Capstone Project – IBM Professional Certificate in Data Science – Coursera

Montréal's Neighborhoods Analysis – Where to open a wellness facility?

By

Antoine Morin

January 29th, 2021

INTRODUCTION

Montréal is the second-most populous city in Canada with a population of almost 2 million inhabitants. It goes without saying that the city offers a lot of opportunities for whoever wants to invest in some sort of business, either small or big. Montréal is home to all kind of venues, from restaurants to cafés, museums, artisanal shops, etc. There is now a clear tendency for people to take care of their health, and Montréal has seen in the past years a steady growth in the number of wellness facilities, for example gyms and yoga studios. As this trend is likely to continue to grow, we are interested in taking a share in this business by opening a wellness facility in Montréal. To have better chances of success, we want to know where the best place is to open this facility. More precisely, in which neighborhood should we invest in our business? To answer this question, we will use the power of data along with machine learning to get crucial insights about the business ecosystem of the city of Montréal.

DATA REQUIEREMENTS

To solve our problem, we will first need to get data about the neighborhoods of Montréal. To identify the neighborhoods, we will use the postal codes of the city, available at https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_H. We will also need the latitude and longitude of each neighborhood to later get the venues with the FourSquare API. Because Nominatim cannot provide us with these latitudes and longitudes using only postal codes and because the Google Maps API needs a paid subscription, we will get manually these latitudes and longitudes using Google Maps. Finally, we need to get the venues in each neighborhood, based on the latitudes and longitudes and a specified radius. To accomplish this step, we will use the FourSquare API. The best radius can be determined with the latitudes and longitudes of each neighborhood by calculating the mean distance between each closest neighborhood. See the Methodology section for more information.

Regarding the venues that we will get through the FourSquare API, we are mainly interested in the number of venues in each category. We will need to define broad categories in order to extract meaningful relationships between the venues. In particular, we want to see correlations between the “Wellness” venues category and other categories. The broad categories that we will use are: restaurants, arts, wellness, fast food, artisanal food, bars, cafés, public area, and essentials.