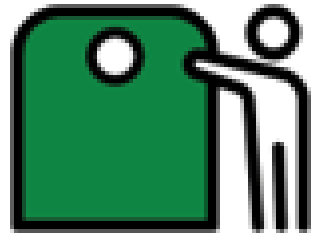


The Battle of the Neighborhoods



Glass Recycling Bins in Lisbon

Applied Data Science Capstone



Introduction

- Lisbon is the capital of Portugal
- This is a marvelous city located by the riverside, with breathtaking views, a rich culture, and full of lovely people
- However, some places in Lisbon have a severe waste management problem near the neighborhoods where the most famous bars are located



Introduction

The Problem

- Let's assume the Municipality of Lisbon wants to increase the number of Glass Recycling Bins (GRBs) in the city and has limited resources to invest
 - Which places have a shortage of GRBs?
 - Are high GRB densities associated with high bar/restaurant densities?
 - Which are the most critical places for GRB installation?



Let's use **Data Science** to answer these questions!



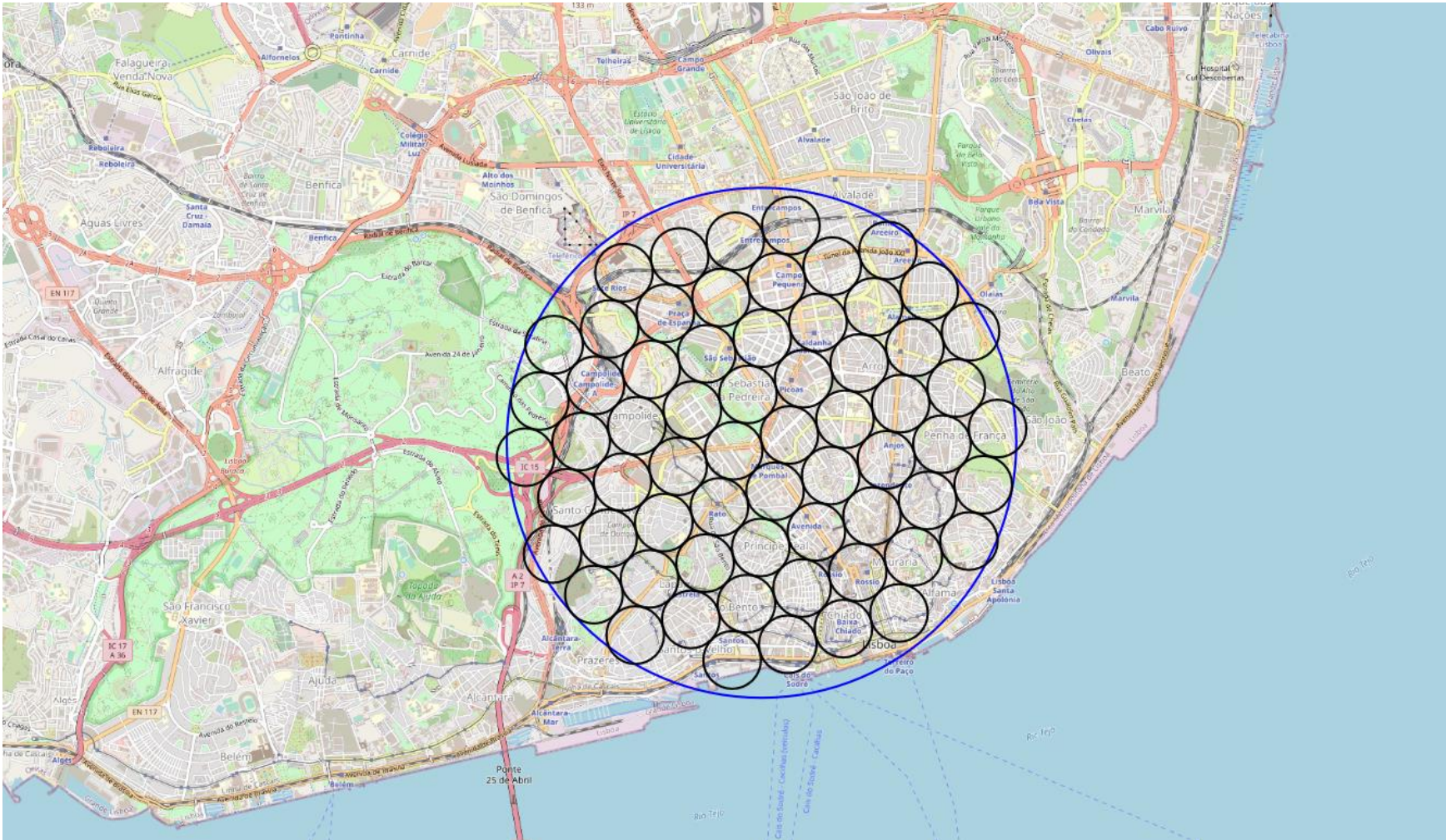
Data

- We define the assessment area as a circle of 2.5 km, centered around the Marquis of Pombal Square in Lisbon (a famous roundabout in the center of the city)
- Data for **GRBs** was retrieved from the official repository for geographical data of the Municipality of Lisbon:

http://geodados.cm-lisboa.pt/datasets/e4af86f9aabe44a1a036a2c677f2755b_5/data

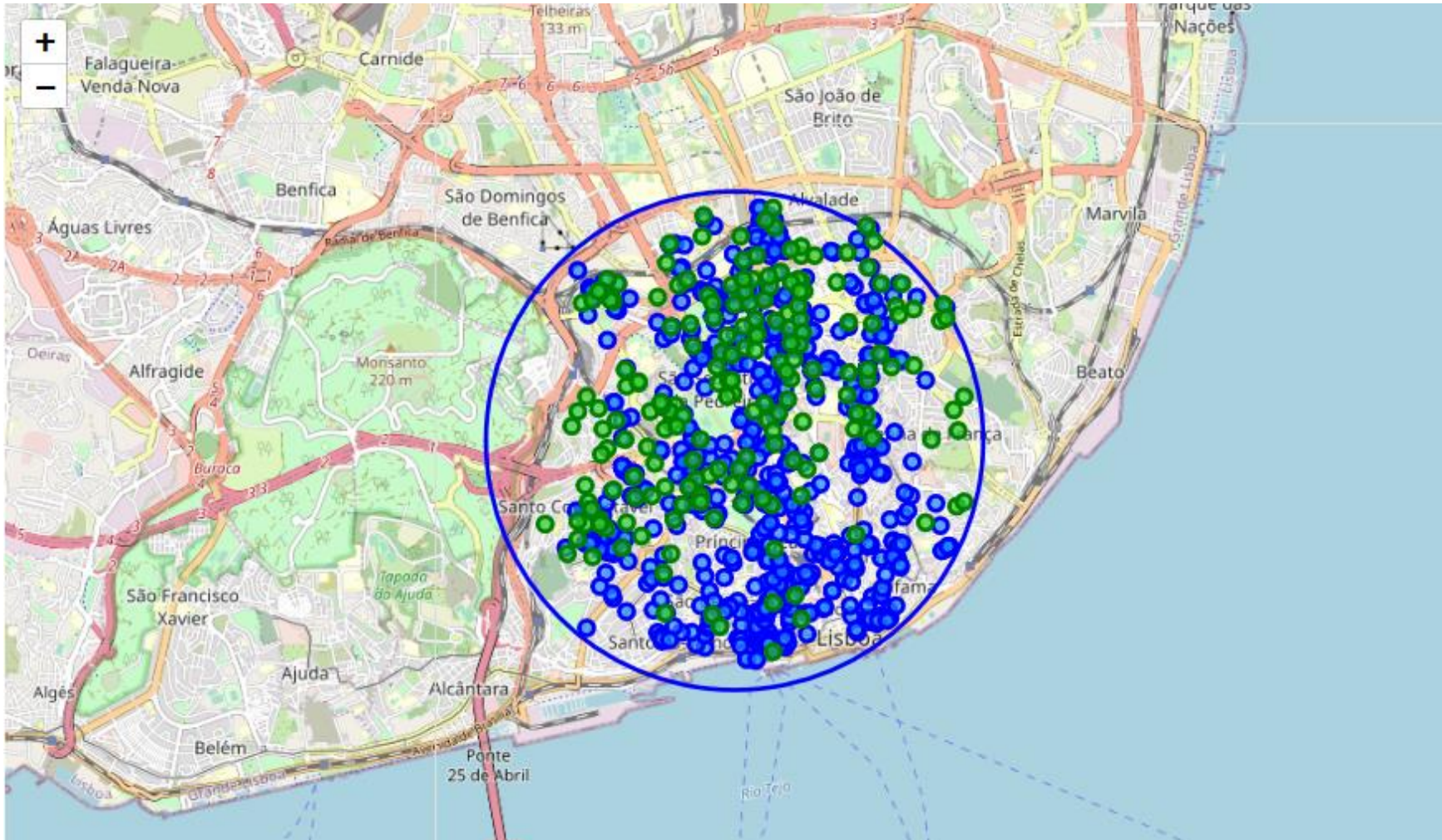
- **Venue** data was retrieved using the Foursquare API. However, due to the limited number of results for each call (100 venues) it was first necessary to divide the assessment area into a subset of smaller circles and make one call for each of those circles, resulting in a greater number of venue data points.
 - Only 'Bars', 'Pubs', 'Restaurants', and 'Cafés' were considered as venues, since they produce the most glass residues

Data



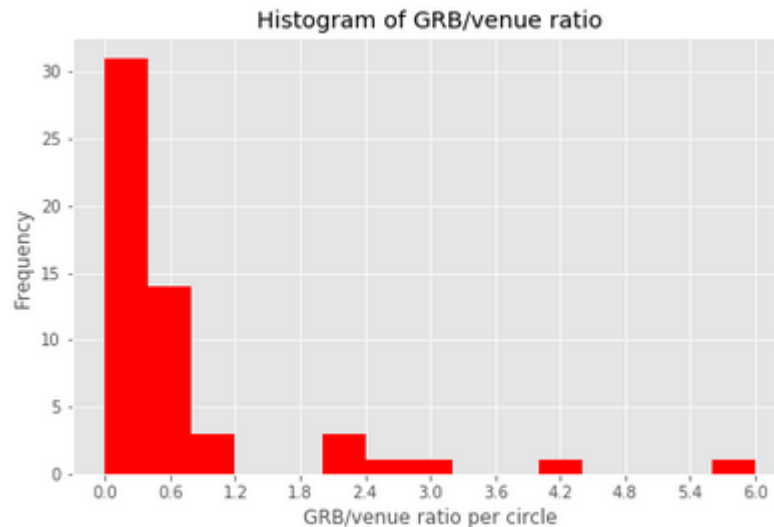
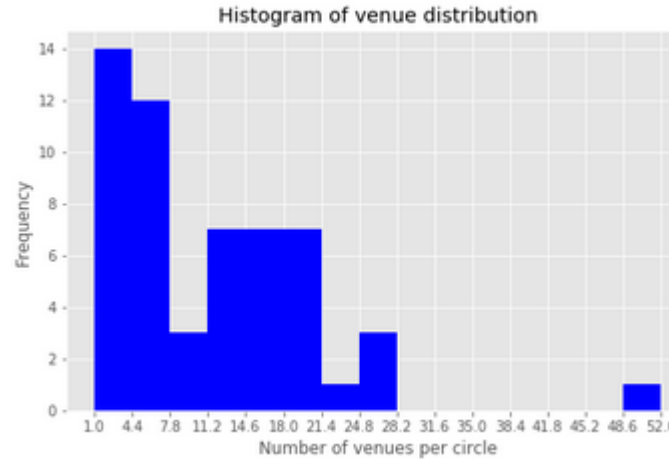
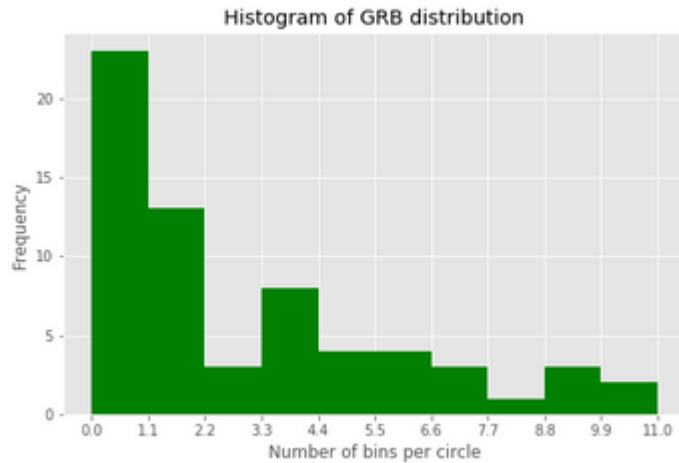
Representation of the subset of small circles, which center coordinates were used to make the API calls

Data



Geospatial data of **GRBs** (green dots) and **venues** (blue dots) located inside our study area in Lisbon (blue circumference)

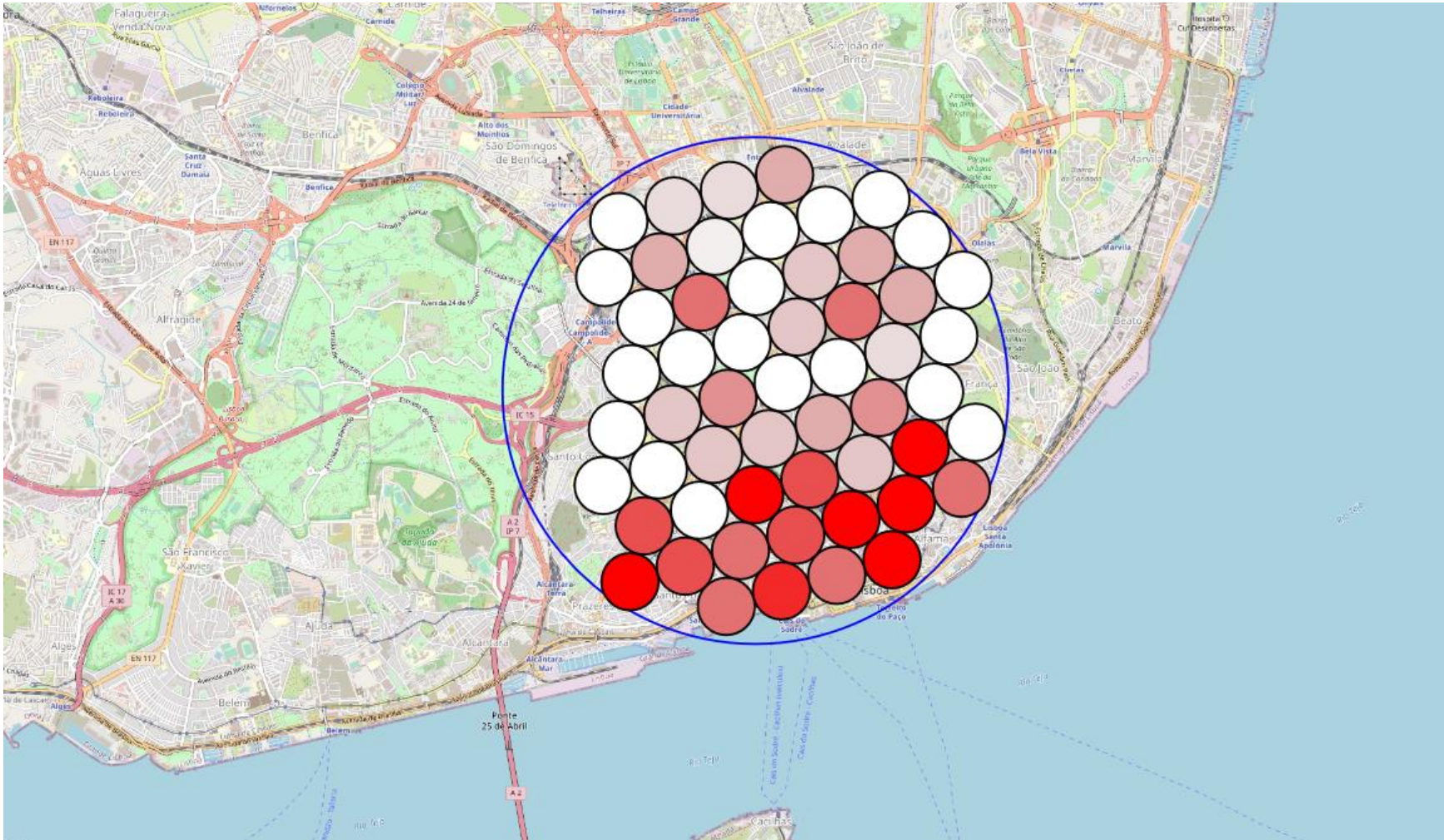
Analysis and Results



I assessed the number of **GRBs** and **venues** inside each circle, as well the number of GRBs per venue (or GRB/venues **ratio**)

The distribution of these values is plotted in the histograms to the left, and served as an exploratory data analysis step

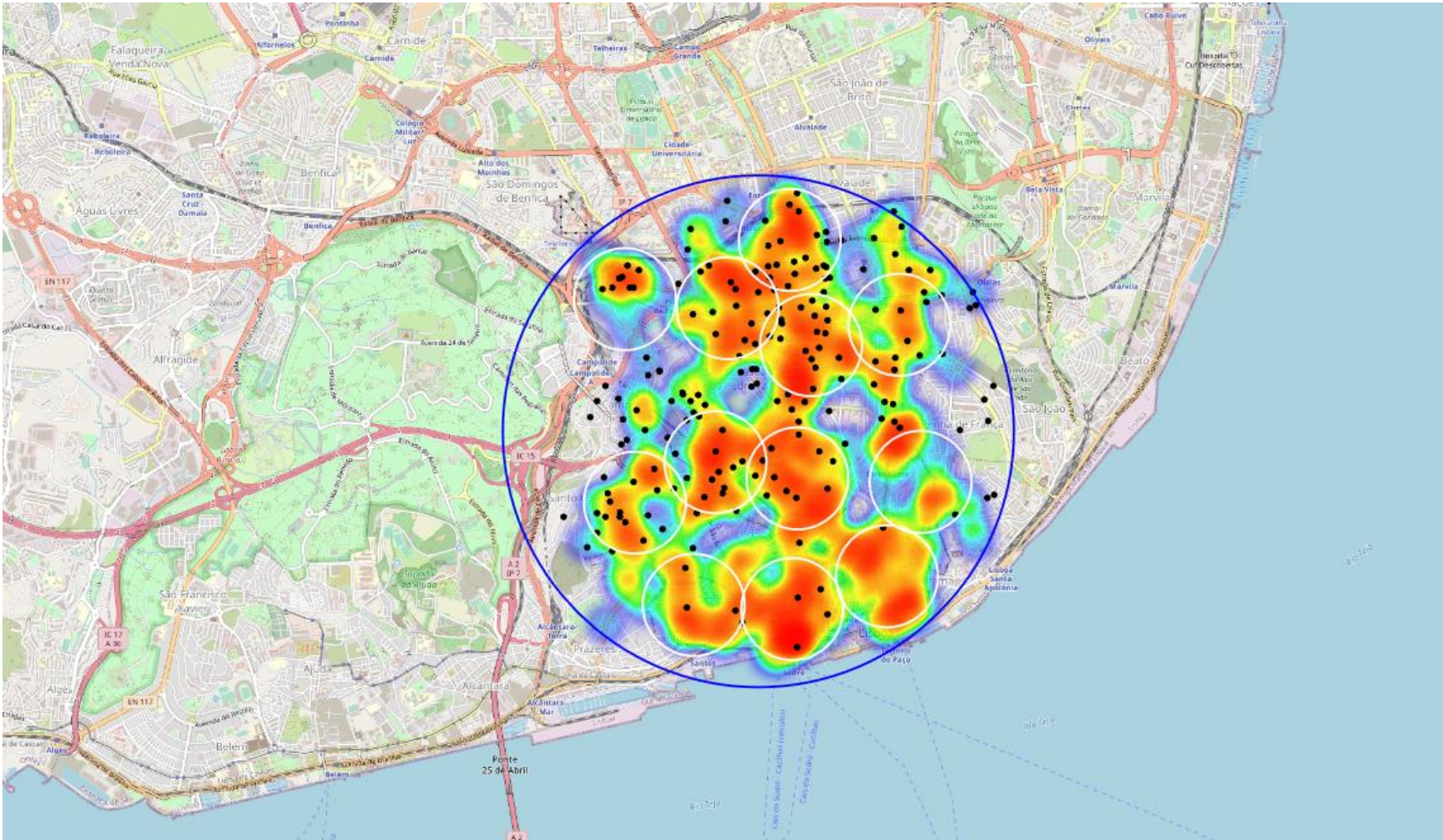
Analysis and Results



Since the ratio is the variable of greatest interest, I plotted this information on the map, using color to convey information about the GRB/venue ratios inside each circle:

A **Red** circle indicates a low ratio, while a **white** indicates a high ratio

Analysis and Results



To have a better idea of the real scenario, I plotted the GRBs dots on top of a Heatmap generated from the venue data

I then used k-means clustering to generate 12 clusters of venues and limited them to circles of 500 m radius in the map

Analysis and Results

cluster	latitude	longitude	grb	venue	ratio
0	38.712706	-9.135395	1	59	0.017
9	38.721005	-9.131586	1	25	0.040
2	38.709823	-9.146072	4	88	0.045
10	38.710188	-9.157193	3	24	0.125
6	38.721312	-9.145484	11	52	0.212

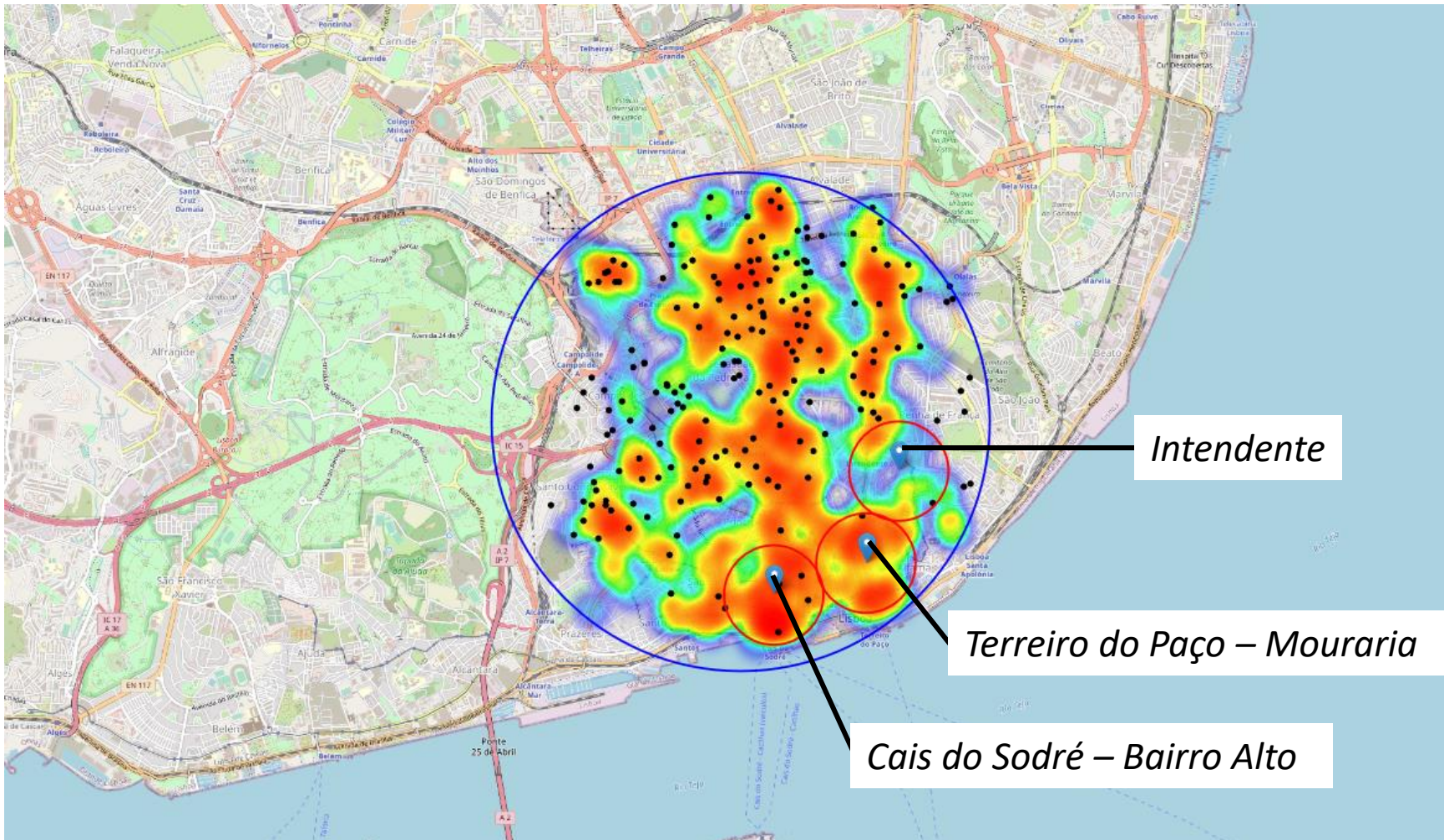
clusters in
need of
new GRB
installation

ratio threshold = 0.1

Next, I calculated the GRB and venue counts for each cluster, and stored this information in a dataframe

Since we are interested in identifying low GRB/venue ratio clusters, I sorted the dataframe by an ascending order of ratio and defined a threshold 0.1 (1 GRB to 10 venues)

Analysis and Results



And finally I plotted only these clusters in the map, delimited by red circles, and identified them

These are all areas with lots of venues, either bars or restaurants, where the amount of GRBs available is not sufficient

Conclusion

- The analysis identified 3 areas in urgent need of installation of more GRBs
- All these areas have a high number of venues associated with a low number of GRBs
- For a future and more detailed analysis, it would be relevant to have access to other types of data, such as the type and amount of residues collected in each location. However, this type of data is usually private and is not provided for free

