Capstone Project - The Battle of the Neighborhoods

Applied Data Science Capstone by IBM/Coursera

By: Bronwyn B.

Table of contents

Introduction

The objective of this project is to find an optimal location for an art gallery in Charlotte, NC. This report will be targeted to stakeholders interested in opening an **Art Gallery** in **Charlotte, NC.**

Based on the current established art galleries in the Charlotte area we will try to detect **locations that are not already crowded with art galleries**. We are mostly interested in **areas with no art galleries 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 tools to generate a few most promising neighborhoods based on this criterion. 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 art galleries in the neighborhood (any type of art gallery)
- number of and distance to art galleries 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 art galleries and their type and location in every neighborhood will be obtained using **Foursquare API**
- coordinate of Charlotte, NC city center will be obtained using **Google Maps API geocoding** of well-known Charlotte, NC location (Spectrum Center)

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 Charlotte, NC city center.

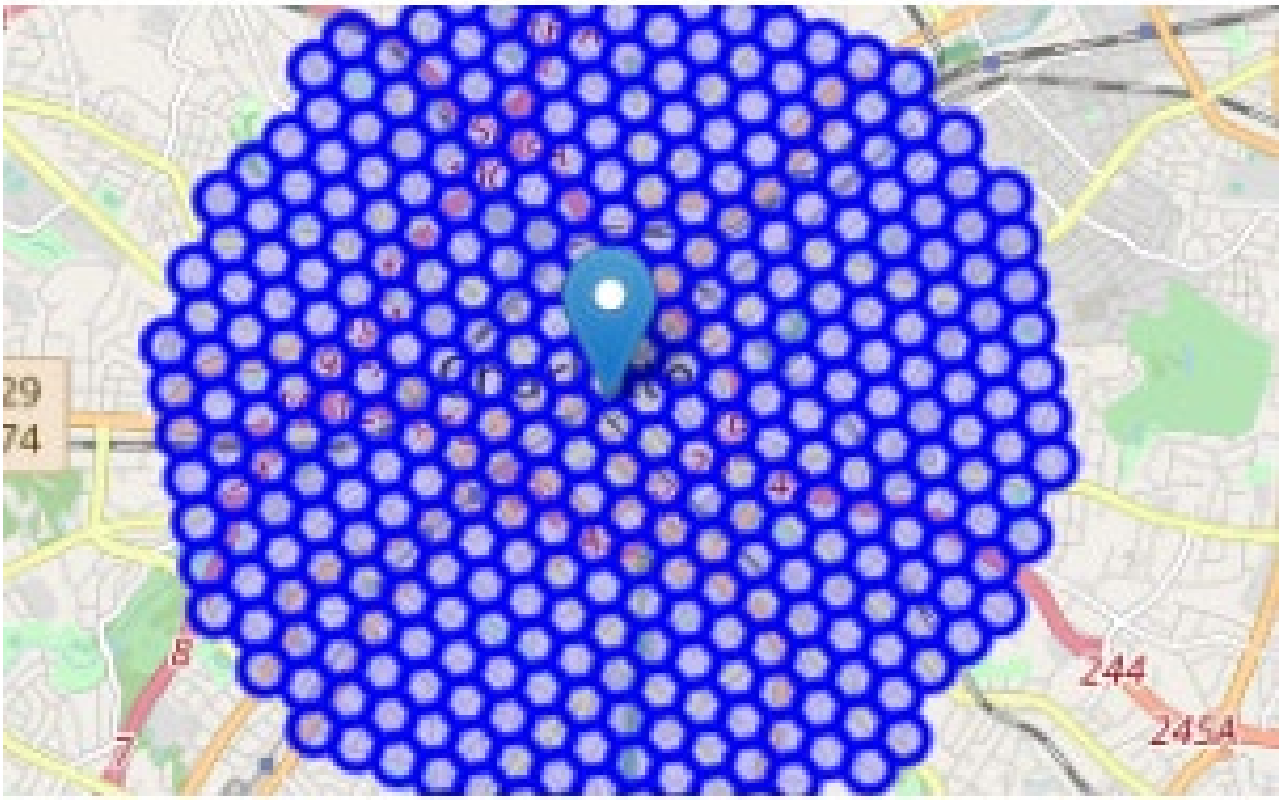
Let's first find the latitude & longitude of Charlotte, NC city center, using specific, well known address and Google Maps geocoding API.

- Coordinate of Spectrum Center, Charlotte, North Carolina: [35.2251809, -80.8393037]

Now let's create a grid of area candidates, equally spaced, centered around city center and within ~6km from Spectrum Center. 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 its neighbors**.



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

	Address	Latitude	Longitude	X	Y	Distance from center
0	1634 Chatham Ave 28205, USA	35.221032	-80.801112	-6.734149e+06	1.090178e+07	5992.495307
1	1918 Winter St 28205, USA	35.224150	-80.801775	-6.733549e+06	1.090178e+07	5840.376700
2	2237 Chambwood Dr 28205, USA	35.227268	-80.802439	-6.732949e+06	1.090178e+07	5747.173218
3	3304 Josephine Ct 28205, USA	35.230386	-80.803103	-6.732349e+06	1.090178e+07	5715.767665
4	2409 Daniel St 28205, USA	35.233505	-80.803767	-6.731749e+06	1.090178e+07	5747.173218
5	2610 Duncan Ave 28205, USA	35.236623	-80.804431	-6.731149e+06	1.090178e+07	5840.376700
6	1120 Leigh Ave 28205, USA	35.239742	-80.805095	-6.730549e+06	1.090178e+07	5992.495307
7	2505 McClintock Rd 28205, USA	35.215884	-80.803406	-6.735049e+06	1.090230e+07	5855.766389
8	2133 Central Ave 28205, USA	35.219001	-80.804070	-6.734449e+06	1.090230e+07	5604.462508
9	2214 Randall St 28205, USA	35.222119	-80.804734	-6.733849e+06	1.090230e+07	5408.326913
10	1923 Kenwood Ave 28205, USA	35.225237	-80.805398	-6.733249e+06	1.090230e+07	5273.518749
11	1908 Belvedere Ave 28205, USA	35.228355	-80.806062	-6.732649e+06	1.090230e+07	5204.805472
12	2213 Thurmond Pl 28205, USA	35.231473	-80.806726	-6.732049e+06	1.090230e+07	5204.805472
13	2412 Lola Ave 28205, USA	35.234592	-80.807391	-6.731449e+06	1.090230e+07	5273.518749
14	933 Drummond Ave 28205, USA	35.237711	-80.808055	-6.730849e+06	1.090230e+07	5408.326913
15	2516 Pinckney Ave 28205, USA	35.240830	-80.808720	-6.730249e+06	1.090230e+07	5604.462508
16	2717 Yadkin Ave 28205, USA	35.243949	-80.809384	-6.729649e+06	1.090230e+07	5855.766389
17	2529 Bay St 28205, USA	35.210736	-80.805700	-6.735949e+06	1.090282e+07	5901.694672
18	2415 Shenandoah Ave 28205, USA	35.213853	-80.806364	-6.735349e+06	1.090282e+07	5556.077753
19	2300 McClintock Rd 28205, USA	35.216971	-80.807028	-6.734749e+06	1.090282e+07	5256.424640
20	1900 Central Ave 28205, USA	35.220088	-80.807692	-6.734149e+06	1.090282e+07	5010.987927

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

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

- Reverse geocoding check: Address of [35.2251809, -80.8393037] is: 290 E 5th St, Charlotte, NC 28202

Foursquare

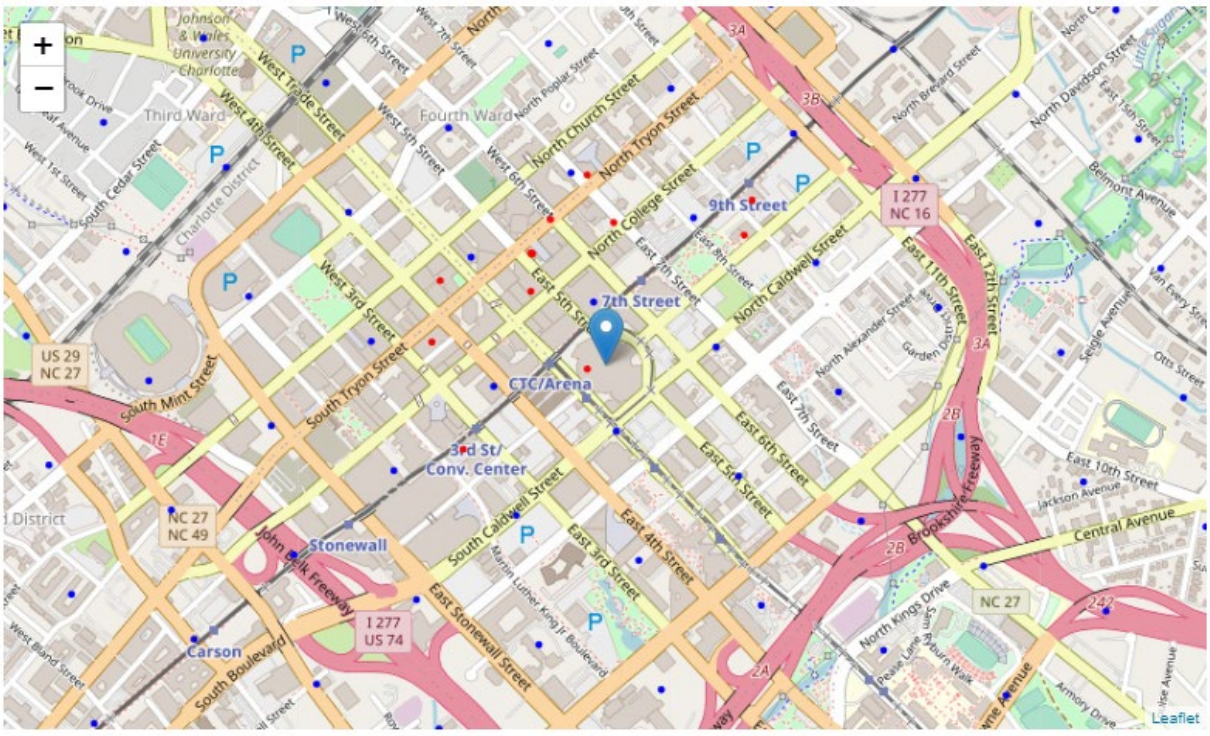
Now that we have our location candidates, let's use Foursquare API to get info on art galleries in each neighborhood.

We're interested in venues in 'art' category. So, we will include in our list only venues that have 'art gallery' in category name.

Art galleries around location

	Name	Lat	Lng	Distance
0	Sip & Stroll	35.225124	-80.839825	63
1	Bank Of America Gallery	35.226996	-80.841495	306
2	Gallery W.D.O.	35.227899	-80.841477	383
3	Foundation For Carolina	35.227910	-80.841443	382
4	Sozo Gallery	35.227922	-80.841501	386
5	Northwest School of the Arts Rocking Chair @ S...	35.228661	-80.839043	403
6	Sonia & Isaac Luski Art Gallery	35.228735	-80.840917	442
7	A Pinata 4 U, LLC	35.223185	-80.843496	446
8	Junior Achievement	35.225775	-80.844425	487
9	UNC gallery	35.228355	-80.835182	515
10	Narasinga jaya reklame	35.227256	-80.844179	521
11	Hodges Taylor Gallery	35.229809	-80.839830	535
12	Clayworks Studio & Gallery	35.229197	-80.834952	599

Let's now see all the collected art galleries in our area of interest on map.



Now we have all the art galleries in area within few kilometers from Spectrum Center. We also know which art gallery 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 art gallery

Methodology

In this project we will direct our efforts on detecting areas of Charlotte that have low art gallery density, particularly those with low number of art galleries. 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 art gallery within 6km from Charlotte center** (Spectrum Center). We have also **identified art gallery** (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of '**art gallery density**' across different areas of Charlotte - we will use **heatmaps** to identify a few promising areas close to center with low number of art galleries in general (*and* no art galleries 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 art galleries in radius of 250 meters**, and we want locations **without art galleries 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 art galleries in every area candidate**:

```
The number of art galleries in every area with radius 600m: 13.0
List of 10 Art Galleries
```

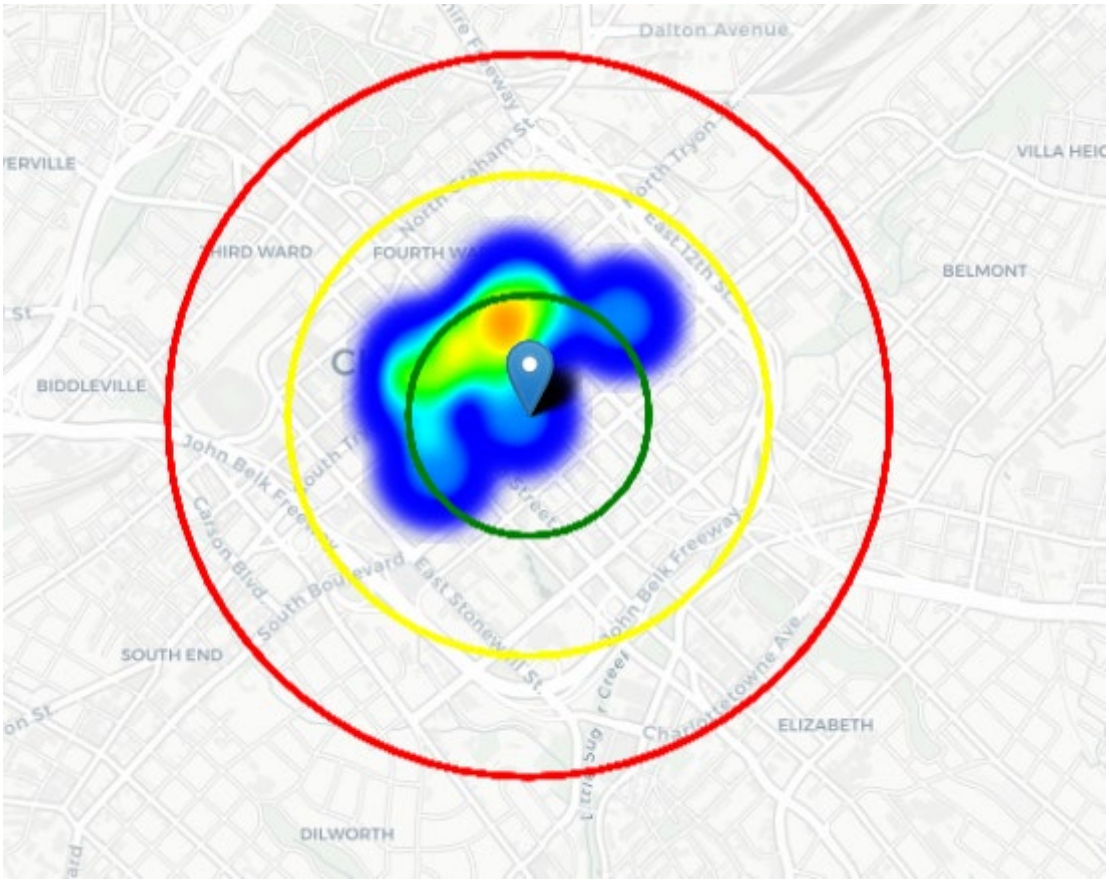
	Address	Latitude	Longitude	X	Y	Distance from center	Art galleries in area
0	1634 Chatham Ave 28205, USA	35.221032	-80.801112	-6.734149e+06	1.090178e+07	5992.495307	13
1	1918 Winter St 28205, USA	35.224150	-80.801775	-6.733549e+06	1.090178e+07	5840.376700	13
2	2237 Chambwood Dr 28205, USA	35.227268	-80.802439	-6.732949e+06	1.090178e+07	5747.173218	13
3	3304 Josephine Ct 28205, USA	35.230386	-80.803103	-6.732349e+06	1.090178e+07	5715.767665	13
4	2409 Daniel St 28205, USA	35.233505	-80.803767	-6.731749e+06	1.090178e+07	5747.173218	13
5	2610 Duncan Ave 28205, USA	35.236623	-80.804431	-6.731149e+06	1.090178e+07	5840.376700	13
6	1120 Leigh Ave 28205, USA	35.239742	-80.805095	-6.730549e+06	1.090178e+07	5992.495307	13
7	2505 McClintock Rd 28205, USA	35.215884	-80.803406	-6.735049e+06	1.090230e+07	5855.766389	13
8	2133 Central Ave 28205, USA	35.219001	-80.804070	-6.734449e+06	1.090230e+07	5604.462508	13
9	2214 Randall St 28205, USA	35.222119	-80.804734	-6.733849e+06	1.090230e+07	5408.326913	13

OK, now let's calculate the **distance to nearest art gallery from every area candidate center** (not only those within 600m - we want distance to closest one, regardless of how distant it is).

- Average distance to closest art gallery from each area center: 13.0

OK, so **on average art gallery can be found within ~600m** from every area center candidate. That's fairly close, so we need to filter our areas carefully!

Let's create a map showing **heatmap / density of art galleries** and try to extract some meaningful info from that. Also, let's show **borders of Charlotte** on our map and a few circles indicating distance of 1km, 2km and 3km from Spectrum Center.

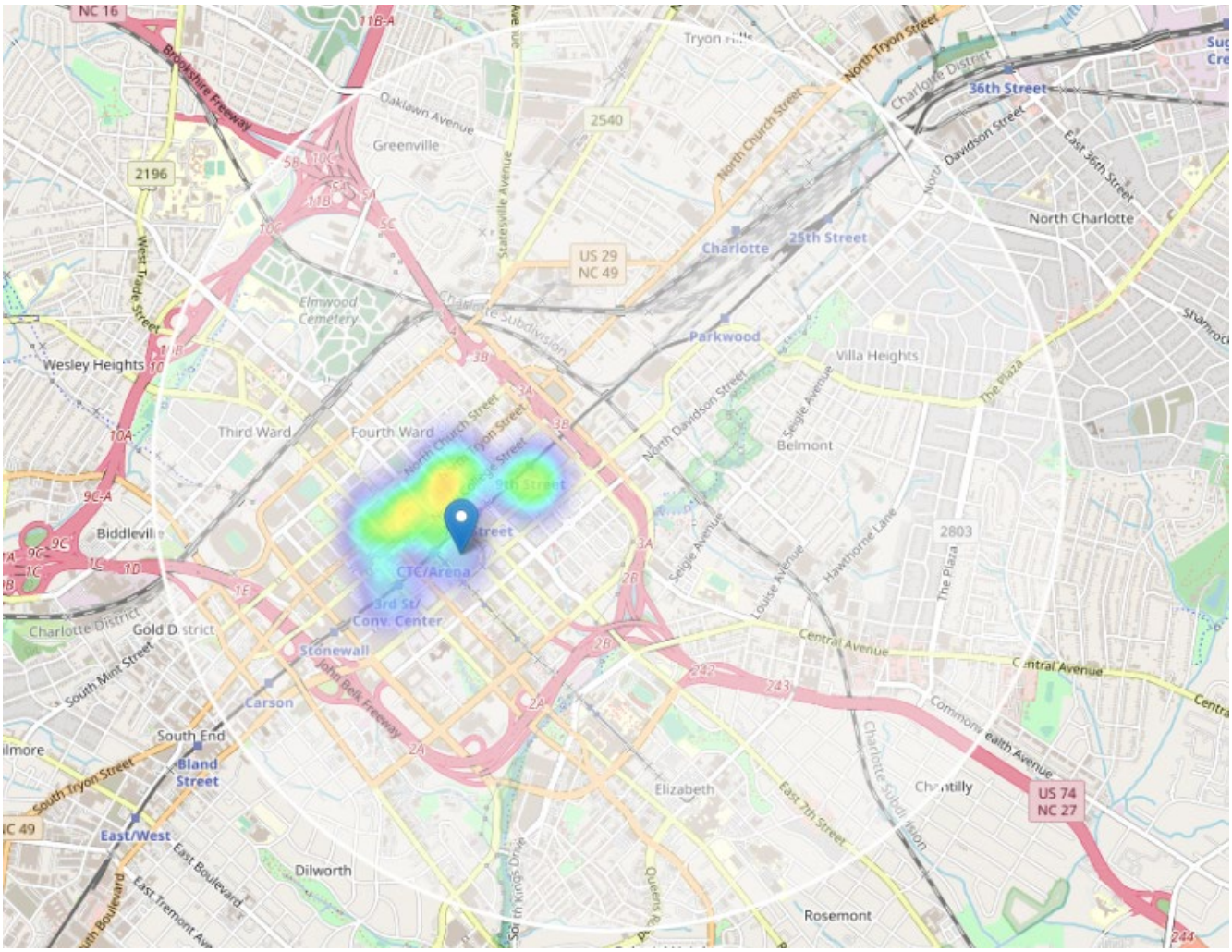


Looks like a few pockets of low art gallery density closest to city center can be found **South from Spectrum Center**.

This map is not so 'hot' (Art galleries represent a subset of ~15% of all art in Charlotte) but it also indicates higher density of existing art galleries directly north and west from Spectrum Center, with closest pockets of **low art galleries density positioned east, south-east and south from city center**.

Based on this we will now focus our analysis on areas *south-west, south, south-east and east from Charlotte center* - we will move the center of our area of interest and reduce its size to have a radius of **2.5km**. This places our location candidates mostly in **South Charlotte** (another potentially interesting borough is **South-East Charlotte** with large low art gallery density south-east from city center.

Let's define new, narrower region of interest, which will include low-art-gallery-count parts of East Charlotte and South Charlotte closest to Spectrum Center.



This nicely covers all the pockets of low art gallery density in South-East Charlotte closest to Charlotte center.

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

OK. Now let's calculate two most important things for each location candidate: **number of art galleries in vicinity** (we'll use radius of **400 meters**) and **distance to closest art galleries**.

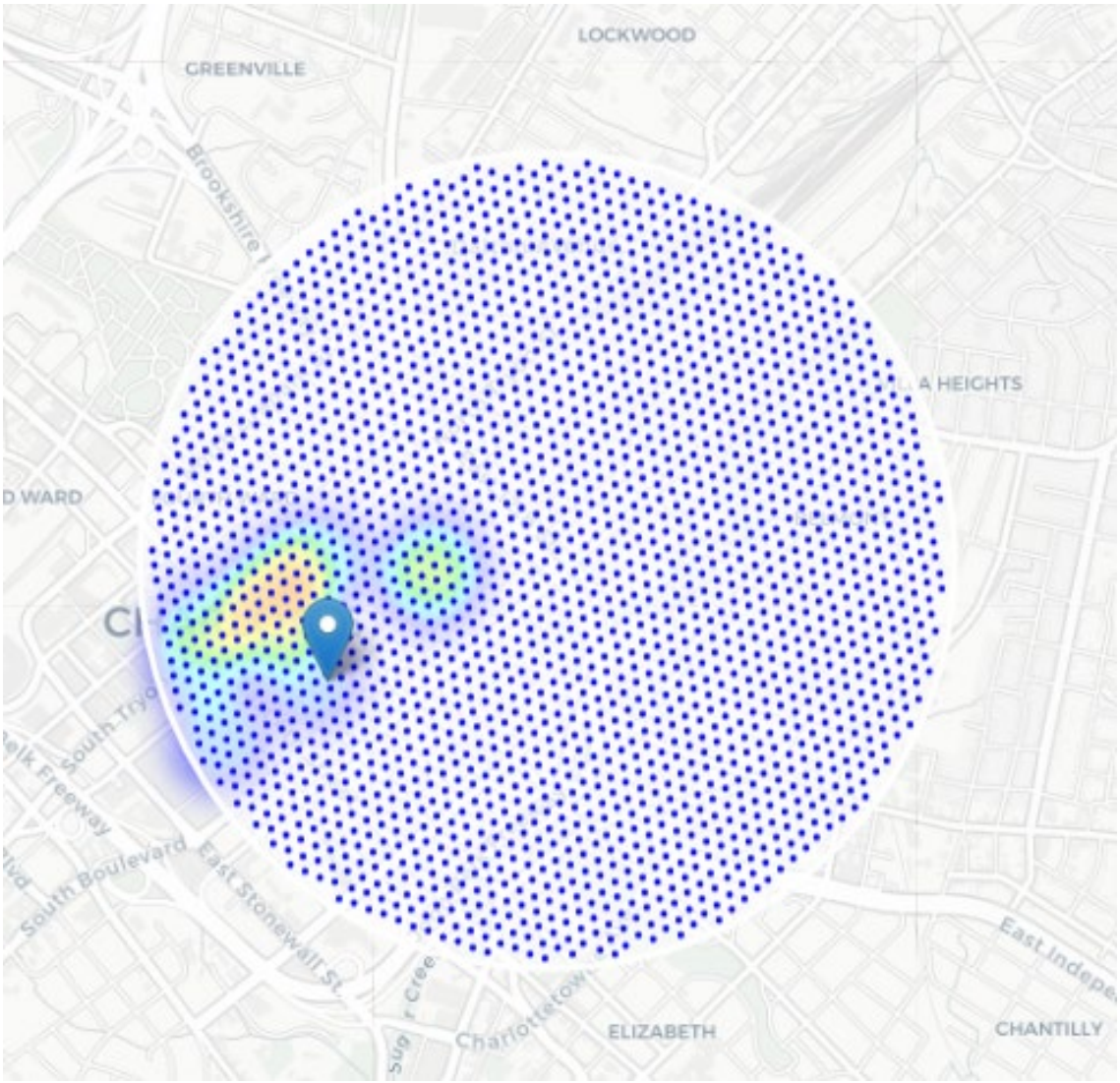
```
Total Art galleries in vicinity
5
```

	Name	Lat	Lng	Distance
0	Sip & Stroll	35.225124	-80.839825	63
1	Bank Of America Gallery	35.226996	-80.841495	306
2	Gallery W.D.O.	35.227899	-80.841477	383
3	Foundation For Carolina	35.227910	-80.841443	382
4	Sozo Gallery	35.227922	-80.841501	386

OK. Let us now **filter** those locations: we're interested only in **locations with no more than two art galleries in radius of 400 meters**, and **no art galleries in radius of 600 meters**.

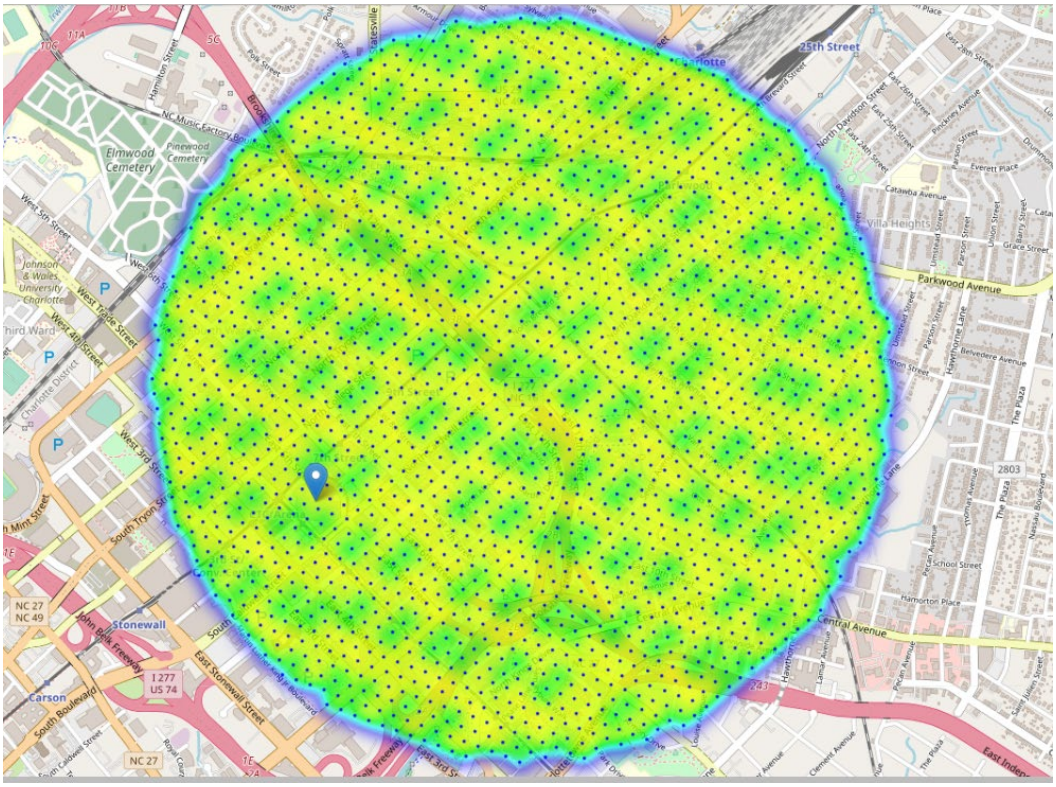
Locations with no more than two art galleries nearby: 2261
Locations with no art galleries within 400m: 2261
Locations with both conditions met: 2261

Let's see how this looks on a map.



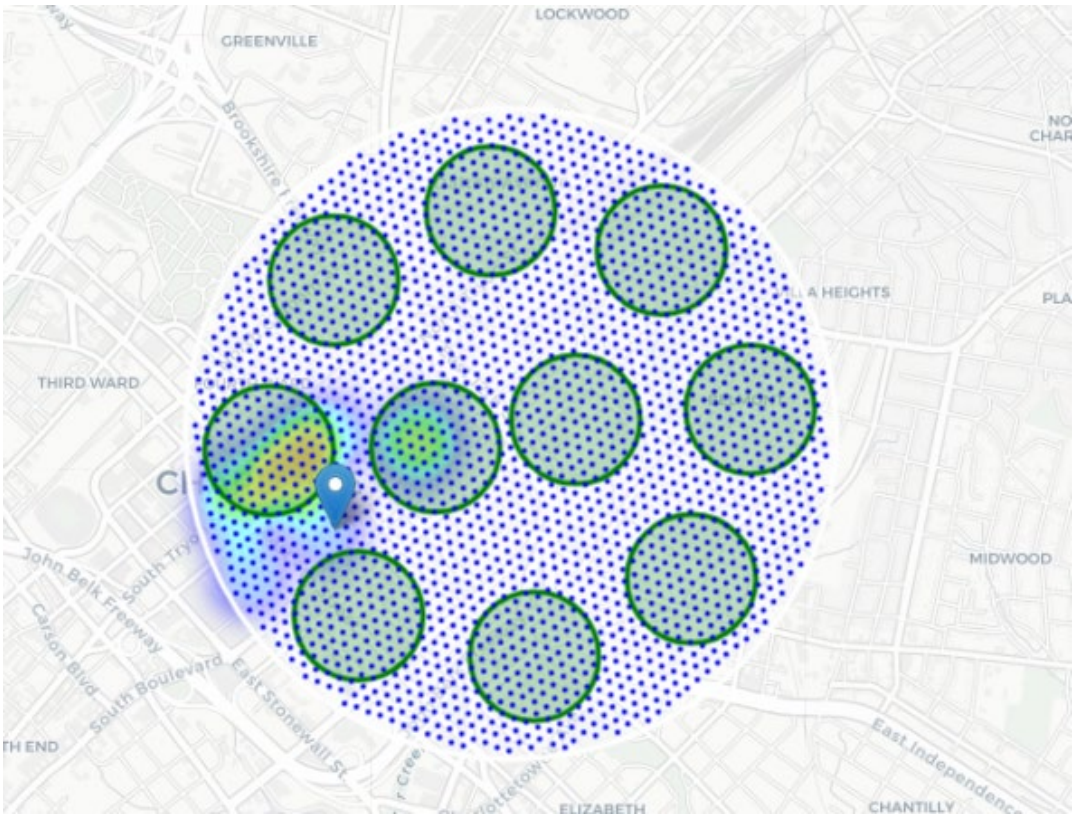
Looking good. We now have a bunch of locations fairly close to Spectrum Center (mostly in ???), and we know that each of those locations has no more than two art galleries in radius of 250m, and no art galleries closer than 400m. Any of those locations is a potential candidate for a new art gallery, 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 art galleries in vicinity, and *no* art galleries 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.

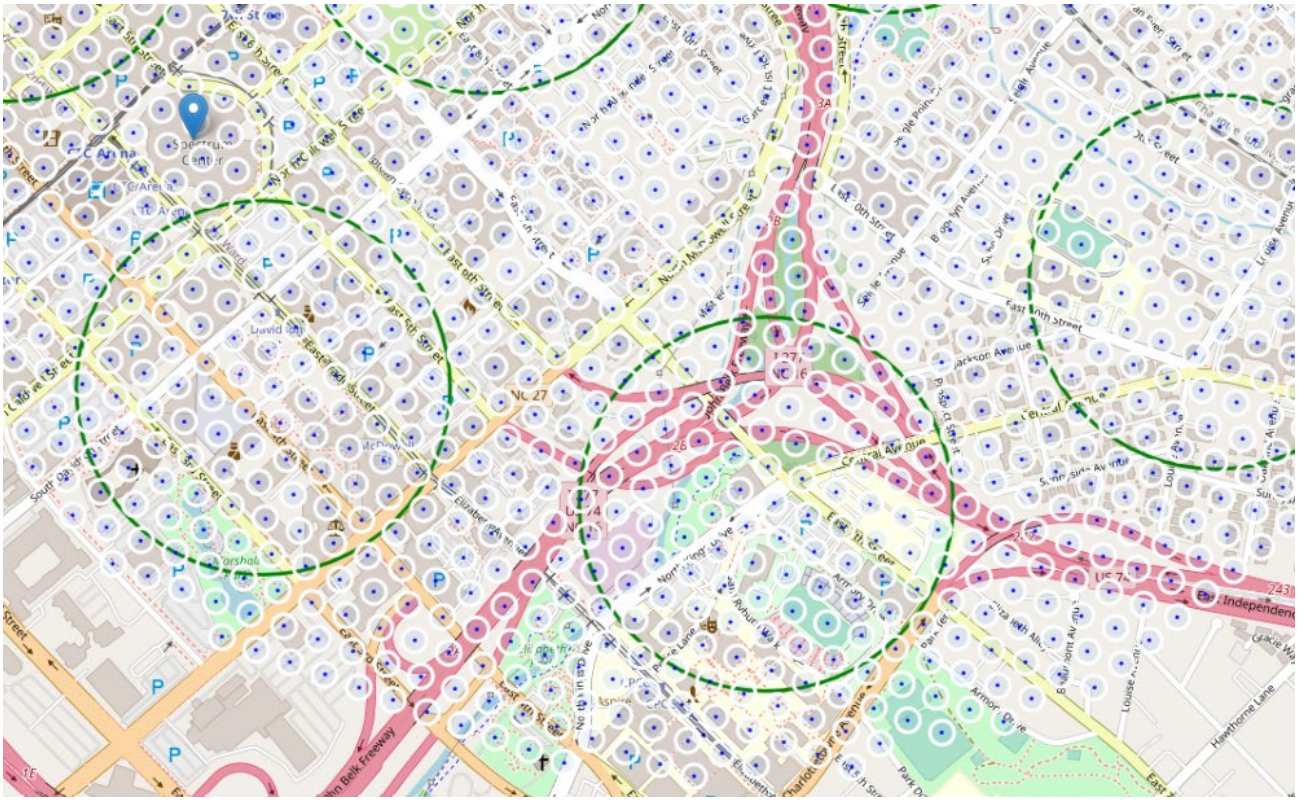


Not bad - our clusters represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates.

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 **South - Downtown Charlotte**:



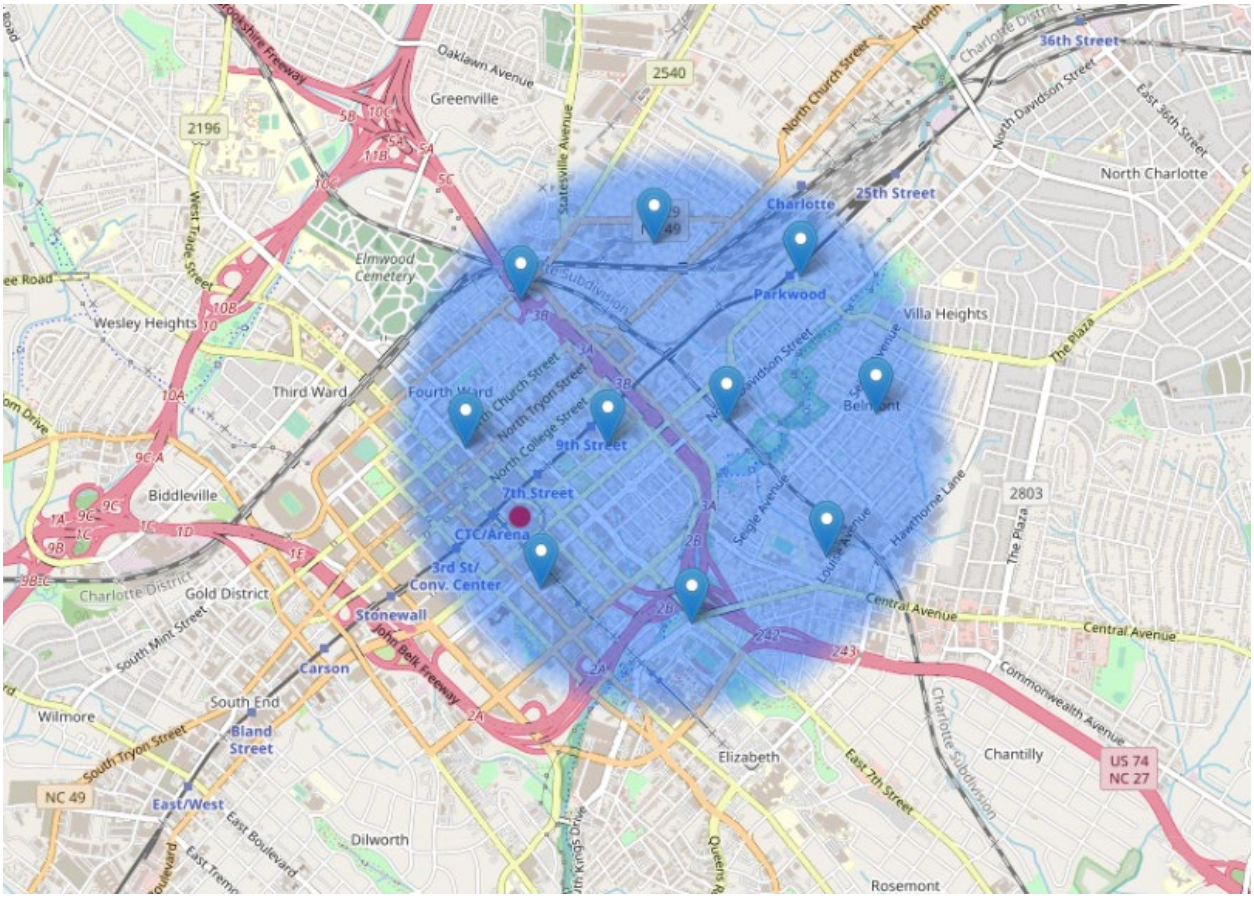
...and candidate areas in **Southeast - Downtown Charlotte**:

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
============================================================

401 E 9th St 28202, USA                 => 1.0km from Spectrum Center
1008 E 17th St 28205, USA               => 3.4km from Spectrum Center
829 Louise Ave 28204, USA               => 2.8km from Spectrum Center
301 Dalton Ave 28206, USA               => 2.8km from Spectrum Center
600 E Trade St 28202, USA               => 0.7km from Spectrum Center
940 N Davidson St 28206, USA            => 2.1km from Spectrum Center
407 N Kings Dr 28204, USA               => 1.9km from Spectrum Center
Parkwood 28206, USA                     => 3.4km from Spectrum Center
700 N Graham St 28202, USA              => 2.0km from Spectrum Center
127 N Tryon St 28202, USA               => 0.8km from Spectrum Center
```

This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of art galleries nearby, all zones being fairly close to city center (all less than 4km from Spectrum Center, and about half of those less than 2km from Spectrum Center). 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 art gallery locations. Most of the zones are located South and Southeast of Downtown Charlotte, 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 there is a small number of art galleries in Charlotte (~2000 in our initial area of interest which was 12x12km around Spectrum Center), there are pockets of low art gallery density fairly close to city center. Highest concentration of art galleries was detected north and west from Spectrum Center, so we focused our attention to areas south and southeast of downtown areas. Another area was identified as potentially interesting (north-east from Spectrum Center), but our attention was focused on South and Southeast downtown areas which offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics *and* a number of pockets of low art gallery density.

After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Spectrum Center) we first created a dense grid of location candidates (spaced 100m apart); those locations were then filtered so that those with more than two art galleries in radius of 250m and those 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 art gallery locations based on number of and distance to existing venues. This, of course, does not imply that those zones are actually optimal locations for a new art gallery! Purpose of this analysis was to only provide info on areas close to Spectrum Center but not crowded with existing art galleries - it is entirely possible that there is a very good reason for small number of art galleries in any of those areas, reasons which would make them unsuitable for a new art galleries 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 considered, and all other relevant conditions met.

Conclusion

Purpose of this project was to identify Charlotte areas close to Spectrum Center with low number of art galleries to aid stakeholders in narrowing down the search for optimal location for a new art gallery. By calculating art gallery density distribution from Foursquare data, we have first identified general areas that justify further analysis (South and Southeast Downtown), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby art galleries. Clustering of those locations was then performed 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 art gallery 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.