Beyond beaches: how to find a good place to enjoy a waterfall in Rio de Janeiro

Overview

Brazil is a famous place in the world, because of the nature, wonderful landscapes, carnival, and other characteristics. Thus, one of the places that translates this perception is Rio de Janeiro, where you can find Copacabana beach, Lapa, samba, and so on.

Even though there are multiple alternatives, sometimes it can be hard to find out reliable information about less popular places.

Defining the situation

Imagine a situation where a foreign tourist arrived in Rio de Janeiro to stay for a few days.

After the first days visiting famous spots, as nature lover, he decided to find a good waterfall to visit, because he read on internet about this.

At first, he struggle to find clearly information, so he decided to use Foursquare API to map all waterfall close to Copacabana (the hostel he's been).

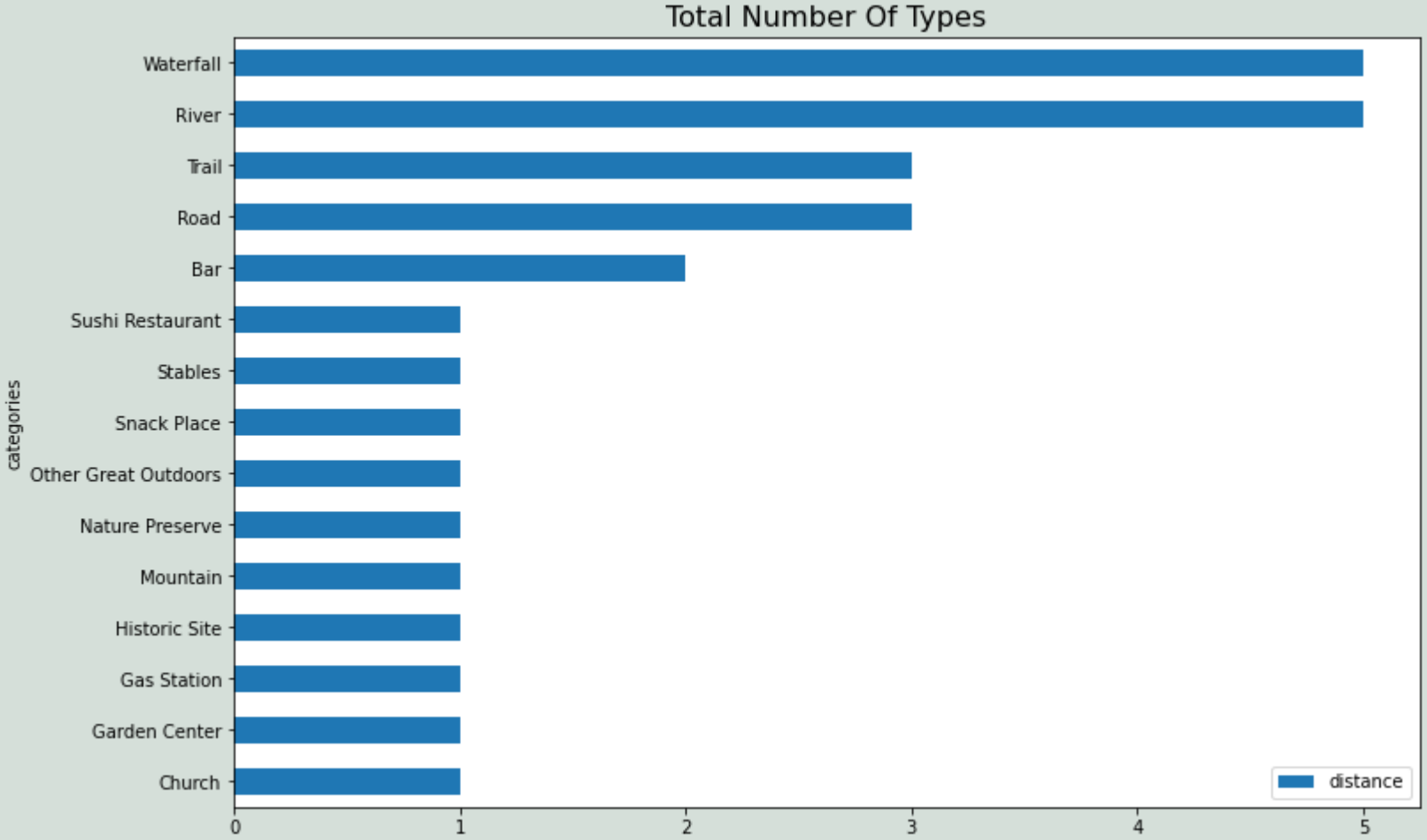
Data collection and data wrangling

1. First, Geolocator must be set to Copacabana as the initial spot.
2. After that, the API is called, in order to give the information that has been looked for. The API will provide latitude and latitude from waterfalls. Then, the results will be filtered to call the API again to get the rating.
3. Pandas is used to create a structure dataset, remove unnecessary data, and to define the dots to show in a plot.
4. Finally, after all the analysis, the results will be shown the chosen places to visit in the map.

Assumptions

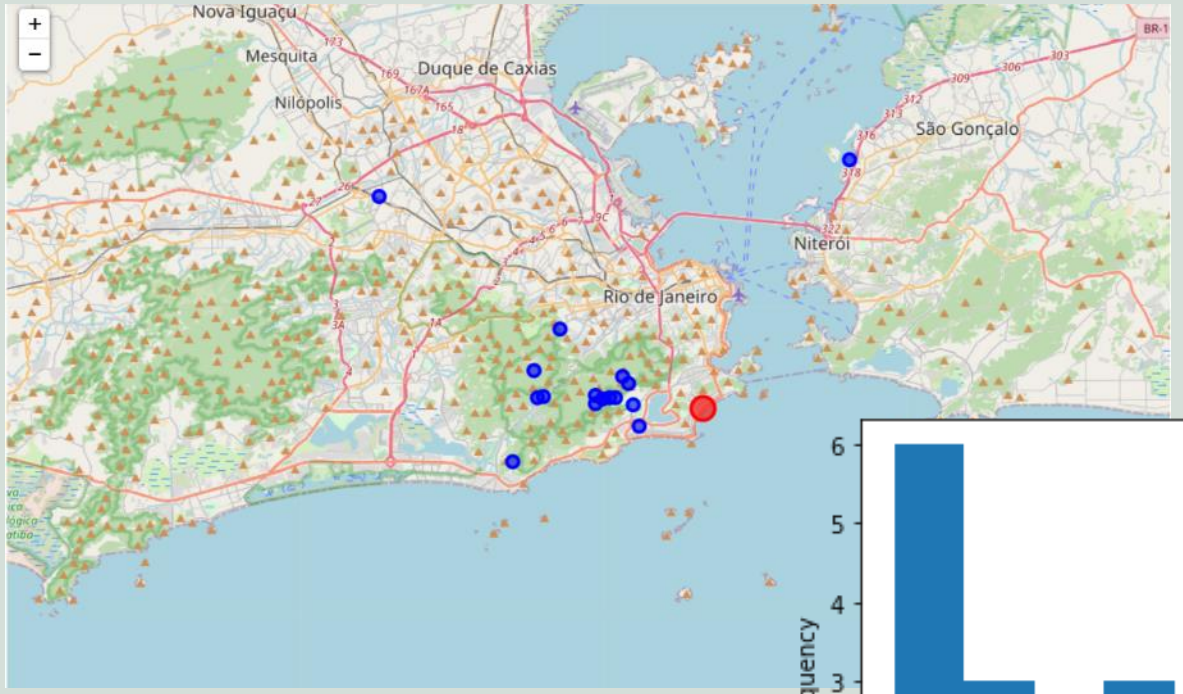
1. Once the city is wide, it was defined, at first, a search in a range of 20 km. After that, the preference is to find spots that is close to the place the tourist has been (up to 10 km).
2. Since Portuguese is the native language in Brazil, the query was defined to search *cachoeira* instead of waterfall.
3. The probability to find multiple results is low, so the constraints in a free account of Foursquare API would not be a problem at all, even to collect the rating data.
4. Results far different from waterfall-related terms will be discarded, in order to improve the analysis.
5. Only good-rating waterfall will be considered to visit (more than 7).

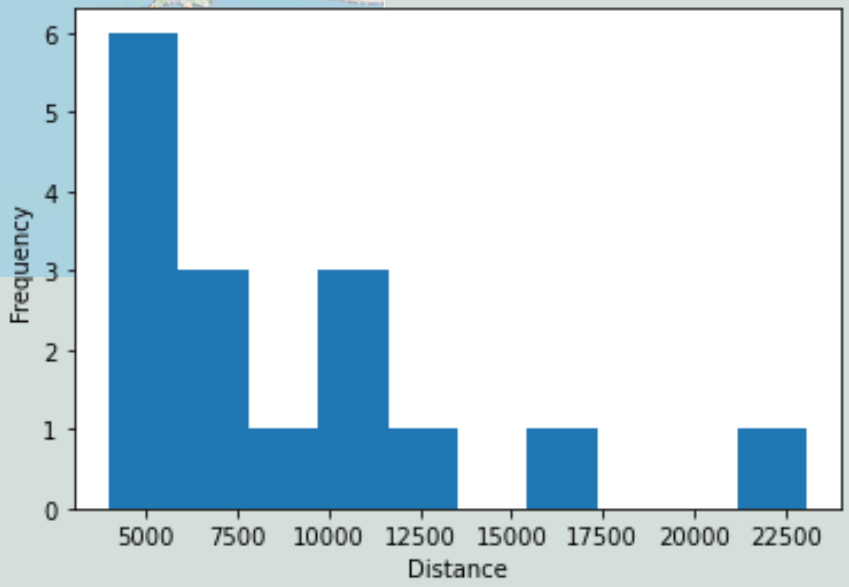
Analysis



After run the query, 15 different types of categories was found, including even church or bar. In order to improve the results, these unrelated categories was removed. The only ones that remained was Waterfall, River, Trail, Other Great Outdoors, Nature Preserve, and Mountain.

Analysis

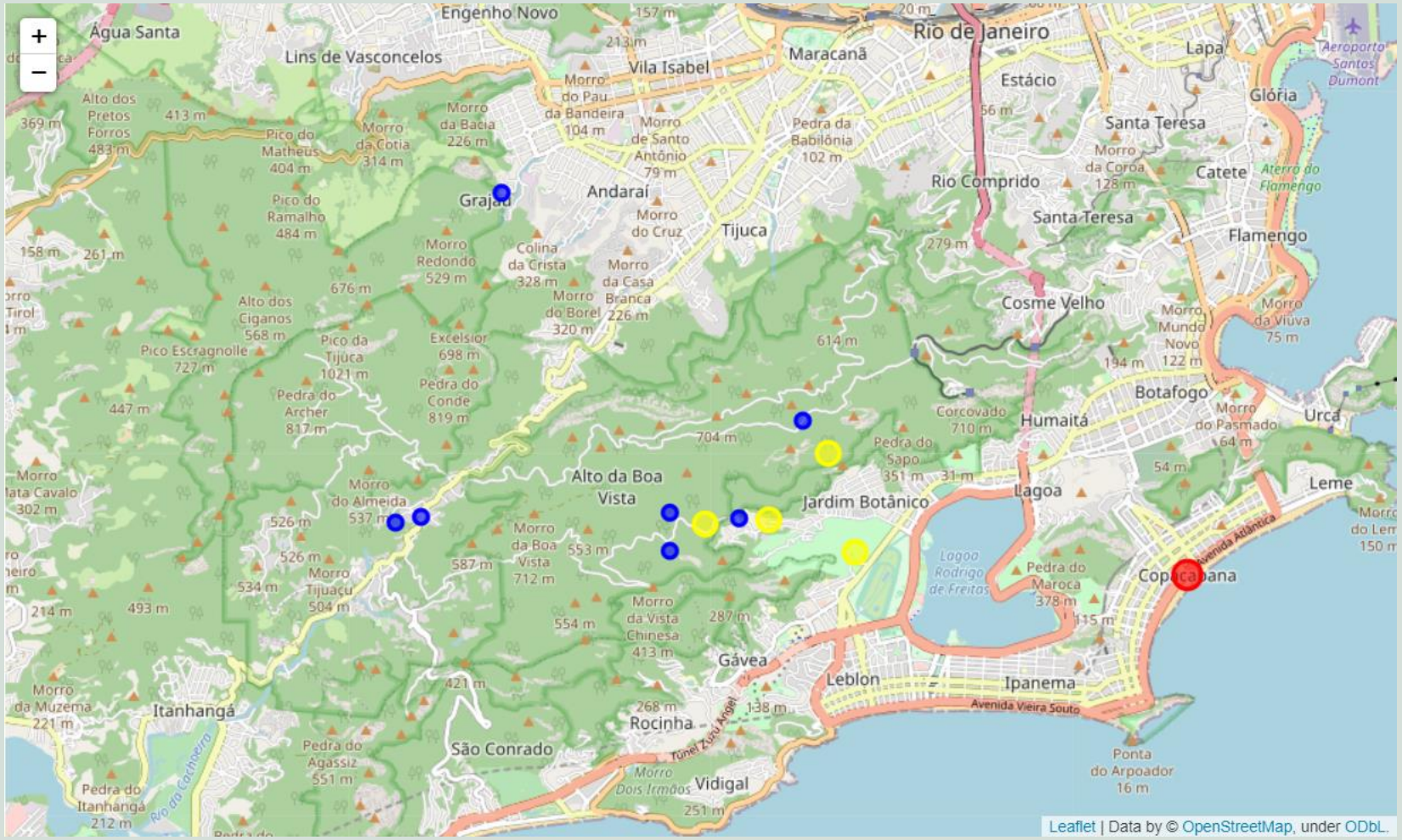




Some of the spots, as we can see in the map, is far from the initial spot (highlighted in red). In this case, everything far from 10km was removed (as set in the assumptions).

It is clear in the histogram, that some spots will be discarded, and the results will be reduced to only 11 spots.

Final Result



Finally, after cleaning the data and including all the ratings, 4 waterfall were chosen. In this case, only these results spotted in yellow (the chosen ones) received any rate.

All spots in blue remains just to show the discarded results. Even though it could be analyzed by different characteristics in the future.