

Capstone Project - Clustering neighborhoods in Belo Horizonte - Brasil

Introduction

In this problem I am trying to discover the best neighborhood to open a Vegan / Vegetarian restaurant in Belo Horizonte.

We will try to find neighborhoods that has lots of restaurants, bars and other comercial place, but no Vegan/Vegetarian restaurants nearby.

Data aquisition and cleaning

Data Sources

First of all, in the Wikipedia site, https://pt.wikipedia.org/wiki/Lista_de_bairros_de_Belo_Horizonte I will get the neighborhood's list parsing with Beautiful Soup.

After that, I will access <http://geocoder.pbh.gov.br/geocoder/0.0.1/address?bairro={ }>, that is a City Hall official site and get the UTM directions of all neighborhoods.

Cleaning and transforming data

Now, we use some panda's tools to clean the data and then use the pyproj library to get the Geographical Coordinates and use the foursquase API.

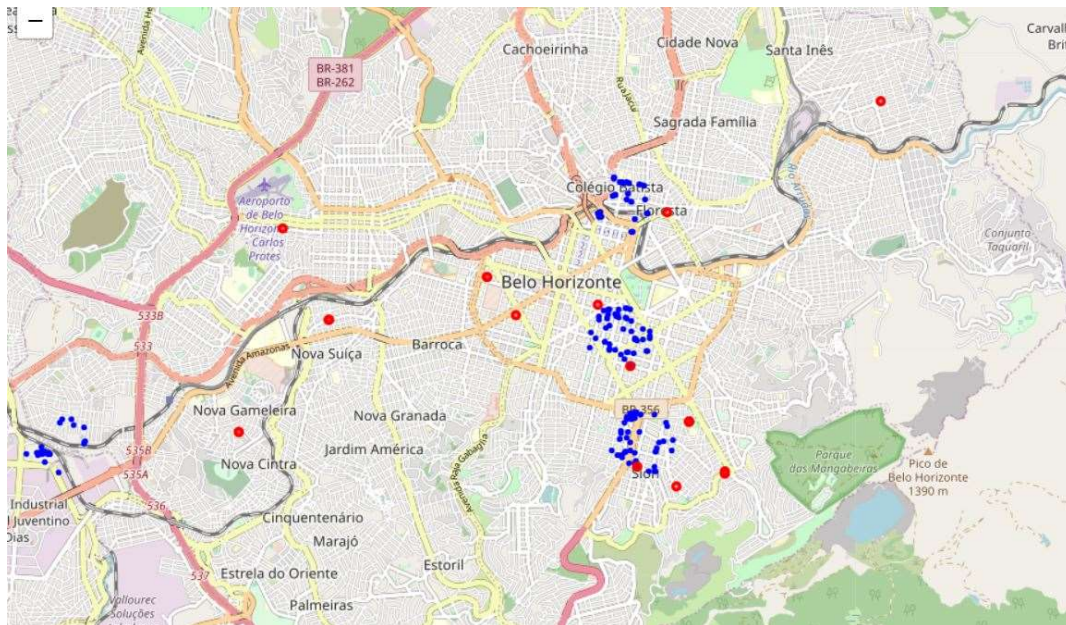
Feature selection

After all the aquisition and transformation, we will get a DataFrame with 496 rows and 5 columns, named as "Bairro", "ZONA", "Regional", "Lat" and "Long".

Now we are ready to access the foursquare data to cluster and define the best place to open this Vegas/Vegetarian Restaurant.

Methodology

Using some graphical and maps library as geocoders, matplotlib and folium we are able to do cluster the neighborhoods, print that mais points in the city maps and find the best places to open this Vegan/Vegetarian Restaurant.



Conclusion

With this map, it's easy to get the conclusion what is the best neighborhood to open the restaurant as it was proposed on the introduction section.