

The Battle of Neighborhood- Best location to open GYM/Fitness Centre in Toronto

Laxmikantha Herle

Introduction

This project aims to utilize all Data Science Concepts learned in the Data Science Professional Course. We define a Business Problem, the data that will be utilized and using that data, we are able to analyze it using Machine Learning tools. In this project, we will go through all the processes in a step by step manner from problem designing, data preparation to final analysis and finally will provide a conclusion that can be leveraged by the business stakeholders to make their decisions.

Table of Contents

- Background & Problem Description
- Data Description
- Methodology
- Machine Learning
- Data Analysis
- Discussion and Conclusion
- References

Background & Problem Description

Gone are the days when gyms targeted only a small segment of the population. Nowadays, gyms are attracting more members. Members can choose a low-budget gym and still get a good sweat in or they could opt for a smaller, more specialized boutique studio to meet their specific fitness goals. Plus, there are 24-hour fitness Centre popping up nationwide to make going to the gym is a lot more convenient for today's consumers. Revenue for the Gym, Health and Fitness Clubs industry in Canada has grown over the five years to 2020 as a result of consumer trends and the proliferation of public health campaigns. With an increasing rate of adult obesity expected during the period, the Public Health Agency of Canada (PHAC) has stressed adherence to fitness and healthy lifestyle choices. According to the PHAC and the Canadian Institute for Health Information, obesity is expected to incur more than \$5.0 billion in costs each year, which must be covered by the healthcare system.

The objective of this project is to determine "what might be the 'best' neighborhood in Toronto to open a GYM/Fitness Centre". Will use foursquare location data and regional clustering of venue information to determine the 'best' neighborhood in Toronto to open a GYM/Fitness Centre. We will find the most suitable location for an entrepreneur to open a new GYM/Fitness Centre in Toronto, Canada.

Target Audience

Information provided by this report would be useful for People who wants open GYM/Fitness Centre in Toronto, Canada. The Objective is to locate and recommend to People which neighborhood of Toronto will be the best choice to open GYM/Fitness Centre.

Data Description

To consider the objective stated above, we can list the below data sources used for the analysis

- I. Toronto Neighborhood Data: The following Wikipedia page was scraped to pull out the necessary information: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
The information obtained i.e. the table of postal codes was transformed into a pandas data frame for further analysis.
- II. Coordinate data for each Neighborhood in Toronto: The following csv file gave us the geographical coordinates of each postal code: http://cocl.us/Geospatial_data
- III. Venue Data in Toronto, Canada. Geographical Coordinates data will be utilized as input for the Foursquare API that will be leveraged to provision venues information for each neighborhood. We will use Foursquare API to explore neighborhood in Toronto, Canada.

Methodology

We used the *BeautifulSoup* package to transform the data in the table on the Wikipedia page into the pandas data frame. After all the data was collected and put into data frames, cleansing and merging of the data was required to start the process of analysis. When getting the data from Wikipedia, there were Boroughs that were not assigned to any neighborhood therefore, the following assumptions were made,

- Only the cells that have an assigned borough will be processed.
- Borough's that were not assigned get ignored.
- More than one neighborhood can exist in one postal code area. For example, in the table on the Wikipedia page, will notice that M5A is listed twice and has two neighborhood: Harbourfront and Regent Park. These two rows will be combined into one row with the neighborhood separated with a comma
- If a cell has a borough but a not assigned neighborhood, then the neighborhood will be the same as the borough.

After the implementation of the following assumptions, the rows were grouped based on the borough as shown below Fig.1

| | Postal Code | Borough | Neighborhood |
|---|-------------|------------------|---|
| 0 | M3A | North York | Parkwoods |
| 1 | M4A | North York | Victoria Village |
| 2 | M5A | Downtown Toronto | Regent Park, Harbourfront |
| 3 | M6A | North York | Lawrence Manor, Lawrence Heights |
| 4 | M7A | Downtown Toronto | Queen's Park, Ontario Provincial Government |

Fig.1 Toronto Neighborhood Data

Load & Explore Geo Coordinate Data

The second source of data provided with the Geographical coordinates of the neighborhood with the respective Postal Codes. The file was in CSV format, so we had to attach it to a Pandas data frame, we merged the two tables together based on Postal Code as shown below Fig.2.

| | Postal Code | Borough | Neighborhood | Latitude | Longitude |
|---|-------------|------------------|---|-----------|------------|
| 0 | M3A | North York | Parkwoods | 43.753259 | -79.329656 |
| 1 | M4A | North York | Victoria Village | 43.725882 | -79.315572 |
| 2 | M5A | Downtown Toronto | Regent Park, Harbourfront | 43.654260 | -79.360636 |
| 3 | M6A | North York | Lawrence Manor, Lawrence Heights | 43.718518 | -79.464763 |
| 4 | M7A | Downtown Toronto | Queen's Park, Ontario Provincial Government | 43.662301 | -79.389494 |

Fig.2 Toronto Neighborhood Data with Coordinates