

I. Data description

To consider the problem we can list the data as below:

- The Toronto Neighborhood Coordinates are obtained from the List of Postal Codes of Canada (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) and the coordinates of each postal code (https://cocl.us/Geospatial_data), clean the data that don't have Borough or Neighborhood, group by Boroughs and Postal code. Then, the Postal Code is matched between the first data and the postal code coordinates, to finally get the coordinates of each neighborhood. And then finally takes only the Center of Toronto.
- The New York Neighborhood Coordinates was downloaded from the Json file (https://cocl.us/new_york_dataset), then capturing the Borough, Neighborhoods and the location (latitude and longitude), and finally takes the Manhattan's information. And then finally takes only the Center of Manhattan.
- The Paris Neighborhood Coordinates is loaded from the CSV file (<https://opendata.paris.fr/explore/dataset/arrondissements/download>) after the information is cleaned and structured in a data frame. And then finally takes only the Center of Paris.
- The venues of each neighborhood are obtained by Foursquare with a limit of 100 venues and a radius of 500 of the center of each neighborhood. For those venues is saved the latitude, longitude and venue category