

Similar Cluster in New York and Toronto.

IBM DATA SCIENCE

IRVIN FERNANDO GUZMÁN GONZÁLEZ

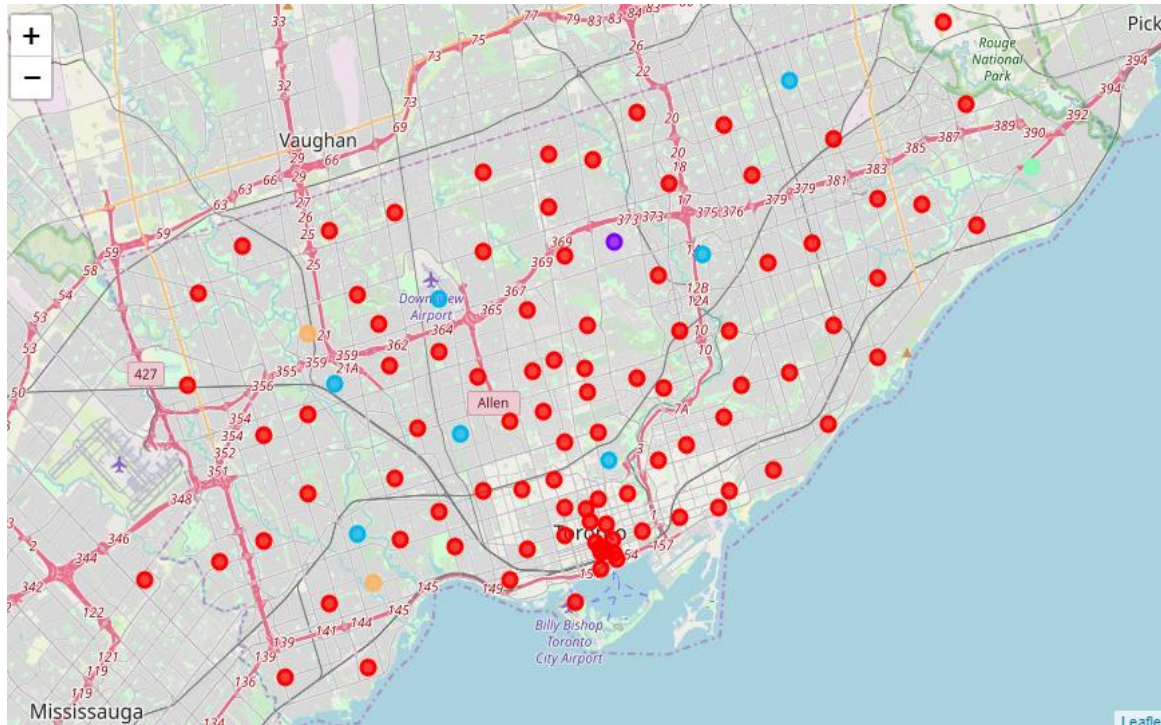
Introduction

- It is supposed to be a commerce located on a certain neighborhood on New York city, and the owner wants to open another one on the city of Toronto, Canada.
- What he wants to know is which section of Toronto is like the one on New York. Recognizing which areas are similar would give an idea where to search for a place to establish the new commerce on Toronto and expect a similar behavior on the sells since both cities are big cities.
- The main problem is to identify which areas are similar between New York and Toronto

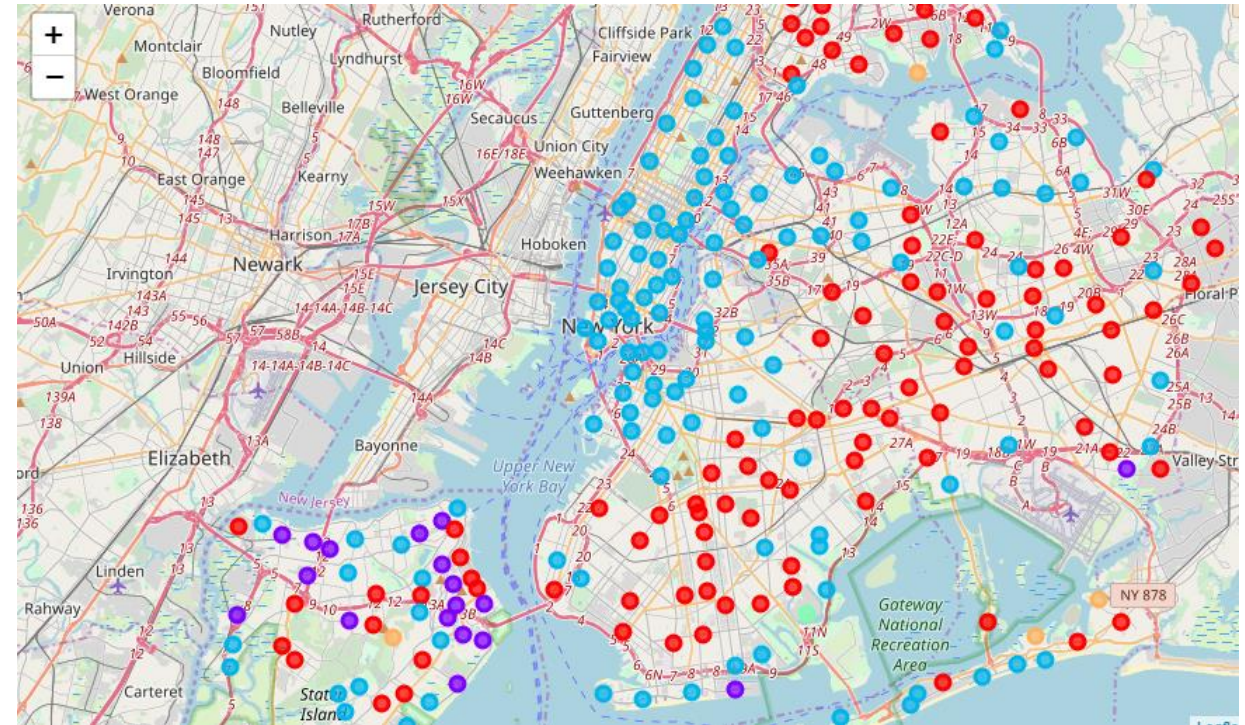
Data

- The data to be used are the geolocalization of postal codes in Toronto, Canada and the geolocalization of the Neighborhoods on New York city.
- The geolocalization will be used to obtain the venues around each localization from the Foursquare database.
- Total of 12453 venues on 461 categories

Maps of 5-means clustering on both cities separated

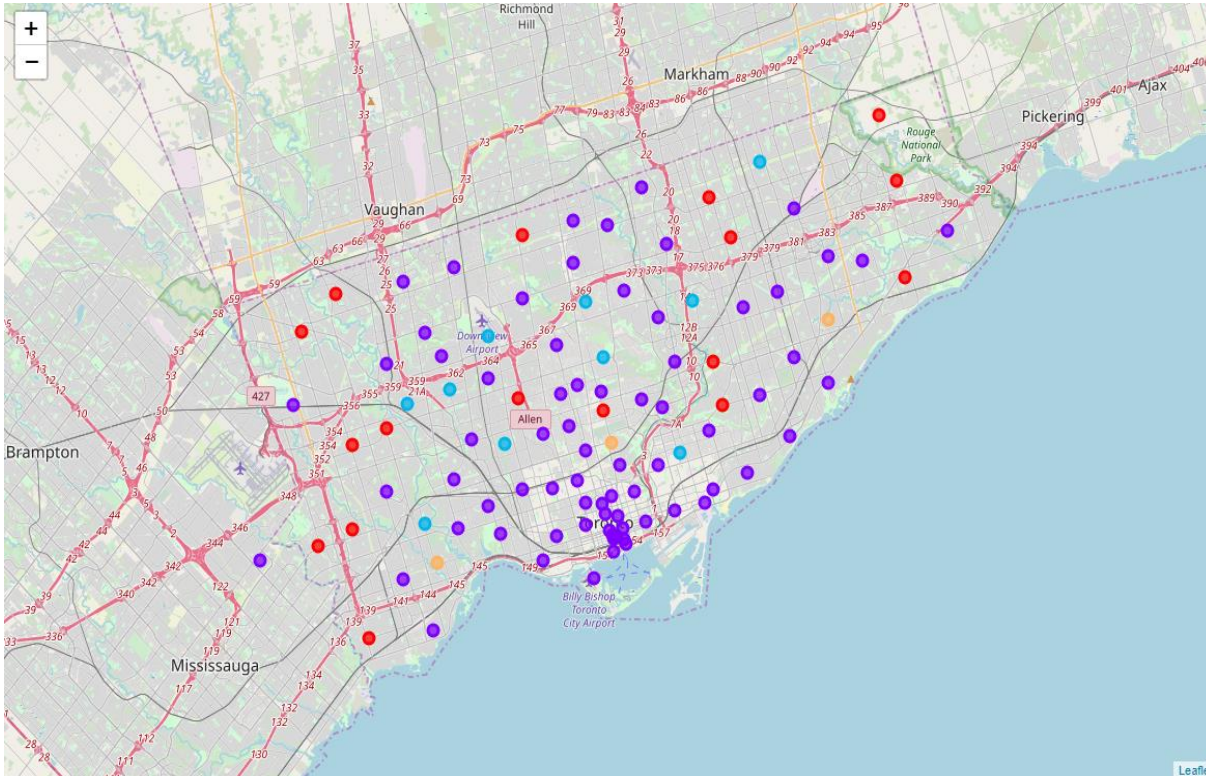


Toronto

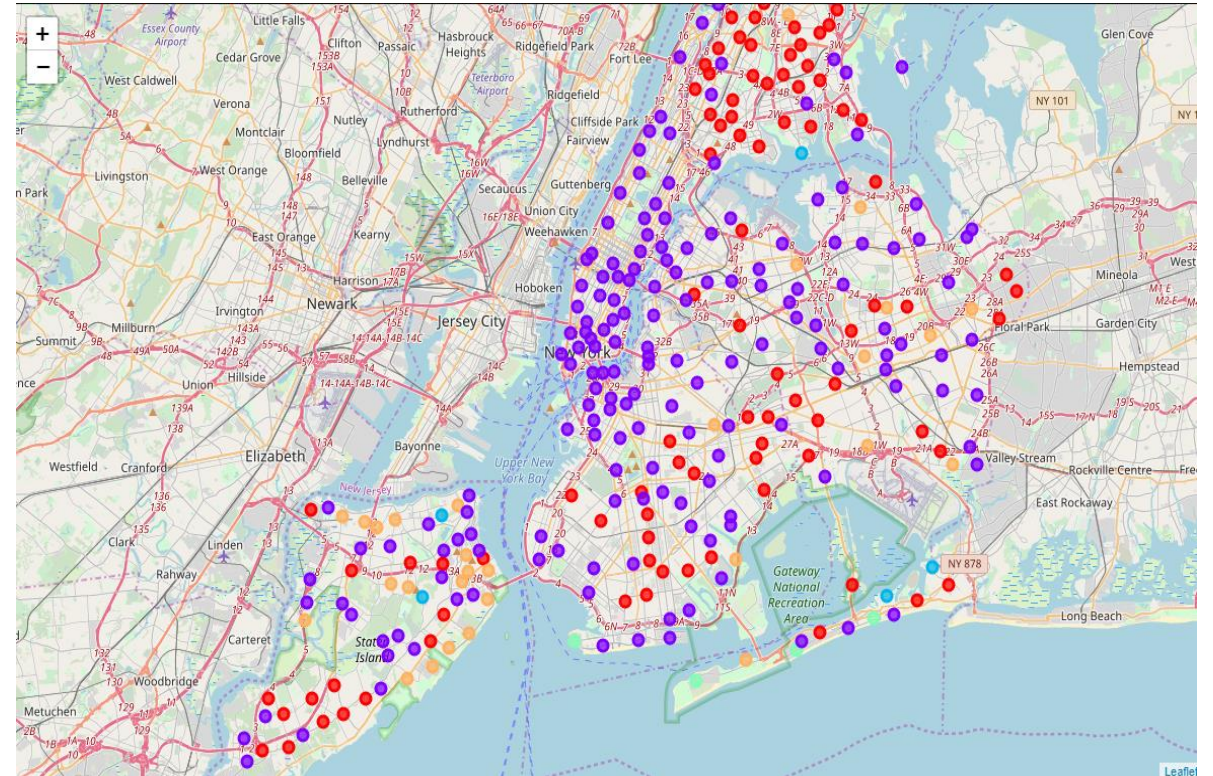


New York

Maps of both cities with 5-means clustering at same time



Toronto



New York

Conclusion

- In conclusion, with the obtained results, can be observed which similarities are between different area of Toronto and New York. It could help to decide where could be a good area to locate a new commerce taking reference from another one that is already establish and was successful. Therefore, if it is assigned a commerce on New York where is successful commerce just left to choose a region on Toronto that its assigned cluster match the color with the New York one.
- To obtain more reliable result is recommendable to obtain another source of data for the venues on Toronto due to a difference on quantity of venues compared to New York. It is thought that is because a lack of information on the Foursquare database about Toronto, and not by an actual considerably difference of venues. Also is desired to design an algorithm or manual process to identify similar types of venues since there are venues on both cities that are very alike but are named differently.