# Introduction

The goal is to help a potential investor from abroad, interested in consumers’ retail businesses in the city of Toronto. I want to provide a support analysis to help them decide what are the most attractive activities in the city, and consequently where possibly to open a new one.

# Data acquisition and cleaning

## Data sources

I will use a geospatial database of Toronto Neighbourhoods (Postal Code, Borough, Neighbourhood, Lat, Long) from Wikipedia and Foursquare geospatial data on venues in the city.

## Data cleaning

The geospatial database for Toronto from Wikipedia was grouped by postal code and had “not assigned value” in the “Borough” columns. Hence, I had to clean it from “not assigned” data and eliminate duplicate in the Neighbourhood column.

For the Foursquare database, I had to call the API to get the Json file and then make it into a pandas dataframe with Python. From it, I got only Name, Category and Latitude and Longitude features.

# Exploratory analysis

To help our potential investor, first step was to extract Top100 most attractive activities in the city of Toronto. It came out that “Sushi Restaurants” are at the top of the list.

Hence, second step has been to look only for sushi restaurants and see where they are. This will help a potential investor to decide where to set a potential new restaurant.

# Conclusion and further directions

From here, the potential investor could look for restaurants ratings, in order to learn from top ratings ones and conversely not opening a new one in an area already served by a top rated restaurant.