

## **Q. A description of the data and how it will be used to solve the problem.**

### 1. Methodology

The XYZ company has emphasized on the neighbourhood with lowest crime rate. So, it is required to find out -

- Which neighbourhoods have the lowest crime?
- Is population density correlated to crime level?

After sorting out the lowest crime rate neighbourhoods, the question to be answered is -

- What venues are most common in different locations within the neighbourhoods with lowest crime rate?

And then we will have a clear picture of neighbourhoods with lowest crime rate but famous for coffee shop. Then the query to be answered is -

- What is the densely populated neighbourhood having fulfilled above requirements?

So, to find out the answer of the above questions, the City of Toronto will be segmented into different neighborhoods using the geographical coordinates of the center of each neighborhood, and then using a combination of location data and machine learning, the neighborhoods will be grouped into clusters. It is required to leverage location data to solve a problem or to get deeper insights into a neighborhood's reputation.

### 2. Data

To understand and explore we will need the following City of Toronto Open Data:

- Toronto Crime by Neighbourhood: [https://opendata.arcgis.com/datasets/af500b5abb7240399853b35a2362d0c0\\_0.geojson](https://opendata.arcgis.com/datasets/af500b5abb7240399853b35a2362d0c0_0.geojson)
- Toronto Neighbourhoods Geo: <https://open.toronto.ca/dataset/neighbourhoods/>
- Foursquare Developers Access to venue data: <https://foursquare.com/>

The 'Toronto Crime by Neighbourhood' data will be used to find out the neighbourhoods having lowest crime rate. It also contains the population info of each neighbourhoods. The combination of Foursquare APIs and the 'Toronto Neighbourhoods Geo' data will then allow us to cluster and quantitatively understand the venues most common to that location.