BEST LOCATION FOR A CHINESE FOOD RESTAURANT

JAVIER WANDURRAGA

APRIL 19, 2020

Introduction

- This project aims to find an optimal place to locate a Chinese food restaurant in Bogota, preferably close to Simon Bolivar Park. However, a site is being sought where the competition is as low as possible and for this it is required to analyze which places do not have nearby Chinese restaurants and if possible close to Simon Bolivar Park.
- To find this location, data science tools will be used to find nearby neighborhoods based on the criteria of having little commercial competition and in the end determine a list of possible places with the best possible location.

Data

- Based on the definition of the problem, the decision factors will be:
  - Number of restaurants in the neighborhood
  - Quantity and distance from Chinese food restaurants, if any.
  - Distance to Simon Bolivar Park
- The data will be taken from:
  - 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 Simon Bolivar will be obtained using Google Maps API geocoding or Foursquare API
  - coordinate of the polygons of the localities will be obtained using data of https://bogota-laburbano.opendatasoft.com/

```
]: address = 'parque simon bolivar Bogota'

   geolocator = Nominatim(user_agent="foursquare_agent")
   location = geolocator.geocode(address)
   latitude = location.latitude
   longitude = location.longitude
   print(latitude, longitude)
```

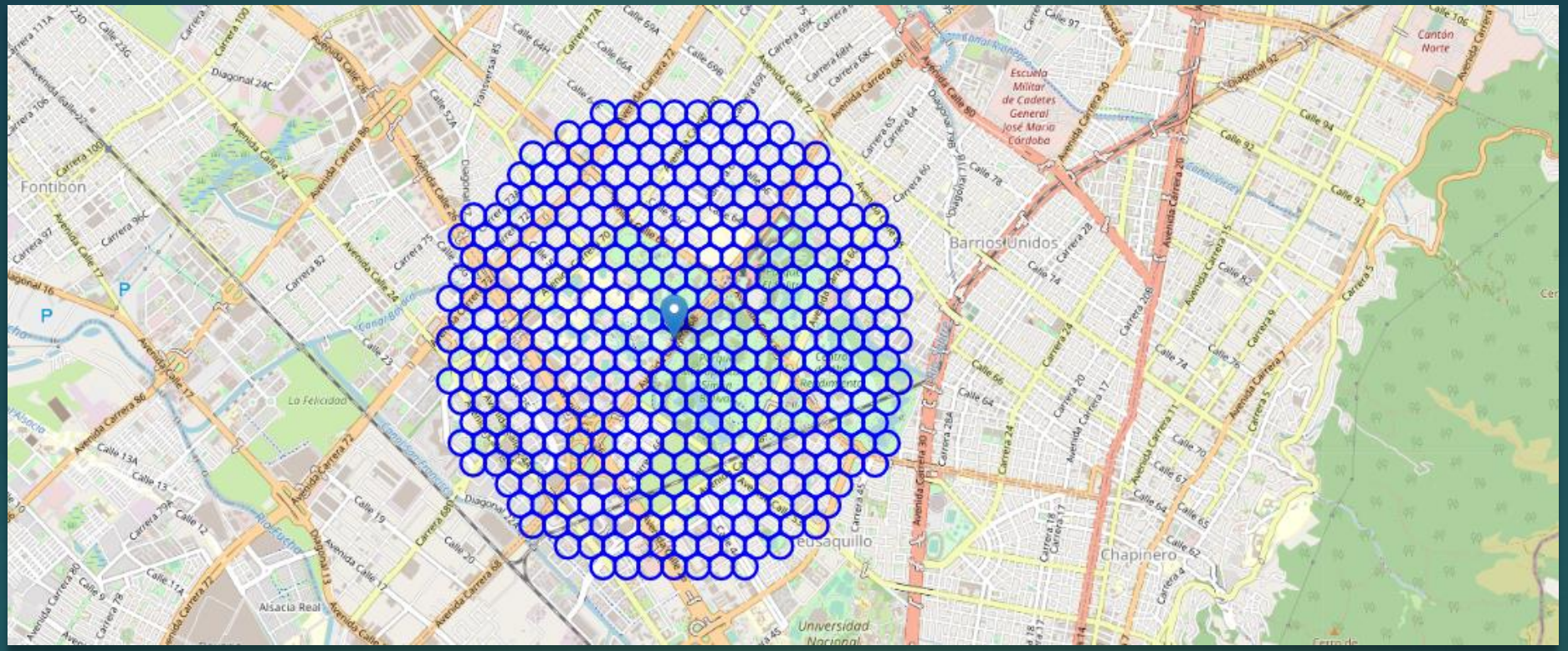
```
   4.6616218 -74.0973687
```

```
]: point_center=[location.latitude,location.longitude]
   point_center
```

```
t[4]: [4.6616218, -74.0973687]
```

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. 4x4 killometers centered around Parque Simon Bolivar Bogota.

Let's first find the latitude & longitude of Simon Bolivar Park



Now let's create a grid of area candidates, equaly spaced, centered around city center and within ~2km from Simon Bolivar Park. Our neighborhoods will be defined as circular areas with a radius of 100 meters, so our neighborhood centers will be 200 meters apart.

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 ~2km from Simon Bolivar Park.

```
df_locations = pd.DataFrame({'Address': addresses,
                             'Latitude': latitudes,
                             'Longitude': longitudes,
                             'X': xs,
                             'Y': ys,
                             'Distance from center': distances_from_center})

df_locations.head(10)
```

Out[12]:

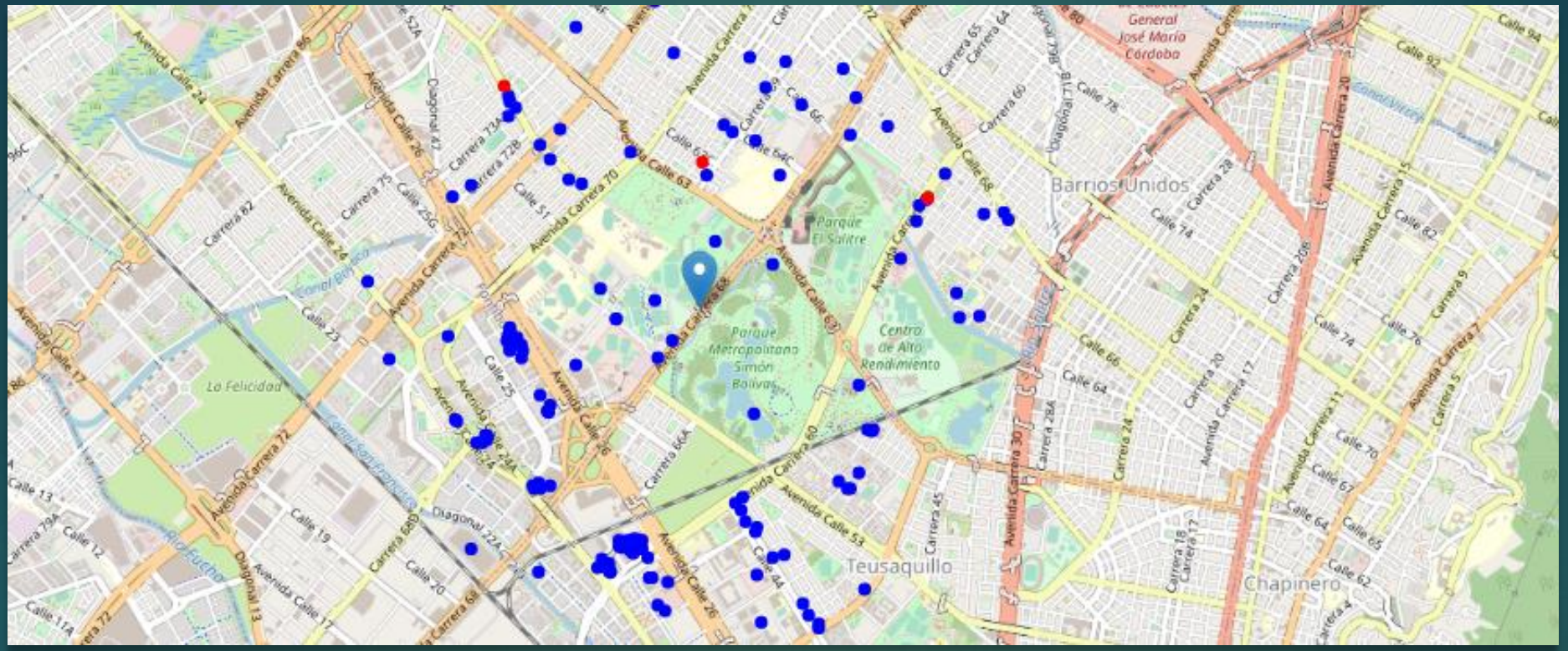
	Address	Latitude	Longitude	X	Y	Distance from center
0	Cl. 23a #59-72, Bogotá	4.644395	-74.102800	599514.340245	513419.512088	1997.498436
1	Cl. 24a #59-59, Bogotá	4.644392	-74.100997	599714.340245	513419.512088	1946.792233
2	Cl. 24a ##57, Bogotá	4.644390	-74.099194	599914.340245	513419.512088	1915.724406
3	Ac. 26 #54-94, Bogotá	4.644388	-74.097391	600114.340245	513419.512088	1905.255888
4	Cra. 54 #26-25, Bogotá	4.644385	-74.095588	600314.340245	513419.512088	1915.724406
5	Cra 45 #24b13, Bogotá	4.644383	-74.093785	600514.340245	513419.512088	1946.792233
6	Cl. 44 #53-54, Bogotá	4.644381	-74.091982	600714.340245	513419.512088	1997.498436
7	Carrera 7 #173-64, Cl. 22b #No. 66 – 46, Bogotá	4.645965	-74.105502	599214.340245	513592.717169	1951.922130
8	AV. Esperanza - AK 60, Bogotá	4.645962	-74.103699	599414.340245	513592.717169	1868.154169

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

In [18]:
```
print('Total number of restaurants:', len(restaurants))
print('Total number of chinese restaurants:', len(chinese_restaurants))
print('Percentage of chinese restaurants: {:.2f}%'.format(len(chinese_restaurants) / len(restaurants) * 100))
print('Average number of restaurants in neighborhood:', np.array([len(r) for r in location_restaurants]).mean())
```

```
Total number of restaurants: 152
Total number of chinese restaurants: 5
Percentage of chinese restaurants: 3.29%
Average number of restaurants in neighborhood: 2.2225274725274726
```

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

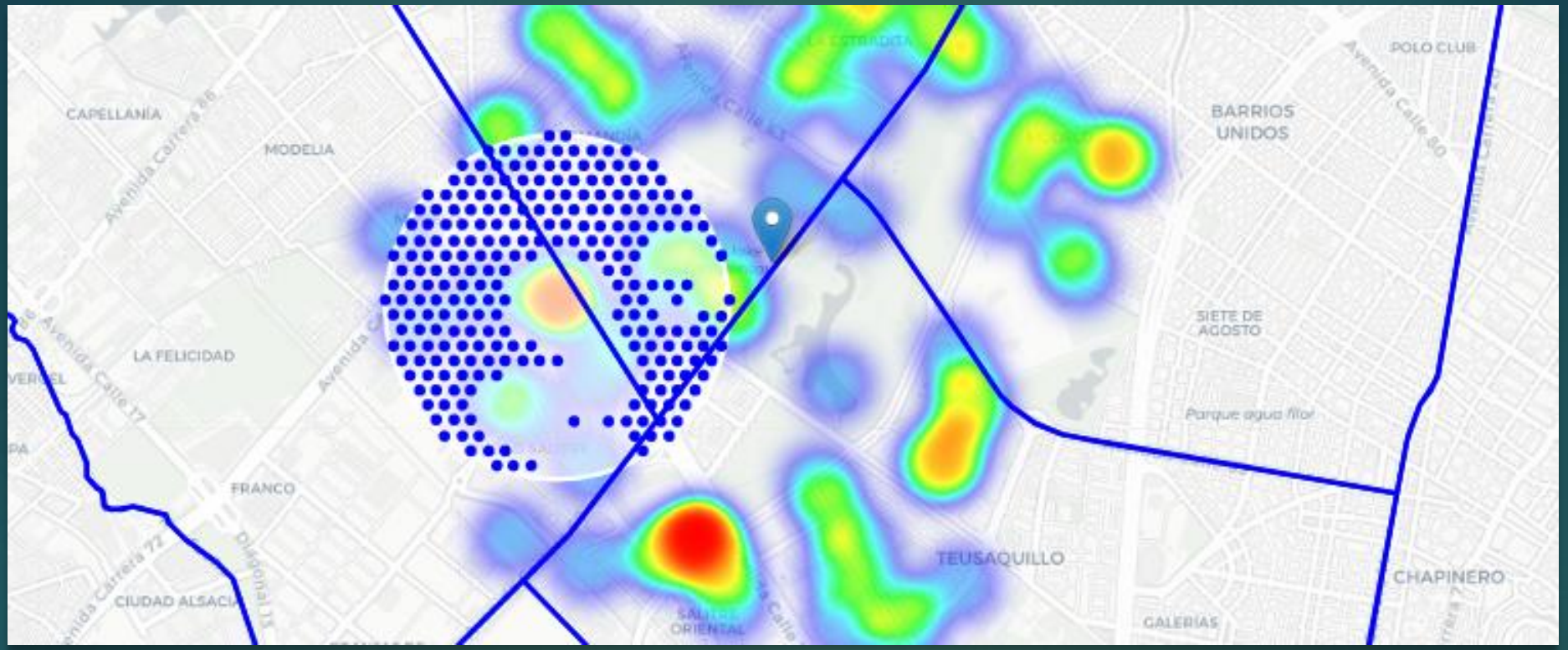


Let's now see all the collected restaurants in our area of interest on map, and let's also show Chinese restaurants in different color.

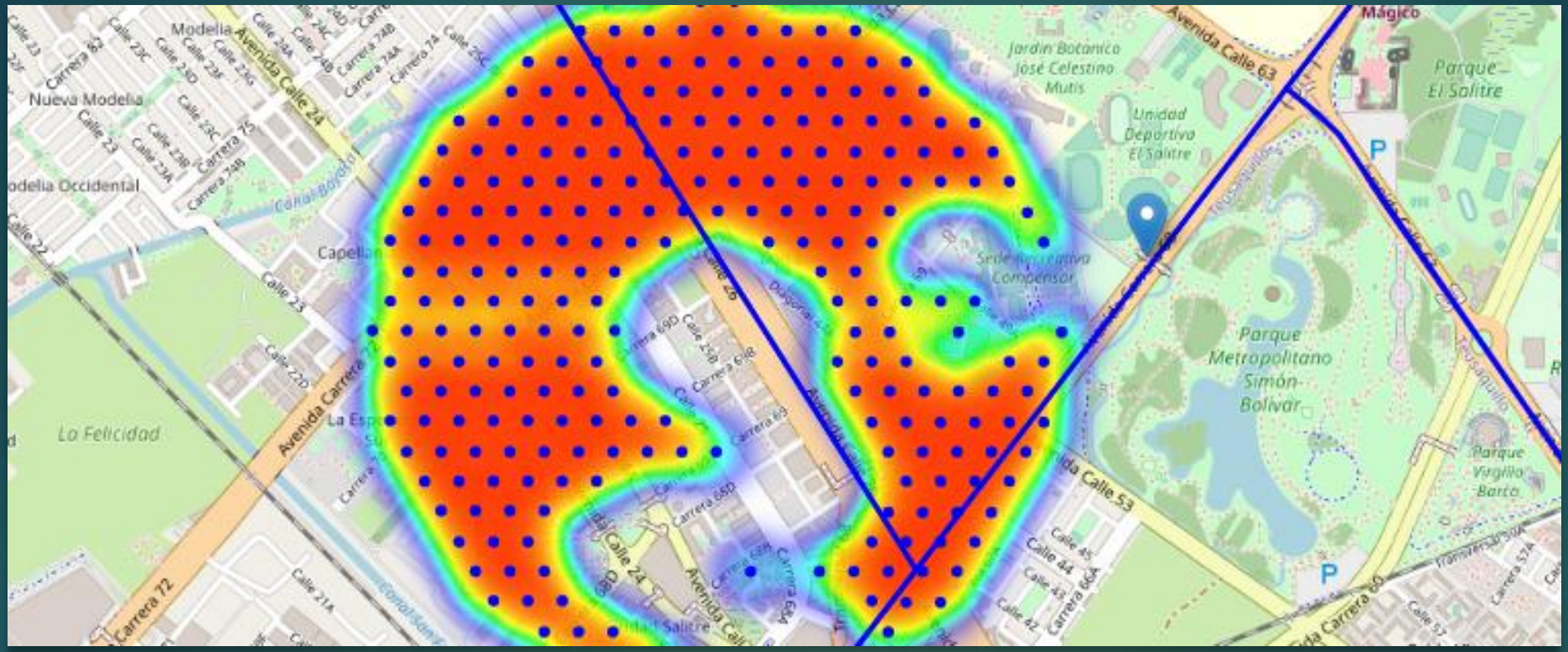
Looks no like a few pockets of low restaurant density closest to city center cant be found

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

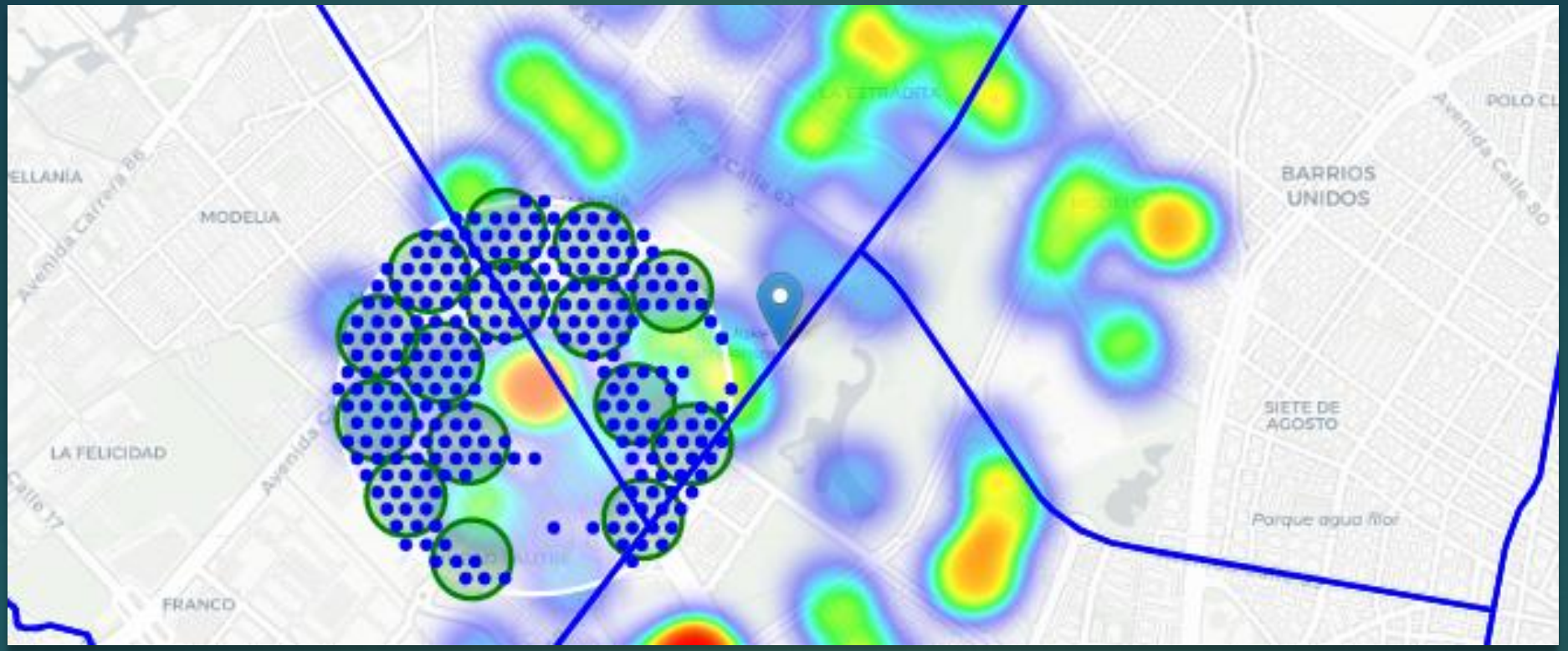
Let's define new, more narrow region of interest, which will include low-restaurant-count parts of Ciudad salitre.



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

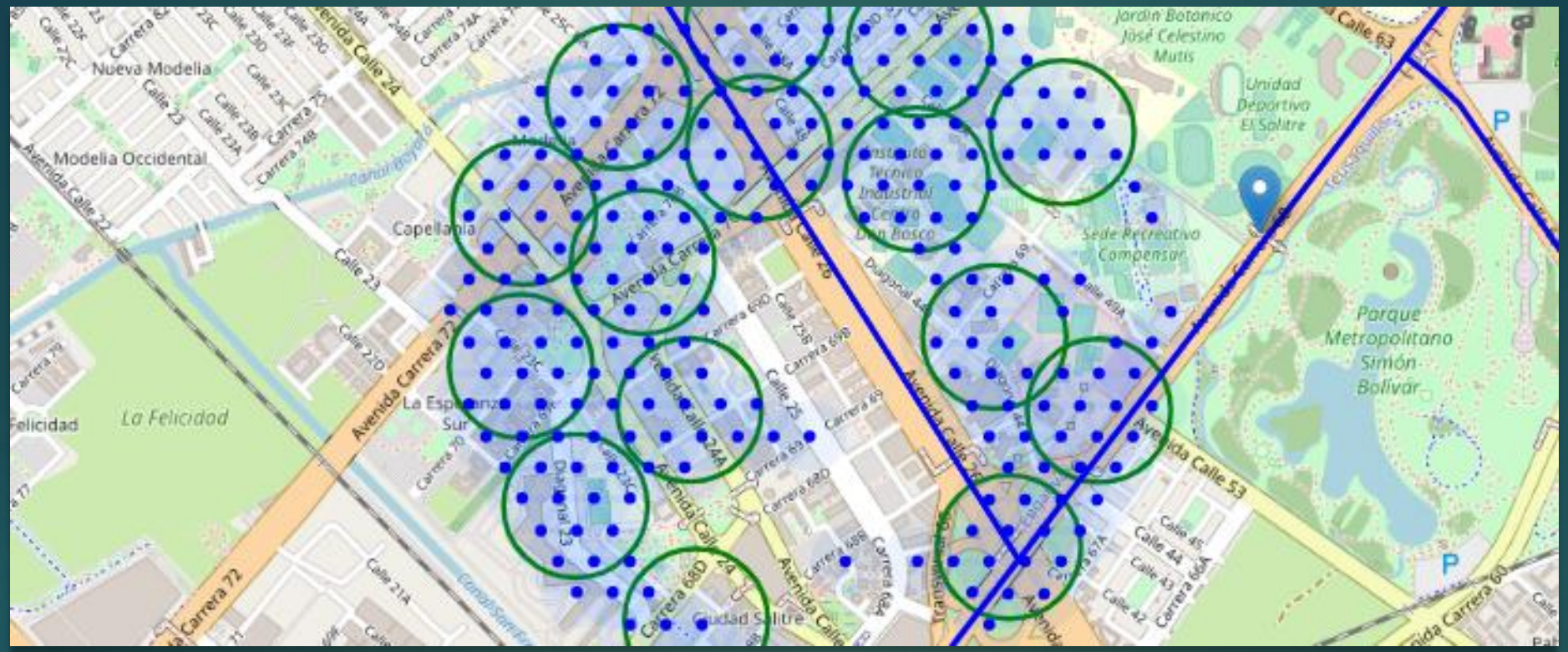


Let's now show those good locations in a form of heatmap:



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.

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

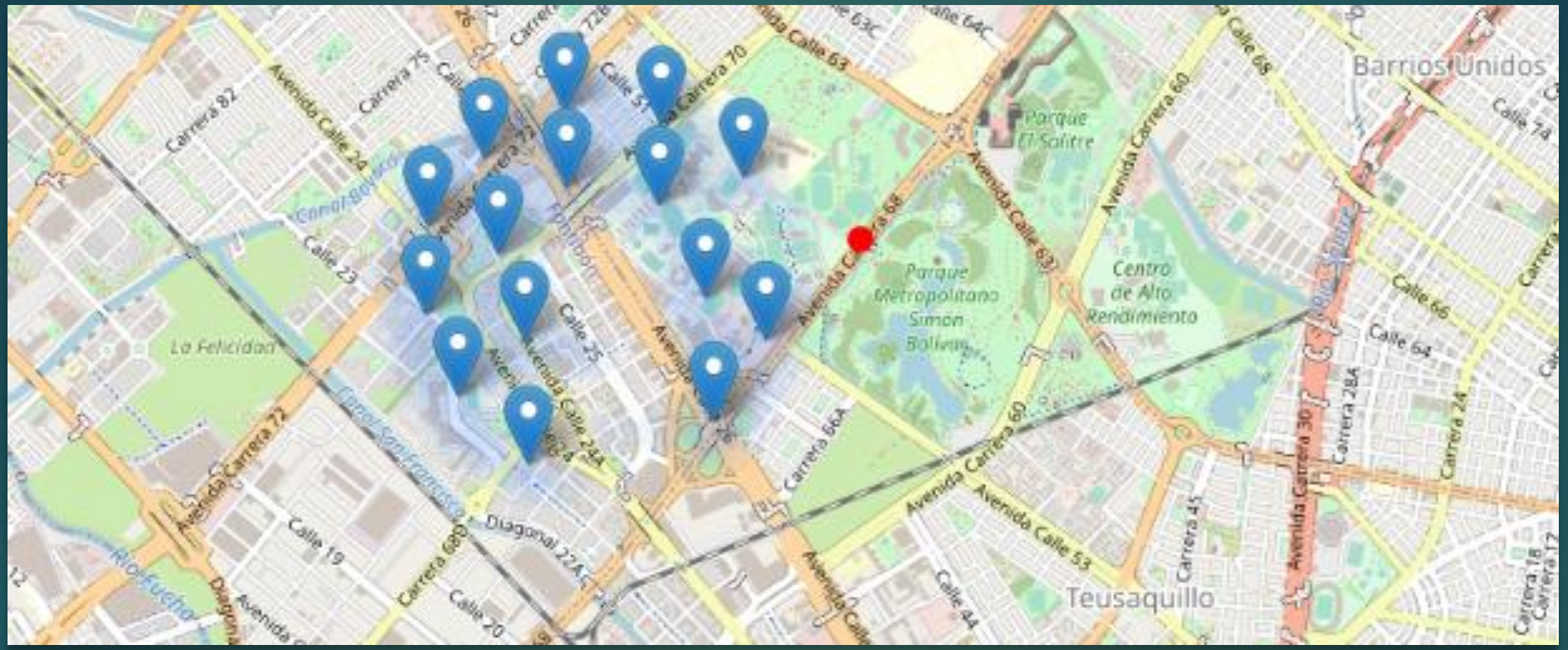


Let's zoom in on candidate areas in **Terminal Ciudad Salitre**:

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

Cra. 69 #5263, Bogotá, Colombia                    => 0.6km from Simon Bolibar Park
Cra. 70 #22d97, Bogotá, Colombia                   => 2.1km from Simon Bolibar Park
Ac. 26 #1001, Bogotá, Cundinamarca, Colombia       => 1.4km from Simon Bolibar Park
Ak 68 #26-22, Bogotá, Colombia                     => 1.1km from Simon Bolibar Park
Cra. 68c #23-31, Bogotá, Colombia                  => 1.9km from Simon Bolibar Park
Cra. 69b #24-51, Bogotá, Colombia                  => 1.7km from Simon Bolibar Park
Dg. 44 ##68b-80, Bogotá, Cundinamarca, Colombia    => 0.8km from Simon Bolibar Park
Ak. 70 #51-14, Bogotá, Cundinamarca, Colombia      => 1.1km from Simon Bolibar Park
007C06, Bogotá, Colombia                           => 2.0km from Simon Bolibar Park
Cra. 71d #4822, Bogotá, Colombia                   => 1.5km from Simon Bolibar Park
Terminal de Transportes de Bogota, Bogotá, Colombia => 2.0km from Simon Bolibar Park
Entrada Occidental a Centro Don Bosco #16, Bogotá, Colombia => 1.0km from Simon Bolibar Park
AK 68 - Cl 46, Bogotá, Colombia                    => 0.7km from Simon Bolibar Park
Ac. 24 #7025, Bogotá, Colombia                     => 1.7km from Simon Bolibar Park
Cl. 25B #72-20, Bogotá, Colombia                   => 1.8km from Simon Bolibar Park
```

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



This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of restaurants and no Chinese restaurants nearby, all zones being fairly close to Simon Bolivar Park (all less than 2km).

Conclusion

- The objective of this project was to identify areas of Bogota near Simon Bolivar Park with a low number of restaurants (particularly Chinese restaurants) to help stakeholders narrow the search for an optimal location for a new Chinese restaurant. When calculating the density distribution of restaurants from Foursquare data, we first identified general districts that warrant further analysis (Barrio Ciudad Salitre), and then generated a broad collection of locations that satisfy some basic requirements regarding nearby restaurants existing. These locations were then grouped together to create the main areas of interest (containing the largest number of potential locations) and the addresses of those area centers were created to be used as starting points for final exploration by the interested.