**Final Capstone Project – High-level Analysis of Toronto Metropolitan Area**

# **Introduction**

I am an employee of a foreign development and investment firm that is seeking different opportunities in Toronto area. There is no limit to what the firm invests as long as it's profitable opportunity. We are looking to enter the Toronto market because it is a big city and there's a lot of traffic.

However, with us being a foreign company and having little local knowledge, we need to do some exploratory analysis of the city before we decide on our entry strategy and projects. I am tasked to do some research to have a high-level understanding of the metropolitan area and present the findings to my supervisor and investors.

# **Business Problem**

In this analysis, I am trying to get a high-level understanding of the city. Since we are hoping to invest in various types of opportunities, we will be looking at different venues in different neighborhoods. There are several neighborhoods in Toronto area. It will be less efficient to examine each neighborhood for a high-level analysis like this. Therefore, I am planning to cluster these neighborhoods into clusters based on their venue category counts. After determining the optimal number of clusters and successfully clustering neighborhoods, I can check within each cluster to see what the majority of the venues are in each cluster. The result of this observation can give me a high-level understanding of each cluster, namely, I can say if cluster/district 1 is a food district and district 2 is a financial district, etc.

To summarize, my business problem is to determine the optimal number of clusters of neighborhoods in Toronto, and then to hopefully examine each cluster to determine what kind of district it is to inform our investment strategy.

# **Data**

In order to solve my business problem, I'll need geological information for all neighborhoods in Toronto, especially latitudes and longitudes. I'll also need the categories of the venues to determine the type of district they are located in. For example, if a cluster has a lot of venues categorized as restaurant or coffee shop, we can say this cluster is more likely a food district.

Fortunately, all the data can be obtained from the internet with a quick parse and download. The venue data can be obtained from the Foursquare data center. I don't need additional data sources. The data collection process is shown below.

Graphical user interface, text, application

Description automatically generated

Figure 1. Geological data of Toronto Neighborhoods

Graphical user interface, text, application

Description automatically generated

Figure 2. Venue data