Identify Niche Areas for New Business Venture

# Introduction

## Background

Every year, hundreds of new businesses are started in every city. Some businesses see higher successes than others. Some of the factors that decide whether a business is successful is the demand for such a business, competitors in the market, availability of patrons, and profit margin. The goal of this project is to offer some insights into how to make some of the decisions pertaining to starting a new venture.

## Problem

Suppose a person is interested in starting a business venture. The person must first decide what type of business to start and where to start it. These two key factors decide all other parameters. Therefore, this project focuses on solving these two questions.

Let us look more closely at these two question. We first select the potential list of cities where this person is interested in starting the business. This selection can be based on his domicile, i.e., where he/she is currently residing or where they want to live long term. In general, people have specific preference on which broad area they want to live in and therefore, we will limit our search space within that area.

In this project, we look at two potential cities – New York City and Toronto. The goal is to find 1) what business is viable in this area and 2) a neighborhood to start this business.

# Data Acquisition and Preprocessing

Details of neighborhoods and boroughs in New York City was obtained from the dataset provided by Coursera at <https://cocl.us/new_york_dataset>. The corresponding data for Toronto was scraped from the Wikipedia page <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>. The BeautifulSoup (bs4) package was used to get the contents of the webpage. The contents were then extracted into a table, preprocessed (stripped of unnecessary values/entries) and structured to obtain a pandas dataframe. The geospatial coordinates of all the neighborhoods were then added to the dataframe.