

Data

In order to get a complete view we will be using the below data sets and initiate our analysis

Neighborhood data - This covers all the details of 5 boroughs and 306 neighborhoods

Geographical data - This covers all the coordinates for the respective neighborhoods, eventually used to plot the maps

Venue Data - This covers all the existing venues/restaurants established in the same business sector, we will use to provide more insights on how they impact our decision

Data Extract and merging plan

The open data source (https://cocl.us/new_york_dataset) will be used to read borough and neighborhoods data, then geographical coordinates will be added to the respective neighborhoods using Python Geocoder Package.

Foursquare API will be used to get the venue data for the neighborhoods and will eventually help us to provide insights to our business problem.

Data Science techniques going to be used

Web Scrapping

Fetching Data from API(json)

Data Cleaning

Grouping

Data Wrangling

Map Plotting