

# Capstone Project - The Battle of the Neighbourhoods

*The Applied Data Science Capstone is a part of **IBM Data Science Professional Certificate** program.*

## Data

### Requirements

1. List of neighbourhoods of Chicago.
2. Location coordinates of all the neighbourhoods.
3. Information of venues in each neighbourhood.

### Sources

The list of neighbourhoods along with community area is available at this Wikipedia page: [https://en.wikipedia.org/wiki/List\\_of\\_neighborhoods\\_in\\_Chicago](https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago).

Geopy libraries are used to acquire the location coordinates of each neighbourhood of Chicago.

Using FourSquare API, the venue information is obtained.

### Data preparation and wrangling

With the use of Pandas library, the list of neighbourhoods and community areas is obtained from the wiki page and then this data is stored as a dataframe. Then using the community area names, the location coordinates are called using Geopy libraries creating two lists: "longitude" and "latitude". It was found that the location coordinates of "Lincoln Square" which the Geopy library returned are not correct, therefore the rows containing "Lincoln Square" were dropped, thus cleaning the data. Then the two lists containing the location data were added to the dataframe as columns.