Where to make Sushi

Levllyn Rocha

15th May 2020

# Introduction

## Background

Asian cuisine is much sought after in Toronto. Since the city is huge location there could be a lot of areas that could be a possible location of a new restaurant. In the restaurant business location matters a lot and since historically the suburbs are not an ideal location this study focusses on a part of the city area.

There are many ways to make a decision on location. This study aims at using cluster analysis on venue / location data to narrow down our choices. Note that the study aims to only broadly define possible locations and further narrowing down would depend on a lot more factors not considered here

## Problem

A budding restauranteur is all set to take his idea of a Japanese restaurant to fruition. This analysis focuses specifically on West Toronto for the optimum location to set up a Japanese Restaurant based on the areas most frequented for eating and locations of existing restaurants.

## Interest

This analysis would be beneficial to budding restaurateurs who have the same use-case of opening a Japanese Restaurants. The logic used could of course be expanded or tweaked to fit the appropriate needs such as other cuisines, other locations or even other business models should the right data be available.

# Data

## Data Sources

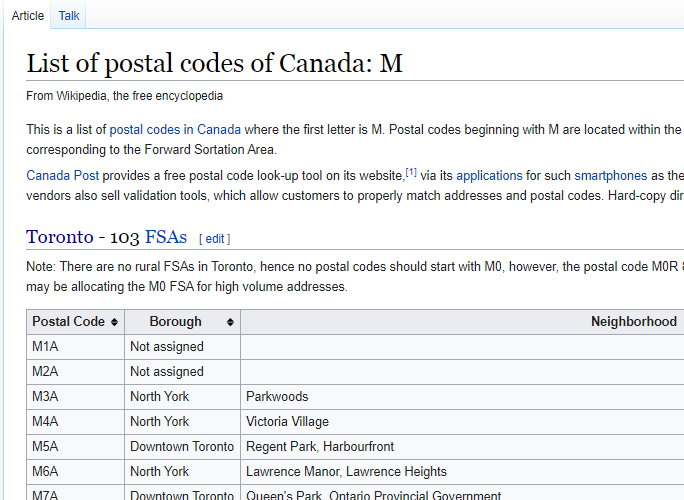
There are 4 primary sources of data both publically available related to

* + The neighborhoods and boroughs in Toronto. Their location and postal code is a necessity.
  + Coordinate data for Each Postal code
  + Details of the most visited places in Toronto where we will be using data from FourSquare API
  + Geopy to get Coordinates where required

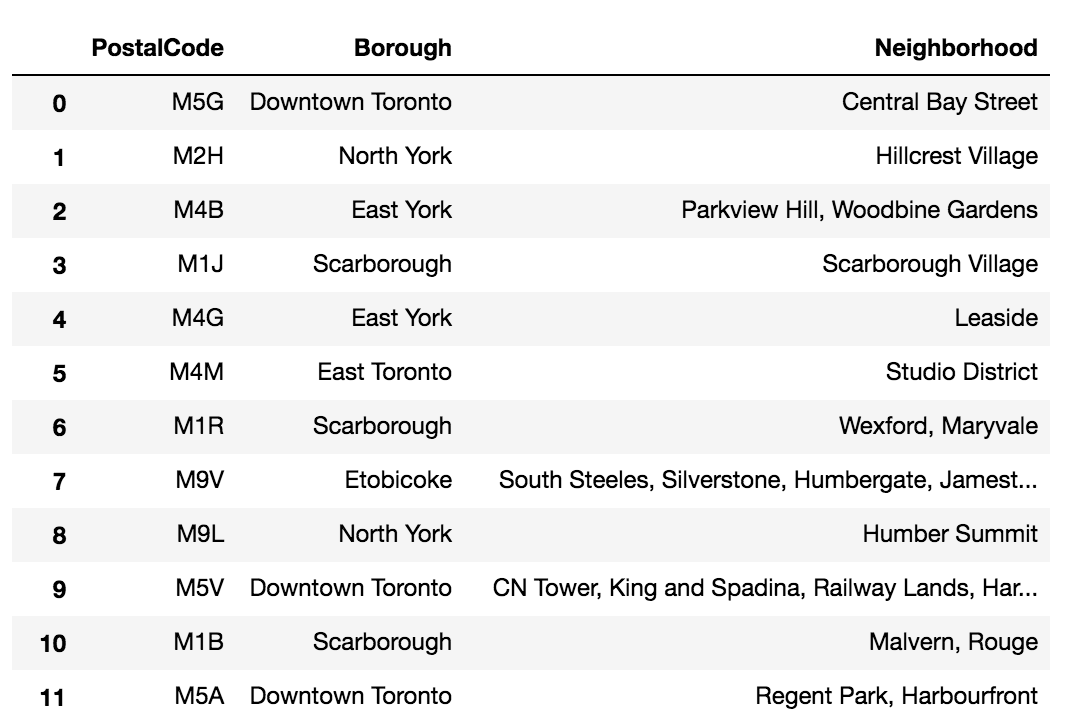
1. Postal Areas with Neighborhoods & Buroughs

We will be using data from Wikipedia on Postal codes beginning with M in Canada which covers all area in the main city of Toronto

[https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M,](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) will be used in order to obtain the data that is in the table of postal codes



To transform the data into a *pandas*dataframe like the one shown below:



While there are different website scraping libraries and packages in Python. For scraping the above table, we will be using pandas to read the table into a pandas dataframe.

1. Coordinate data for postal codes is got from the below location:

<http://cocl.us/Geospatial_data>

This is stored into a CSV to obtain and subseauently added to the data fra

1. Venue Data from Foursquare

Foursquare is a technology company that built a massive dataset of location data. They actually crowd-sourced their data and had people use their app to build their dataset and add venues and complete any missing information they had in their dataset. Currently its location data is the most comprehensive out there, and quite accurate that it powers location data for many popular services like Apple Maps, Uber, Snapchat, Twitter and many others

Calls to the API return structured data like below:



1. JSON file from Geopy to get coordinate data



## Data Cleaning & Selection

Basic cleansing steps like analyzing and removal of NULLS is required.

Data from Postal code is taken as the base 🡪 Coordinate data is added to this from COCL.US site 🡪 This is used to pull venue data from each location.

Since we are focusing on West Toronto, the CN Tower is taken as border all coordinate to the West of CN Tower are retained.

One-hot encoding is done from the venues obtained from Foursquare which is the base of the analysis.

The data is changed into frequency and modeled so we get the top 10 venues for each location after the require grouping is done.