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. Several factors may have an impact when choosing a place to open a restaurant. One is the population density. Rio de Janeiro's population is not well distributed. It is due its districts having so many discrepant areas and sizes. It means there's a lot of places where the population still has room to grow, and that is good for a long-term vision. Others are crime-rate, land value, people's income and points of interest.

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 population of each district will also be viewed and compared. One thing to be assessed is that the bigger the population, bigger is the number of restaurants. But we are aiming for top level ones, so we cannot just rely on the numbers of people. 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.

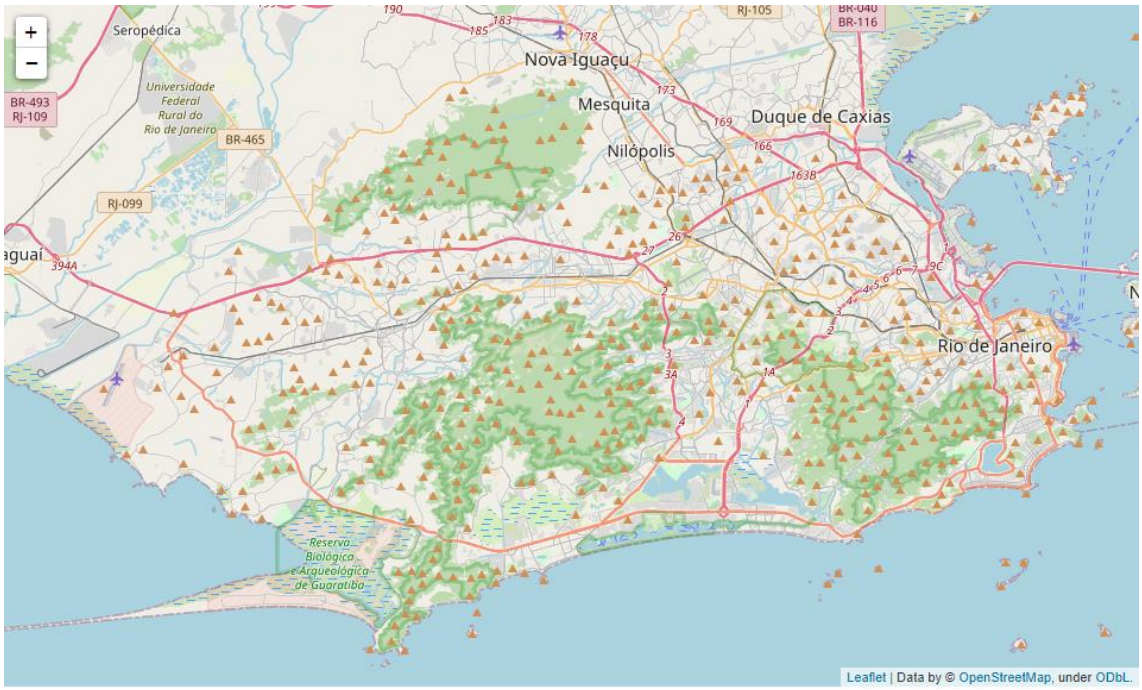

Figure 1 – Rio City map with Folium.

So, to visualize them, we will 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 can produce an interactive map, but it uses modules like **bokeh** to work with geometries obtained by the dependencies.

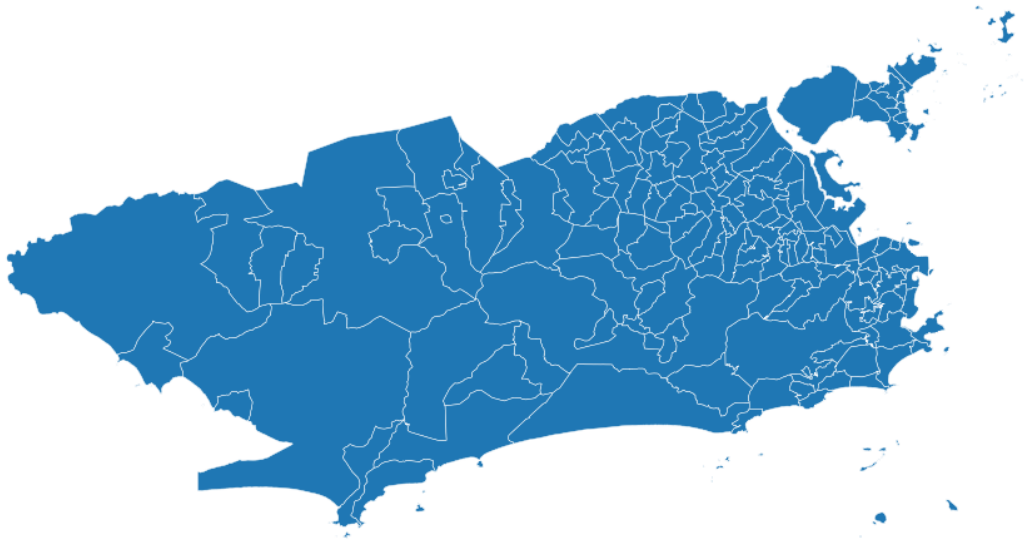

Figure 2 - Rio City map using Geopandas

To get the restaurants, I will use Foursquare search engine 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. There are some particular inconsistences in the Foursquare API that will be discussed in the Data section.

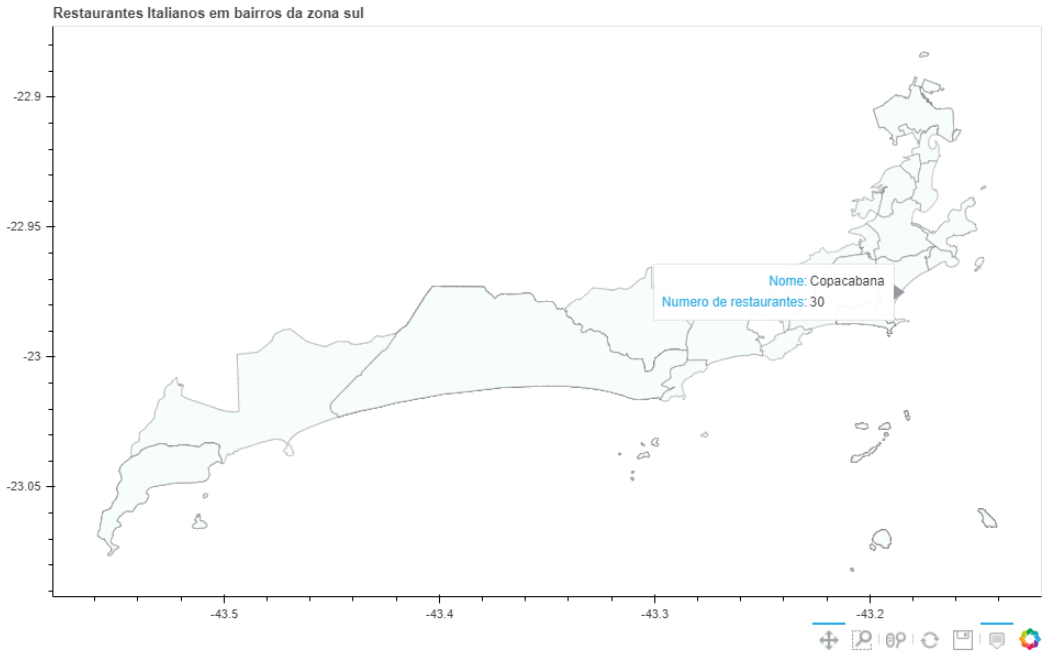

Figure 3 - Example of an interactive plot using bokeh