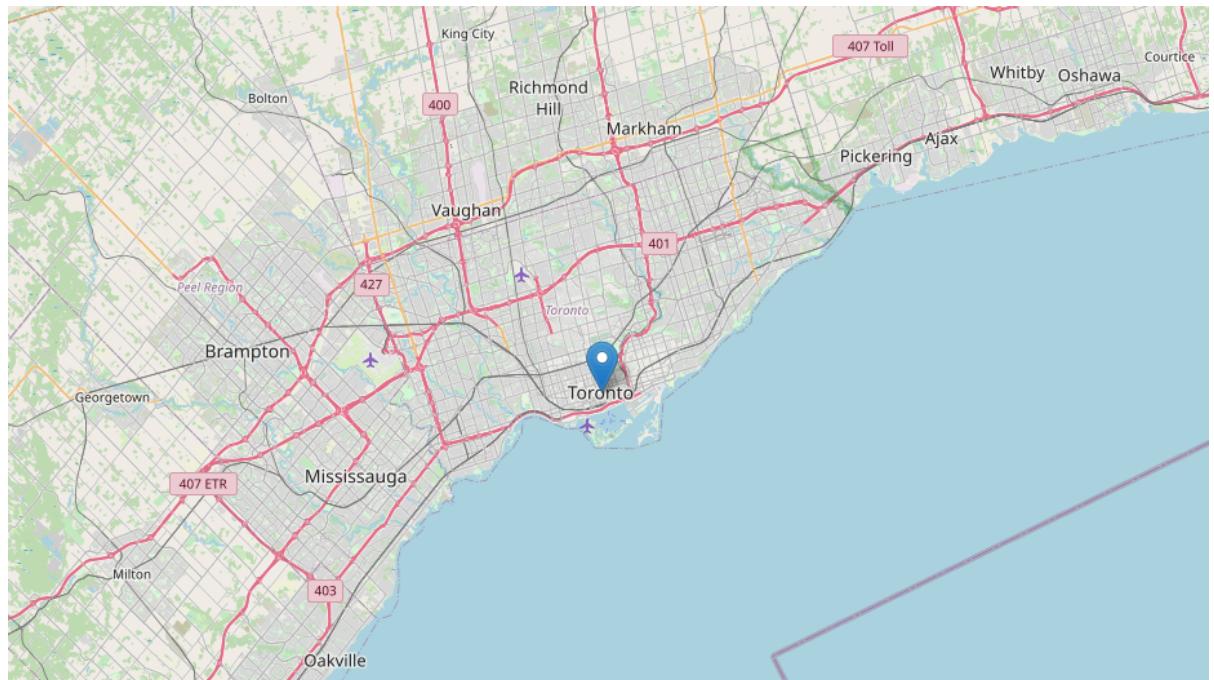
This report will use Google Geolocation and Foursquare APIs.

Foursquare API uses latitudes and longitudes to identify the locations in the REST API calls. So, as a first step we will have to get the location from Google Geolocation API to be used in all the subsequent calls of foursquare APIs.

3.1 Toronto Geolocation:

Coordinate of Toronto, Ontario, CA: [43.653226, -79.3831843]

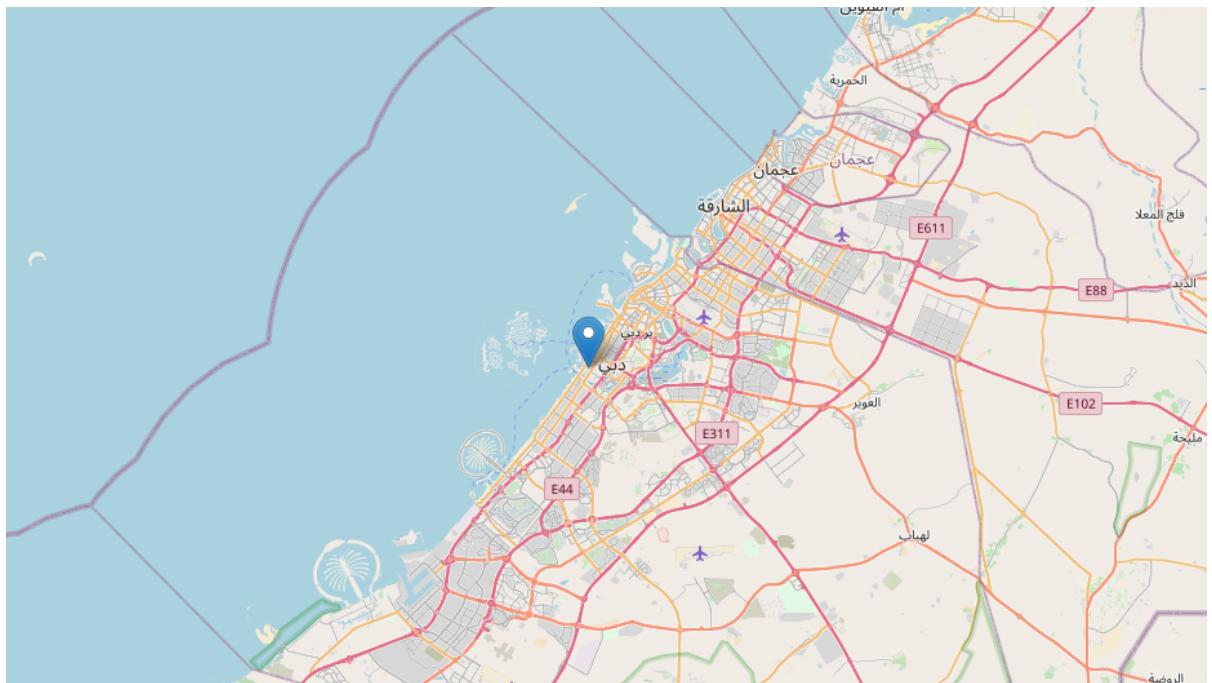
It can be displayed on the map as follows:



3.2 Dubai Geolocation:

Coordinate of Jumeirah, Dubai, UAE: [25.2016428, 55.2452567]

It can be displayed on the map as follows:



3.3 Categories of Toronto Venues

“The Foursquare Places API offers real-time access to Foursquare’s global database of rich venue data and user content to power your location-based experiences in your app or website”. According to the service description from foursquare website, we will use the Places API to extract the categories of all venues in Toronto city. Below is the summary and head section of the data frame (using python pandas library):

There are " 451 " Subcategories in Toronto.

451

| | id | Category | Subcategory | pluralName | shortName |
|---|--------------------------|----------------------|--------------------|-------------------|------------------|
| 0 | 56aa371be4b08b9a8d5734db | Arts & Entertainment | Amphitheater | Amphitheaters | Amphitheater |
| 1 | 4fceea171983d5d06c3e9823 | Arts & Entertainment | Aquarium | Aquariums | Aquarium |
| 2 | 4bf58dd8d48988d1e1931735 | Arts & Entertainment | Arcade | Arcades | Arcade |
| 3 | 4bf58dd8d48988d1e2931735 | Arts & Entertainment | Art Gallery | Art Galleries | Art Gallery |
| 4 | 4bf58dd8d48988d1e4931735 | Arts & Entertainment | Bowling Alley | Bowling Alleys | Bowling Alley |

3.4 Categories of Dubai Venues

Below is the summary and head section of the data frame for Dubai (using python pandas library):

| There are " 448 " Subcategories in Dubai | | | | | |
|--|--------------------------|----------------------|--------------------|-------------------|------------------|
| | id | Category | Subcategory | pluralName | shortName |
| 0 | 56aa371be4b08b9a8d5734db | Arts & Entertainment | Amphitheater | Amphitheaters | Amphitheater |
| 1 | 4fceea171983d5d06c3e9823 | Arts & Entertainment | Aquarium | Aquariums | Aquarium |
| 2 | 4bf58dd8d48988d1e1931735 | Arts & Entertainment | Arcade | Arcades | Arcade |
| 3 | 4bf58dd8d48988d1e2931735 | Arts & Entertainment | Art Gallery | Art Galleries | Art Gallery |
| 4 | 4bf58dd8d48988d1e4931735 | Arts & Entertainment | Bowling Alley | Bowling Alleys | Bowling Alley |

3.5 Compare Subcategories

Let's compare both subcategories using python method. It will show three subcategories that are not available in Dubai city. Only two of them can be used Dubai due to local regulations:

- Mac & Cheese Joint
- Poutine Place
- Marijuana Dispensary

Since both of these subcategories are like fast food and mainly to offer French fries, we will have to check all venues in Jumeirah Area to find out the best place where we can start this new business. The new idea is to offer this type of food as item in Café menu better than a fast food venue. So, we will target the areas with a smaller number of Café venue in Jumeirah area.

3.6 Venues of Dubai

Using Places API, we will extract venues using below criteria:

- All venues around Jumeirah in radius of 10KM *
- All venues with subcategory Café venues around Jumeirah in radius of 10KM

* Limit of 100 venues for demonstration only.

3.6.1 All Venues

Display the head section of the data frame using python pandas library:

| | name | lat | lng |
|---|-------------------|-----------|-----------|
| 0 | Coya | 25.202581 | 55.240430 |
| 1 | Scalini | 25.202483 | 55.240503 |
| 2 | Saddle Dubai | 25.204722 | 55.252573 |
| 3 | 3 Fils Restaurant | 25.210281 | 55.243483 |
| 4 | Nusr-Et (نصرت) | 25.202223 | 55.240285 |

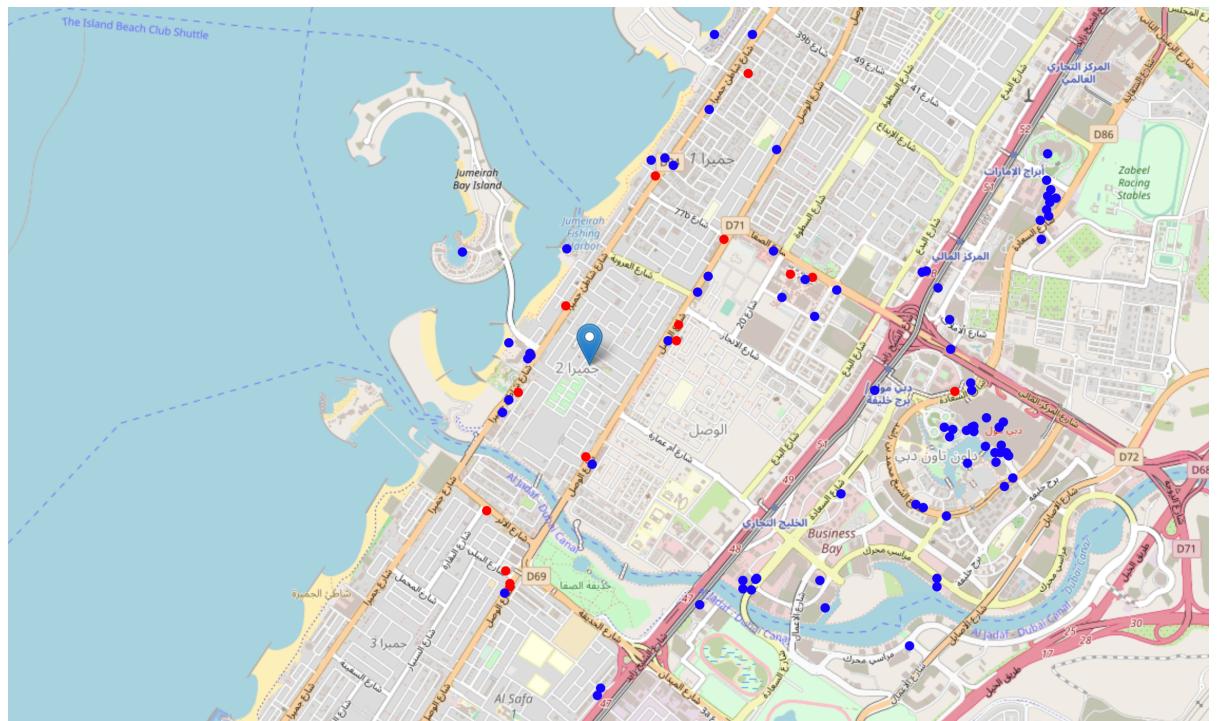
3.6.1 Café Venues

Display the head section of the data frame using python pandas library:

| | name | lat | lng |
|---|--------------------------------|-----------|-----------|
| 0 | Saddle Dubai | 25.204722 | 55.252573 |
| 1 | Alchemy Coffee | 25.195002 | 55.244968 |
| 2 | Al Falamanki (الفلمنكي) | 25.199752 | 55.239512 |
| 3 | RX Coffee Apothecary & Kitchen | 25.191058 | 55.236944 |
| 4 | Mikel Coffee Company | 25.206137 | 55.243395 |

3.6.1 Visualise both data frames on Dubai map

Data visualization is very important in Machine Learning, so using python Folium library we will display all restaurants in Dubai and will use a different color for Café venues.



4. Methodology

There are many libraries in python libraries are ready to be used by data scientists that will save a lot of time to work on data analysis and build the accurate models.

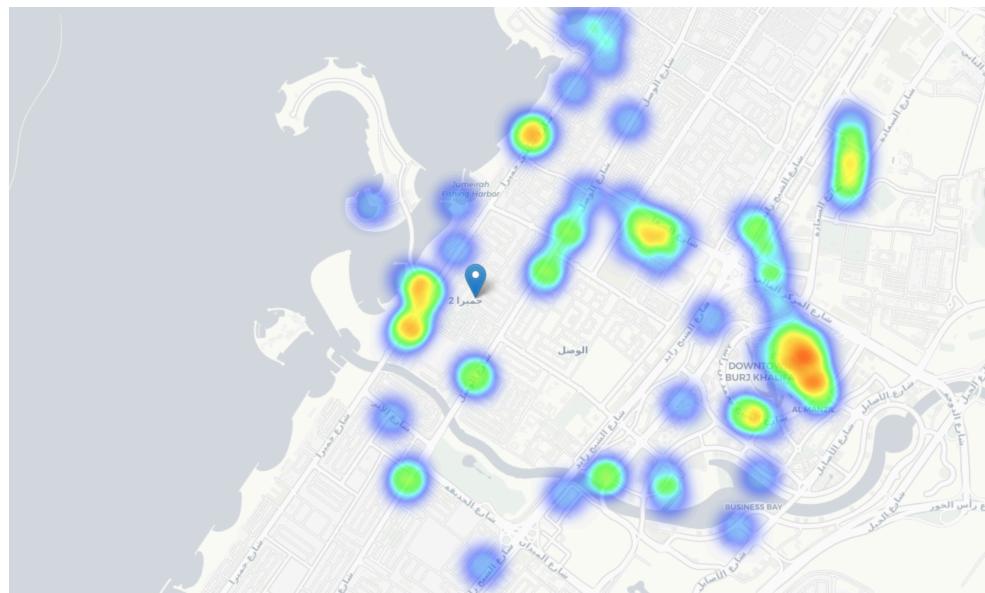
We will use unsupervised Machine Learning Model, and one of the useful techniques is clustering which will help us identify the locations and do some analysis to select the most suitable location for the new Café. In this study, and for simplicity, we will use locations latitudes and longitudes only to define the clustering centre points, however, we can use many other attributes and data frames to get more accurate results i.e. populations, interests, customer feedback, customer check in information.... etc.

Simply, we will use below methodology to define our suggested location:

- Use heat map to define the locations with the greatest number of restaurants
- Use heat map to define the locations with a smaller number of cafés
- Define the coordinate of the area with the greatest number of restaurants using the Machine Learning clustering (KMean)
- Define the coordinate of the area with the smaller number of cafés using the Machine Learning clustering (KMean)
- Display the suggested area using the intersections of two circle on Dubai map

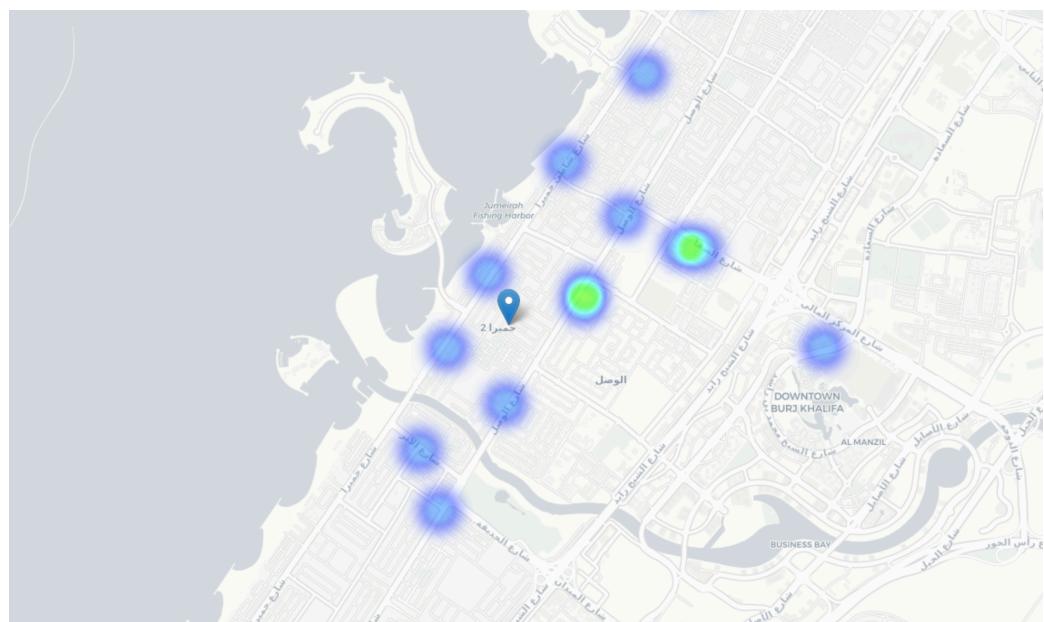
4.1 Heat map to define the locations of restaurants in Dubai

Using python Folium library will display the distribution of all restaurants in Dubai map. As expected, down town (next to Burj Khalifa) is the area with the greatest number of restaurants.



4.2 Heat map to define the locations of Cafés in Dubai

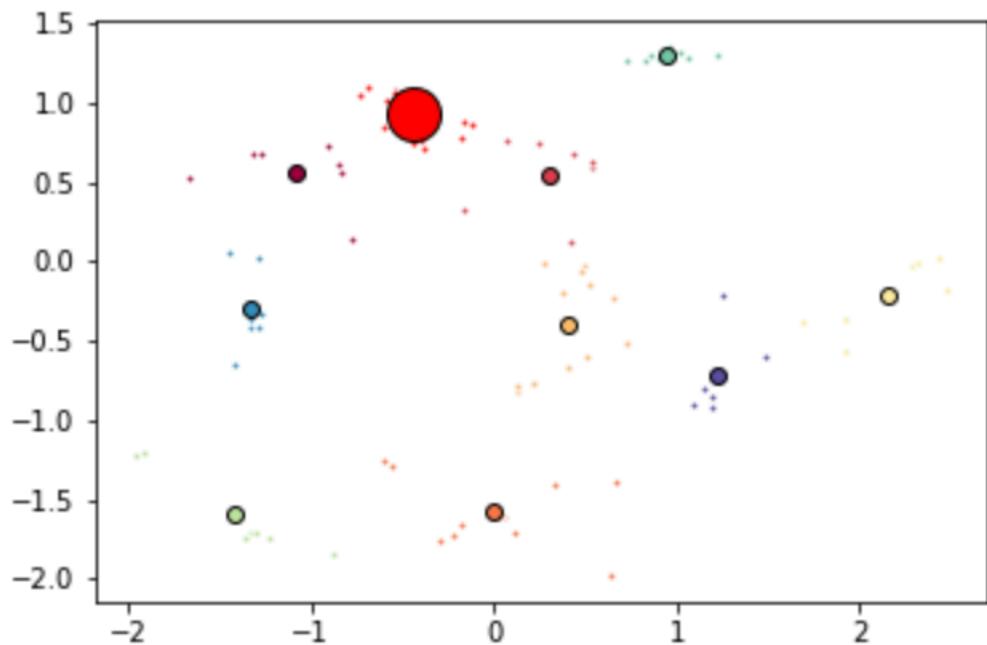
Again, using python Folium library will display the distribution of Café restaurants only in Dubai map.



4.3 Define the coordinate of the area with the greatest number of restaurants using the Machine Learning clustering

Using python sklearn library will train the model using the coordinates information we prepared in the data section to get the coordinate of the area (cluster) with the greatest number of restaurants.

We will use 10 clusters & 12 iterations which will help us to get more accurate locations from the algorithm then display the cluster and the centroids points as per below:

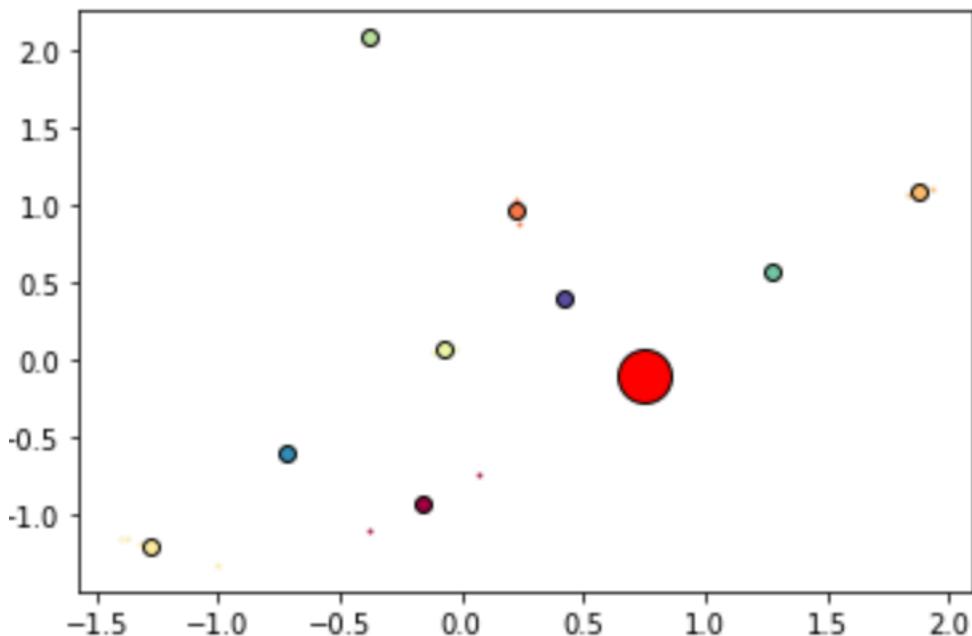


Now we identified the cluster number, we will can get the average of the coordinates in this cluster to define the coordinates of the suggested location as (Lat: 25.1964, Lnt: 55.2773).

4.4 Define the coordinate of the area with the smaller number of Café restaurants using the Machine Learning clustering

Again, using python sklearn library will train the model using the coordinates information we prepared in the data section to get the coordinate of the area (cluster) with the smaller number of restaurants.

We will use 10 clusters & 12 iterations which will help us to get more accurate locations from the algorithm then display the cluster and the centroids points as per below:

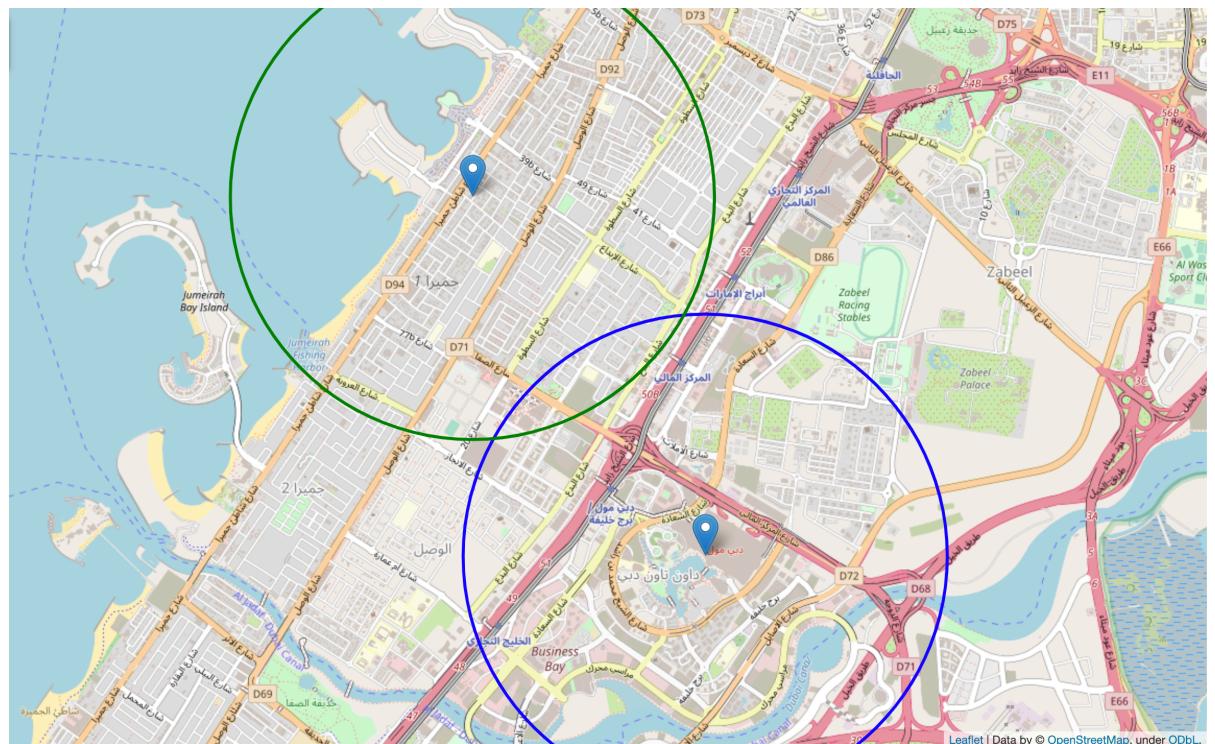


Now we identified the cluster number, we will can get the average of the coordinates in this cluster to define the coordinates of the suggested location as (Lat: 25.2156, Lnt: 55.2507).

5. Result

5.1 Display both locations on the map:

Using python Folium library will display both locations on Dubai map.



6. Discussions

We could identify the two locations using Machine Learning clustering technique. The Places API data were very useful to identify the venues and categories in both cities. Also, python saved a lot of time to cluster the data and get the coordinates of the two locations. Also, we used python to visualize the results of the methodologies to identify the suggested area where the business owner can open the new Café restaurant.

Data is important and the more data we get from trusted source the more accurate results we can get. Although, the data sets are small compared to real life scenarios, but it is good to demonstrate the methodology and the capabilities of Machine Learning.

7. Conclusion

According to the results section, it is suggested to open this new venue in down town e.g. Dubai Mall. In real life scenario, we can add other data sets to get another location based on other attributes as explained in Data section.

Annex 1 – References

- [Places API](#)
- [Google API](#)