**Applied Data Science Capstone Project - The Battle of Neighbourhoods Report**

**Introduction:**

This project will analyse the New York City data. The City of New York, is the most populous city in the United States. It is diverse and is the financial capital of USA. It is multicultural. It provides lot of business opportunities and business friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce.

The city is a major center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in the United States. This also means that the market is highly competitive. As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analysed carefully.

The insights derived from analysis will give good understanding of the business environment which helps in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

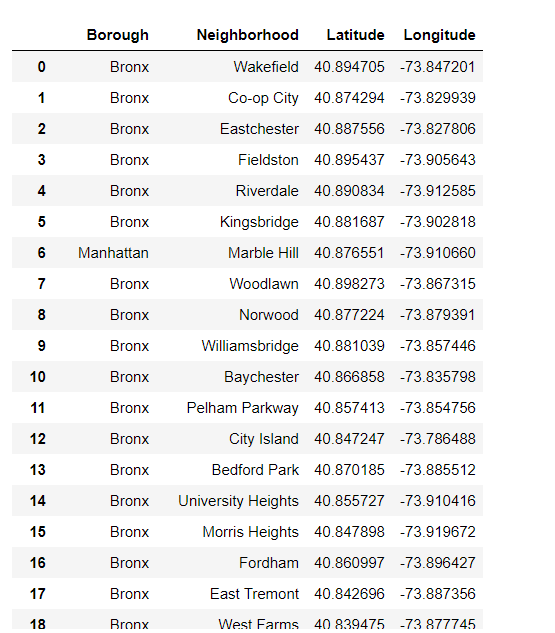
**Problem Statement:**

The City of New York is famous for its excellent cuisine. Its food culture includes an array of international cuisines influenced by the city's immigrant history. Sushi restaurants have become so popular in the United States now it seems that there is one on every corner, not only in major cities but also in smaller cities. Starting a sushi restaurant can be a great business opportunity, but you need to distinguish yourself from others to enjoy long-term success.

If you plan a real restaurant that can demand higher prices for fresh fish, delivered daily from Japan, focus on neighborhoods and outlets that already attract a sophisticated Japanese client. If you plan a cheap buffet restaurant, points to the masses looking for affordable high-traffic locations with large shopping centers and other local points of interest.

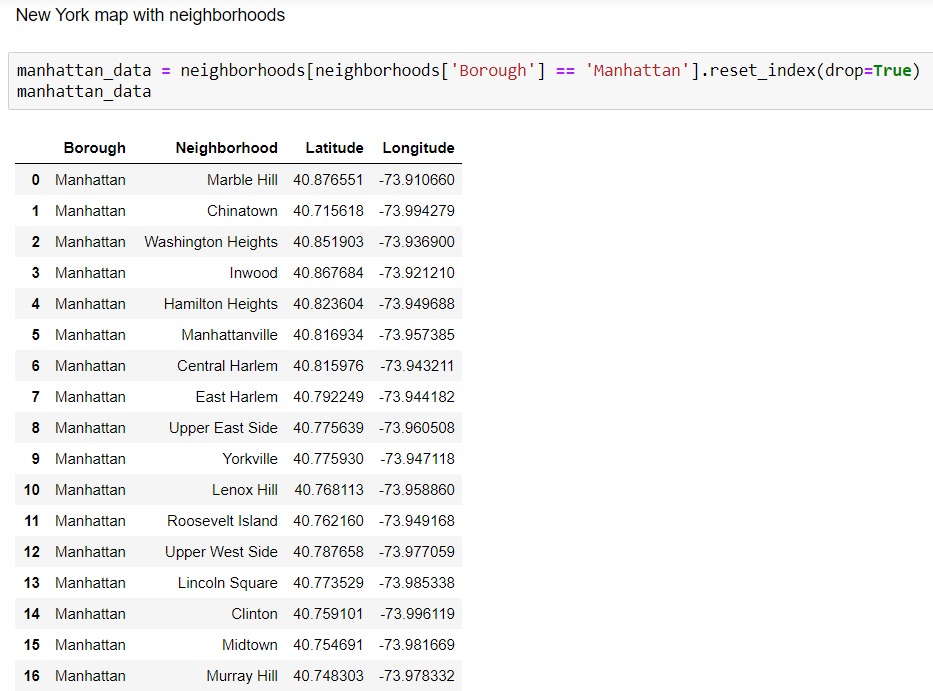
**Data Section:**

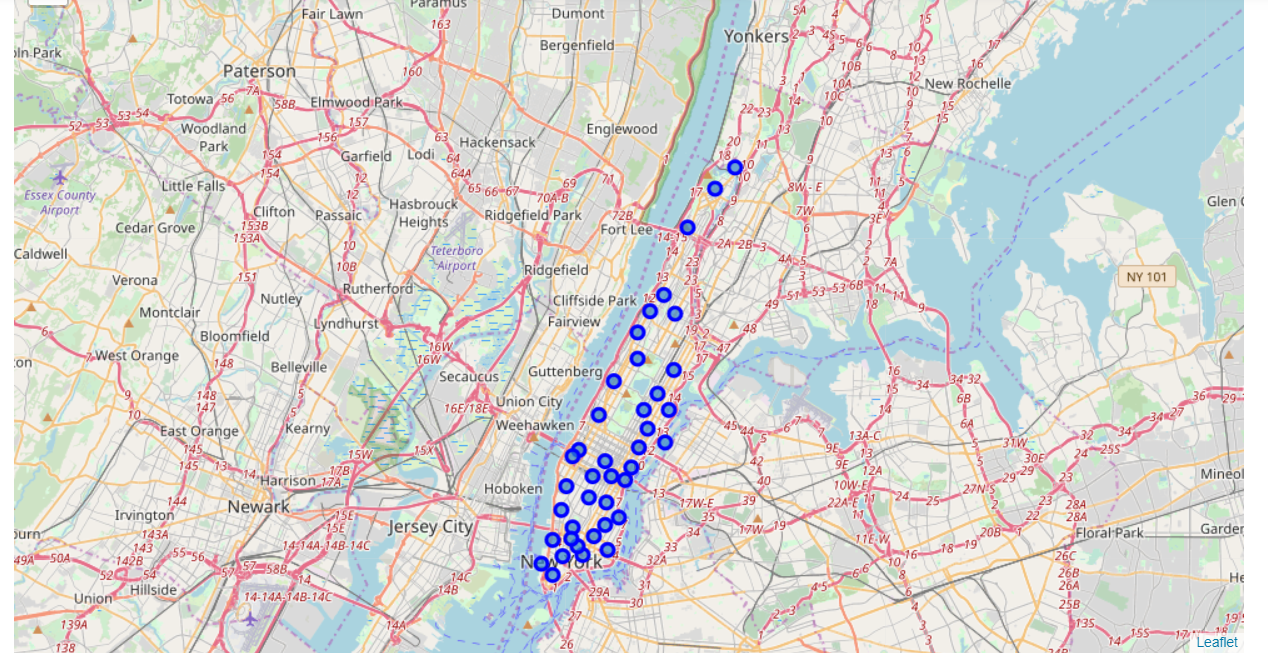
The downloaded dataset link here: <https://geo.nyu.edu/catalog/nyu_2451_34572>. It contains 306 rows × 4 columns of details like ID, Borough, Neighbourhood, Latitude, and Longitude.



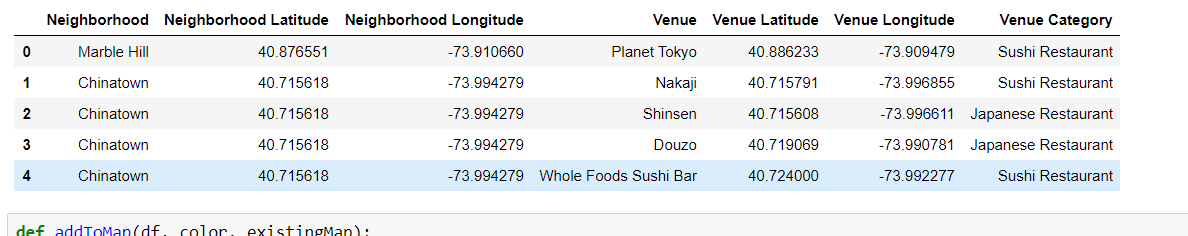
Newyork City geographical coordinates data will be utilized as input for the Foursquare API that will be leveraged to provision venues information for each neighbourhood. We will use the Foursquare API to explore neighborhoods in New York City. The below is image of the Foursquare API data. In addition, Sushi category Id 4bf58dd8d48988d1d2941735 is used for retrieving data from Foursquare API.

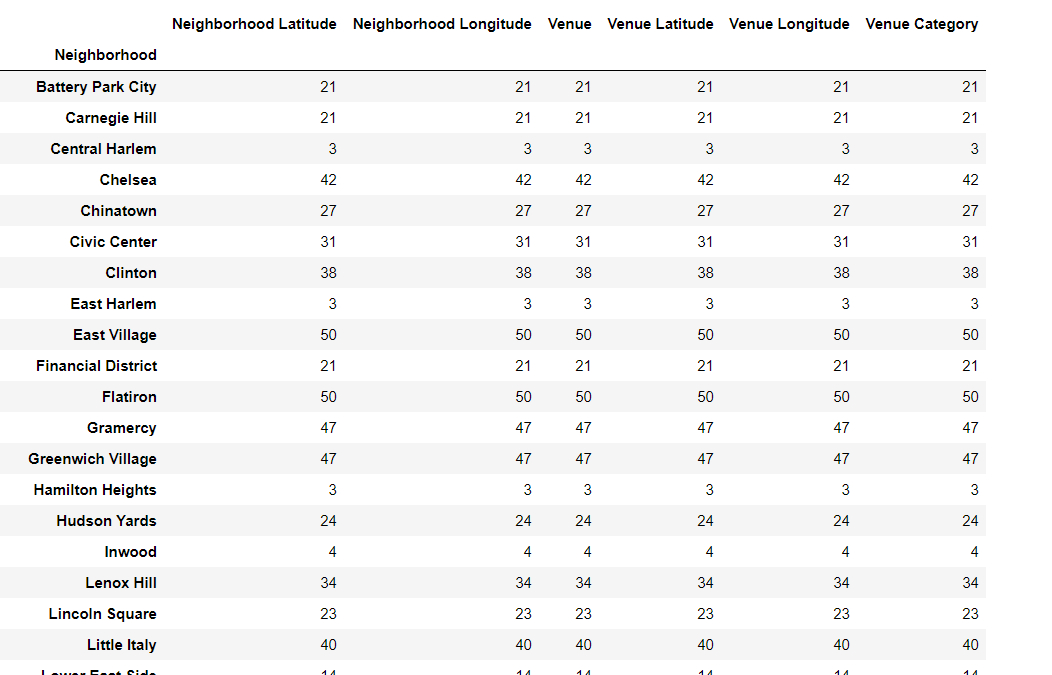
**Methodology:**



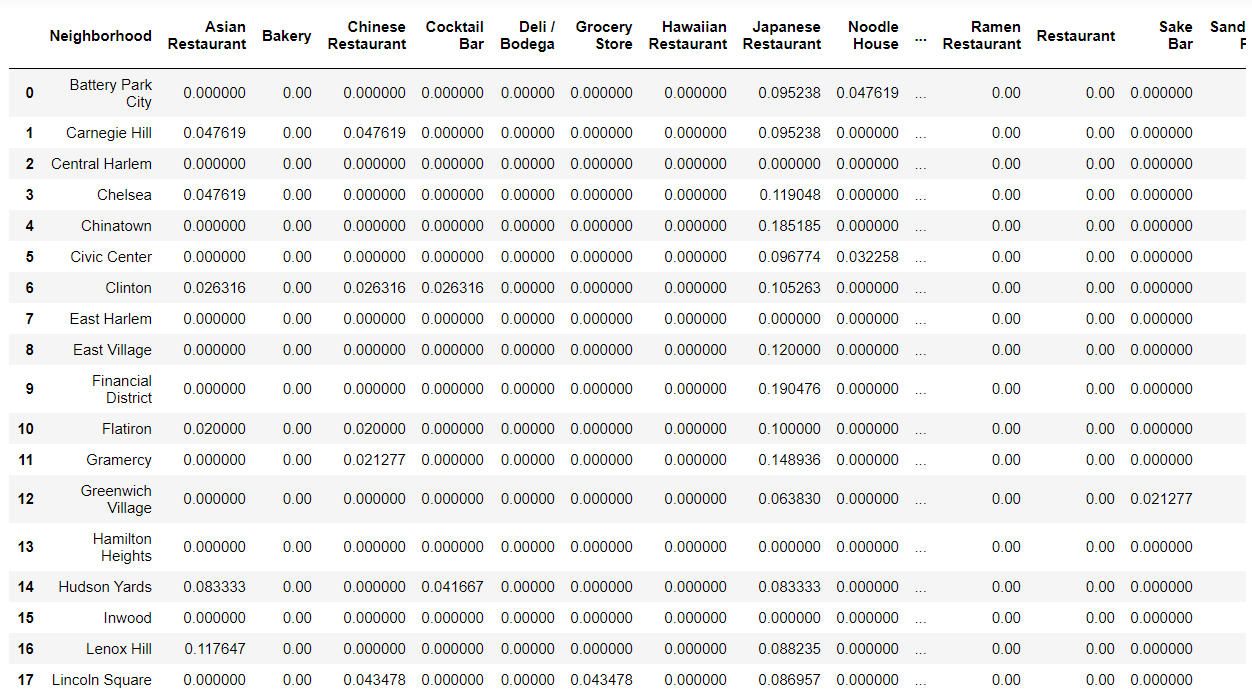


Above, I have done convert addresses into their equivalent latitude and longitude values. Then we will use the Foursquare API to explore neighborhoods in Manhattan, New York. After that, explore function to get sushi restaurant categories in each neighborhood.





Then use this feature to group the neighborhoods into clusters K-means clustering algorithm will be use to complete this task. And also, the Folium library to visualize the neighborhoods in Manhattan and its emerging clusters.



**Results:**

K-mean Cluster Using K-mean to clustering data area with less number of sushi bars

**Cluster 0**



**Cluster 1**



**Cluster 3**



**Cluster 4**



Based on data frame analysis above Cluster 3 and Cluster 4 areas are the best places to open a new sushi bar business.

**Discussion:**

In this section, I would be discussing the observations I have noted and the recommendation that I can make based on the results. This analysis is performed on limited data. This may be right or may be wrong. But if good amount of data is available there is scope to come up with better results.

There is high competition in Midtown and Soho so it is very risky to open business in these areas. Central Harlem has also potential where closes to Morningside Heights area.

It can be done more detailed analysis by adding other factors such as transportation, demographics of inhabitants.

Finally, Four Square proved to be a good source of data but frustrating at times. Despite having a Developer account I regularly exceeded my hourly limit locking me out for the day.

**Conclusion:**

Although all of the goals of this project were met there is definitely room for further improvement and development as noted below. However, the goals of the project were met and, with some more work, could easily be developed into a fully pledged application that could support the opening a business idea in an unknown location.

As per the neighbourhood or restaurant type mentioned like Sushi restaurants analysis can be checked. A venue with lowest risk and competition can be identified.