

What is Foursquare opinion on splitting up Belgium?

CAPSTONE PROJECT FOR COURSERA IBM DATA SCIENCE
SPECIALIZATION

ADRIEN ALLART

Contents

1. Introduction	2
2. Data	3
2.1. Dataset	3
2.2. Data Cleaning / Data Manipulation	4

1. Introduction

An over-simplification of Belgium structure as a country could be visualize as 3 regions (Flanders in the North and Wallonia in the South, Brussels in the middle) with two different language (Dutch and French) living under one flag. The idea of splitting the country is a topic that is often discussed with a lot of arguments on both sides. It involves a lot of geo-political, socio-economic and cultural questions.

Answering whether this would or would not be a good idea is obviously out of the scope of this project as it is an extremely complex problem. The idea of this project is to compare how similar or dissimilar cities from the regions are based on location data from Foursquare. It will also extend the analysis to neighboring countries (The Netherlands and France) which are often considered as viable options for regions of Belgium with which they share a language.

This will hopefully provide another perspective and be an objective data-driven argument in the discussion. This might be interesting for decision maker that have to handle that question, might also be of use of anyone wanting to build their opinion on the question or to validate/challenge their existing opinion. It might also give some insights on how similar some of the founding members of the EU are.

More information about the partition of Belgium might be found at:

https://en.wikipedia.org/wiki/Partition_of_Belgium

2. Data

2.1. Dataset

There are two main source of data that are used in the following analysis:

The first dataset required is a list of cities for each country (Belgium – France – The Netherlands) with their longitude – latitude coordinates. The list of city with their location can be found in a Kaggle dataset from MaxMind [here](#) . The same dataset give regions for each country which for the case of Belgium matches the country provinces. That information is used to separate the country by its regions.

The raw data look like this:

```
1 Country, City, AccentCity, Region, Population, Latitude, Longitude
2 ad, aixas, Aixàs, 06, , 42.4833333, 1.4666667
3 ad, aixirivali, Aixirivali, 06, , 42.466666700000005, 1.5
4 ad, aixirivall, Aixirivall, 06, , 42.466666700000005, 1.5
5 ad, aixirvall, Aixirvall, 06, , 42.466666700000005, 1.5
6 ad, aixovall, Aixovall, 06, , 42.466666700000005, 1.4833333
7 ad, andorra, Andorra, 07, , 42.5, 1.5166667
8 ad, andorra la vella, Andorra la Vella, 07, 20430.0, 42.5, 1.5166667
9 ad, andorra-vieille, Andorra-Vieille, 07, , 42.5, 1.5166667
10 ad, andorre, Andorre, 07, , 42.5, 1.5166667
11 ad, andorre-la-vieille, Andorre-la-Vieille, 07, , 42.5, 1.5166667
12 ad, andorre-vieille, Andorre-Vieille, 07, , 42.5, 1.5166667
13 ad, ansalonga, Ansalonga, 04, , 42.5666667, 1.5166667
14 ad, anyos, Anyòs, 05, , 42.533333299999995, 1.5333333
15 ad, arans, Arans, 04, , 42.5833333, 1.5166667
16 ad, arinsal, Arinsal, 04, , 42.5666667, 1.4833333
17 ad, aubinya, Aubinya, 06, , 42.45, 1.5
18 ad, auvinya, Auvinya, 06, , 42.45, 1.5
19 ad, bicisarri, Biçisarri, 06, , 42.4833333, 1.4666667
20 ad, bixessarri, Bixessarri, 06, , 42.4833333, 1.4666667
21 ad, bixisarri, Bixisarri, 06, , 42.4833333, 1.4666667
22 ad, canillo, Canillo, 02, 3292.0, 42.5666667, 1.6
23 ad, casas vila, Casas Vila, 03, , 42.533333299999995, 1.5666666999999999
24 ad, certers, Certers, 06, , 42.466666700000005, 1.5
25 ad, certes, Certés, 06, , 42.466666700000005, 1.5
26 ad, eixirivall, Eixirivall, 06, , 42.466666700000005, 1.5
27 ad, el pui, El Pui, 04, , 42.55, 1.5166667
28 ad, els bons, Els Bons, 03, , 42.533333299999995, 1.5833333
29 ad, el serrat, El Serrat, 04, , 42.6166667, 1.55
30 ad, els plans, Els Plans, 02, , 42.5833333, 1.6333333
31 ad, el tarter, El Tarter, 02, , 42.5833333, 1.65
32 ad, el tremat, El Tremat, 03, , 42.55, 1.5833333
```

The second set of data required for building the tool is [Foursquare](#) API that will be used to retrieve information of surrounding venues for the geolocation data chosen from the previous set of data.

The API transform longitude and latitude data into a JSON with list of venues located around the position and the category of the venue (as well as a lot of other information that are not used for this project).

Not critical for the overall result, but geojson file of Belgium provinces is used to draw the border of each provinces on a couple of map. The data was found on [digitalwallonia.be](#).

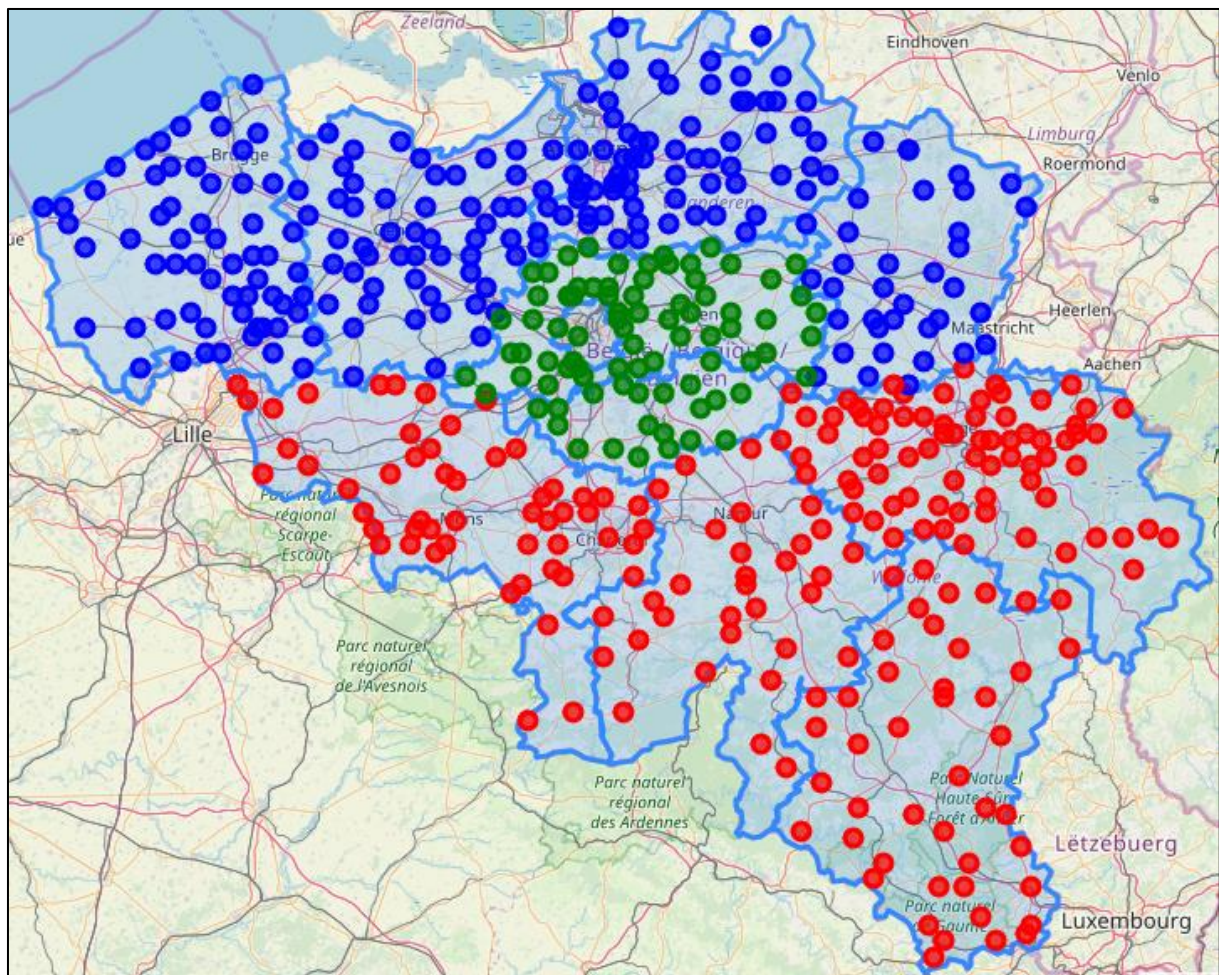
2.2. Data Cleaning / Data Manipulation

Some data cleaning has to be done on the dataset that are used:

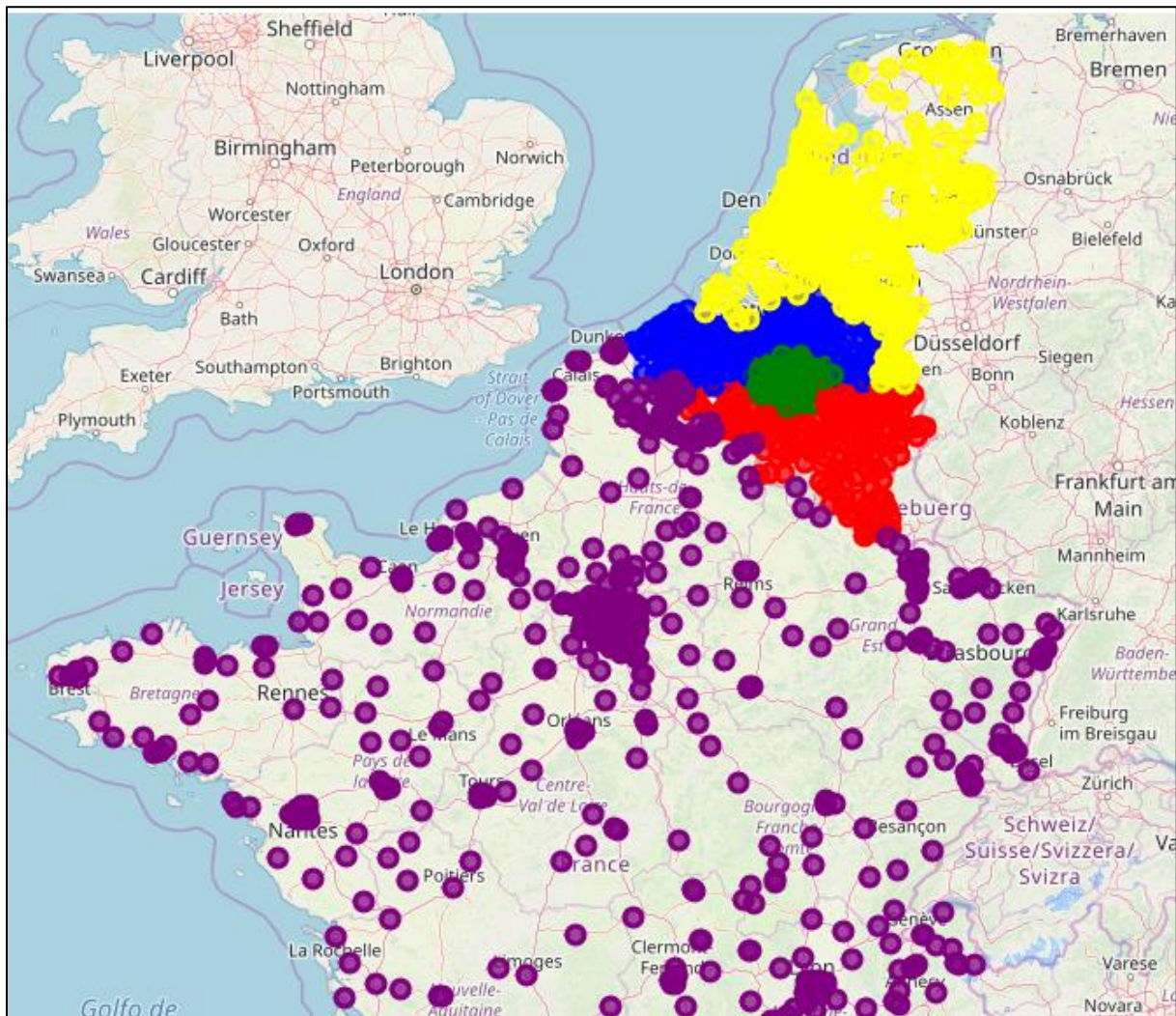
1. The cities dataset has to be rescope to keep only the country of interest for the analysis (Belgium – France – The Netherlands)
2. The numerical value for the region corresponding to Belgium in the cities dataset had to be transform in either Flanders ,Wallonia or Brussels. There was some issue with the province Brabant which should be split into 2 (Flanders Brabant and Wallonia Brabant) but the dataset had most cities grouped into a single number with a couple cities on their own. To simplify the analysis I decided to include all of it into the Brussels region. The table I used for the transformation is as below (source: me, am Belgian):

01	:	Antwerp	:	Flanders
02	:	Brabant (except couple cities)	:	BXL
03	:	Hainaut	:	Wallonia
04	:	Liege	:	Wallonia
05	:	Limbourg	:	Flanders
06	:	Luxembourg	:	Wallonia
07	:	Namur	:	Wallonia
08	:	Flandre or	:	Flanders
09	:	Flandre occ	:	Flanders
10	:	single city	:	BXL
11	:	BXL	:	BXL
12	:	single city	:	BXL

3. Visualization of the cities on the map allowed to spot an error in the dataset and fix it



Similar visualization was done for the second part of the analysis with The Netherlands and France:



4. The city data is used to create a dataset of the venues using foursquare:

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	aalst	50.933333	4.033333	Basic-Fit	50.934496	4.034168	Gym / Fitness Center
1	aalst	50.933333	4.033333	De Prins Drinkt Koffie	50.934941	4.038068	Coffee Shop
2	aalst	50.933333	4.033333	Tang's Palace	50.935172	4.038079	Chinese Restaurant
3	aalst	50.933333	4.033333	Healthfactor Gym	50.930120	4.035189	Gym / Fitness Center
4	aalst	50.933333	4.033333	Pizza Talia	50.932481	4.030968	Pizza Place