

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

Applied Data Science Capstone by IBM

By Diego Hidalgo

Table of contents

- Introduction: Business Problem
- Data
- Methodology
- Analysis
- Results and Discussion
- Conclusion

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 **Chicago, IL US**.

Since there are lots of restaurants in Chicago 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 east 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 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 Chicago center will be obtained using **Google Maps API geocoding**

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 around Chicago city east.

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

Coordinate of Chicago, IL, USA: [41.8781136, -87.6297982]

Now let's create a grid of area candidates, equally spaced, centered around city center and within ~6km from Chicago East. 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).

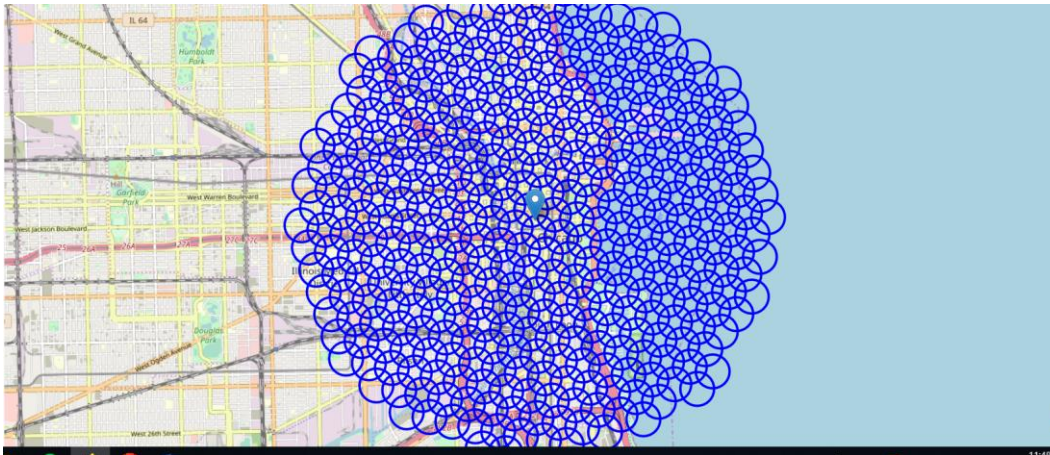
Coordinate transformation check

```
-----
Chicago center longitude=-87.6297982, latitude=41.8781136
Chicago center UTM X=-5379210.510659022, Y=11522384.948403545
Chicago center longitude=-87.62979820000015,
latitude=41.87811360000047
```

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 it's neighbors is exactly the same) and within ~6km from Chicago Center.

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

```
[ '1330 N LaSalle Dr, Chicago, IL 60610',
  '1534 N Wells St, Chicago, IL 60610',
  '242 W St Paul Ave, Chicago, IL 60614',
  '25 Fort Dearborn Dr, Chicago, IL 60616',
  'S Lake Shore Dr, Chicago, IL 60616',
  'S Lake Shore Dr, Chicago, IL 60616',
  'E 18th Dr, Chicago, IL 60605',
  '13 S Lake Shore Dr, Chicago, IL 60605',
  '1388 S Lake Shore Dr, Chicago, IL 60605',
  '1158 S Columbus Dr, Chicago, IL 60605',
  '150 E 11th Pl, Chicago, IL 60605',
  'Michigan & Ida B Wells Drive, Chicago, IL 60605',
  '215 S Wabash Ave, Chicago, IL 60604',
  '17 N State St, Chicago, IL 60602',
  '35 W Wacker Dr, Chicago, IL 60601',
  '59 W Hubbard St, Chicago, IL 60654',
  '631 N LaSalle Dr, Chicago, IL 60654',
  '168198 W Chicago Ave, Chicago, IL 60654',
  '219 W Oak St, Chicago, IL 60610',
  '300 W Division, Chicago, IL 60610']
```

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

| | Address | Latitude | Longitude | X | Y | Distance from center |
|---|-------------------------------------|-----------|------------|---------------|--------------|----------------------|
| 0 | Cook County, IL | 41.878741 | -87.580104 | -5.381011e+06 | 1.151667e+07 | 5992.495307 |
| 1 | 1000 E Grand Ave, Chicago, IL 60611 | 41.882270 | -87.581676 | -5.380411e+06 | 1.151667e+07 | 5840.376700 |

| | Address | Latitude | Longitude | X | Y | Distance from center |
|---|-------------------------------------|-----------|------------|---------------|--------------|----------------------|
| 2 | 1000 E Grand Ave, Chicago, IL 60611 | 41.885798 | -87.583248 | -5.379811e+06 | 1.151667e+07 | 5747.173218 |
| 3 | 1000 E Grand Ave, Chicago, IL 60611 | 41.889327 | -87.584821 | -5.379211e+06 | 1.151667e+07 | 5715.767665 |
| 4 | 1000 E Grand Ave, Chicago, IL 60611 | 41.892856 | -87.586394 | -5.378611e+06 | 1.151667e+07 | 5747.173218 |
| 5 | 1000 E Grand Ave, Chicago, IL 60611 | 41.896385 | -87.587967 | -5.378011e+06 | 1.151667e+07 | 5840.376700 |
| 6 | 1003 E Ohio St, Chicago, IL 60611 | 41.899915 | -87.589540 | -5.377411e+06 | 1.151667e+07 | 5992.495307 |
| 7 | Cook County, IL | 41.872432 | -87.581835 | -5.381911e+06 | 1.151719e+07 | 5855.766389 |
| 8 | Cook County, IL | 41.875960 | -87.583407 | -5.381311e+06 | 1.151719e+07 | 5604.462508 |
| 9 | 1000 E Grand Ave, Chicago, IL 60611 | 41.879488 | -87.584979 | -5.380711e+06 | 1.151719e+07 | 5408.326913 |

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

```
df_locations.to_pickle('./locations.pkl')
```

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 - coffe 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 'Italian restaurant' category, as we need info on Italian restaurants in the neighborhood.

```
# Category IDs corresponding to Italian restaurants were taken from Foursquare
```

```
food_category = '4d4b7105d754a06374d81259' # 'Root' category for all food-related venues
```

```
italian_restaurant_categories = ['4bf58dd8d48988d110941735', '55a5a1ebe4b013909087cbb6', '55a5a1ebe4b013909087cb7c',
```

'55a5a1ebe4b013909087cba7','55a5a1ebe4b013909087cba1','55a5a1ebe4b013909087cba4',
'55a5a1ebe4b013909087cb95','55a5a1ebe4b013909087cb89','55a5a1ebe4b013909087cb9b',
'55a5a1ebe4b013909087cb98','55a5a1ebe4b013909087cbbf','55a5a1ebe4b013909087cb79',
'55a5a1ebe4b013909087cbb0','55a5a1ebe4b013909087cbb3','55a5a1ebe4b013909087cb74',
'55a5a1ebe4b013909087cbaa','55a5a1ebe4b013909087cb83','55a5a1ebe4b013909087cb8c',
'55a5a1ebe4b013909087cb92','55a5a1ebe4b013909087cb8f','55a5a1ebe4b013909087cb86',
'55a5a1ebe4b013909087cbb9','55a5a1ebe4b013909087cb7f','55a5a1ebe4b013909087cbbc',
'55a5a1ebe4b013909087cb9e','55a5a1ebe4b013909087cbc2','55a5a1ebe4b013909087cbad']

Total number of restaurants: 1042

Total number of Italian restaurants: 101

Percentage of Italian restaurants: 9.69%

Average number of restaurants in neighborhood: 4.837912087912088

List of Italian restaurants

('59de862d5ba0465cb17712fa', 'Buona', 41.89321035992132, -
87.61765524594212, '613 N. McClurg Court, Chicago, IL 60611, United
States', 183, True, -5377364.632266688, 11520225.170559894)
('5893d651a8b75947f57a394c', 'Coco Pazzo Cafe', 41.89262157193581, -
87.62210929928445, '212 E Ohio St, Chicago, IL 60611, United States',
329, True, -5377284.706283157, 11520764.353846686)
('51929ff1498e88f22e86e0db', 'Tre Soldi', 41.89263532067262, -
87.62199460296691, '212 E Ohio St, Chicago, IL 60611, United States',
332, True, -5377286.9814082, 11520750.541650802)
('4ec316c49a524f6c471eaa9a', 'Francesca's on Chestnut',
41.89841184265746, -87.62213776268716, '200 E Chestnut St, Chicago, IL
60611, United States', 275, True, -5376397.890122146,
11520471.80924886)
('597ce4e0dd84420c834bb970', 'Torali - Italian Steak',
41.89845436210819, -87.6224268612446, '160 E Pearson St (Michigan),
Chicago, IL 60611, United States', 251, True, -5376380.353447471,
11520502.675198432)
('4a26e91bf964a520ed7e1fe3', 'Spiaggia', 41.90071807326922, -
87.62428580523864, '980 N Michigan Ave (at Oak St), Chicago, IL 60611,
United States', 288, True, -5375963.159628295, 11520599.466788312)
('4f8ca8bde4b013a983855b4d', 'Cafe Spiaggia', 41.90034640572532, -
87.62426070727044, '980 N Michigan Ave, Chicago, IL 60611, United
States', 317, True, -5376020.965068638, 11520615.585188359)
('4e99c7cf4fc602a58ade9ec1', 'Radisson Blu - Filini Restaurant and
Bar', 41.886385, -87.620102, '221 N Columbus Dr (Radisson Blu Aqua
Hotel), Chicago, IL 60601, United States', 338, True, -
5378315.438141202, 11520853.535203282)

```
('4b215953f964a520843a24e3', 'Sopraffina', 41.88476886982222, -  
87.62153948006629, '200 E Randolph, Chicago, IL 60601, United States',  
216, True, -5378507.808726274, 11521100.458613314)  
('4abc187ff964a520588620e3', 'Volare Ristorante Italiano',  
41.89172060243636, -87.62259046540213, '201 E Grand Ave (at St Clair  
St), Chicago, IL 60611, United States', 27, True, -5377404.164462861,  
11520865.383932121)
```

Total: 101

Restaurants around location

Restaurants around location 101: Safari Cafe

Restaurants around location 102:

Restaurants around location 103:

Restaurants around location 104:

Restaurants around location 105:

Restaurants around location 106:

Restaurants around location 107: III Forks Prime Steakhouse

Restaurants around location 108: Carsons Steak And Ribs, Buona, Niu

Japanese Fusion Lounge, Bombay Wraps, Bellwether Meeting House &

Eatery, Shula's Steak House, Flamingo Bar and Grill

Restaurants around location 109: LYFE Kitchen, Buona, the Albert,

Beatrix, Bombay Wraps, GRK Greek Kitchen, Nando's Peri-Peri,

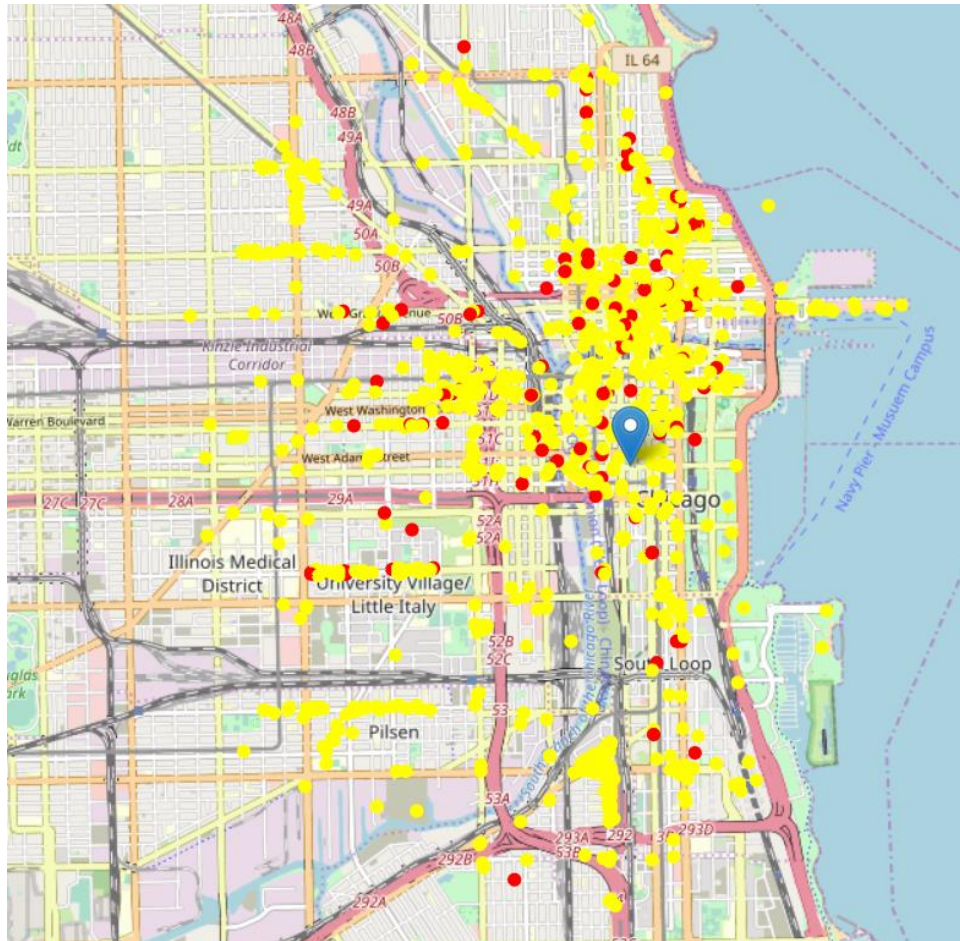
Markethouse Restaurant

Restaurants around location 110: Marisol, Cafecito, Francesca's on

Chestnut, The Signature Room at the 95th, Mity Nice, Foodease, Le

Petit Paris, Harry Caray's 7th Inning Stretch

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 East Chicago, 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 Chicago 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 East Chicago** . 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 Chicago - 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=300m:
4.837912087912088

| | Addresses | Latitude | Longitude | X | Y | Distance from center | Restaurants in area |
|---|-------------------------------------|-----------|------------|---------------|--------------|----------------------|---------------------|
| 0 | Cook County, IL | 41.878741 | -87.580104 | -5.381011e+06 | 1.151667e+07 | 5992.495307 | 0 |
| 1 | 1000 E Grand Ave, Chicago, IL 60611 | 41.882270 | -87.581676 | -5.380411e+06 | 1.151667e+07 | 5840.376700 | 0 |
| 2 | 1000 E Grand Ave, Chicago, IL 60611 | 41.885798 | -87.583248 | -5.379811e+06 | 1.151667e+07 | 5747.173218 | 0 |
| 3 | 1000 E Grand Ave, Chicago, IL 60611 | 41.889327 | -87.584821 | -5.379211e+06 | 1.151667e+07 | 5715.767665 | 0 |
| 4 | 1000 E Grand Ave, Chicago, IL 60611 | 41.892856 | -87.586394 | -5.378611e+06 | 1.151667e+07 | 5747.173218 | 0 |

| | Addresses | Latitude | Longitude | X | Y | Distance from center | Restaurants in area |
|---|-------------------------------------|-----------|------------|---------------|--------------|----------------------|---------------------|
| 5 | 1000 E Grand Ave, Chicago, IL 60611 | 41.896385 | -87.587967 | -5.378011e+06 | 1.151667e+07 | 5840.376700 | 0 |
| 6 | 1003 E Ohio St, Chicago, IL 60611 | 41.899915 | -87.589540 | -5.377411e+06 | 1.151667e+07 | 5992.495307 | 0 |
| 7 | Cook County, IL | 41.872432 | -87.581835 | -5.381911e+06 | 1.151719e+07 | 5855.766389 | 0 |
| 8 | Cook County, IL | 41.875960 | -87.583407 | -5.381311e+06 | 1.151719e+07 | 5604.462508 | 0 |
| 9 | 1000 E Grand Ave, Chicago, IL 60611 | 41.879488 | -87.584979 | -5.380711e+06 | 1.151719e+07 | 5408.326913 | 0 |

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

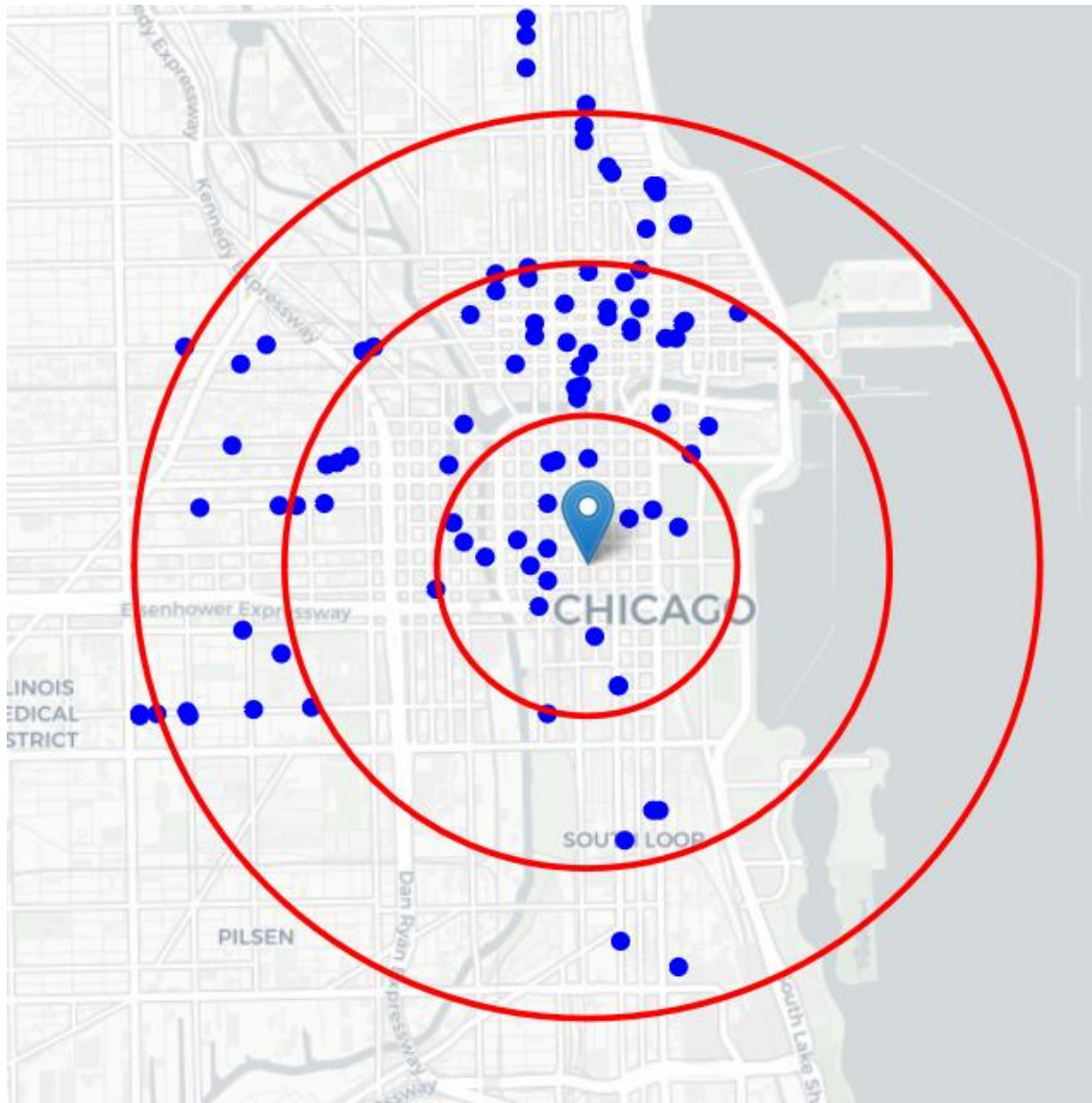
| | Address | Latitude | Longitude | X | Y | Distance from center | Restaurants in area | Distance to Italian restaurant |
|---|-------------------------------------|-----------|------------|---------------|--------------|----------------------|---------------------|--------------------------------|
| 0 | Cook County, IL | 41.878741 | -87.580104 | -5.381011e+06 | 1.151667e+07 | 5992.495307 | 0 | 4977.171703 |
| 1 | 1000 E Grand Ave, Chicago, IL 60611 | 41.882270 | -87.581676 | -5.380411e+06 | 1.151667e+07 | 5840.376700 | 0 | 4679.546040 |

| | Address | Latitude | Longitude | X | Y | Distance from center | Restaurants in area | Distance to Italian restaurant |
|---|-------------------------------------|-----------|------------|---------------|--------------|----------------------|---------------------|--------------------------------|
| 2 | 1000 E Grand Ave, Chicago, IL 60611 | 41.885798 | -87.583248 | -5.379811e+06 | 1.151667e+07 | 5747.173218 | 0 | 4315.945403 |
| 3 | 1000 E Grand Ave, Chicago, IL 60611 | 41.889327 | -87.584821 | -5.379211e+06 | 1.151667e+07 | 5715.767665 | 0 | 4006.535991 |
| 4 | 1000 E Grand Ave, Chicago, IL 60611 | 41.892856 | -87.586394 | -5.378611e+06 | 1.151667e+07 | 5747.173218 | 0 | 3767.927358 |
| 5 | 1000 E Grand Ave, Chicago, IL 60611 | 41.896385 | -87.587967 | -5.378011e+06 | 1.151667e+07 | 5840.376700 | 0 | 3614.169684 |
| 6 | 1003 E Ohio St, Chicago, IL 60611 | 41.899915 | -87.589540 | -5.377411e+06 | 1.151667e+07 | 5992.495307 | 0 | 3556.285764 |
| 7 | Cook County, IL | 41.872432 | -87.581835 | -5.381911e+06 | 1.151719e+07 | 5855.766389 | 0 | 5062.904184 |
| 8 | Cook County, IL | 41.875960 | -87.583407 | -5.381311e+06 | 1.151719e+07 | 5604.462508 | 0 | 4732.945484 |
| 9 | 1000 E Grand Ave, Chicago, IL 60611 | 41.879488 | -87.584979 | -5.380711e+06 | 1.151719e+07 | 5408.326913 | 0 | 4377.977379 |

Average distance to closest Italian restaurant from each area center:
1419.4455654086992

OK, so **on average Italian restaurant can be found within ~1418m** from every area center candidate.

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



Looks like a few pockets of low restaurant density closest to city center can be found **south, south-west and east from Chicago**.

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

Based on this we will now focus our analysis on areas *south-west, south, south-west and west from East Chicago* - 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 with large low restaurant density north-east from city center, however this borough is less interesting to stakeholders as it's mostly residential and less popular with tourists).

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.

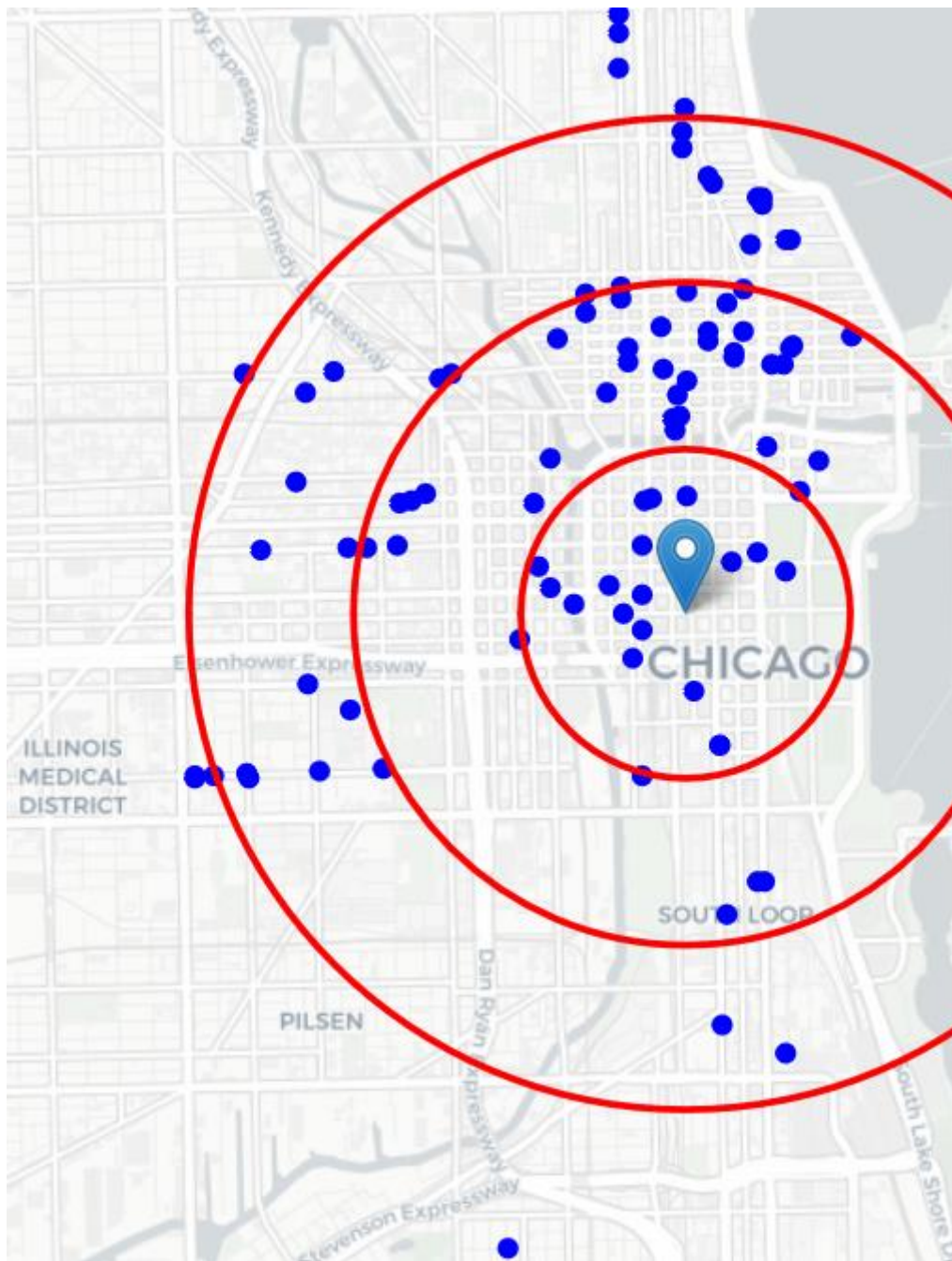
OK. Now let's calculate two most important things for each location candidate: **number of restaurants in vicinity** (we'll use radius of **250 meters**) and **distance to closest Italian restaurant**.

| | Latitude | Longitude | X | Y | Restaurants nearby | Distance to Italian restaurant |
|---|-----------|------------|---------------|--------------|--------------------|--------------------------------|
| 0 | 41.888612 | -87.599506 | -5.378761e+06 | 1.151838e+07 | 0 | 2309.738962 |
| 1 | 41.889200 | -87.599768 | -5.378661e+06 | 1.151838e+07 | 0 | 2250.715085 |
| 2 | 41.885207 | -87.598744 | -5.379311e+06 | 1.151847e+07 | 0 | 2581.475998 |
| 3 | 41.885795 | -87.599006 | -5.379211e+06 | 1.151847e+07 | 0 | 2544.602881 |
| 4 | 41.886383 | -87.599269 | -5.379111e+06 | 1.151847e+07 | 0 | 2474.524827 |
| 5 | 41.886972 | -87.599531 | -5.379011e+06 | 1.151847e+07 | 0 | 2405.015060 |
| 6 | 41.887560 | -87.599794 | -5.378911e+06 | 1.151847e+07 | 0 | 2337.717212 |
| 7 | 41.888148 | -87.600056 | -5.378811e+06 | 1.151847e+07 | 0 | 2272.827772 |
| 8 | 41.888736 | -87.600318 | -5.378711e+06 | 1.151847e+07 | 0 | 2210.558844 |
| 9 | 41.889324 | -87.600581 | -5.378611e+06 | 1.151847e+07 | 0 | 2151.138007 |

OK. Let us now **filter** those locations: we're interested only in **locations with no more than two restaurants in radius of 250 meters**, and **no Italian restaurants in radius of 400 meters**.

Locations with no more than two restaurants nearby: 1265
Locations with no Italian restaurants within 400m: 1262
Locations with both conditions met: 1110

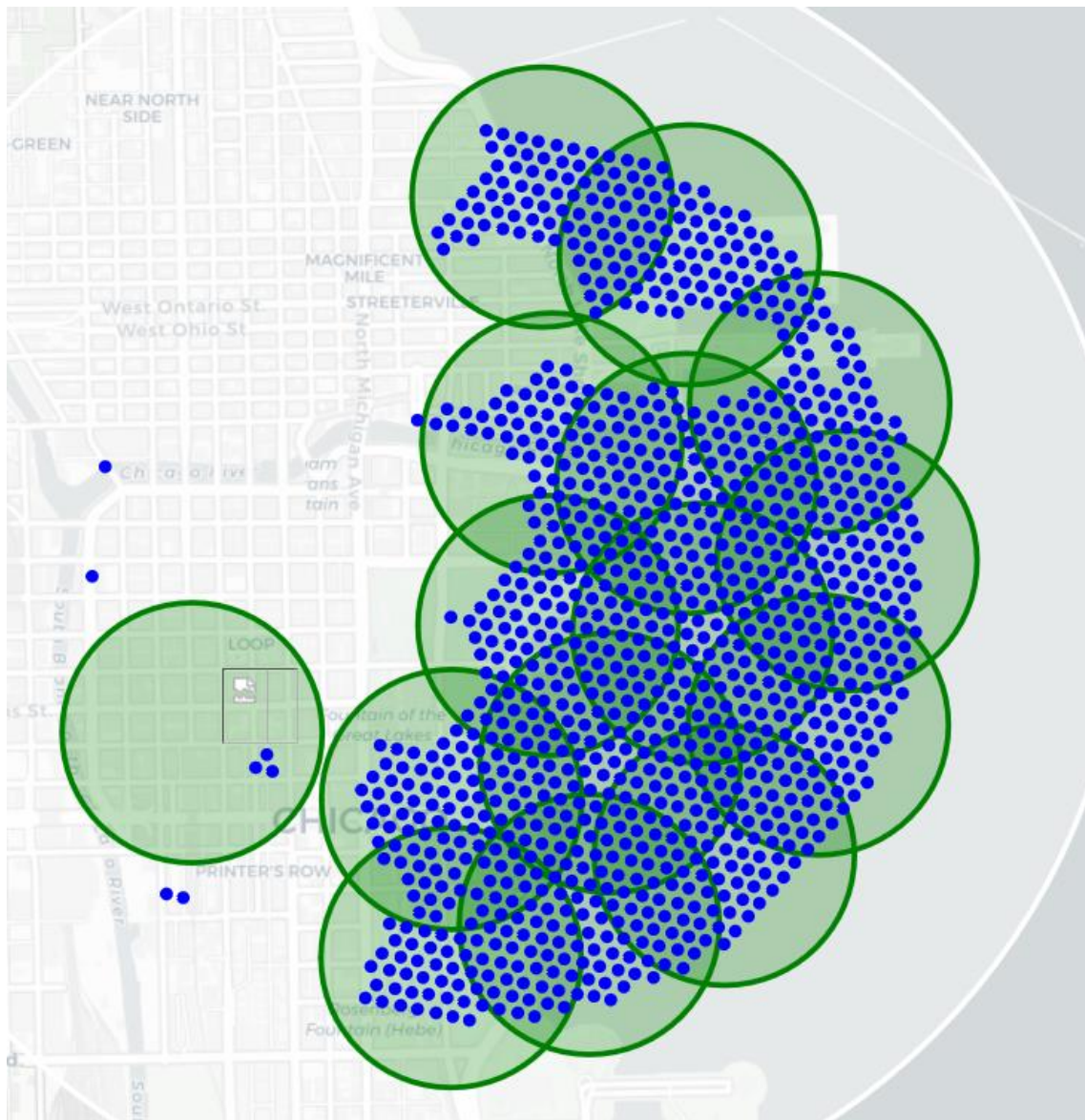
Let's see how this looks on a map.



Looking good. We now have a bunch of locations fairly close to East Chicago (mostly in Quincy and south-west corner of Mitte boroughs), and we know that each of those locations has no more than two restaurants in radius of 250m, and no Italian restaurant closer than 400m. Any of those locations is a potential candidate for a new Italian restaurant, at least based on nearby competition.

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.

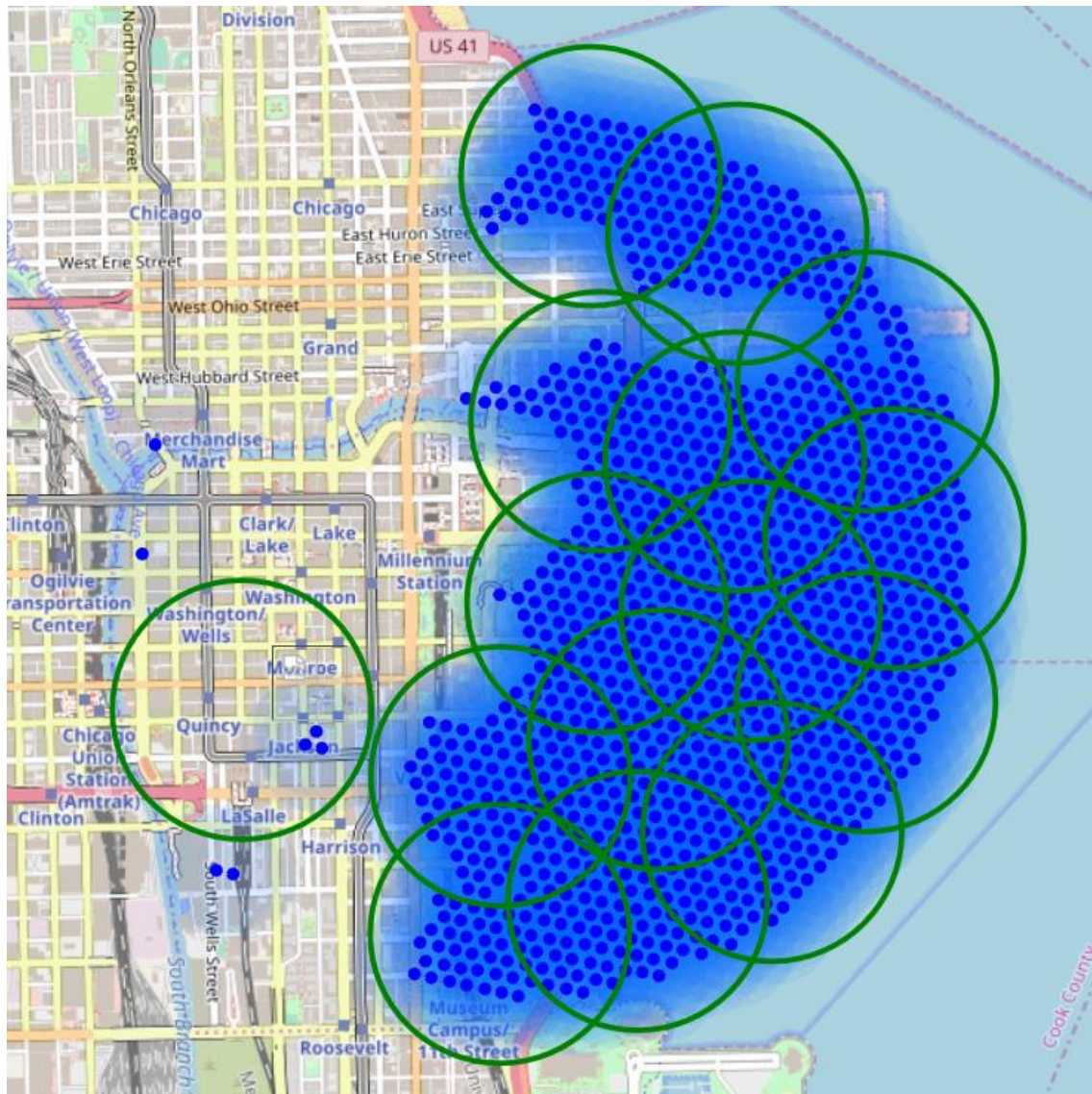
Let us now **cluster** those locations to create **centers of zones containing good locations**. Those zones, their centers and addresses will be the final result of our analysis.



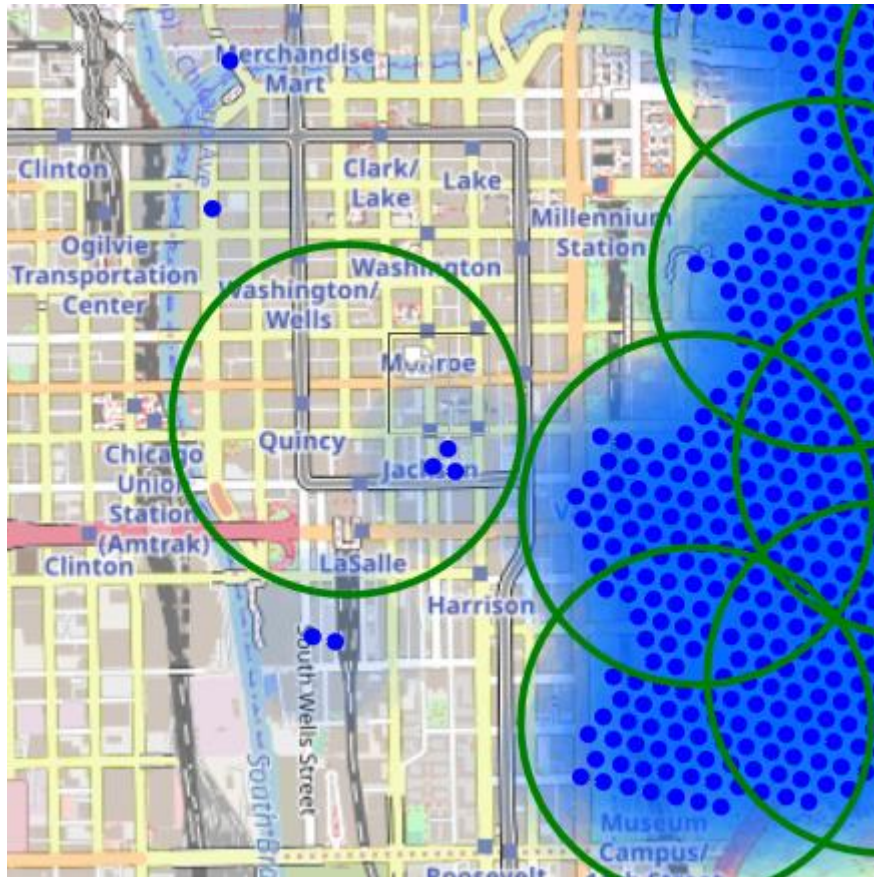
Not bad - our clusters represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates.

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 **Quincy**:



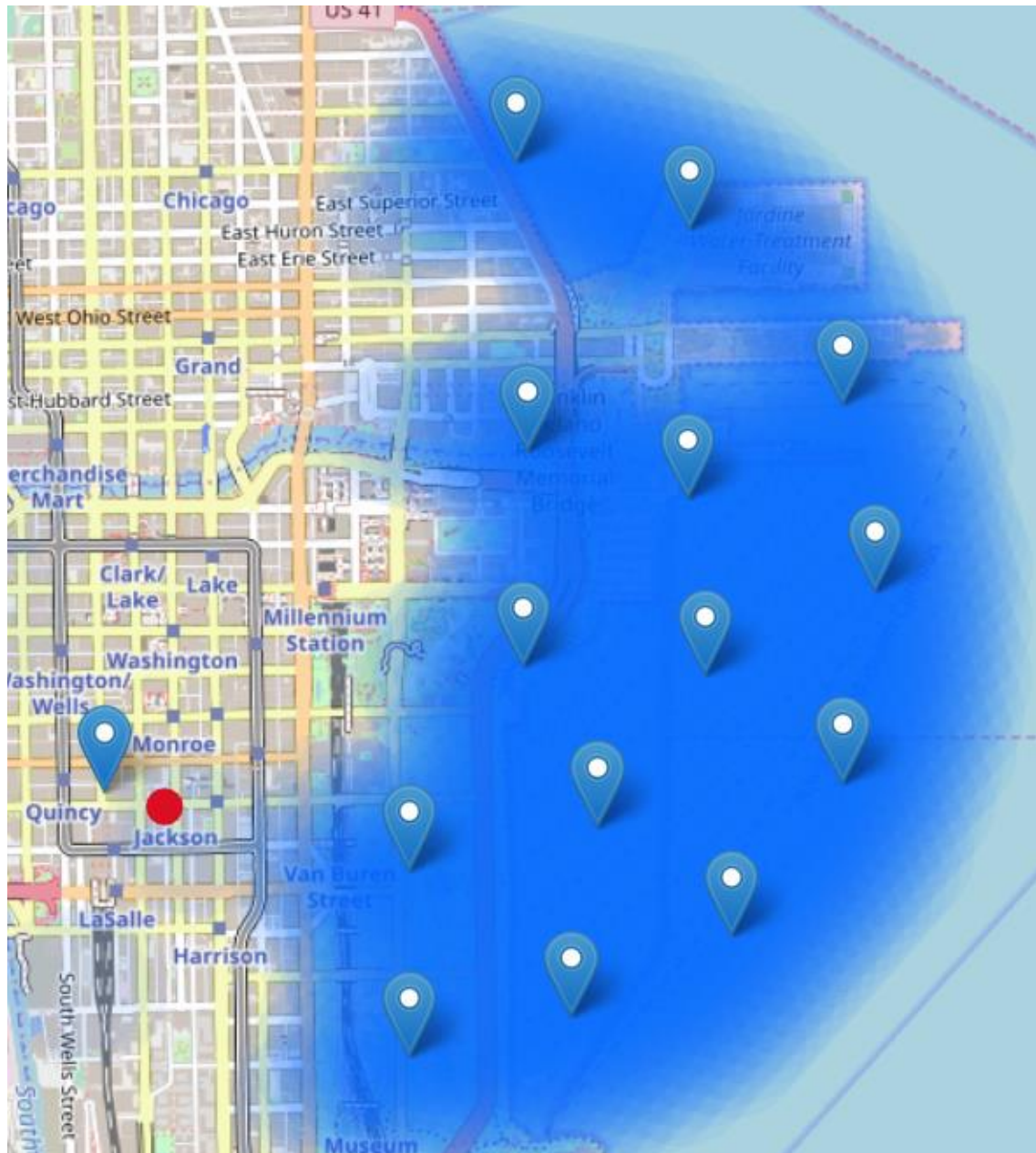
Finally, let's **reverse geocode** those candidate area centers to get the addresses which can be presented to stakeholders.

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

| | |
|---|--------------------|
| 880 S Lake Shore Dr, Chicago, IL 60611 | => 2.2km from East |
| Chicago | |
| Unnamed Road, Chicago, IL 60601 | => 2.6km from East |
| Chicago | |
| Unnamed Road, Chicago, IL 60611 | => 3.7km from East |
| Chicago | |
| Chicago, IL 60605 | => 2.7km from East |
| Chicago | |
| 447 S Columbus Dr, Chicago, IL 60612 | => 1.2km from East |
| Chicago | |
| Cook County, IL | => 3.5km from East |
| Chicago | |
| 603 E North Water St, Chicago, IL 60611 | => 2.4km from East |
| Chicago | |
| 840 E Grand Ave, Chicago, IL 60611 | => 3.7km from East |
| Chicago | |
| 235 S LaSalle St, Chicago, IL 60604 | => 0.3km from East |
| Chicago | |
| Cook County, IL | => 3.2km from East |
| Chicago | |
| 1000 S Columbus Dr, Chicago, IL 60605 | => 1.7km from East |
| Chicago | |

| | |
|---|--------------------|
| 101 N Lake Shore Dr, Chicago, IL 60601 Chicago | => 1.8km from East |
| Unnamed Road, Chicago, IL 60601 Chicago | => 2.0km from East |
| 791 N Lake Shore Dr, Chicago, IL 60611 Chicago | => 3.5km from East |
| Heliport, Chicago, IL 60601 Chicago | => 2.9km from East |

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 east Chicago, and about half of those less than 2km from east Chicago). 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 zone is located in Quincy 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 Chicago (~2000 in our initial area of interest which was 12x12km around East Chicago), there are pockets of low restaurant density fairly close to city center. Highest concentration of restaurants was detected north and east from East Chicago, so we focused our attention to areas south, south-west and west, corresponding to borough Quincy, but our attention was focused on Quincy 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-west from East Chicago) 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 East Chicago 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 Chicago 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 (Quincy), 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. Another important decision factor is the economic factor, in this study budgets have not been taken into account, since the area that may need more money can be determined.