Vegan Franchise

SL – Capstone Project – Vegan Franchising

June 2020

Table of Contents

1. INTRODUCTION 3

2. DATA 6

3. METHODOLOGY AND MODEL 11

4. RESULTS AND FINDINGS 14

Introduction

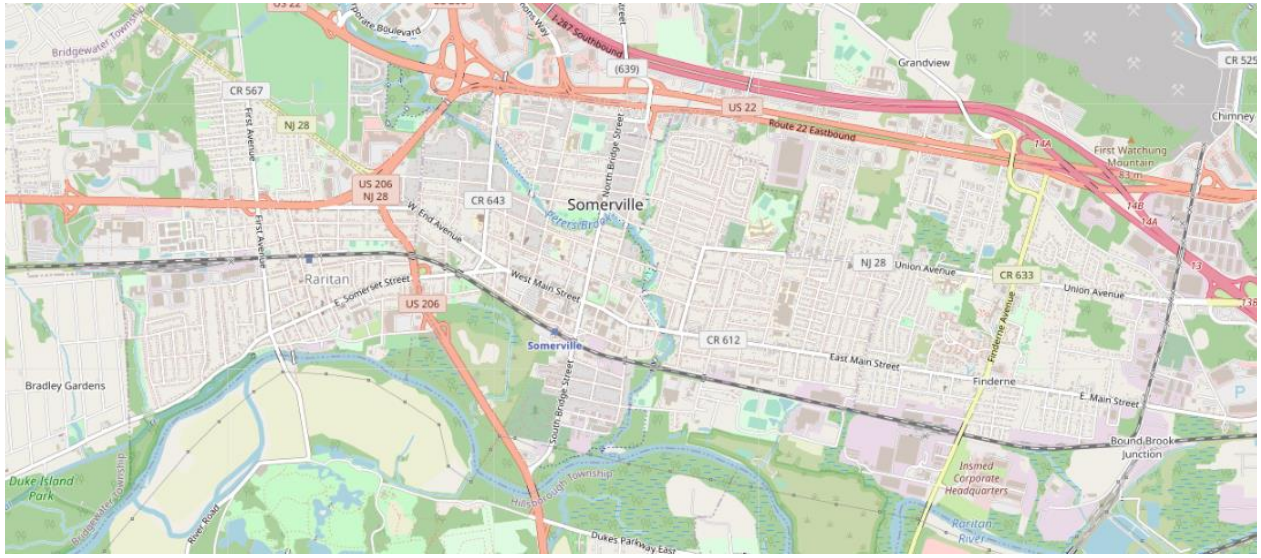
Introduction

Background

- It is widely understood that cities have their own individual culture and vibe. However, it is important to note that individual neighborhoods have their own sub-cultures.
- The specific attributes of the neighborhood may have more of an impact on the success of a business than the impact of the city itself. This is one of the key themes of this paper.

Somerville

- Somerville is an example of a unique neighborhood in New York.
- New York Times recently published an article highlighting the benefits and status of the neighborhood, which has been under the radar for a long time. https://www.nytimes.com/2019/08/14/realestate/somerville-nj-a-walkable-suburban-alternative.html



Introduction

Fresh Tiki

- Fresh Tiki, a vegan franchise.
- Mr. Xi operates a famous vegan restaurant in Somerville, Fresh Tiki. This restaurant has received excellent reviews from food critics and the company is planning to open in numerous locations globally.
- Recently, Mr. Xi was approached by Ms.Varga, a prominent Toronto businesswoman, to open five locations of Fresh Tiki in Toronto. Mr. Xi is amenable to opening such locations on a franchise basis, but wants to provide his consent regarding locations.

The Problem

- Mr. Xi is a firm believer in Somerville and he believes that the eclectic nature of the neighborhood, with its high walkability and urban / suburban feel, have been strong contributors to the success of Fresh Tiki. Therefore, one of his requirements is that Fresh Tiki be opened in similar locations in Toronto.
- Mr. Xi wants to shortlist up to 10 locations to serve as the initial neighborhoods for opening Fresh Tiki in Toronto. The requirements for this shortlist are that the selected locations 'must be similar to Summerville'.
- **The Assignment:** He has engaged our data science firm to assist with this real world problem.

The Data

The Data

Three key sources of data

- **Data 1: Toronto neighborhood data**
  - Data on 103 neighborhoods across multiple boroughs.
  - Source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
  - Sample shown below (after clean-up).

	Borough	Neighborhood	Latitude	Longitude
0	North York	Parkwoods	43.753259	-79.329656
1	North York	Victoria Village	43.725882	-79.315572
2	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

The Data

Three key sources of data (continued)

- **Data 2: New York neighborhood data**
  - Data on 306 neighborhoods across multiple boroughs.
  - Source: https://cocl.us/new_york_dataset
  - Sample shown below (after clean-up).

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

The Data

Three key sources of data (continued)

- **Data 3: Data from the Foursquare API**
  - Prior to applying the venue data from Foursquare (using the explore function), we combined the dataframes from Toronto and New York in a single data frame.
  - Then we collected data from Foursquare for the types of venues located within a 500 meter radius of the neighborhood.
  - This captured 12,089 venues across 395 individual neighborhoods in New York and Toronto.
  - There are 463 unique categories of data.
  - Sample shown below (after clean-up).

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Agincourt	5	5	5	5	5	5
Alderwood, Long Branch	9	9	9	9	9	9
Allerton	26	26	26	26	26	26
Annadale	14	14	14	14	14	14
Arden Heights	4	4	4	4	4	4
Arlington	8	8	8	8	8	8
Arrochar	18	18	18	18	18	18
Arverne	19	19	19	19	19	19
Astoria	100	100	100	100	100	100
Astoria Heights	15	15	15	15	15	15

The Data

Grouping and sorting of data so that it was ready for the model

- **We grouped and sorted our data so that it was ready for our model.**

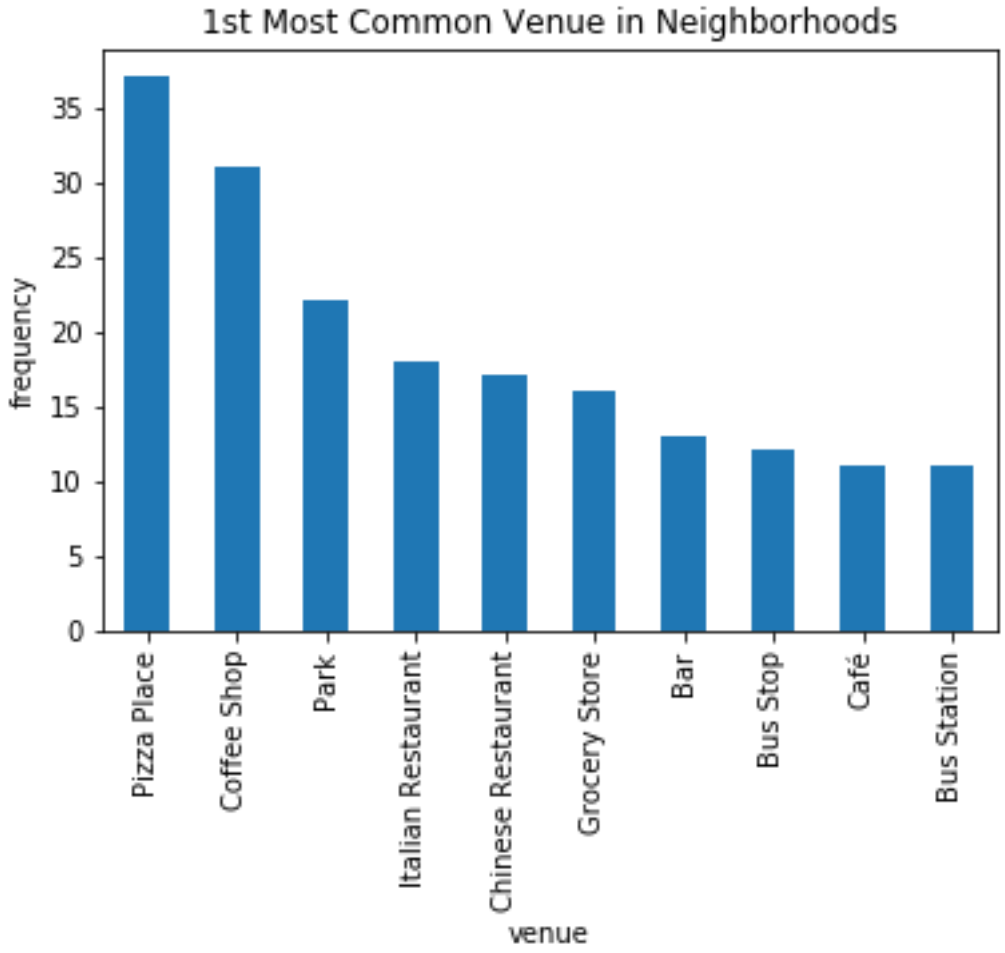
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Lounge	Breakfast Spot	Clothing Store	Latin American Restaurant	Skating Rink	Factory	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant
1	Alderwood, Long Branch	Pizza Place	Pub	Dance Studio	Athletics & Sports	Pharmacy	Sandwich Place	Coffee Shop	Gym	Filipino Restaurant	Entertainment Service
2	Allerton	Pizza Place	Supermarket	Deli / Bodega	Breakfast Spot	Dessert Shop	Bakery	Martial Arts Dojo	Grocery Store	Chinese Restaurant	Donut Shop
3	Annadale	Pizza Place	Food	American Restaurant	Diner	Pharmacy	Deli / Bodega	Restaurant	Dance Studio	Sports Bar	Park
4	Arden Heights	Bus Stop	Pharmacy	Coffee Shop	Pizza Place	Falafel Restaurant	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant

Methodology and Model

Methodology and Model

Pre-model analysis

- Once our data set is ready for further analysis, we completed some preliminary data analysis through running some basic statistics regarding features of the data. This included assessing the most frequent location types in the neighborhoods as well as the most concentrated neighborhoods.
    - Our data set consisted of 12,089 different venues in 395 individual neighborhoods.
    - There are, in fact, 463 unique categories of venues within those neighborhoods.
    - We are able to group and sort data so that we can assess the most frequent venues in the individual neighborhoods.
    - The most frequent venues in Neighborhoods are shown in the chart below.



Methodology and Model

Model implementation

- In order to solve our problem, we will use the K-means Clustering Algorithm. This is an unsupervised machine learning approach, especially useful when we have unlabeled data. The idea is to cluster our neighborhoods into various groups, without providing any guidance related to what is similar about these neighborhoods. A key input into this clustering approach is the need to specify the number of cetroids, i.e. K. We used K=5, i.e. to cluster neighborhoods in New York and Toronto in five clusters.

- Once the clusters were properly formed, we:
    - Assessed the individual clusters to assess composition and similarities.
    - Compared neighborhoods within clusters as well as across clusters.

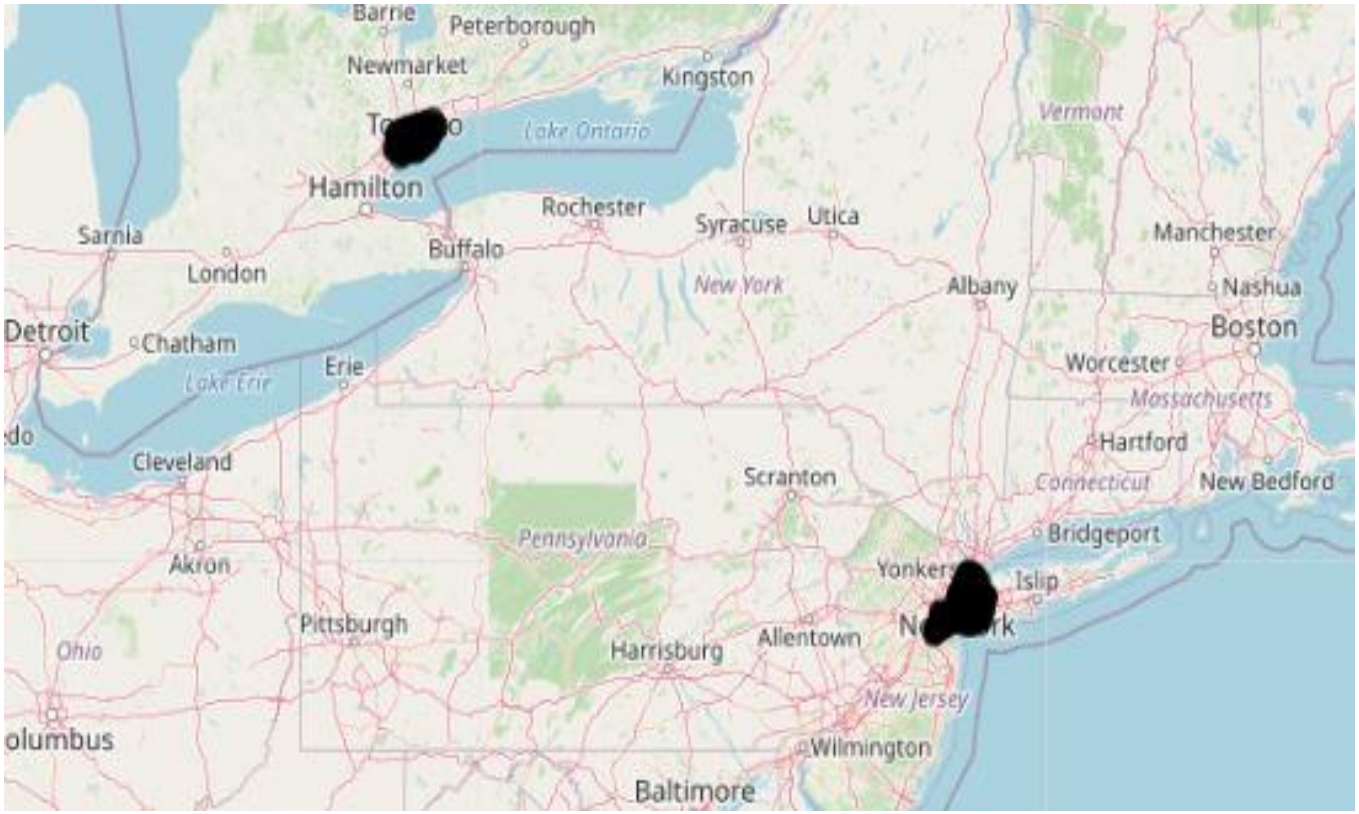
- The focus of the analysis was to assess the cluster consisting of Somerville, a neighborhood in New York to find similar neighborhoods in Toronto. We were successful in finding several neighborhoods in Toronto, which were in the same cluster.

Results and Findings

Results and Findings

Results

- The 395 neighborhoods fell within five clusters.
- Some clusters were far more concentrated than other clusters.
- One cluster was purely focused on Toronto.
- Other clusters were purely New York clusters.



Results and Findings

Results

- Important cluster information

	Number of Neighborhoods	Key features
Cluster 1	12	Closeness to parks, fitness studios, farms, farmers' markets, women's stores and ethnic restaurants
Cluster 2	300	Emphasis on cafes, fast food restaurants, bars, sushi restaurants, etc.
Cluster 3	2	Baseball field centric clusters in Toronto
Cluster 4	86	Clusters with a strong emphasis on pizza places, pubs and cafes.
Cluster 5	4	Beach / pier centric neighborhoods in New York.

Results and Findings

Results

- Our cluster of interest is Cluster 1, the cluster of Somerville.

		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
Toronto	0	Parkwoods	Food & Drink Shop	Park	Fireworks Store	Falafel Restaurant	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service
	21	Caledonia-Fairbanks	Park	Women's Store	Pool	Farm	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service
	35	East Toronto, Broadview North (Old East York)	Convenience Store	Metro Station	Park	Intersection	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service
	61	Lawrence Park	Bus Line	Swim School	Park	Flower Shop	Flea Market	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant
	64	Weston	Park	Convenience Store	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant
	66	York Mills West	Convenience Store	Park	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant
	83	Moore Park, Summerhill East	Gym	Park	Women's Store	Falafel Restaurant	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service
	85	Milliken, Agincourt North, Steeles East, L'Amo...	Park	Playground	Women's Store	Farm	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service
	98	The Kingsway, Montgomery Road, Old Mill North	Park	River	Women's Store	Farm	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service
New York	27	Clason Point	Park	Convenience Store	Pool	South American Restaurant	Bus Stop	Boat or Ferry	Grocery Store	Egyptian Restaurant	Eastern European Restaurant	Electronics Store
	192	Somerville	Park	Women's Store	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant
	203	Todt Hill	Park	Trail	Women's Store	Falafel Restaurant	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service

Results and Findings

Findings

- This cluster of interest, the “Somerville” cluster (Cluster 1) seems to validate Mr. Xi’s assertion that Somerville is a fairly unique neighborhood in New York. There are only three other neighborhoods in New York with similar characteristics / vibe.
- This cluster can be characterized as:
    - Being close to parks, gyms, women's stores.
    - Having a high concentration of Mediterranean, Eastern European and empanada restaurants.
    - Open spaces.
    - Unique areas, e.g. flea markets, flowers shops, farms.
- **There are 9 locations similar to Somerville neighborhood in Toronto.**

Results and Findings

Shortlist

- As Fresh Tiki considers opening various locations in Toronto, these are the 9 possible neighborhoods that provide the same feel. These neighborhoods include:
    - Parkwoods
    - Caledonia-Fairbanks
    - East Toronto, Broadview North
    - Lawrence Park
    - Weston
    - York Mills West
    - Moore Park, Summerhill East
    - Mililikwn, Agincourt North, Steeles East
    - The Kingsway, Montgomery Road, Old Mill North

Thank You