**Introduction**

Toronto is the capital city of Ontario, Canada with 103 neighborhoods and the most populous city in the country with a population of 2.7 million. I decided to do this project based of this city owing to the good data availability of this city.

An American company is looking to open a branch of their baseball accessories/souvenirs shop in Toronto, Canada. They are looking for the most ideal location of their new shop to maximize visitors and profits. To do this, they decided to hire a team of data scientists to leverage the power of data to gain insights about where them should open their branch of baseball shop in the city

**Data**

There are three sources of data that we are going to be using for this project. The first data is a list of neighborhoods, boroughs, and postal code of Toronto. We obtained the dataset by scraping this Wikipedia page <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M> using BeautifulSoup.

Text

Description automatically generated

Figure 1. Preview of the dataset.

The next data we will be using is geospatial data which contains coordinates for each neighborhood in Toronto. The data is available in this link <https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs_v1/Geospatial_Coordinates.csv>

Table

Description automatically generated

Figure 2. Coordinates data

Both of these dataframes are combined to create a single data frame, in order to query information from our third data source, Foursquare API. Foursquare Developer API provides the means to query a list of trending venues within the radius of specific coordinates. In the notebook, a function to obtain JSON values from the API into a dataframe is created.