



# Applied Data Science Capstone

## **Data**

The most valuable thing that we need is good and accurate data. So to be precise, for this project there are two data sources that are going to be used. The first containing information about the different areas of Toronto (i.e. postal code, city areas and boroughs). This dataset is available on Wikipedia and it also used in the previous weeks. This data are going to be used to navigate in Toronto with the different area names.

Once this is possible there will be the need of some extra information about each Neighborhood. This information is necessary to find out how many restaurants are located in each area. Foursquare is the solution to this problem as it provides information for every area only by getting the coordinates and area name.

So to sum up, this project requires:

- List of Neighborhoods of Toronto, Canada
- Coordinates of these neighborhoods
- Venue data related to Greek restaurants. This will help us find the neighborhoods that are more suitable to open a Greek Restaurant

## **EXTRACTING THE DATA**

- Scraping Wikipedia webpage to retrieve Toronto neighborhoods
- Getting Latitude and Longitude data of these neighborhoods via geocoder package
- Using Foursquare API to get venue data related to these neighborhoods