Data Sources

Since this project is data driven, it is important to select and include all the required data. However, in this project only two data sources were used which are listed below:

- New York Neighborhood Dataset which includes all New York Boroughs, Neighborhoods and locations (latitude, longitude). This dataset will help in finding the required venues and information from Foursquare API. It will also support in clustering.
  https://cocl.us/new_york_dataset
- Foursquare Places API which gives real-time access to Foursquare's global database of rich venue data and user content. This will help in analyzing the neighborhoods in terms of the popular venues and food venues. It provides features regarding venues such as rating and price which will be helpful in clustering.
  https://developer.foursquare.com/docs/places-api/

There are various data sources that may be used to enrich this project and improve the clustering and in return enhance the predictions. For example, Foursquare Places Database may be used to get some restaurant/café features that are not available in the Foursquare Places API Also, obtaining data about the demographic, social and economic characteristics of the people in Manhattan would've helped to analyze and understand their interests and tendencies.