

Final Capstone Project – The Battle of Neighbours: Week 1.

Introduction

Business problem: Rio de Janeiro is a worldwide known city and it has a great potential for tourism. There's a huge number of restaurants located in several areas, but most of them are small and have low price profile. Rio de Janeiro has a community of Italian immigrants, not as big as São Paulo's community, but significantly big compared with the rest of the country. There is a challenge when trying to open a good Italian restaurant in Brazil, because of the competition among restaurants. I am of Italian descendant, so I would like to open an original restaurant, using data to see where is the best location to open one in Rio.

The objective of this report is to define a strategy to choose the best location to open an Italian restaurant in Rio's noble areas. Basically, the project will search for restaurants in each selected district and the location will be picked in the area with less restaurants in order to avoid competition. The selected districts will have the same potential for attracting clients and are geographically similar.

Methodology

Using Python's map modules along with Foursquare API to solve this problem is interesting because Python has a lot of libraries to work with maps, like Folium. Folium can make leaflet maps, wrapping from Leaflet.JS. The problem with the interactive map is that I cannot see the districts, at least in Rio de Janeiro's area.

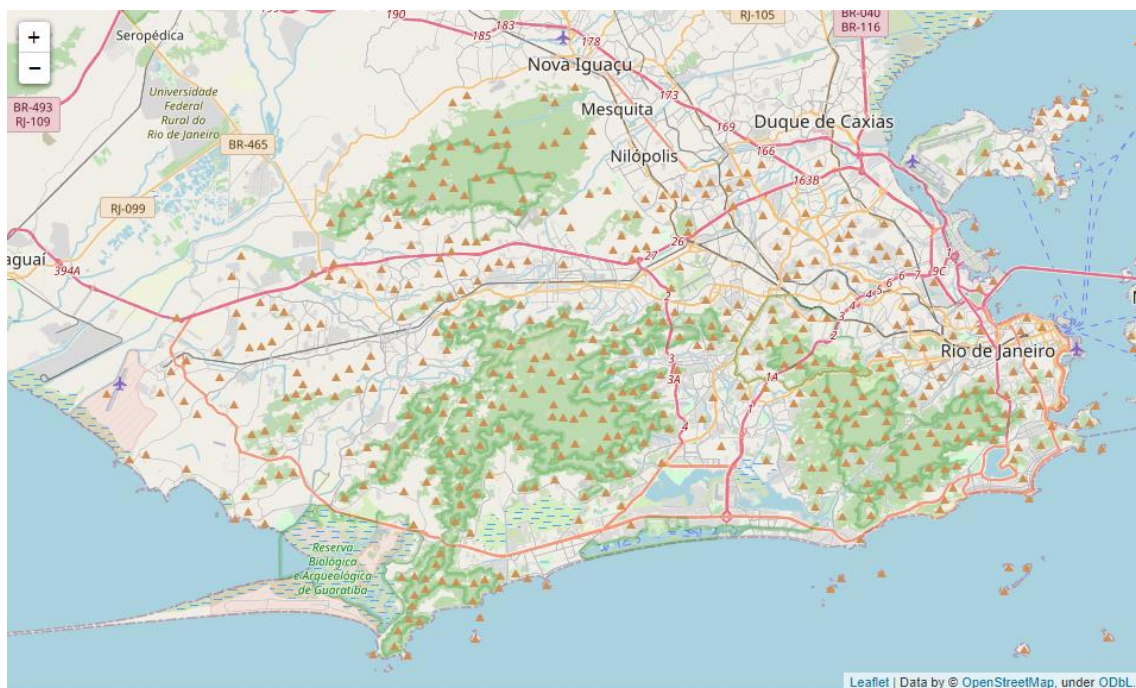


Figure 1 – Rio City map with Folium.

So, to visualize them, I chose Geopandas, a module that works with geospatial data in python. GeoPandas extends the datatypes used by pandas to allow spatial operations on geometric types. Geopandas is a bit tricky to be installed. It needs some dependencies like GDAL (Geospatial Data Abstraction Library), Fiona, which reads and writes geographic data files, pyproj - Python interface to PROJ (cartographic projections and coordinate transformations library), Rtree, which is a Python wrapper of libspatialindex that provides a number of advanced spatial indexing features, and Shapely, used for manipulation and analysis of planar geometric objects. Geopandas does not produce an interactive map, but it works with geometries obtained by the dependencies.



Figure 2 - Rio City map using Geopandas

To get the restaurants, I used Foursquare search API to see how many restaurants are in the districts. Rio de Janeiro City has 163 districts, but most of them are suburbs or low-income zones. So, to open an expensive Italian restaurant, the study will focus in the noblest areas, basically close to beaches in front of the open sea. I will use Foursquare search engine to look for Italian restaurants in each district. Foursquare needs a point and a radius to search for a keyword, so my objective is to use Python to define the center in each district, and then search in a specific radius.

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.

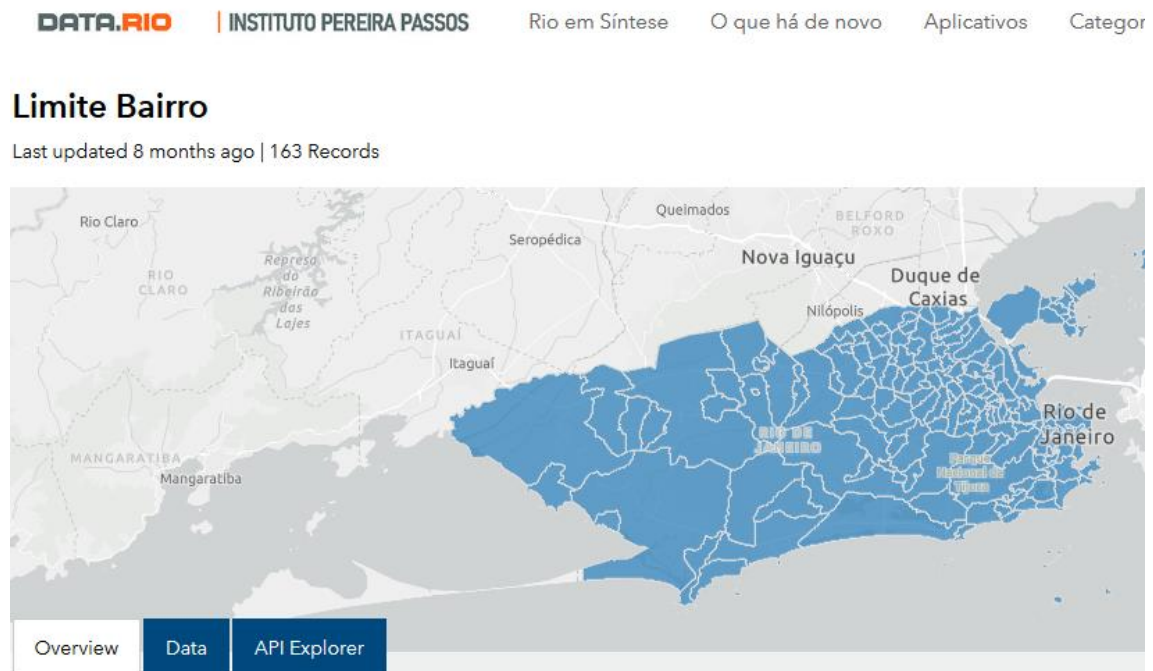


Figure 3 - Data-Rio website

After that, I will create a dataframe containing the Foursquare search data, and will use a for loop to sweep through the districts. 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.

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.

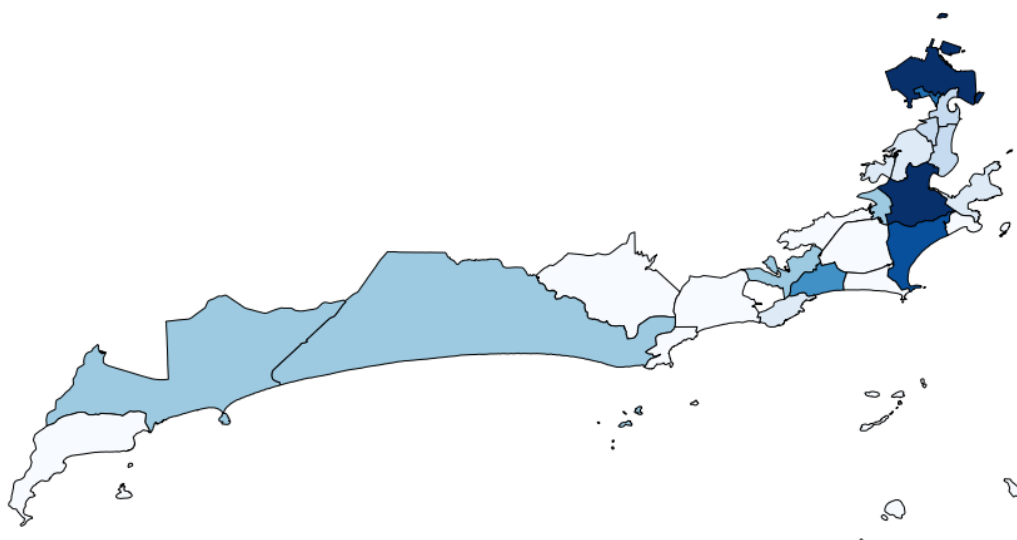


Figure 4 - Final result showing a choropleth map