

New York City and the city of Toronto - Similarities and Disimilarities

Data Sources

Based on the most common places captured from Foursquare, we are going to segment areas of Toronto and New York. Two randomly neighbourhoods will be picked and analysed the top 10 most common venues in each of those two neighbourhoods based on the number of visits by people in each of those places. K-mean clustering unsupervised machine learning algorithm will cluster the venues based on the place category such as restaurants, park, coffee shop, gym etc.

The following urls will be used in order to collect the data necessary for analysis:

- Toronto Postal Codes from Wikipedia:

[https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

- New York dataset that contains the 5 boroughs and the neighbourhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

- Foursquare API: (<https://foursquare.com/>)

Performing location search, location sharing and details about a business. Photos, tips and reviews.

- Folium- Python visualization library used to visualize the neighbourhoods cluster distribution

Various libraries, such as pandas (for data analyses), json (to handle json files), geopy (to retrieve location data), sklearn (machine learning library), etc.