**A comparison between New York and Toronto**

* **A Battle of Neighbourhoods**

Vibhu Sood

1. **Introduction**
   1. **Background**

Toronto and New York are the financial capitals, respectively, of the second and third largest countries in the world in terms of area.

Toronto is the provincial capital of [Ontario](https://en.wikipedia.org/wiki/Ontario) and the [most populous city in Canada](https://en.wikipedia.org/wiki/List_of_the_100_largest_municipalities_in_Canada_by_population), with a population of 2,731,571 spread across 630,2 Km2.

Before 1998, Toronto was a much smaller municipality and formed part of [Metropolitan Toronto](https://en.wikipedia.org/wiki/Metropolitan_Toronto). When the [city amalgamated that year](https://en.wikipedia.org/wiki/Amalgamation_of_Toronto), Toronto grew to encompass the boroughs of [York](https://en.wikipedia.org/wiki/York,_Toronto), [East York](https://en.wikipedia.org/wiki/East_York), [North York](https://en.wikipedia.org/wiki/North_York), [Etobicoke](https://en.wikipedia.org/wiki/Etobicoke), and [Scarborough](https://en.wikipedia.org/wiki/Scarborough,_Toronto) i.e. about 103 neighbourhoods.

New York City (NYC) is the [most populous city](https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population) in the [United States](https://en.wikipedia.org/wiki/United_States) with an estimated population of 8,398,748 distributed over about 784 Km2. New York is also the [most densely populated](https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population_density) major city in the United States.

New York City is composed of five [boroughs](https://en.wikipedia.org/wiki/Boroughs_of_New_York_City), each of which is a [county of the State of New York](https://en.wikipedia.org/wiki/List_of_counties_in_New_York). The five boroughs–[Brooklyn](https://en.wikipedia.org/wiki/Brooklyn), [Queens](https://en.wikipedia.org/wiki/Queens), [Manhattan](https://en.wikipedia.org/wiki/Manhattan), [the Bronx](https://en.wikipedia.org/wiki/The_Bronx), and [Staten Island](https://en.wikipedia.org/wiki/Staten_Island)–were consolidated into a single city in 1898 composing of about 306 neighbourhoods.

Every time one plans to travel, one faces a dilemma about which city/country to choose. One way to tackle this dilemma is contact people who stay there and get an idea about the different venues present in the city, which are worth visiting and then comparing different cities. A better way to make comparisons is exactly what I tried to achieve using this project.

* 1. **Interested audience**

The target audience for such a project is twofold.

Firstly, any person who plans to visit a country in the western region of the globe. This can help him make the decision regarding which city to visit.

Secondly, a company can use this information to create a website or a mobile application, which is updated on a regular basis, to allow individuals to compare two cities by expanding this to several cities.

1. **Data**
   1. **Data Sources**

For Toronto, I scraped the dataset from the following Wikipedia page : <https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=942851379>

And the location data was available at ‘<http://cocl.us/Geospatial_data>”

I used these two files and joined them into one pandas data-frame.

For New York, a JSON file was luckily available at <https://cocl.us/new_york_dataset>

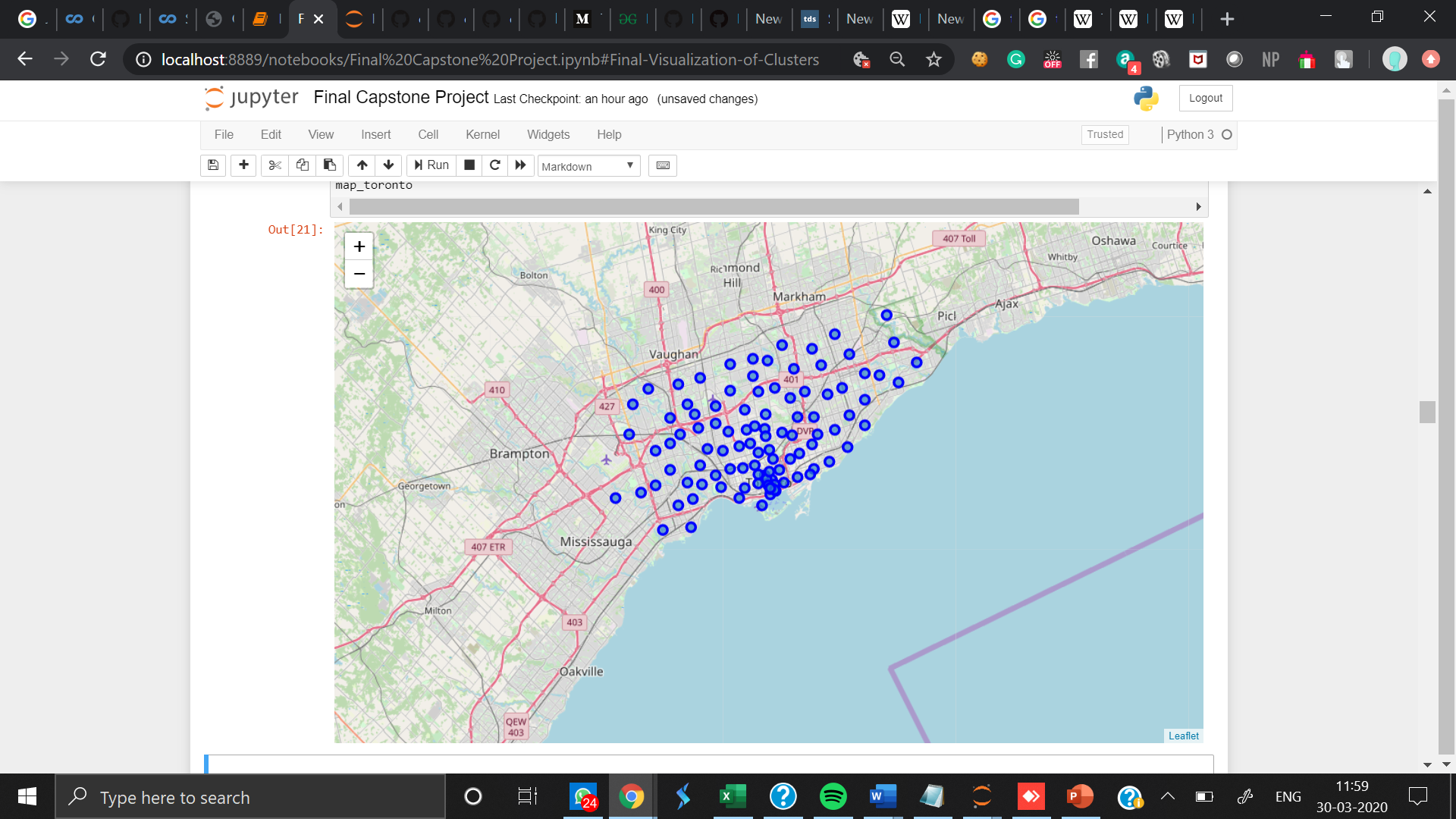
I converted the JSON file into a pandas data-frame for further analysis.

From the Foursquare API , I retrieved the following for each venue in proximity to all the neighbourhoods:

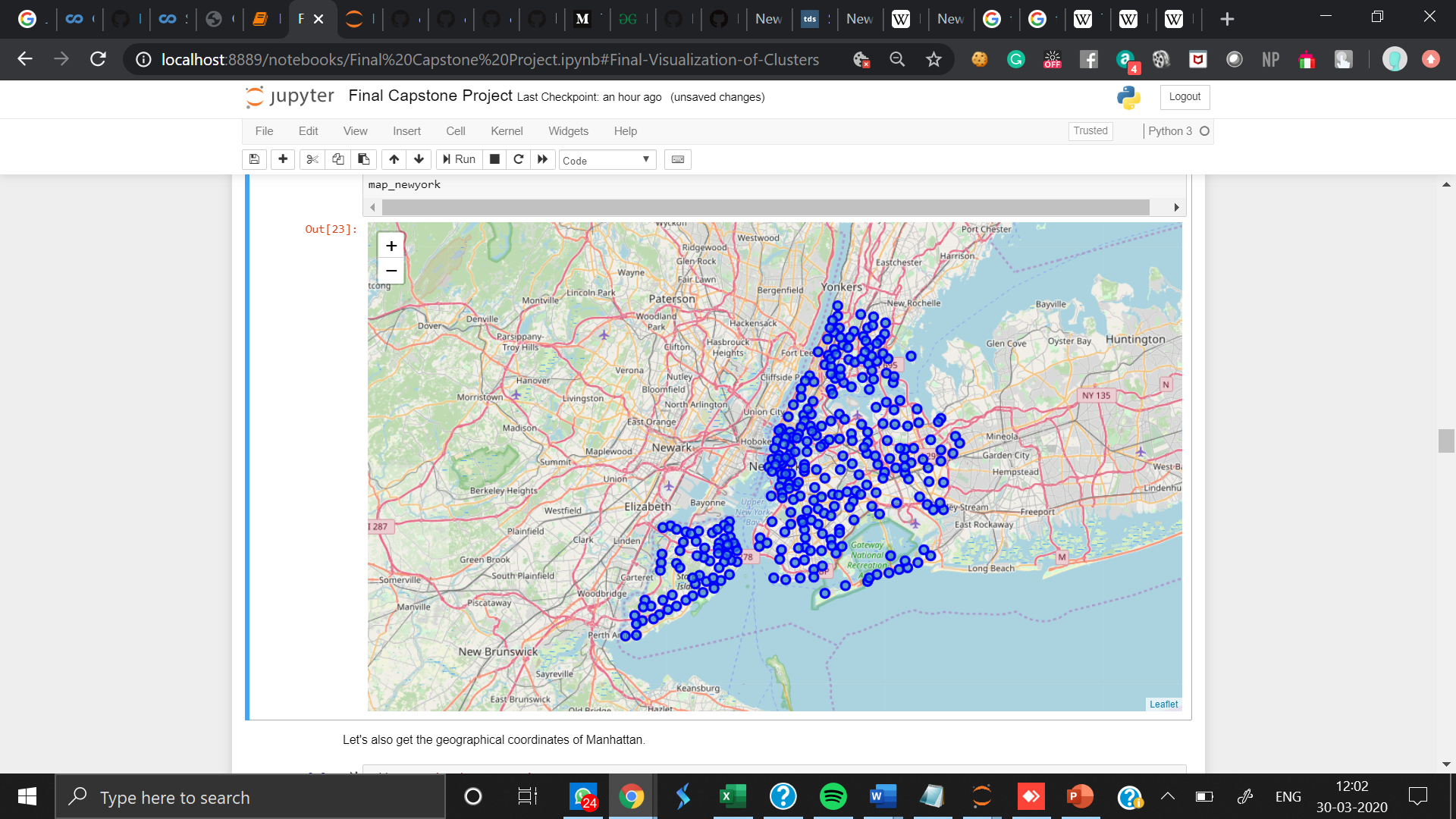
* **Name:** The name of the venue.
* **Category:** The category type as defined by the API.
* **Latitude:** The latitude value of the venue.
* **Longitude:** The longitude value of the venue.

And lastly, GEOPY library was used to fetch the longitude and latitude values of both Toronto and New York.

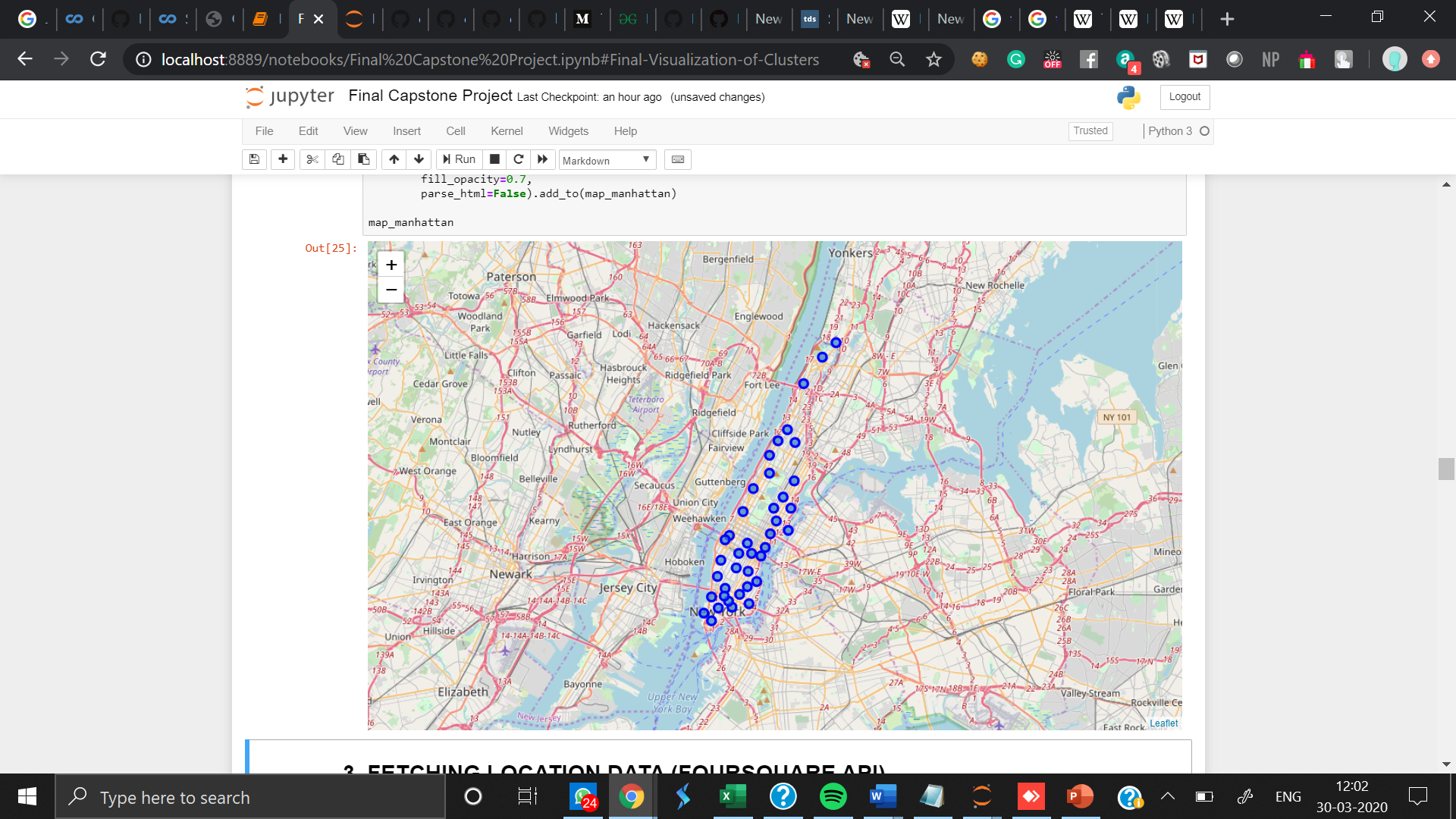
* 1. **Data Visualization:**



A map of Toronto with neighbourhoods superimposed on top.



A map of New York City with neighbourhoods superimposed on top.



A map of Manhattan, New York City with neighbourhoods superimposed on top.

1. **Methodology and Exploratory Data Analysis**

As a first step, I retrieve the venues in Toronto and New York from Foursquare API.

I extract the location data from the Foursquare API for all venues up to a distance of 500 m from each of the neighborhood of Toronto as well as New York City.

After analyzing the venues returned I observed that Toronto had an estimated total of 2188 venues whereas NYC has 10259 of the same.

Clearly comparing 2188 venues to 10259 venues doesn't make sense.

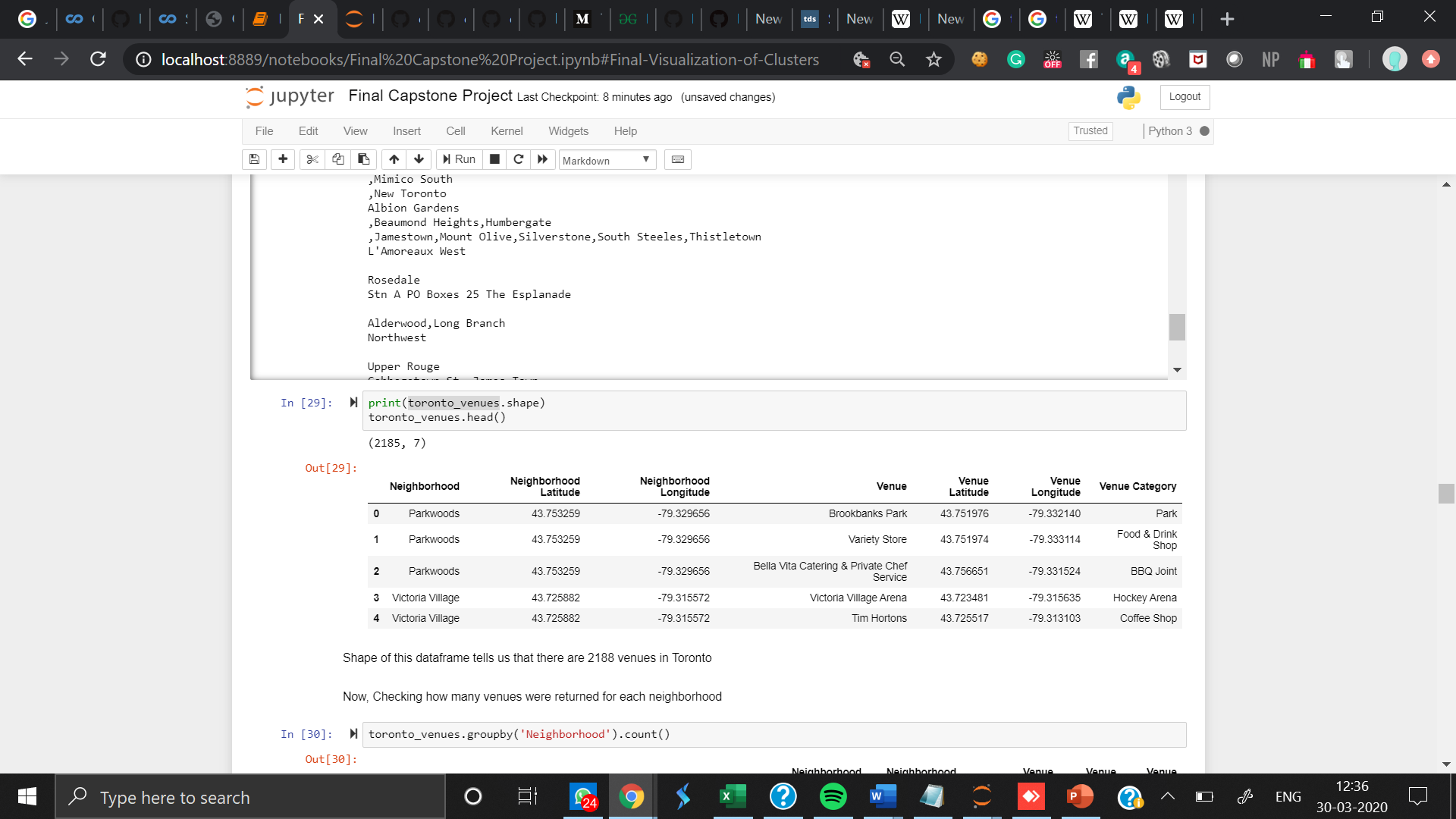
Thus, I decided on going ahead with comparing Manhattan's Neighbourhoods with those of Toronto instead of NYC

Then I used K-mean clustering see if we can draw meaningful information out of what kind of venues exist in New York and Toronto.

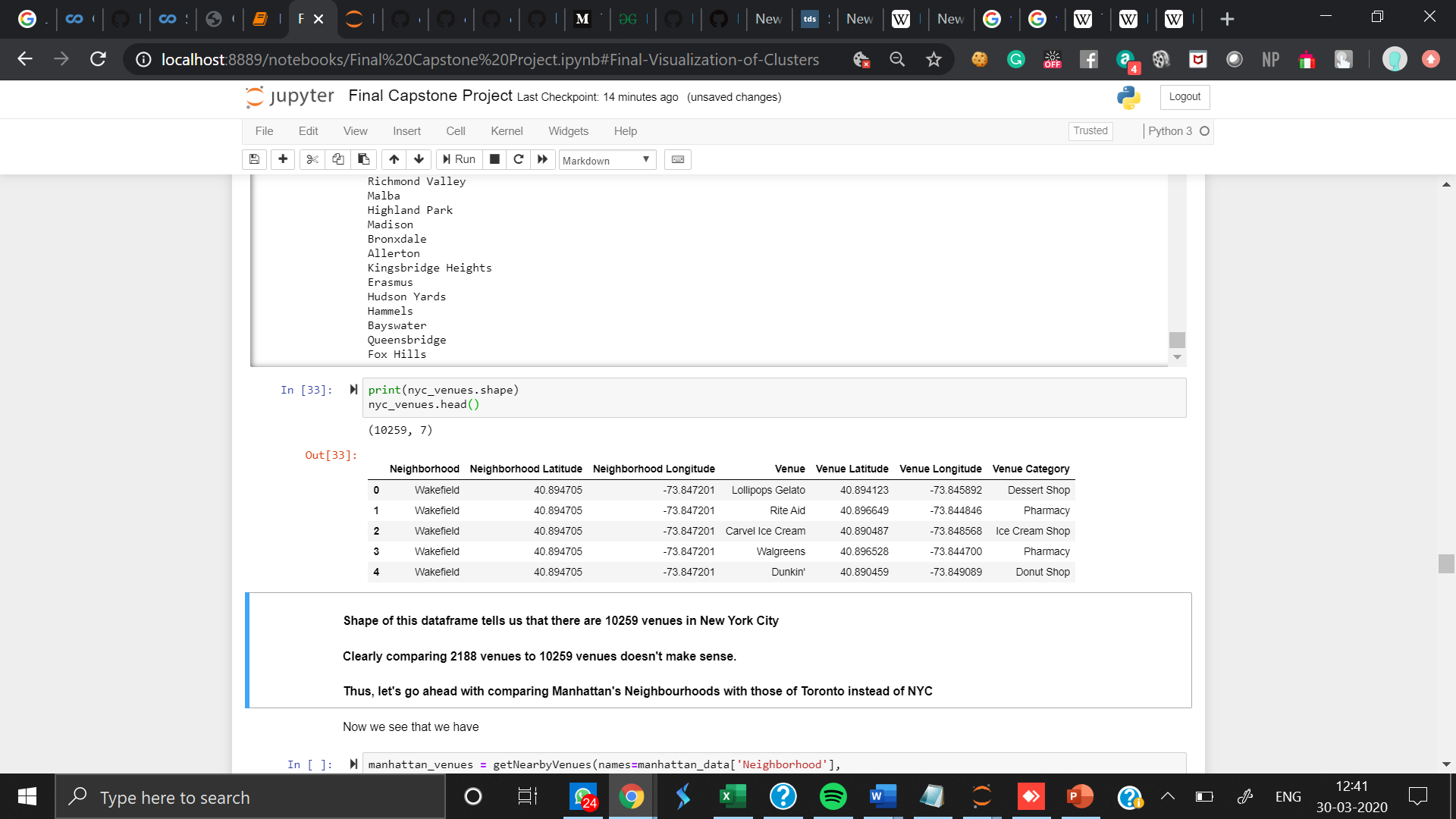
As a final step, I will analyse these plots and try to draw conclusions about differences between the two cities.

* 1. **Exploring the venues around neighbourhoods**

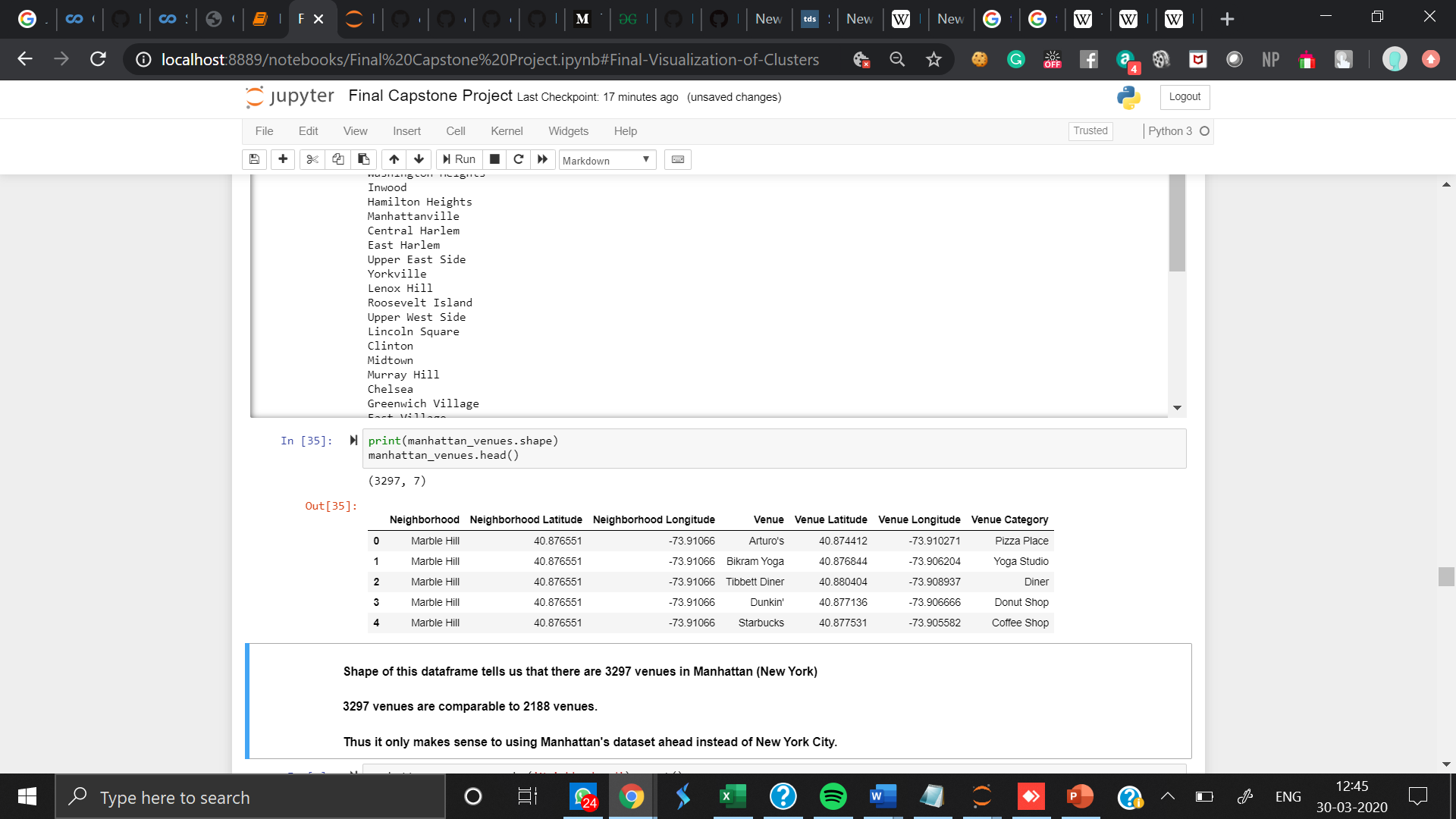
1. TORONTO



1. NEW YORK CITY

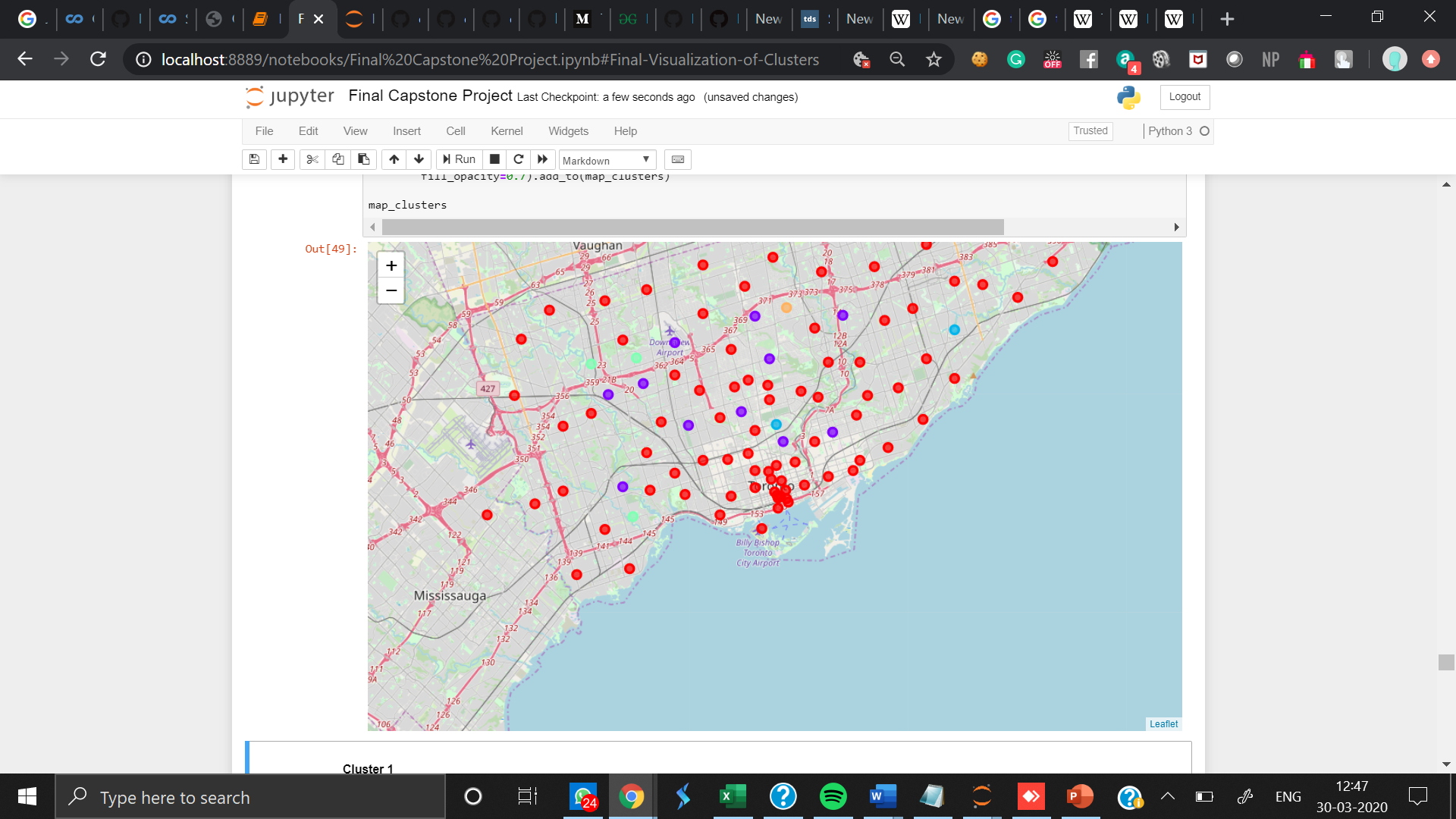


1. MANHATTAN

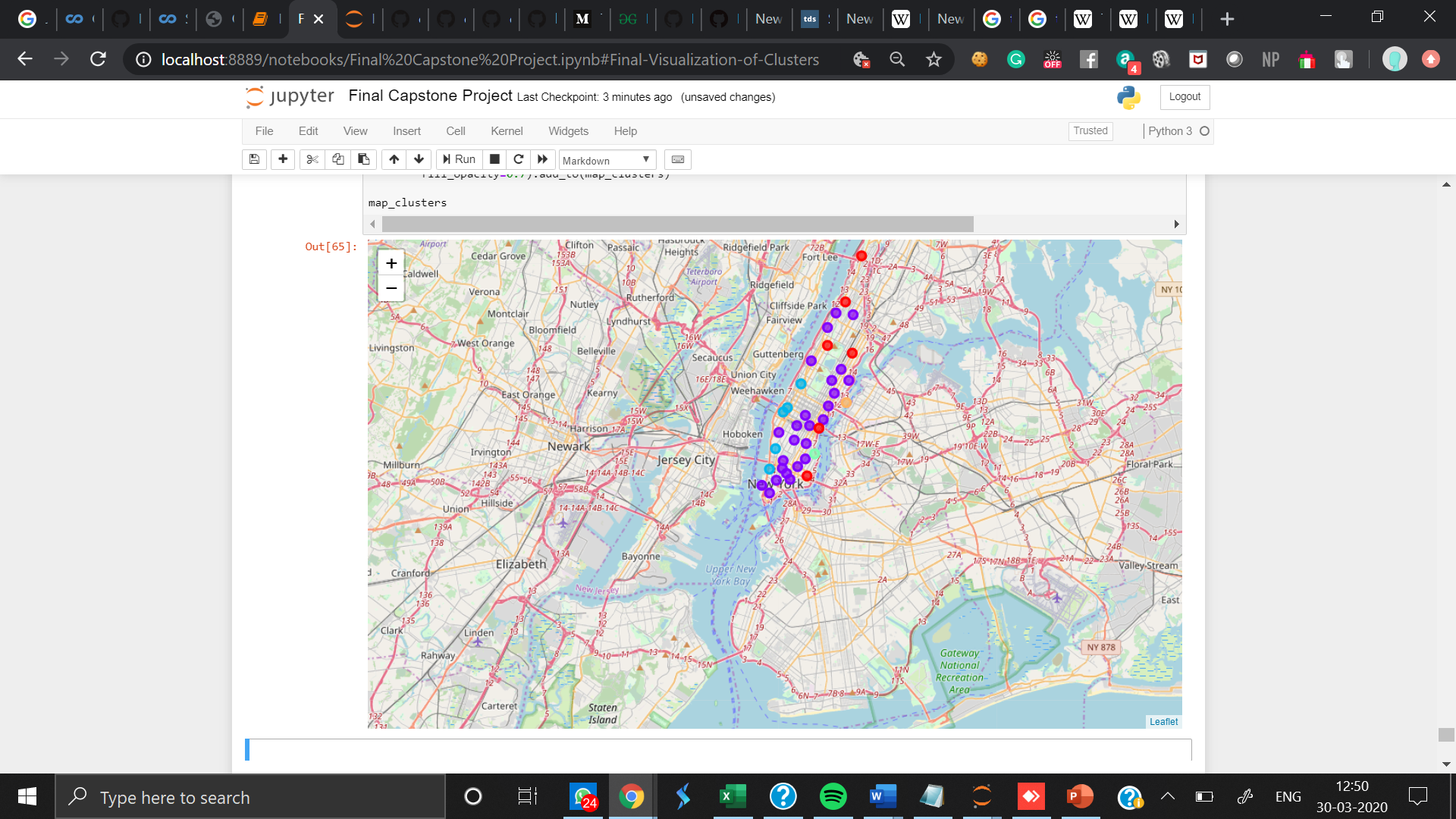


* 1. **CLUSTERING**

Finally, I cluster all the venues based on their category, location and more to identify similar venues and the relationship amongst them. I used KMeans clustering and decided to cluster the venues into two separate groups.



Clustering of neighbourhoods in Toronto



Clustering of neighbourhoods in Manhattan

1. **Results and Discussion**
2. In Toronto we observe varied clusters, while on the other hand Manhattan had repetitive clusters containing food chains and coffee shop. Thus, we can conclude that Manhattan has higher number of venues serving food.

Thus for a young foodie person, Manhattan would be a better place to visit.

1. We can also conclude that night life in Manhattan would be better than Toronto because we observe more clusters containing variety of venues like Restaurants and Pubs.
2. In Toronto we observe that is an entire cluster of Parks, whereas in Manhattan we can find Parks and Playgrounds among clusters of other venues. Thus, we can conclude

that for a health-cautious person, Toronto would be better place to settle.

1. In Toronto we observe that is an entire cluster of Baseball fields, whereas in Manhattan we hardly find any sports complexes among clusters of other venues. Thus, we can conclude that for a sports-oriented person, Toronto would be better place to settle.
2. One common thing that we observe is that in both the cities, people prefer Italian cuisines and coffee shops over other types of Restaurants