David López Companys

FINDING THE OPTIMAL LOCATION FOR A NEW STORE

Content:

1. Problem description

In this project, the problem attempted to solve will be to find the best possible location or the most optimal, for a latin supermarket in the city of Barcelona (Catalunya, Spain).

To achieve this task, an analytical approach will be used, techniques used will be focused on clustering to have a group of optimal neighborhoods where index of immigration and rental costs fit the supermarket needs and capabilities.

During the process of analysis, several data transformations will be performed, in order the find the best possible data format for the machine learning model to ingest. Once the data is set up and prepared, a modeling process will be carried out, and this statistical analysis will provide the best possible places to open this new supermarket.

2. Data presentation

The data used for this analytical process will come from dispare open data sources, which are:

1. <u>Foursquare API</u>: This data will be accessed via Python and used to obtain the most common venues per neighborhood in the city of Barcelona. This way, it is possible to have a taste of how the city's venues are distributed, what are the most common places for leisure, and in general, it will provide an idea of what people's likes are.

2. <u>% of immigration by nationality and neighbourhood</u>: This information is provided as an open data source by the Barcelona's city hall, the information provided includes:
   a. Year
   b. Month
   c. Neighbourhood
   d. Country / Region of immigrant
   e. Number of immigrants

   This information can be downloaded as csv file in the following url: https://opendata-ajuntament.barcelona.cat/data/es/dataset/est-demo-immigrants-lloc-procedencia

3. <u>Average rental price per neighbourhood:</u> This information is also provided by the Barcelona's city hall as statistical data, the information contained in this dataset is as follows:
   a. Neighbourhood
   b. Average price per sqm / per year

   This dataset can be acceced via the following url: https://www.bcn.cat/estadistica/angles/dades/timm/ipreus/hab2mave/evo/t2mab.htm

This three data sources will be used to cluster the different neighbourhoods of barcelona in different groups depending on the:
- %immigration by precedence living in every single neighbourhood and venue.
- rental price

With this information, first the optimal number of clusters will be found via elbow method (or others, to be defined which fits better in development), after the optimal number of clusters is found, different clustering methods will be applied:
- Partitioning clustering: K-means
- Hierarchical clustering: Bottom up

With that, venues will be clustered in how they perform similar in the different variables analysed, having that information and optimal cluster will be chosen to place the new supermarket depending on budget available and immigration.