Data

Data sources

In order to start collecting data and ask questions about neighborhoods, we first need to get the locations of the neighborhoods.

San Francisco neighborhoods location

We will use a geojason file containing the boarders of the neighborhoods in San Francisco.

We will have to calculate the center of each neighborhood border in order to get a specific coordinate of the neighbourhood.

The geojason file can be downloaded from the following link:

https://cocl.us/sanfran_geojson

Manhattan neighborhoods location

We will use a geojason file containing the neighborhoods locations and names of New-York area.

This geojason file already contains the coordinates of the neighbourhoods, however it does not contains information on the boarders of the neighbourhoods.

We will use only a subset of this neighbourhoods which are localized to Manhatta,

The geojason file can be downloaded from the following link:
https://geo.nyu.edu/catalog/nyu_2451_34572

Foursquare location data

In order to be able to compare between different neighborhoods we will have to retrieve information about them. The Foursquare location data will be used to collect data regarding the venues in the neighborhoods.

Foursquare is a location-based online social network that collects data regarding venues and their location. The data is being collected using more than 30 million people worldwide.

The Foursquare Places API will be used to access to Foursquare's database of venue data.

A link for Foursquare API:

https://api.foursquare.com

Information about the venues in each neighbourhood will be retrieved from the Foursquare API. The relative abundance of the different venues types in the neighborhoods will be used to determine the similarities between neighborhoods. The details will be described at the methodology section.