

Correlation between Malaga and Toronto Neighborhoods

Juan Antonio Garcia Fernandez

Coursera | IBM Data Science Capstone Project | January 2020

Contents

[**Introduction** 2](#_Toc29662309)

[**Problem description** 2](#_Toc29662310)

[**Target audience** 2](#_Toc29662311)

[**Success criteria** 2](#_Toc29662312)

[**Data Description** 3](#_Toc29662313)

[Data sources 3](#_Toc29662314)

[**Methodology** 4](#_Toc29662315)

[**Results** 4](#_Toc29662316)

[**Discussion** 4](#_Toc29662317)

[**Conclusion** 4](#_Toc29662318)

# **Introduction**

After centuries of technological progress and advances in international cooperation, the world is more connected than ever. **Globalization** is the word used to describe the growing interdependence of the world’s economies, cultures, and populations, brought about by cross-border trade in goods and services, technology, and flows of investment, people, and information.

**“Flows of people”** means that people can now travel more often in less time, but also **live in many different countries**. This is not always easy because the new country can be a great unknown to that level at the beginning. When you move from one city to another in a distant country, you would like to find a place, a neighborhood, relatively similar to the one you were living in. **People usually like to keep their living standards**.

Therefore, the idea of this project comes from the effort required for finding a neighborhood to live after moving to another city in a different country. Specifically, this report is focused on **people moving from Toronto (Canada) to Malaga (Spain)** in order to illustrate the process for developing a **Machine Learning model to identify the best possible neighborhood to move in to.**

## **Problem description**

The goal of this project is to explore the neighborhoods of Malaga and Toronto in order to identify similarities and correlations between them. In short, the question to be answered is: **Given a neighborhood X in the city of Toronto, which neighborhoods in Malaga have the same kind of venues?**

## **Target audience**

The target audience of this project are:

* **Real estate agencies in Málaga or Toronto** that are paid by people who are moving from Malaga to Canada (or vice versa) to search for a proper accommodation, either sale or rental, in a convenient neighborhood similar to the one they are coming from.
* **International companies** hiring people from another country (Spain/Canada) need to offer to their new employees a place to stay. Preferably, a place where they feel comfortable and at home.
* **Families** that want to live an adventure abroad, but also maintain an environment close to their lifestyle.

## **Success criteria**

The success criteria of this project will be a good recommendation of similar neighborhoods, i.e., same kind of venues, leisure opportunities, green areas and so on, to people moving from Malaga to Canada (or vice versa).

# **Data Description**

Malaga and Toronto neighborhoods are the main scope of the project. We need to explore, segment and cluster them based on the types and amounts of venues offered by each of them. Accordingly, we will need the following data:

* **List of neighborhoods in Malaga and Toronto.** Note that Malaga is a medium city in the South of Spain, while Toronto is a big city located on Lake Ontario in Canada.
* **Coordinates of the neighborhoods.** This information is needed to determine the available venues in each neighborhood, and also to be able to plot the neighborhoods together with the clustering outcome provided by the algorithm on a map.
* **Venue data.** This is key to feed the clustering algorithm that aims to process the data and find clusters of neighborhoods, if they exist in the data.

## **Data sources**

* For the **Toronto neighborhood data**, a **Wikipedia** page exists that has all the information we need to explore and cluster the neighborhoods in Toronto (<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>). On the other hand, for the **Malaga neighborhood data**, there is an **OpenStreetMap** page with the same kind of information that we need to explore and cluster the neighborhoods in Malaga (<https://wiki.openstreetmap.org/wiki/ES:Lugares_en_M%C3%A1laga>). We will scrape these webpages, wrangle the data, clean it, and then read it into a pandas dataframe so that it is in a structured format. Once the data is in a structured format, we will analyze the dataset.
* **Geopy** is an excellent Python library for (among others) geocoding and reverse geocoding that supports many APIs. In this project, we will use the **Nominatim** API, which is based on OpenStreetMap (OSM) data. Nominatim allows us **to convert an address into latitude and longitude values**. In this manner, we will get the coordinates of the neighborhoods.
* **FourSquare API** is an API to interact with the Foursquare platform. Foursquare is a social location service that allows users to explore the world around them. One of the available API methods **returns the surrounding venues given a specific location** (<https://developer.foursquare.com/docs/api/venues/search>).
* **Folium** is a Python library for producing maps with Leaflet.js. With this, we will be able to create maps where to visualize the clusters provided by the Machine Learning model.

# **Methodology**

Rrr

# **Results**

Dd

# **Discussion**

Adffsdf

# **Conclusion**

aefesgd