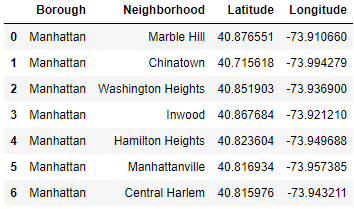
**Analyzing a suitable location for a new restaurant near Manhattan**

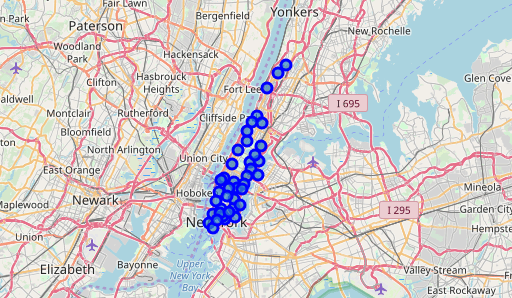
**Introduction/Business Problem**

According to the [*World Population Review*](http://worldpopulationreview.com/boroughs/manhattan-population/) 2019:  Manhattan is the most densely populated of the five boroughs in New York. It has an estimated population of 1.63 million people, all living in an area of just 23 square miles. This gives Manhattan a population density of 70,826 people per square mile, or 27,346 per square mile, which is denser than any city in the country. It is also one of the most densely populated areas on earth. Manhattan has the highest cost of living in the United States and the only county in the country with a per capita income of more than $100,000. Being heart to several iconic locations such as the Time Square, NASDAQ, the New York Stock Exchange and Wall Street, it is one of the commercial, financial and cultural centers in the world. Given all the above information, a client decides to open a restaurant near Manhattan. This project aims to find a suitable location for the new restaurant.  
  
  
**Data Acquisition and Preparation**

This analysis mainly uses a data set freely available on [*https://geo.nyu.edu/catalog/nyu\_2451\_34572*](https://geo.nyu.edu/catalog/nyu_2451_34572), which captures the neighborhood names and latitude-longitude information for all 5 boroughs and 306 neighborhoods of New York. An extract from the relevant data for the Manhattan Borough, extracted to a pandas dataframe is as follows:



A visual representation of the neighbourhoods considered are shown below, using a Folium map:

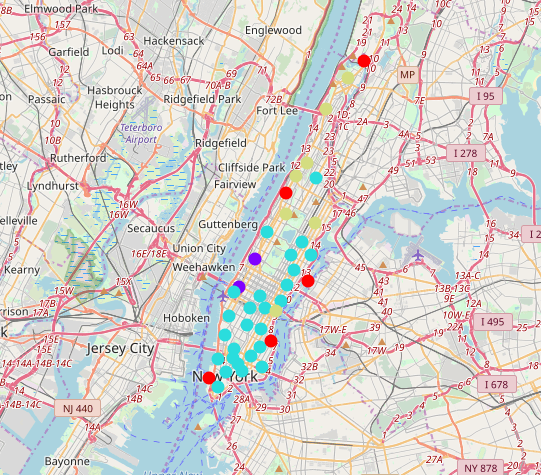


Following the initial data acquisition step above, [*Foursquare*](https://foursquare.com/) API was used to extract information on the popular venues in the neighborhoods of Manhattan which were processed into a pandas dataframe displaying the top (most frequent) 5 venues for each neighbourhood. An extract of the resulting dataframe is shown below:

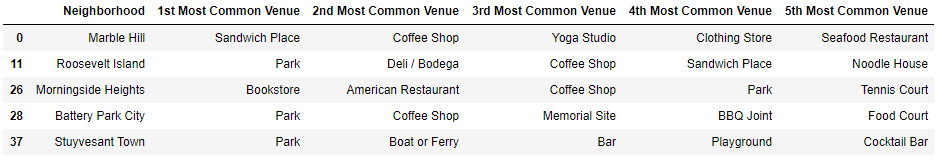


**Data Analysis**

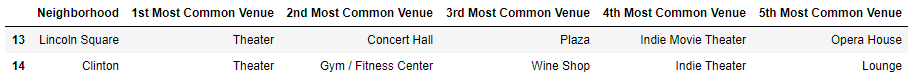
Following the data preparation phase, K-means clustering (a popular unsupervised machine learning algorithms) was used to divide the neighbourhoods of Manhattan into 4 clusters based on the similarity of venue occurrences. The visualization map and the tabulations of the 4 clusters are as follows:



**Cluster 1 (red circles)**



**Cluster 2 (purple circles)**



**Cluster 3 (light blue circles)**



**Cluster 4 (light yellow circles)**



**Conclusion**

From the above tabulations, it is evident that the two neighborhoods Lincoln Square and Clinton (in cluster 2) provide the ideal locations for a new restaurant, mainly due to the ample availability in leisure activity locations such as theatres (after visiting which, people are highly likely to dine outside), and due to the lowest competition from other restaurants.