**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.**

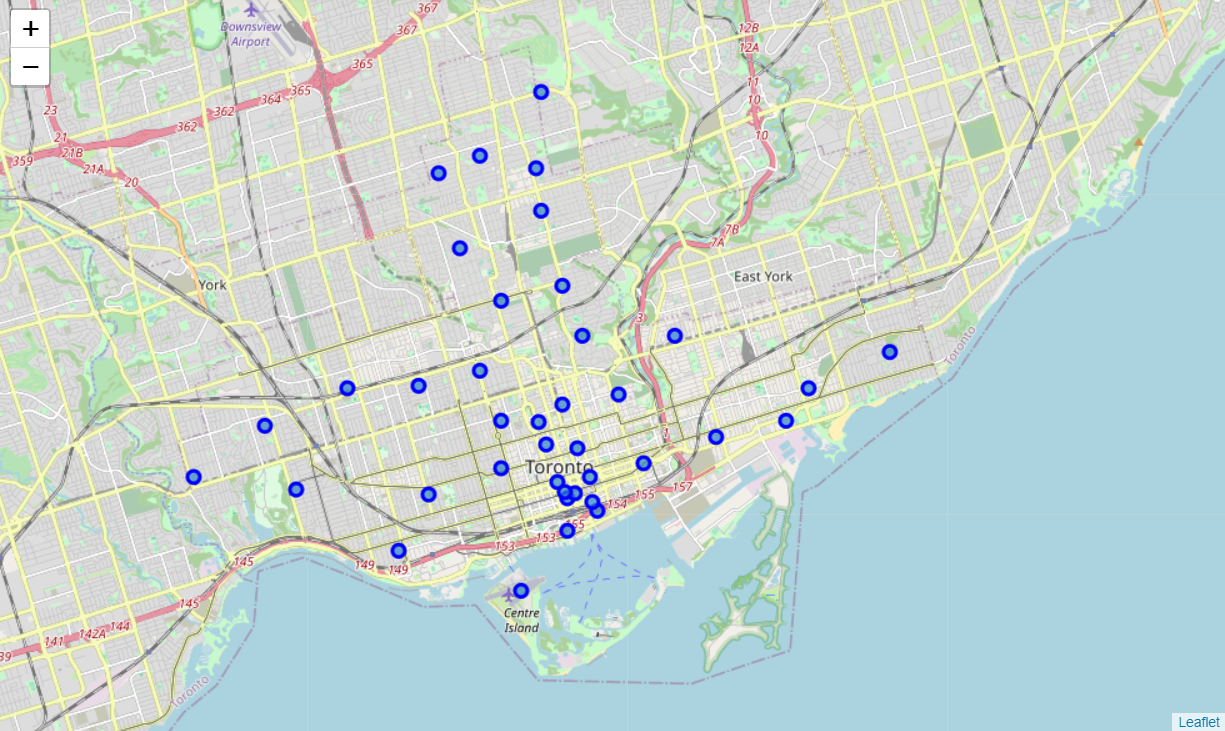
**->Collecting the data provided by Wikipedia about the list of Toronto neighborhoods with web scraping. Data of Latitudes and Longitudes of these neighborhoods with the help of Geocoder package. Finally, to get the venue data related to neighborhood using Foursquare API.**

**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. **

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

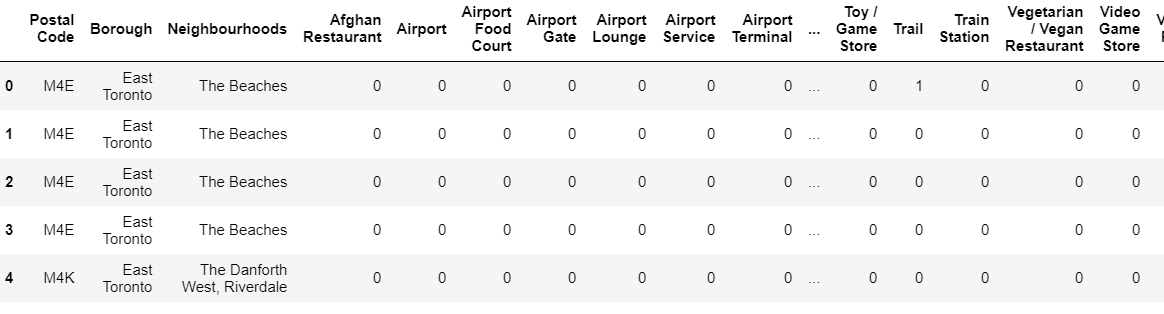
**I used to Folium to visualize the data:**

****

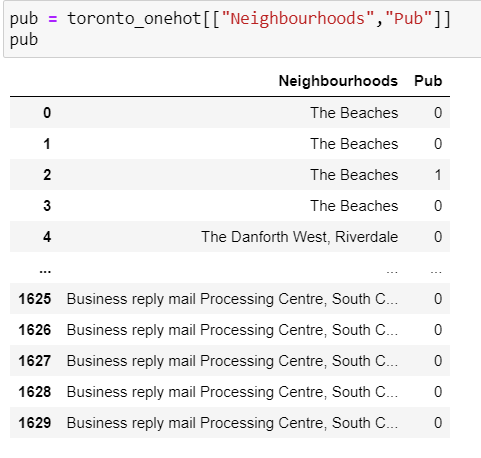
**These are the nearest neighbourhoods of Toronto. So, now our major objective is to find places with a large number of pubs. To do so, I used the Foursquare API to get the venue data. And then group by columns (i.e. neighbourhood, venue, Borough).**

**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.**

****

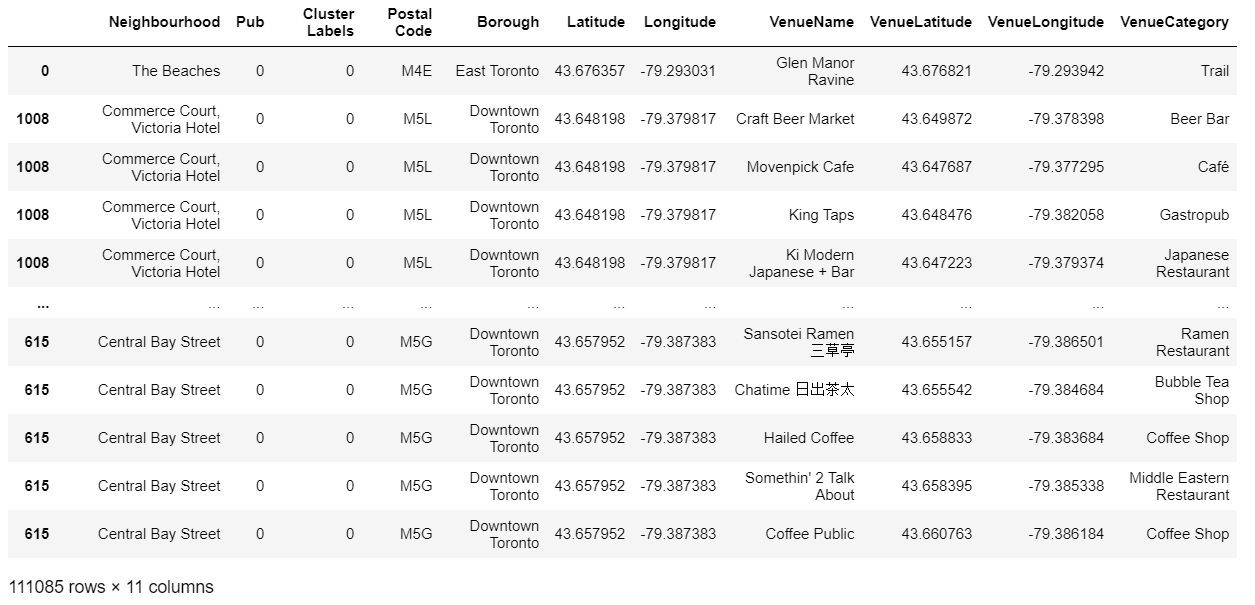
**Hereafter we analyze the data, it’s important to understand the categorical representation of pubs in these neighborhoods.**

****

**Result:**

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

**Cluster 1:**

****

**Cluster 2:**

****

**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.**