COURSERA CAPSTONE: BATTLE OF NEIGHBORHOODS

Promoting my newest beer in Belgium

By: DEVOS Tore

June 2020

INHOUDSOPGAVE

1   INTRO

1.1   Introduction

Belgium's beer-making history goes back centuries and it's famous the world over as being a top beer country. Belgium has wel over 200 000 cafés which meant there is a café for every five houses. Also beer consumption at home is still a common thing in Belgium. It's the favourite drink of the Belgium people. Together with some friends, I developed a new beer named 'Kelderkopke'. We want to promote this beer with the Belgium people. It's important to promote the beer in the right neighbourhoods. If the beer gets know in the places where people drink a lot of beer then our demand and popularity for the beer can grow a lot because of that. We only have a few people in our team so finding the best location to start with our promoting campaign will be very important if we want this to have success. if we promote the beer in the wrong places, we can incur a lot of costs without much result.

1.2   Business Problem

The objective of the Capstone Project is to analyse the citys in Belgium and their neighbourhoods and select the best locations to promote our new beer 'Kelderkopke'. Using data science methodes and machine learning techniques we can find solutions to answer our business problem**: If we want to promote our new beer in Belgium, in what location should we start ?**

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 the internet) 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.

3 METHODOLOGY

3.1   City Selection

Data containing name data of all the different citys in Belgium was found on Wikipedia. The data was manually extracted from Wikipedia and on a Data website of Belgium the coordinates were found. The data need to be cleaned and some useless information was dropped. This data was merged into one dataset. Because some city datasets have more subcity parts, the data need to be cleaned again. After analysing the data, a way was found to one keep the major part of the city. This resulted in our first cleaned dataset (as shown in Figure 1).

	postal	City	Latitude	Longitude	Population
0	1000	Brussel	50.842750	4.351550	168576
1	1500	Halle	50.732485	4.234913	37104
3	1800	Vilvoorde	50.927251	4.425787	41432
4	2000	Antwerpen	51.219877	4.401136	507911
11	2200	Herentals	51.176879	4.835648	27438

Figure 1: geodata_df CITYS OF BELGIUM

With this data a map is made to visually see where all those City's are. The interest for those city's is high if the population in the city is high, because we need to reach a lot of people with our promotion campaign. The data in figure 1 is used to select the 20 city's with the largest population. The city's were selected using 'nlargest' on population. Using the matplotlib pyplot we made a good visualisation of those city's (as shown in figure 2).
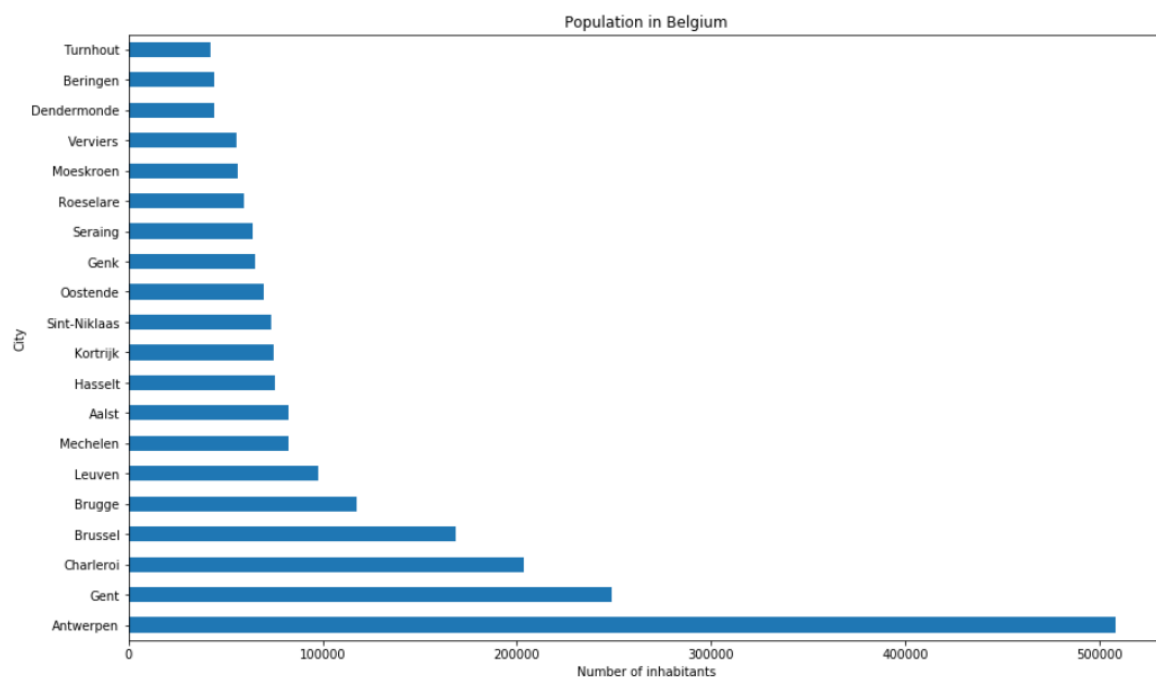

Figure 2: Barplot of the largest populated Belgium city's

To get a better visualisation of our selection two maps were made to show how we went from all the Belgium city's to the top 20 (As shown in figure 3).
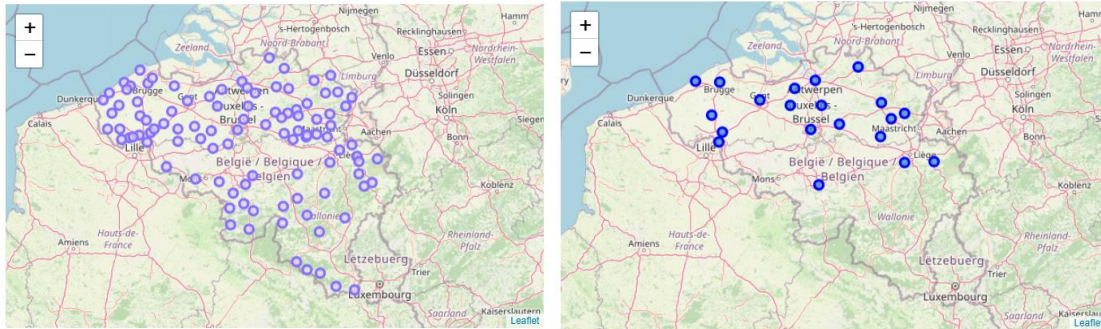

Figure 3:map transformation from all the citys to the largest populated Belgium city's

Next, Using the Foursquare API for each city in a radius of 2000 meters from the city centers, the venues of those 20 locations were analysed. The amount of venues that were returned per city is limited with only the top 100. All the venues were analysed if they had something to do with the usage of beer. For easy use it is called the Number of Bars and made a barplot with this data (as shown in Figure 4).
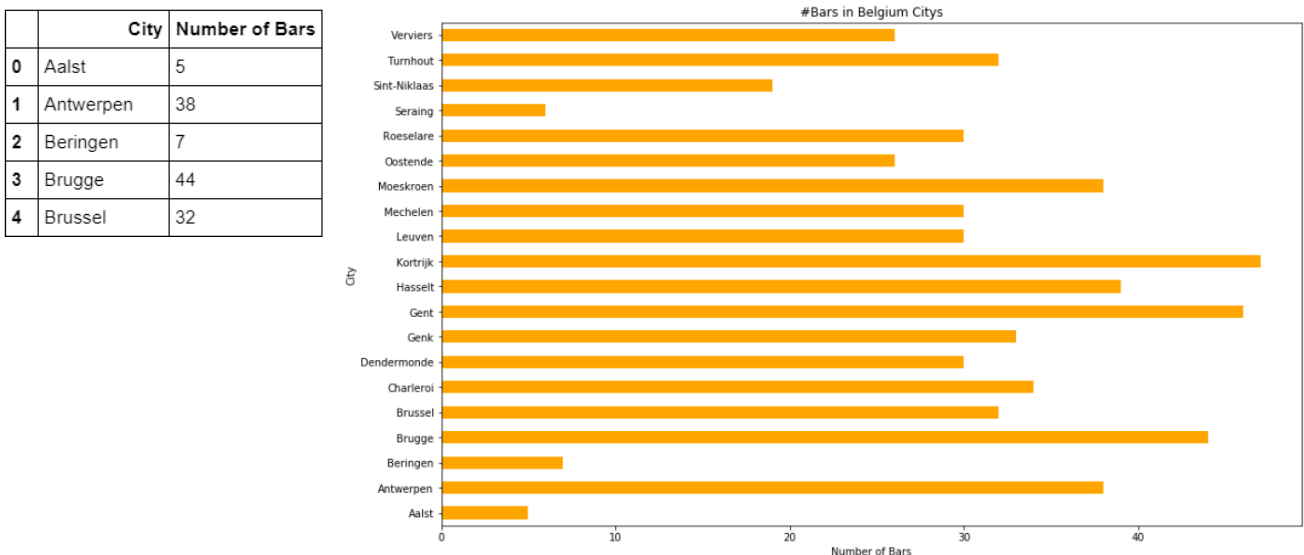
	City	Number of Bars
0	Aalst	5
1	Antwerpen	38
2	Beringen	7
3	Brugge	44
4	Brussel	32

Figure 4: Head of bars dataframe and barplot

Analysing this data the 3 city's with the most beer related venues per 100 venues were selected.
It is clear that Gent has the biggest population of the 3 city's, so our campaign and beer promoting is going to start in gent. Now it's a matter of finding the best location for our Beer promoting campaign inside of Gent (as shown in Figure 5).

	postal	City	Latitude	Longitude	Population	Number of Bars
100	9000	Gent	51.067831	3.729091	248813	46.0
79	8000	Brugge	51.209349	3.224701	117577	44.0
85	8500	Kortrijk	50.819489	3.257708	75120	47.0

Figure 5: TOP 3 CITY'S

3.2 Neighbourhood Selection

By Scraping data from the internet, the dataset for the Neighbourhoods was made. A dataset from the data website of gent delivered the names and coordinates of all the Neighbourhoods in Gent.
This dataset (https://data.stad.gent/explore/dataset/stadswijken-gent/table/?flg=nl) was uploaded to github to gain easy access. After cleaning up the data a map was made of all the Neighbourhoods.
(as Shown in Figure 6)

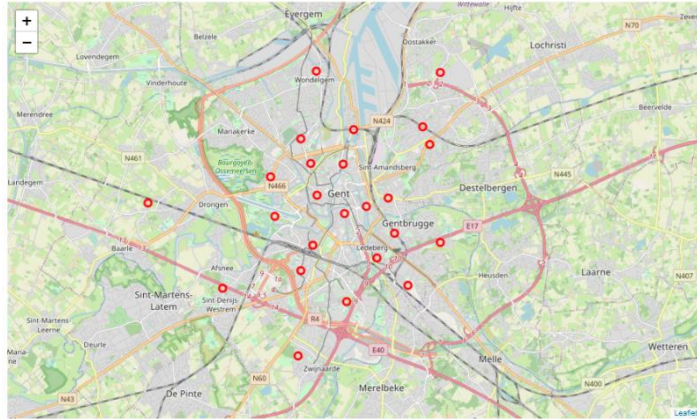|	Neighbourhood	Latitude	Longitude
0	Binnenstad	51.0477602841	3.72765891938
1	Mariakerke	51.0741260697	3.76585871261
2	Moscou - Vogelhoek	51.0256748878	3.75871589607
3	Stationsbuurt-Zuid	51.0302467102	3.70661267376
4	Nieuw Gent - UZ	51.0207649274	3.72885601604

Figure 6: Head of Neighbourhood datafram and map of Gent

With Foursquare API the data was explored to see where the most beer related venues were located. We selected the venues related with beer cafetaria café restaurant and bar.
With the use of a machine learning algorithm using k-means clustering 5 clusters were made.
Going from Beer related area's to not related area's. Those labels were matched with the right Neighbourhoods and plotted onto the citymap of Gent (as shown on Figure 7).

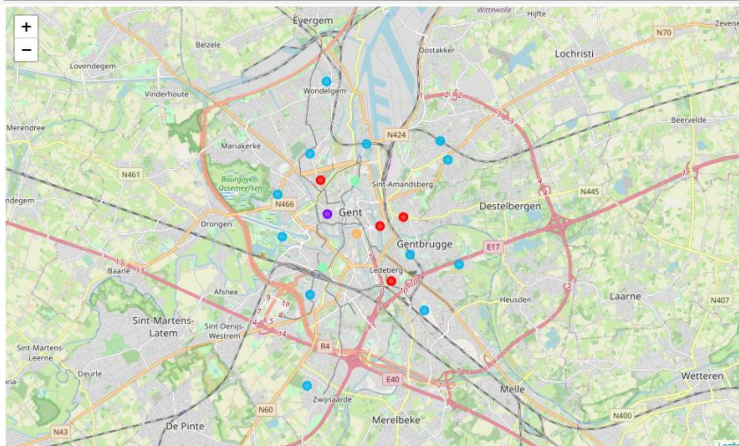|	Neighbourhood	Number of Bars	Latitude	Longitude	Cluster Labels
0	Binnenstad	35	51.047760	3.727659	4
1	Bloemekenswijk	1	51.070534	3.706548	2
2	Brugse Poort - Rooigem	4	51.058966	3.691731	2
3	Dampoort	9	51.052336	3.749105	0
4	Elisabethbegijnhof - Prinsenhof - Papegaai - S...	30	51.053335	3.714365	1

Figure 7: Clusters of Neighbourhoods

4 RESULT/DISCUSSION

Most of the restaurants and pubs are concentrated in the central area of Gent , with the highest number in the orange cluster and moderate number in cluster purple and green. On the other hand, the blue cluster has very low number to totally no beer related venues in the neighbourhoods. To promote our beer we want to start in the area with a lot of pubs and restaurants. We could start in the city center and then promote our beer in the different parts of the city near the purple and green clusters. Those areas will also be good. We need to stay away from the blue areas. There is almost no beer related venues in this area. Therefore, this will not reach the right target group of peoples. Another good idea might be to explore the other candidate city's we had in our top 3: 'Kortrijk' and 'Brugge'. If we could find the neighbourhoods in those city's were beer related venues are, we can make advantage of that. instead of staying in Gent and expand our promoting area in one city, we could move to the most beer related parts of 'Kortrijk' and 'Brugge'. in that way we can reach more different people with the same group of people. Finally , Gent binnenstad is the most attractive neighbourhood to start with our Beer promoting campaign after this analysis of Belgium.

5 CONCLUSION

In this project I've extracted data from the internet. I was able to get the data clean and plot the data onto a map. Next, the data was analysed to find the perfect city for our promoting campaign for our new beer. The Neighbourhood data of Gent was also cleaned and analysed. This time we used a Cluster Algorithm to find the best Neighbourhood to promote our beer. This application can be used by any person who likes a good drink. Having the possibility to find good locations in an unknown country can be very handy to get the best target group possible. For example, if we want to promote our beer in other countrys we can use this tool to find the city and neighbourhood where we should go to to get the best possible future consumers. Our system can be made better to improve recommendations in the future. Not only bars and restaurants etc plays a role in what would be the best Neighbourhood. Also Age, Income (Poverty and wealth), Popularity of a City and other factors plays a role. This project only used Foursquare and location data and did a good job with the info.