Data and Methodology

1. Process data about neighborhoods in New York

Data about neighborhoods in New York can be collected from official online sources such as
https://cocl.us/new_york_dataset

The neighborhood data includes critical information such as names, latitude and longitude.

The neighborhood data must be cleaned to reach criteria of having unique identification, no duplicate data.

2. Process data about venues of these neighborhoods

Based on the neighborhood data with certain latitudes and longitude, data about venues are collected from source of Foursquare.

To avoid processing data too long, the venue data must be not excessive, so some limitations are:

   i. Radius of a particular location (neighborhood with latitude and longitude) is 1000 kilometers
   ii. Maximum number of venues of a particular neighborhood is 100

The venue data must show the interested rate of each venue category of a particular neighborhood

3. Find the most interested venue category (the service is most interested in New York)

The most interested service in New York is the venue category having maximum rate

4. Analyze the data to find neighborhoods (areas) having no the most interested venue category

Neighborhoods are clustered based on their similarity of venues

The answer is the cluster including no the most interested venue category