



# IBM data science professional certificate capstone – Battle of neighborhoods

Project Report

# Table of contents

As a part of the final capstone assignment, this is going to be my final report blog which will have the following components :

- ▶ Introduction/business problem
- ▶ Data gathering
- ▶ Data science methodology
- ▶ Results
- ▶ Discussions/observations
- ▶ Conclusion

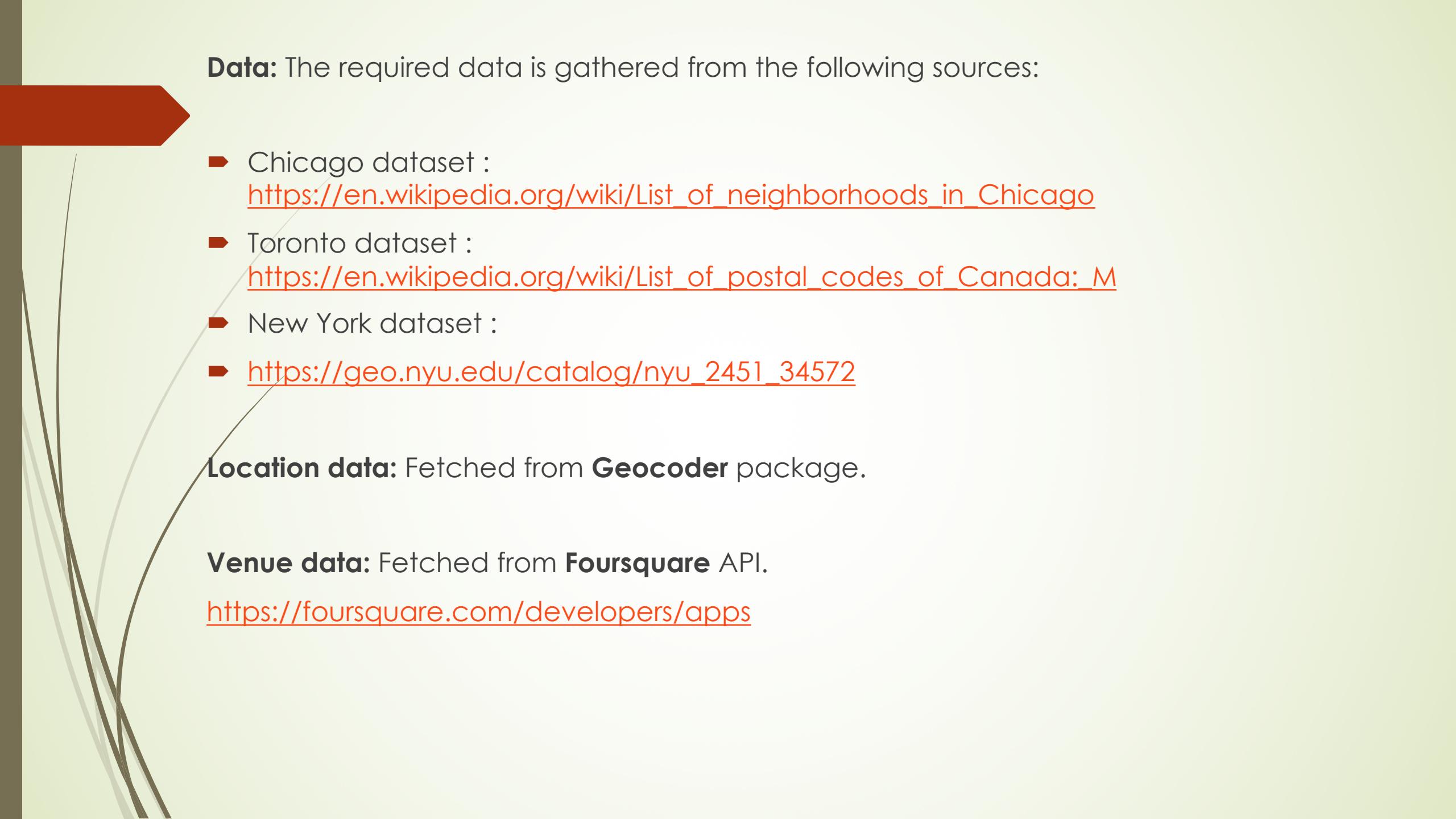


**Introduction:** I am going to explore the city of Chicago on basis of architecture and tourist outlooks, food habits, recreation and sports and cultural outlook, so that either the residents or tourists have readily available access to city's neighbourhood exploration and the observations would be substantiated with sufficient data. Lastly I will compare this exploration data with similar data of neighbourhoods of two other cities namely New York and Toronto, to cite the similarities as well as the differences between the people's culture of surrounding cities as well as countries.

**Background:** Chicago is a city in the state of Illinois in United States and is the third most populated city in the US with approximate population of 2.7 million. Chicago is widely known for its food, architecture and loyal sports fan, it also houses a few of the most reputed college and high school institutions in the world. So I am going to explore the neighbourhoods of the city on the basis of following pointers:

- ▶ Educational institutions—High schools, universities
- ▶ Cuisines—Restaurants, bars, stalls and cafes
- ▶ Transportation— Public : Buses, subways and trains
- ▶ Sports—Stadiums, arenas and clubs
- ▶ Healthcare—Hospitals and clinics
- ▶ Tourist attractions— Architectural buildings, museums and skyscrapers

- 
- Once the city's neighborhoods exploration is complete, I will compare these neighborhoods' on similar grounds with those of New York and Toronto.
  - I will then cluster similar group of observations from each city's neighborhood and compare it with a different city and report my observations regarding the same.
  - These observations would base as useful insights as to why a particular city is better than the other or moreover as a ground point from which anyone can start driving their decisions based on data.
  - These observations would be useful for either the residents or tourists who would like to explore the city before arriving at a final decision, with an advantage that now the observations would be backed up by data compared to earlier hunches.
  - The cluster readings would clearly help us identify the most popular, most visited and most trending places in each of the boroughs of their respective cities, which then can be pushed in a travel blog to help residents and tourists.
  - Finally the cluster readings would also help us determine the nature of the venues so that it would be easy to differentiate if the particular venue has a cultural or a business outlook.



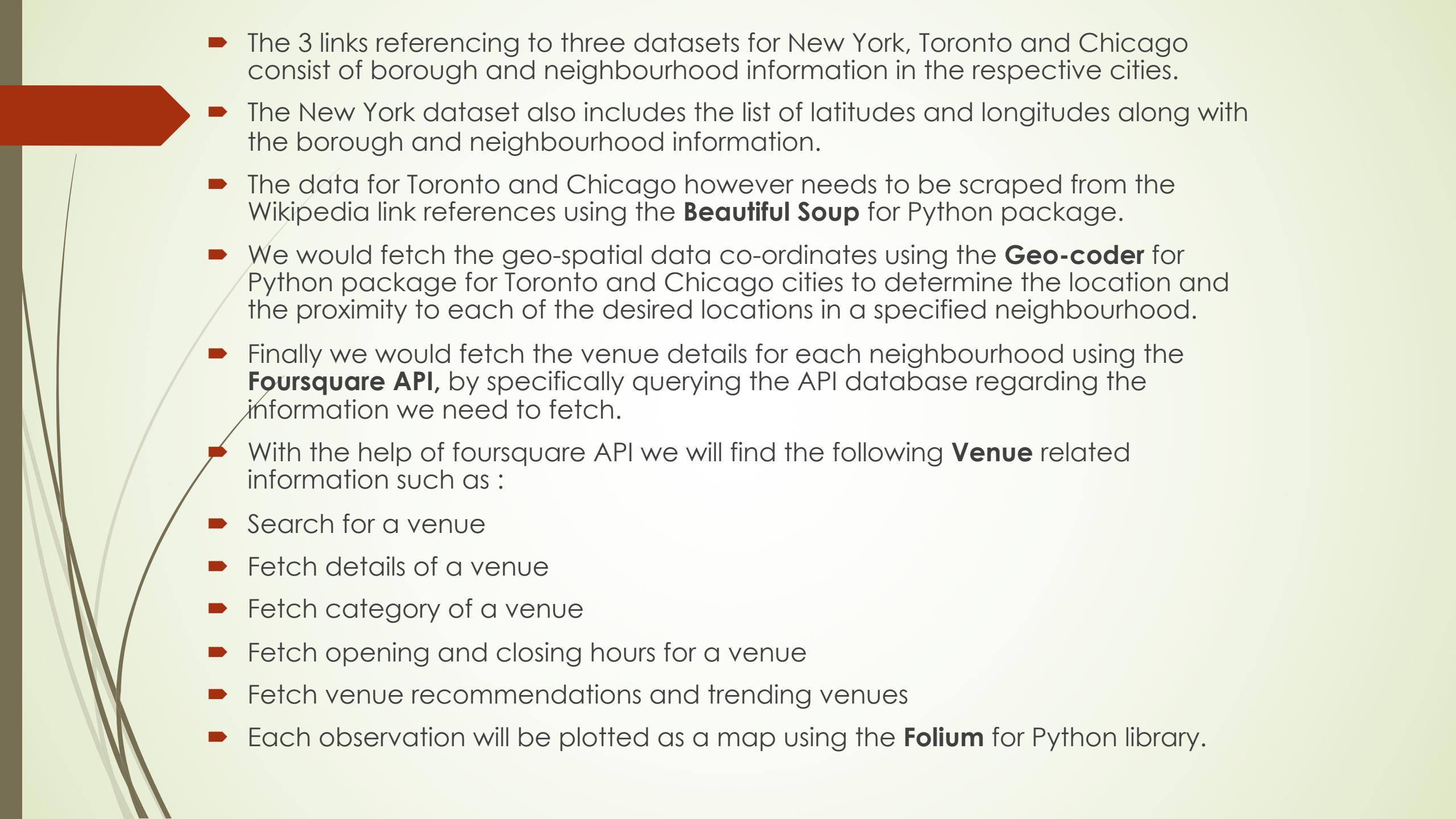
**Data:** The required data is gathered from the following sources:

- ▶ Chicago dataset :  
[https://en.wikipedia.org/wiki/List\\_of\\_neighborhoods\\_in\\_Chicago](https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago)
- ▶ Toronto dataset :  
[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 :  
[https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)

**Location data:** Fetched from **Geocoder** package.

**Venue data:** Fetched from **Foursquare** API.

<https://foursquare.com/developers/apps>

- 
- ▶ The 3 links referencing to three datasets for New York, Toronto and Chicago consist of borough and neighbourhood information in the respective cities.
  - ▶ The New York dataset also includes the list of latitudes and longitudes along with the borough and neighbourhood information.
  - ▶ The data for Toronto and Chicago however needs to be scraped from the Wikipedia link references using the **Beautiful Soup** for Python package.
  - ▶ We would fetch the geo-spatial data co-ordinates using the **Geo-coder** for Python package for Toronto and Chicago cities to determine the location and the proximity to each of the desired locations in a specified neighbourhood.
  - ▶ Finally we would fetch the venue details for each neighbourhood using the **Foursquare API**, by specifically querying the API database regarding the information we need to fetch.
  - ▶ With the help of foursquare API we will find the following **Venue** related information such as :
    - ▶ Search for a venue
    - ▶ Fetch details of a venue
    - ▶ Fetch category of a venue
    - ▶ Fetch opening and closing hours for a venue
    - ▶ Fetch venue recommendations and trending venues
    - ▶ Each observation will be plotted as a map using the **Folium** for Python library.



Let's discuss the approach for week 2 of the capstone project.

- ▶ Import the required packages to execute our code.
- ▶ Fetch the data from the respective sources and store it in data-frames.
- ▶ Work on the data-frame to fetch the desired insights.
- ▶ Map the data from the data-frame with geo-spatial co-ordinates.
- ▶ Plot the data with their co-ordinates as a map using **Folium**.
- ▶ Superimpose the map with venue and location details fetched from the **Foursquare** API.
- ▶ Query to API database for a specific output return value such as search, tip, user details, photos, recommendations etc.
- ▶ Explore the neighbourhoods within a specified vicinity.
- ▶ Cluster the neighbourhoods based on similarity
- ▶ Analyse each of the clusters for each neighbourhood in each of the cities.
- ▶ Present the analysis, observations and conclusions in a report.