Data Section

In this project, each famous borough will be selected from the Toronto City and from New York City respectively for analyses and exploration. Downtown Toronto from Toronto City and Manhattan from New York City be selected for the carrying of this data exploration project.

For the New York City Location Dataset,

- The relevant data (Borough, Neighbourhood, Latitude and Longitude) which saved in JSON format file could be obtained from NYU Spatial Data Repository (https://geo.nyu.edu/catalog/nyu_2451_34572). Since the original dataset file is in JSON format, it is needed to download the JSON file and convert it into *pandas* dataframe.
- Foursquare API is one of the location data providers used to retrieve the data from the Foursquare database through the application of API to explore a location even to get trending venues around a point of interest. We use the venues group with explore endpoint, pass the latitude and longitude coordinates of neighbourhoods in Manhattan along with our credentials, then make the call to the database to get a list of the popular spots around the neighbourhoods in return. The explored categories of venues which are in the *items* key could be obtained, clean the JSON file and structured them into *pandas* dataframe.
- With the important data such as Borough, Neighbourhoods with their geographical coordinates, and the venues categories, the analysing, clustering and examining of the data processes could be carried out.

For the Toronto City Location Dataset,

- The relevant data such as Postal Code, Borough and Neighbourhood are scrapped from the Wikipedia webpage (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) using Beautiful Soup package. The read data then transform into *pandas* dataframe.
- In order to utilize the Foursquare location data, we need to get the geographical coordinates of each neighbourhood. These data in csv file could be accessed through https://cocl.us/Geospatial_data then convert them into *pandas* dataframe.
- With the application of Foursquare API, we use the venues group with explore endpoint, pass the latitude and longitude coordinates of neighbourhoods in Downtown Toronto along with our credentials, then make the call to the database to get a list of the popular spots around the neighbourhoods in return. The explored categories of venues which are in the *items* key could be obtained, clean the JSON file and structured them into *pandas* dataframe.
- With the important data such as Borough, Neighbourhoods with their geographical coordinates, and the venues categories, the analysing, clustering and examining of the data processes could be carried out.