Capstone Project - The Battle of Neighborhoods (Week 4&5) Project: A Restaurant Recommendation System in North York, Canada

Part 1 - Introduction

Business Problem Statement

North York is an administrative division in Toronto, Ontario, Canada. It is located directly north of Old Toronto, between Etobicoke to the west and Scarborough to the east. As of the 2011 Census, it had a population of 655,913 [1] . The neighborhoods of North York are highly diverse, inhabited by people of many different cultures. The North York neighbourhood with the largest percentage of immigrants in is the Bathurst – Steeles area of Westminster – Branson, where 73% of its population were not born in Canada. Based on the cultural diversity, a variety of cuisines could be easily found in North York.

This project is aimed at providing recommendations to those people who are interested opening restaurants in North York, such as where would be the most suitable locations to start the business, and which kinds of cuisine should be considered as the first options. The recommendation system is built on Foursquare location, venue information, geographical data, neighborhood data in North York from the internet.

Data Overview

The dataset used in this project is mainly from Wikipedia page in tabular format, including postcodes and neighborhoods information, as well as the geographical information (longitude and latitude data) from the Geospatial_Coordinates.csv file. Besides that, thanks to the information provided by Foursqaure which helps to solve the main problem by giving detailed venue information.

First of all, the data from Wikipedia page in tabular format should be combined with Geospatial_Coordinates.csv file according to the same postcodes, and then we need to filter all the neighborhoods needed in this project. After we get the combined dataset, we need to acquire venue information from Foursquare dataset based on the geographical data. Finally, this project relies heavily on the venue information provided by Foursquare, including the venue name, venue category, summary, and etc.
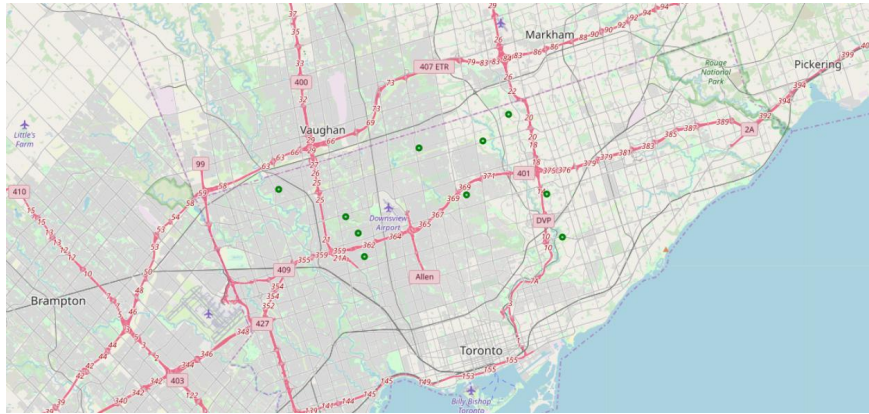
Methodology

1. Download and transpose the data into python pandas dataframe

2. Clean the data such as dropping data and grouping data

3. Combine the neighborhood data with the geographical (longitude and latitude) data based on the same postcode

4. Import folium library and create a map add markers to show all neighborhoods in North York

5. Use Foursquare client id and secret to create a new dataset contains venue information in all neighborhoods in North York

6. Select all the restaurants information in the venue category as required

7. One hot coding to see the number of restaurants in each type

8. Use k-means clustering to cluster the neighborhoods based on the number of restaurants and connect the cluster group level with neighborhood names

9. Create a map and add markers to these neighborhoods with lower cluster group

10. Explore top 10 venues(restaurants) in these neighborhoods with less restaurants and check the cuisine information

Results

1. After exploring venue information in North York, it is discovered that less restaurants could be found in these neighborhoods: Bayview Village, Downsview Central, Downsview West, Downsview, Hillcrest Village, Humber Summit, Parkwoods, Victoria Village, Willowdale West, and York Mills West as shown in the map below.

2. Top 10 venues (restaurants) in each neighborhood have been explored to clarify the cuisine issue. For example, in Bayview Village, most popular restaurants are Asian restaurants (Chinese & Japan), fast-food restaurants, and European restaurants (Eastern European & Middle European).

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Bayview Village	Japanese Restaurant	Café	Chinese Restaurant	Fast Food Restaurant	Wings Joint	Eastern European Restaurant	Greek Restaurant	Fried Chicken Joint	French Restaurant	Food Court

Discussion

1. Based on the result that found in this project, it is suggested that the neighborhoods mentioned above are more suitable for investors to open restaurants because there are fewer counterparts in these areas

2. The most popular cuisines in each neighborhood are different, for instance, opening a restaurant specialized in Asian cuisines or fast food might be the first option in Hillcrest Village.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
12	Hillcrest Village	Korean Restaurant	Fast Food Restaurant	Pizza Place	Chinese Restaurant	Sandwich Place	Wings Joint	Dim Sum Restaurant	Fried Chicken Joint	French Restaurant

Bayview Village: Asian cuisines (Chinese & Japanese) are the most popular, followed by fast food and European cuisines

Downsview Central & Downsview West: Southeast Asian cuisines are the most popular, followed by fast food and European cuisines

Downsview: Chinese cuisines are the most popular, followed by Mediterranean cuisines, and fast food

Hillcrest Village: Asian cuisines (Korean and Chinese cuisines), followed by fast food

Humber Summit: Italian cuisines are the most popular, followed by fast food

Parkwoods & Victoria Village & York Mills West: Fast food is the most popular, followed by European cuisines

Willowdale West: European cuisines are the most popular, followed by fast food

References:

1. https://en.wikipedia.org/wiki/North_York

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