

2 DATA

2.1 The Data We Need

To find the perfect promoting place we need data. What data do we need?

- A list of all the cities in Belgium.
- Latitude and longitude coordinates of the city's. We need this data to plot the map and to search for the city's with the most and best related venues.
- if we find the perfect city to promote our beer we need information about the neighbourhoods of that city.
- Latitude and longitude coordinates of the neighbourhoods in this city.
- Venue data relate to beer promoting. We will use this data to find the best neighbourhood to promote our beer with clustering.

2.2 The Source of The data

The Wikipedia page (https://nl.wikipedia.org/wiki/Lijst_van_steden_in_Belgi%C3%AB) gives us a list of all the city's in Belgium. the data is gonna be extracted from this list with web scraping techniques and Python. The latitude and longitude coordinates of those citys with the Python Geocoder Package. The Foursquare API is used to get the venue data for the city's. Foursquare has a huge database with over 100+ million places. The API will give us categories of the venue data and we need the categories linked with our cause. That is finding the best city to promote our beer and what neighbourhood we need to promote the beer in. After finding the best city we need a neighbourhood databank of the neighbourhoods (also on Wikipedia). Repeat the Foursquare process more local in the best city. Our data science skills will be needed. Web scraping, working with Foursquare API , data cleaning, data wrangling, machine learning techniques and map visualization (Folium) will be used during this research.