# **New Business Locational Data**

## Johanna Miller

## March 19, 2021

1. **Introduction**

-Background

There are many issues to consider when one is thinking about opening a business. One of the key issues an entrepreneur must think about is the business location. The location will be critical because it will determine:

1. Whether the business receives enough customer traffic
2. Whether there is too much competition in the area, making it difficult for a new player to enter the market and build sufficient market share.

Our hypothetical client is an entrepreneur who wants to open a new coffee shop. We will use the data at hand to uncover areas within the city of Toronto for him to consider for his location.

**2.Data acquisition and cleaning**

-Data sources

For this project, I used two sources:

1. The Wikipedia page for Toronto. This table contained a list of the postal codes and neighborhood information for the city and can be found at: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
2. Data from the Foursquare API. I narrowed the venue information returned to only include businesses with venue category “Coffee Shop.” This resulted in a dataset of 173 records and seven features.

-Cleaning

The Wikipedia table was already nearly ready to use. I needed to remove a handful of missing values and shape it into the form I wanted to use. Additionally, I removed the postal codes without an assigned borough/neighborhood. The Foursquare API returned JSON data, which needed to be shaped and normalized but was otherwise relatively clean.

Due to the nature of the small dataset, there was no specific feature selection required. I created a few tables that included specific columns, but no features were ultimately eliminated.

3. **Analysis**

-The main goal of the analysis was to narrow down the potential locations to a few areas with a moderate number of coffee shops already existing. I accomplished this by eliminating neighborhoods with fewer than ten and more than three shops. This reduced the number of candidate neighborhoods to ten. I then mapped the remaining neighborhoods using Folium. This gives a good visual for our client to understand where in the city these neighborhoods are.