**Thai Restaurant density segmentation in Toronto**

Theethat Anuraksoontorn

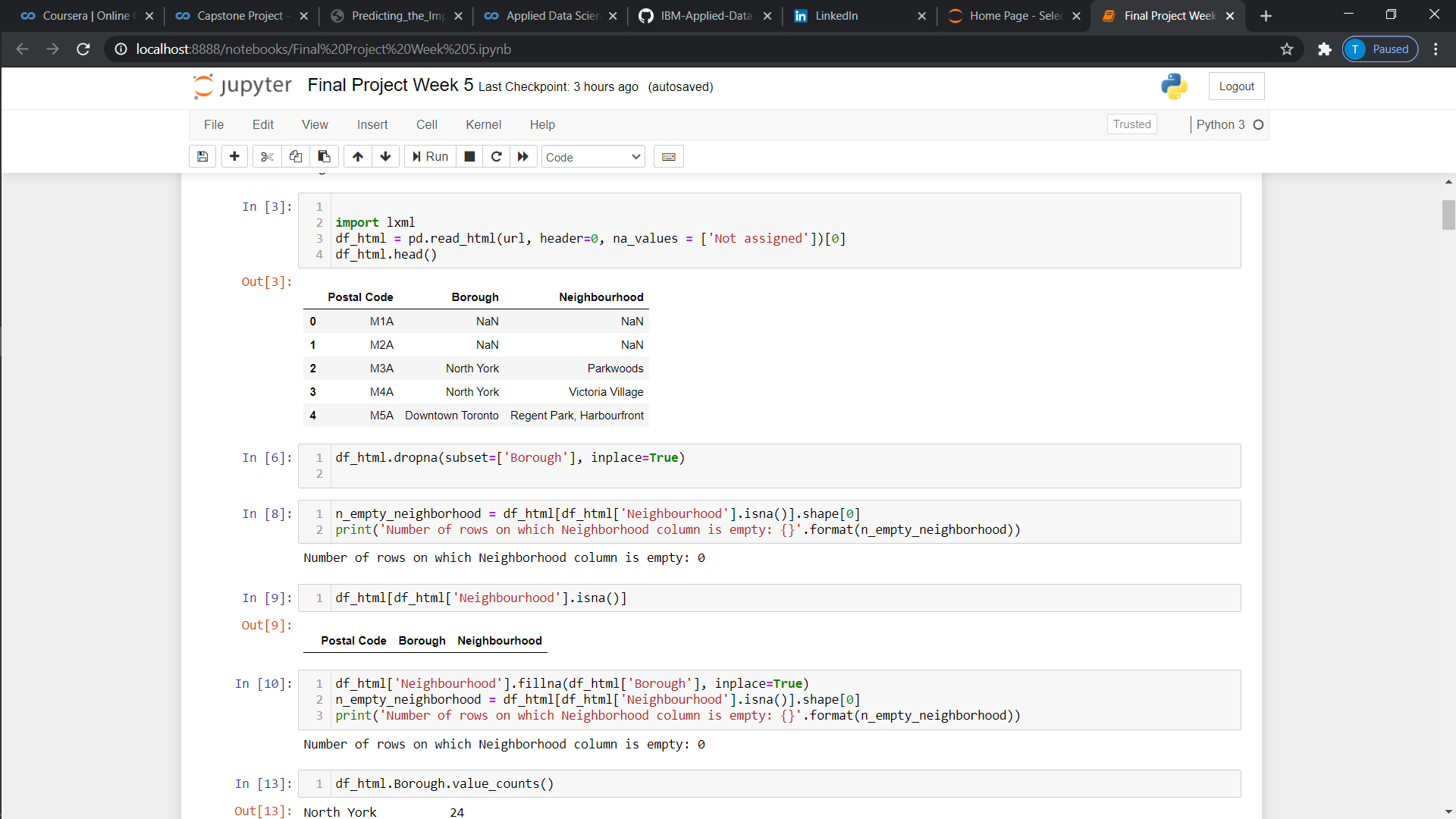
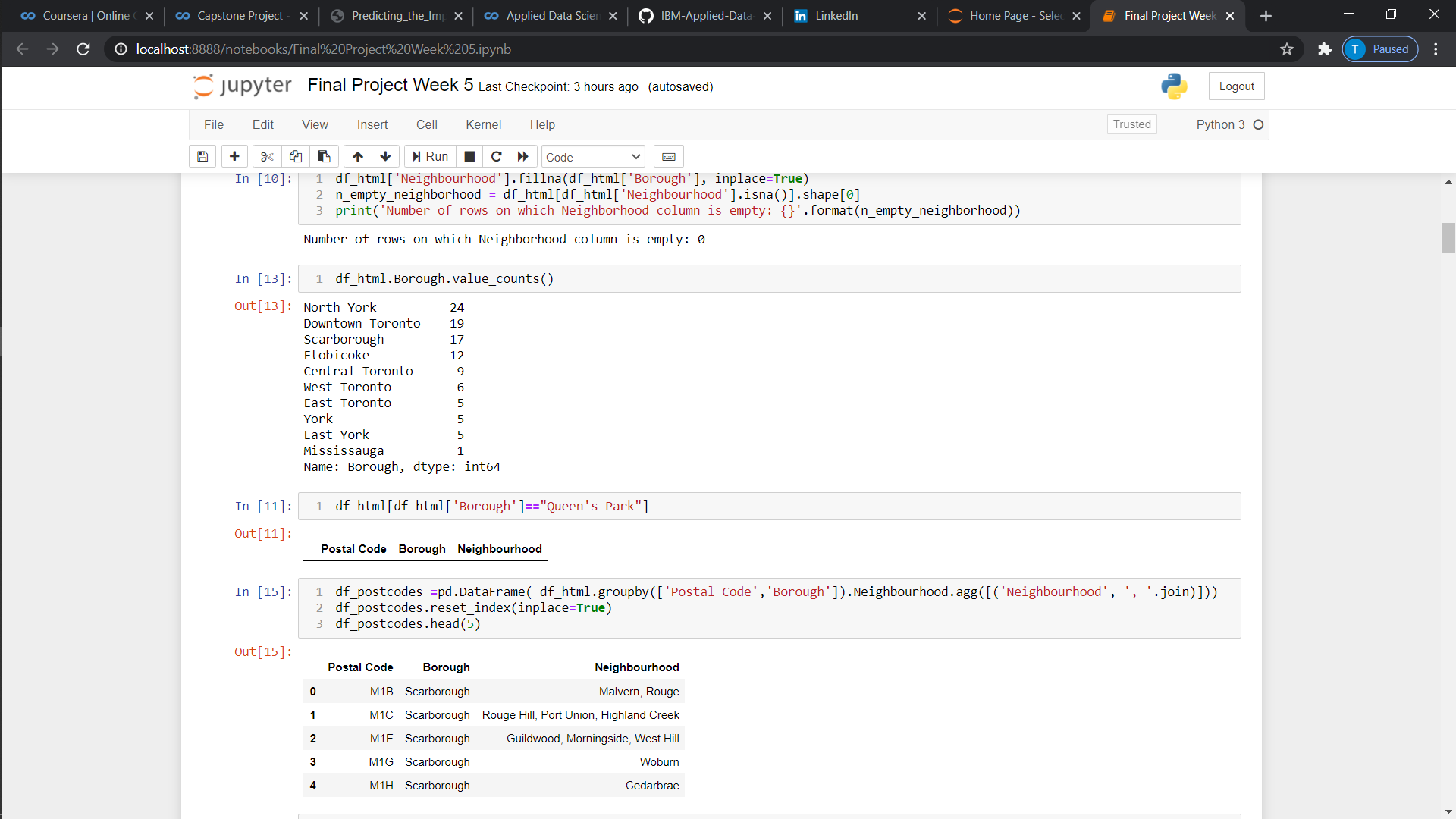
**Introduction**

Thailand is undoubtedly one of the richest cultures in a variety of nature, art , and most obvious and important, cuisine. Thai cuisine is not simply beautiful and tasteful, it is embedded with Thai culture and story along with its deliciousness. Westerns have been crazy about Thai food, especially Canada. They just could not get enough of it, so a lot of Thai restaurants popped up in Canada neighbourhoods. However, the popped Thai restaurant does not seem to be authentic cuisine at all, they adapt for American tongue too much such that it has no story and thai culture at all. Regarding this information, I envision the opportunity for opening a Thai restaurant in that place in Toronto, Canada, at the place where there is low Thai restaurant density.

**Data**

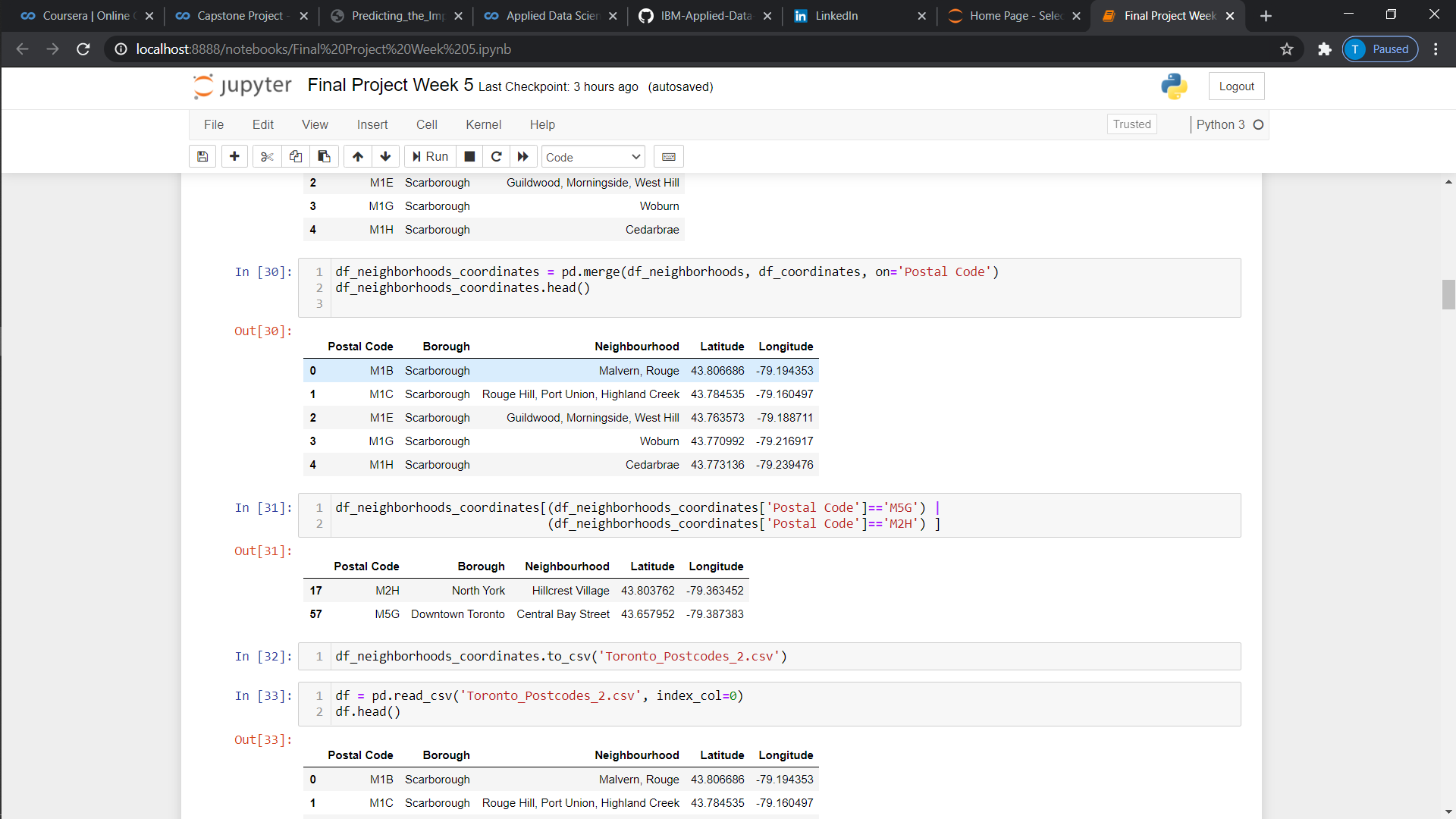
The data that we need would be simply all the restaurants in Toronto composed of

1. Neighborhood names of City of Toronto

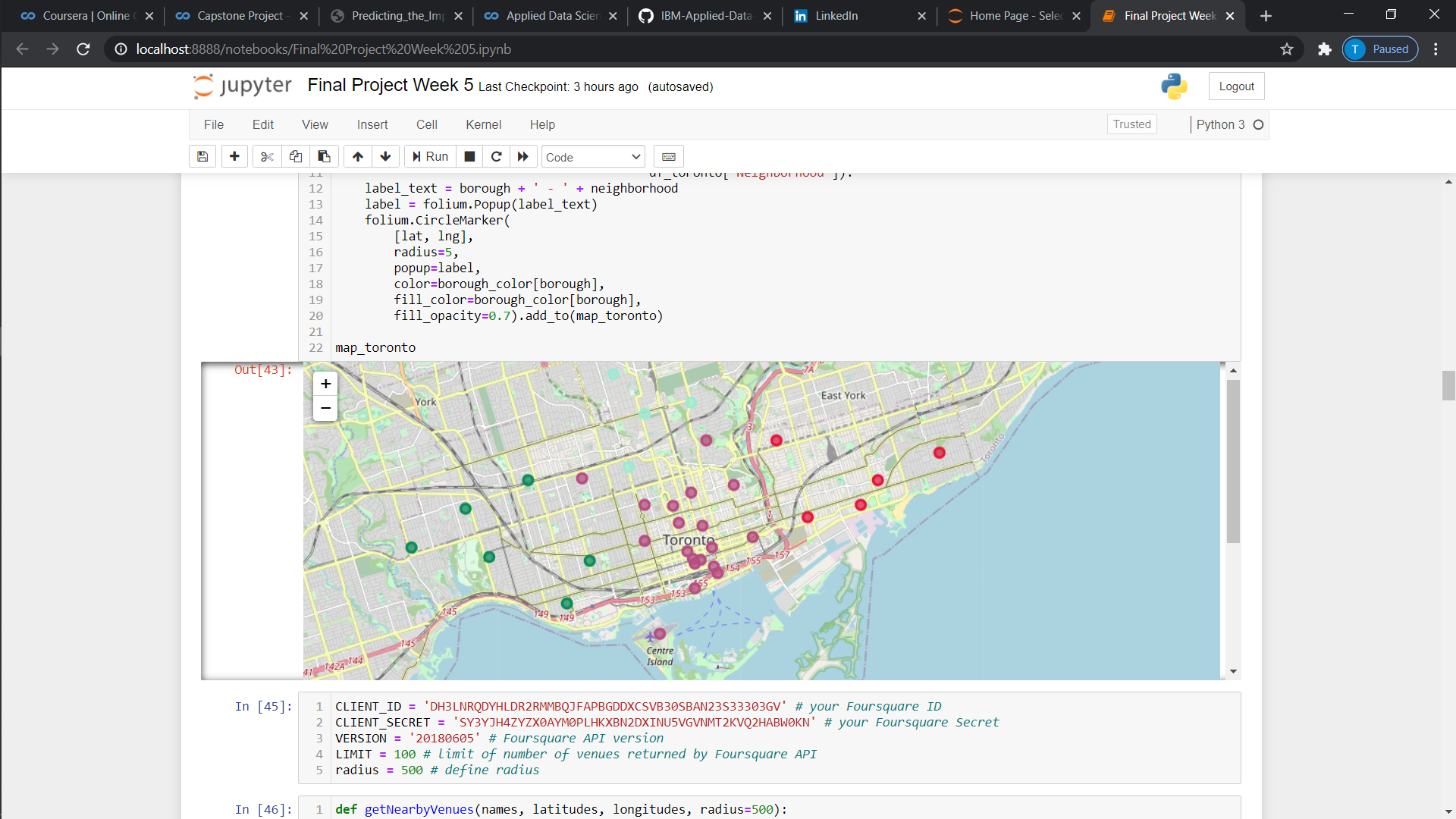
This piece of data can be obtained from the wikipedia “List of postal codes of Canada” in this link (<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>). I simply import using pandas library so that I don’t need to import other web scraping library to slow down my project. 

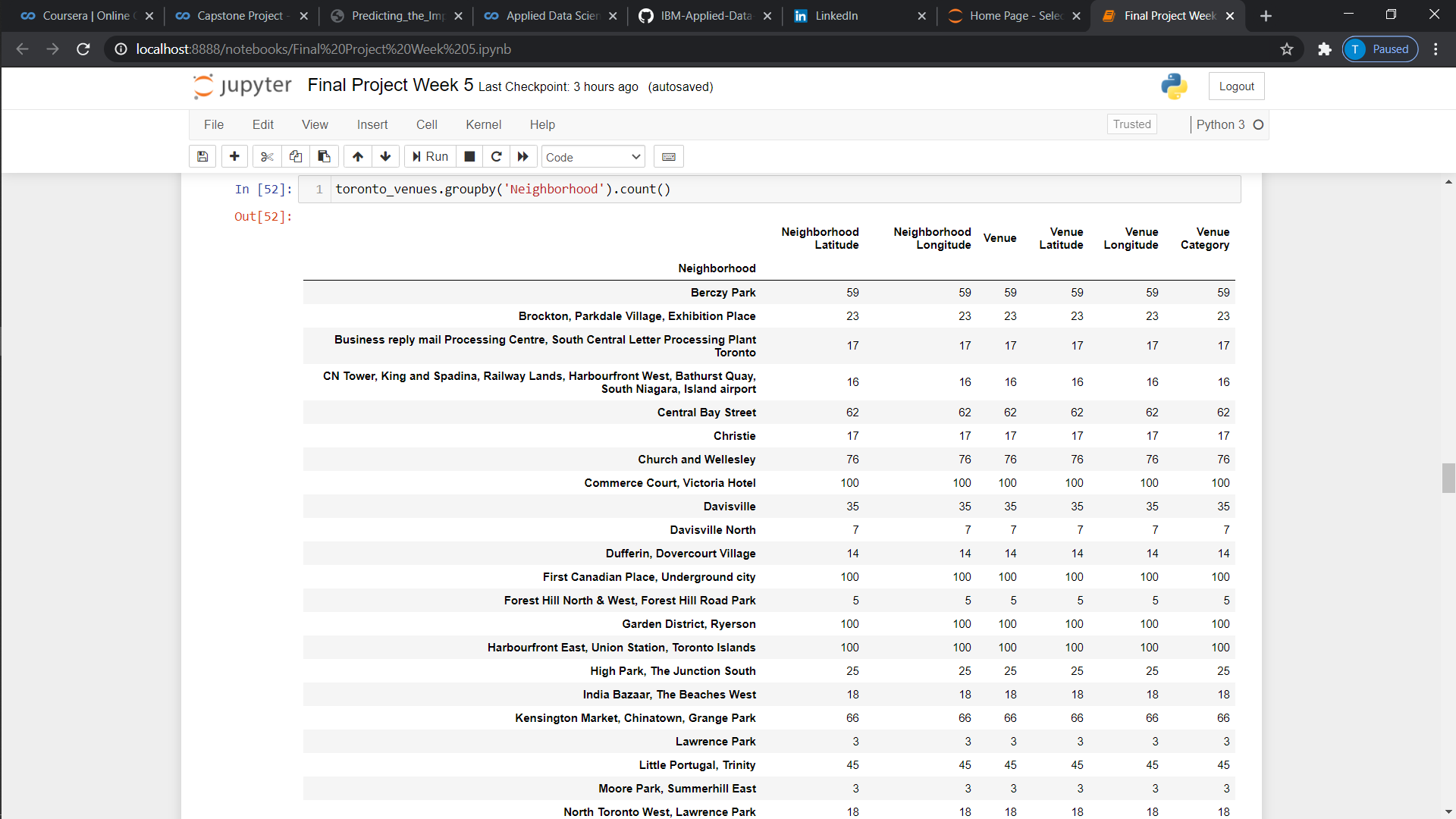
1. Geopy library to find the latitude and longitude of each neighborhood

After I have acquired the postal codes for Toronto, Canada, I mapping the dataset with the latitude and longitude using Geopy library for each neighbourhood.



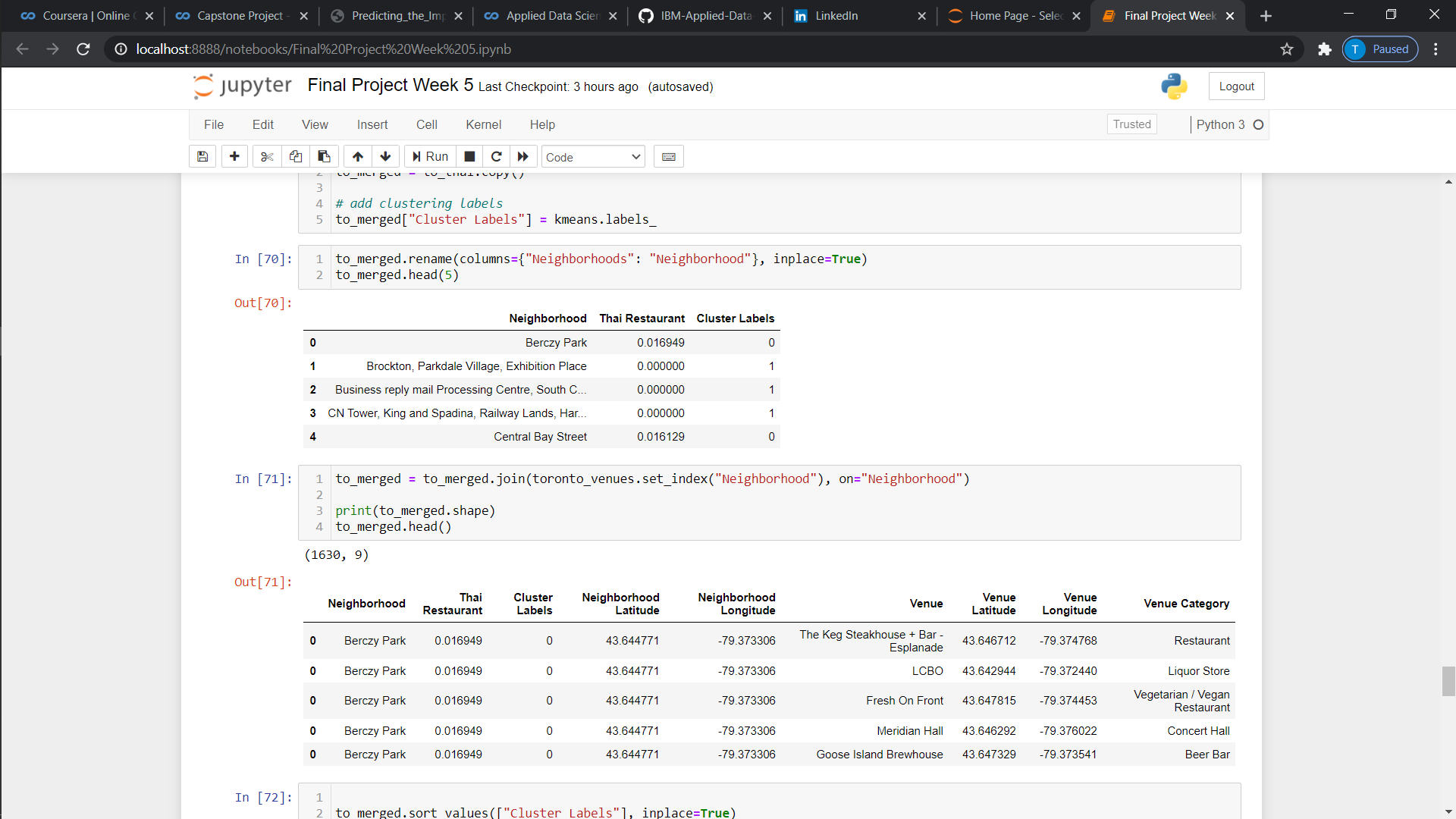
1. Foursquare location data to find venues in each neighborhood

After I have all the information ready to extract from the Toronto restaurant, I use foursquare API to track and acquire data of all restaurants in Toronto whether it is Thai or not. And with these three, I eventually narrow our interest into Thai and non-Thai data points for each restaurant.

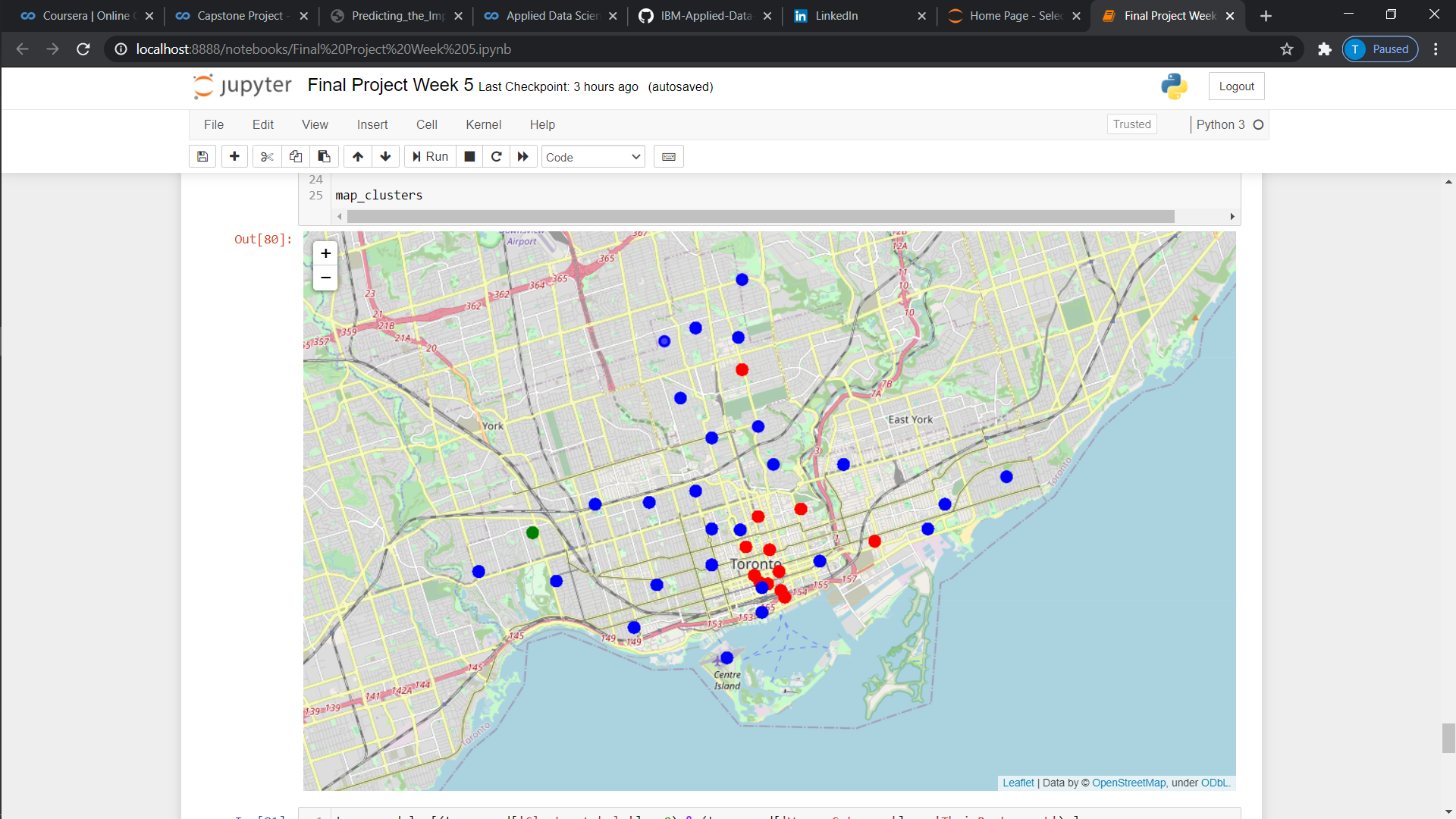


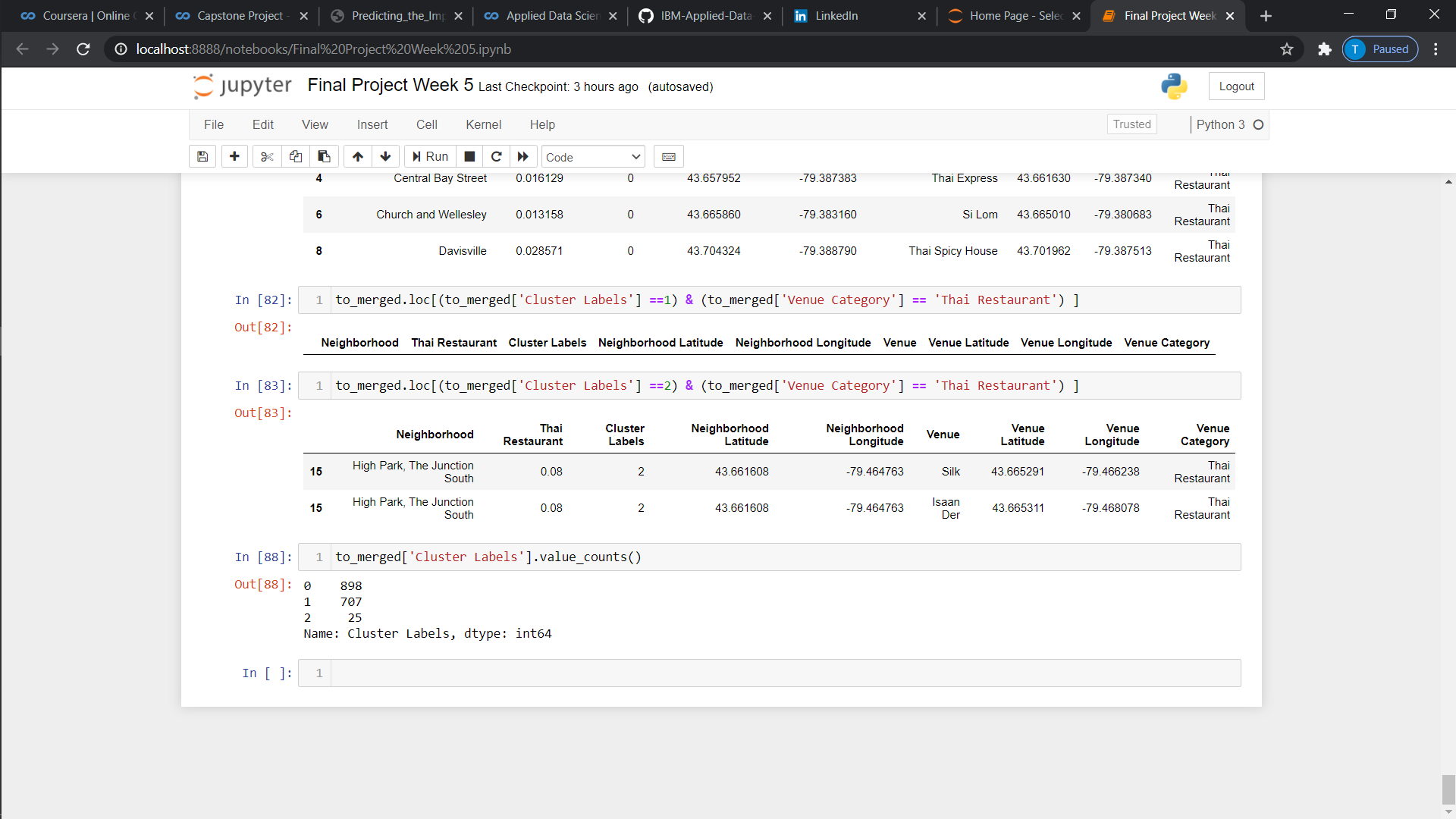
Methodology

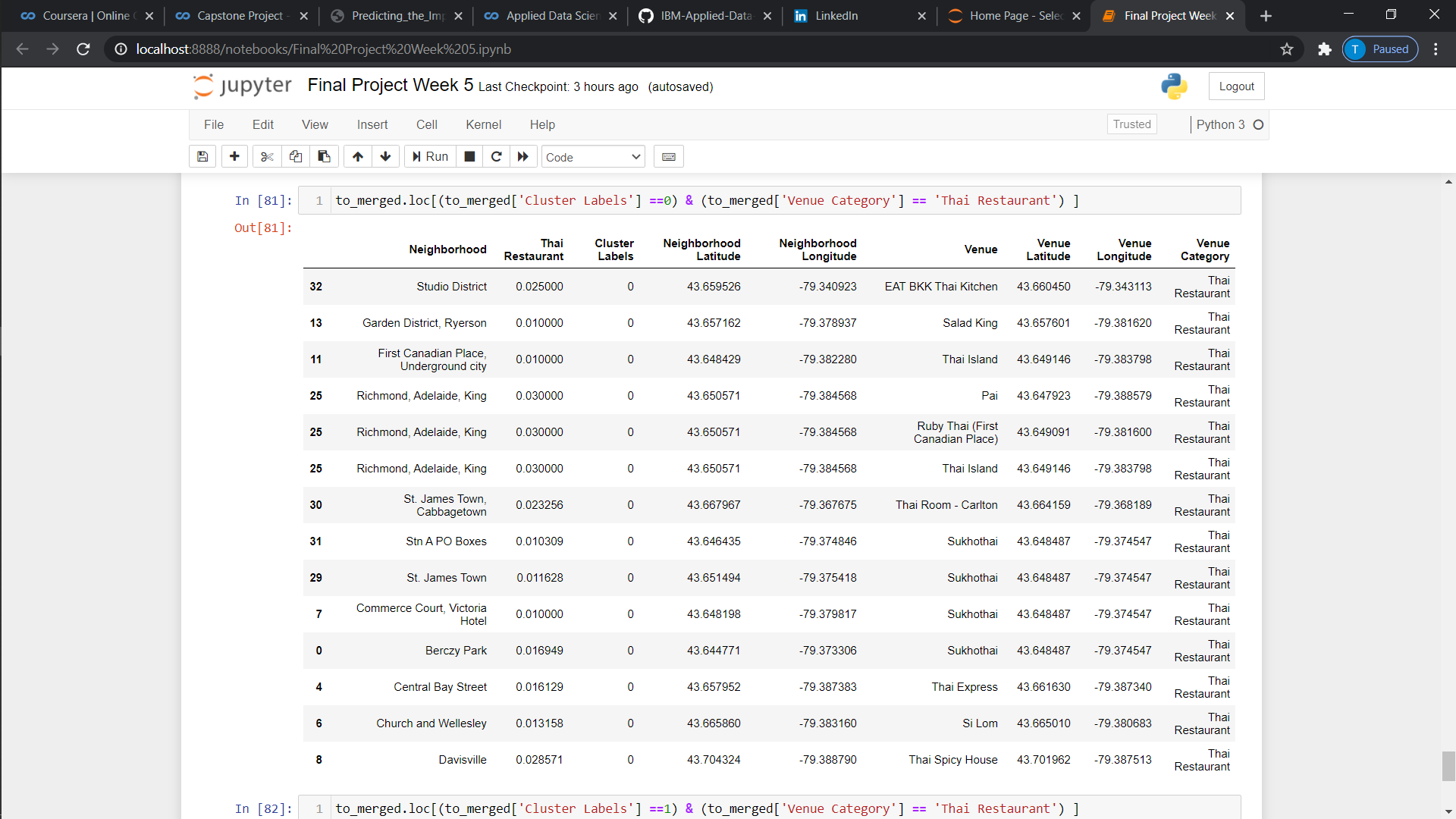
Our objective is to open a Thai restaurant in Toronto by finding the magnitude of Thai restaurant density. And that is basically segmentation of neighbourhood into high, low, and zero density. The best method should be using the clustering because I do not know exactly what the label for portions of high, low or zero should be. And it is best for the algorithm to define the rule of thumb for me. The one of the simple clustering, Kmeans would fit such task, because hierarchical clustering is too complicated and not need in this low dimension clustering, DBSCAN is quite not good because I still want to keep the outiner, i.e. the place where there is no Thai restaurant.



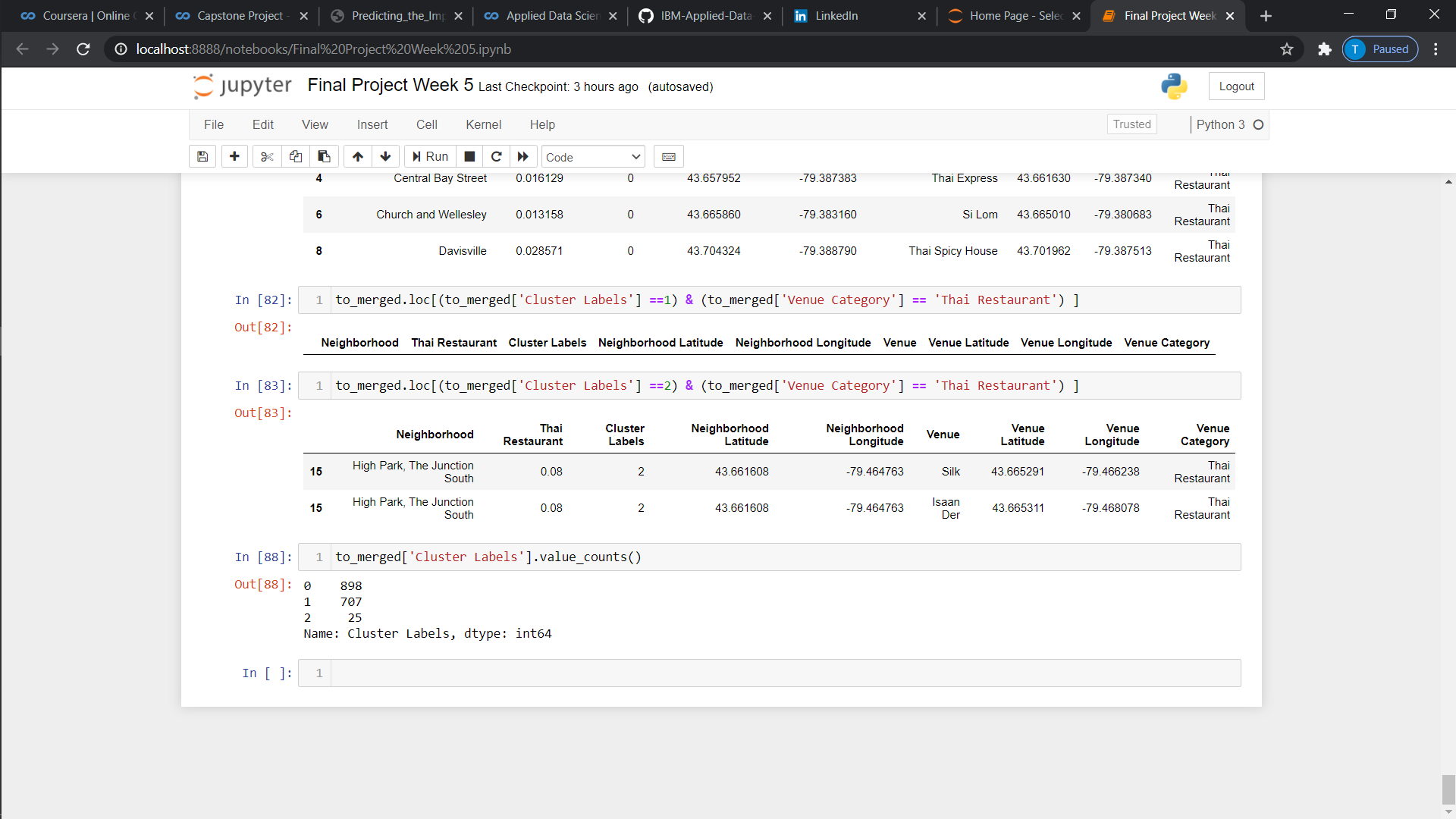
Results

The map below shows the density of restaurants based on Thai restaurant density, there are three clusters: red, green and blue clusters. 

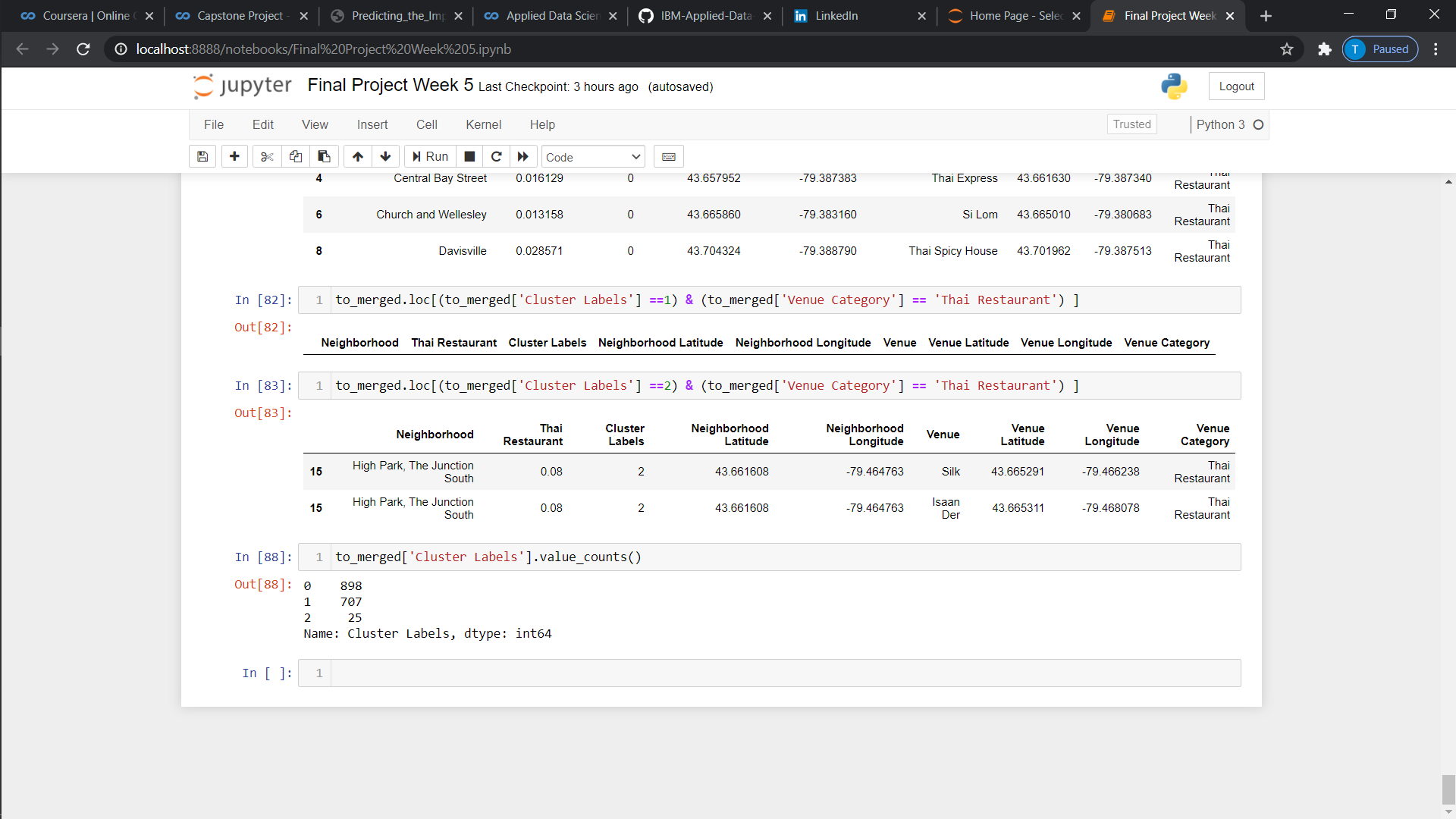
Most of thai restaurants are in red clusters which mean this red cluster is a high density cluster, this is the warning that there is too much competition in this area, and suggest not to open here.



A few restaurants are located in green clusters, this implies that the green cluster is a low-density cluster, this may be good to differentiate ourselves with the potential market and a good location to open the restaurants.



The last cluster is the blue one, there is no single Thai restaurant in this area and could be worse: because there is no one interested in Thai food here, or best: there is no one seeing the opportunity yet.



Conclusion

From the three clusters, it still needs a little more data based on the customer preference instead of location data to verify the best location for opening a restaurant. However, with the given data it would seem to be enough to some extent to suggest open restaurants either in green or blue clusters. The reason that it still needs more data is that the revenue generated from the restaurant is not purely based on the location of the restaurant itself, but based on the location of the potential customers. In this sense, the red area has high-density would imply that there are a lot of customers who like Thai cuisine, however opening up there might result in an intense competition that will only drive the cost eventually instead of profit. The superior place should be the subarea (The area near the red cluster, but is not red) which is the blue area surrounding the red area. However, we still not know the exact location where should be open in blue area because it is too big, but for the green area which is surprisingly small but contain a guarantee potential customer, this might be a gold mine or just gold nuggets that a small number of restaurant is taking all the profit and by entering this market might result in low or no profit at all. For this perspective, I believe that blue cluster that is close to red cluster is that best location to open restaurant.