The Battle of Neighborhoods

Exploring Toronto for Italian Restaurant

Introduction:

As a country of immigrants, when families came to Canada, they typically brought with them a taste of home. These traditional recipes merged with Canadian customs and ingredients to create cuisines that can now be found in all manner of restaurants in Canada **.** The multiculturalism is seen through the various neighborhoods including; Chinatown, Corso Italia, Little India, Kensington Market, Little Italy, Koreatown and many more. Downtown Toronto being the hub of interactions between ethnicities, brings many opportunities for entrepreneurs to start or grow their business. It is a place where people can try the best of each culture, either while they work or just passing through. Toronto is well known for its great food.

The **objective of this project** is to use Foursquare location data and regional clustering of venue information to determine what might be the 'best' neighborhood in Toronto to open a restaurant. Pizza and Pasta are one of the most bought dishes in Toronto originating from Italy. Toronto being the fourth largest home to Italians with a population over 500k, there are numerous opportunities to open a new Italian restaurant. Through this project we will find the most suitable location for an entrepreneur to open a new Italian restaurant in Toronto, Canada

Target Audience:
• Entrepreneurs who want to open an Italian Restaurant or pizzeria in Toronto and are looking for a suitable neighbourhood.

Data Overview:

The data set required for the following project was acquired from three different data sources.

The three data sources are listed below,

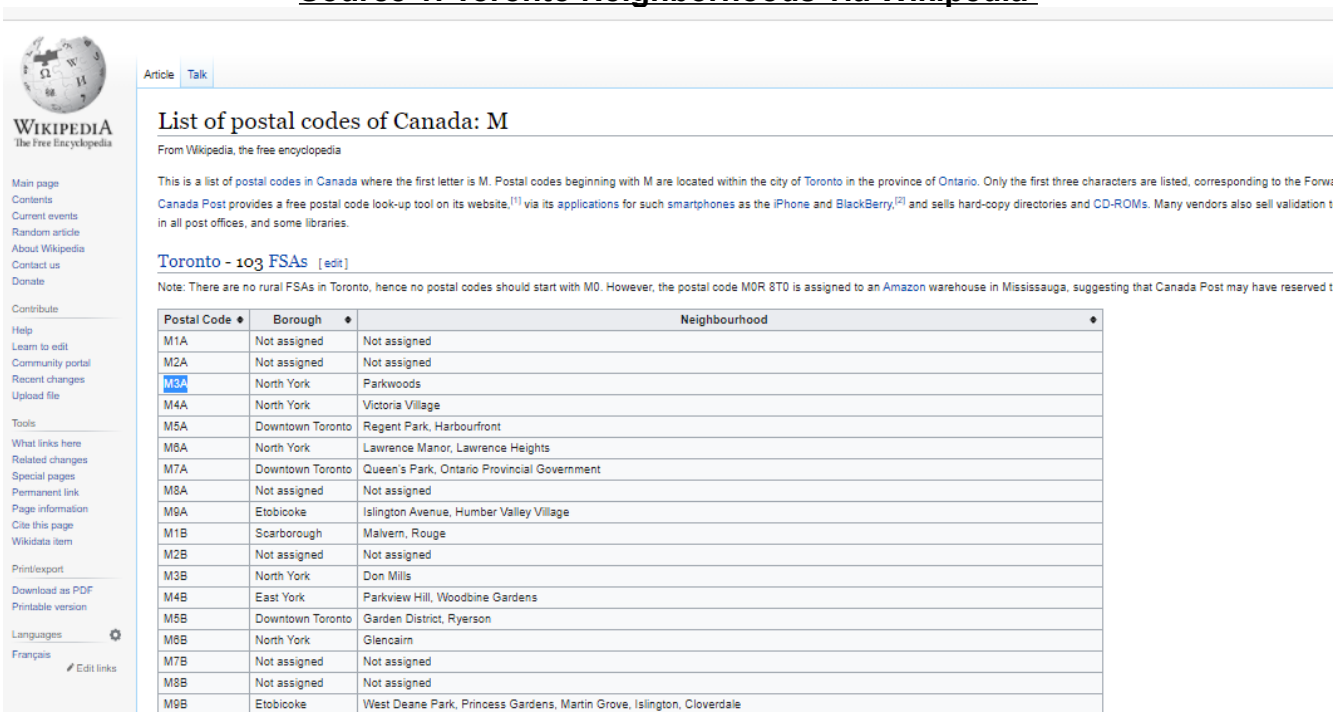
1. A Wikipedia Page to fetch boroughs and neighborhoods of Toronto city.
2. A csv file to fetch latitudes and longitudes corresponding to each postal code.
3. The foursquare api to fetch different public venues in the vicinity of the neighborhood.

The data that will be required will be a combination of CSV files that have been prepared for the purposes of the analysis from multiple sources which will provide the list of neighborhoods in Toronto (via Wikipedia), the Geographical location of the neighborhoods (via Geocoder package) and Venue data pertaining to Italian restaurants (via Foursquare). The Venue data will help find which neighborhood is best suitable to open an Italian restaurant.

Methodology:

First, we will need to extract the data from the data sources:

Source 1: Toronto Neighborhoods via Wikipedia



WIKIPEDIA
The Free Encyclopedia

Article Talk

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 Forwa
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 t
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 t

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
M6A	North York	Lawrence Manor, Lawrence Heights
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
M8A	Not assigned	Not assigned
M9A	Etobicoke	Islington Avenue, Humber Valley Village
M1B	Scarborough	Malvern, Rouge
M2B	Not assigned	Not assigned
M3B	North York	Don Mills
M4B	East York	Parkview Hill, Woodbine Gardens
M5B	Downtown Toronto	Garden District, Ryerson
M6B	North York	Glencairn
M7B	Not assigned	Not assigned
M8B	Not assigned	Not assigned
M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grove, Islington, Cloverdale

Figure 1:Wikipedia Page showing List of Neighborhoods in Toronto with respective Postal Codes

The Wikipedia site (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) 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	Neighbourhood
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

Source2: Geographical Location data using Geocoder Package

	A	B	C
1	Postal Code	Latitude	Longitude
2	M1B	43.8066863	-79.1943534
3	M1C	43.7845351	-79.1604971
4	M1E	43.7635726	-79.1887115
5	M1G	43.7709921	-79.2169174
6	M1H	43.773136	-79.2394761
7	M1J	43.7447342	-79.2394761

Figure 3: Geographical data of Neighborhoods in Toronto

The second source of data provided (https://cocl.us/Geospatial_data) us with the Geographical coordinates of the neighborhoods with the respective Postal Codes which were converted into data frame

Source3: Venue Data using Foursquare

The retrieval of the location, name and category about the various venues in Toronto was collected through the Foursquare explore API. To obtain the data, it was required to make an account where it would provide a 'Secret Key' as well as 'Client ID' which will allow me to pull any data

Data Pre-processing

Below are the steps performed

- The first step I performed was to scrape data from the Wikipedia page that consisted of all the boroughs and neighborhoods along with their postal codes and convert them into data frame so that we can do analysis using visualization techniques.
- Dropped rows having missing values in the data frame as missing values can cause discrepancy
- Importing data from a Geospatial_Coordinates.csv file. The .csv file consisted of latitude and longitude coordinates of each postal code. This .csv file was imported into a data frame for ease of analysis in the later stage. Using the Latitude and Longitude collected from the Geocoder package, we merged the two tables together based on Postal Code.

```
In [19]: # Merging the Data
         df = pd.merge(df, geo_df, on='PostalCode')
         df.head()
```

Out[19]:

	PostalCode	Borough	Neighbourhood	Latitude	Longitude
0	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790
1	M4P	Central Toronto	Davisville North	43.712751	-79.390197
2	M4R	Central Toronto	North Toronto West, Lawrence Park	43.715383	-79.405678
3	M4S	Central Toronto	Davisville	43.704324	-79.388790
4	M4T	Central Toronto	Moore Park, Summerhill East	43.689574	-79.383160

Figure 4: Merging tables together 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 shown below.

```
#Get venues for all neighborhoods in our dataset
toronto_venues = getNearbyVenues(names=df_toronto['Neighbourhood'],
                                 latitudes=df_toronto['Latitude'],
                                 longitudes=df_toronto['Longitude'])
```

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Lawrence Park	43.728020	-79.388790	Lawrence Park Ravine	43.726963	-79.394382	Park
1	Lawrence Park	43.728020	-79.388790	Zodiac Swim School	43.728532	-79.382860	Swim School
2	Lawrence Park	43.728020	-79.388790	TTC Bus #162 - Lawrence-Donway	43.728026	-79.382805	Bus Line
3	Davisville North	43.712751	-79.390197	Homeway Restaurant & Brunch	43.712641	-79.391557	Breakfast Spot
4	Davisville North	43.712751	-79.390197	Sherwood Park	43.716551	-79.387776	Park

Figure 5: Local Venues near the respective Neighborhood

Now after cleansing the data, the next step was to analyze the data which will be presented in our next weeks report.