Capstone Project

Tiago Caldeira

Introduction

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 **Italian restaurant** in **Dubai**, United Arab Emirates. This is based on Batte of Neighborhoods Sample project.

We will use our data science powers to generate a few most promissing 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 Italian restaurants in the neighborhood, if any
- distance of neighborhood from city center
- We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Data

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 (Emirates Hills)

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 aprox. 12x12 killometers centered around Dubai Hills.

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

Coordinate of Emirates Hills, Dubai, United Arab Emirates:

[25.0687174, 55.1734594]

Neighborhood Candidates

- 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).

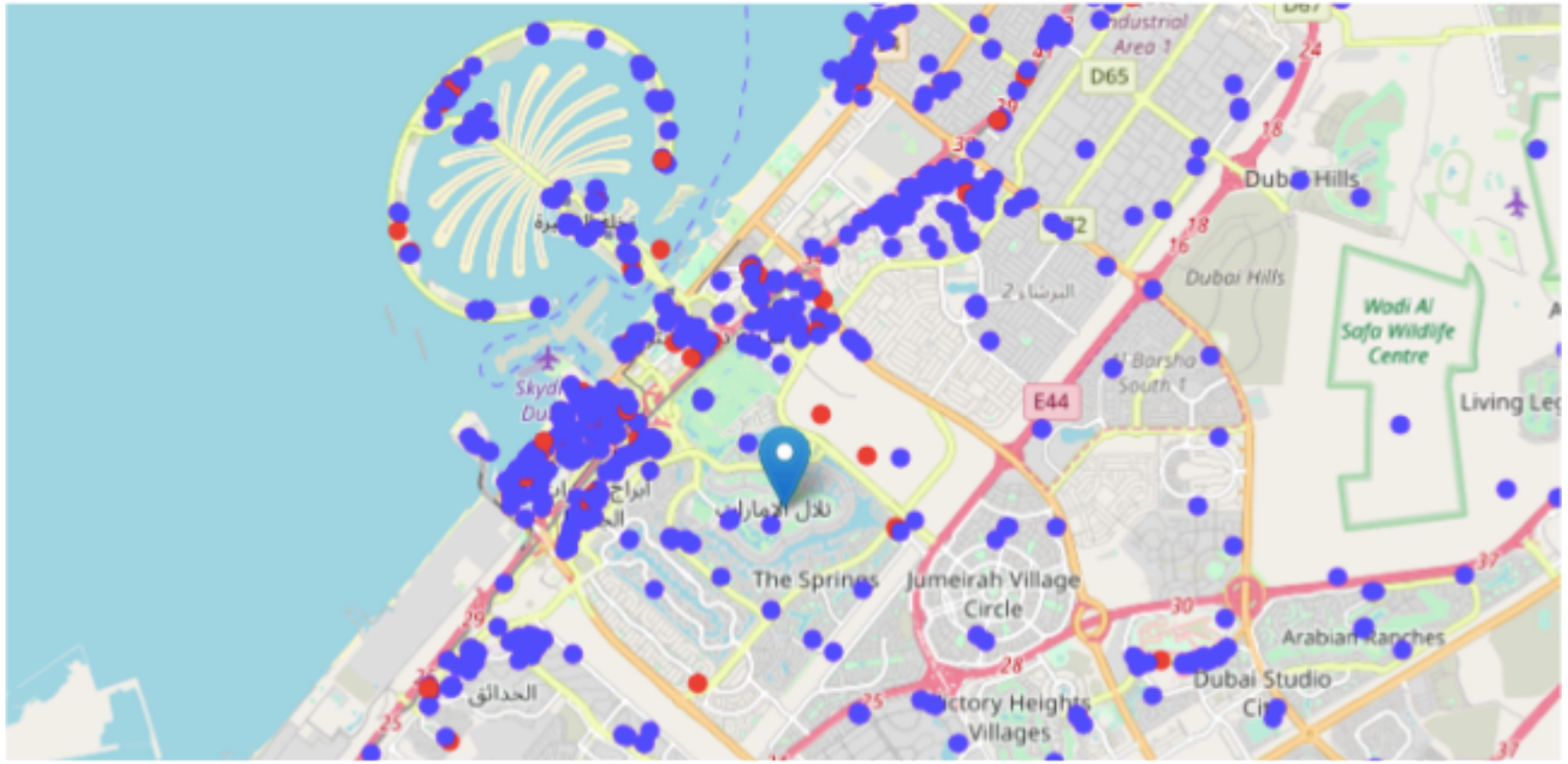
Coordinate transformation check

Dubai center longitude=55.1734594, latitude=25.0687174

Dubai center UTM X=4769373.5528520495, Y=3487130.1045077746

Dubai center longitude=55.173459400000006, latitude=25.068717399999986

Neighborhood Candidates



Neighborhood Candidates

- Panda DataFrame (top 10 elements):

	Address	Distance from center	Latitude	Longitude	X	Y
0	S210, Corner S219, Building A, South Zone 2 - ...	19974.984355	24.952123	55.076533	4.763374e+06	3.468078e+06
1	Unnamed Road - Dubai	19467.922334	24.947210	55.091669	4.765374e+06	3.468078e+06
2	Unnamed Road - Dubai	19157.244061	24.942296	55.106802	4.767374e+06	3.468078e+06
3	Sheikh Mohammed Bin Zayed Rd - Dubai	19052.558883	24.937382	55.121930	4.769374e+06	3.468078e+06
4	Unnamed Road - Dubai	19157.244061	24.932467	55.137054	4.771374e+06	3.468078e+06
5	Unnamed Road - Dubai	19467.922334	24.927551	55.152174	4.773374e+06	3.468078e+06
6	Unnamed Road - Dubai	19974.984355	24.922635	55.167291	4.775374e+06	3.468078e+06
7	Unnamed Road - Dubai	19519.221296	24.971447	55.058489	4.760374e+06	3.469810e+06
8	DCFC Logistic 2 - Dubai	18681.541692	24.966532	55.073632	4.762374e+06	3.469810e+06
9	Transworld Group Of Companies - Dubai	18027.756377	24.961617	55.088770	4.764374e+06	3.469810e+06

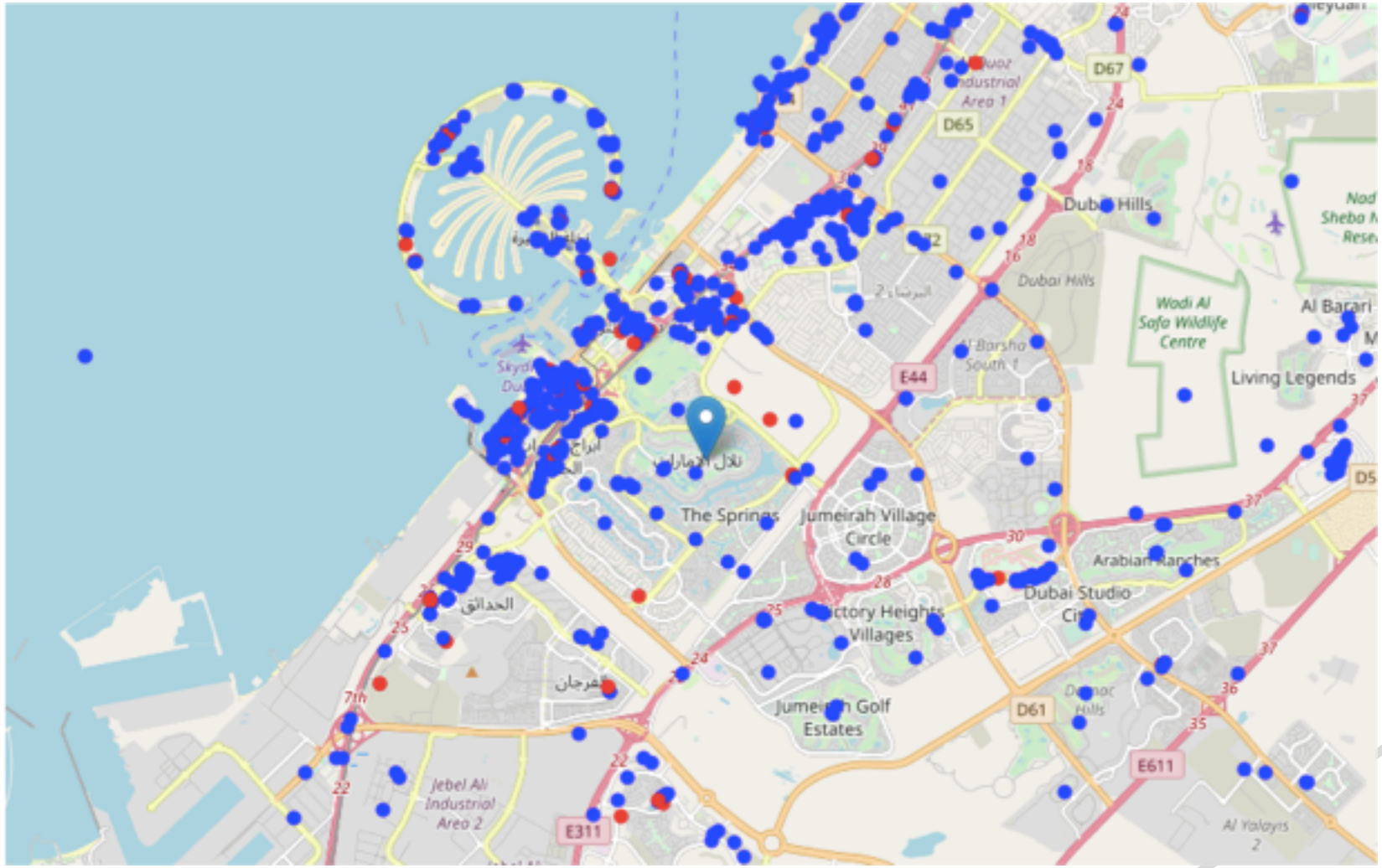
Foursquare

- 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/out list only venues that have 'restaurant' in category name, and we'll make sure to detect and include all the subcategories of specific 'Italian restaurant' category, as we need info on Italian restaurants in the neighborhood.

List of all restaurants

- Let's now see all the collected restaurants in our area of interest on map, and let's also show Italian restaurants in different color.
  - Blue – Restaurant
  - Red – Italian Restaurant

Foursquare


D67
D65
Dubai Hills
E44
Wadi Al Safa Wildlife Centre
Al Barari
Living Legends
The Springs
Jumeirah Village Circle
Arabian Ranches
Dubai Studio City
Jumeirah Golf Estates
D61
E611
E311
Jebel Ali Industrial Area 2
Al Yalayis 2

Methodology

- In this project we will direct our efforts on detecting areas of Dubai that have low restaurant density, particularly those with low number of Italian restaurants. We will limit our analysis to area ~6km around the Emirates Hills.

Methodology

- In first step we have collected the required **data: location and type (category) of every restaurant within 6km from Dubai center** (Emirates Hills). We have also **identified Italian restaurants** (according to Foursquare categorization).
- Second step in our analysis will be calculation and exploration of '**restaurant density**' across different areas of Dubai - we will use **heatmaps** to identify a few promising areas close to center with low number of restaurants in general (*and* no Italian 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 Italian restaurants in radius of 400 meters**.

Analysis

- 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**:

	Address	Distance from center	Latitude	Longitude	X	Y	Restaurants in area
0	S210, Corner S219, Building A, South Zone 2 - ...	19974.984355	24.952123	55.076533	4.763374e+06	3.468078e+06	2
1	Unnamed Road - Dubai	19467.922334	24.947210	55.091669	4.765374e+06	3.468078e+06	1
2	Unnamed Road - Dubai	19157.244061	24.942296	55.106802	4.767374e+06	3.468078e+06	0
3	Sheikh Mohammed Bin Zayed Rd - Dubai	19052.558883	24.937382	55.121930	4.769374e+06	3.468078e+06	0
4	Unnamed Road - Dubai	19157.244061	24.932467	55.137054	4.771374e+06	3.468078e+06	0
5	Unnamed Road - Dubai	19467.922334	24.927551	55.152174	4.773374e+06	3.468078e+06	0
6	Unnamed Road - Dubai	19974.984355	24.922635	55.167291	4.775374e+06	3.468078e+06	0
7	Unnamed Road - Dubai	19519.221296	24.971447	55.058489	4.760374e+06	3.469810e+06	1
8	DCFC Logistic 2 - Dubai	18681.541692	24.966532	55.073632	4.762374e+06	3.469810e+06	4
9	Transworld Group Of Companies - Dubai	18027.756377	24.961617	55.088770	4.764374e+06	3.469810e+06	3

Analysis

- OK, now let's calculate the **distance to nearest Italian restaurant from every area candidate center** (not only those within 1000m - we want distance to closest one, regardless of how distant it is).

Average distance to closest Italian restaurant from each area center:
3948.92935700017

- OK, so **on average Italian restaurant can be found within ~4000m** from every area center candidate.
- That's close enough, so we need to filter our areas carefully!

Analysis

- Let's crete a map showing **heatmap / 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 Emirates Hills.

Analysis

- Looks like a few pockets of low restaurant density closest to city center can be found **north, north-east from Emirates Hills**.
- Let's create another heatmap map showing **heatmap/density of Italian restaurants** only:

Analysis

- This map is not so 'hot' (Italian restaurants represent a subset of ~15% of all restaurants in Dubai) but it also indicates higher density of existing Italian restaurants directly north and west from Emirates Hills, with closest pockets of **low Italian restaurant density positioned east, south-east and south from Hills**.

- 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:

Analysis

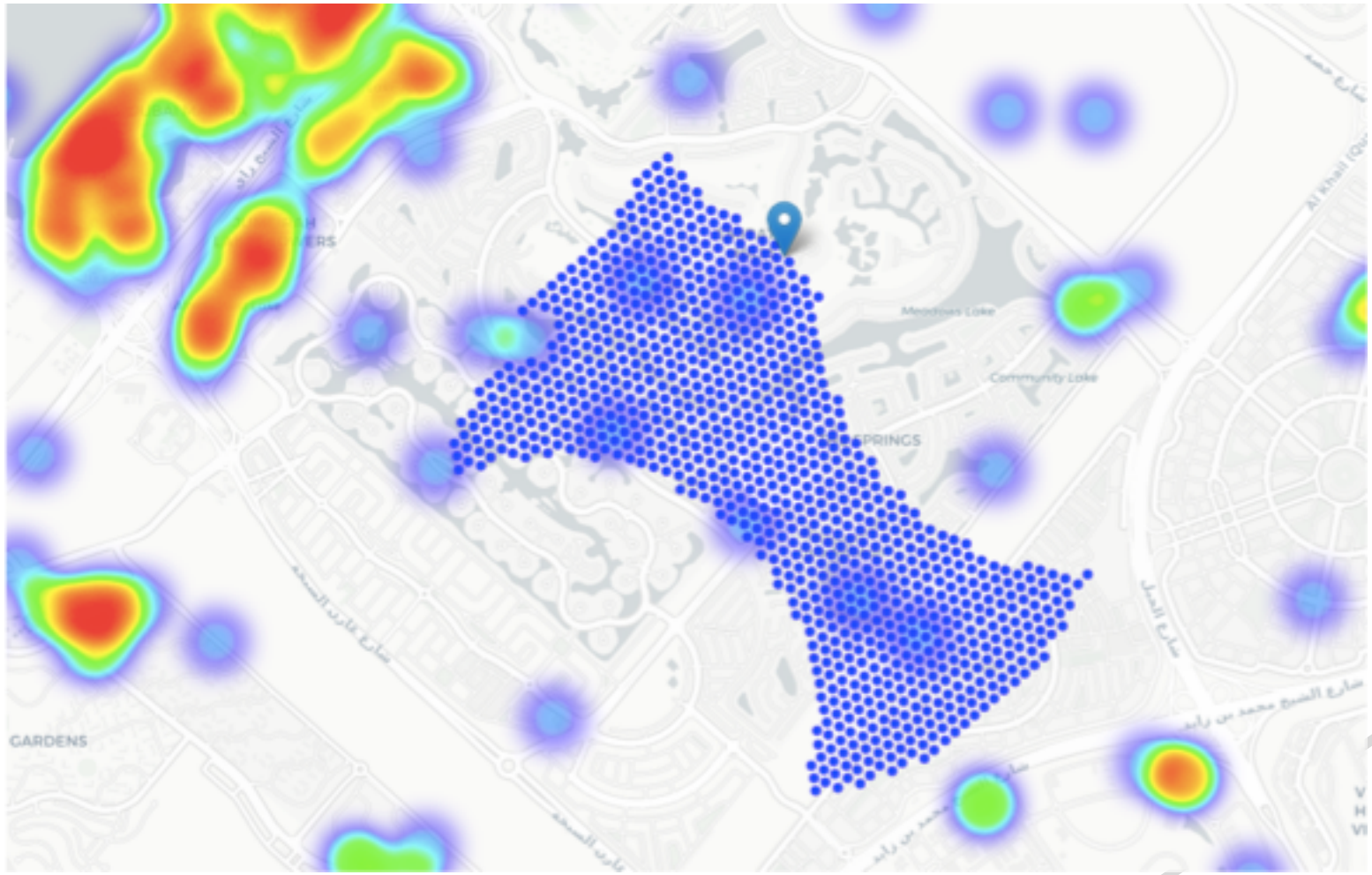
Optimization

	Distance to Italian restaurant	Latitude	Longitude	Restaurants nearby	X	Y
0	934.188316	25.040041	55.166015	0	4.769824e+06	3.483130e+06
1	1033.649940	25.039795	55.166771	0	4.769924e+06	3.483130e+06
2	378.927984	25.041996	55.162089	0	4.769274e+06	3.483217e+06
3	478.864080	25.041749	55.162846	0	4.769374e+06	3.483217e+06
4	578.822250	25.041502	55.163602	0	4.769474e+06	3.483217e+06
5	678.792742	25.041256	55.164359	0	4.769574e+06	3.483217e+06
6	778.770811	25.041009	55.165115	0	4.769674e+06	3.483217e+06
7	878.753871	25.040762	55.165872	0	4.769774e+06	3.483217e+06
8	978.740393	25.040515	55.166628	0	4.769874e+06	3.483217e+06
9	1078.729413	25.040268	55.167384	0	4.769974e+06	3.483217e+06

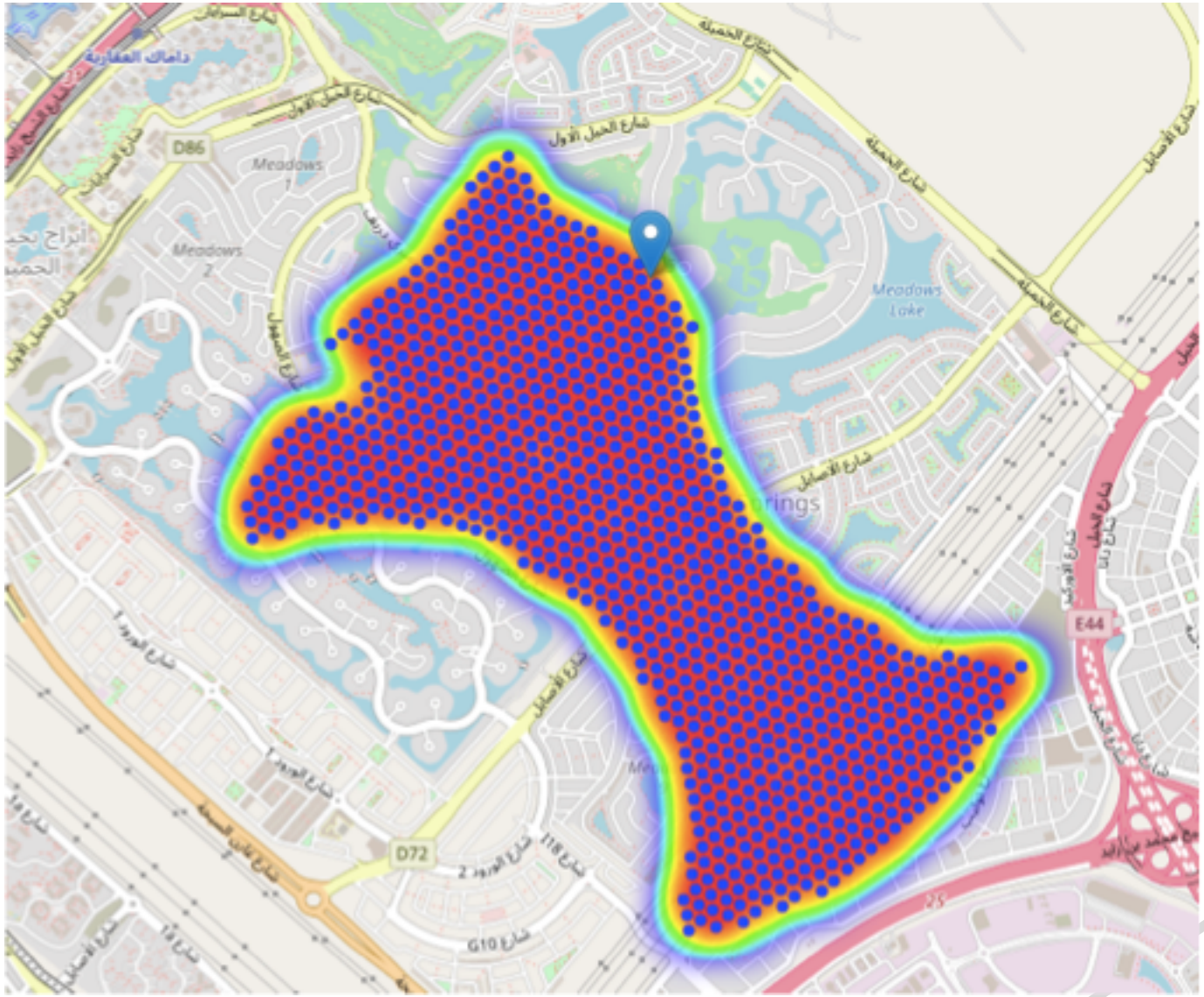
Optimization

- OK. Let us now **filter** those locations: we're interested only in **locations with no more than two restaurants in radius of 1000 meters**, and **no Italian restaurants in radius of 2000 meters.**
-
-
- Locations with no more than two restaurants nearby: 2236
- Locations with no Italian restaurants within 2000m: 991
- Locations with both conditions met: 978

Optimization



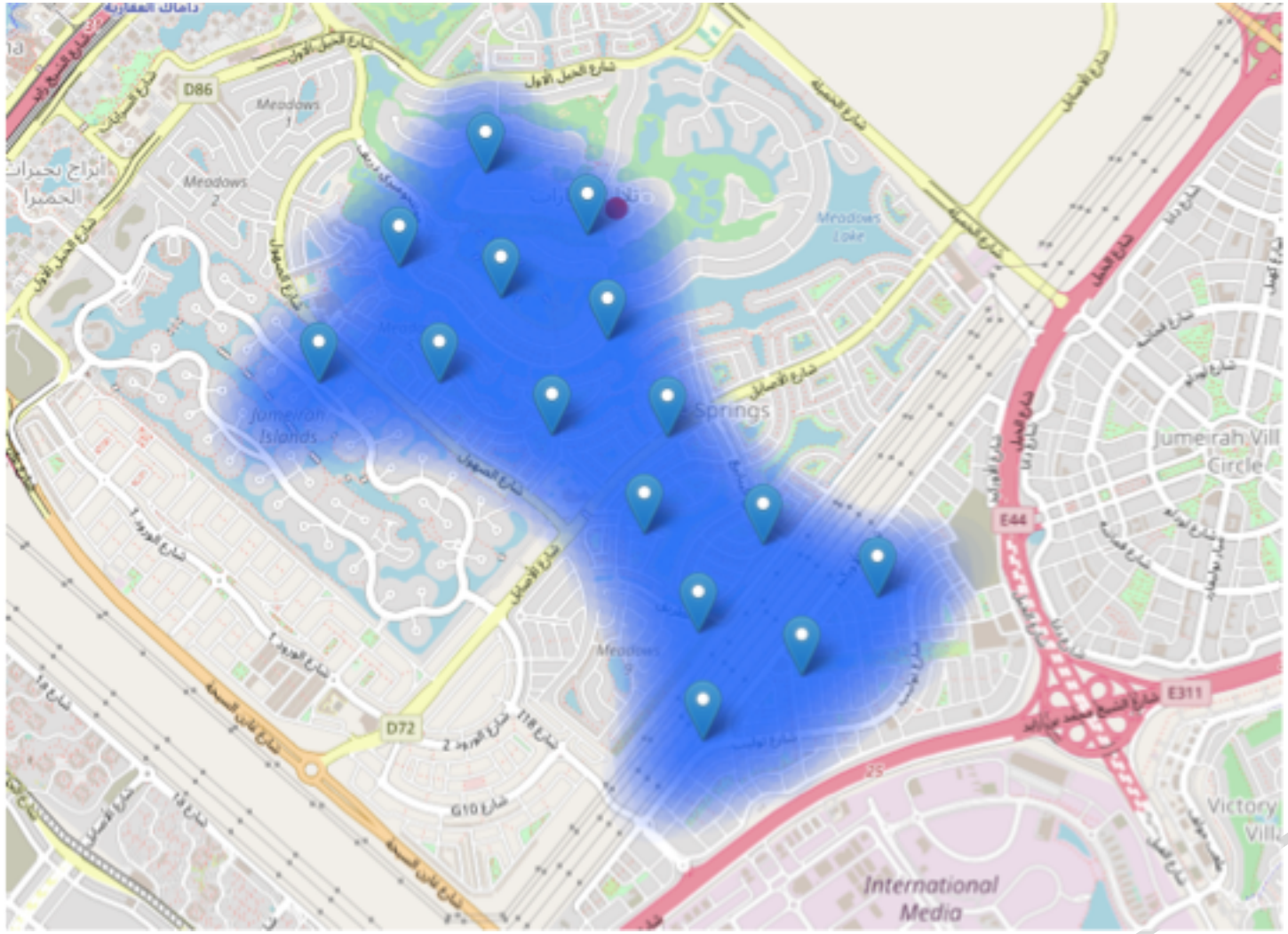
Optimization



Optimization

- This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of restaurants and no Italian restaurants nearby, all zones being fairly close to city center (all less than 4km from Emirates Hills, and about half of those less than 2km from Emirates Hills).
- 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

Results and Discussion



Results and Discussion

- Our analysis shows that although there is a great number of restaurants in Dubai (~2000 in our initial area of interest which was 12x12km around Emirates Hills), there are pockets of low restaurant density fairly close to city center. Highest concentration of restaurants was detected north and west from Emirates Hills, so we focused our attention to areas south, south-east and east.

Results and Discussion

- After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Emirates Hills)
- 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 Italian restaurant closer than 400m were removed.

Results and Discussion

- 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 Italian 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 Dubai center but not crowded with existing restaurants (particularly Italian)

Conclusion

- Purpose of this project was to identify Dubai areas close to center with low number of restaurants (particularly Italian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Italian restaurant.
- By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis, 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.

Conclusion

- 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, etc.

- A Trial code was run searching for Halal restaurants, but the information in Foursquare isn't complete enough.

Future Work

- In the Dubizzle platform, there are many organized data about houses, containing things like zone, type, price, year, condition, among others. Using the data from :
- https://dubai.dubizzle.com/en/property-for-rent/residential/apartmentflat/ (42.000 registers) and the information in
- https://dubai.dubizzle.com/en/property-for-sale/residential/apartment/ (25.000 registers),

- I could connect the price of buying and selling, with the renting prices. This will allow people to know where to buy (how far away from their job), and if it is better to buy or rent.