Data Science Capstone

North York, Toronto

Sankhadeep Dutta

Introduction / Business Problem

- The Brazilian tourism agency plans to organize their first trip to Toronto in Canada, and they plan to start with the city of North York.

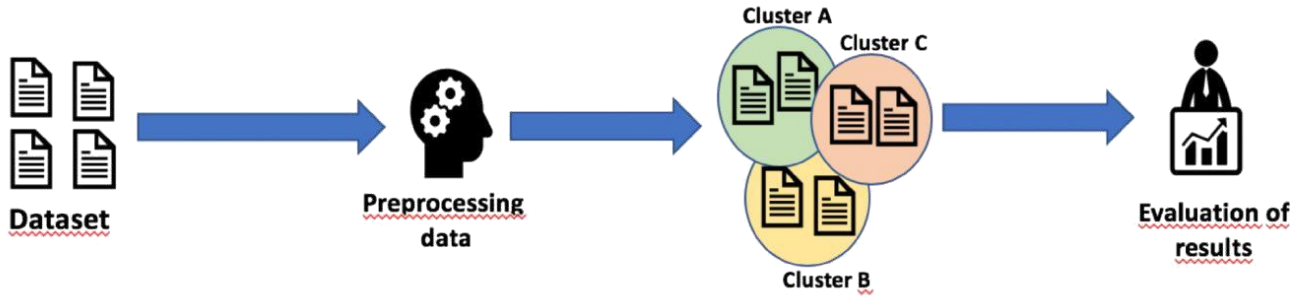
  To create the best itinerary for your clients, the travel agency is looking for which districts near North York visit, for that, he had commissioned an analysis of characteristics and the most relevant places in these neighborhoods. Mainly, the tourism agency focuses on hotels, restaurants, parks, shops, places, squares, etc.

- Question: So, what are the characteristics of the neighborhoods neighboring North York and what places should the Brazilian tourism agency visit to provide the best tour to its clients?

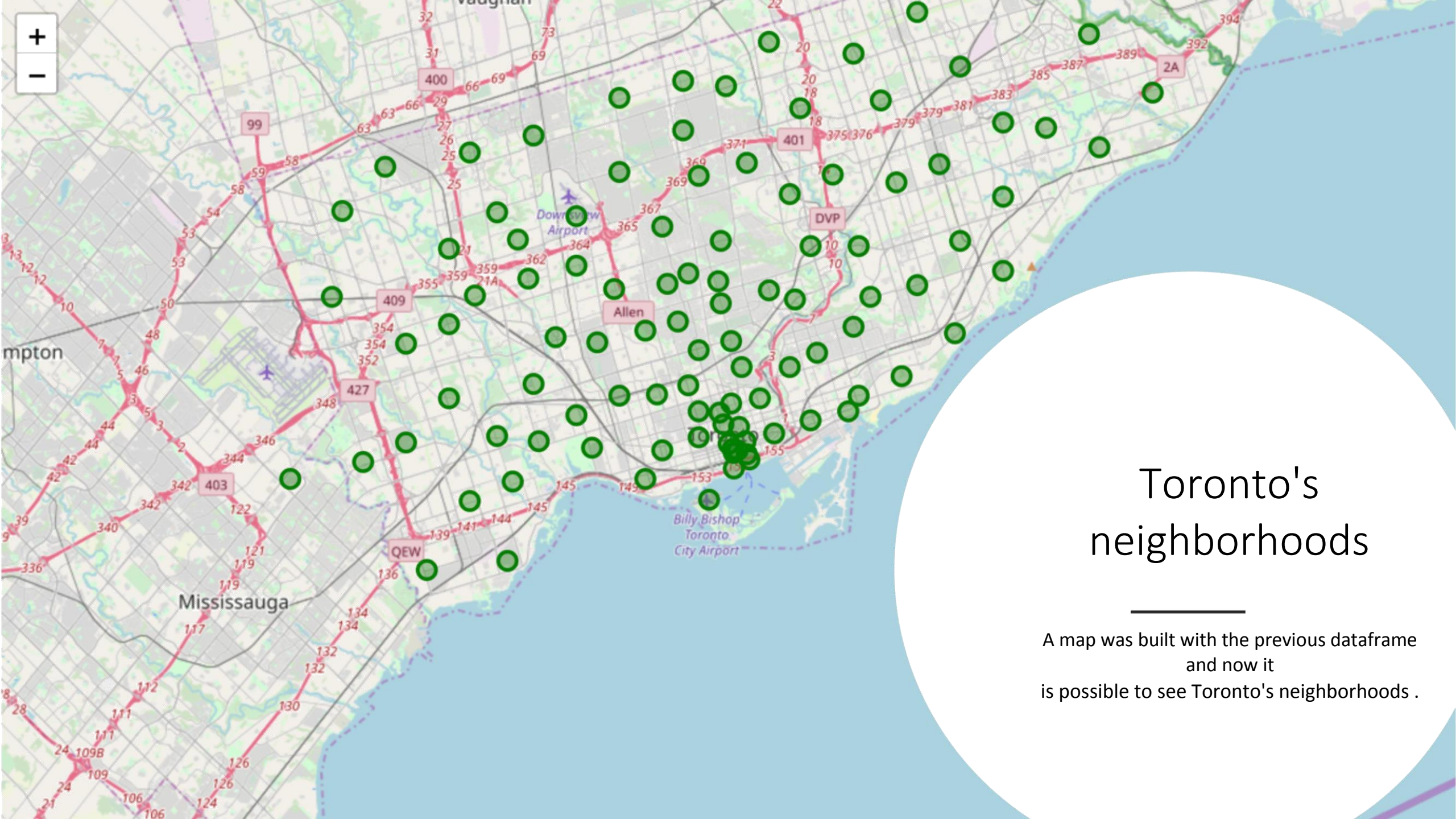
Data description

- To solve this problem, I'll first get the data on the neighborhoods of North York (Toronto): Borough, Latitude, Longitude. This data is available on Wikipedia, so with a simple script, I was able to access the data in tabular format. Below is the example of this dataframe.
- The second step, with the help of the API of Foursquare, got the information of the best places of the neighborhood and its neighbors. With a new dataframe with the complete information, you're all set to cluster the entire dataset.

Methodology



- This section describes what steps have been taken during project development. Here you will find pictures about the workflow, maps, explanations of procedures performed and other relevant information.



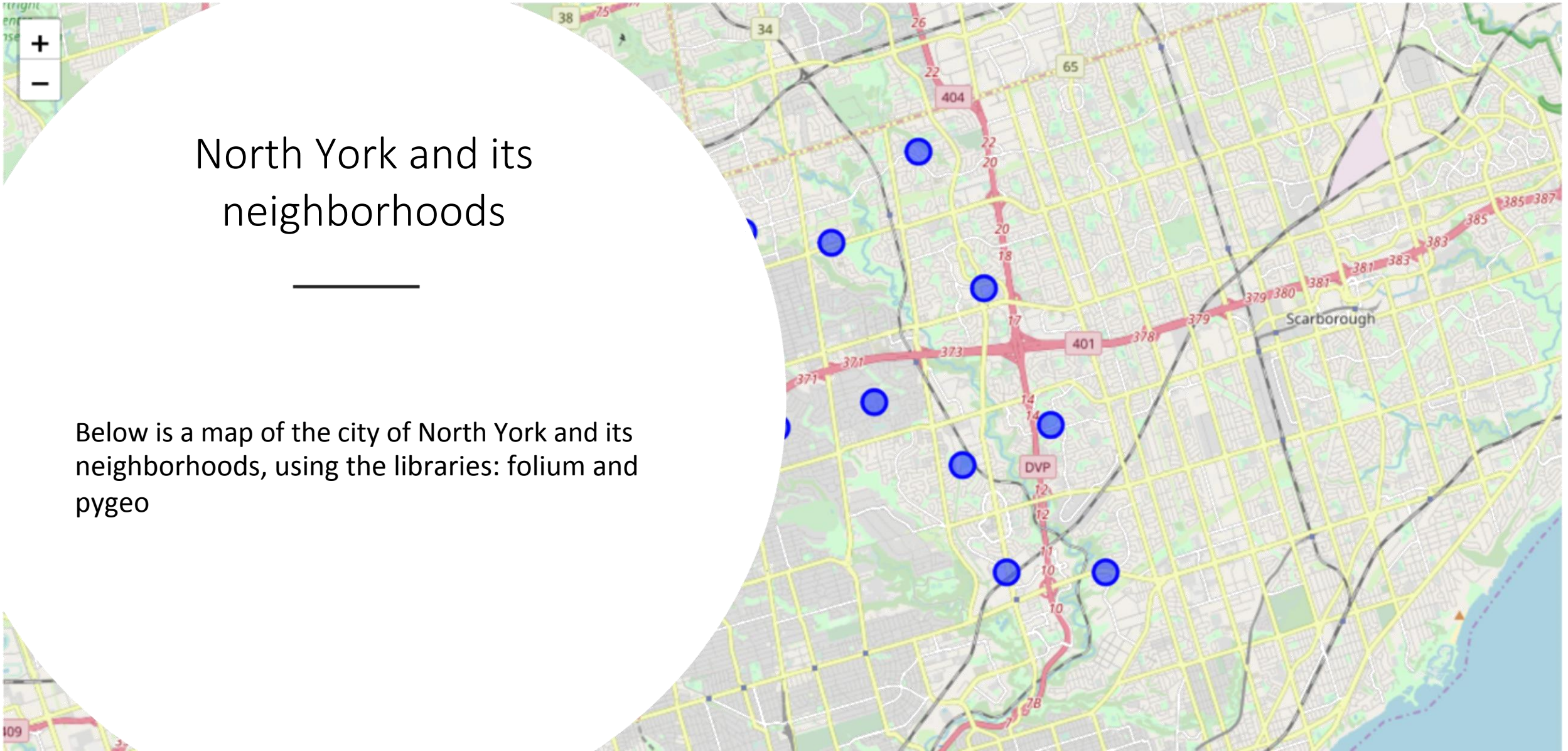
Toronto's neighborhoods

A map was built with the previous dataframe and now it
is possible to see Toronto's neighborhoods .

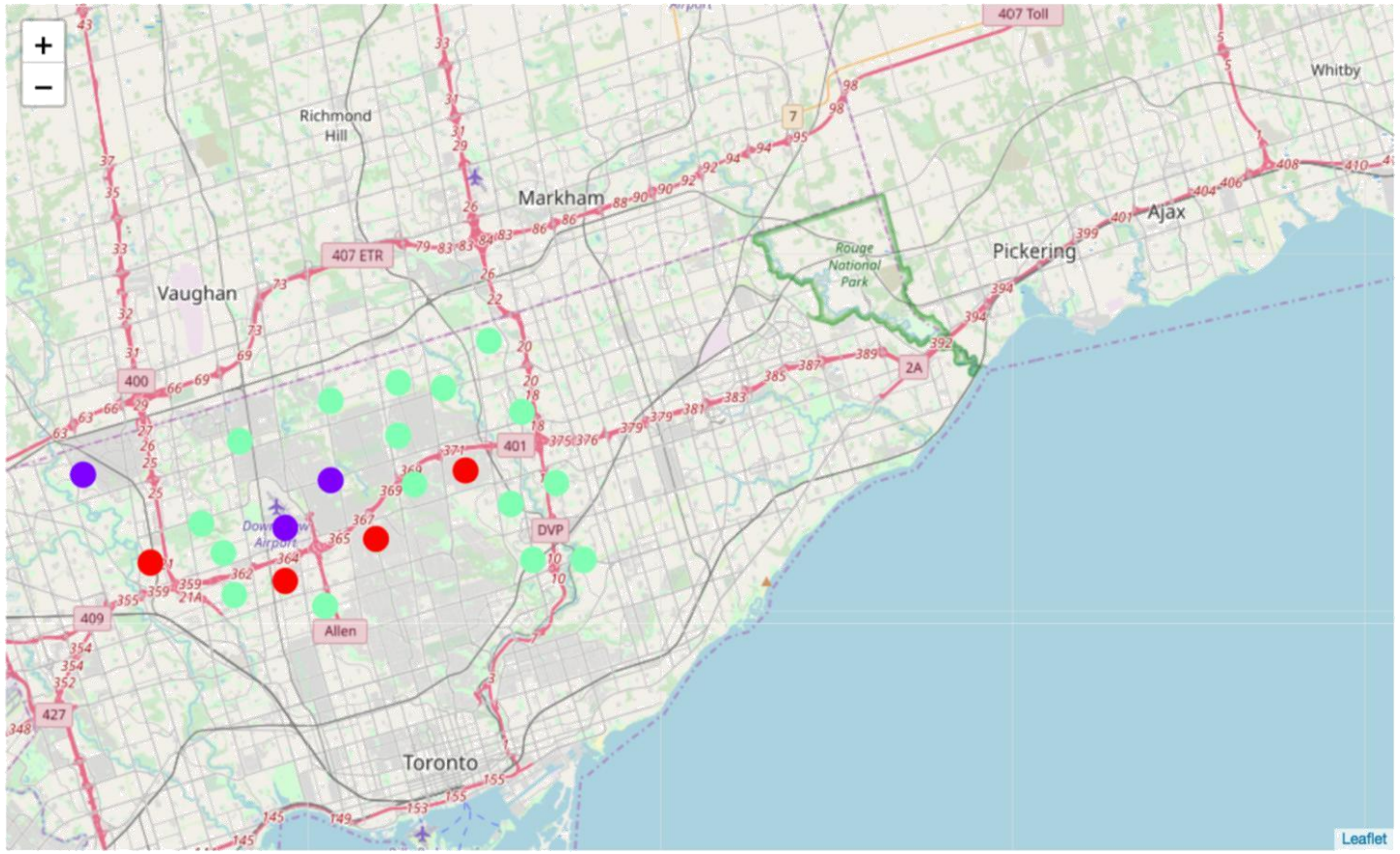
```
        popup=label,
        color='blue',
        fill=True,
        fill_opacity=0.5).add_to(map_nyork)

map_nyork
```

North York and its neighborhoods

Below is a map of the city of North York and its neighborhoods, using the libraries: folium and pygeo



Results



- Below, the map plotted with clusters generated by Kmeans. We can notice the similarity of several neighborhoods with regard to the category of venues

Cluster label 0

```
nyork_merged.loc[nyork_merged['Cluster Labels'] == 0, nyork_merged.columns[[1] + list(range(5, nyork_merged.shape[1]))]]
```

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	North York	0	Park	Martial Arts Dojo	Clothing Store	Coffee Shop	Comfort Food Restaurant	Construction & Landscaping	Cosmetics Shop	Deli / Bodega	Department Store	Dim Sum Restaurant
18	North York	0	Coffee Shop	Fast Food Restaurant	Italian Restaurant	Indian Restaurant	Comfort Food Restaurant	Café	Liquor Store	Pharmacy	Pizza Place	Juice Bar
19	North York	0	Clothing Store	Furniture / Home Store	Accessories Store	Event Space	Miscellaneous Shop	Boutique	Coffee Shop	Vietnamese Restaurant	Fried Chicken Joint	Food Truck
23	North York	0	Furniture / Home Store	Baseball Field	Women's Store	Empanada Restaurant	Comfort Food Restaurant	Construction & Landscaping	Cosmetics Shop	Deli / Bodega	Department Store	Dim Sum Restaurant

Cluster label 1

```
nyork_merged.loc[nyork_merged['Cluster Labels'] == 1, nyork_merged.columns[[1] + list(range(5, nyork_merged.shape[1]))]]
```

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
11	North York	1	Coffee Shop	Middle Eastern Restaurant	Bank	Bridal Shop	Restaurant	Diner	Sandwich Place	Shopping Mall	Pizza Place	Pharmacy
13	North York	1	Park	Airport	Bus Stop	Electronics Store	Coffee Shop	Comfort Food Restaurant	Construction & Landscaping	Cosmetics Shop	Deli / Bodega	Department Store
22	North York	1	Pizza Place	Empanada Restaurant	Dog Run	Clothing Store	Coffee Shop	Comfort Food Restaurant	Construction & Landscaping	Cosmetics Shop	Deli / Bodega	Department Store

Cluster label 2

```
nyork_merged.loc[nyork_merged['Cluster Labels'] == 2, nyork_merged.columns[[1] + list(range(5, nyork_merged.shape[1]))]]
```

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	North York	2	Dog Run	Golf Course	Pool	Athletics & Sports	Mediterranean Restaurant	Women's Store	Comfort Food Restaurant	Construction & Landscaping	Cosmetics Shop	Deli / Bodega
1	North York	2	Clothing Store	Fast Food Restaurant	Women's Store	Coffee Shop	Food Court	Restaurant	Tea Room	Kids Store	Bakery	Health Food Store
2	North York	2	Chinese Restaurant	Café	Bank	Japanese Restaurant	Electronics Store	Comfort Food Restaurant	Construction & Landscaping	Cosmetics Shop	Deli / Bodega	Department Store
4	North York	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	North York	2	Ramen Restaurant	Restaurant	Sandwich Place	Café	Fast Food Restaurant	Coffee Shop	Pizza Place	Japanese Restaurant	Middle Eastern Restaurant	Indonesian Restaurant
6	North York	2	Park	Bank	Electronics Store	Coffee Shop	Comfort Food Restaurant	Construction & Landscaping	Cosmetics Shop	Deli / Bodega	Department Store	Dim Sum Restaurant
7	North York	2	Pharmacy	Pizza Place	Coffee Shop	Butcher	Dog Run	Comfort Food Restaurant	Construction & Landscaping	Cosmetics Shop	Deli / Bodega	Department Store
8	North York	2	Park	Food & Drink Shop	Fast Food Restaurant	Women's Store	Dog Run	Coffee Shop	Comfort Food Restaurant	Construction & Landscaping	Cosmetics Shop	Deli / Bodega
9	North York	2	Caribbean Restaurant	Gym / Fitness Center	Café	Pool	Basketball Court	Japanese Restaurant	Dog Run	Comfort Food Restaurant	Construction & Landscaping	Cosmetics Shop
10	North York	2	Gym	Asian Restaurant	Coffee Shop	Beer Store	Bike Shop	Grocery Store	Fast Food Restaurant	Italian Restaurant	Japanese Restaurant	Dim Sum Restaurant
12	North York	2	Caribbean Restaurant	Furniture / Home Store	Miscellaneous Shop	Bar	Massage Studio	Coffee Shop	Women's Store	Dog Run	Construction & Landscaping	Cosmetics Shop
14	North York	2	Moving Target	Shopping Mall	Grocery Store	Bank	Coffee Shop	Comfort Food Restaurant	Construction & Landscaping	Cosmetics Shop	Deli / Bodega	Department Store
15	North York	2	Business Service	Food Truck	Baseball Field	Women's Store	Electronics Store	Comfort Food Restaurant	Construction & Landscaping	Cosmetics Shop	Deli / Bodega	Department Store
17	North York	2	Coffee Shop	Hockey Arena	Portuguese Restaurant	Intersection	Women's Store	Dog Run	Comfort Food Restaurant	Construction & Landscaping	Cosmetics Shop	Deli / Bodega
20	North York	2	Pizza Place	Metro Station	Pub	Japanese Restaurant	Park	Women's Store	Diner	Clothing Store	Coffee Shop	Comfort Food Restaurant
21	North York	2	Park	Construction & Landscaping	Bakery	Basketball Court	Electronics Store	Coffee Shop	Comfort Food Restaurant	Cosmetics Shop	Deli / Bodega	Department Store

- The main challenge of data science applications is certainly data manipulation. We need to get data from different sources, which are not always friendly and easy to manipulate.
- When the data preparation step is unsuccessful, we need to choose the appropriate technique to be applied. The word is to experiment and analyze options.
- Finally, the challenge of a data scientist is to organize the chronology of facts and tell the story. In the course, we learned that the data scientist, in addition to the various skills, need to be able to tell the story and persuade the person involved in the project.

Discussion

Conclusion

- This project analyzed data from different sources and was able to obtain results for the client. The applied process also proved that with a more comprehensive database, we could extrapolate the Toronto boundaries and get insightful insights from other places as well.
- The challenge always ends up being: the data. It is not always possible to get them easily. But with deep research and the curiosity of a data scientist who wants to solve the customer's problem, the sky may be the limit.