

# **IBM Applied Data Science Capstone Project**

## **Evaluating Restaurant Locations in Toronto By Using Major Crime Indicators (MCI)**

**Richard Xue, MBA, MS**

**4/2/2020**

## Table of Contents

<b>1. Introduction .....</b>	<b>3</b>
Business Problem .....	3
Audience & Stakeholders .....	3
<b>2. Data .....</b>	<b>3</b>
Data Sources .....	3
Data Gathering & Cleaning .....	4
Data Feature Selections.....	5
<b>3. Methodology.....</b>	<b>6</b>
<b>4. Results .....</b>	<b>6</b>
Choropleth Crime Neighborhood Map.....	6
Choropleth Venue Neighborhood Map.....	7
Comparison of Choropleth Crime and Venue Neighborhood Maps .....	8
<b>5. Discussion .....</b>	<b>8</b>
<b>6. Recommendations.....</b>	<b>9</b>
<b>7. Conclusion.....</b>	<b>9</b>
<b>8. Reference .....</b>	<b>9</b>

# 1. Introduction

This project uses Toronto Major Crime Indicators (MCI) data to identify neighborhoods in Toronto that may be good locations to build new restaurants.

## Business Problem

Demand for new restaurants with more menu options is growing in Toronto. Healthy restaurants are emerging to attempt to meet the demand for meatless cuisine, vegetarian or vegan tasting menus.<sup>1</sup>

Selecting new venue locations is always a challenge for restaurants owners and investors. Few factors such as health-conscious and safe neighborhoods need to be carefully considered before making the final decision of new venue locations.

This project explores data insights specifically to identify restaurant locations in safe neighborhoods in Toronto that may be under-served by current venues. Detail data analysis is performed to explain the factors that lead to such data insights on selecting of new restaurant locations.

## Audience & Stakeholders

The audience for this project is restaurants owners, investors and anyone servicing the fast-food and sit-in food segments. These stakeholders have a vested interest in responding to market demand by making investment and operations decisions based on data insights. This project explores some of those data insights and provides guidance specific to venue location.

# 2. Data

This section describes the data sources used for this project, as well as the data cleansing and data preparation for subsequent data exploration steps.

## Data Sources

This project sources and integrates data from Toronto Policy Service website<sup>2</sup> as well as Foursquare data. This section describes each of these data sources and provides examples of the data.

### 2.1.1 Major Crime Indicators

Toronto Policy Service website publishes Open Data including Homicide, Traffic, Fatal Traffic Collisions, Field Information Report, and Major Crime Indicators (MCI) data in Toronto. The datasets contain MCI, premise type, occurrence date and frequency, neighborhood, longitude, and latitude, etc. They are downloadable as CSV/JSON files or via API calls at <http://data.torontopolice.on.ca/datasets/mci-2014-to-2019/data?geometry=-79.404%2C43.714%2C-79.359%2C43.725>. We use MCI indicators for each of Toronto's neighborhoods for this project.

The following table illustrates the sample records of the MCI data for each neighborhood in Toronto:

premisetype	ucr_code	ucr_ext	offence	...	occurrencedayofyear	occurrencedayofweek	occurrencehour	MCI	Division
Outside	1610	220	Robbery - Other	...	283	Saturday	1	Robbery	D53
Outside	1610	200	Robbery - Mugging	...	283	Saturday	1	Robbery	D53
House	2120	200	B&E	...	84	Wednesday	10	Break and Enter	D53
House	2120	200	B&E	...	280	Wednesday	17	Break and Enter	D53
House	2120	200	B&E	...	14	Wednesday	14	Break and Enter	D53

### 2.1.2 Foursquare Data

Foursquare allows users to search for near-by venues, view venue information, and see ratings and comments from other reviewers via mobile app. Foursquare also provides various venues data for software developers, data scientists and others who are interested retrieving venues data.

This project uses the Foursquare venue data in Toronto neighborhoods. The Foursquare venue data will be used in particularly to seek the restaurant type in the existing locations in Toronto neighborhoods. We access the Foursquare venue data via Foursquare Application Programming Interface (API) through a free developer account at <https://developer.foursquare.com/docs/places-api/>.

These venue data along with the MCI data will be applied for subsequent data analysis, data comparison, data categorization, data classification to provide detail insights to the business problem.

The following table shows the sample records of the venue name, category, and geographic location data for each neighborhood in Toronto:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park / Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park / Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park / Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park / Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Regent Park / Harbourfront	43.65426	-79.360636	Morning Glory Cafe	43.653947	-79.361149	Breakfast Spot

## Data Gathering & Cleaning

### 2.1.3 Load Toronto MCI data from Toronto Policy Service

The MCI data is downloaded from Toronto Policy Service website as a CSV file and stored locally for subsequent access. Non-essential data columns were removed. The data includes neighborhood names, crime data and geocodes.

#### 2.1.4 Load Neighborhood Geocodes

The neighborhood geocode data is retrieved from Wikipedia website at [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada). The geocode web content then is parsed using Python BeautifulSoup package to get Postal Code, Borough, and Neighborhood information.

The following table shows the sample records of the borough and postal code data for each neighborhood in Toronto:

	Postal Code	Borough	Neighborhood
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

#### 2.1.5 Retrieve Foursquare Food Venues Data

We access the Foursquare Food Venues data via Foursquare Application Programming Interface (API) through a free developer account at <https://developer.foursquare.com/docs/places-api/>. Each venue is given a category for each neighborhood in Toronto.

### Data Feature Selections

The final clean and composite neighborhood datasets from Toronto Policy Service MCI data and Foursquare Food Venues data. Neighborhood and MCI columns from MCI data are selected as features for consideration of existing restaurant types near-by. Neighborhood and category columns from Foursquare Food Venues data are choose as features for consideration of restaurant location for safety concerns.

### 3. Methodology

This section describes the data exploration, inferences, and data exploration that are conducted and how they relate to the original business problem of gaining data insights specifically to identify restaurant locations in safe neighborhoods in Toronto that may be under- served by existing venues.

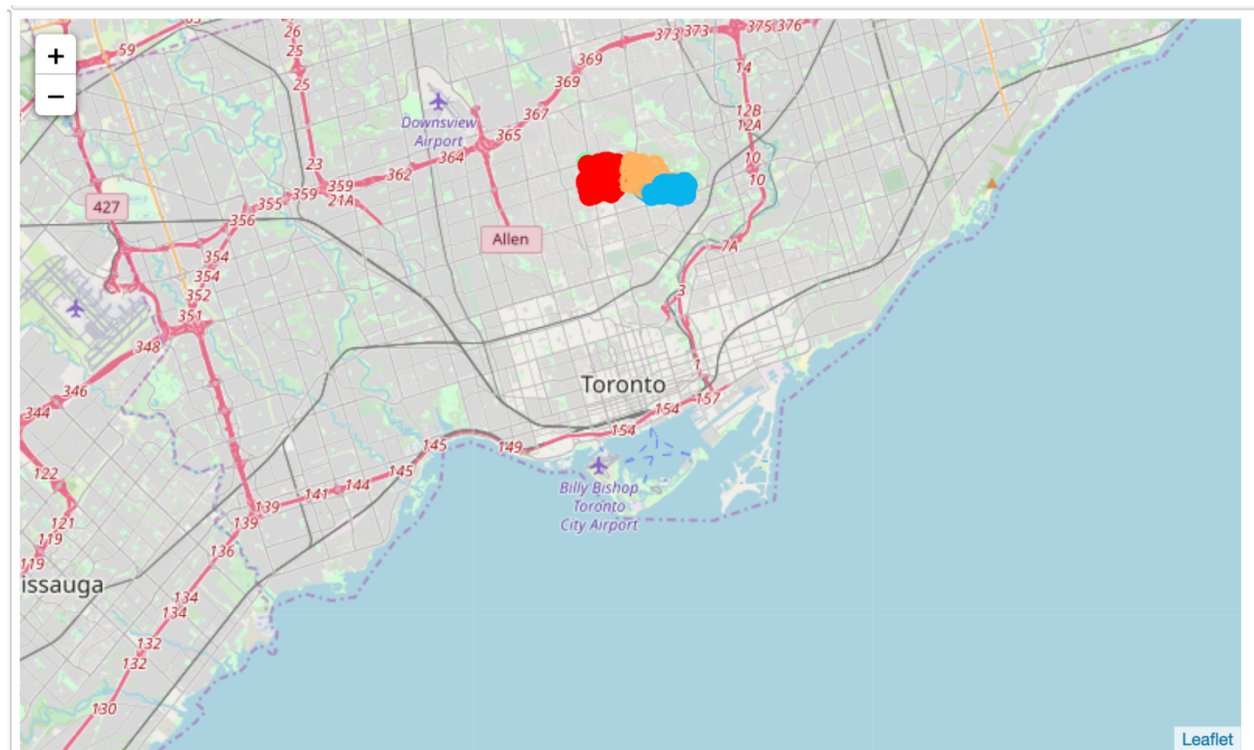
The methodology applied for this project includes:

- Data Preparation
- Data Exploration
- Data Categorization
- Clustering of Neighborhoods
- Choropleth Crime Neighborhood Map
- Choropleth Venue Neighborhood Map
- Data Comparison
- Examine the Clusters

### 4. Results

#### Choropleth Crime Neighborhood Map

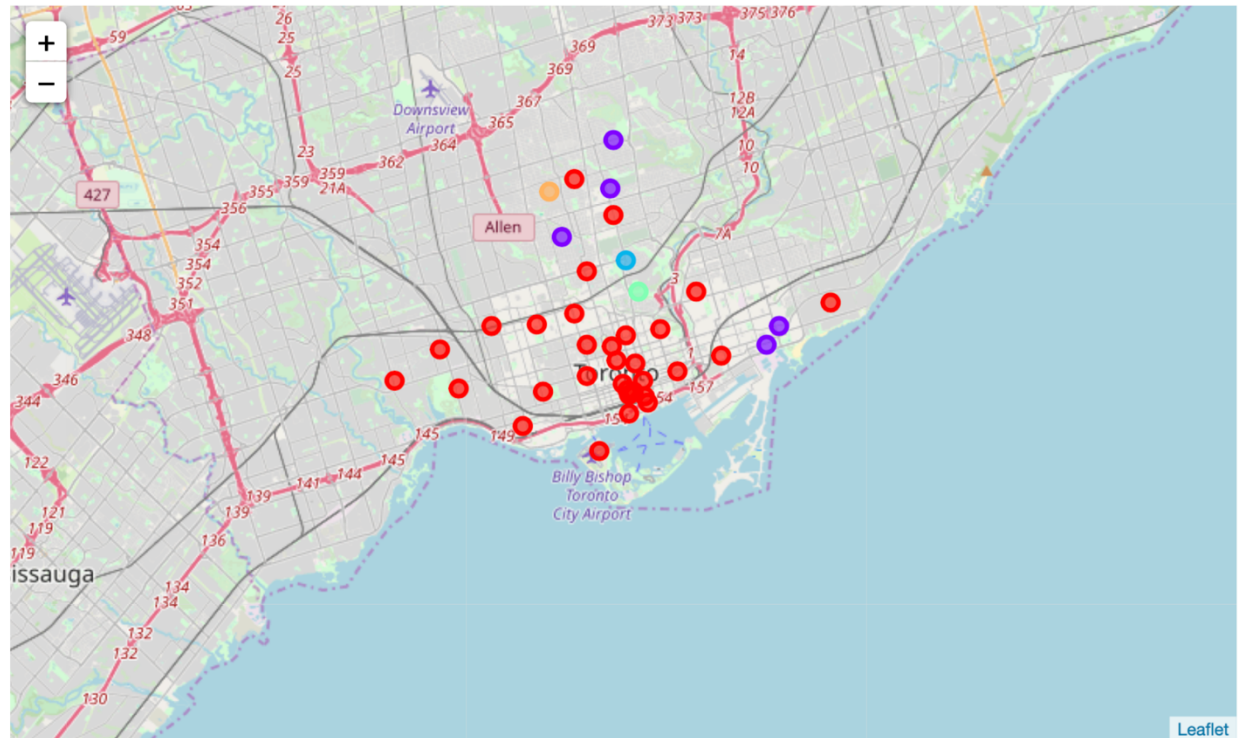
The following is a Choropleth Crime Neighborhood map composited with MCI data.



By visualizing the Choropleth Crime Neighborhood map composited with MCI data, we can easily identify the three top clusters highlighted in red, orange and blue. These three top clusters are obviously located outside of the downtown area of Toronto. The downtown area of Toronto are relative safer comparing to these three clusters.

## Choropleth Venue Neighborhood Map

The following is a Choropleth Venue Neighborhood map composited with Foursquare Venue data:



By visualizing the Choropleth Venue Neighborhood map composited with Foursquare Venue data, we can quickly identify the top cluster highlighted in red. This top cluster concentrates on the downtown area of Toronto.

The top venue cluster represents most common venues such as *Coffee Shop*, *Pub*, *Italian Restaurant*, *Mexico Restaurant*, *Japanese Restaurant*, etc. (See the table below)

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	M5A	Downtown Toronto	Regent Park / Harbourfront	43.654260	-79.360636	0	Coffee Shop	Pub	Park	Theater	Mexican Restaurant
1	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government	43.662301	-79.389494	0	Coffee Shop	Gym	Diner	Music Venue	Mexican Restaurant
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	0	Clothing Store	Coffee Shop	Café	Bubble Tea Shop	Japanese Restaurant
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	0	Coffee Shop	Italian Restaurant	Café	Restaurant	Clutch Shop
4	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Trail	Pub	Health Food Store	Donut Shop	Deli
5	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306	0	Coffee Shop	Cocktail Bar	Bakery	Cheese Shop	Cafe
6	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	0	Coffee Shop	Italian Restaurant	Sandwich Place	Japanese Restaurant	Restaurant
7	M6G	Downtown Toronto	Christie	43.669542	-79.422564	0	Grocery Store	Café	Park	Gas Station	Baby Store
8	M5H	Downtown Toronto	Richmond / Adelaide / King	43.650571	-79.384568	0	Coffee Shop	Restaurant	Café	Bakery	
9	M6H	West Toronto	Dufferin / Dovercourt Village	43.669005	-79.442259	0	Bakery	Pharmacy	Bar	Pool	Event Space
10	M5J	Downtown Toronto	Harbourfront East / Union Station /	43.640816	-79.381752	0	Coffee Shop	Aquarium	Italian Restaurant	Café	Hotel

## Comparison of Choropleth Crime and Venue Neighborhood Maps

As presented above, the top venue cluster in downtown Toronto neighborhoods illustrates the higher business density, especially in food industry. This area attracts many consumers to dine in or have a drink. It is one of good indicators to selecting new restaurants in the downtown Toronto neighborhoods.

By examining and comparing the top clusters from both Venue and Crime Neighborhood Maps, the downtown Toronto neighborhoods which have low crimes reports in past five year (2014 – 2019) are idea locations for selecting new restaurants. This MCI indicator is another key factors to locate new venues for new businesses.

## 5. Discussion

Although, number of venues and safety are among the key reasons for restaurant owner and investors to consider when they select a location to open a restaurant. There are other factors we may want to consider in our data analytic and model methodology. For example, should we consider Neighborhood Venue Average Health Ratings (VAHR) when selecting a location beside the safety reason? Should we consider the existing competitors with the same food type?



Population to Venue Ratio, Population Density and Population Growth Rate may also be good indicators for restaurant owners or investors to look into details on the restaurant, fast-food, or dine-in food market opportunities.

## 6. Recommendations

The in-depth data analysis and machine learning of this project provide several decision factors and the insights gained by the data analysis, visualizations and machine learning. These factors and insights together highlight a framework in considering locations for restaurant owners, investors and stakeholders seeking to open new venues in neighborhoods of Toronto.

The data analyses, visualizations, and machine learning in this project suggest how the MCI data and venue data can be considered as part of a decision framework for locating new venues. One key contribution is the data comparison of Toronto crime data with venue data.

## 7. Conclusion

In conclusion, this project explores data from various data sources, composites data, and create Choropleth maps for neighborhood venues and crime data. The data analysis, visualization of choropleth maps along with insights contribute to the identification of restaurant locations in safe neighborhoods in Toronto. These data analysis and insights may be useful for restaurant owners, investor and stakeholders as considering factors for new restaurant locations. These data analysis, insights and factors could also be applied for evaluating existing venues to rebrand or renovate their venues with respect to neighborhood food category.

## 8. Reference

1. Jennifer Gould Keil, <https://nypost.com/2020/01/05/restaurants-adapt-as-demand-for-meatless-cuisine-heats-up/>
2. Toronto Policy Service, <http://data.torontopolice.on.ca/datasets/mci-2014-to-2019/data?geometry=-79.404%2C43.714%2C-79.359%2C43.725>