

The Battle of the Neighborhoods

Applied Data Science Capstone

Recycling Glass Bins in Lisbon

João Romão

IBM Data Science Professional Certificate by Coursera

1. Introduction

1.2 Background

If you are ever in Portugal, just finished a great meal and are now drinking an espresso by the riverside, enjoying the beautiful view, all the while you are listening to a beautiful *Fado* playing from one of the open windows of a house nearby, chances are you are in Lisbon. The capital of Portugal, from where Vasco da Gama set sail to discover the sea route to India in 1497, is a city full of culture, beautiful sites, and nice people. However, not everything in Lisbon is so wonderful. Filled with restaurants, pubs, and nightclubs, the city streets close to the venues are quick to fill with trash, especially on the weekends. This issue is particularly concerning for those who experience the nightlife, where a broken bottle on the ground could spell disaster for anyone who had a few drinks too much that evening. Unfortunately, waste management in certain areas of Lisbon is still an issue, and even though the Municipality has implemented some measures to reduce the piling up of broken bottles near full bins, not much has been solved. Popular places like *Bairro Alto*, *Cais do Sodré*, *Terreiro do Paço*, are as filled with trash by Saturday morning as their bars were full of customers the night before. If one wants to see where the best pubs are, just follow the trail of broken bottles.

1.2 The Problem

In the present report, I imagined the Lisbon Municipality wanted to implement a better waste management plan, particularly concerning the disposal of glass residues. For that, the stakeholders need to understand which areas have a high density of venues related to the production of glass residues (Bars, Restaurants, etc.), and which are lacking glass recycling bins (GRBs). Although any local resident can pinpoint a lot of candidate places for intervention, let us imagine for the sake of this study that the Municipality only has the budget for a limited number of GRBs. In that case, they need to understand which areas are most severely affected by this problem and invest their efforts and resources accordingly. Therefore, on this report I will present the analysis I used to identify the areas in need of more urgent action concerning the installation of GRBs. This will include the data analysis and leverage of the Foursquare Places API and the knowledge obtained in the various courses of the IBM Data Science Professional Certificate.

2. Data

Before starting to address the problem, we need to collect and prepare our data. The required data is pertaining to both GRBs and relevant venues in Lisbon, particularly geospatial information about individual datapoints. Nonetheless, and since the number of calls with the Foursquare API are limited, I restricted the assessment area to a 2.5 km radius circle centered around the Marquis of Pombal Square, a very famous roundabout in Lisbon which is situated roughly at the geographical center of Lisbon, although it is not the official city center.

GRB data was obtained from the website http://geodados.cm-lisboa.pt/datasets/e4af86f9aabe44a1a036a2c677f2755b_5/data, provided and curated by the Municipality of Lisbon. Although the website displays a table that can be scraped from the html code, I used the .csv file available and uploaded it directly to the Notebook. I selected this approach since it provided two additional columns with Latitude and Longitude data for each GRB, that were absent from the html table. Once I imported the table, I converted it to a dataframe and discarded all information except for the Latitude and Longitude values. Furthermore, I also selected the GRBs located inside our study area and discarded those located outside.

I used the Foursquare Places API to collect venue data. However, since each API call only returns 100 venues, I had to use a different strategy to collect enough information and adequately populate the assessment area with venue data. For that, I first divided the 2.5 km radius circle in a subset of smaller circles with an approximate area of 0.25 km², all packed together in an hexagonal fashion. After that, I made individual API calls using the coordinates of the center of each circle and restricted the call to a radius of 284 m, which was the same as the radius of the small circles, and stored the results in a database. From there, I needed to restrict the venue data, excluding irrelevant datapoints from the evaluation. For that, I filtered the venue categories and kept those which had the keywords 'Bar', 'Pub', 'Restaurant', or 'Café', storing the relevant data in a new dataframe.

By the end of the data wrangling process, I was left with GRB and venue data located inside our study area. You can have a better sense of the geospatial location of the data by inspecting Figure 1.

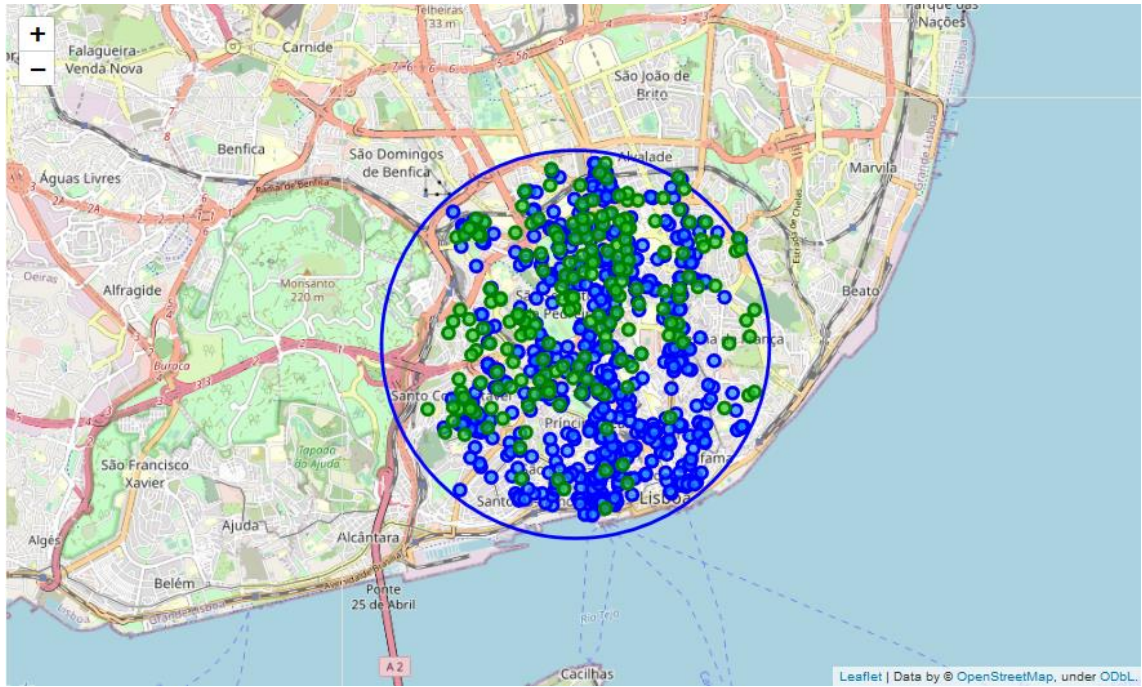


Figure 1. Geospatial data of GRBs (green dots) and venues (blue dots) located inside our study area in Lisbon (blue circumference)