

Data Sets:

Following datasets and information will be required for this project:

- 1) For this project, I will be using postal code information on neighborhoods within different in Toronto. This dataset is available at following location:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- 2) The coordinates for each postal code using csv file obtained from following URL:
https://cocl.us/Geospatial_data
- 3) Foursquare API to extract information on Venue location and Venue type within Toronto city.

Plan of Action:

Following steps need to be followed for successful execution of this project:

- 1) Download and preprocess data using different libraries like beatifulsoup4, requests, numpy and pandas.
- 2) Save data in tabular format as a dataframe.
- 3) Then calculate the distance of each venue from Toronto city center (43.6532, -79.3832) using distance function from 'geopy' python library to calculate distance of each venue from City center.
- 4) Then create a master table by merging all the data-table. The master table will contain several important features.

For Example:

- i) Borough name
- ii) Neighborhood name
- iii) Neighborhood coordinates

- iv) Venue name
 - v) Venue Coordinates
 - vi) Venue Category
 - vii) Distance from city center, *etc.*
-
- 5) Estimation of Pearson's correlation coefficient and regression analysis between the distance from city center and number of venues will be calculated using 'scipy'
 - 6) Then using k-mean clustering method, cluster all the venues based on their location and check for most common type of food channels for each cluster.
 - 7) Visualize all these cluster on the map using Foursquare API and folium library
 - 8) Recommend the stakeholder a location proximal to city center with least number of competition, and also warn them about most common type of food channels/competitors in the area.