**Data Science Capstone Project Report:**

**Clustering and Segmenting and Analysis the New York and Toronto neighbourhood**

Contents

[1. Data Requirement 3](#_Toc25400853)

This file contains information about the project being executed, in this case, the Clustering and Analysis of Neighbourhood data of New York and Toronto. It is organized according to the capstone project process flow.

# Data Requirement

.

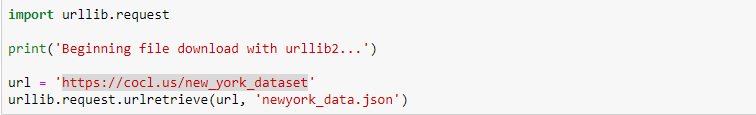
The dataset for this project is collected from various sources.

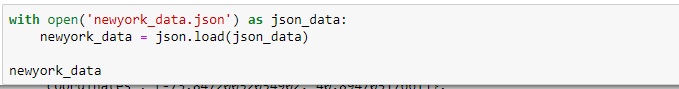
**New York City Data:**

New York City data is collected from the below link:

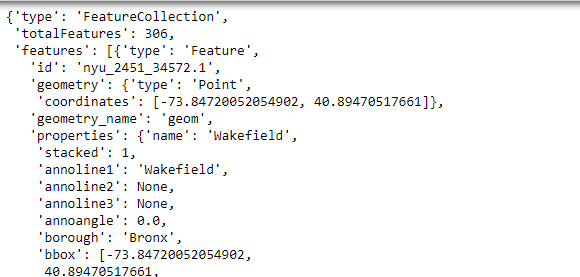
<https://cocl.us/new_york_dataset>

This is a json file. So this needs to read the data in json format and loaded locally in newyork\_data.json file. The read the data from the local json file.

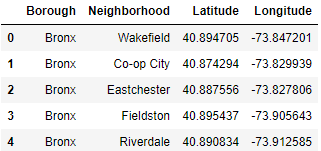




Sample json data:



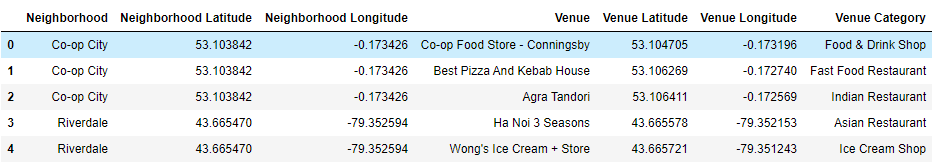
This json data then transformed and pre-processed into data frame with columns Borough, neighbourhood, latitude and longitude.



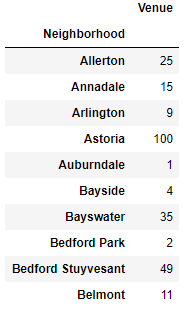
The dataframe has 5 boroughs and 306 neighborhoods.

Foursquare location data used along with this data to find out the venues, popularity, location etc.

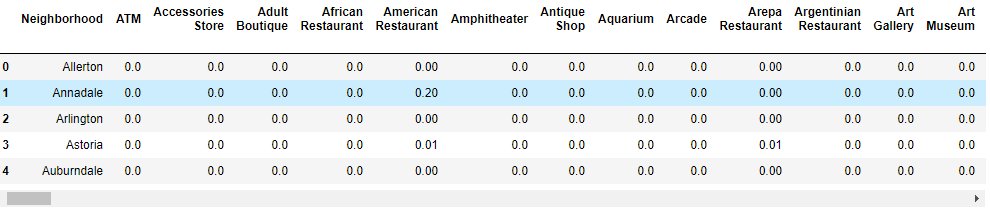
Foursquare location data is used to venue location, details and category.



How many venue in which location.



Mean venue count per neighbourhood:



Top 10 most common venues per neighbourhood:

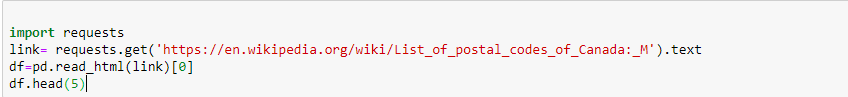


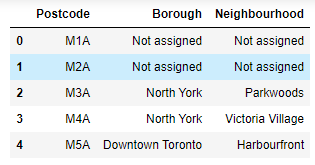
**Toronto Data:**

Toronto data is collected from the below link:

<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

This Wikipedia contains the postal codes of the neighbourhood of Toronto. The table read from the Wikipedia as below and loaded as data frame.

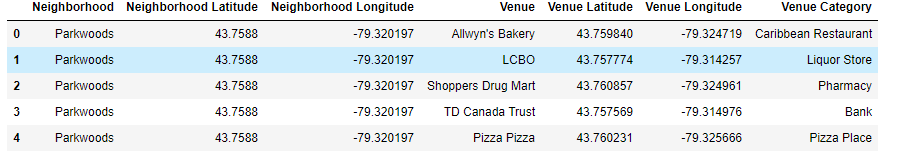




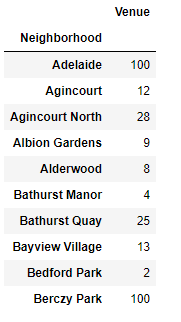
The dataframe has 287 records.

Next foursquare location data used along with this data to find out the venues, popularity, location etc. And used for further analysis, clustering and segmentation.

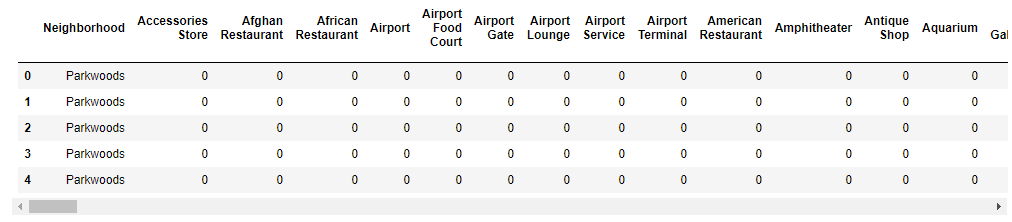
Foursquare location data is used to venue location, details and category.



How many venue in which location.



Mean venue count per neighbourhood:



Top 10 most common venues per neighbourhood:

