DATA SCIENCE PROJECT: BATTLE OF NEIGHBORHOOD

COURSERA CAPSTONE

- **Introduction**

- **Problem solving**

- **Processing Data**

- **Cluster map**
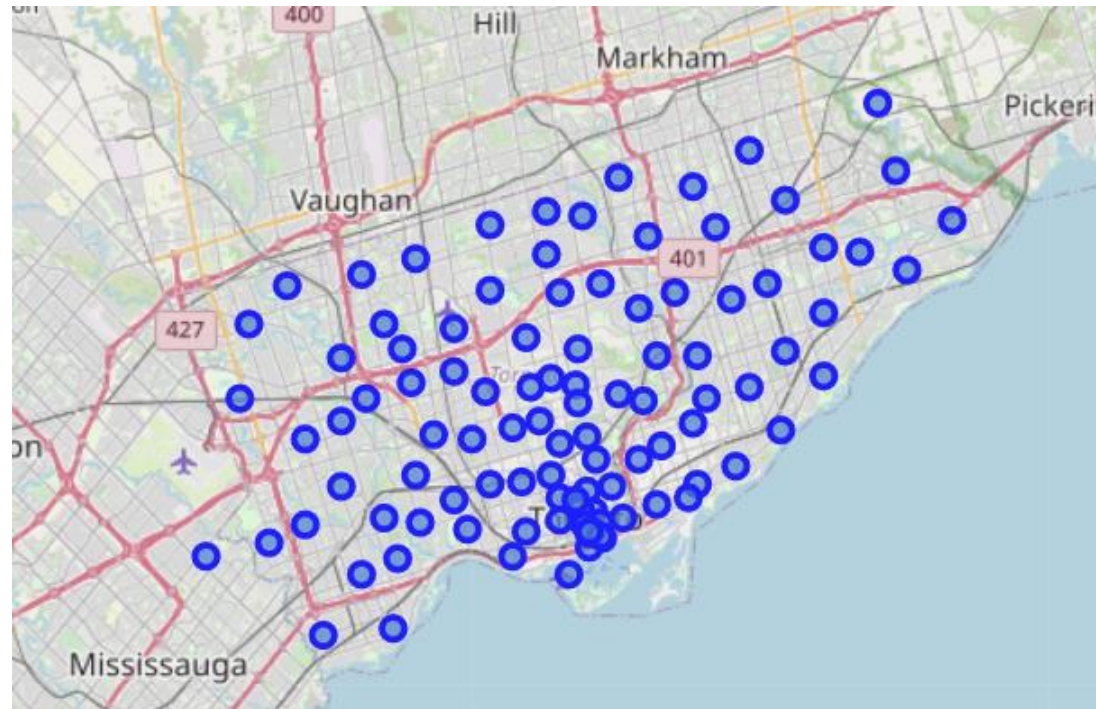
- **Discussion**

- **Conclusion**

INTRODUCTION

- I will be required to explore, segment, and cluster the neighborhoods in the city of NEW York by using Foursquare API

- Then I will also explore the data in the city of Toronto .However, unlike New York, the neighborhood data is not readily available on the internet. What is interesting about the field of data science is that each project can be challenging in its unique way, so you need to learn to be agile and refine the skill to learn new libraries and tools quickly depending on the project.

- For the Toronto neighborhood data, a Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto. You will be required to scrape the Wikipedia page and wrangle the data, clean it, and then read it into a pandas dataframe so that it is in a structured format like the New York dataset.

-

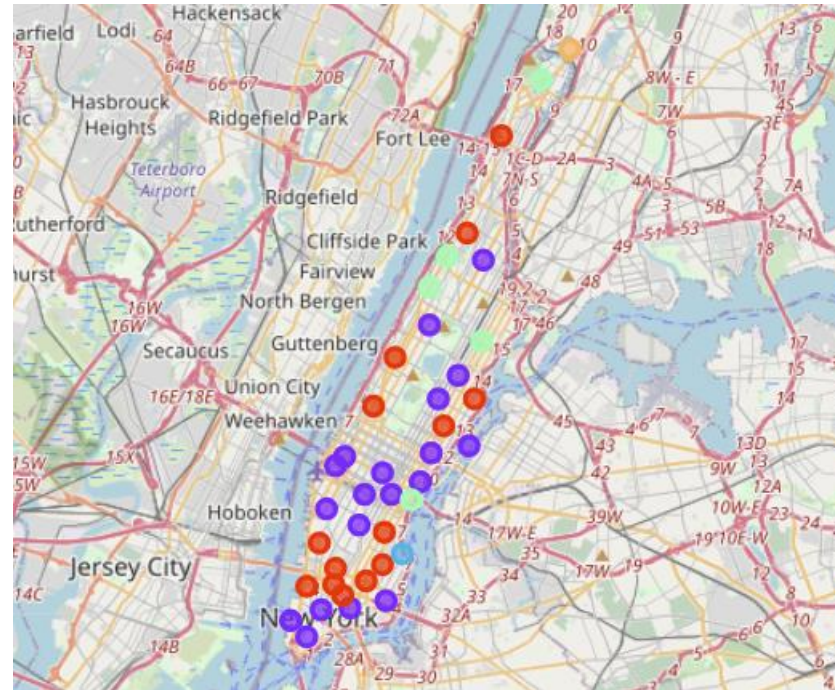# CLUSTING OF TORONTO

FOLIUM.MAP

CLUSTING OF NEW YORK

FOLIUM.MAP

DISCUSSION&CONCLUSION

This solution is based purely on the geographic location, which is quite good in most cases. Since in the business of cafe, the right location is almost THE important decision to make. Nevertheless, if we could find more information on each cluster, it will improve the solution further.

The clustering is very useful and intuitive to solve the problem related to geographic data. We should also pay attention to the outcomes of the clustering, especially when it yields a result different ideas of a 'business circle' in a city. Especially it shown somehow concentration near the streamside. It may because the travel covnience(via water way) and landscape.

THANKS FOR WATCHING