

# **REPORT**

## **IBM Data Science Capstone Project**

**Using ML to help Drug Enforcement Agency to Stop the drug supply chains in the Toronto area.**

### **1.Introduction**

#### **1.1 Background**

The drug enforcement Agency wants to search for areas where drugs are been excessively used and to stop the supply chains. The idea behind this project is to help the department in Toronto with the needed information, that the drug dealers might be running their supply chains in or around pubs. With this project its easy to find the locations of the pubs and the nearest neighborhoods. So, that the drug enforcement agency could break the supply chain Since it's the matter of youth and nation interest to stop the drugs.

#### **1.2. Business problem**

The aim objective of this capstone project is to help the Drug enforcement Agency to find and stop the drug dealers in the city of Toronto. With machine learning algorithms like clustering we can solve this problem. It's illegal so that the pub owner can lose their license and to face further worst consequences. And this data also helps NGO'S and other organization to organize Awareness camps. If DAE got to know that there are people who wants to get out the drug addiction could help organizations to open re-habitation centers.

#### **1.3 Target Audience**

The Drug enforcement Agency, Health commission, NGO'S and other public related organization

## 2. Data and Data Extraction

1. List of neighborhoods in Toronto, Canada.
2. Latitude and Longitude of these neighborhoods.
3. Venue data related to Pub's in Toronto, Canada.

->Collecting the data provided by Wikipedia about the list of Toronto neighborhoods with web scraping. Data of Latitudes and Longitudes of these neighborhoods with the help of Geocoder package. Finally, to get the venue data related to neighborhood using Foursquare API.

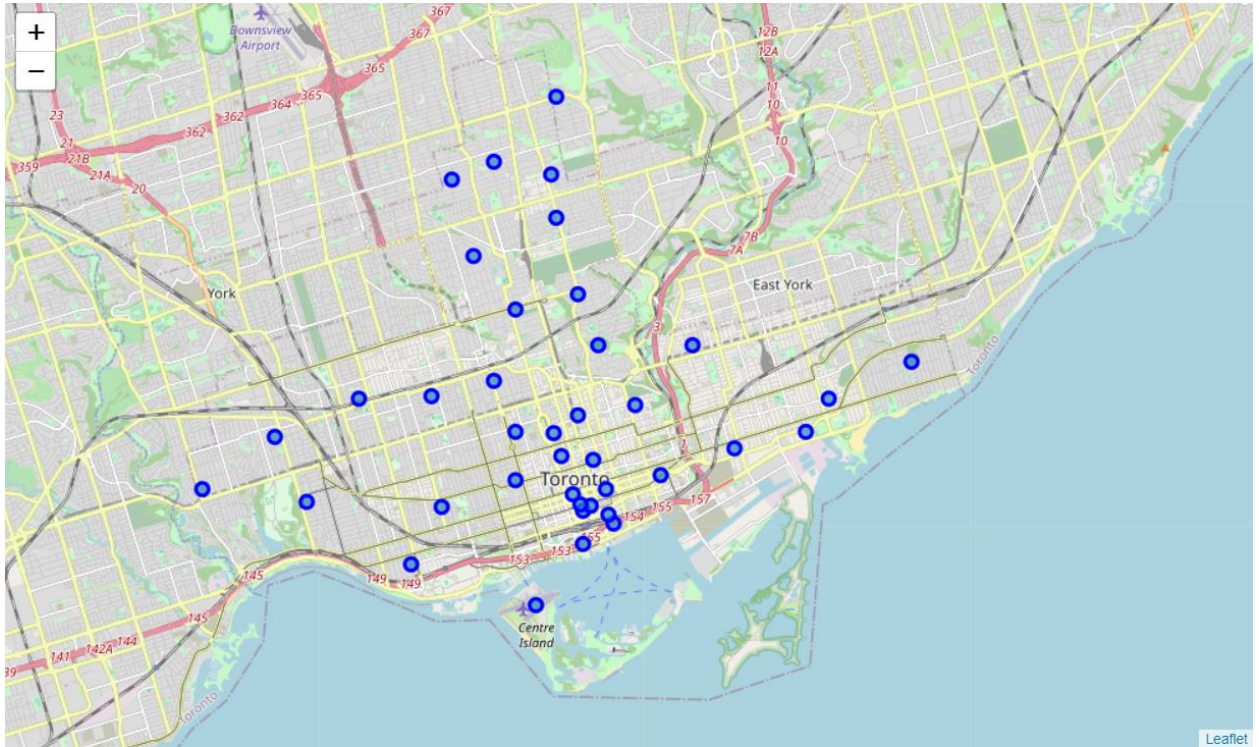
## 3. Methodology

In this project, I used Clustering a machine learning algorithm. Clustering is a Machine Learning technique that involves the grouping of data points. Firstly, web scrap the data provided by Wikipedia. And further, I did data wrangling a technique to clean the dataset, as few columns had improper values. It's important that we clean the data before we analyze.

index	Postal Code	Borough		Neighbourhood
0	2	M3A	North York	Parkwoods
1	3	M4A	North York	Victoria Village
2	4	M5A	Downtown Toronto	Regent Park, Harbourfront
3	5	M6A	North York	Lawrence Manor, Lawrence Heights
4	6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
5	8	M9A	Etobicoke	Islington Avenue, Humber Valley Village
6	9	M1B	Scarborough	Malvern, Rouge
7	11	M3B	North York	Don Mills
8	12	M4B	East York	Parkview Hill, Woodbine Gardens
9	13	M5B	Downtown Toronto	Garden District, Ryerson
10	14	M6B	North York	Glencairn
11	17	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...

**Our main objective is to solve the problem in Toronto and other no taken in to the consideration.**

**I used to Folium to visualize the data:**



**These are the nearest neighbourhoods of Toronto. So, now our major objective is to find places with a large number of pubs. To do so, I used the Foursquare API to get the venue data. And then group by columns (i.e. neighbourhood, venue, Borough).**

**One hot encoding:**

**This means that categorical data must be converted to a numerical form. If the categorical variable is an output variable, you may also want to convert predictions by the model back into a categorical form in order to present them or use them in some application.**

	Postal Code	Borough	Neighbourhoods	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	...	Toy / Game Store	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Game Store	V
0	M4E	East Toronto	The Beaches	0	0	0	0	0	0	0	...	0	1	0	0	0	
1	M4E	East Toronto	The Beaches	0	0	0	0	0	0	0	...	0	0	0	0	0	
2	M4E	East Toronto	The Beaches	0	0	0	0	0	0	0	...	0	0	0	0	0	
3	M4E	East Toronto	The Beaches	0	0	0	0	0	0	0	...	0	0	0	0	0	
4	M4K	East Toronto	The Danforth West, Riverdale	0	0	0	0	0	0	0	...	0	0	0	0	0	

Hereafter we analyze the data, it's important to understand the categorical representation of pubs in these neighborhoods.

```
pub = toronto_onehot[["Neighbourhoods", "Pub"]]
pub
```

	Neighbourhoods	Pub
0	The Beaches	0
1	The Beaches	0
2	The Beaches	1
3	The Beaches	0
4	The Danforth West, Riverdale	0
...	...	...
1625	Business reply mail Processing Centre, South C...	0
1626	Business reply mail Processing Centre, South C...	0
1627	Business reply mail Processing Centre, South C...	0
1628	Business reply mail Processing Centre, South C...	0
1629	Business reply mail Processing Centre, South C...	0

## Result:

We use clustering with k=2,( verified with Elbow method )

### Cluster 1:

	Neighbourhood	Pub	Cluster Labels	Postal Code	Borough	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	The Beaches	0	0	M4E	East Toronto	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1008	Commerce Court, Victoria Hotel	0	0	M5L	Downtown Toronto	43.648198	-79.379817	Craft Beer Market	43.649872	-79.378398	Beer Bar
1008	Commerce Court, Victoria Hotel	0	0	M5L	Downtown Toronto	43.648198	-79.379817	Movenpick Cafe	43.647687	-79.377295	Café
1008	Commerce Court, Victoria Hotel	0	0	M5L	Downtown Toronto	43.648198	-79.379817	King Taps	43.648476	-79.382058	Gastropub
1008	Commerce Court, Victoria Hotel	0	0	M5L	Downtown Toronto	43.648198	-79.379817	Ki Modern Japanese + Bar	43.647223	-79.379374	Japanese Restaurant
...	...	...	...	...	...	...	...	...	...	...	...
615	Central Bay Street	0	0	M5G	Downtown Toronto	43.657952	-79.387383	Sansotei Ramen 三草亭	43.655157	-79.386501	Ramen Restaurant
615	Central Bay Street	0	0	M5G	Downtown Toronto	43.657952	-79.387383	Chatime 日出茶太	43.655542	-79.384684	Bubble Tea Shop
615	Central Bay Street	0	0	M5G	Downtown Toronto	43.657952	-79.387383	Hailed Coffee	43.658833	-79.383684	Coffee Shop
615	Central Bay Street	0	0	M5G	Downtown Toronto	43.657952	-79.387383	Somethin' 2 Talk About	43.658395	-79.385338	Middle Eastern Restaurant
615	Central Bay Street	0	0	M5G	Downtown Toronto	43.657952	-79.387383	Coffee Public	43.660763	-79.386184	Coffee Shop

111085 rows × 11 columns

### Cluster 2:

	Neighbourhood	Pub	Cluster Labels	Postal Code	Borough	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
341	Regent Park, Harbourfront	1	1	M5A	Downtown Toronto	43.654260	-79.360636	Wine Rack	43.656573	-79.356928	Wine Shop
177	Summerhill West, Rathnelly, South Hill, Forest...	1	1	M4V	Central Toronto	43.686412	-79.400049	The Market By Longo's	43.686711	-79.399536	Supermarket
177	Summerhill West, Rathnelly, South Hill, Forest...	1	1	M4V	Central Toronto	43.686412	-79.400049	LCBO	43.686991	-79.399238	Liquor Store
341	Regent Park, Harbourfront	1	1	M5A	Downtown Toronto	43.654260	-79.360636	The Beer Store	43.657773	-79.357463	Beer Store
177	Summerhill West, Rathnelly, South Hill, Forest...	1	1	M4V	Central Toronto	43.686412	-79.400049	Mary Be Kitchen	43.687708	-79.395062	Restaurant
...	...	...	...	...	...	...	...	...	...	...	...
970	Commerce Court, Victoria Hotel	1	1	M5L	Downtown Toronto	43.648198	-79.379817	Crepe TO	43.650063	-79.374587	Creperie
970	Commerce Court, Victoria Hotel	1	1	M5L	Downtown Toronto	43.648198	-79.379817	DAVIDsTEA	43.650547	-79.383385	Tea Room
970	Commerce Court, Victoria Hotel	1	1	M5L	Downtown Toronto	43.648198	-79.379817	The Poké Box	43.650469	-79.376317	Poke Place
970	Commerce Court, Victoria Hotel	1	1	M5L	Downtown Toronto	43.648198	-79.379817	Hudson's Bay	43.652040	-79.380391	Department Store
222	St. James Town, Cabbagetown	1	1	M4X	Downtown Toronto	43.667967	-79.367675	Pizza Pizza	43.668090	-79.370274	Pizza Place

**Observation:** In cluster 2, pubs are more so its easy to say that the chances of drug supply are more and neighborhoods can be located according to the intensity of drug usage. Whereas in cluster 1, pubs are less and public parks and markets are more. So, less probability that DAE could find the drug dealers.

