**Introduction**

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 provides lot of business opportunities with lot of income sources through profits. It has got so many different kinds of businesses into the market. It is a economical hub of business and commerce. The city is a major center for banking and finance, markets, world trade, transport, tourism, culture, living, theater, concerts, **gymnasiums** and the arts in the United States. This vast market is highly competitive for new comers. As it is highly costly environment the ease of business is also costly. Thus, any new business needs to be check and take proper descions. The data derived from analysis will give good vision for the future business development which helps in reduction of risk. And the Return on Investment will be reasonable.

**Business problem**

The City of New York has a numerous population and much part of the population’s combination of teenagers and middle aged people. With the increase in health problems due to lack of fitness people are interested of joining in gyms and develop physical fitness. So by this people need a good well equipped gym in the city. This is a good start of any new comer to introduce to the market, so we have to find out a suitable lace for the gym.

**To Find a suitable location**

To be a business successful ,we have to find a best place to start and establish a gym the place where the environment is so nice to people come out of the traffic noises and increase their fitness levels so the place should get in this condition and it should be a proper safety place in the city

**Target Audience:**

My Client wants to open his business in NewYork , so I only focus on that borough during my analysis. The main aim is to find and suggest to the management which neighborhood of Newyork city will be best option for the gym.

**Data Data 1:**

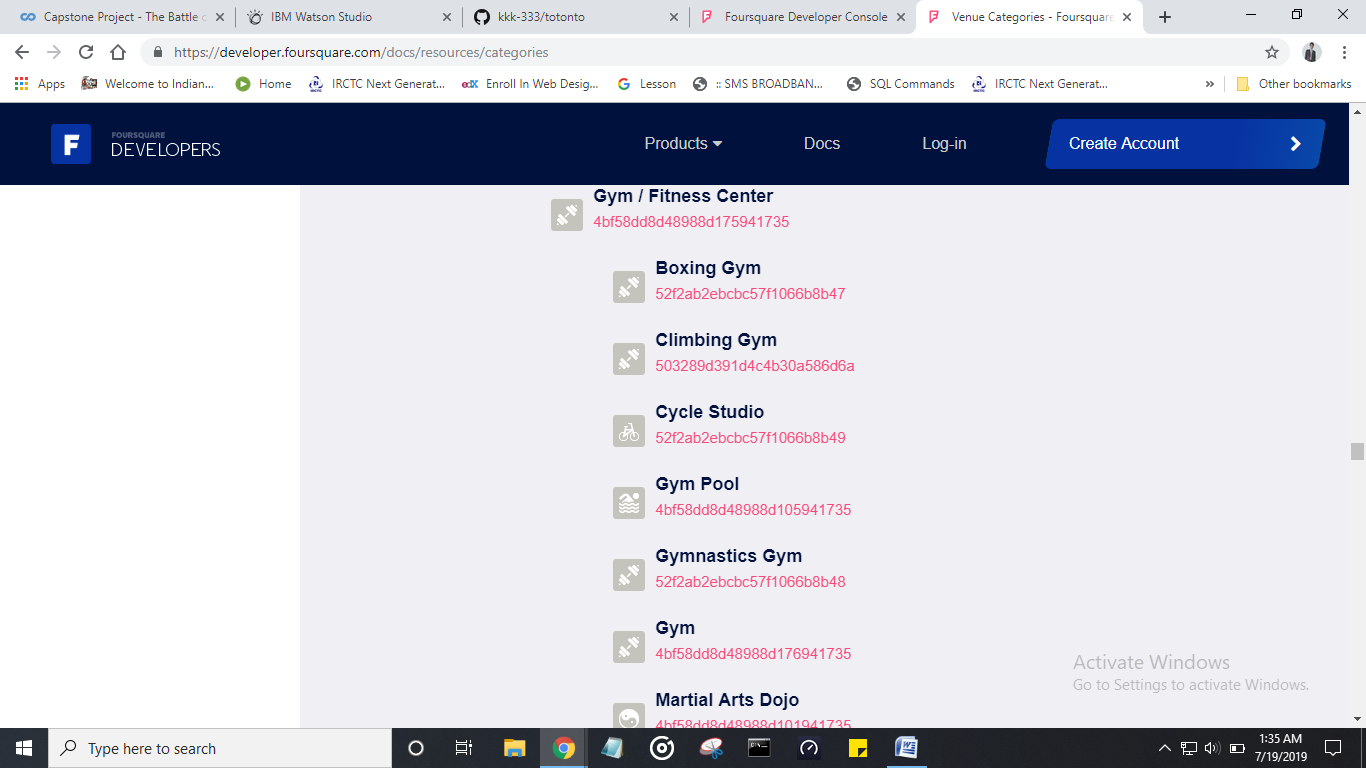
Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to divide the neighborhoods and search over them, we will mainly need a dataset that contains the 5 boroughs and the neighborhoods that contain each borough with the latitude and longitude coordinates of each neighborhood. This dataset exists for free on the web. Link to the dataset is:

https://geo.nyu.edu/ catalog/nyu\_2451\_34572.



**Data2:**

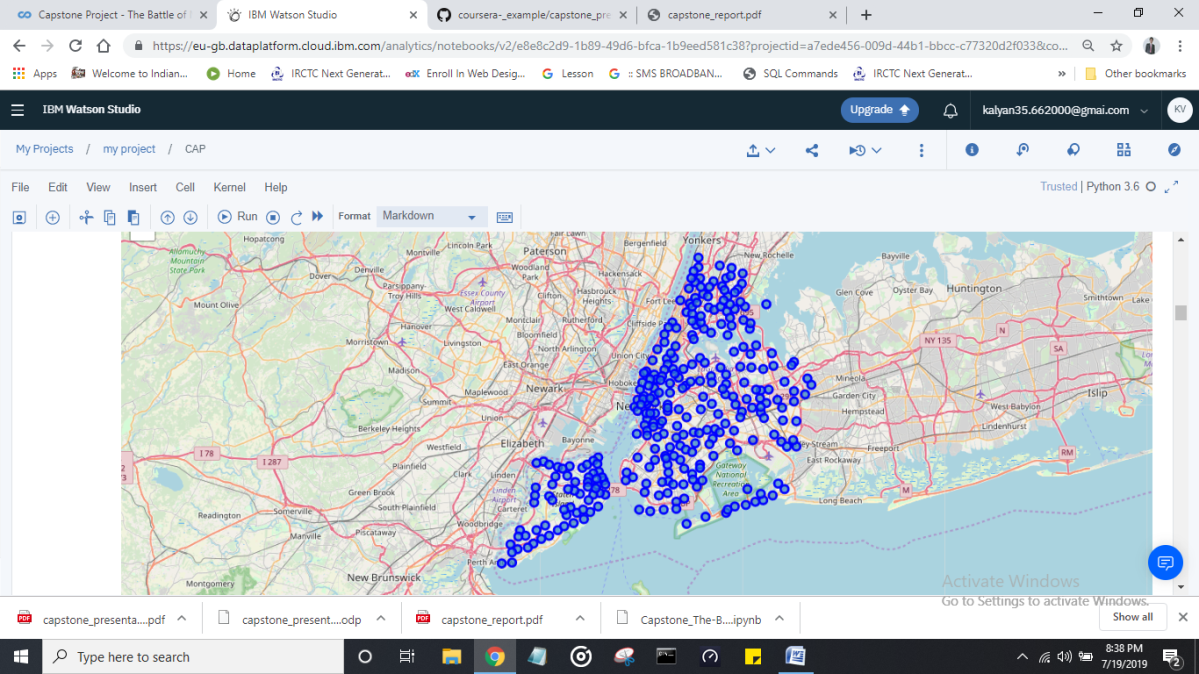
New York City geographical coordinate’s data will be utilized as input for the Foursquare API that will be helpful in getting venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in New York City



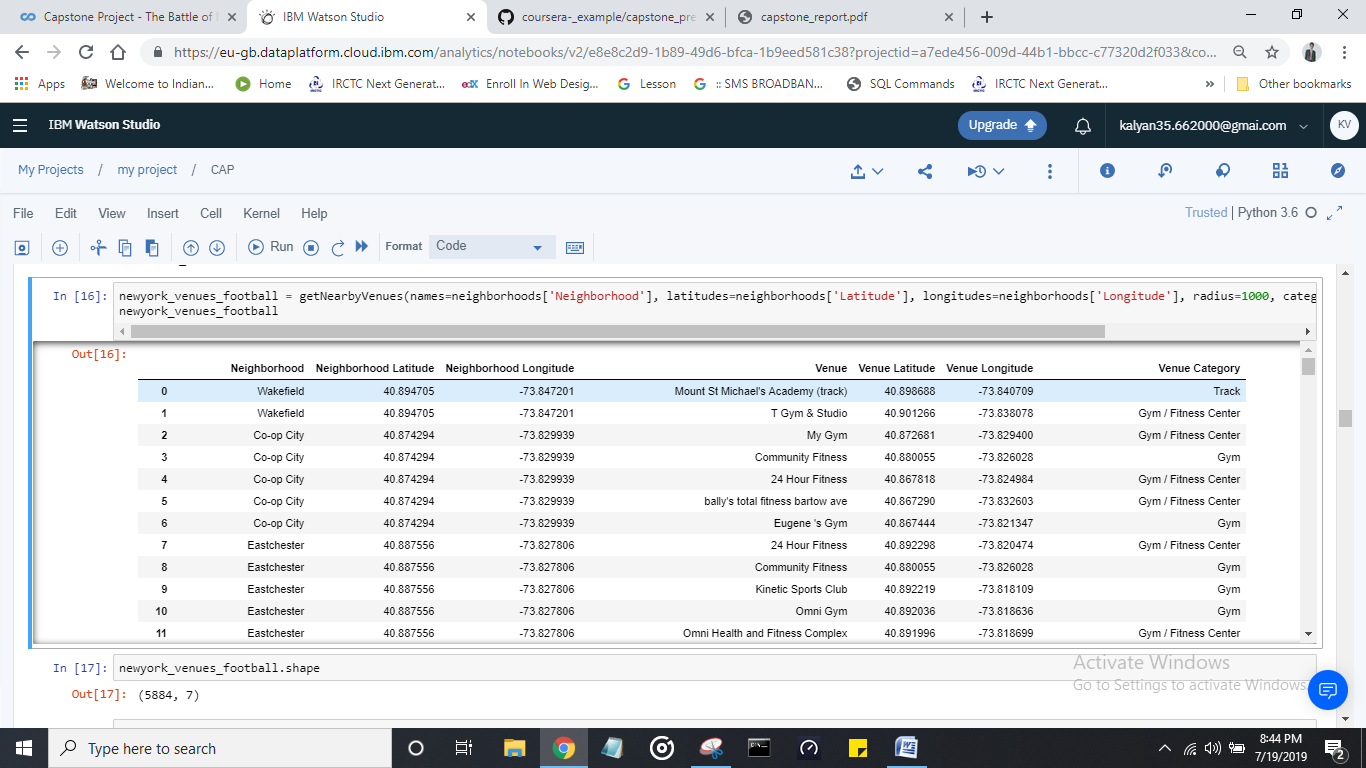
By this data we can get the places of the gyms already existing and by that we can get to cear where should we lace our new gymnasium.

**Methodology:** First we will convert data from the various resources into a data frame,

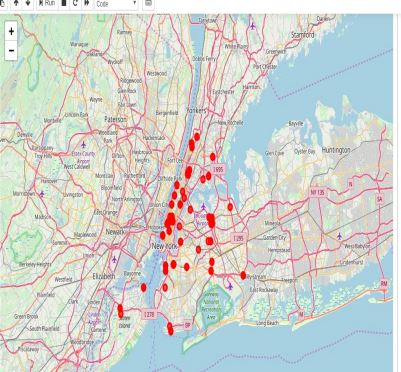
The data frame looks like this



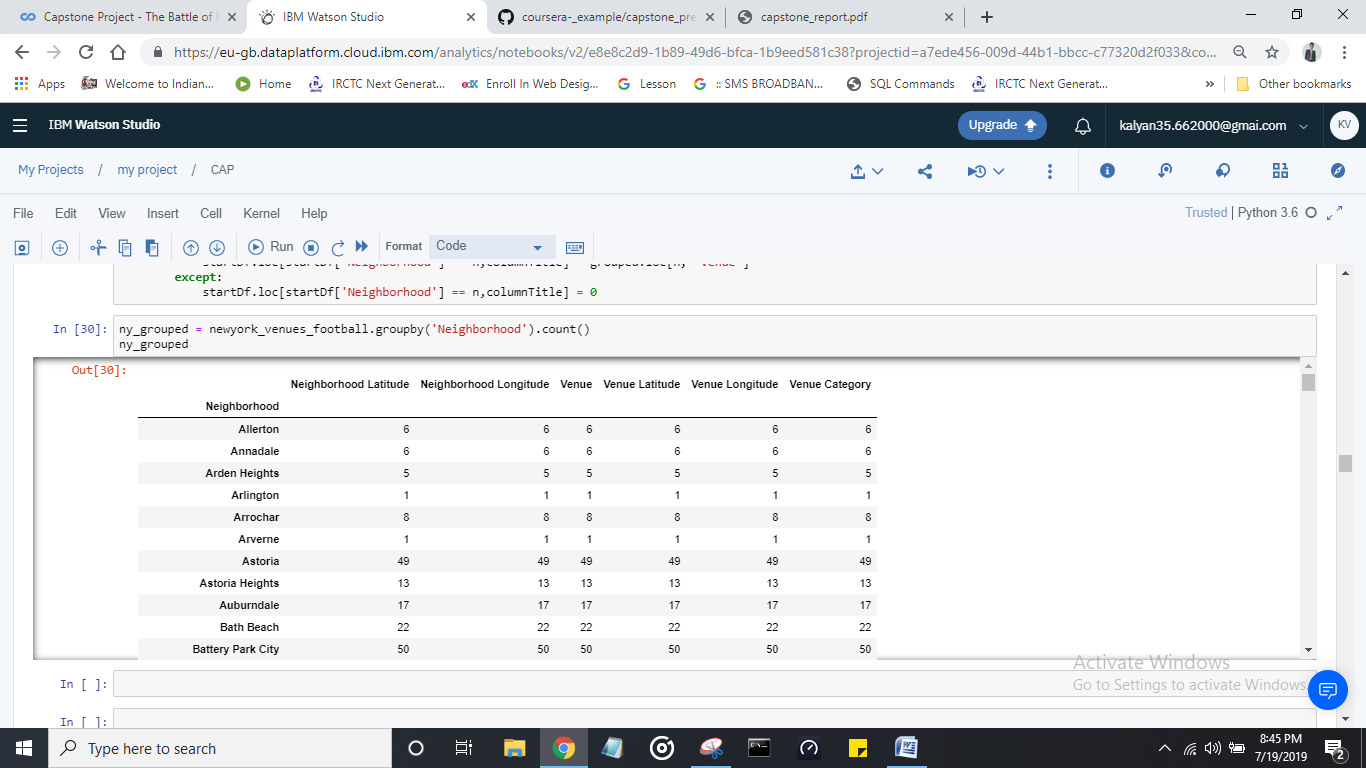
A map with all the neighborhoods imposed on the map of new York



This data is retrived from the foursquare api where we got the different gym facilities located at different localities.

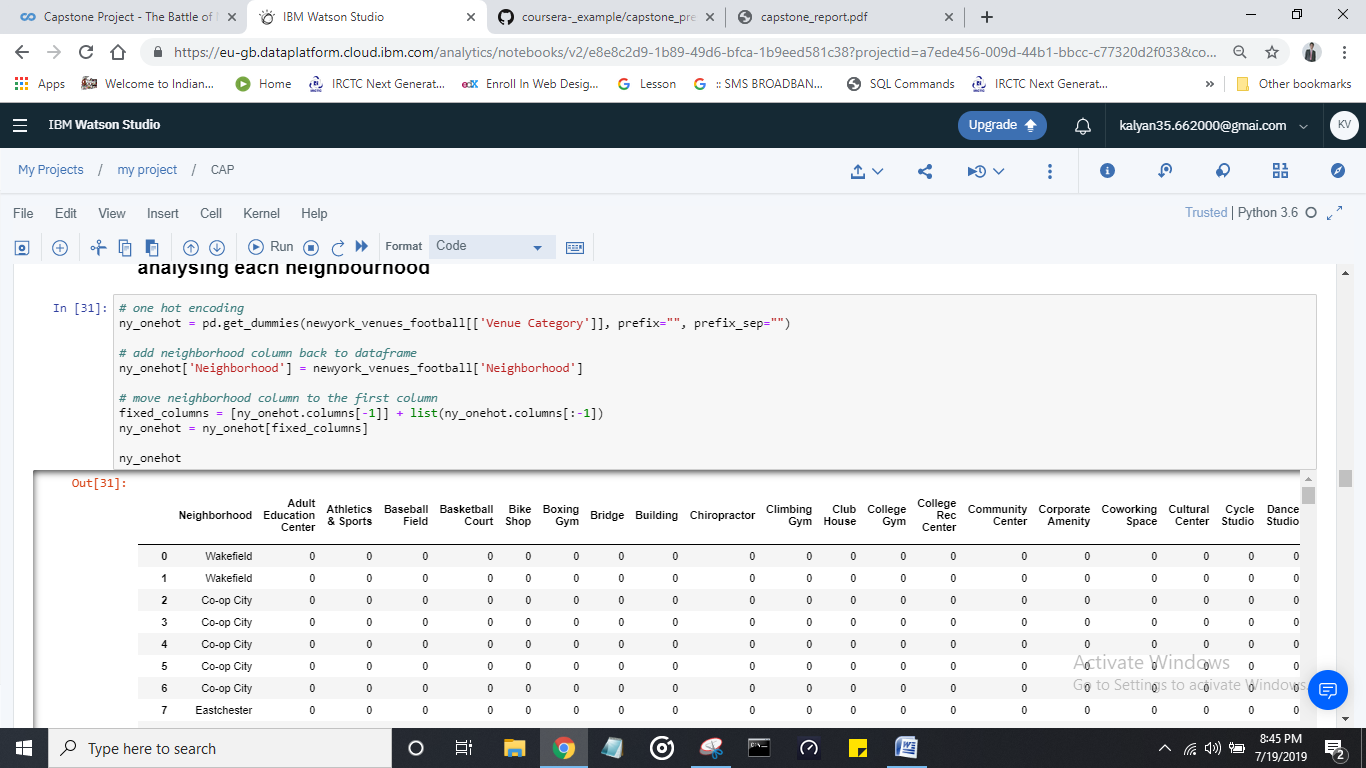


This is the map of New York superimposing the places where the gym facilities are present.



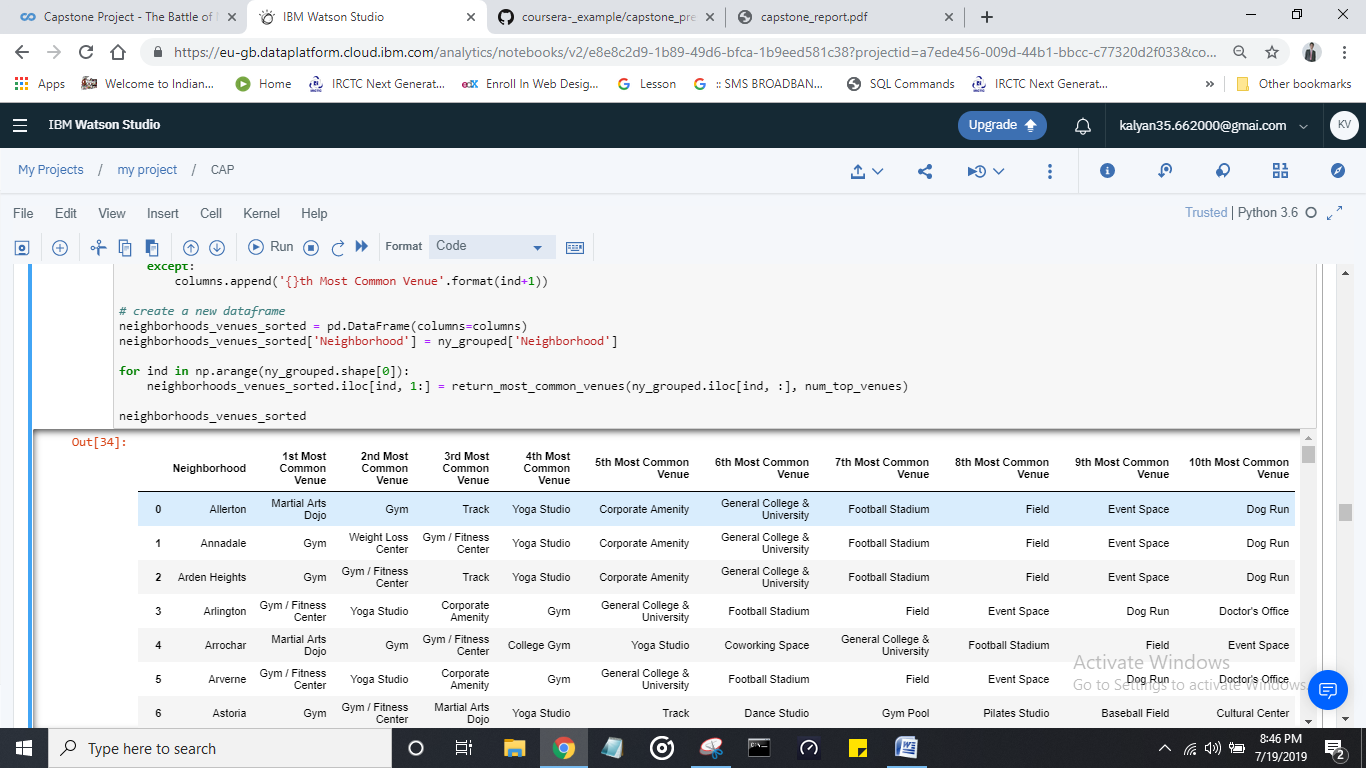
We will group the data according to the neighborhood and the data will look like this with each neighborhood is having categories like Neighborhood Latitude, Neighborhood, Longitude, Venue, Venue Latitude, Venue Longitude and Venue Category

Later we will analyses each neighborhood as follows and we will get the following result

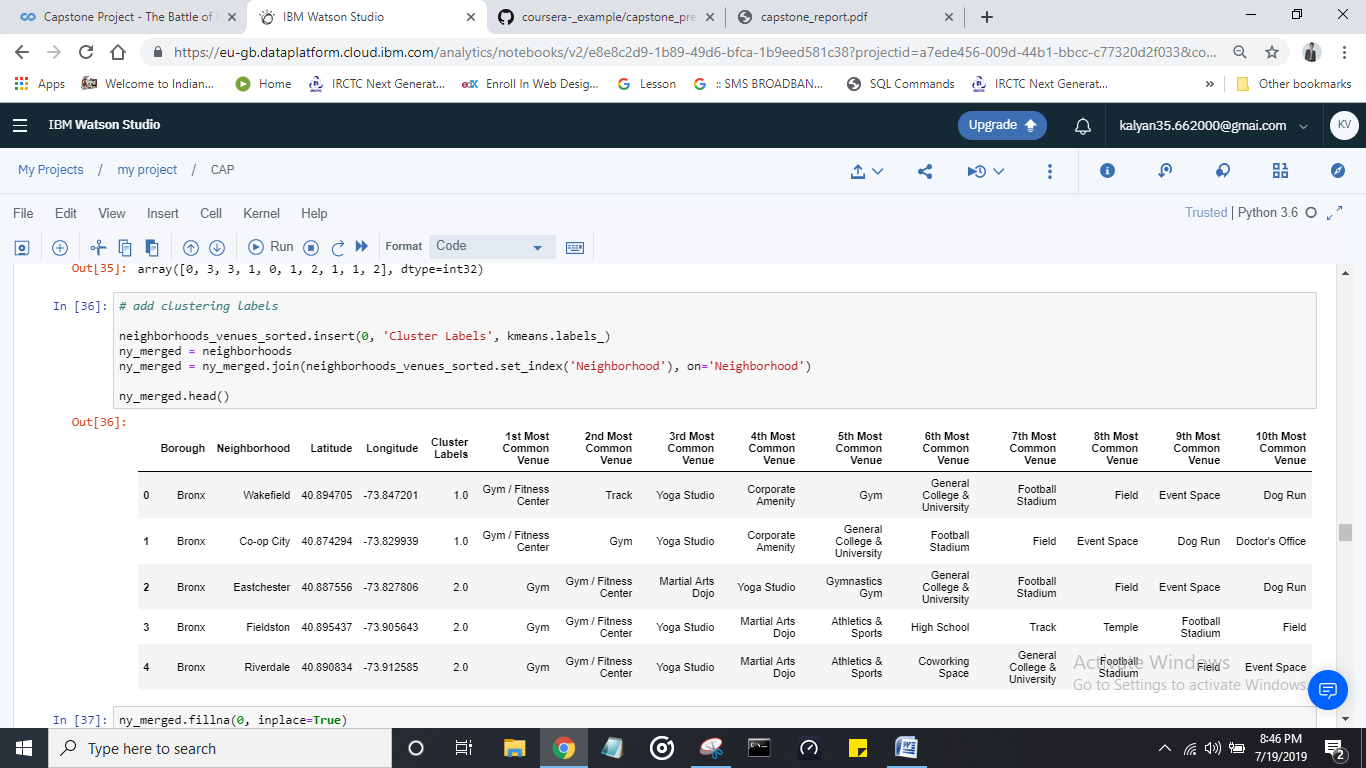


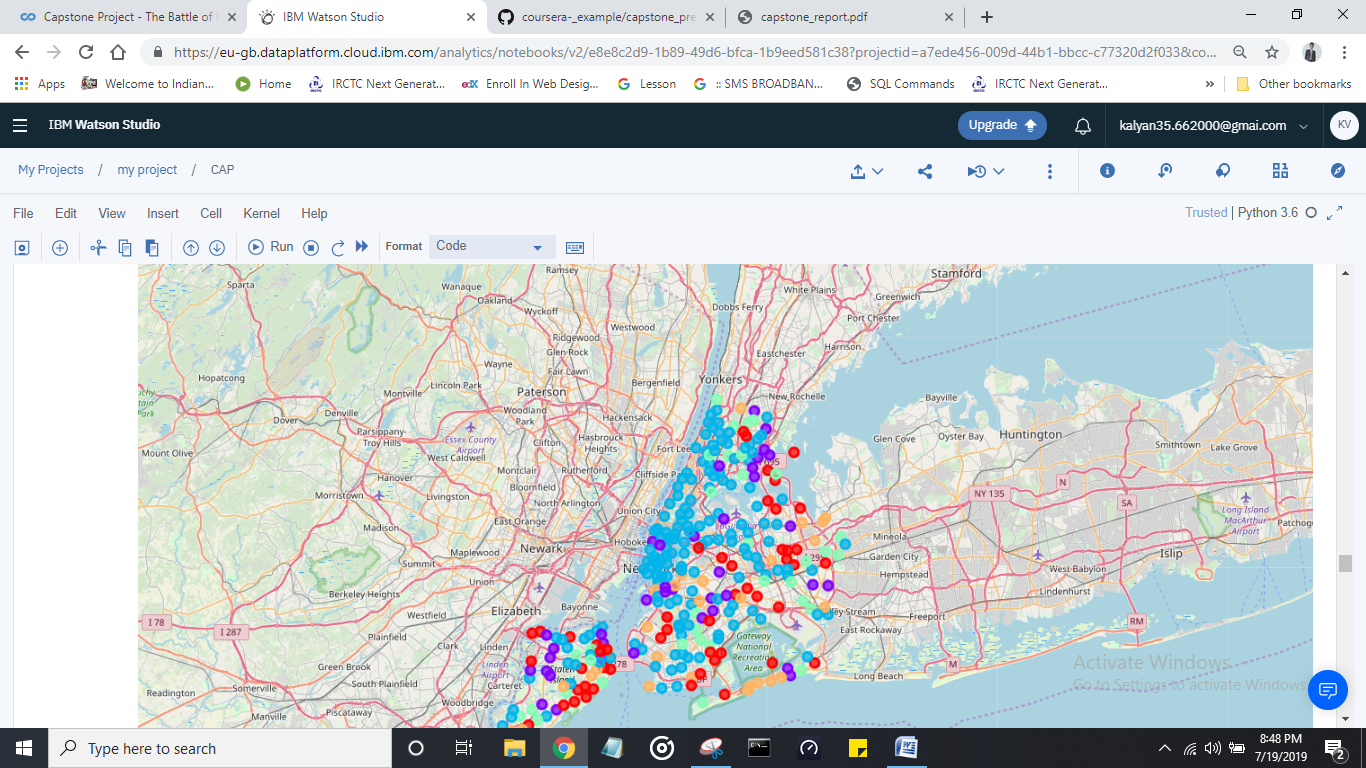
Where the 1’s represent there is a stadium in that neighborhood and 0’s represent there is no stadium in that neighborhood.

Now we consider the 10 most common places.



This is the major part of the project applying famous machine learning algorithm called **k-means clustering** which helps to forms different clusters from the given data. Here we took 5 clusters; we can take any number of clusters according to our requirement



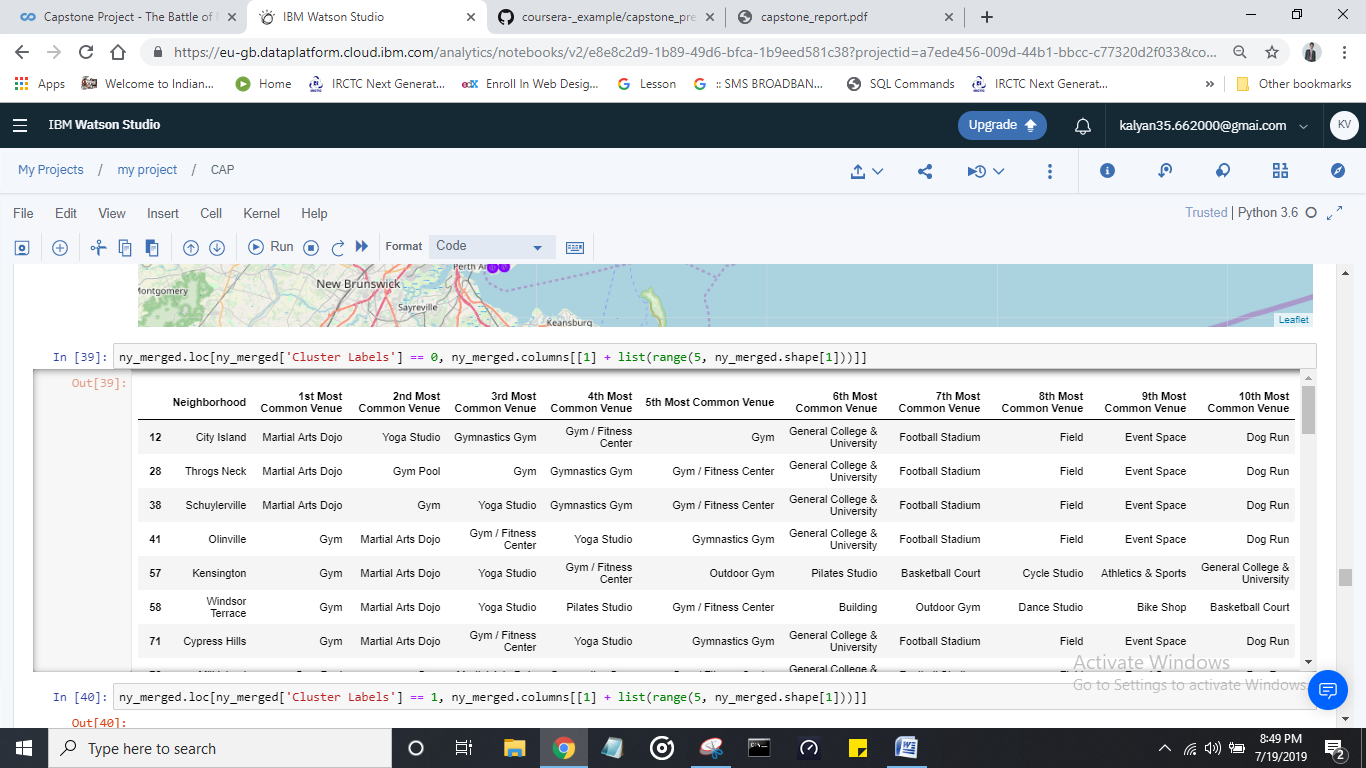
After we apply k-means algorithm to the data we can visualize a map where we can find different clusters on the New York City map.

This is how the map looks like and different colors represent different clusters.

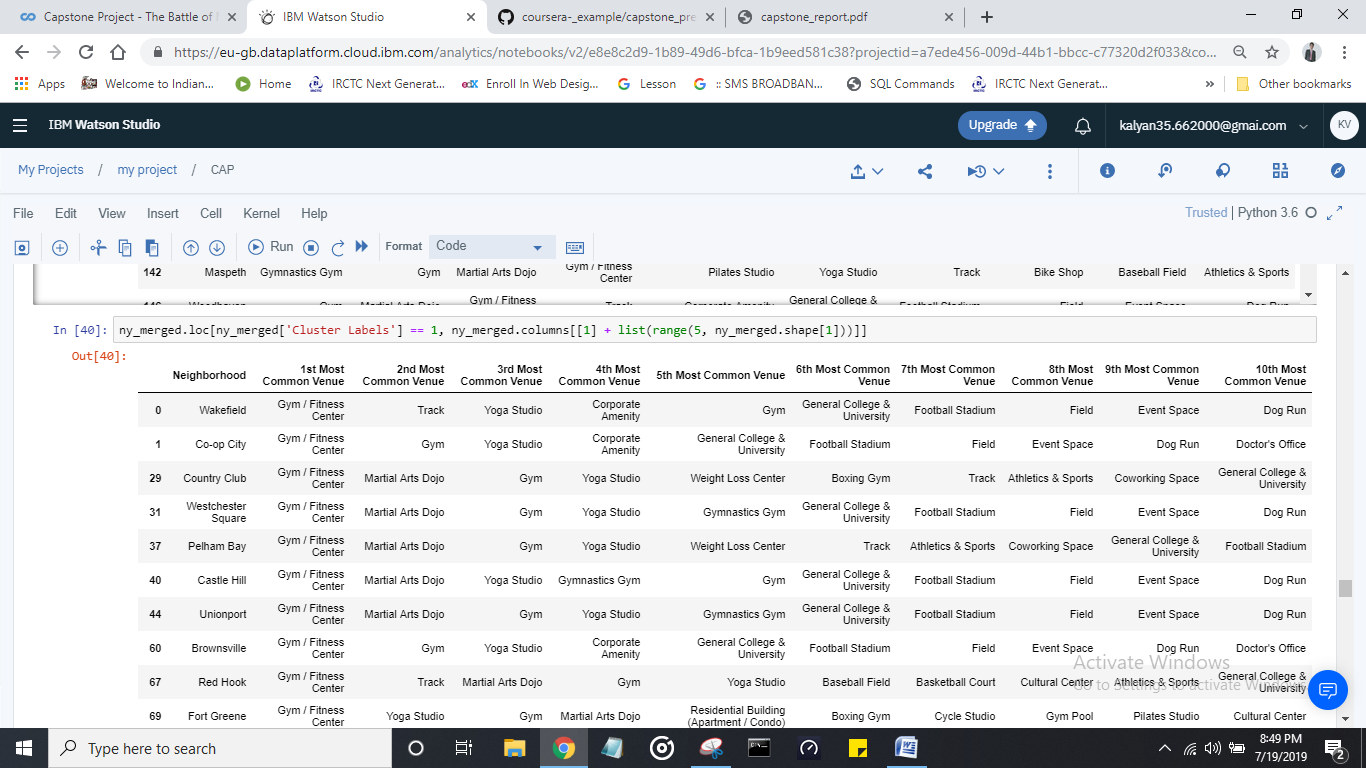
**RESULTS**

Now we will check each cluster results individually.

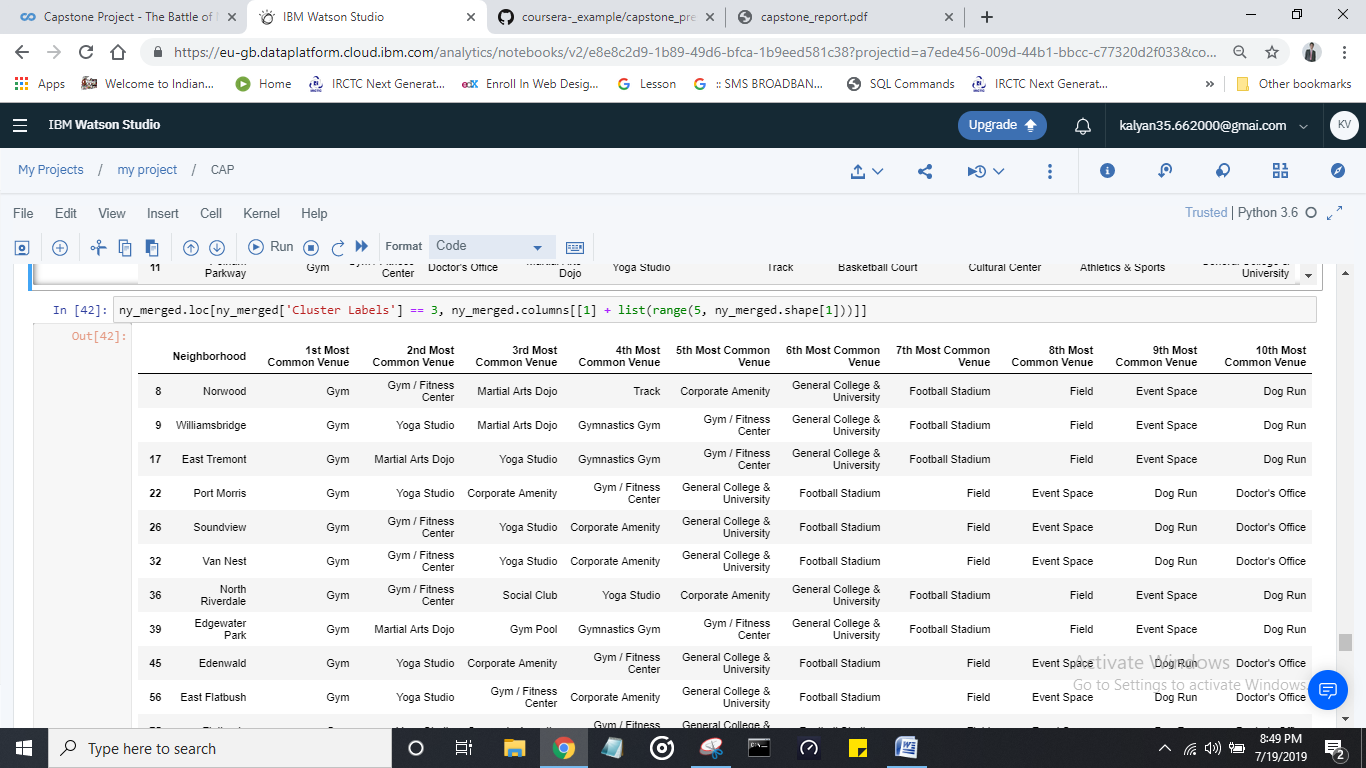
**First cluster:**

****

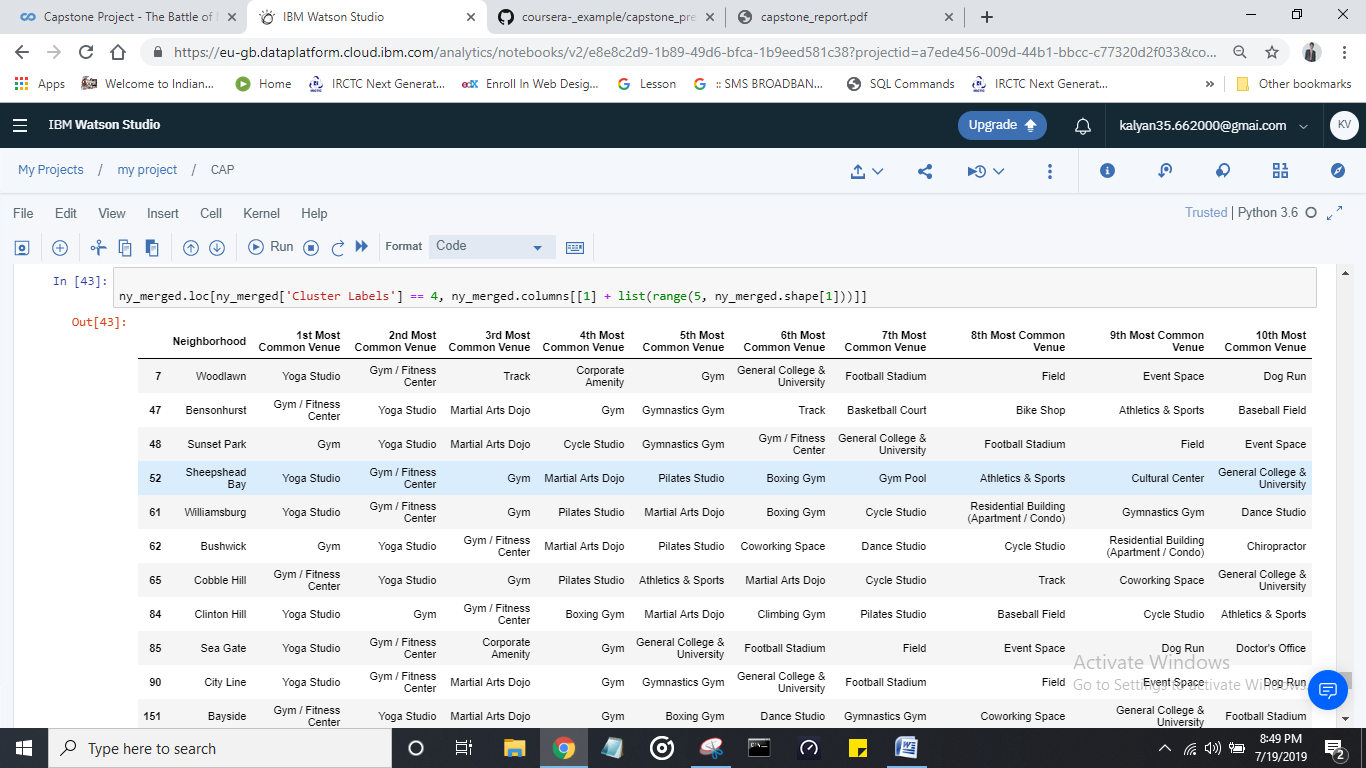
**Second cluster :**

****

**Third cluster:**

****

**Fourth cluster:**

****

Here we can observe the different common venues for the different clusters. By observing the clusters, if my client want to build a gymnasium then cluster 2 and 3 will be the best places where there is less competition for the gymnasium and there people are very interested in participating in fitness increasing activities.

**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.

**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.