APPLIED DATA SCIENCE CAPSTONE

Final Report

Justin Yu

Introduction

The purpose of this report is to showcase the application of data science in the business world. More specifically, this report aims to illustrate how advanced data analytics create value to the business sector by conducting in depth analysis and generating actionable recommendations. Through understanding the implication of the analysis and the recommendation, stakeholders can make a better business decision.

Business Problem

The business problem that the analysis aims to solve is where in the Greater Toronto Area is the ideal location to open a new gym / fitness center. The trend of living a healthy lifestyle has transformed the daily lives of many, and the popularity of working out is growing exponentially. In specific, the percentage of people going to the gym has increased significantly in recent years in Ontario. As a result, there are many gyms in the Greater Toronto Area now and the intense competition is a headache to many gym owners. Therefore, this analysis will help gym owners to identify neighborhoods in the Greater Toronto Area that has a low prevalence of gyms, so as to open up a new gym in a profitable location.

Target Audience

The target audience of the analysis are gym owners. Gym owners include owners of gym chains, owners of boutique gyms, and potential gym owners who are looking for an opportunity to start a gym. Among all the factors to be considered, the location of the gym is a critical element contributing to the success of the gym. As gym memberships are relatively long, entering a neighborhood with less competition significantly enhances the customer loyalty. On the other hand, enter a saturated neighborhood will likely lead to the failure of an investment. In essence, the analysis can guide the (potential) gym owners to strategically open a new gym at a specific location that has compelling prospects.

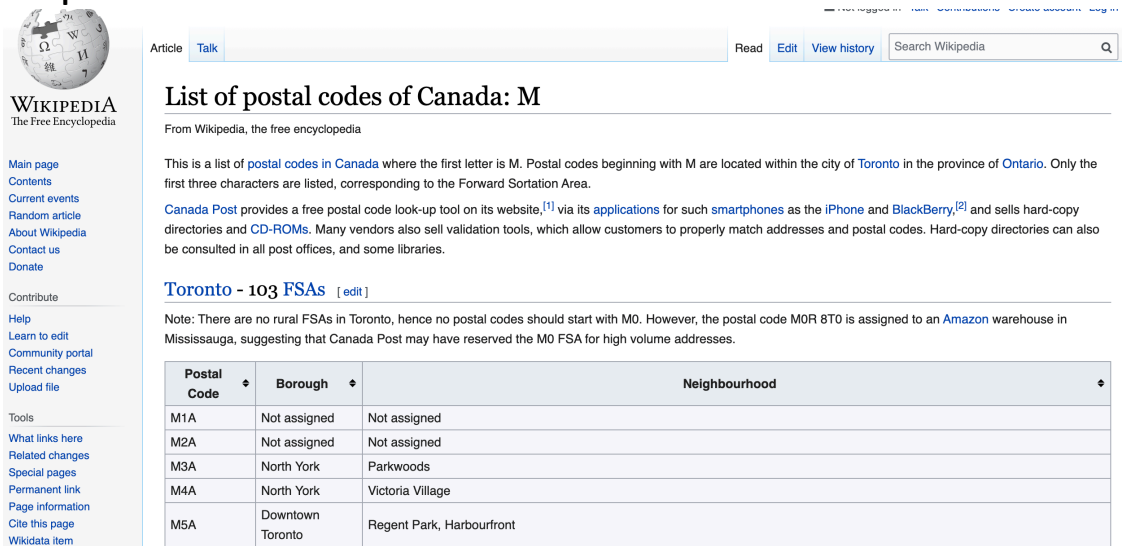
Data

The analysis will require two sets of data. The first set of data includes geographical information, latitude and longitude, of all the neighborhoods in the Greater Toronto Area. The second set of data is the Foursquare location data, which includes the venues of the neighborhoods.

Neighborhoods in Toronto

The first set of the data is obtained by utilizing Wikipedia and Geospatial data from a csv file provided, which are illustrated as follows:

Wikipedia :



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Cite this page
Wikidata item

Article Talk | Read Edit View history | Search Wikipedia

List of postal codes of Canada: M

From Wikipedia, the free encyclopedia

This is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. Only the first three characters are listed, corresponding to the Forward Sortation Area.

Canada Post provides a free postal code look-up tool on its website,[1] via its applications for such smartphones as the iPhone and BlackBerry,[2] and sells hard-copy directories and CD-ROMs. Many vendors also sell validation tools, which allow customers to properly match addresses and postal codes. Hard-copy directories can also be consulted in all post offices, and some libraries.

Toronto - 103 FSAs [edit]

Note: There are no rural FSAs in Toronto, hence no postal codes should start with M0. However, the postal code M0R 8T0 is assigned to an Amazon warehouse in Mississauga, suggesting that Canada Post may have reserved the M0 FSA for high volume addresses.

Postal Code	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront

Source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

CSV File:

Postal Code	Latitude	Longitude
M1B	43.806686	-79.194353
M1C	43.784535	-79.160497
M1E	43.763573	-79.188711
M1G	43.770992	-79.216917
M1H	43.773136	-79.239476
M1J	43.744734	-79.239476
M1K	43.727929	-79.262029
M1L	43.711112	-79.284577
M1M	43.716316	-79.239476
M1N	43.692657	-79.264848
M1P	43.757410	-79.273304

Source: http://cocl.us/Geospatial_data

Venues in Different Neighborhoods

Foursquare location data is leveraged in this project in order to obtain all the information required to conduct the analysis. In specific, the scope covers the venues in different neighborhoods in order to better understand the neighborhood, and the data is illustrated as follows:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Parkwoods	43.753259	-79.329656	Allwyn's Bakery	43.759840	-79.324719	Caribbean Restaurant
Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
Parkwoods	43.753259	-79.329656	Tim Hortons	43.760668	-79.326368	Café
Parkwoods	43.753259	-79.329656	A&W	43.760643	-79.326865	Fast Food Restaurant
Parkwoods	43.753259	-79.329656	Bruno's valu-mart	43.746143	-79.324630	Grocery Store

Data Explanation

Combining all three sources of data, a comprehensive dataset can be created. In particular, the combined dataset will include all the venues in different neighborhoods in Toronto, including geospatial data, the name of the venue, and the category of the venue. Hence, the dataset can facilitate the process of identifying the most ideal location to start a gym, which is the purpose of the analysis.