

Clustering Barcelona neighborhoods

Ian Riera Smolinska

April 06, 2020

1 Introduction

1.1 Background

Historically cities have been divided administratively into districts and neighborhoods, and so is the case of Barcelona. The city is divided into 10 districts and 73 neighborhoods. However, society is constantly changing and the demography of the cities as well. In the case of Barcelona, there was a huge before and after the Olympic games back in 1992 and nowadays is the destination for 20 million tourists every year. In fact, many decided not just to visit it but to move to Barcelona, as the statistics recall that, in 2019, 20 % of the inhabitants of the city were foreigners. [1]

This constant flux of travelers makes opening a restaurant, shop and other commercial venues, a tempting investment to get profit from the visitors. In 2019, there were 80.500 street-level commercial venues in Barcelona. [2]

As a consequence of the previous two, a huge amount of user rating on local venues is generated every day in platforms like Tripadvisor or Foursquare.

1.2 Problem

The administrative division of the city provides a handful distribution, but not a representative one of the inhabitants' characteristics. The economic power, commercial sectors, age or unemployment among others, would provide a richer approach to segment the neighborhoods of the city. This project aims to cluster the neighborhoods of the city basing on different parameters, detect if there are any correlations between them and test how fit a user-generated social media platform as Foursquare would be for segmenting a city.

1.3 Interest

The target group for this analysis is any company or individual that is considering opening a commercial venue in Barcelona. Basing on the existing data, it can compare how the similar venues are distributed around the city and the characteristics of each neighborhood, choosing to settle in the one that fits more their products.

Note that this data is previous to the Covid-19 pandemic, that might carry an economical crisis and transformation. Several venues might close due to it. This project will be retaken in 2021, when data from 2020 is available and compared with the obtained results for 2019.

2 Data preparation

2.1 Data source

The Barcelona City Council publish statistics openly on their webpage. This includes demographic, economic, political and laboral information grouped by neighborhoods. The following parameters for each neighborhood have been used in this project:

- Nationality. [1]
- Unemployment. [3]
- Local elections voting. [4]
- Age. [5]
- Family Income Available index. [6] [7]
- Street-level commercial venues. [2]

In order to be able to work on these datasets, the Beautiful Soup [8] library for web scrapping has been used to obtain the information to fill the created Panda data frames [9].

Additionally, the Foursquare API [10] is used to access the platform information.

2.2 Data processing

First of all, the information published by the Barcelona's city council is either in Catalan or Spanish, so all the tables had to be translated to English. Additionally, the table for commercial had the column and rows transposed so had to be corrected.

For the data analysis, the format of the provided tables allows an easier extraction of information, but in order to segment the population some restructuring is required. For example, the age information is presented by individual years and it will have to be grouped in decades for segmenting. Moreover, the average age for each neighborhood is calculated, as the information is presented as number of individuals for each age.

Family Income Available is an index that represents the clean income a family have after taxes, to either spend or save. In this case the latest information published by neighborhood dates back to 2017. The index for Barcelona is 100, with a reference value of 22.390 € per person. [7] The information about the neighborhoods is presented in relation to the Barcelona index, so the value in euros has to be calculated from the reference value.

For the political scenario, two different approaches can be taken: grouping by left-center-right wings, as of they define themselves, or grouping by Catalan nationalists or Spanish nationalists. For the clustering, the first division is used.

Finally, we have two datasets that contain information about commercial venues. From the Foursquare API we will need to calculate the more repeated venues per neighborhood and select the top 10 for each one in descending order. From the city council source, two data tables are presented, one with absolute values and another with percentual. The second one will be used.

Once the different segmentations with each single parameter are completed, a new table with the neighborhoods and the corresponding cluster index per segmentation is created and a correlation matrix generated to detect new data relationships.

3 Bibliography

- [1] "Barcelona city council. Statistics by neighborhood: Nationality.," 2019. [Online]. Available:
<https://www.bcn.cat/estadistica/castella/dades/barris/tpob/pad/ine/a2019/ine17.htm>.
- [2] "Barcelona city council. Statistics by neighborhood: Commercial venues.," 2019. [Online]. Available:
<https://www.bcn.cat/estadistica/castella/dades/barris/economia/tacteco/gacbarri19.htm>.
- [3] "Barcelona city council. Statistics by neighborhood: Unemployment.," 2019. [Online]. Available:
<https://www.bcn.cat/estadistica/castella/dades/barris/ttreball/atur/durada/durbar19.htm>.
- [4] "Barcelona city council. Statistics by neighborhood: Elections results.," 2019. [Online]. Available:
<https://www.bcn.cat/estadistica/castella/dades/barris/elec/loc/a2019.htm>.
- [5] "Barcelona city council. Statistics by neighborhood: Age.," 2019. [Online]. Available:
<https://www.bcn.cat/estadistica/castella/dades/barris/tpob/pad/ine/a2019/ine02.htm>.
- [6] "Barcelona city council. Statistics by neighborhood: Family income available.," 2019. [Online]. Available:
<https://www.bcn.cat/estadistica/castella/dades/barris/economia/renda/rdfamiliar/a2017.htm>.
- [7] "Diputació de Barcelona. Hermes project.," 2017. [Online]. Available:
https://www.diba.cat/hg2/presentacio.asp?prId=888&idioma=cat&codi_any=2017&format=pantalla.
- [8] "Beautiful Soup Documentation.," [Online]. Available:
<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>.
- [9] "Panda DataFrame.," [Online]. Available: <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html>.
- [10] "Foresquare developer.," [Online]. Available: <https://developer.foursquare.com/>.