

# *CAPSTONE*

# *PROJECT*

Battle of the neighbourhoods

## Table of Contents

<b>Introduction .....</b>	<b>3</b>
Selection criteria.....	3
<b>Data .....</b>	<b>5</b>
Approach .....	5
<b>Methodology .....</b>	<b>6</b>
Data acquisition and cleansing .....	6
Data preparation .....	7
Feature selection .....	7
Clustering.....	7
<b>Results .....</b>	<b>8</b>
<b>Discussion .....</b>	<b>12</b>
<b>Results .....</b>	<b>12</b>

## Introduction

Although immigration has become a hot topic over the past few years with more governments seeking more restrictions on immigrants and refugees, the general trend of immigration into Canada has been one of on the rise. With more people entering the country from different parts of the world comes more diversity of culture, values and shifts in social norms. Increased population also brings opportunity from a business perspective. A restaurant with a cuisine deemed foreign some decades ago may have been less popular since little was known about it and its origins. However, over time, with more diversity and knowledge sharing and understanding, what was once foreign is now less unfamiliar and as a result, such a restaurant in this day and age may not only be accepted but thrive.

The aim of this project is to be able to explore the possibilities of where a restaurant or food vendor with an emerging cuisine type could be setup and be embraced in Toronto and hopefully thrive. A major benefit of pursuing this type of project is that if it is successful, it could help future restaurateurs to determine where next to focus their efforts in bringing a not-so-mainstream cuisine to the population of Toronto.

## Selection criteria

For the purposes of this project, the definition of an emerging cuisine, is one that has an appreciable commercial presence within a given community but is not:

1. Associated with a geographic location
2. Associated with a major fast food chain

Toronto, like many other global cities, has areas where specific cuisines are dominant ([https://en.wikipedia.org/wiki/Cuisine\\_in\\_Toronto](https://en.wikipedia.org/wiki/Cuisine_in_Toronto)):

- Eglinton West - Caribbean food
- Roncesvalles - Polish cuisine
- Chinatown - Chinese and Vietnamese food
- Kensington Market - Latin American and others
- Little Italy and Corso Italia - Italian
- Gerrard India Bazaar and Rexdale - Indian, Punjabi, Pakistani
- Agincourt - Chinese, Korean, Japanese, Vietnamese, South Indian, Sri Lankan
- Koreatown - Korean
- Little Portugal/Rua Acores - Portuguese
- The Danforth - Greek
- The Annex – Hungarian

As such, since these cuisines are synonymous with location, they are assumed to be established cuisines and thus fail condition 1 of the selection criteria.

The following were identified as fast food chains in Canada  
[\(\[https://en.wikipedia.org/wiki/List\\\_of\\\_fast-food\\\_chains\\\_in\\\_Canada\]\(https://en.wikipedia.org/wiki/List\_of\_fast-food\_chains\_in\_Canada\)\)](https://en.wikipedia.org/wiki/List_of_fast-food_chains_in_Canada):

A&W	Dairy Queen	KFC	Robin's Donuts
Arby's	Dic Ann's Hamburgers	Krispy Kreme	Subway
Baker's Dozen Donuts	Dixie Lee Fried Chicken	Lafleur Restaurants	Swiss Chalet
La Belle Province	Domino's	Lick's Homeburgers	Taco Bell
Blimpie	Donut Diner	Manchu Wok	Taco del Mar
Booster Juice	Edo Japan	Mary Brown's	Taco Time
Burger Baron	Extreme Pita	McDonald's Canada	Thai Express
Burger King	Fast Eddies	Mr. Sub	Tim Hortons
Captain Submarine	Five Guys	New York Fries	Valentine
Carl's Jr.	Fryer's	Orange Julius	Wendy's
Chez Ashton	Goji's	Papa John's	White Spot
Chicken Delight	Greco Pizza	Pita Pit	Williams Fresh Cafe
Cinnabon	Harvey's	Pizza Hut	Yogen Früz
Coffee Time	Hero Certified Burgers	Popeyes Louisiana Kitchen	
Country Style	Jimmy the Greek	Quiznos	

As such, cuisines associated with these chains will also be excluded from the emerging cuisines category.

Lastly, for a cuisine to be emerging, it needs to be already present within the community as opposed to completely new which would make the cuisine a pioneering one. If no restaurant/food vendor for a specific cuisine exists e.g. Nauru ethinc food, such a cuisine is disqualified.

Following all the selection criteria, the choice for this project was chosen to be cuisines of **Ethiopian** origin.

## Data

This project will require knowledge of the different neighbourhoods in Toronto and the cuisines they do or do not possess. As such the neighbourhood data required will be:

- Neighbourhood location in terms of latitude and longitude
- Restaurants/food vendor within the neighbourhood. For each reach restaurant/food vendor in the neighbourhood, the following data will also be required:
  - Whether the cuisine is emerging or not

Neighbourhood locations can be retrieved by pooling resources from Wikipedia ([https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)) which possesses a list of neighbourhoods in the Toronto area and use of a geo locator to determine the latitude and longitude of each area.

Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, tips relating to these venues, menus and even photos. As such, the Foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

## Approach

The intention is to explore similarities between neighbourhoods based on the restaurants/food vendors present. From this, it may then be possible to ascertain whether the emerging cuisine in one neighbourhood could be established in another neighbourhood. It might also be possible to ascertain if the emerging food could increase its presence in a neighbourhood where it already exists.

## Methodology

Steps taken were:

1. Data acquisition and cleansing
2. Data preparation
3. Feature selection
4. Clustering

### Data acquisition and cleansing

Data acquisition was a 2-step process:

1. Obtaining the postcodes for neighbourhoods in Toronto
2. Obtaining venues within these neighbourhoods

Wikipedia provided a page containing post codes of neighbourhoods in Toronto. In order to utilise this information, web scraping was employed. Some values needed to be removed from the extracted data due to insufficient information. The remaining postcodes were then run through a geocoder in order to obtain the corresponding longitudes and latitudes.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

- Neighbourhood
- Neighbourhood Latitude
- Neighbourhood Longitude
- Venue
  - Name of the venue e.g. the name of a store or restaurant
- Venue Latitude
- Venue Longitude
- Venue Category
  - Group assigned to the venue by Foursquare e.g. Pharmacy X might be in category 'Drug Store.'
- Icon Prefix
  - Icon used on the website to help identify the category of the venue

## Data preparation

Having successfully retrieved the data, it was necessary to filter out everything except for the restaurants/food vendors. In order to achieve this, the 'Icon Prefix' was used as it consistently showed food information regardless of the venue category.

Venue	Venue Category	Icon Prefix
Brookbanks Park	Park	<a href="https://.../img/categories_v2/parks_o...">https://.../img/categories_v2/parks_o...</a>
KFC	Fast Food Restaurant	<a href="https://.../img/categories_v2/food/fa...">https://.../img/categories_v2/food/fa...</a>
Variety Store	Food & Drink Shop	<a href="https://.../img/categories_v2/shops/f...">https://.../img/categories_v2/shops/f...</a>
Victoria Village Arena	Hockey Arena	<a href="https://.../img/categories_v2/arts_en...">https://.../img/categories_v2/arts_en...</a>
Tim Hortons	Coffee Shop	<a href="https://.../img/categories_v2/food/co...">https://.../img/categories_v2/food/co...</a>
Portugril	Portuguese Restaurant	<a href="https://.../img/categories_v2/food/po...">https://.../img/categories_v2/food/po...</a>

Looking at the table above showing some sample results, it can be seen that rows 2,5 and 6 are specifically food sellers but there is a lot of variation in the Venue Category. However, the 'Icon Prefix' column contains the '/food/' path for all three and thus proves to be a more identifier of restaurants/food vendors

With food venues (restaurants/food vendors) extracted from the main data body, the next task involved categorising food venues into emerging or non-emerging cuisine groups. For the purpose of this project, only food venues serving food of Ethiopian origin/inspiration were considered to be in the merging category.

## Feature selection

This was achieved through the use of correlation between the emerging cuisine food venues and all other venues. Pearson correlation was used. From this, the top 5 food venues with the strongest negative correlations and the 5 top venues with the strongest positive correlation were selected as the features to retain (10 features in total)

## Clustering

Using the selected features, K-mean clustering was performed. The number of clusters chosen was 3:

1. A cluster to capture neighbourhoods likely to embrace a new emerging cuisine (Ethiopian) food venue
2. A cluster to capture neighbourhoods unlikely to embrace a new emerging cuisine food venue
3. A cluster to capture neighbourhoods neither likely or unlikely to embrace a new emerging cuisine food venue

## Results

From the process of data retrieval and cleaning, the following postcodes were obtained for the Toronto area. Each was then investigated to determine the venues it possessed.

M3A	M5E	M3K	M3N	M5S	M7Y
M4A	M6E	M4K	M4N	M6S	M8Y
M5A	M1G	M5K	M5N	M1T	M8Z
M6A	M4G	M6K	M6N	M4T	
M7A	M5G	M1L	M9N	M5T	
M9A	M6G	M2L	M1P	M1V	
M1B	M1H	M3L	M2P	M4V	
M3B	M2H	M4L	M4P	M5V	
M4B	M3H	M5L	M5P	M8V	
M5B	M4H	M6L	M6P	M9V	
M6B	M5H	M9L	M9P	M1W	
M9B	M6H	M1M	M1R	M4W	
M1C	M1J	M2M	M2R	M5W	
M3C	M2J	M3M	M4R	M8W	
M4C	M3J	M4M	M5R	M9W	
M5C	M4J	M5M	M6R	M1X	
M6C	M5J	M6M	M7R	M4X	
M9C	M6J	M9M	M9R	M5X	
M1E	M1K	M1N	M1S	M8X	
M4E	M2K	M2N	M4S	M4Y	

Following the retrieval of venues from Foursquare and filtering for restaurant/food vendors using the 'Icon Prefix' column, it was found that there were 1298 food venues throughout Toronto within 500 metres of a neighbourhood postcode.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0 Parkwoods	43.753259	-79.329656	KFC	43.754387	-79.333021	Fast Food Restaurant
1 Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
2 Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
3 Regent Park, Harbourfront	43.654260	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
4 Regent Park, Harbourfront	43.654260	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop

```
toronto_food_venues.shape
```

```
(1298, 7)
```

Investigation into identifying Ethiopian food venues resulted in the discovery that only 2 areas had these venues and each venue had only 1 Ethiopian food venues:

1. Garden District, Ryerson
2. Church and Wellesley

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Emerging Cuisine
Garden District, Ryerson	43.657162	-79.378937	Ethiopiques	43.656513	-79.377078	Ethiopian Restaurant	1
Church and Wellesley	43.665860	-79.383160	Ethiopian House	43.666599	-79.385669	Ethiopian Restaurant	1

Correlation of food venue categories identified the following as possessing the strongest negative and positive correlations against the Ethiopian food venues:

```
# Top 5 negatively correlated categories with Ethiopian Restaurant
top5_neg_corr = list(toronto_emerging_corr[0:5].index)
toronto_emerging_corr[0:5]
```

```
Bakery           -0.088593
Fried Chicken Joint -0.059880
Brewery          -0.055152
French Restaurant -0.053179
New American Restaurant -0.051988
Name: Ethiopian Restaurant, dtype: float64
```

```
# Top 5 negatively correlated categories with Ethiopian Restaurant
top5_pos_corr = list(toronto_emerging_corr[-5:].index)
toronto_emerging_corr[-5:]
```

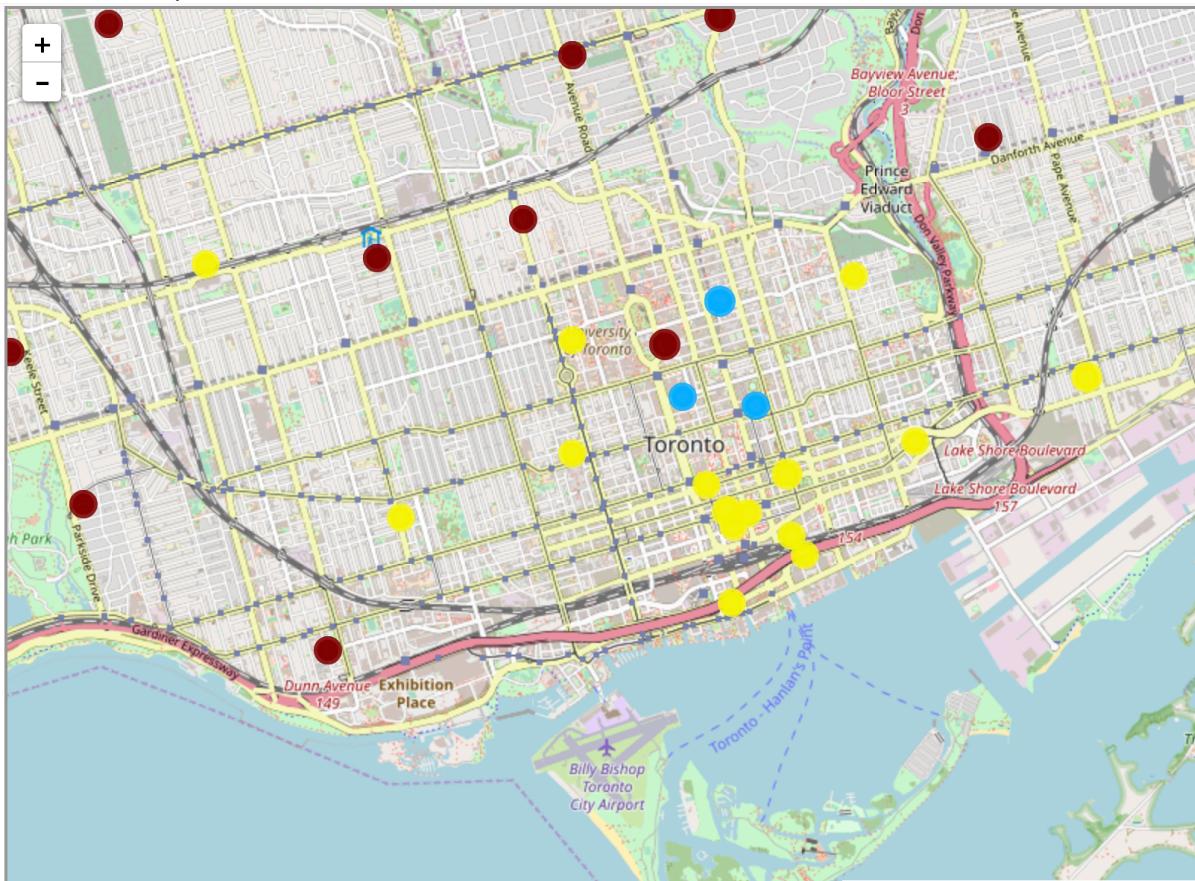
```
Juice Bar        0.562731
Bubble Tea Shop  0.574324
Japanese Restaurant 0.644599
Afghan Restaurant 0.702728
Ethiopian Restaurant 1.000000
Name: Ethiopian Restaurant, dtype: float64
```

These food venue categories would be the features selected for the clustering process, all other food venue categories were excluded.

Clustering into the following 3 categories yielded the following:

Cluster ID	Description	Cluster population
0	A cluster to capture neighbourhoods likely to embrace a new emerging cuisine (Ethiopian) food venue	63
1	A cluster to capture neighbourhoods unlikely to embrace a new emerging cuisine food venue	3
2	A cluster to capture neighbourhoods neither likely or unlikely to embrace a new emerging cuisine food venue	16

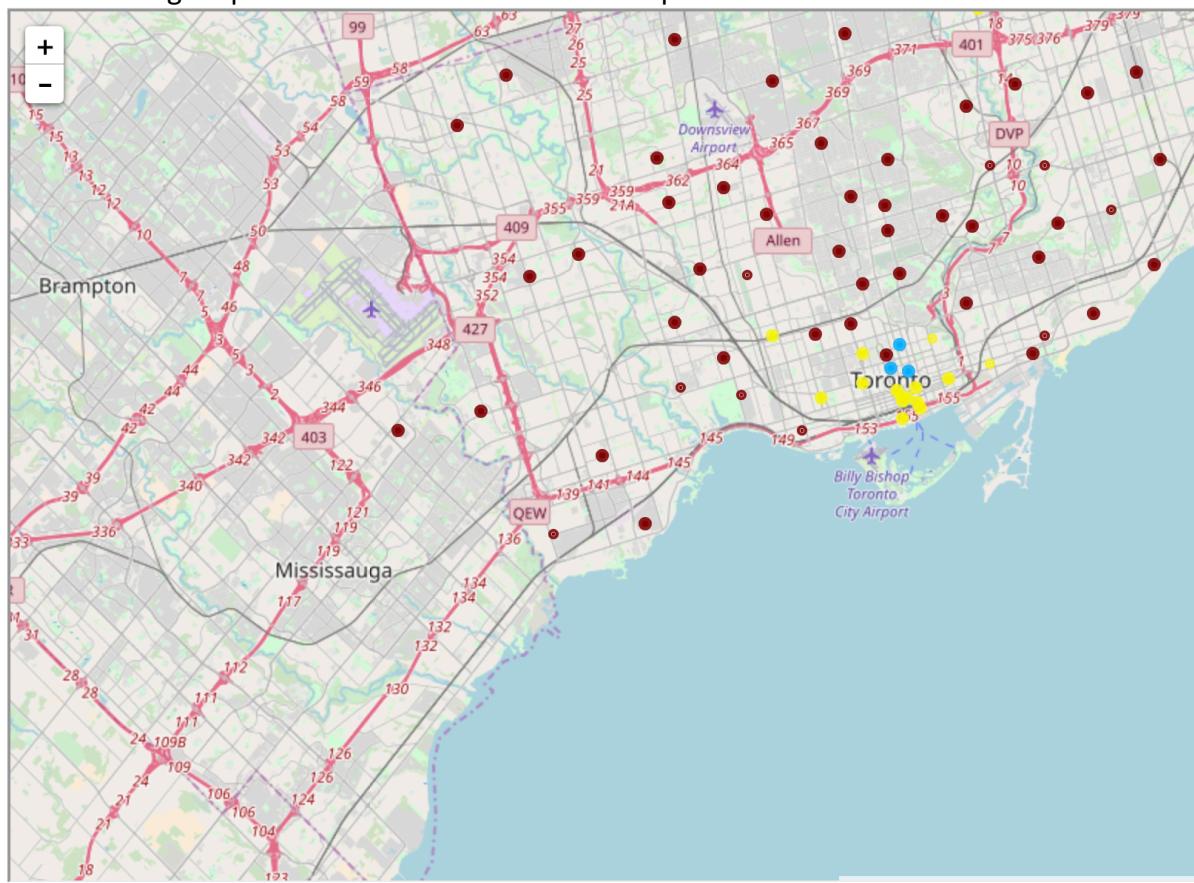
The following map is a zoomed in view of a subset of the clusters. Red is cluster 0, blue is cluster 1 and yellow is cluster 2



The cluster group of primary interest, cluster 1 corresponded to the following neighbourhoods:

1. Garden District, Ryerson
2. Church and Wellesley
3. Central Bay Street

The following map is a zoomed out view of the map:



## Discussion

From the final result, it can be seen that the clustering performed reasonably well. It was able to assign neighbourhoods already containing Ethiopian food vendors (Garden District, Ryerson and Church and Wellesley) into the same cluster (cluster 1) and identify an additional neighbourhood with similar characteristics to that cluster (Central Bay Street). Based on the food venue characteristics of these areas, the clustering process would suggest that these would be likely to embrace a new Ethiopian food vendor in the area.

It is interesting to note that the presence of certain food vendors in certain proportions can act as possible indicators of the likeliness of an area to adopt an Ethiopian restaurant. These food vendor categories were found to be:

- Ethiopian Restaurant
- Afghan Restaurant
- Japanese restaurant
- Bubble Tea shop
- Juice bar

It is also interesting to note that the presence of certain food vendors in certain proportions can act as possible indicators of the unlikeliness of an area to adopt an Ethiopian restaurant. These food vendor categories were found to be:

- Bakery
- Fried Chicken Joint
- Brewery
- French Restaurant
- New American Restaurant

## Results

It was shown that clustering has the potential to identify areas likely to embrace an emerging cuisine which in this case was chosen to be the Ethiopian cuisine. It was also shown through correlation, it is possible to identify food venue categories that are both positively and negatively associated with the emerging cuisine, implying the respective likeliness and unlikeliness of the willingness to embrace the cuisine.

As neighbourhoods continue to evolve, the food venue categories and proportions that they exist within the area may serve to change the indicators for likeliness and unlikeliness to embrace Ethiopian. As such the process would be repeated as newer information is made available. Additionally, supplementary information such as area demographics e.g. ethnicity would be useful information to have in the future to aid in the process.