

IBM DATA SCIENCE CAPSTONE PROJECT

MATHEUS MAFRA OLIVEIRA ANDRADE

INTRODUCTION

Belgium is the 13rd most populous country in Europe and the 11th most developed in the continent. It is also known by its linguistic and cultural diversity, and we cannot forget the Belgian tradition in producing handcrafted beer. Its capital, Brussels, is famous by the cafés, pubs, and bars, where the locals and tourists can spend time with friends and enjoy a good Belgian beer.

With this in mind, an entrepreneur has a dream of opening a new bar in Brussels where he can sell different beers, but he does not know where he should open it. His priority is to avoid strong competition, because he wants to gain trust from clients and a good market share before expanding his business. Thus, he wants to open the bar somewhere distant to other bars.

He is now analyzing the municipalities in the city's metropolitan area and, with the help of Foursquare dataset about places in Brussels, the entrepreneur will cluster the city's municipalities to check in which of them there are less bars in comparison to other types of venues, and, in other words, the best municipalities to open a new bar.

DATA DESCRIPTION

In order to complete the entrepreneur's study, he will need data from three different sources:

A table containing the names of all Brussels' municipalities, based on data retrieved in 2015 by the Brussels Regional Informatics Centre (CIBG/CIRB). The table can be found on the Wikipedia page:

https://en.wikipedia.org/wiki/List_of_municipalities_of_the_Brussels-Capital_Region;

A table containing the names of venues in Brussels, along with their category, latitude, and longitude;

Coordinates for each municipality, collected in the Python Geopy library.