**Final Report: Making a Sushi Trial in NYC**

Rohan Ramdoss

Business Understanding:

The City of New York is the most populous country in the United States. It is multicultural and has always been on my travel bucket list. New York’s multiculturalism invites the possibility of a high range and diversity of multicultural cuisines. The insights from this analysis will help create a travel itinerary of 1 day visiting the many sushi restaurants in NYC. Why Sushi? Because it’s fantastic.

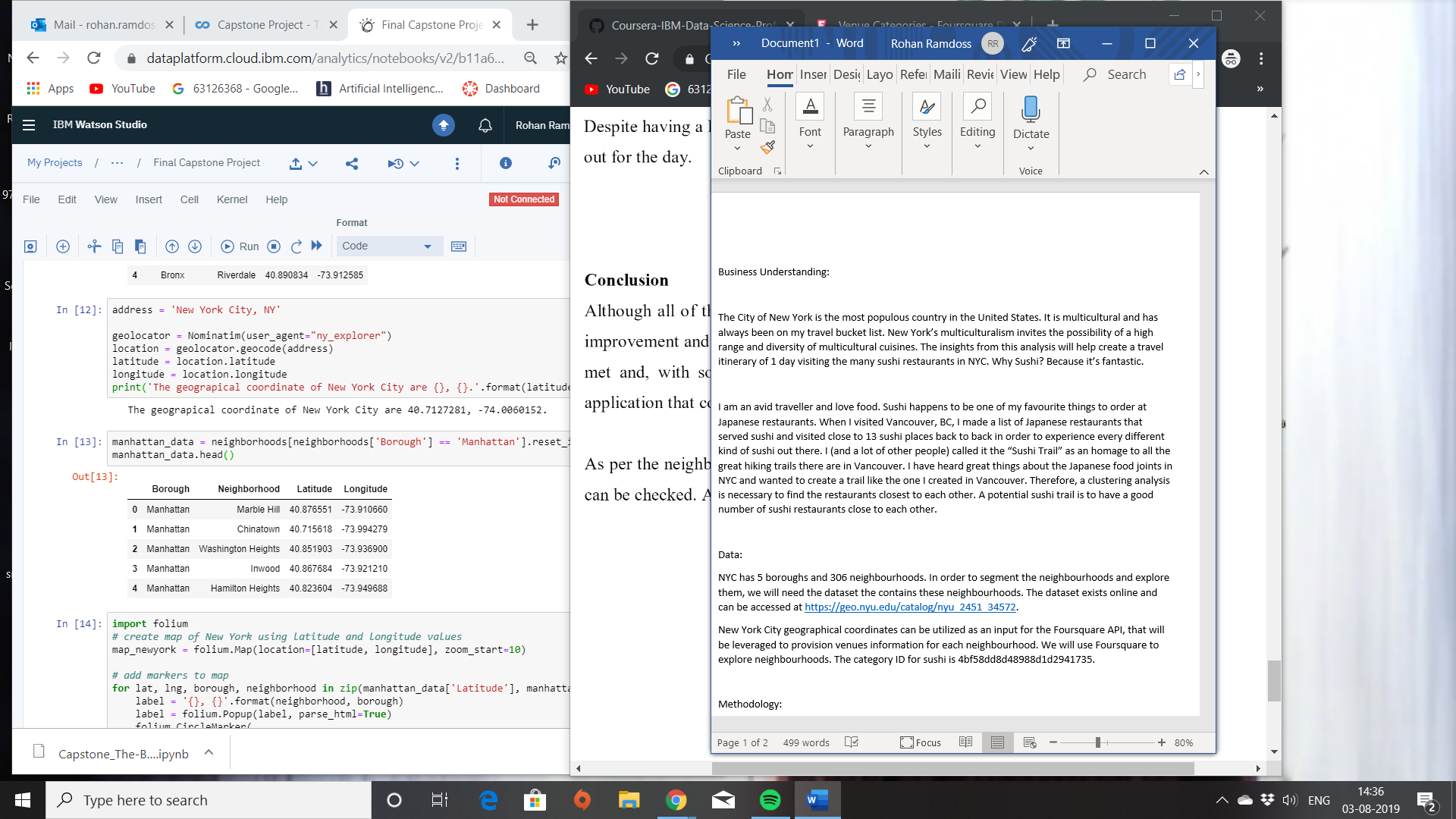
I am an avid traveller and love food. Sushi happens to be one of my favourite things to order at Japanese restaurants. When I visited Vancouver, BC, I made a list of Japanese restaurants that served sushi and visited close to 13 sushi places back to back in order to experience every different kind of sushi out there. I (and a lot of other people) called it the “Sushi Trail” as an homage to all the great hiking trails there are in Vancouver. I have heard great things about the Japanese food joints in NYC and wanted to create a trail like the one I created in Vancouver. Therefore, a clustering analysis is necessary to find the restaurants closest to each other. A potential sushi trail is to have a good number of sushi restaurants close to each other.

Data:

NYC has 5 boroughs and 306 neighbourhoods. In order to segment the neighbourhoods and explore them, we will need the dataset the contains these neighbourhoods. The dataset exists online and can be accessed at <https://geo.nyu.edu/catalog/nyu_2451_34572>.

The borough chosen was Manhattan due to proximity from housing accommodation.

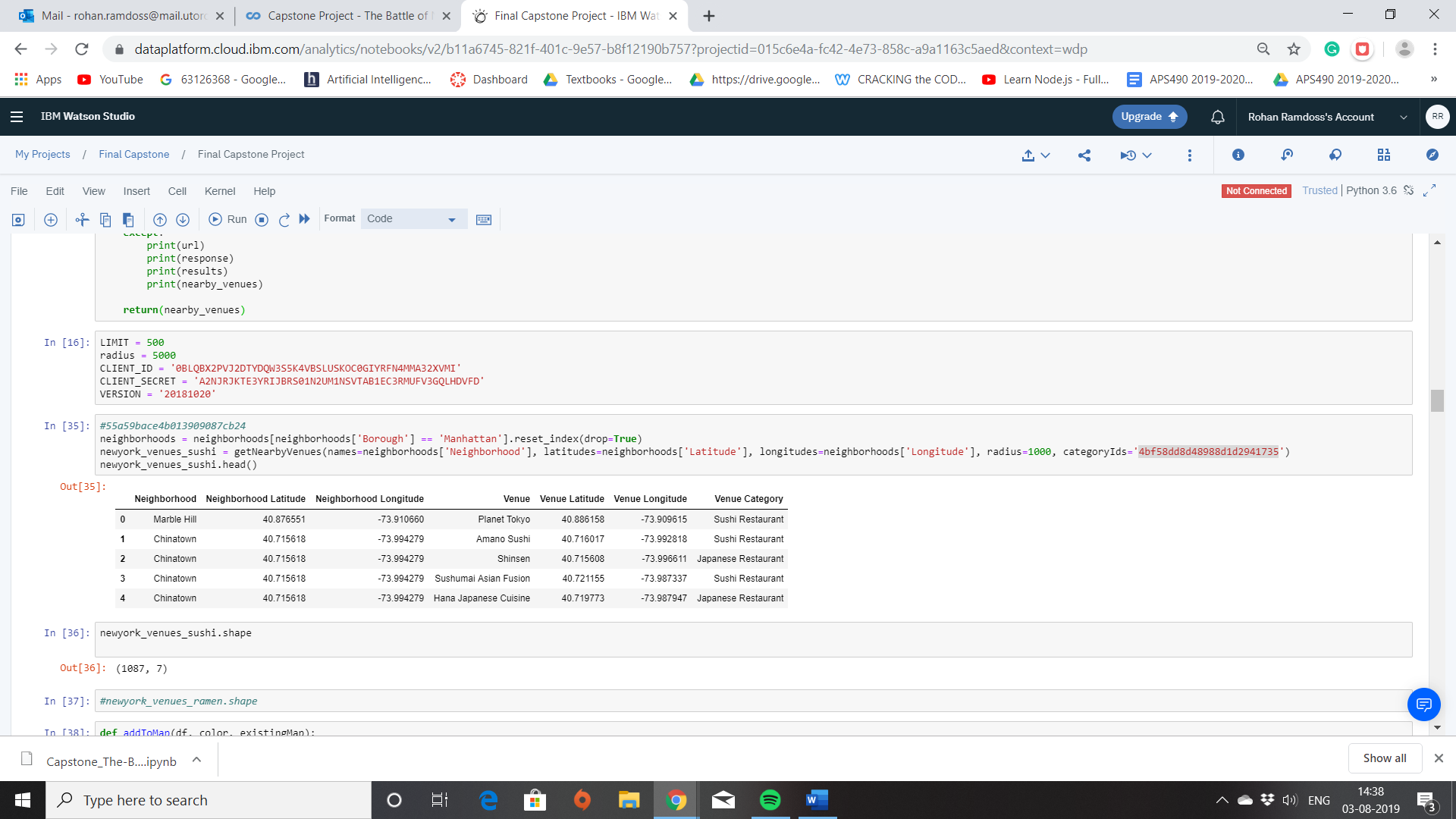
Once the data is transformed into data frames, it looks like the picture shown below

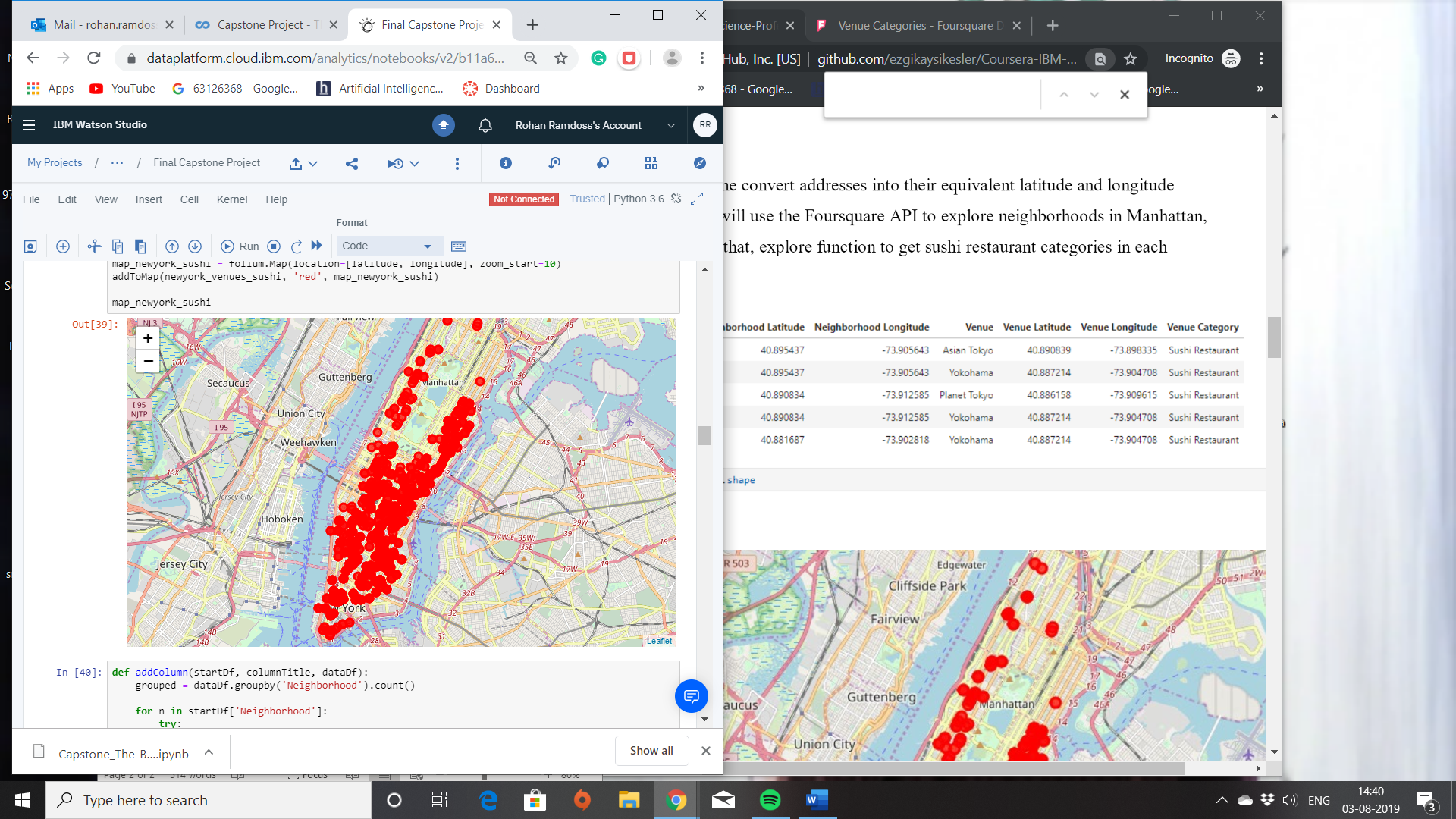


New York City geographical coordinates can be utilized as an input for the Foursquare API, that will be leveraged to provision venues information for each neighbourhood. We will use Foursquare to explore neighbourhoods. The category ID for sushi is 4bf58dd8d48988d1d2941735.

Methodology:

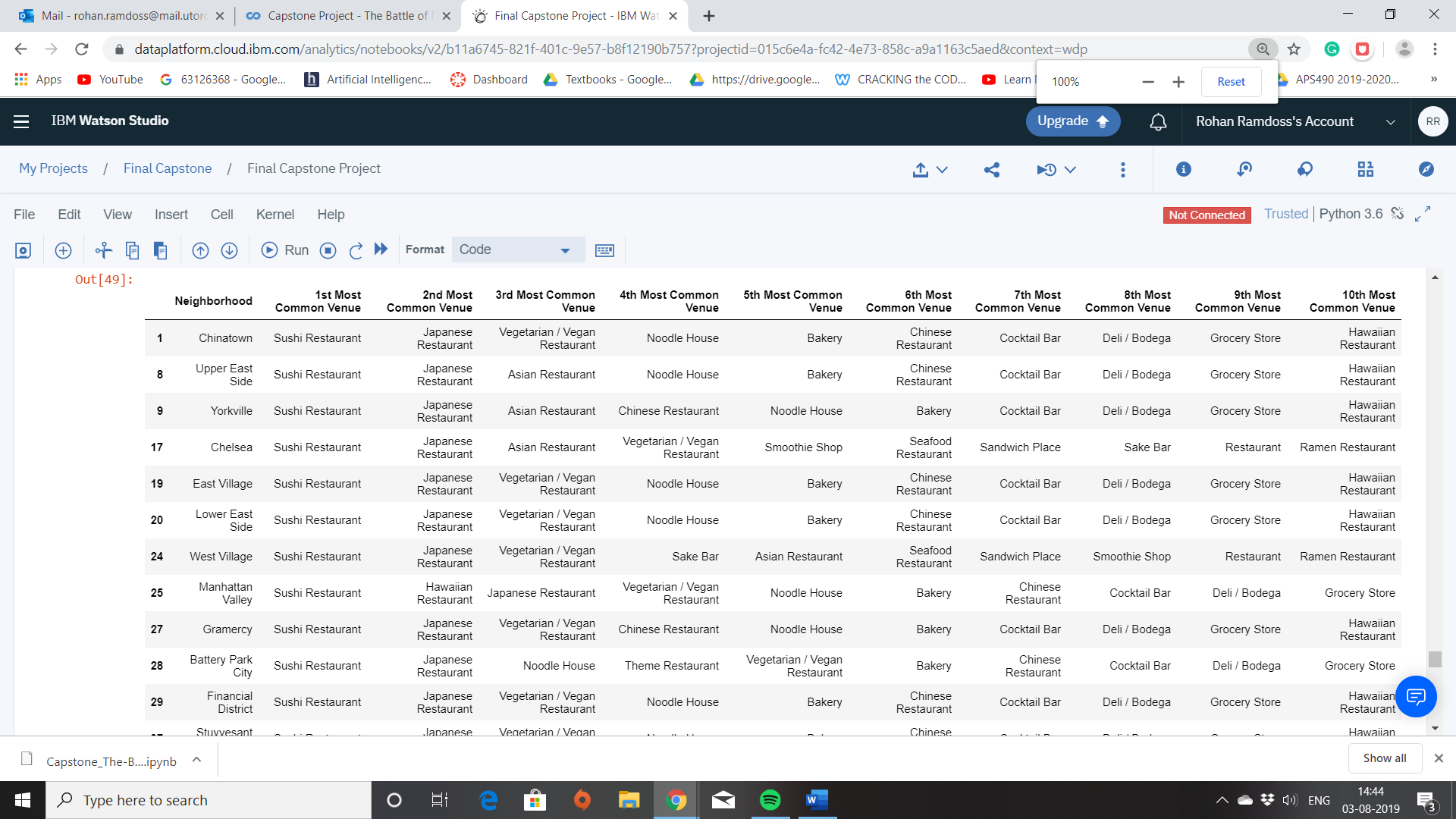
This follows the methodology of the lab done in week 3 with analysing neighbourhoods in Toronto. After converting addresses into their equivalent latitude and longitude values, I used the Foursquare API to explore neighbourhoods in Manhattan New York. After one hot encoding, we use the feature group for K-means clustering algorithm and folium to make clusters and visualize Manhattan Clusters.

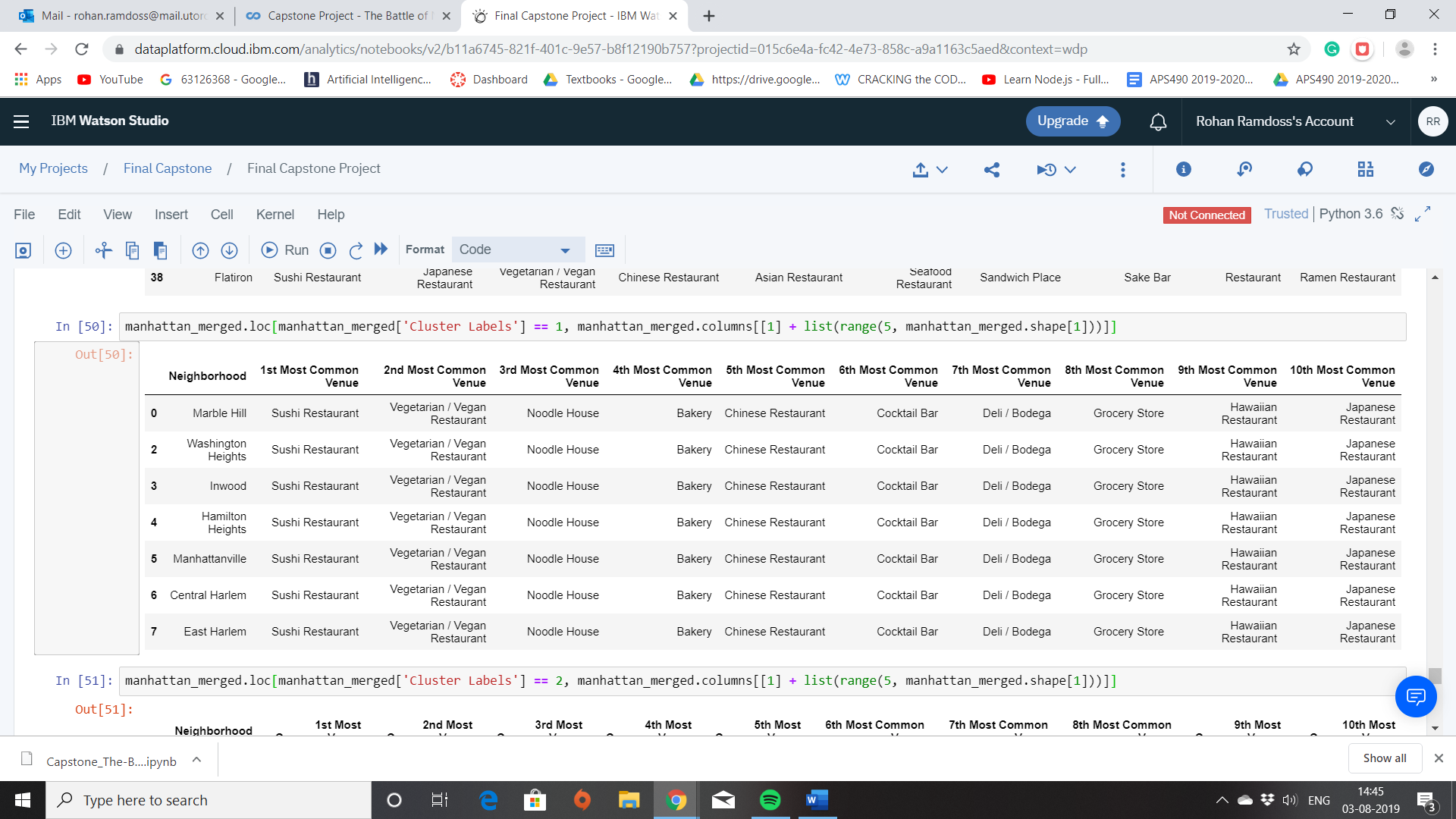


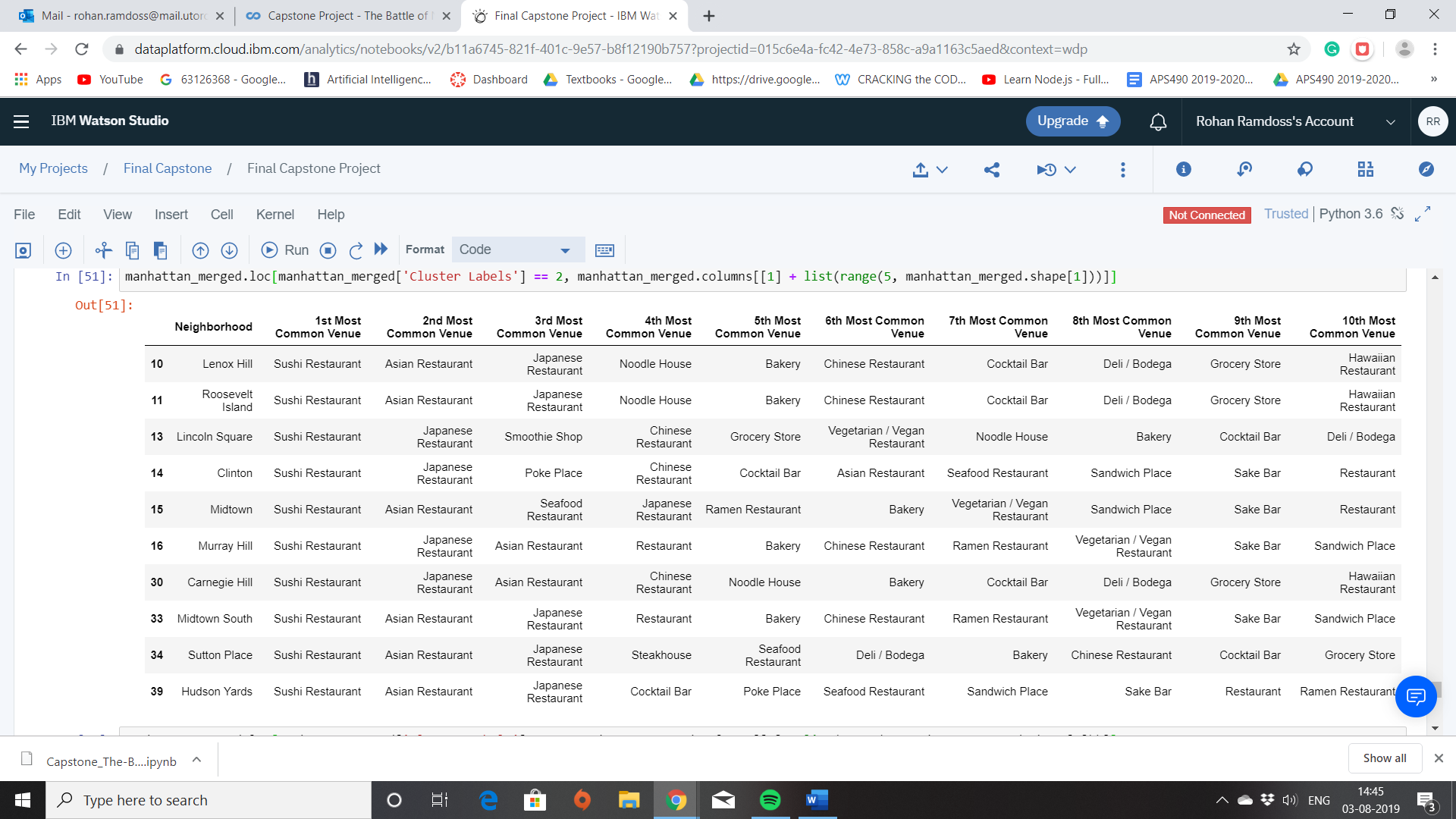


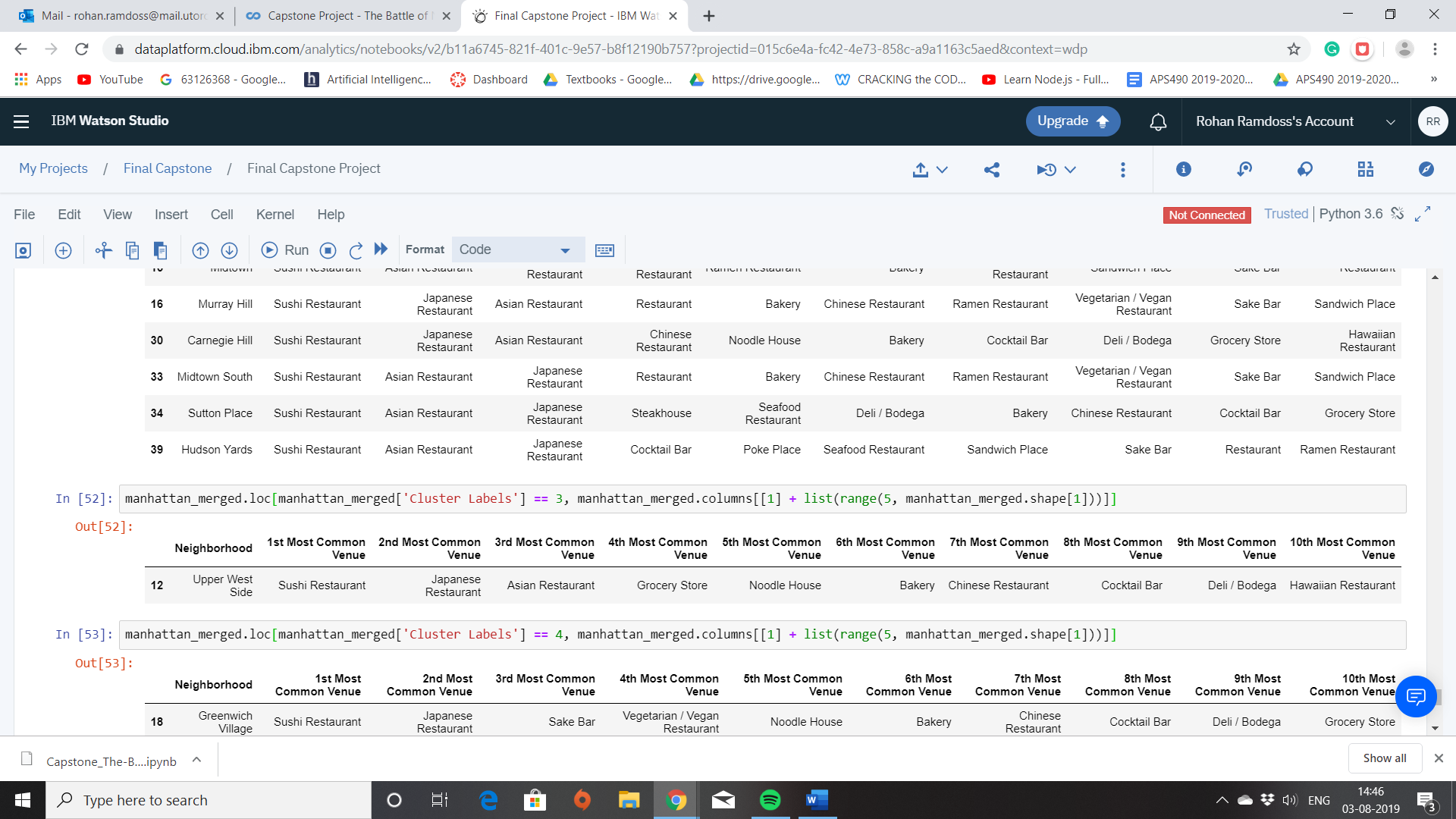
Results:

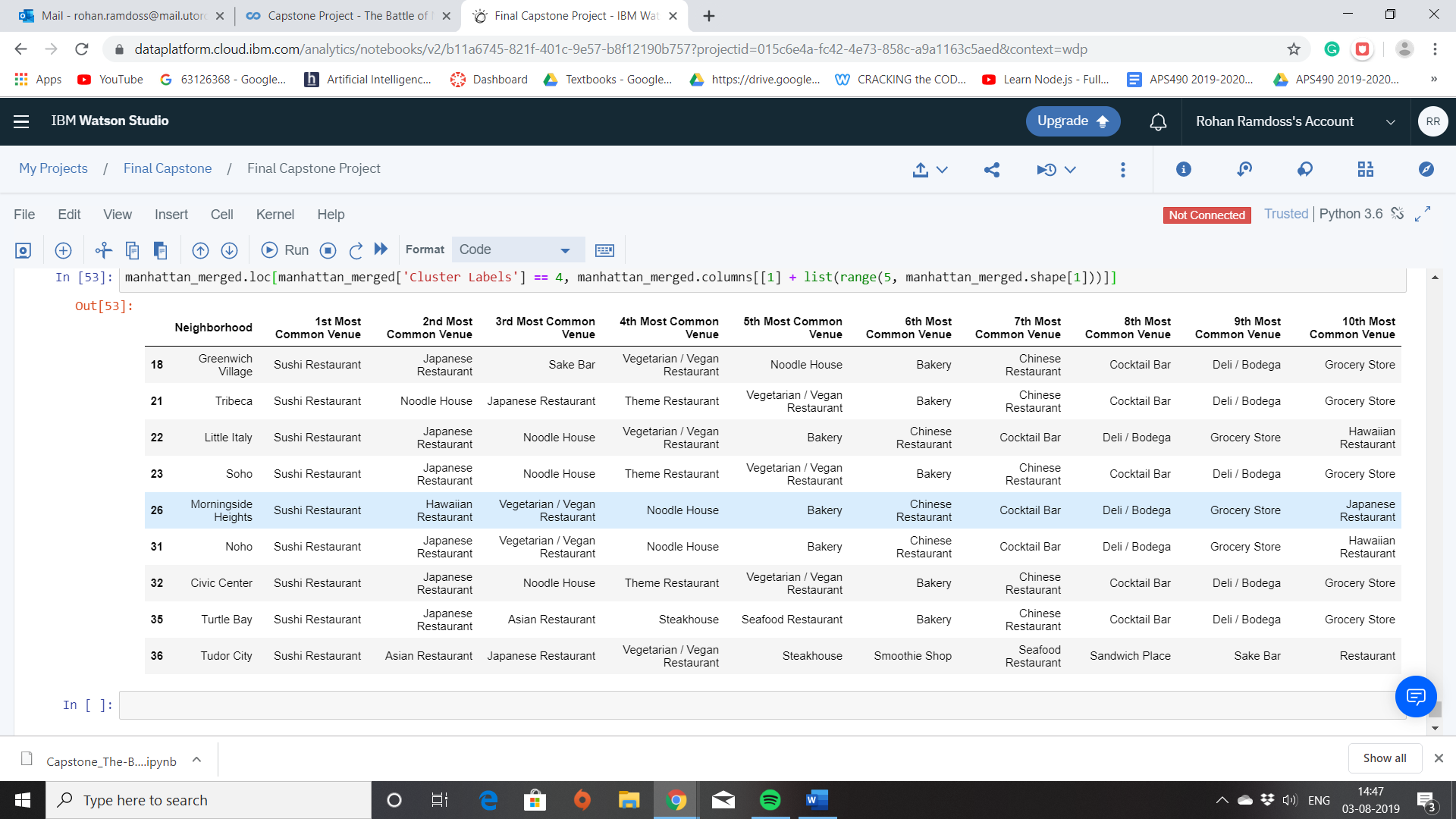
Using K means Clustering I have 5 clusters and used folium to display the clusters:

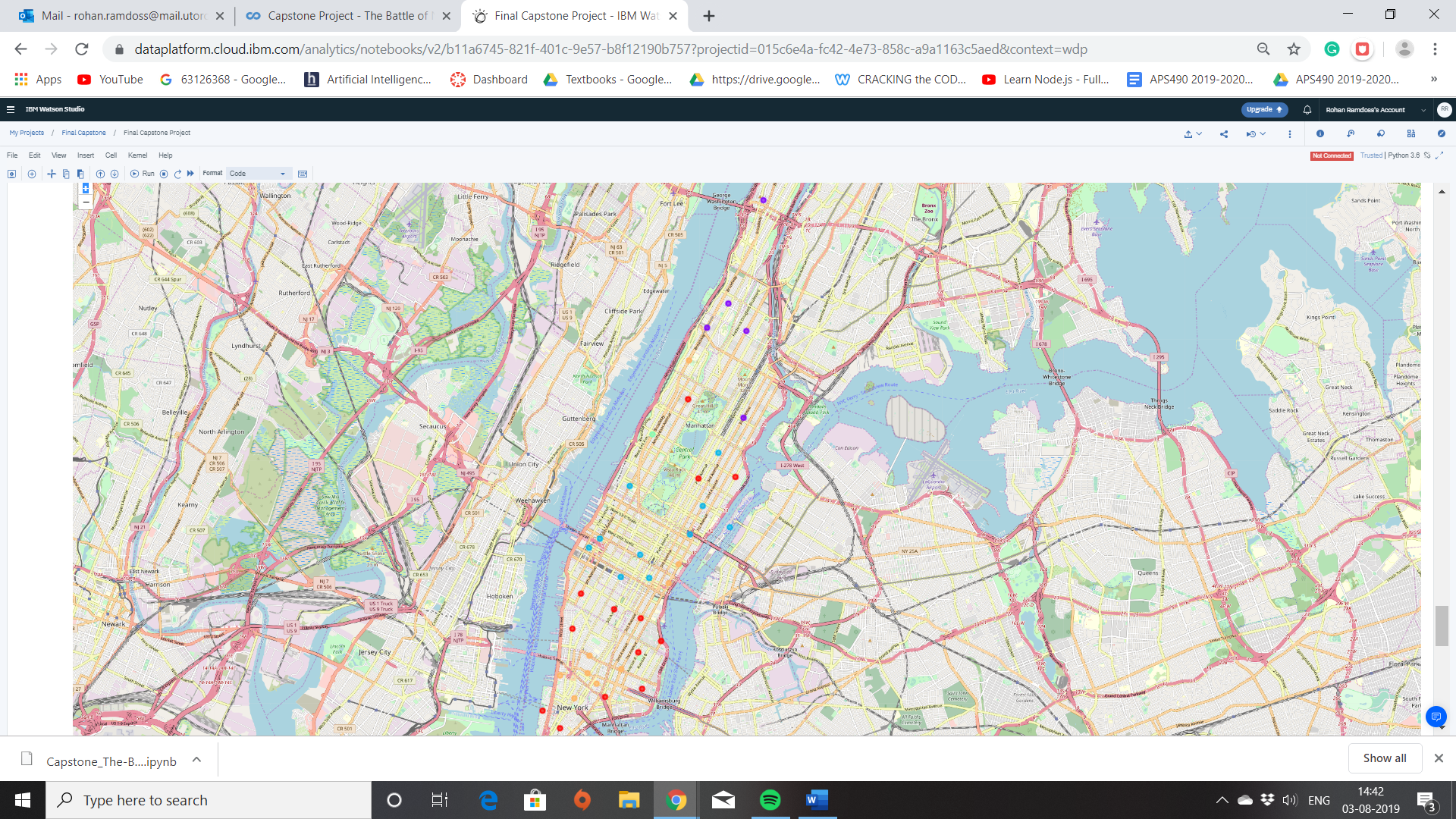












The pictures above show the clusters and the folium representation.

The results will be discussed in the following section.

Discussion and Conclusion:

Foursquare proved to be a good source of data. Cluster 2,0, and 4 have the highest number of sushi restaurants. Cluster 3 has the lowest number of restaurants and cluster 1 has a good number of restaurants. Cluster 2’s potential trail is not a good choice since the distance between each restaurant is relatively large. Cluster 1, however, offers the perfect mix between Asian vs Japanese restaurants, a decent number of restaurants, and the relative distances between them are much smaller in comparison to Cluster 2,0, and 4.

The goal of this project was to be able to create a “Sushi Trail”. The final cluster consists of 13 restaurants and one outlier which can be discarded. I am confident that this trail comes with good times and a full belly!