

Coursera IBM Capstone Project

A wide-angle photograph of a suspension bridge, likely the Bay Bridge in San Francisco, during twilight. The sky is a deep blue, transitioning to orange and yellow near the horizon. The bridge's towers and cables are silhouetted against the bright city lights of the San Francisco skyline in the background. The water in the foreground is dark and reflects the bridge's structure.

The Battle of Neighborhoods



Purpose

This article is part of my final peer reviewed assignment for the IBM Data Science Professional Certificate program —[Applied Data Science Capstone](#).

This is an overview of the project, the project can be found [here](#).

Introduction



Photo by [Alex Paganelli](#) on [Unsplash](#)

Lisbon is the capital and main city of Portugal. It is a vibrant city with many cultural and touristic attractions and with a significant number of college campus spread across the city. In the last few years there was a boom on the number of foreign exchange students and in the same period there was increasing of foreign tourists visiting the city.

In consequence, we see now more and more restaurants, cafés and bars opening in the main neighborhoods of the city. The demand for more sophisticated places than can provide an improved experience and attract these new customers.

Business Problem

A group of investors is looking to start a new Café concept of addressed to a young and international audience. They also want to place their first location in upper scale neighborhoods, either in the downtown area or in the neighborhoods that have gone through a renovation in recent years.

Another requirement is to place these cafés in areas where already exist other similar venues so they can attract those customer to their new concept of Café. We provide a recommendation based on these criteria.

Photo by [Jonas Jacobsson](#) on [Unsplash](#)

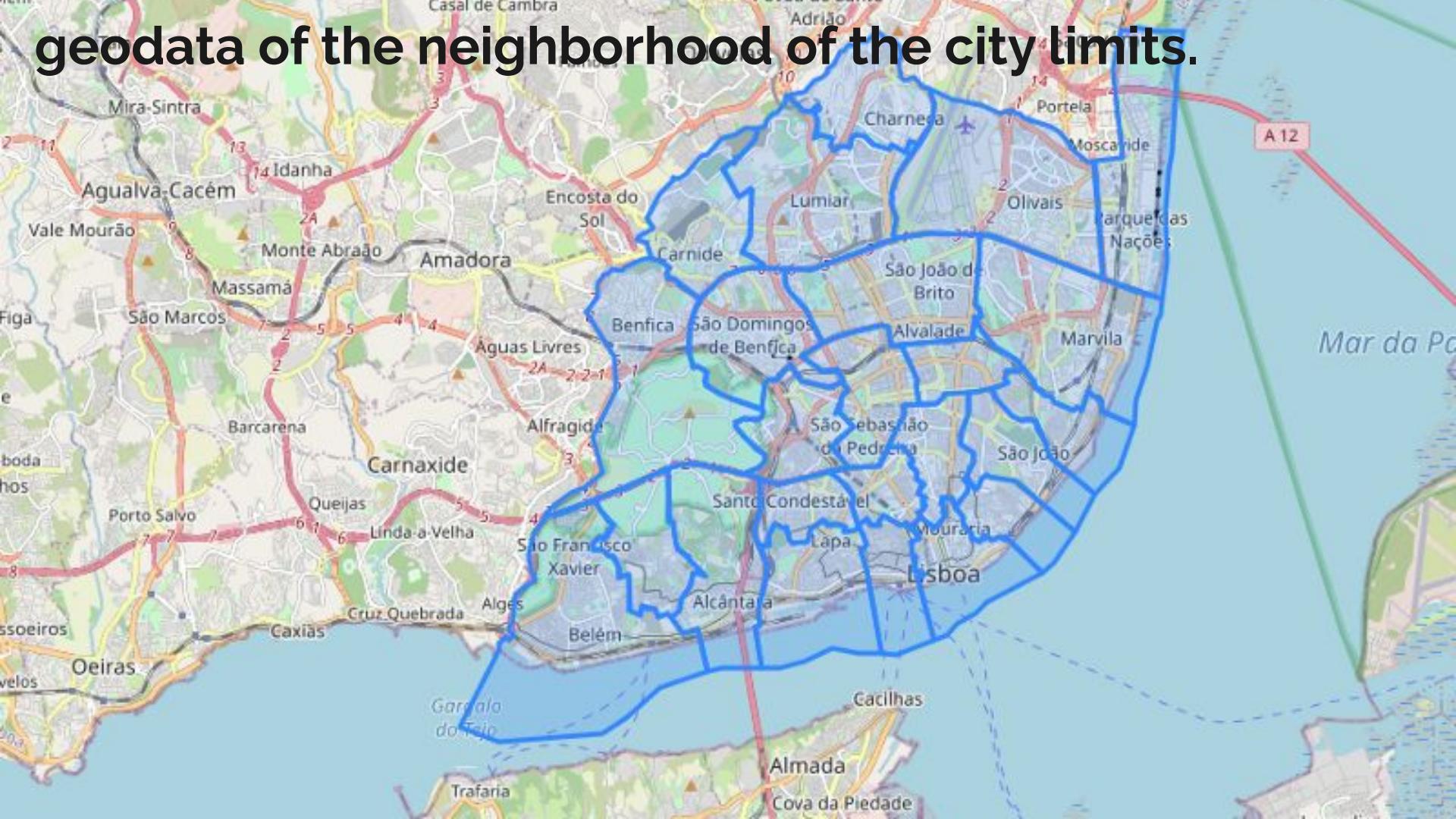




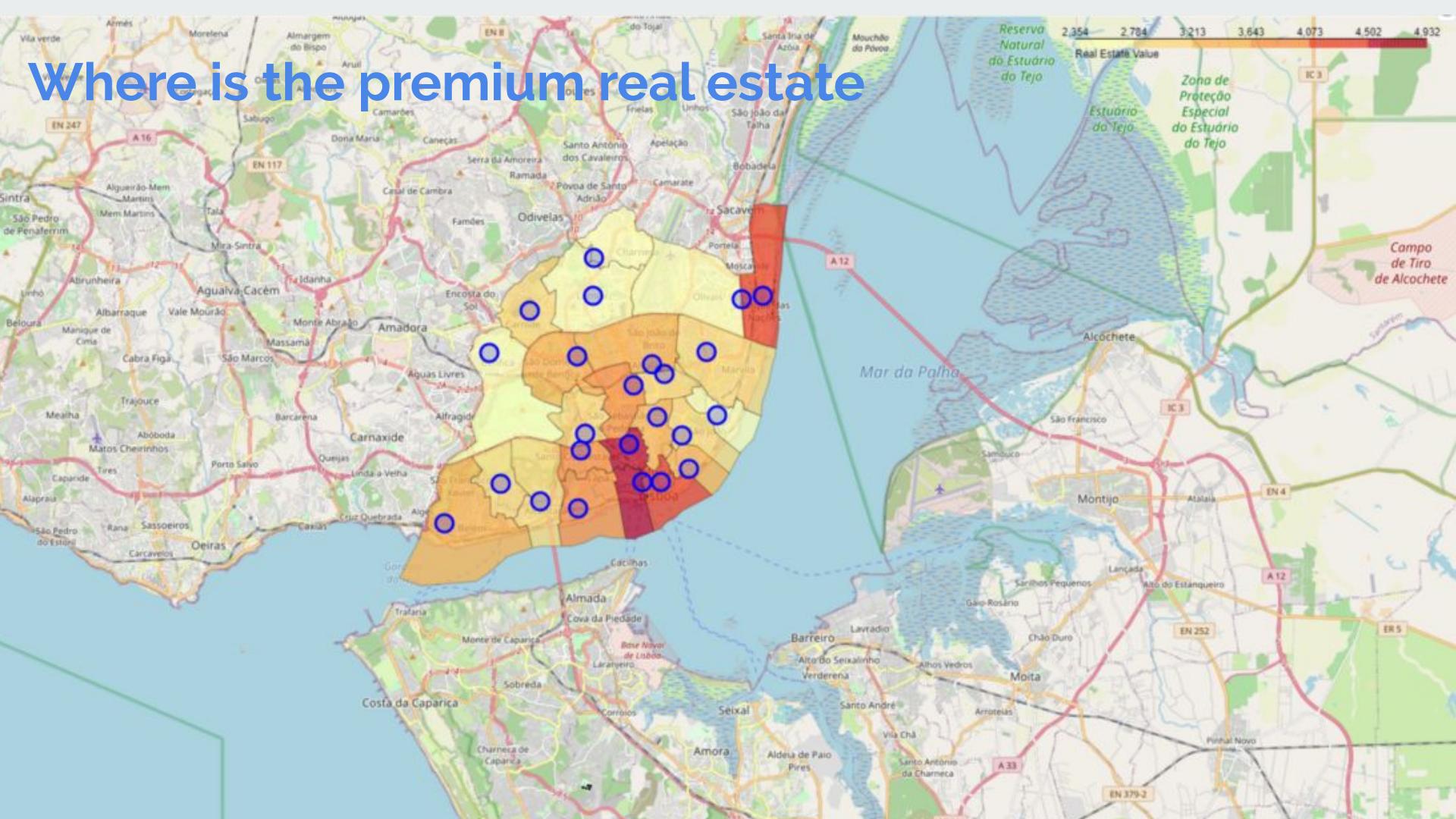
Data

- Foursquare venue data available through the Foursquare API.
 - Demographic data for the city of Lisbon from the Portuguese national statistics bureau.
 - Real estate data from the Portuguese national statistics bureau.
 - Spatial data for the Lisbon neighborhoods from the Lisbon city council.
 - Lisbon neighborhoods historic data from the Wikipedia.

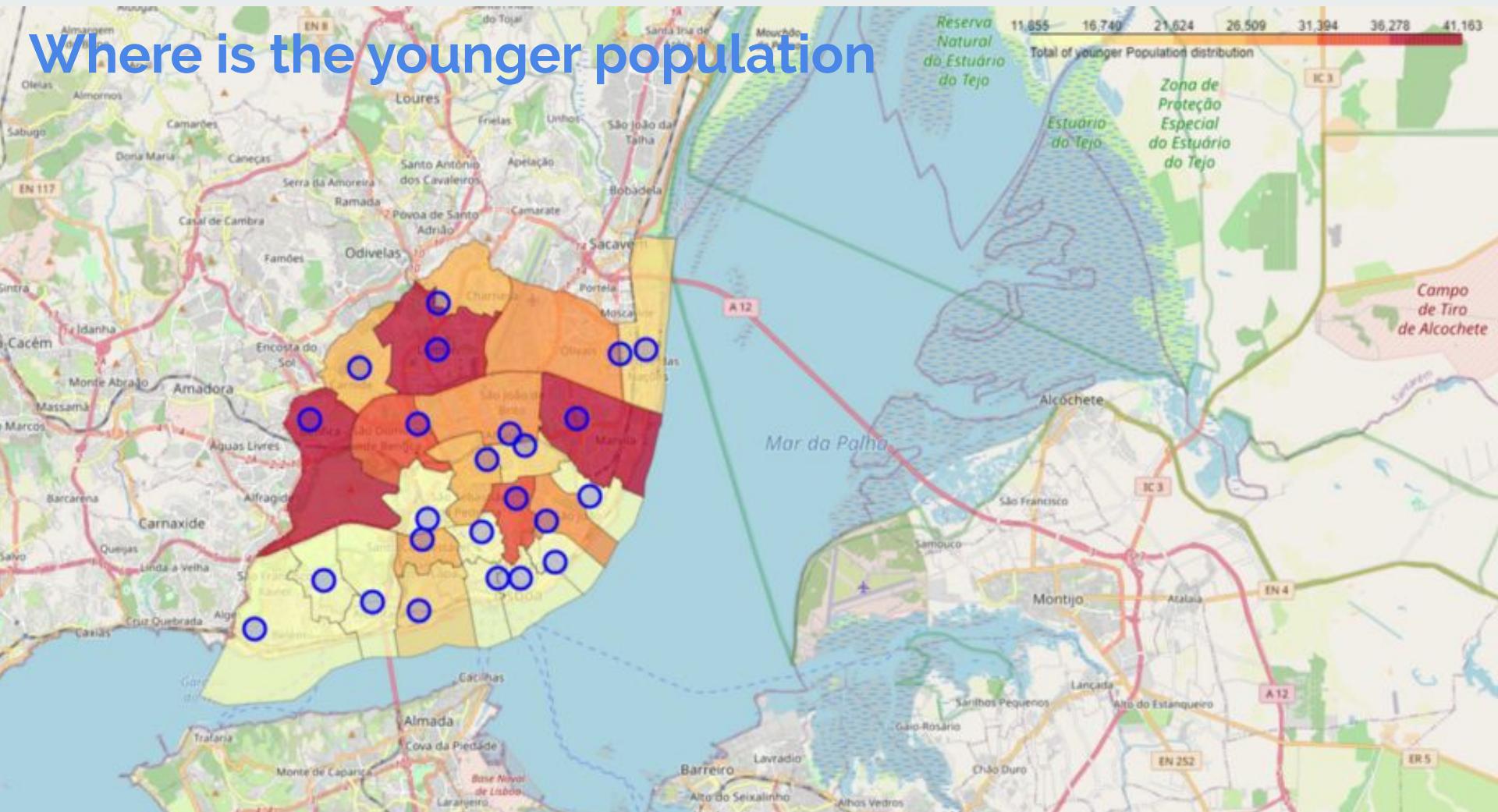
geodata of the neighborhood of the city limits.



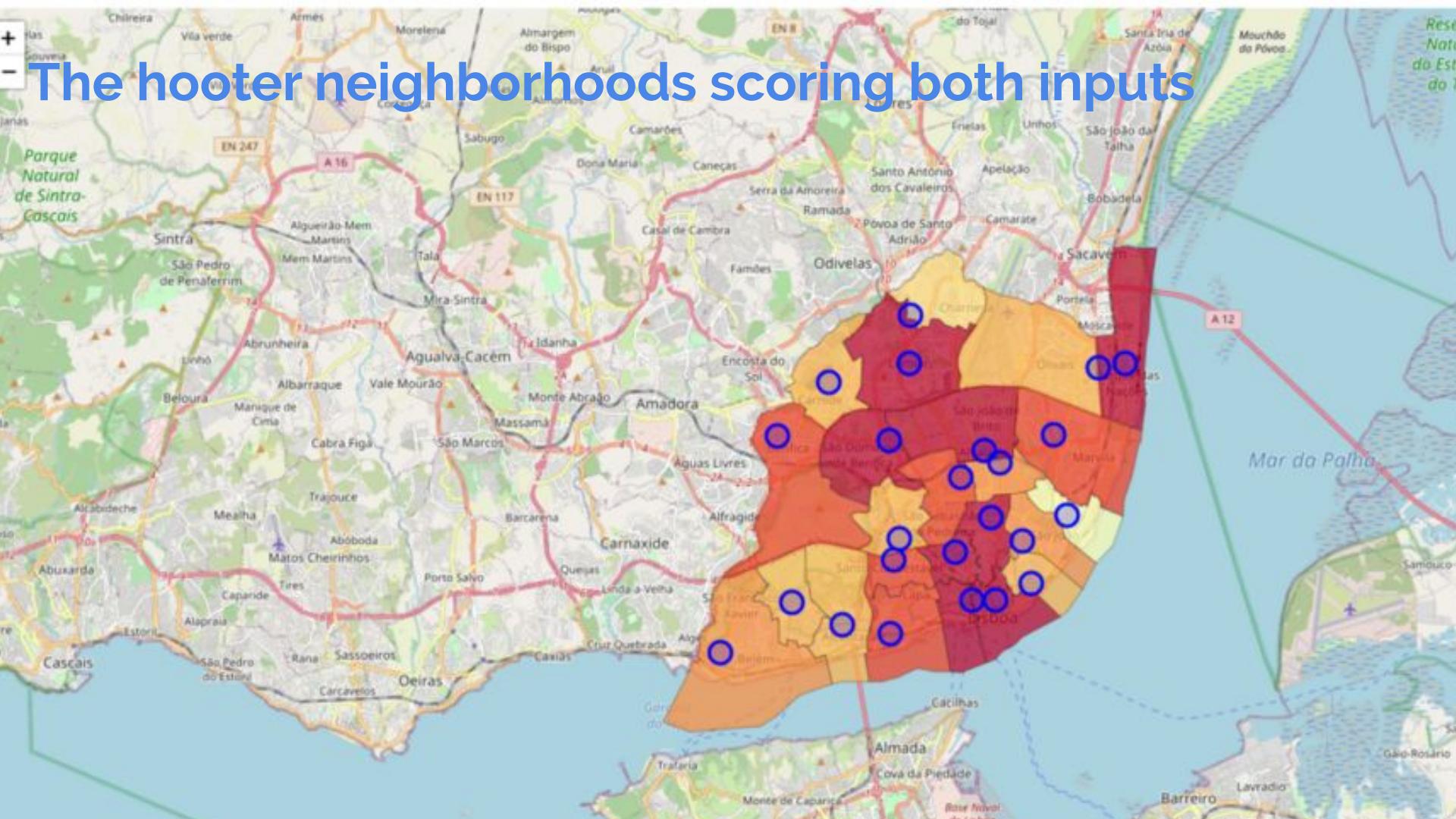
Where is the premium real estate



Where is the younger population



The hooter neighborhoods scoring both inputs



Modelling

Next we will see some violins.

I mean, the venue distribution by neighborhood based of the three categories we are most interested:

- Restaurants
- Cafés
- Bars

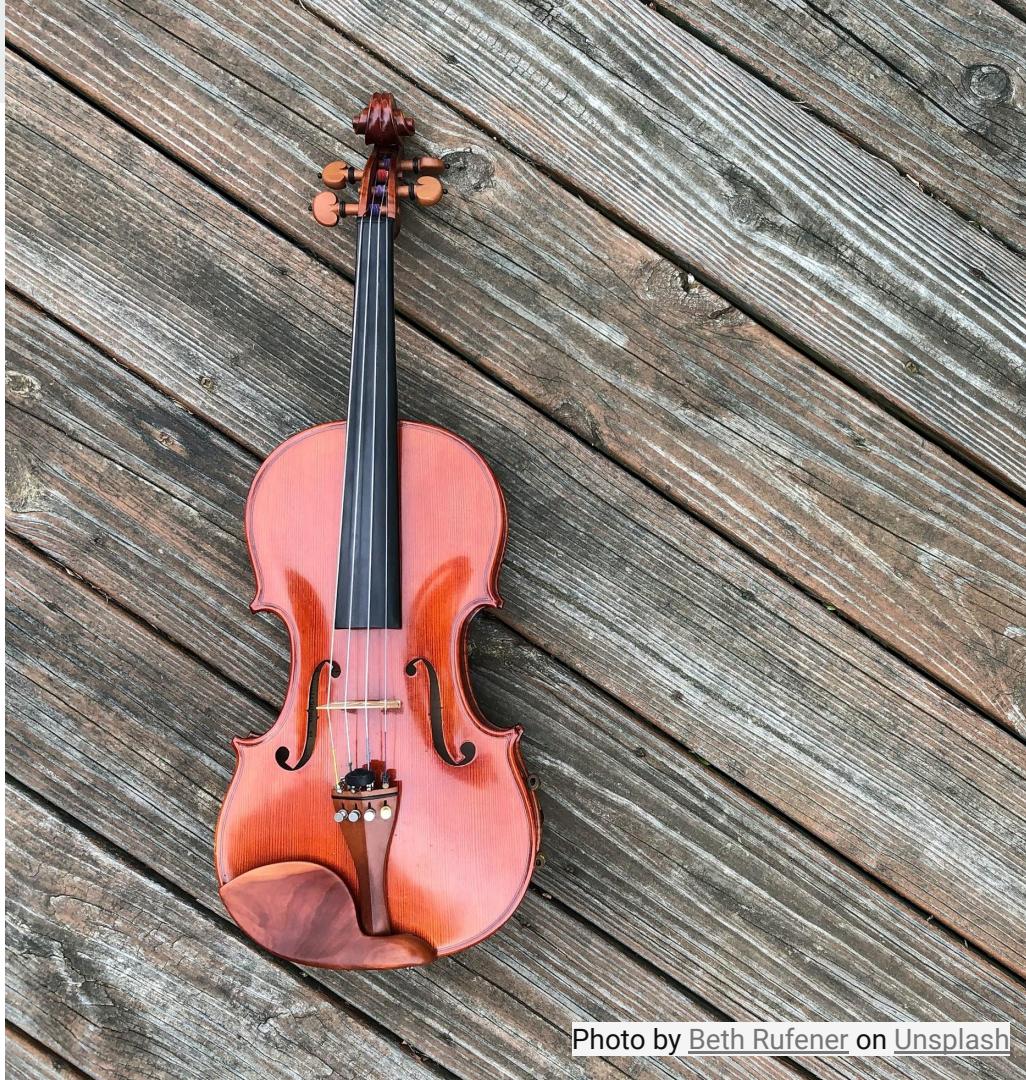
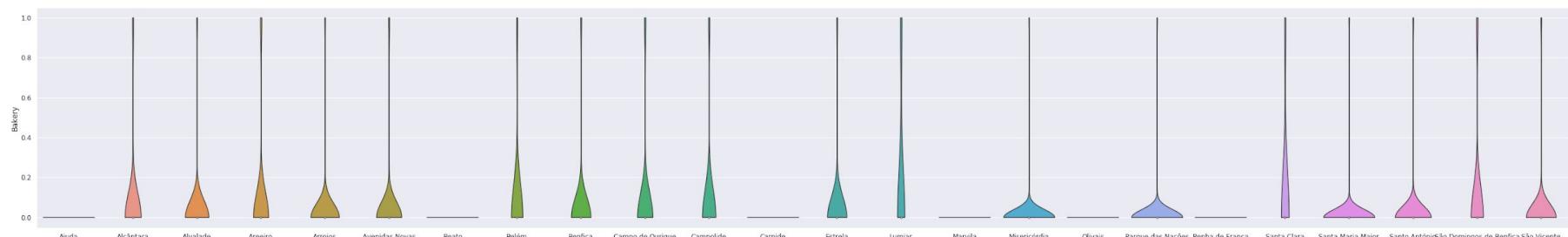
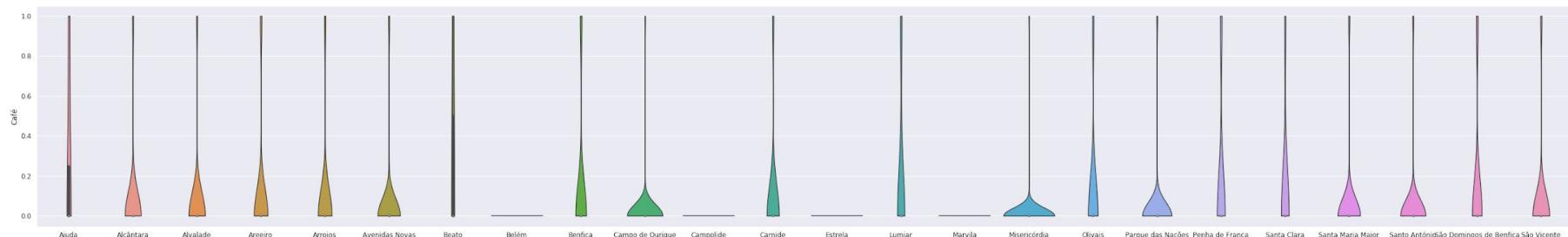
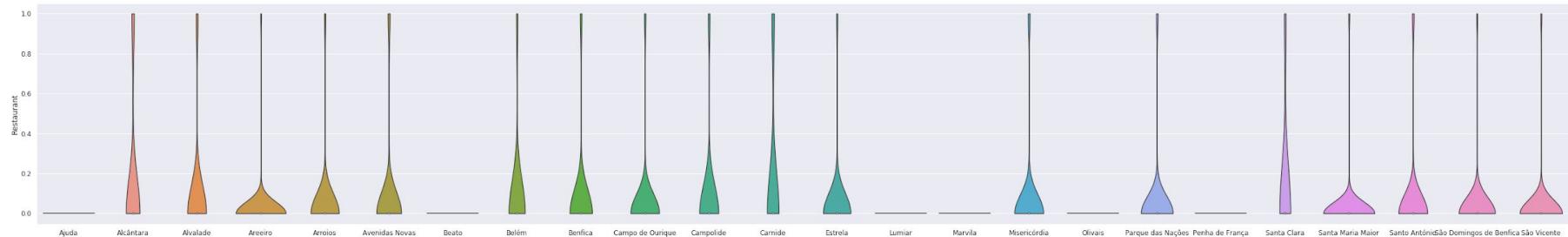


Photo by Beth Rufener on Unsplash

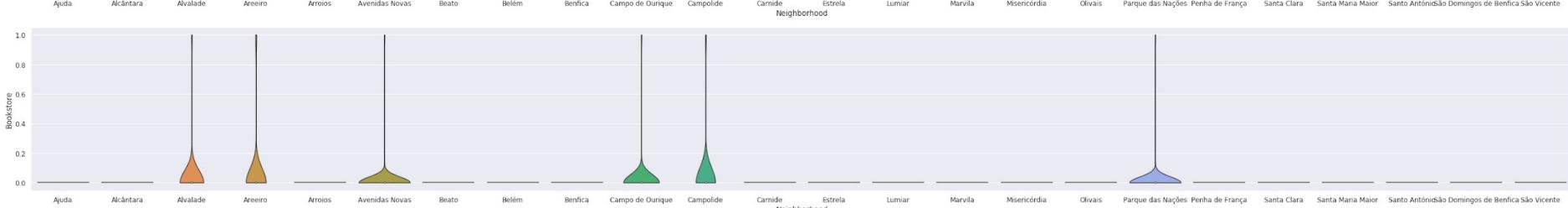
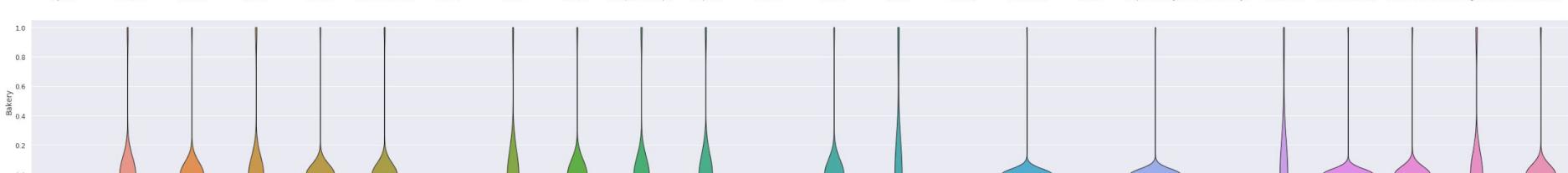
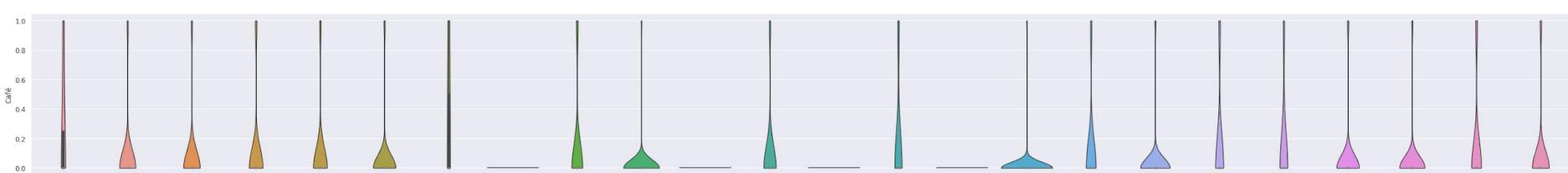
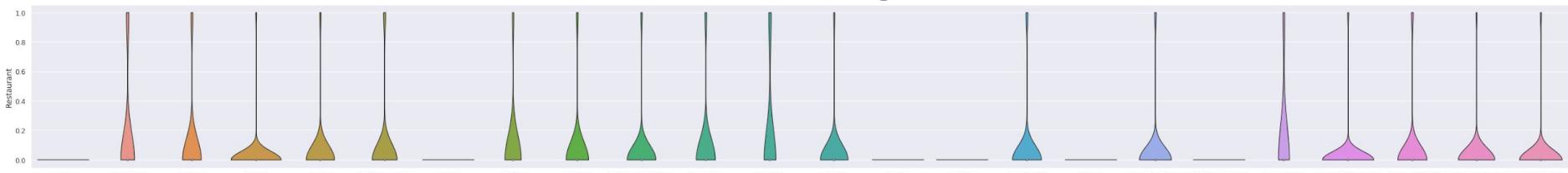
Frequency distribution for the 3 target venue categories by neighborhood

Too many neighborhood to pick from



Frequency distribution for the 3 target venue categories by neighborhood (plus bookstores)

Another of our preferred “neighbors”: bookstores



Final results

When we put together the results from both analysis we reach a short list of candidates

Where is the treasure? Our target neighborhoods

