Business Problem

New York is a magnet for tourists from all around the globe. It allures international spotlight where it is one of the most sought after travel destination due to its cultural, ethnic and natural diversity, world best museums and art galleries, developed infrastructure and fine educational institutions. Also, New York is the heart of trade as economic growth as well as the best technological, medical and scientific minds in the world which makes it a strong competitor on the world map. This project will focus on Manhattan because the possibilities are endless where it has a dense population, beautiful skyscrapers, lavish shopping, tourist attractions, iconic historical structures, fine and performing arts, beautiful parks, recreational facilities and some of best restaurants in the world.

Since New York, Manhattan specifically, is host to culinary experts from all across the globe and has one of the most competitive and diverse restaurant scenes in the world, it will not be easy to casually predict if opening a certain restaurant/café in Manhattan will be successful or not. This is where this project makes a breakthrough in helping food business seekers to decide the best locations for their restaurant/café. So, the aim if this project is to use clustering techniques to group neighborhoods in Manhattan and support food business seekers to decide which neighborhood will be best suited for their business.

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. 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.