

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 ² 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.

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.