Capstone Project - The Battle of the Neighborhoods (Week 2)

Indian Restaurants In Al Karama Neighborhoods

Table of contents

Introduction: Business Problem

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an **Indian restaurant** in **Karama**, Dubai, U.A.E.

Since there are lots of restaurants in Karama, Dubai, we will try to detect **locations that are not already crowded with restaurants**. We are also particularly interested in **areas with no Indian restaurants in vicinity**. We would also prefer locations **as close to city center as possible**, assuming that first two conditions are met.

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Data

Based on definition of our problem, factors that will influence our decision are:

- number of existing restaurants in the neighborhood (any type of restaurant)
- number of and distance to Indian restaurants in the neighborhood, if any
- distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered on city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Google Maps API reverse geocoding**
- number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**
- coordinate of Dubai center will be obtained using **Google Maps API geocoding** of well-known Dubai location (Karama)

Neighborhood Candidates

Let's create latitude & longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is approx. 12x12 kilometers centered on Al Karama, Dubai.

Let's first find the latitude & longitude of Al Karama, Dubai, using specific, well known address and Google Maps geocoding API.

Now let's create a grid of area candidates, equally spaced, centered around city center and within ~6km from Al Karama. Our neighborhoods will be defined as circular areas with a radius of 300 meters, so our neighborhood centers will be 600 meters apart.

To accurately calculate distances we need to create our grid of locations in Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we'll project those coordinates back to latitude/longitude degrees to be shown on Folium map. So let's create functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

Let's create a **hexagonal grid of cells**: we offset every other row, and adjust vertical row spacing so that **every cell center is equally distant from all it's neighbors**.

364 candidate neighborhood centers generated.

Let's visualize the data we have so far: city center location and candidate neighborhood centers:

OK, we now have the coordinates of centers of neighborhoods/areas to be evaluated, equally spaced (distance from every point to its neighbors is exactly the same) and within ~6km from Al Karama.

<u>Foursquare</u>

Now that we have our location candidates, let's use Foursquare API to get info on restaurants in each neighborhood.

We're interested in venues in 'food' category, but only those that are proper restaurants - coffee shops, pizza places, bakeries etc. are not direct competitors so we don't care about those. So we will include in our list only venues that have 'restaurant' in category name, and we'll make sure to detect and

include all the subcategories of specific 'Indian restaurant' category, as we need info on Indian restaurants in the neighborhood.

Let's now see all the collected restaurants in our area of interest on map, and let's also show Indian restaurants in different color.

Looking good. So now we have all the restaurants in area within few kilometers from Karama Plaza, and we know which ones are Indian restaurants! We also know which restaurants exactly are in vicinity of every neighborhood candidate center.

This concludes the data gathering phase - we're now ready to use this data for analysis to produce the report on optimal locations for a new Indian restaurant!

Methodology

In this project we will direct our efforts on detecting areas of Karama, Dubai that have low restaurant density, particularly those with low number of Indian restaurants. We will limit our analysis to area ~6km around city center.

In first step we have collected the required **data: location and type (category) of every restaurant within 6km from Karama, Dubai** (Karama Plaza). We have also **identified Indian restaurants** (according to foursquare categorization).

Second step in our analysis will be calculation and exploration of '**restaurant density**' across different areas of Karama, Dubai - we will use **heat maps** to identify a few promising areas close to center with low number of restaurants in general (*and* no Indian restaurants in vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements** established in discussion with stakeholders: we will take into consideration locations with **no more than two restaurants in radius of 250 meters**, and we want locations **without Indian restaurants in radius of 400 meters**. We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

<u>Analysis</u>

Let's perform some basic explanatory data analysis and derive some additional info from our raw data.

First let's count the **number of restaurants in every area candidate**:

Average number of restaurants in every area with radius=300m: 3.3324175824175826

OK, now let's calculate the **distance to nearest Indian restaurant from every area candidate center** (not only those within 300m - we want distance to closest one, regardless of how distant it is).

Average distance to closest Indian restaurant from each area center: 773.2144956740684

OK, so **on average Indian restaurant can be found within ~500m** from every area center candidate. That's fairly close, so we need to filter our areas carefully!

Let's create a map showing **heat map / density of restaurants** and try to extract some meaningfull info from that. Also, let's show **borders of Dubai boroughs** on our map and a few circles indicating distance of 1km, 2km and 3km from Karama.
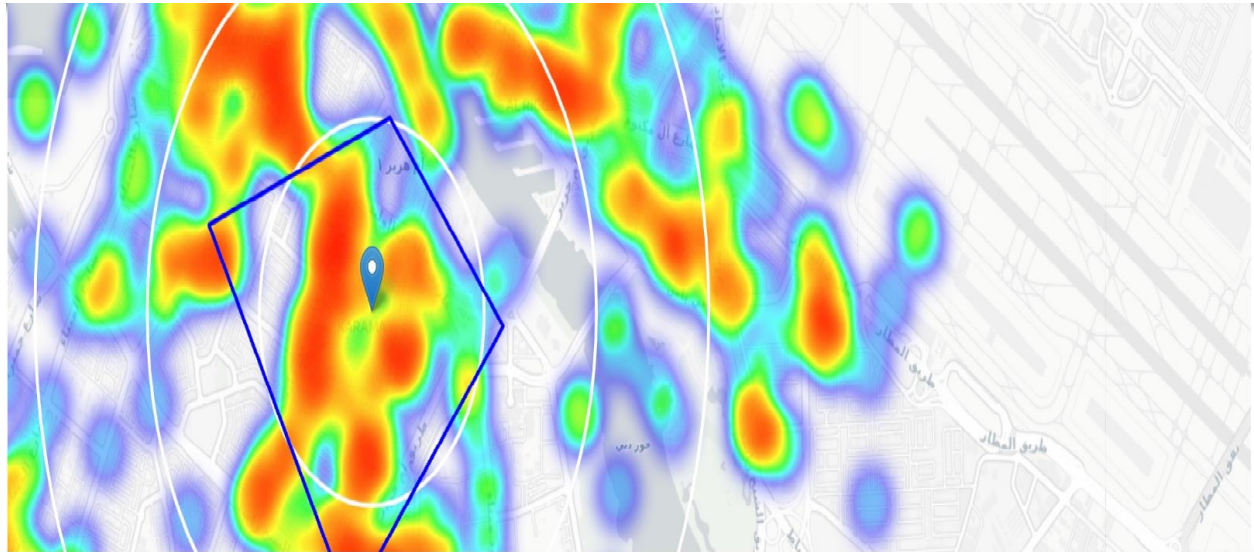

Fig 2 Heat Map -Resturants Density in Al Karama

Looks like a few pockets of low restaurant density closest to karama plaza can be found **south, south-east and east from karama**.

Let's create another heatmap map showing **heatmap/density of Indian restaurants** only.
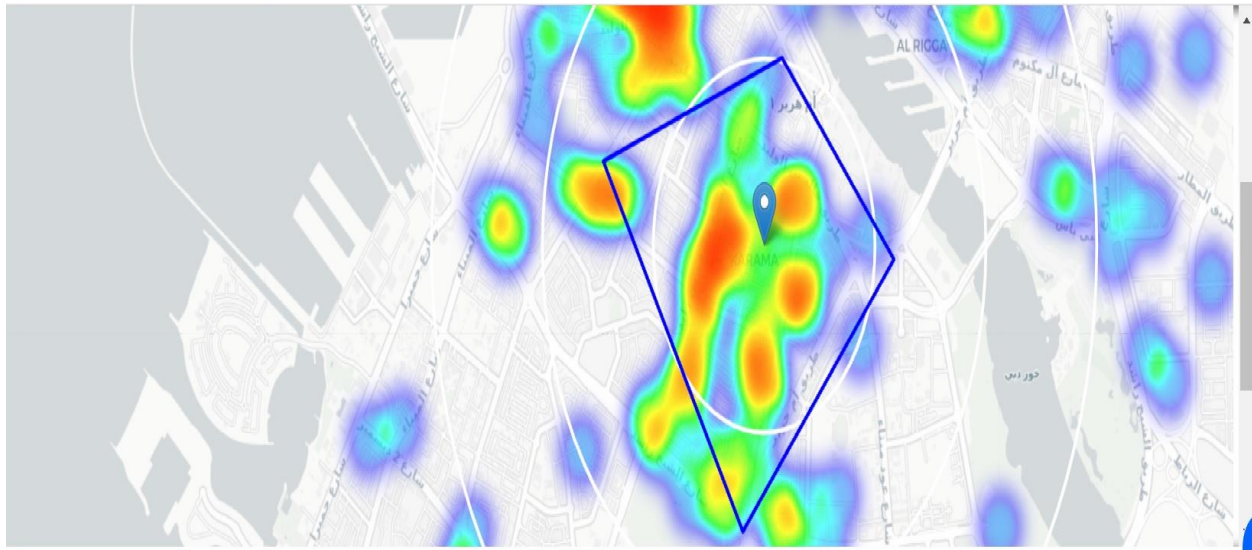

Fig 3 Heat Map- Indian Restuarant Density

This map is not so 'hot' (Indian restaurants represent a subset of ~20% of all restaurants in Karama, Dubai) but it also indicates higher density of existing Indian restaurants directly South and west from Karama, with closest pockets of **low Indian restaurant density positioned east, south-east and south from Karama**.
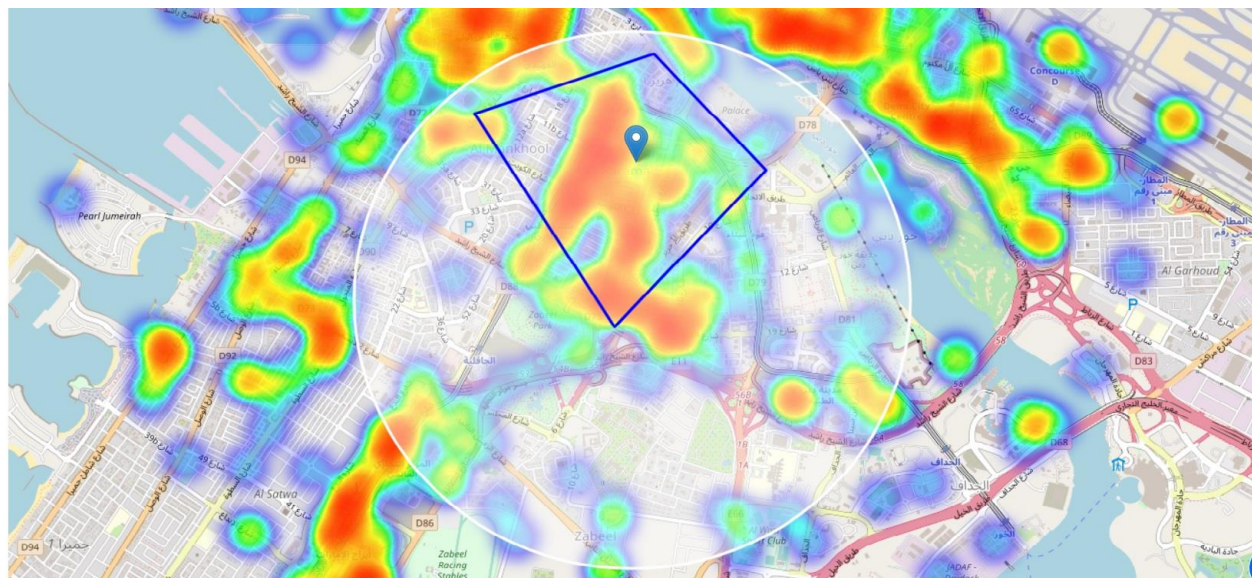

Fig 4 Heat Map - Resturants Density in Neighbourhood of Al Karama

Based on this we will now focus our analysis on areas *south-west, south, south-east and east from Al*

Karama - we will move the center of our area of interest and reduce it's size to have a radius of **2.5km**. This places our location candidates mostly in boroughs **Al Rigga and Al Mankool**

Al Rigga and Al Mankool

Analysis of popular travel guides and web sites often mention Al Rigga and Al Mankool as beautiful, interesting, rich with culture, 'hip' and 'cool' Dubai neighborhoods popular with tourists and loved by Dubaiers.

"Bold and brazen, Dubai's creative people, places, and spaces might challenge your paradigm." Tags: Nightlife, Artsy, Dining, Trendy, Loved by "Dubai"ers, Great Transit (airbnb.com)

"Al Rigga has long been revered for its diverse cultural life and as a part of Dubai where alternative lifestyles have flourished. Envisioning the glamorous yet gritty nature of Dubai often conjures up scenes from this neighborhood, where cultures, movements and artistic flare adorn the walls of building and fills the air. Brimming with nightclubs, street food, and art galleries, Al Rigga is the place to be for Dubai's young and trendy." (theculturetrip.com)

"Imagine an art gallery turned inside out and you'll begin to envision Al Mankool. Single walls aren't canvases for creative works, entire buildings are canvases. This zealously expressive east Dubai neighborhood forgoes social norms" Tags: Artsy, Nightlife, Trendy, Dining, Touristy, Shopping, Great Transit, and Loved by "Dubai"ers (airbnb.com)

Popular with tourists, alternative and bohemian but booming and trendy, relatively close to city center and well connected, those boroughs appear to justify further analysis.

Let's define new, narrower region of interest, which will include low-restaurant-count parts of Al Rigga and Al Mankool closest to Al Karama.
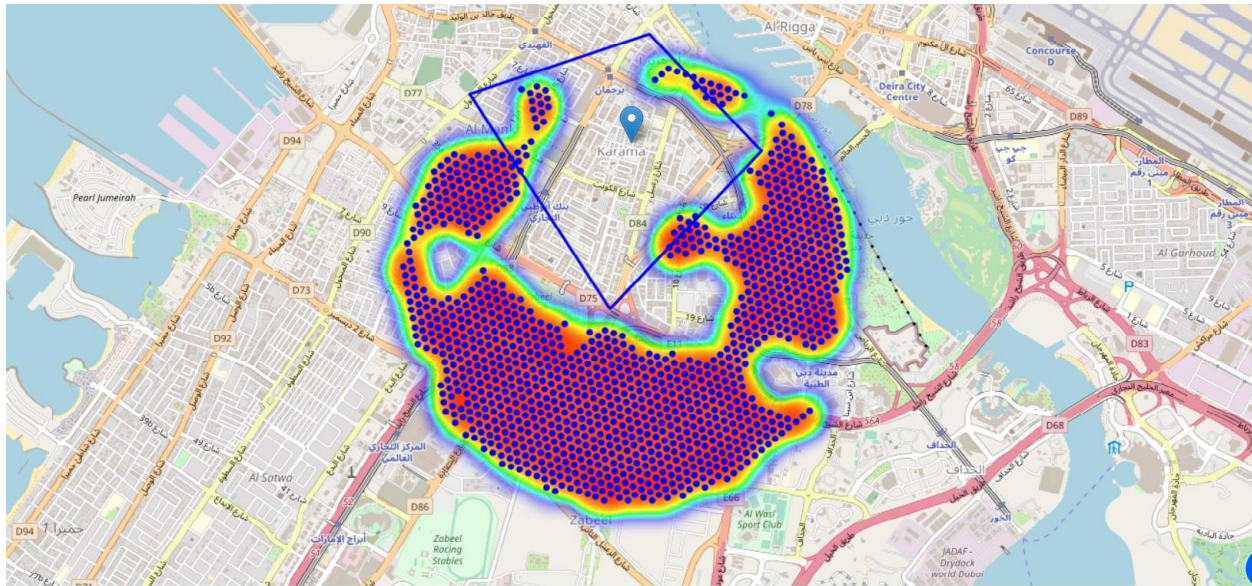
Fig 5 Zero Indian Restaurants Zone on the neighbourhood of Al karama

Not bad - this nicely covers all the pockets of low restaurant density in Al Rigga and Al Mankool closest to Karama.

Let's also create new, more dense grid of location candidates restricted to our new region of interest (let's make our location candidates 100m apart).
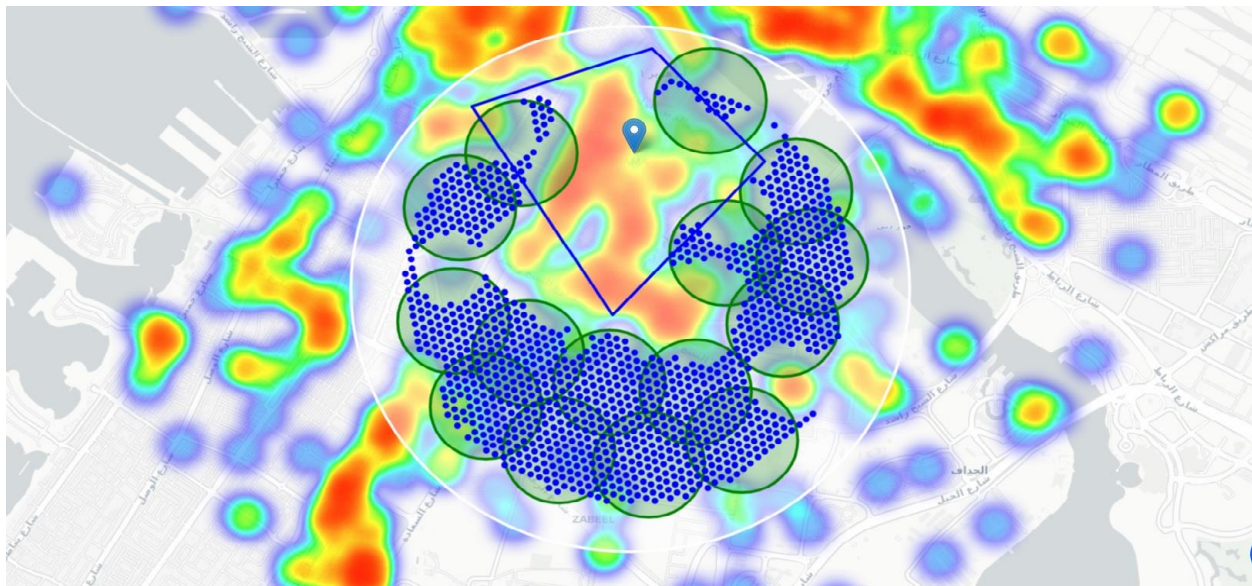

Fig 6 Kmeans Clustering of Indian Resturants in Al Karama Neighborhood

OK. Now let's calculate two most important things for each location candidate: **number of restaurants in vicinity** (we'll use radius of **250 meters**) and **distance to closest Indian restaurant**.

We now have a bunch of locations fairly close to Karama (mostly in Al Mankool, Al Rigga and south-east corner of neighborhood boroughs), and we know that each of those locations has no more than two restaurants in radius of 250m, and no Indian restaurant closer than 400m. Any of those locations is a potential candidate for a new Indian restaurant, at least based on nearby competition.

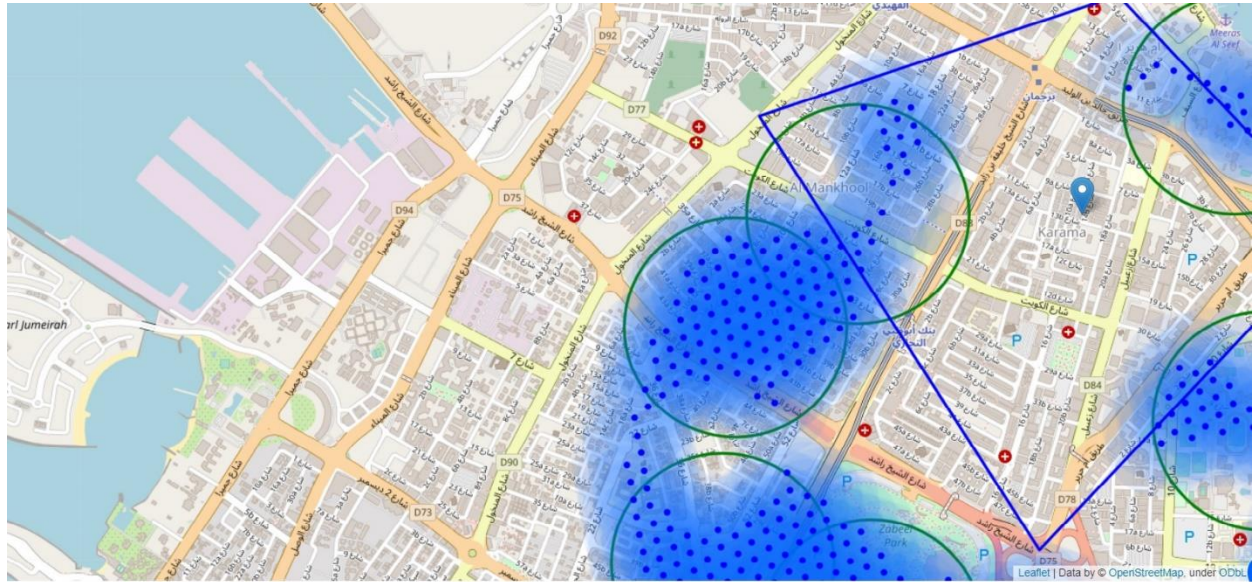Let's now show those good locations in a form of heatmap:

Fig 8 Al Mankool Neighbourhood of Al Karama in Clustering Map

Looking good. What we have now is a clear indication of zones with low number of restaurants in vicinity, and *no* Indian restaurants at all nearby.

Let us now **cluster** those locations to create **centers of zones containing good locations**. Those zones, their centers and addresses will be the final result of our analysis.

Fig 7 Clustering without Heat map
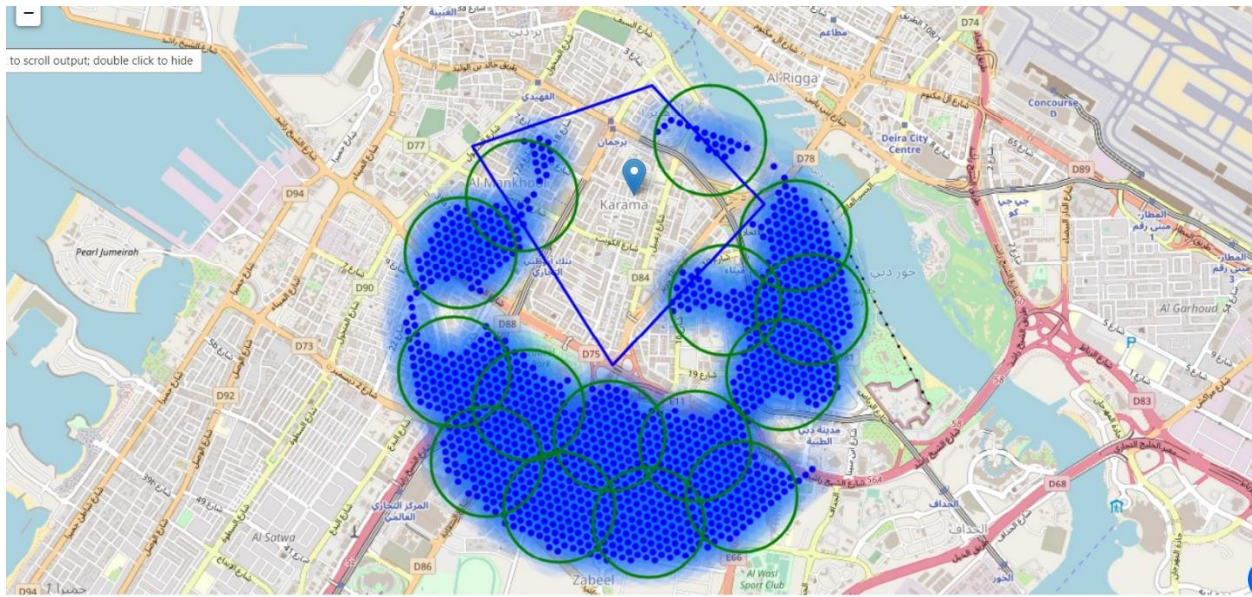
Not bad - our clusters represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates.

Fig 10 Suitable location for Indian Resturants in AL Karama Neighbourhood
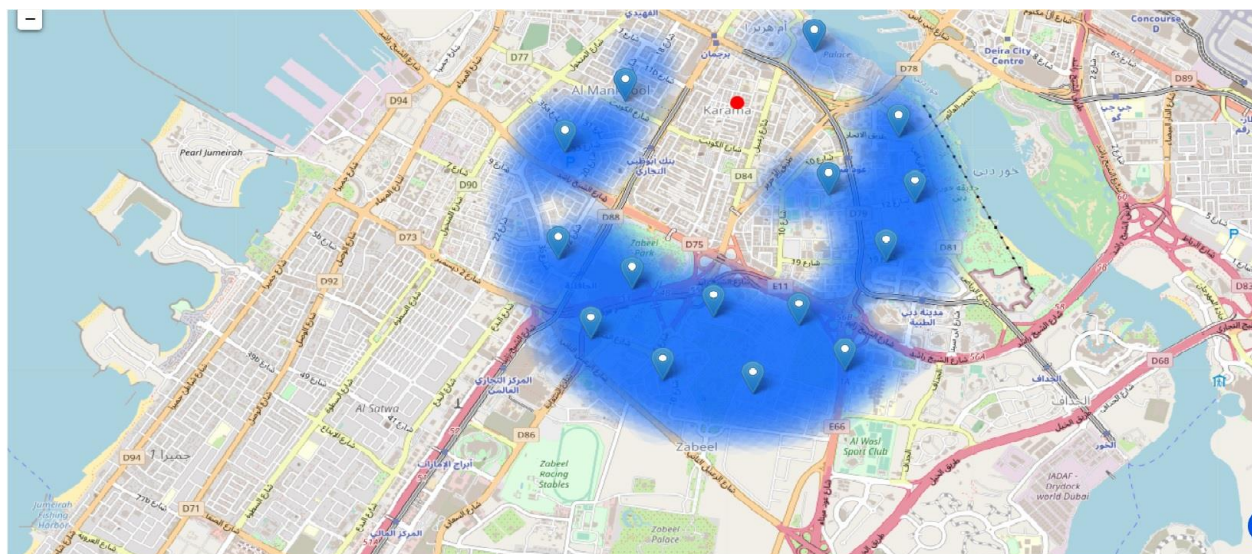
Addresses of those cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics.

This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of restaurants and no Indian restaurants nearby, all zones being fairly close

to city center (all less than 4km from Al Karama, and about half of those less than 2km from Al Karama). Although zones are shown on map with a radius of ~500 meters (green circles), their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential restaurant locations. Most of the zones are located in Al Mankool and Al Rigga boroughs, which we have identified as interesting due to being popular with tourists, fairly close to city center and well connected by public transport.

Results and Discussion

Our analysis shows that although there is a great number of restaurants in Karama, Dubai (~2000 in our initial area of interest which was 12x12km around Al Karama), there are pockets of low restaurant density fairly close to Karama Center. Highest concentration of restaurants was detected north and west from Al Karama, so we focused our attention to areas south, south-east and east, corresponding to boroughs Al Mankool, Al Rigga and south-east corner of central Neighborhood borough.

Another borough was identified as potentially interesting (Zabeel, north-east from Al Karama), but our attention was focused on Al Mankool and Al Rigga which offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics and a number of pockets of low restaurant density.

After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Al Karama) we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with an Indian restaurant closer than 400m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 15 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Indian restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant!

Purpose of this analysis was to only provide info on areas close to AL Karama but not crowded with existing restaurants (particularly Indian) - it is entirely possible that there is a very good reason for

small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area.

Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

Conclusion

Purpose of this project was to identify Dubai areas close to center with low number of restaurants (particularly Indian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Indian restaurant.

By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis (Al Mankool and Al Rigga), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants.

Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.