**The Battle of the Neighborhoods**

1. **Introduction and Business Problem :**

**Background:**

New York City is one of the most populated cities in the United States. It is diverse, mulitcultural and known as the financial capital of America. There are many business opportunities and it is a business friendly environment . It has attracted many players into the market and is a global hub for business and commerce. It is a major center for banking and finance, retail, world trade, transportation, tourism, real estate, new and traditional media, advertising, legal services, accountancy, insurance, entertainment, fashion and the arts in the United States. This also means that the market is highly competitive and the cost of doing business is also very expensive. Hence, when it comes to a business venture or expansion, it needs to be analyzed carefully to reduce risk and help return on investment be reasonable . The insights derived from analysis will give a good understanding of the business environment which help in strategically targeting the market.

**Problem Description:**

A restaurant is a business which prepares and serves food and drinks to customers in return for money, either paid in before the meal, after the meal, or with an open account. New York City is famous for its excellent cuisine. Its food culture includes an array of international cuisines influenced by the city’s immigrant history.

1. Central and Eastern European immigrants, especially Jewish immigrants - bagels, cheesecakes, hot dogs, knishes, and delicatessens.
2. Italian immigrants - New York style pizza and Italian cuisine.
3. Jewish Immigrants and Irish immigrants - pastrami and corned beef.
4. Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city.
5. There are about 4000 licensed mobile food vendors in the city.
6. Middle Eastern foods such as falafel and kebabs are examples of modern New York street food.
7. The city is also known for its fine dining Michelin starred restaurants and is home to “nearly one thousand of the finest and most diverse haute cuisine restaurants in the world” according to Michelin.

So it is evident that in order to survive in such a competitive market, it is important to strategically plan. Various factors need to be studied in order to decide on the Location such as:

* New York City population
* New York City demographics
* Are there any Farmers markets., wholesale markets, etc. nearby so that the ingredient can be purchased fresh to maintain quality and cost?
* Are there any venues like Gyms, Entertainment zones, parks, etc. nearby where the floating population is high?
* Who are the competitors in that location?
* Cuisine served/ menu of the competitors
* Segmentation of the Borough

The list can go on. Even the well funded ABC Company Ltd. needs to choose the correct location to start its first venture. If this is successful, they can replicate the same in other locations. First step in setting up a venture is very important, hence location holds a heavy importance.

**Target Audience:**

The objective is to locate and recommend to the management which neighborhood of New York City will be the best choice to open a restaurant. The management also expects to understand the rationale of the recommendations made. This would interest anyone who wants to start a new restaurant in New York City.

**Success Criteria:**

The success criteria of the project will be a good recommendation of the borough/neighborhood choice to ABC Company Ltd. based on lack of such restaurants in that location and nearest suppliers of ingredients.

**2. Data:**

One city will be analyzed in this project: New York City.

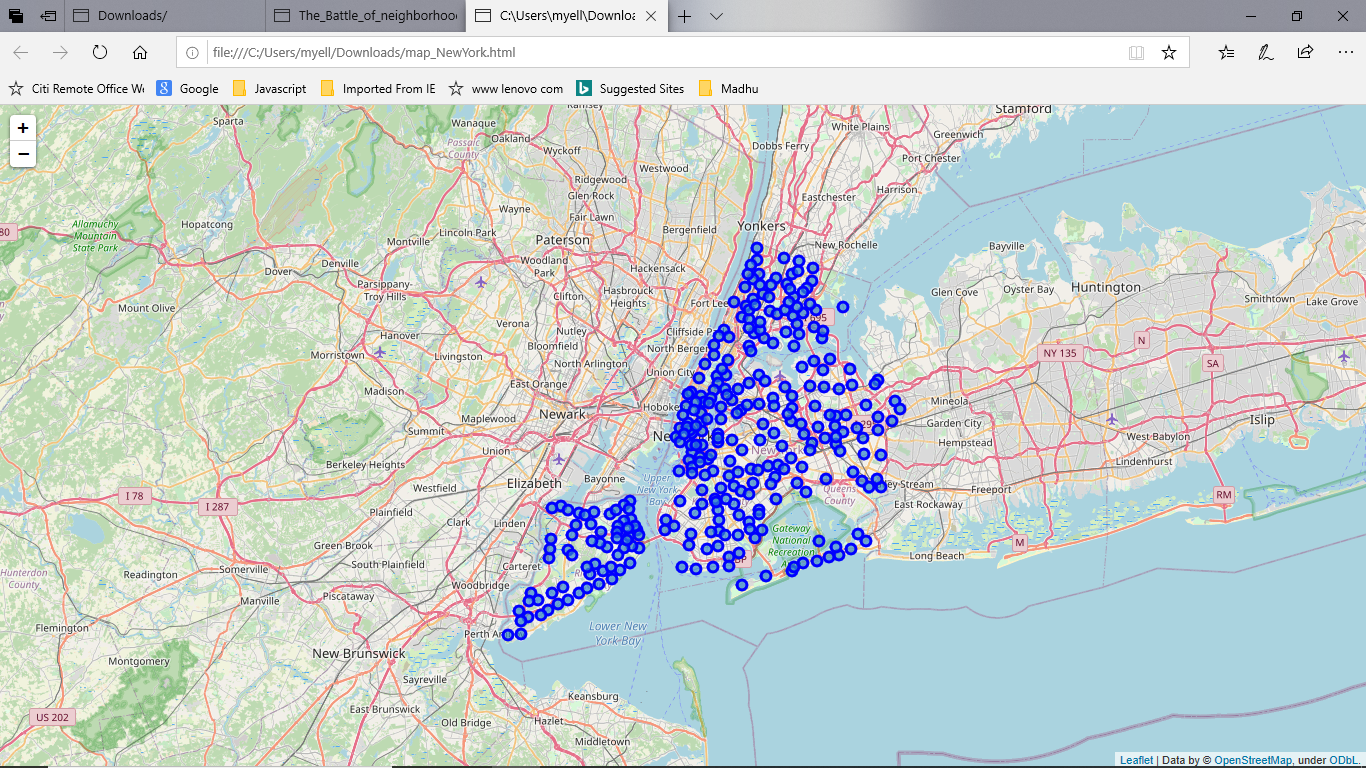
We will be using the below datasets for analyzing New York City:

**Data 1:**

Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

This dataset exists for free on the web. The link is below:

<https://geo.nyu.edu/catalog/nyu_2451_34572>



**Data 2:**

For the below analysis, we’ll be getting data from Wikipedia and the City of New York website. They will be covering:

1. New York Population
2. New York Demographics
3. Cuisine of New York City

* <https://en.wikipedia.org/wiki/New_York_City>
* <https://en.wikipedia.org/wiki/Demographic_history_of_New_York_City>
* <https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j/data>

**Data 3:**

The New York City geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venue information for each neighborhood. We’ll be using the Foursquare API to explore neighborhoods in New York City.

**3. Methodology:**

**Business Understanding:**

Our main goal is to get an optimum location for a new restaurant business in New York City for ABC Company Ltd.

**Analytic Approach:**

New York City neighborhood has a total of 5 boroughs and 306 neighborhoods. In this project, the first part is clustering Manhattan and Brooklyn. The second part is clustering Bronx, Queens and Staten Island.

**Exploratory Data Analysis :**

**Data 1: New York City Geographical Coordinates Data.**

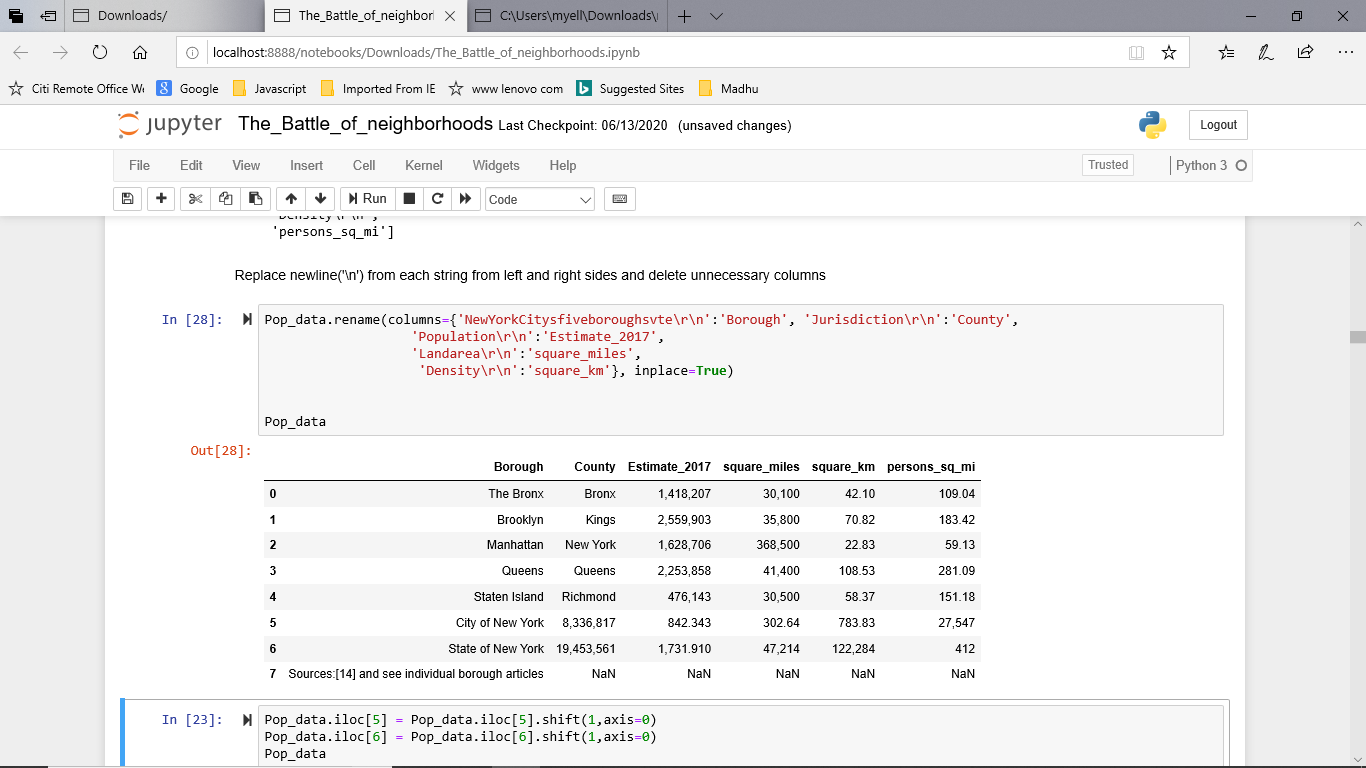
1. In this we load the data and explore data from newyork\_data.json file
2. Transform the data of nested python dictionaries into a pandas dataframe
3. This dataframe contains the geographical coordinates of New York City neighborhoods.
4. This data will be used to get Venues data from Foursquare.
5. We used geopy and folium libraries to create a map of New York City with neighborhoods superimposed on top

**Data 2:**

To analyze New York City population, demographics and cuisine, scrapped the data from Wikipedia pages given above in the data section. We used the BeautifulSoup python library. BeautifulSoup in python package for parsing HTML and XML documents. It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping.

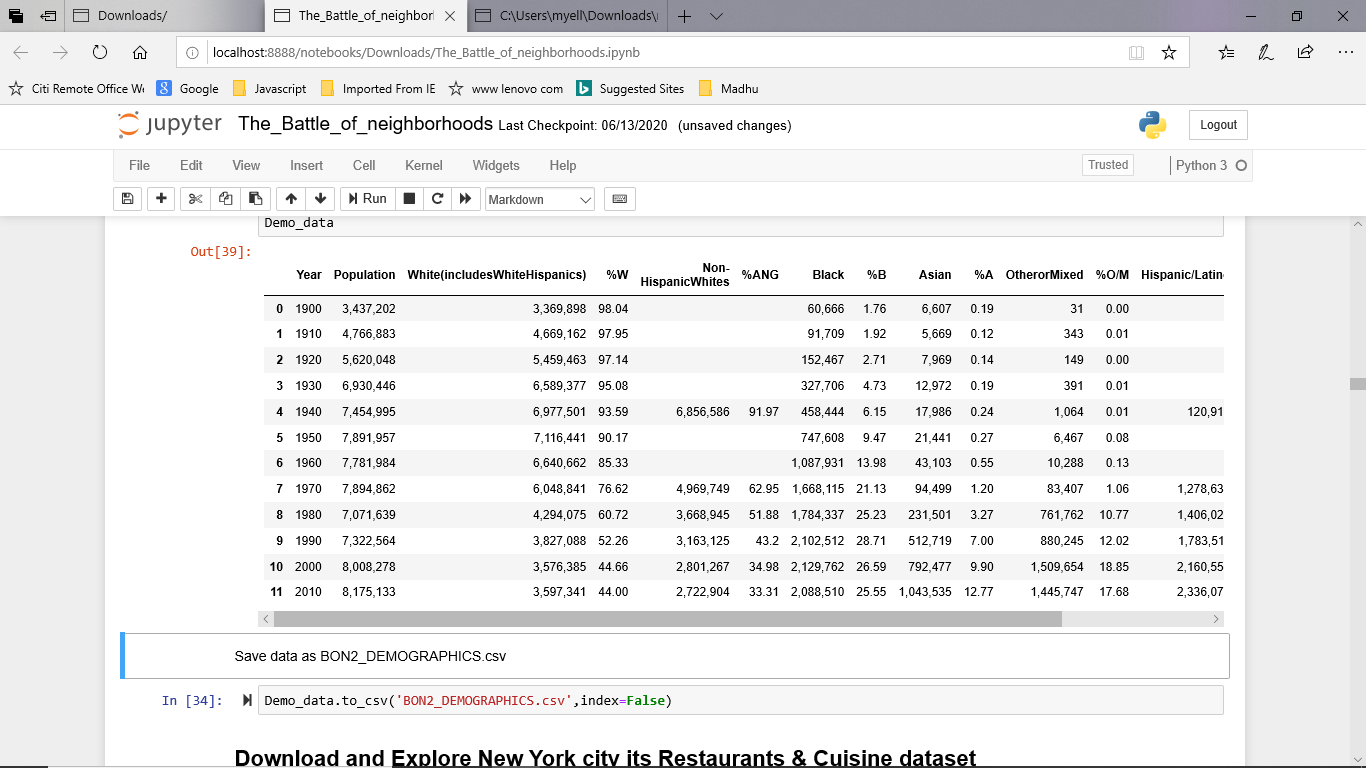
1. NYC Population, Insights from the data:

* Manhattan is geographically the smallest but most densely populated borough.
* Manhattan’s population density of 368,500 people per square mile in 2017 makes it the highest of any county in the United States and higher than the density of any individual American city.
* Brooklyn, on the western tip of Long Island, is the city’s most populous borough.
* Queens, on Long Island north and east of Brooklyn is geographically the largest borough.



1. NYC Demographics:

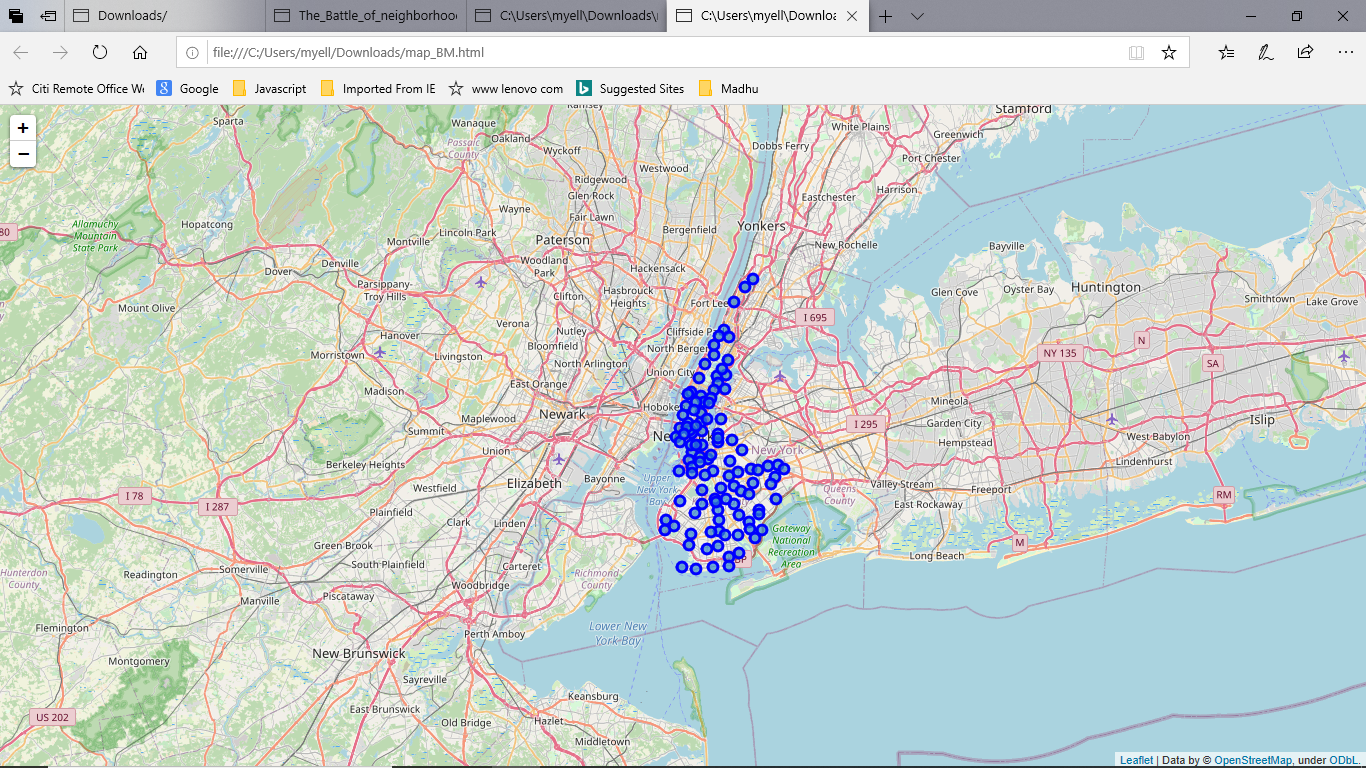
New York City is the most populous city in the United States, with an estimated record high of 8, 336,817 residents as of 2017, incorporating more immigration into the city than outmigration since the 2010 United States Census. The racial composition is as given below. This is the reason New York City has restaurants serving cuisine from many countries such as Indian, African , Japanese, etc. This also increases the scope for restaurant business in New York City.



Data 3:

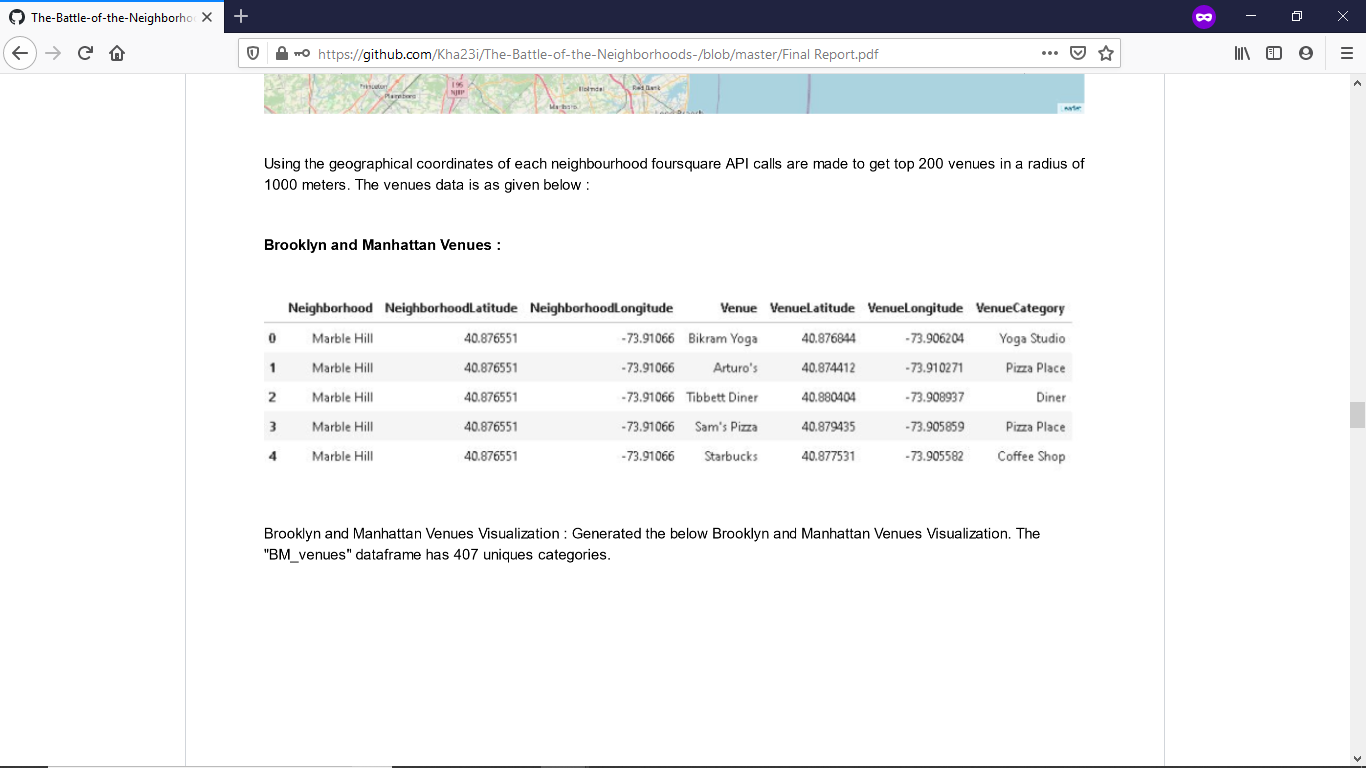
New York City geographical coordinates data has to be utilized as input for the Foursquare API, that has been leveraged to provision venues information for each neighborhood. We used the Foursquare API data to explore neighborhoods in New York City.

**Brooklyn and Manhattan:**

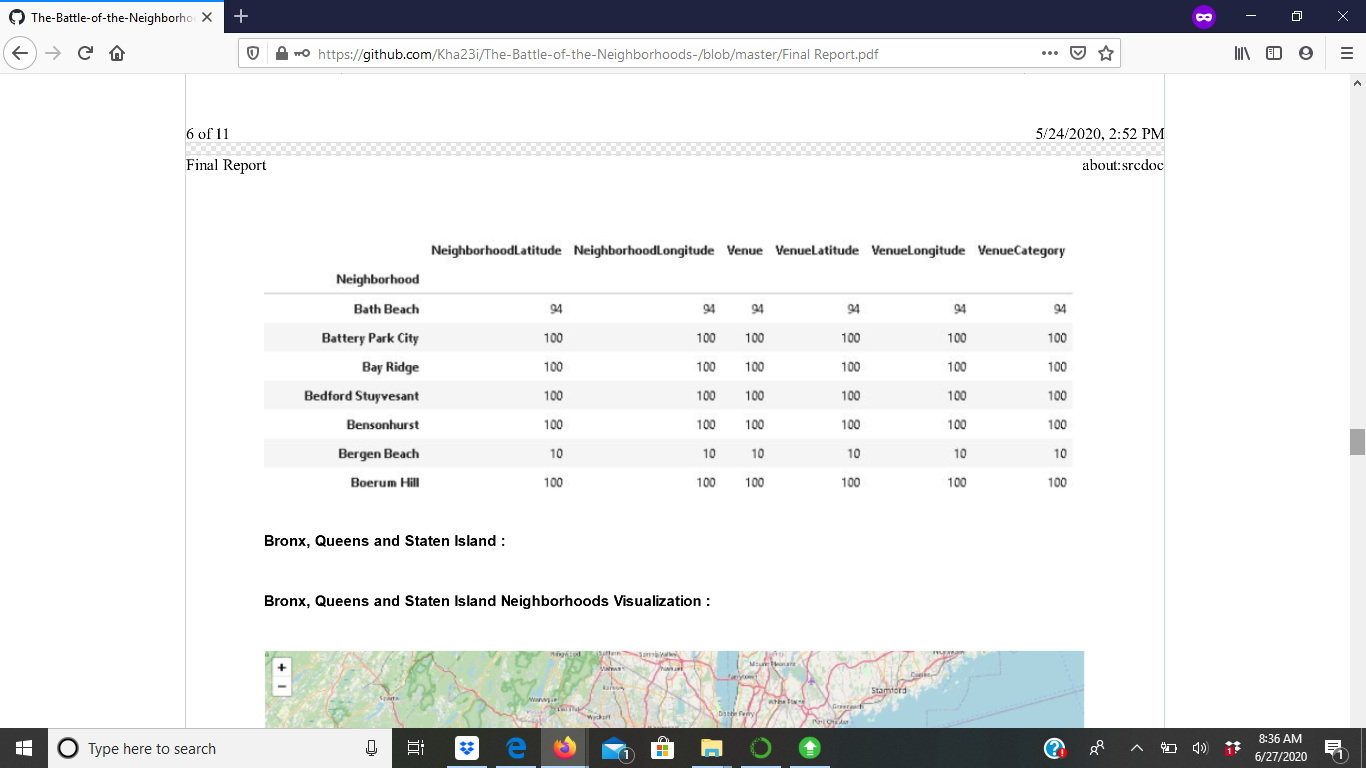


Using the geographical coordinates of each neighborhood foursquare API calls are made to get top 200 venues in a radius of 1000 meters.

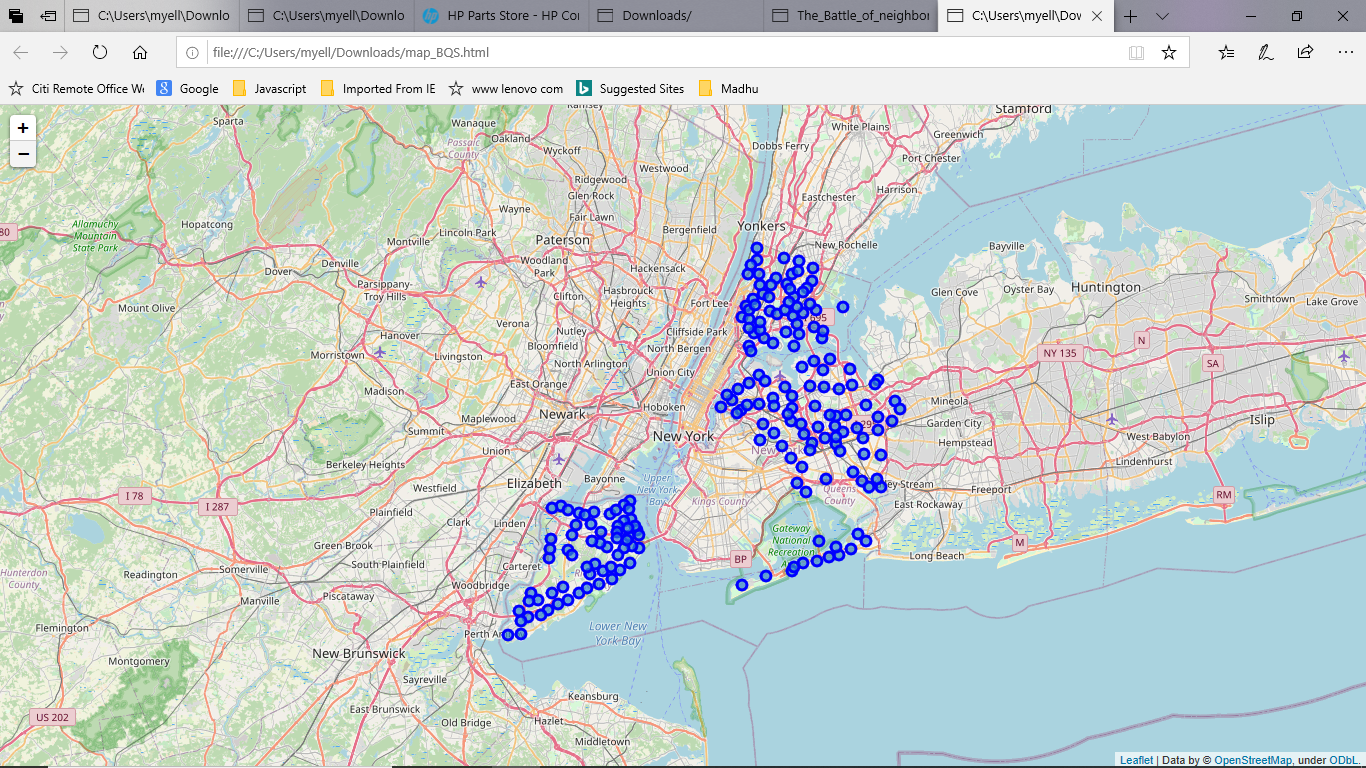
**Brooklyn and Manhattan Venues:**



Brooklyn and Manhattan Venues Visualization: The “BM\_Venues” dataframe has 407 unique categories.

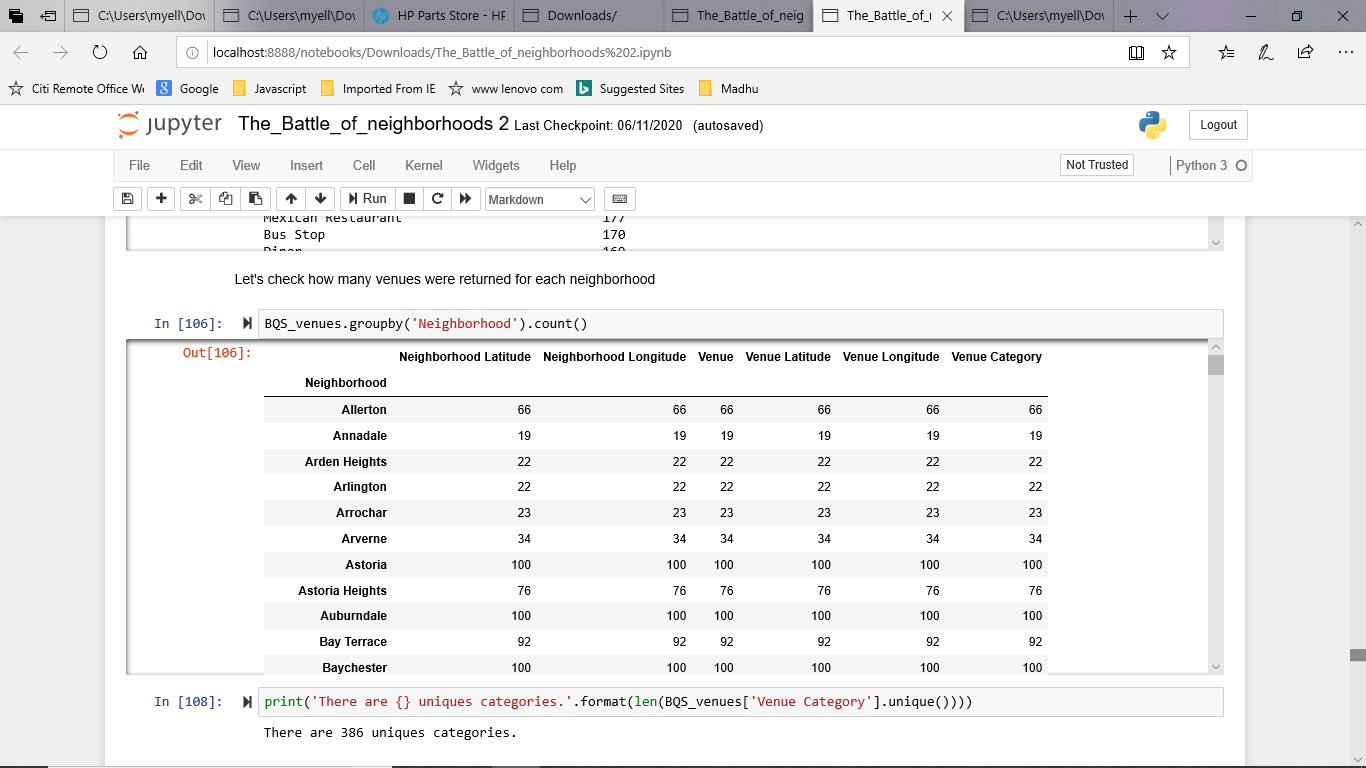


Bronx, Queens and Staten Island:



Bronx, Queens and Staten Island Venues Visualization: The “BQS\_venues” dataframe has 386 categories.

**Bronx, Queens and Staten Island Venues:**



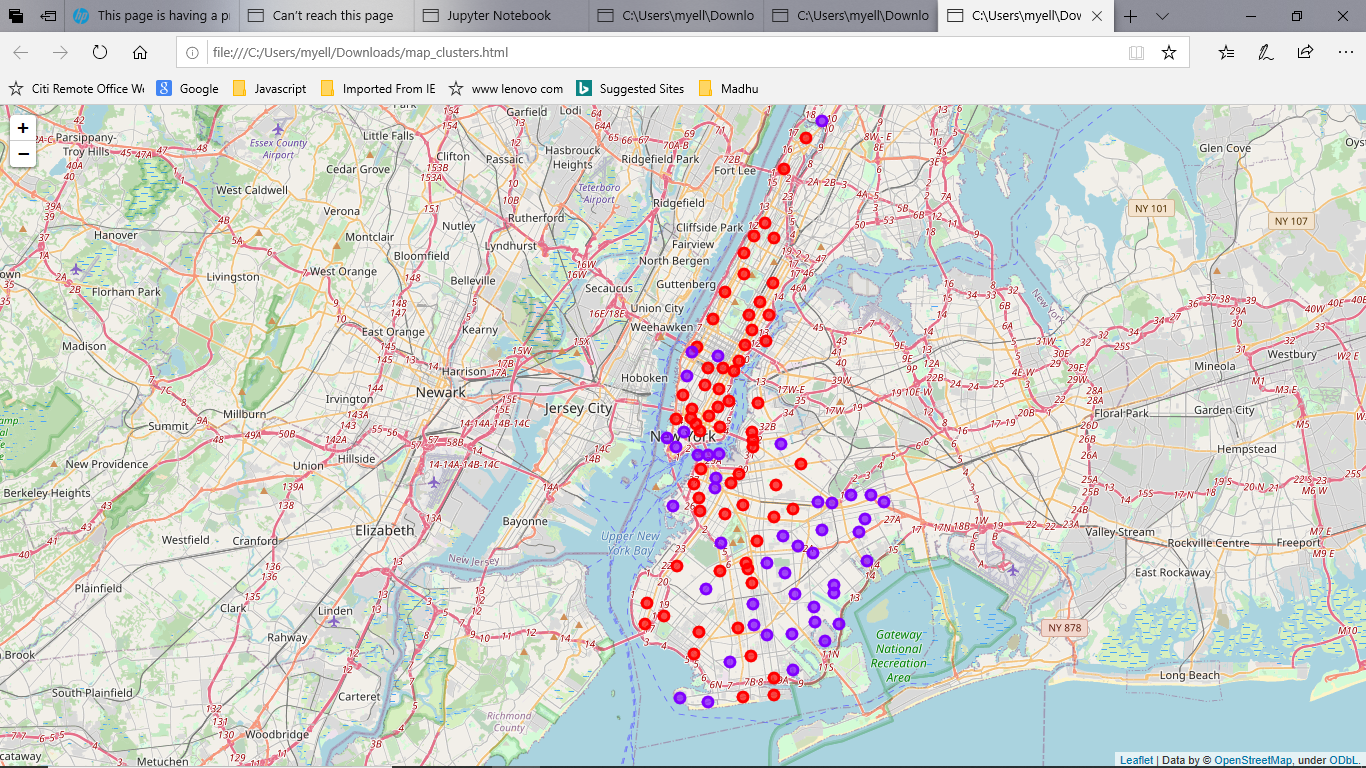
**4. RESULTS:**

From this venues data, we filtered and used only the restaurant data for Brooklyn and Manhattan clustering and Bronx, Queens and Staten Island clustering, as we focused only on the restaurant business.

**Neighborhood and K-means clustering based on mean occurrence of venue category:**

To cluster the neighborhoods into two clusters, we used the K-means clustering algorithm. K-means clustering aims to partition n observations into k clusters in which, each observation belongs to the cluster with the nearest mean, It uses an iterative refinement approach.

**Brooklyn and Manhattan:**

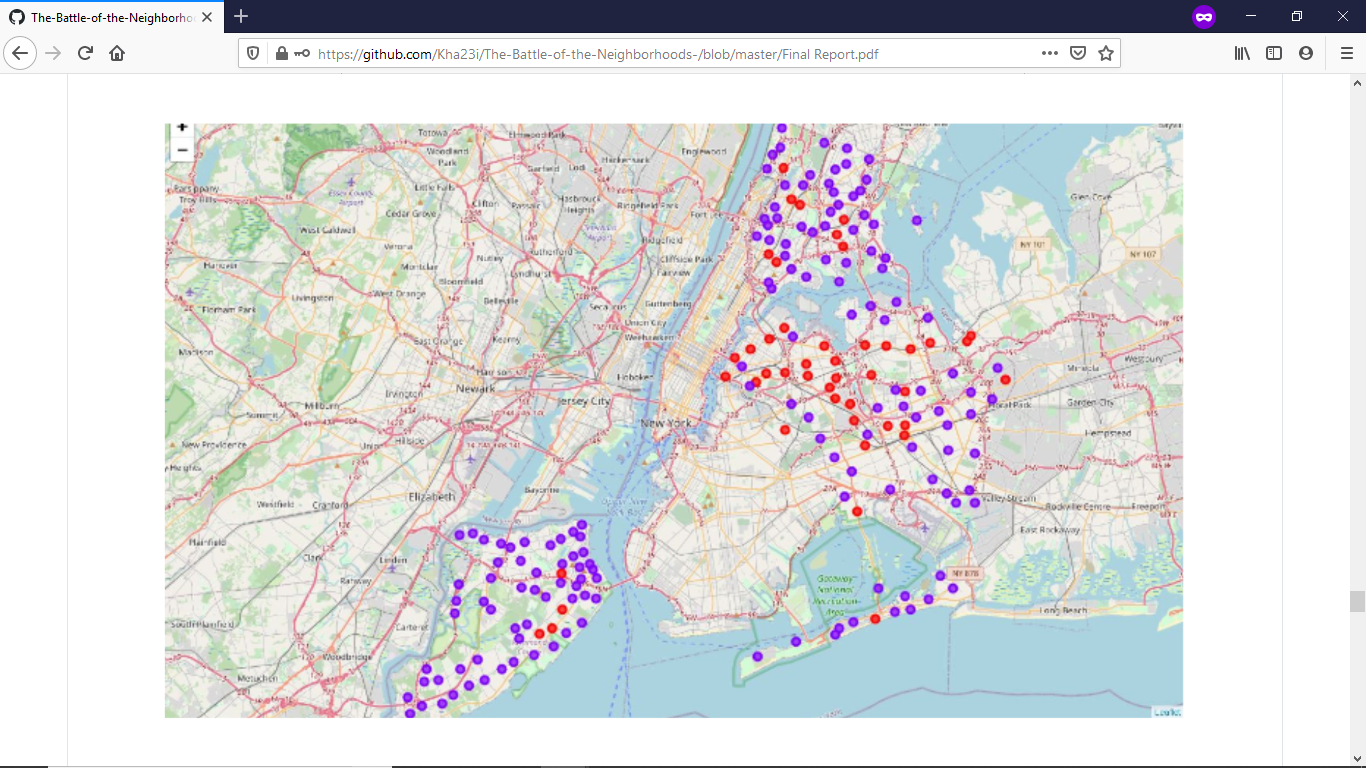


**Cluster0:** The Total and Total Sum of cluster0 has the smallest value. It shows that the market is not saturated.

**Cluster1:**  The Total and Total Sum of cluster1 has the highest value. It shows that the markets are saturated. The number of restaurants is very high.

There are no untapped neighborhoods in Brooklyn and Manhattan.

**Bronx , Queens and Staten Island:**



**Cluster0:** The Total and Total Sum of cluster0 has the smallest value. It shows that the market is not saturated.

**Cluster1:**  The Total and Total Sum of cluster1 has the highest value. It shows that the markets are saturated. The number of restaurants is very high.

**5. Discussion:**

1. There is scope to increase Farmers markets in Bronx, Queens and Staten Island.
2. There is scope to explore cuisines of various countries in Bronx, Queens and Staten Island.
3. In Manhattan and Brooklyn restaurants of cuisines of many countries are available. So risk can be taken with a great menu on board. It also shows people love eating cuisines of various countries.

**6. Conclusion:**

This analysis is performed on limited data. This may be right or may be wrong. But if a good amount of data is available, there is scope to come up with better results. If there are alot of restaurants, then there must be alot of demand. Brooklyn and Manhattan have a high concentration of restaurant business and a very competitive market. Bronx, Queens, and Staten Island also have a good number of restaurants, but not as many compared to Manhattan and Brooklyn. This can be further explored. As per the neighborhood or restaurant type, for example, say Indian restaurant analysis can be checked. A venue with lowest risk and competition can be identified.