



A. Introduction

This project aims to utilize all Data Science Concepts learned in the IBM Data Science Professional Course. I start by defining a Business Problem, then discuss the data that will be utilized. Using that data, I will be able to analyze it using Machine Learning tools. In this project, I will go through all the processes in a step by step manner from problem designing, data preparation to final analysis.

Table of Contents

1. Introduction
2. Target Audience
3. Data Overview

4. Methodology

5. Discussion

6. Conclusion

1. Identifying the Business Problem (Introduction):

Food is a big deal in Toronto, a very big deal. If you are a foodie, Toronto should be a “must-visit” destination on your bucket list.

Toronto is one of the most densely populated areas in Canada. It is located in the province of Ontario, Canada and it is also the capital of Ontario. It has a recorded population of over 6 million people. It is not only the most populous city in Canada, it is also the fourth most populous city in North America. Toronto is an international center of business, finance, arts, and culture, and I would add food. It is recognized as one of the most multicultural and cosmopolitan cities in the world. It is without a doubt a foodie destination.

For over 200 years, Immigrants from South America, Asia, Africa, the Caribbean and Europe have been bringing their culinary cultures to Toronto. Over time, the mixing and blending of these cultures give birth to food that can only be found where cultures harmoniously reside. One such example is Rasta Pasta restaurant, which is collision of Italian and Jamaican flavors.

There are also traditional Canadian foods that are not just cheese, curd and gravy, e.g Pow Wow Café which features the First People of Canada cuisine such as Ojibwe-style fry bread etc. There is also Antler Kitchen and Bar that features seasonal menus of authentic Canadian game such as bison, venison, duck and wild-caught halibut etc.

Multiculturalism is seen through the various neighborhoods including; Chinatown, Corso Italia, Little India, Kensington Market, Little Italy, Koreatown and many more.



2. Target Audience


This project is aimed towards Entrepreneurs or Business owners, who may want to understand the food climate in Toronto and explore their options of investing in the food business in Toronto. The analysis will provide vital information that can be used by the target audience.

3. Data Overview

The data required will be a combination of CSV files that have been prepared for the purposes of the analysis from multiple sources. These files will provide the list of neighborhoods in Toronto (via Wikipedia), the Geographical location of the neighborhoods (via Geocoder package) and Venue data pertaining to restaurants (via Foursquare).

3.1 — Data acquisition:

Source 1: Toronto Neighborhoods via Wikipedia



WIKIPEDIA
The Free Encyclopedia

Main page

Contents

Featured content

Current events

Random article

Donate to Wikipedia

Wikipedia store

Interaction

Help

About Wikipedia

Community portal

Recent changes

Contact page

Tools

What links here

Related changes

Upload file

Special pages

Permanent link

Page information

Wikidata item

Cite this page

Print/export

Create a book

Download as PDF

Printable version

Article

Talk

Read

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, 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 v 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

Postal Code	Borough	Neighborhood
M1A	Not assigned	
M2A	Not assigned	
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront
M6A	North York	Lawrence Manor, Lawrence Heights
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
M8A	Not assigned	
M9A	Etobicoke	Islington Avenue
M1B	Scarborough	Malvern, Rouge
M2B	Not assigned	
M3B	North York	Don Mills
M4B	East York	Parkview Hill, Woodbine Gardens

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

The Wikipedia site shown above provided almost all the information about the neighborhoods. It included the postal code, borough and the name of the neighborhoods present in Toronto. Since the data is not in a format that is suitable for analysis, scraping of the data was done from this site (shown in *figure2*).

	PostalCode	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

Figure 2: Data that was scraped from Wikipedia site and put into Pandas data frame

Source 2: Geographical Location data using Geocoder Package

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Fig 3 Conversion of the file into Pandas data frame

2. https://cocl.us/Geospatial_data

The second source of data provided us with the Geographical coordinates of the neighborhoods with the respective Postal Codes. The file was in CSV format, so we had to attach it to a Pandas data frame (shown in figure 3).

Source 3: Venue Data using Foursquare

4. Methodology

4.1 — Data Cleansing

After all the data was collected, they were put into data frames. I then cleansed and merged the data so it can be analyzed. In the data retrieved from Wikipedia, there were Boroughs that were not assigned to any neighborhood therefore, the following assumptions were made:

1. Only the cells that have an assigned borough will be processed.

2. Multiple neighborhoods can share the same postal code, e.g. For example, M5A is listed twice and has two neighborhoods: Harbourfront and Regent Park. These two rows were combined into one row with the neighborhoods separated with a comma.

3. Some borough were not assigned a 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.

	Postcode	Borough	Neighbourhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Fig. 4 Grouped together by Borough

Using the Latitude and Longitude collected from the Geocoder package, we merged the two tables together based on Postal Code.

	PostalCode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Fig 5 Merged tables based on Postal Code

After, the venue data pulled from the Foursquare API was merged with the table above providing us with the local venue within a 500-meter radius.

4.2 — Data Exploration

Now after cleansing the data, I analyzed it. I then created a map using Folium and color-coded each Neighborhood depending on what Borough it was located.

Next, I used the Foursquare API to get a list of all the Venues in Toronto which included Parks, Schools, Café Shops, Asian Restaurants etc. This data was crucial to analyze the number and different types Restaurants in Toronto.

This data will be used for the machine learning, k-means clustering and Data Analysis portion of this project.

