Capstone Project

Tiago Caldeira

based on Batte of Neighborhoods Sample project

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

Since there are lots of restaurants in Dubai we will try to detect **locations that are not already crowded with restaurants**. We are also particularly interested in **areas with no Italian 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 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 decission 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.

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 city center.

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]
```

Now let's create a grid of area candidates, equaly spaced, centered around city center and within ~6km from Emirates Hills. Our neighborhoods will be defined as circular areas with a radius of 500 meters, so our neighborhood centers will be 1000 meters apart.

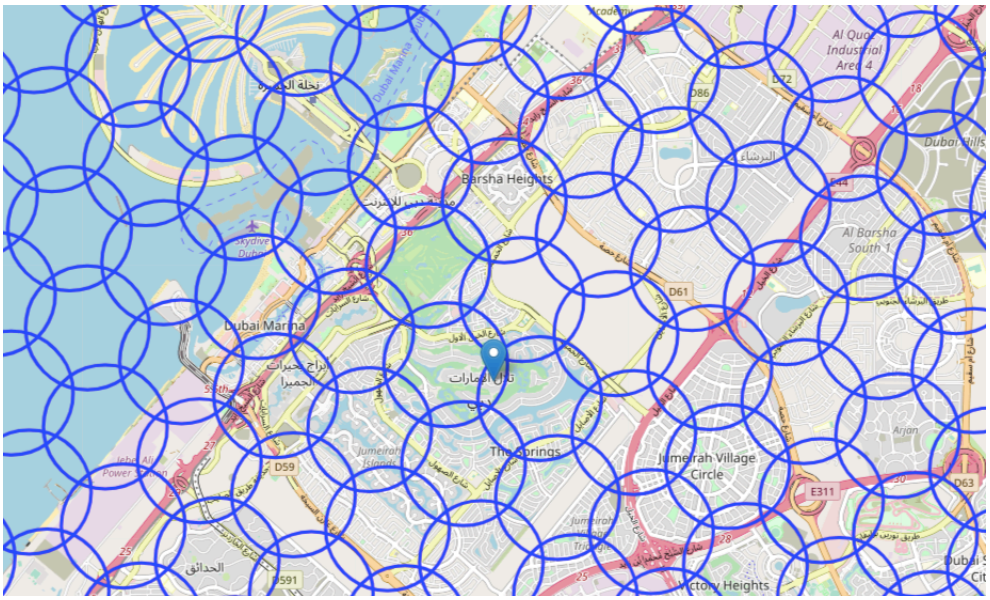
To accurately calculate distances we need to create our grid of locations in Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we'll project those coordinates back to latitude/longitude degrees to be shown on Folium map. So let's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

```
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
```

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 its 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 it's neighbors is exactly the same) and within ~6km from Emirates Hills.

Let's now use Google Maps API to get approximate addresses of those locations.

```
Reverse geocoding check
-----------------------
Address of [25.0687174, 55.1734594] is: L 37 Lailak3 St - Dubai - United Ar
ab Emirates

Obtaining location addresses:  . . . . . .
. . . . . .
. . . . . .
. . . . . .
. . . . . done.

364 Addresses Located Successfully!
```

Looking good. Let's now place all this into a Pandas dataframe.

	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

...and let's now save/persist this data into local file.

Foursquare

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/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.
Foursquare credentials are defined in hidden cell bellow.

```
Restaurant data loaded.
Obtaining venues around candidate locations: . . . . . .
. . . . . .
. . . . . .
. . . . . done.
```

```
List of all restaurants
-----------------------
('529dea9911d2a130373f71af', 'Goto King', 25.07079550519343, 55.30093378933
527, 'The Global Village, 766 ,'الإمارات العربية المتحدة', False, 4784211.815
098009, 3492767.5557770417)
('4c3a00e81e06d13a3b807a3e', 'Hummus Wa Falafel 25.157690614 ,'(حمص و فلافل)
70935, 55.20578222068266, 'Jumaira Road (شارع جميرا), دبي, الإمارات العربية
610 ,'المتحدة', False, 4769044.282004146, 3499926.9894962977)
('5493e957498ed521a27f6ea1', 'Dolma house', 25.071929654241114, 55.29239272
879815, '254 ,'الإمارات العربية المتحدة', False, 4783158.214367357, 3492552.7
683752156)
('52230a2211d2f5e7fdbf9c3a', 'Arabian Fame', 25.167633596177676, 55.2492846
2850204, 'Al Quoz (Near Dubai Bowling Center), دبي, الإمارات العربية المتحدة
', 355, False, 4773671.6781430235, 3503040.2256060936)
('507156d7e4b00022a1bfa6bd', 'Tong Thai', 25.186097225212052, 55.2588778778
1253, 'JW Marriott Marquis Dubai, 544 ,'الإمارات العربية المتحدة', False, 477
3936.466532904, 3505818.6201815754)
('52bc5921498ee6a30000619b', 'Doner Kebab - German Doner Kebab دونر كباب الأ
لماني', 25.078799233141982, 55.14304223372914, 'الإمارات العربية المتحدة', 8
32, False, 4765347.312789697, 3487148.9718169854)
('5166aa43e4b0ffd5d0dc2250', 'Ines', 25.185704304224537, 55.24106162163222,
'Al Wasl Square (Jumirah), 615 ,'الإمارات العربية المتحدة', دبي, False, 47718
71.304806152, 3505015.3919359534)
('52ac32a0498e1c9576dfd0e5', 'Chicken Now', 25.03402929563716, 55.202271953
08487, 'Dubai Marina Mall, 175 ,'الإمارات العربية المتحدة', دبي مارينا, False
, 4774349.265483891, 3483879.7819207828)
('4e10288c8877ac23b655c6a4', 'Oregano', 24.981091518070865, 55.093726392011
54, '662 ,'الإمارات العربية المتحدة', دبي, True, 4764061.904417273, 3472521.2
75694559)
('4b164b33f964a5200db823e3', "Chili's", 25.197375853976013, 55.237244384093
72, 'Jumeirah Beach Road - opposite Jumeirah Beach Park (at Jumeirah Rd.),
421 ,'الإمارات العربية المتحدة', دبي, False, 4770883.223349414, 3506353.95095
40205)
...
Total: 1099
```

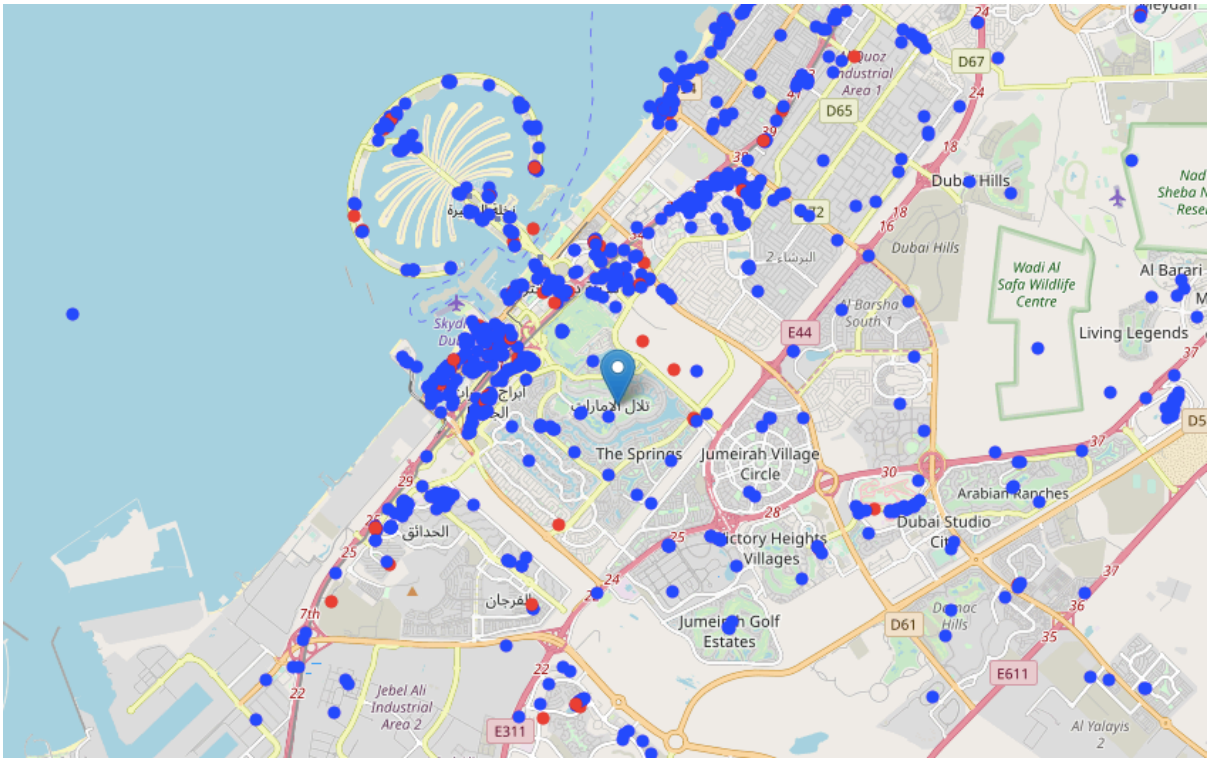
List of Italian restaurants

('5a68abb82aff3150185770ea', 'C House Milano', 25.034225, 55.118454, 'دبي, الإمارات العربية المتحدة', True, 4764519.721284089, 3480387.9672200275)
('4d64a1966ccdba7a69569ac3', 'Sicilia', 25.042199970319306, 55.11506390558833, 'Mövenpick Hotel Ibn Bautatta Gate (Near Gardens), دبي, الإمارات العربي, ة المتحدة', True, 4763756.006726628, 3481271.2289360412)
('4cee3601821254816ca263a1', 'Al Dente Restaurant', 25.097779894577346, 55.173482321863666, 'Byblos Hotel, 353 , دبي, الإمارات العربية المتحدة', True, 4768035.109128352, 3490867.3407083596)
('5814ce4d38fa462c358b3311', 'II Capo', 25.184294, 55.255688, 'دبي, الإمارا, ت العربية المتحدة', True, 4773647.176164293, 3505452.056156282)
('4d21eea76e8c3704101f07a0', 'Pregos', 25.10007887365995, 55.17986285804715, 'Media Rotana (Hessa St), 921 , دبي, الإمارات العربية المتحدة', True, 4768675.808256671, 3491431.132048977)
('4f9a5190e4b033dc295a0546', 'Znkol', 25.082977, 55.14616, 'الإمارات العربي, ة المتحدة', True, 4765519.742793849, 3487816.831388589)
('4c21d1177e85c9285c03bb21', 'Segafredo Zanetti', 25.10474564510248, 55.16799566351403, 'Office park building, Dubai Internet City, دبي, الإمارات العرب, ية المتحدة', True, 4767071.3653601995, 3491532.067513119)
('4ba7a38af964a52004a539e3', 'BiCE, Hilton Dubai Jumeirah Resort', 25.079041691956416, 55.13377528711668, 'Hilton Dubai Jumeirah Resort, دبي, الإمارات العربية المتحدة', True, 4764251.349299353, 3486791.300140736)
('4c077057cf8c76b073013c65', 'Ronda Locatelli', 25.131477282199103, 55.11876168530195, 'Atlantis The Palm (Crescent Rd), دبي, الإمارات العربية المتحدة', 973, True, 4760078.35043683, 3492899.063011324)
('4d3572d82c76a14373998ec7', "Stefano's", 25.004241767880902, 55.164251284423706, 'Green Community, 88 , دبي, الإمارات العربية المتحدة', True, 4771265.392500846, 3478451.5913260398)
...
Total: 80

Restaurants around location

Restaurants around location 200: Latest Recipe, Bussola, Fish Beach Taverna
Restaurants around location 202: SanaBonts, Pizzaro
Restaurants around location 203: koshi bashi, Pizzaro, Dolma land
Restaurants around location 204: Wide Range Restaurant
Restaurants around location 205: McDonald's
Restaurants around location 206: McDonald's
Restaurants around location 207: Cravings
Restaurants around location 208: Ranches Restaurant and Bar, Maison Mathis
Restaurants around location 209: Terranova

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.



Looking good. So now we have all the restaurants in area within few kilometers from Emirates Hills, and we know which ones are Italian 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 Italian restaurant!

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 city center.

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**. 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.

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**:

```
Average number of restaurants in every area with radius=1000m: 3.870879120
```

	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

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).

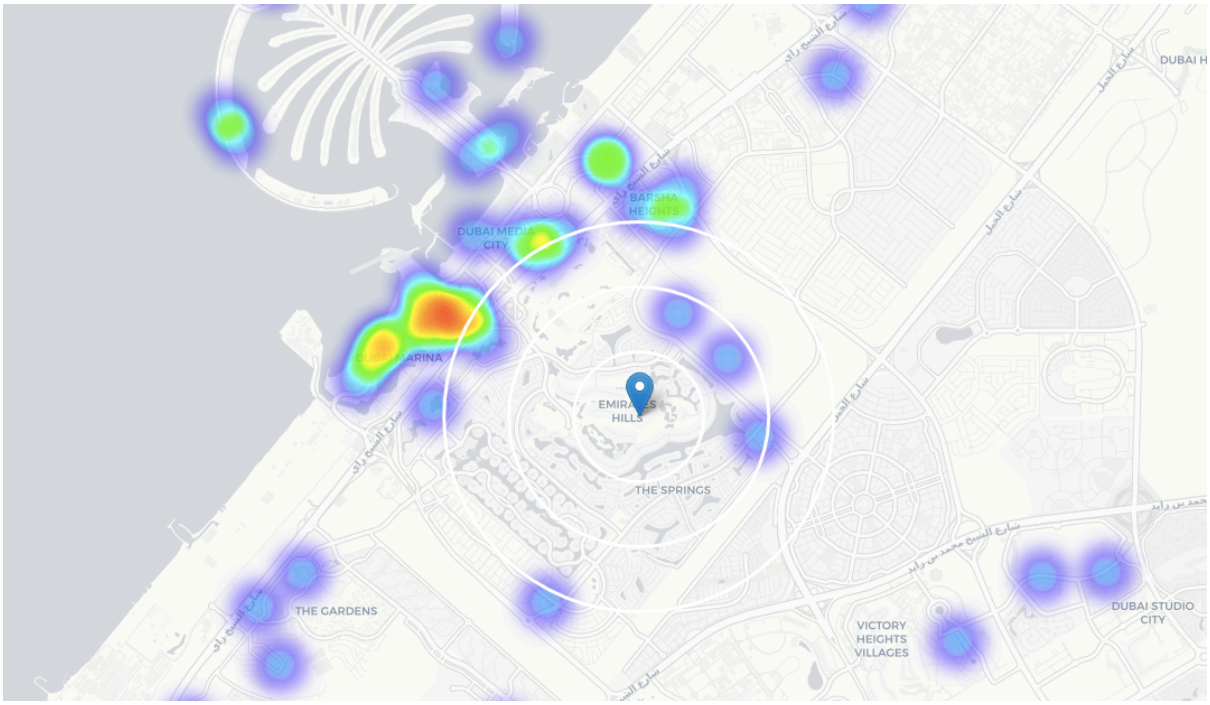
```
Average distance to closest Italian restaurant from each area center: 3948.
92935700017
```

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

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.

Looks like a few pockets of low restaurant density closest to city center can be found **north, north-east from Emirates Hills**.

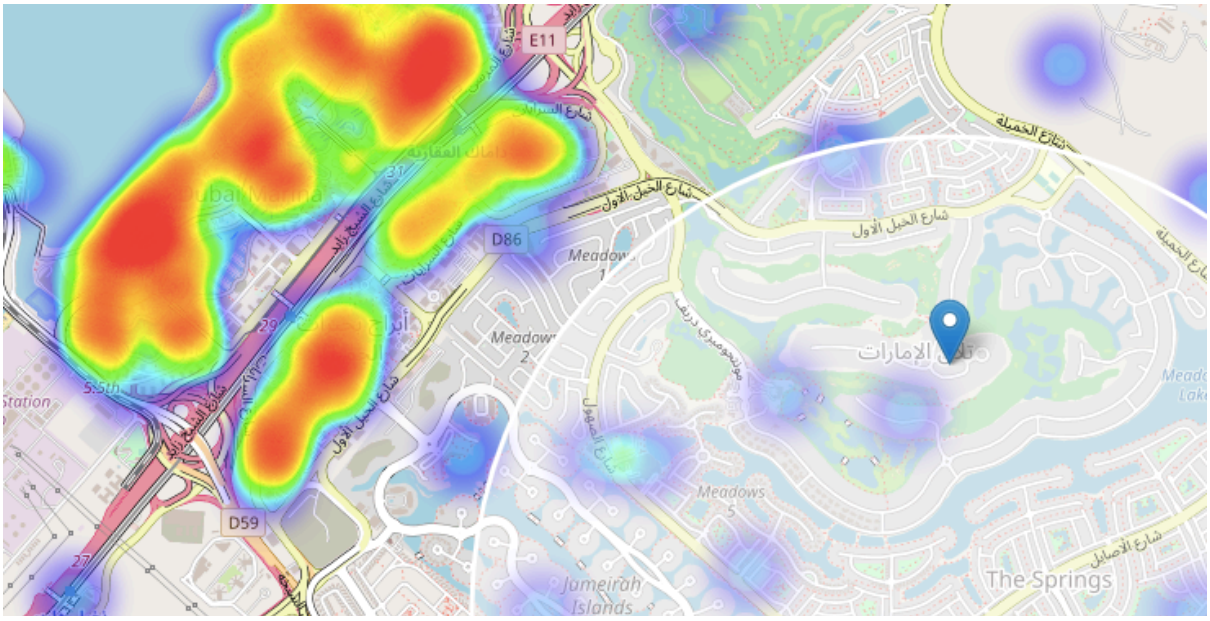
Let's create another heatmap map showing **heatmap/density of Italian restaurants** only.



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 city center**.

Analysis of popular travel guides and web sites often mention Kreuzberg and Friedrichshain as beautifull, interesting, rich with culture, 'hip' and 'cool' Dubai neighborhoods popular with tourists and loved by Dubaiers.

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.



Not bad - this nicely covers all the pockets of low restaurant density in Dubai closest to Emirates Hills in Dubai.

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 appart).

```
2261 candidate neighborhood centers generated.
```

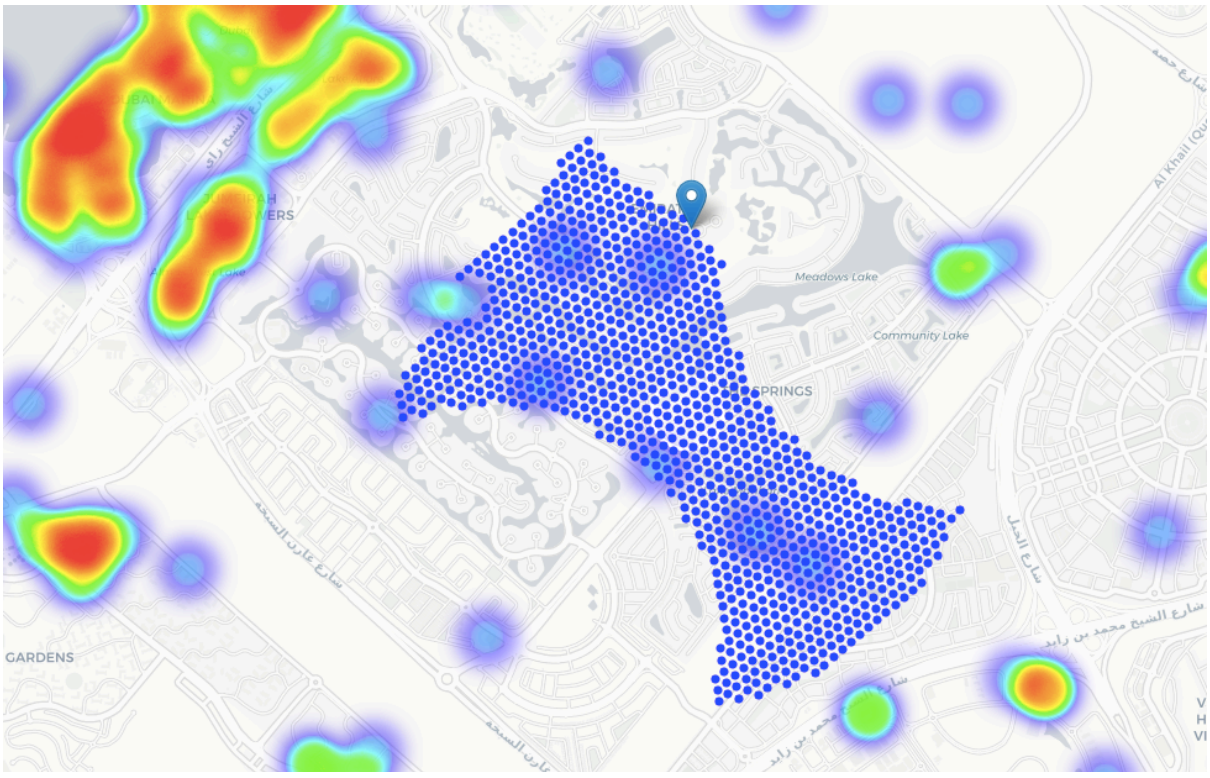
OK. Now let's calculate two most important things for each location candidate: **number of restaurants in vicinity** (we'll use radius of **250 meters**) and **distance to closest Italian restaurant**.

	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

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**.

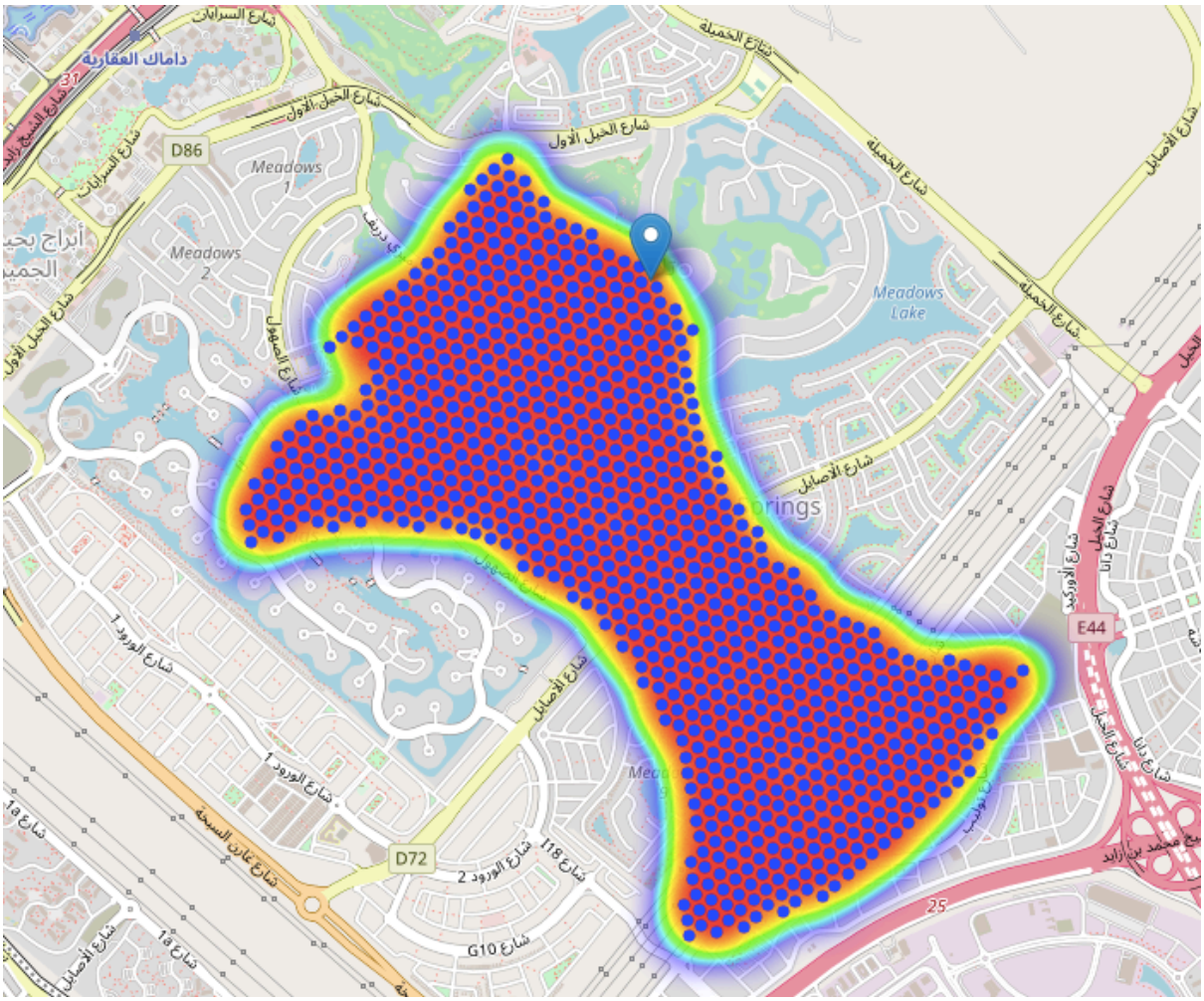
```
Locations with no more than two restaurants nearby: 2236
Locations with no Italian restaurants within 2000m: 991
Locations with both conditions met: 978
```

Let's see how this looks on a map.



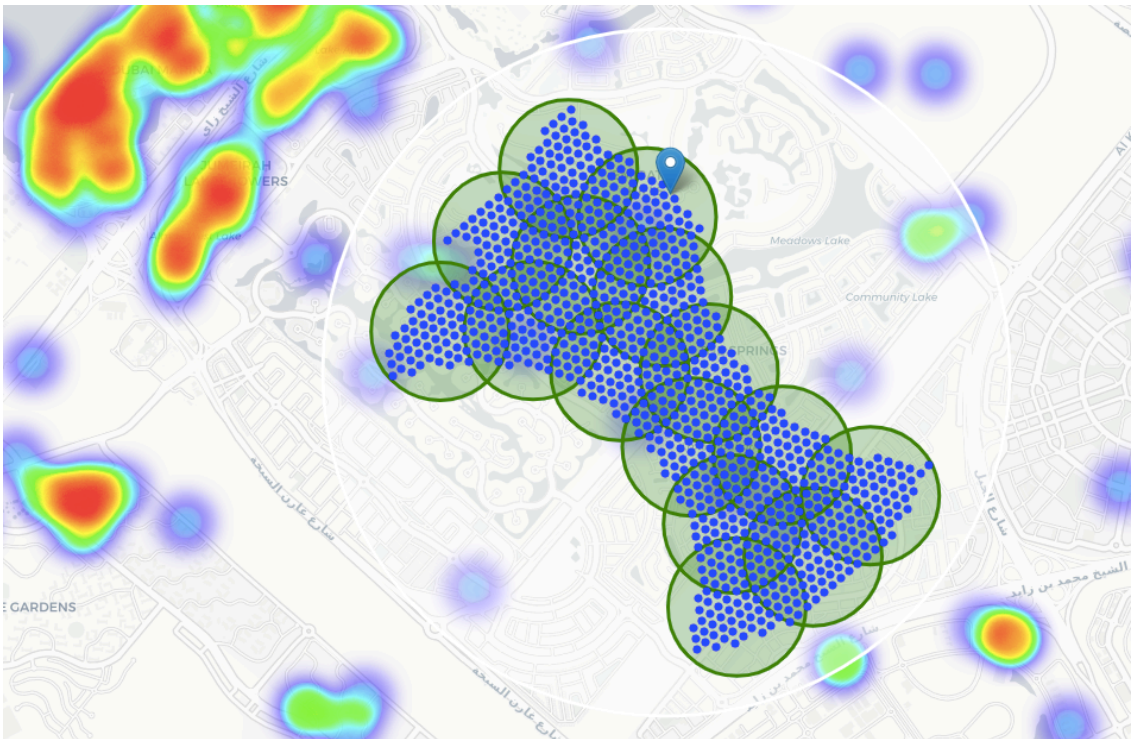
Looking good. We now have a bunch of locations fairly close to Emirates Hills, and we know that each of those locations has no more than two restaurants in radius of 1000m, and no Italian restaurant closer than 4000m. Any of those locations is a potential candidate for a new Italian restaurant, at least based on nearby competition.

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



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

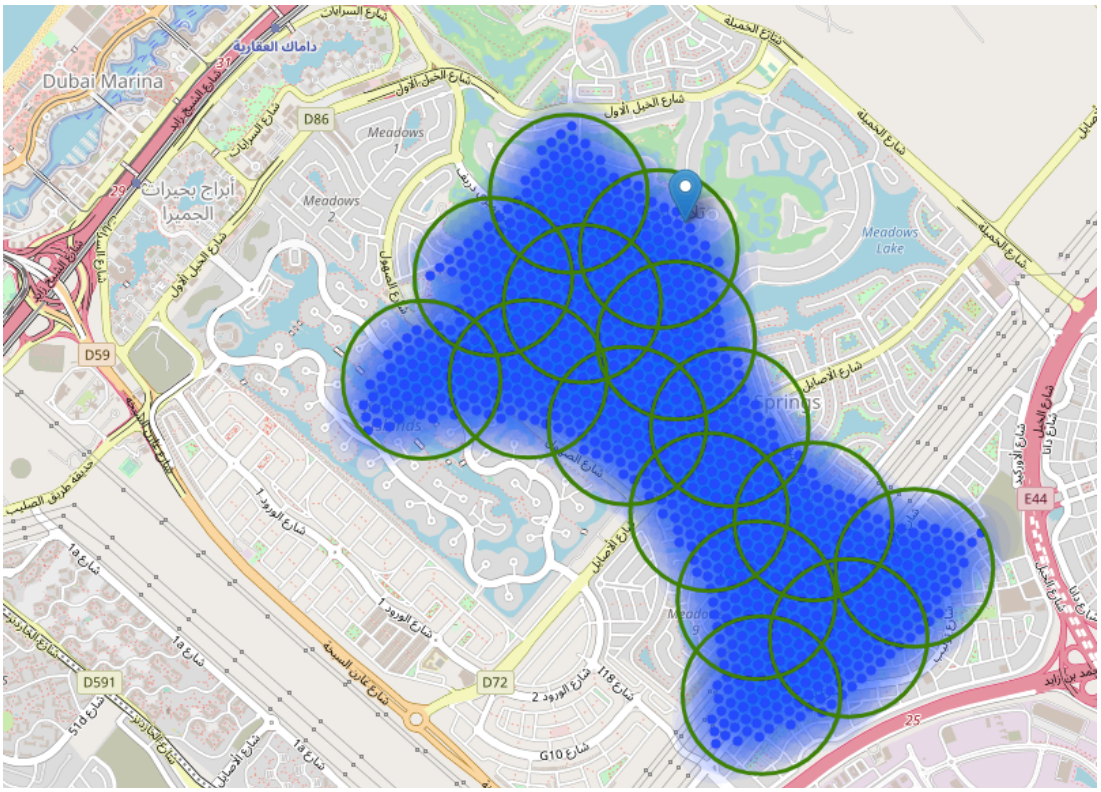
Let us now **cluster** those locations to create **centers of zones containing good locations**. Those zones, their centers and addresses will be the final result of our analysis.



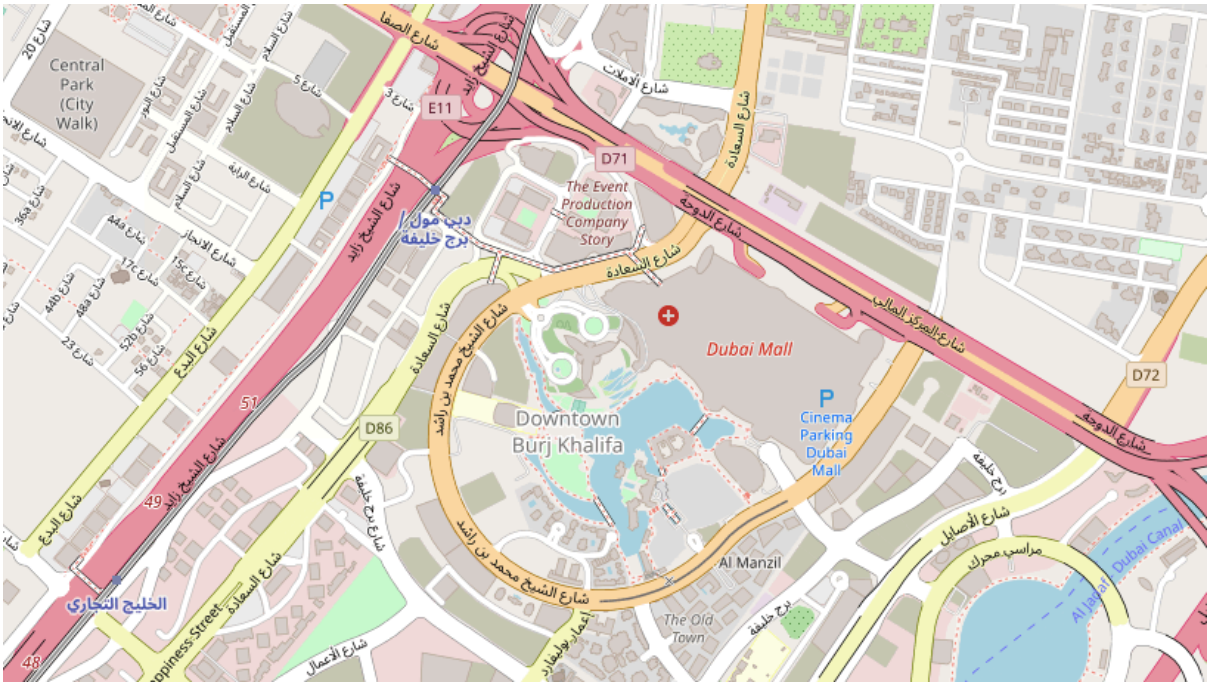
Not bad - our clusters represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates.

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.

Let's see those zones on a city map without heatmap, using shaded areas to indicate our clusters:



Let's zoom in on candidate areas in **Dubai Downtown**:



...and candidate areas in **Dubai Palm**:

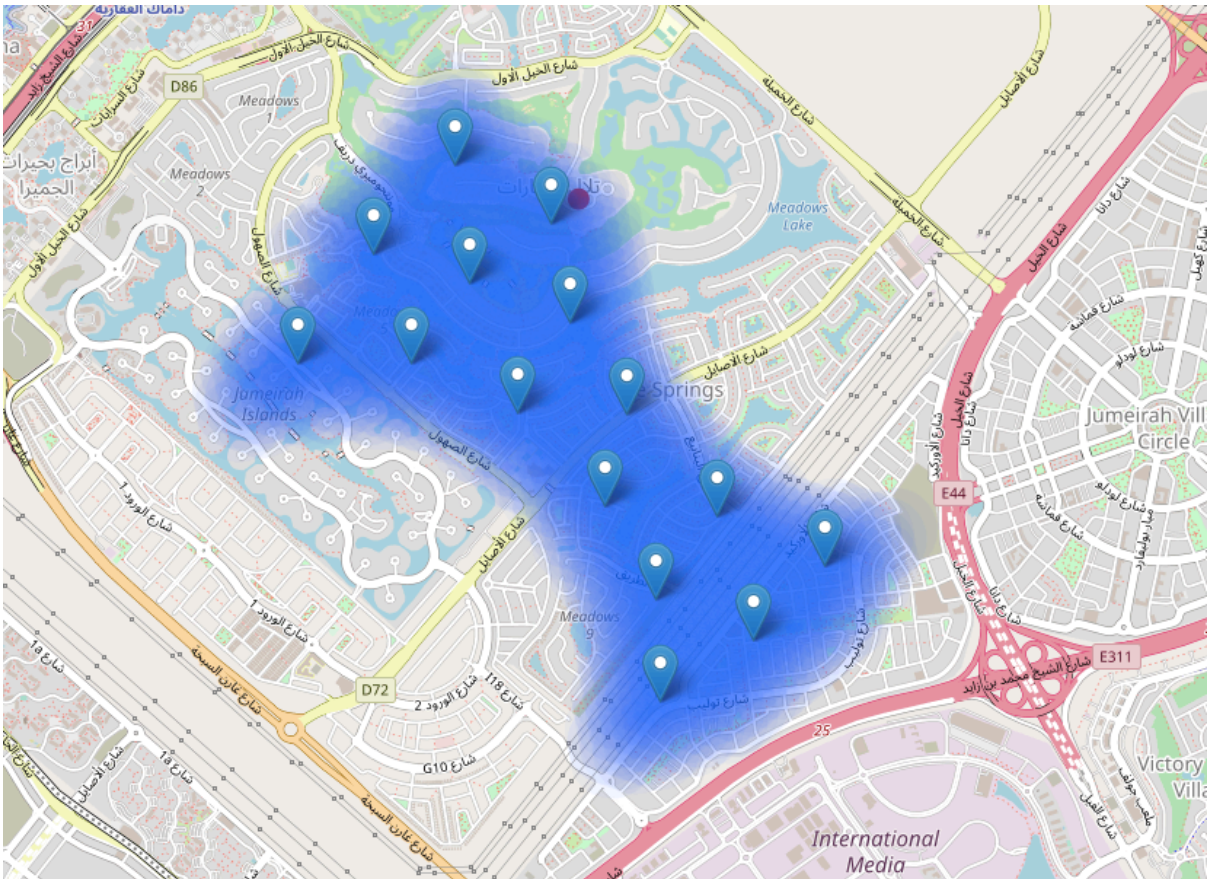
Finaly, let's **reverse geocode those candidate area centers to get the addresses** which can be presented to stakeholders.

```
============================================================
Addresses of centers of areas recommended for further analysis
============================================================

7 Street 2 - Dubai - United Arab Emirates          => 1.6km from Emirates Hills, Dubai
L 19 Lailak2 St - Dubai - United Arab Emirates     => 0.9km from Emirates Hills, Dubai
Unnamed Road - Dubai - United Arab Emirates        => 3.2km from Emirates Hills, Dubai
14 13 St - Dubai - United Arab Emirates            => 1.7km from Emirates Hills, Dubai
```

```
22 3 St - Dubai - United Arab Emirates              => 3.0km from Emirates Hills, Dubai
Unnamed Road - Dubai - United Arab Emirates        => 2.4km from Emirates Hills, Dubai
E8 Corniche St - Dubai - United Arab Emirates      => 0.9km from Emirates Hills, Dubai
29 6 St - Dubai - United Arab Emirates             => 2.3km from Emirates Hills, Dubai
E172 Yasmen1 St - Dubai - United Arab Emirates     => 1.0km from Emirates Hills, Dubai
8 Narjes St - Dubai - United Arab Emirates         => 0.3km from Emirates Hills, Dubai
Unnamed Road - Dubai - United Arab Emirates        => 3.5km from Emirates Hills, Dubai
51 9 St - Dubai - United Arab Emirates             => 1.6km from Emirates Hills, Dubai
223 Tulip St - Dubai - United Arab Emirates        => 3.7km from Emirates Hills, Dubai
7 7 St - Dubai - United Arab Emirates              => 1.6km from Emirates Hills, Dubai
3 Street 1 - Dubai - United Arab Emirates          => 2.5km from Emirates Hills, Dubai
```

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. Most of the zones are located in Kreuzberg and Friedrichshain 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 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, corresponding to boroughs Kreuzberg, Friedrichshain and south-east corner of central Mitte borough. Another borough was identified as potentially interesting (Prenzlauer Berg, north-east from Emirates Hills), but our attention was focused on Kreuzberg and Friedrichshain 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 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.

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 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) - 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 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.

Final decission 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.

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