Data Acquisition - Districts

To get the data about Rio de Janeiro's districts, I looked for the prefecture website, with no success. I found relevant data about Rio' geographic data from the Data-Rio website, which has all the information about districts geography, and a lot of other ones. They use interactive maps powered by Here. The API information can be downloaded from the https://www.data.rio/datasets/limite-bairro URL and it is public. The collected data will be treated and cleaned, then they will be displayed in a dataframe describing the districts and their geometry.

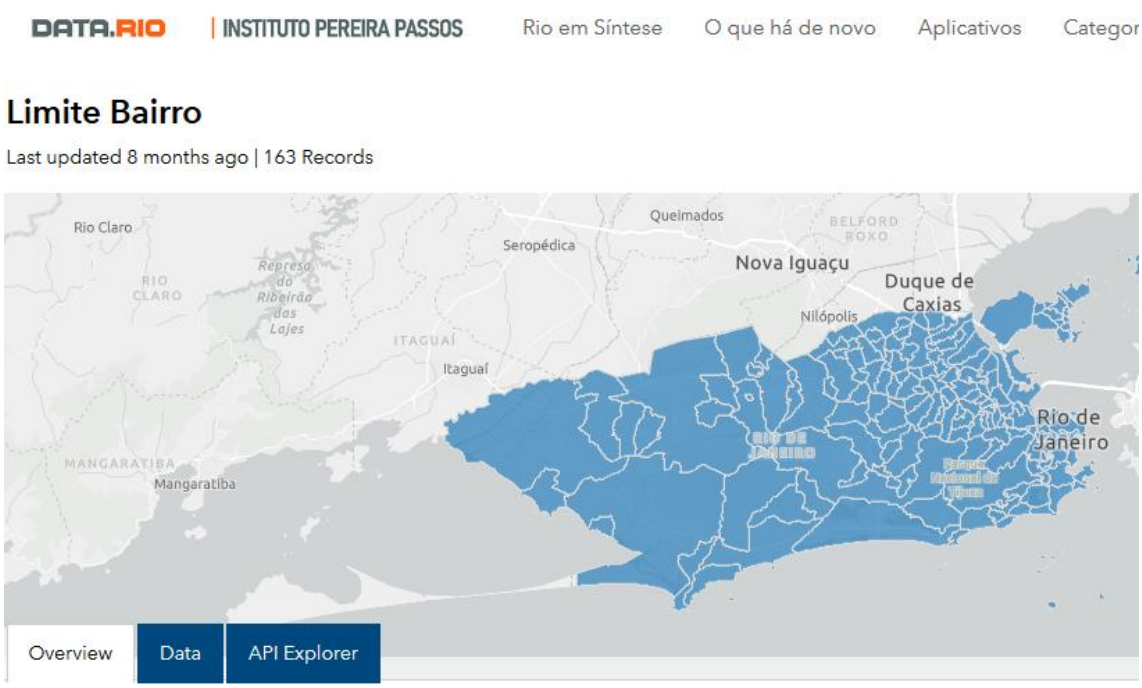

Figure 1 - Data-Rio website

After that, I will create a dataframe containing the Foursquare search data, and will use a for loop to sweep through the districts. To use Foursquare, one must create an account as a developer

I will define a "geometric center" for them, specify a search radius and create a new dataframe. After that, I will concatenate the relevant information in a new dataframe and will display them as a choropleth map.

In the same website, I get data about Rio de Janeiro's population per district as an Excel spreadsheet. That will compose our complete dataframe, seeing most populated areas with more restaurants. It is hard to find the population income in each district; the same with the crime rate, but maybe it is good to perform further research of these data.

Concerning Foursquare API, we can see some inconsistences with their search engine and the API. We can see that when using the API, the results are not as good as the ones obtained in the website. I've been looking for some way to refine this search mechanism but with no avail. The analysis shows some results that do not correspond to the real world when coding this in Python, but with precise ones in the website. As an example we can cite restaurants that are closed but the search API says they aren't, or Italian restaurants ranked in another category.

Conclusion

The code for this survey will be posted in a Github repository. It will be a basic assessment, because most of the tools can be enhanced, and Foursquare is a limited tool inside South America. I will point some problems that can be solved just stressing a little bit more the code, but the core to make a fair choice is there. The choropleth map is a good visual tool to check statistical densities and provides an easy way to visualize how a variable varies across a geographic area or show the level of variability within a region. The other part of the analysis is a subject one, depending on a prior knowledge of the region. Some districts are better than others because they have intrinsic characteristics that make them suitable to be chosen. For example, one place may seem good to place a restaurant, but most of its territory is on a high ground, with difficult access. This project will not cover this type of analysis, even though it may be possible using data.

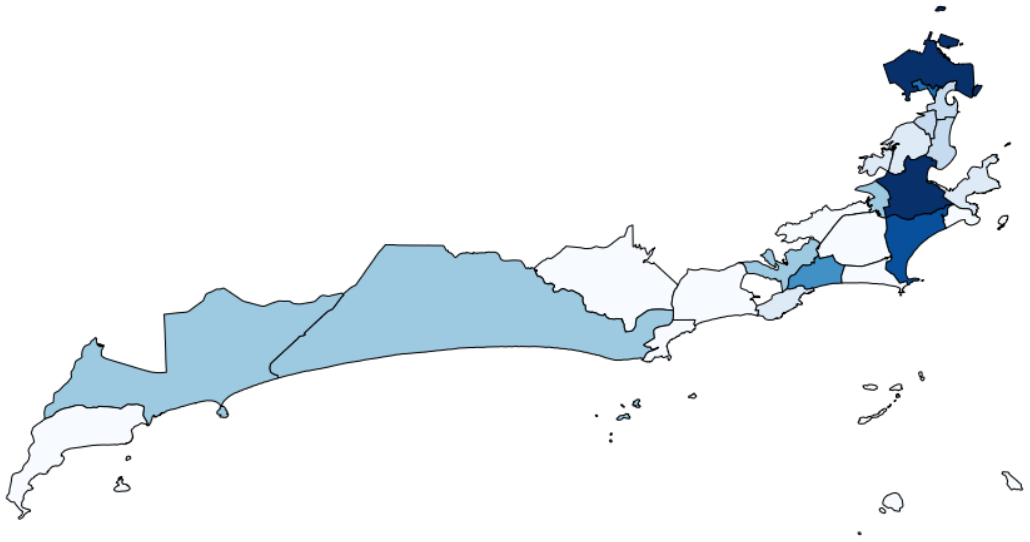

Figure 2 - Final result showing a choropleth map