Battle of tri-cities: New York, Toronto, Paris

Explore similarity of neighborhoods between the cities

Introduction: where you discuss the business problem and who would be interested in this project.

- The business problem is the overview of the neighborhoods for each city; the similarities / dissimilarities between them

- A traveler or a new employee in the cities

Data: where you describe the data that will be used to solve the problem and the source of the data.

- The data for the three cities include neighborhoods, Borough, Geographical Latitude and Longitude; Foursquare is used for exploring cities and neighborhoods.

- *Data source:*
  - **New York City:** 'newyork_data.json' downloaded from https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json
  - **Toronto:** url 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M' and 'Geospatial_Coordinates.csv'
  - **Paris:** correspondances-code-insee-code-postal.json downloaded from https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e

Methodology: which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

- Mainly *BeautifulSoup* library was used to wrangler the data and arranged data into Pandas dataframe

- K-means machine learning was use to segmentation and clustering the neighborhoods

- This unsupervised machine learning method was used because we did no know the outcome feature priori

- Examined the clustering

- Category set was compared

Results: where you discuss the results

- All the neighborhoods in each city are similar and clustered into one segment

- New York city and Toronto are very similar, which in turn was clustered into one segment

- However, New York city and Paris are quite dissimilar:
  - First, the unique categories is much more for Paris than for New York
  - Second, k-means clustered the common categories between the two city also showed quite dissimilarity, hence, segmented into two clusters

- Plots showed the similarity/dissimilarity by k-means clustering in next few slides
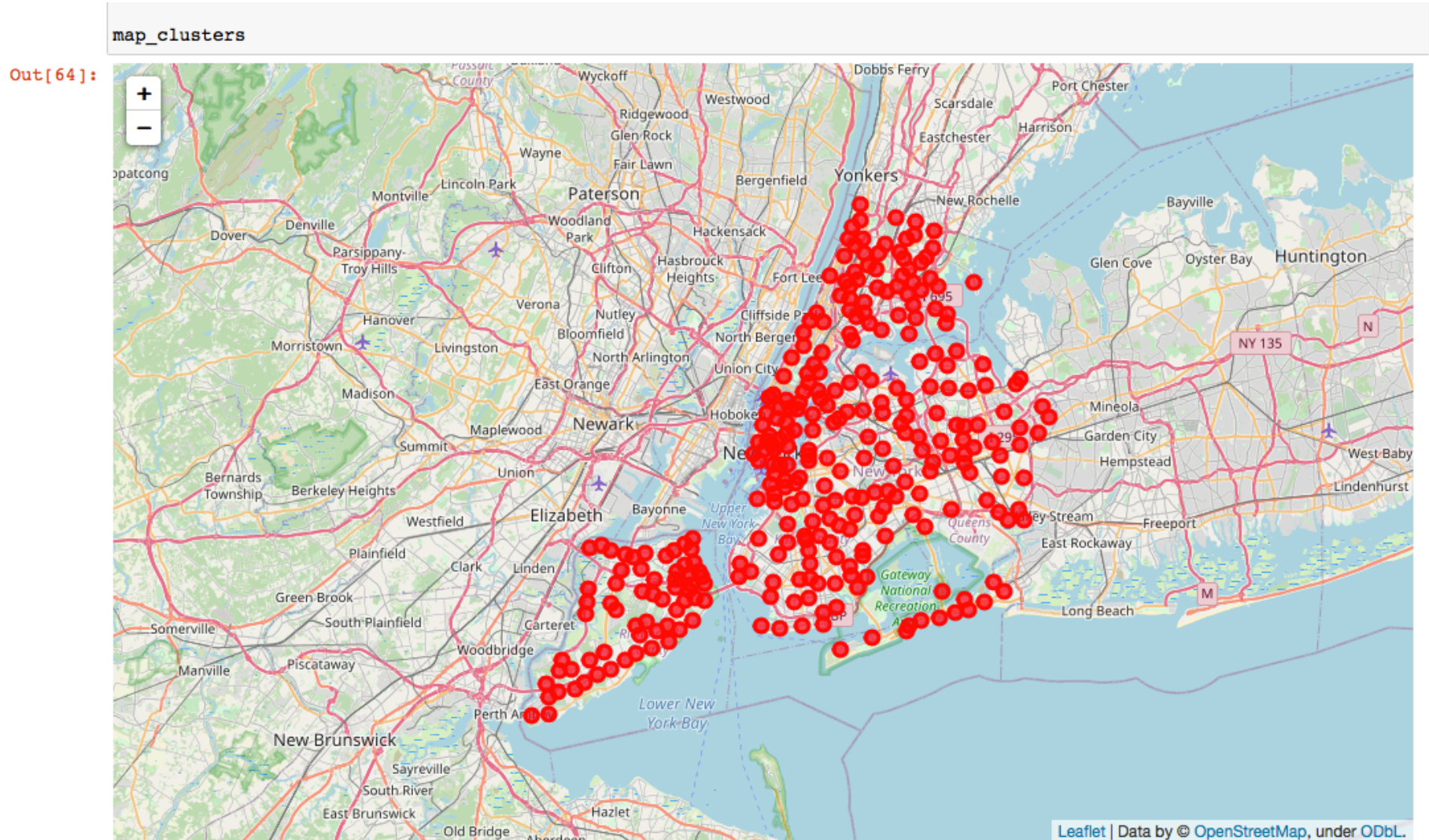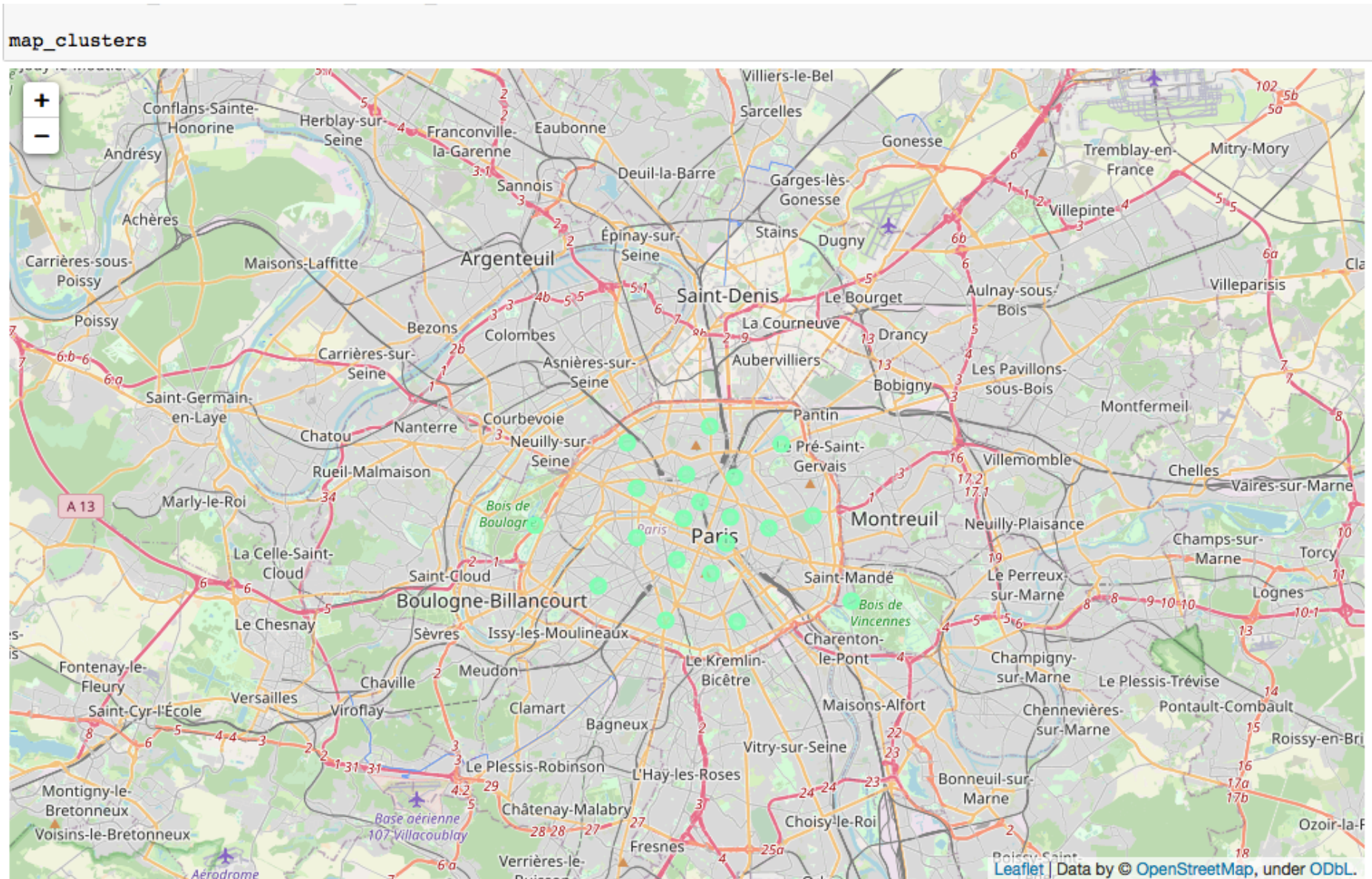
Toronto neighborhoods: similar in one segment
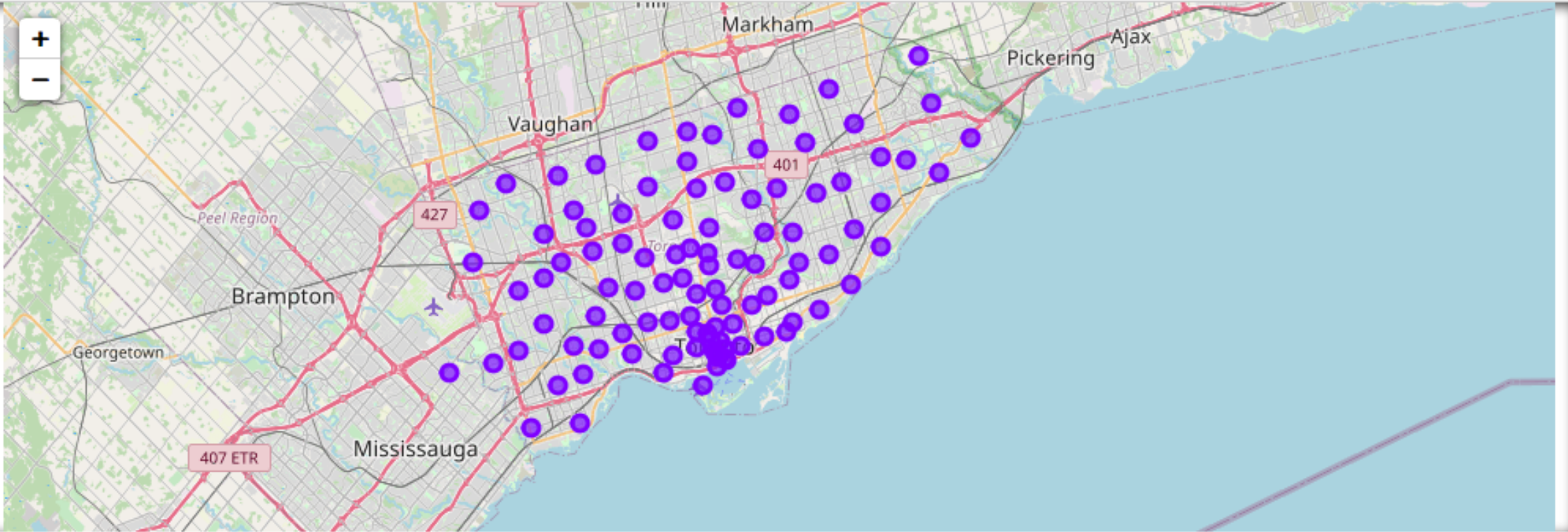
New York city neighborhoods: similar in one segment

Paris neighborhoods: similar in one segment

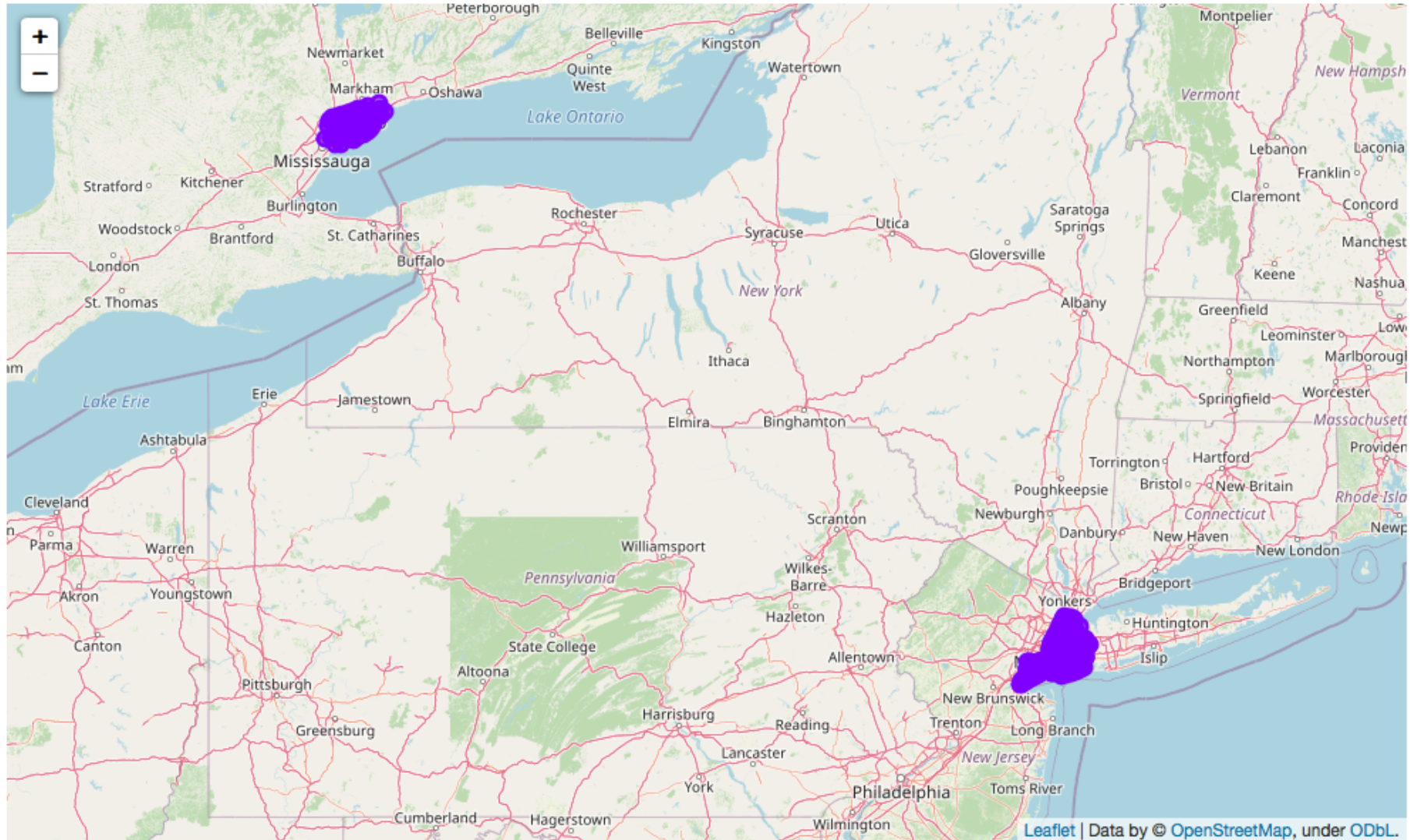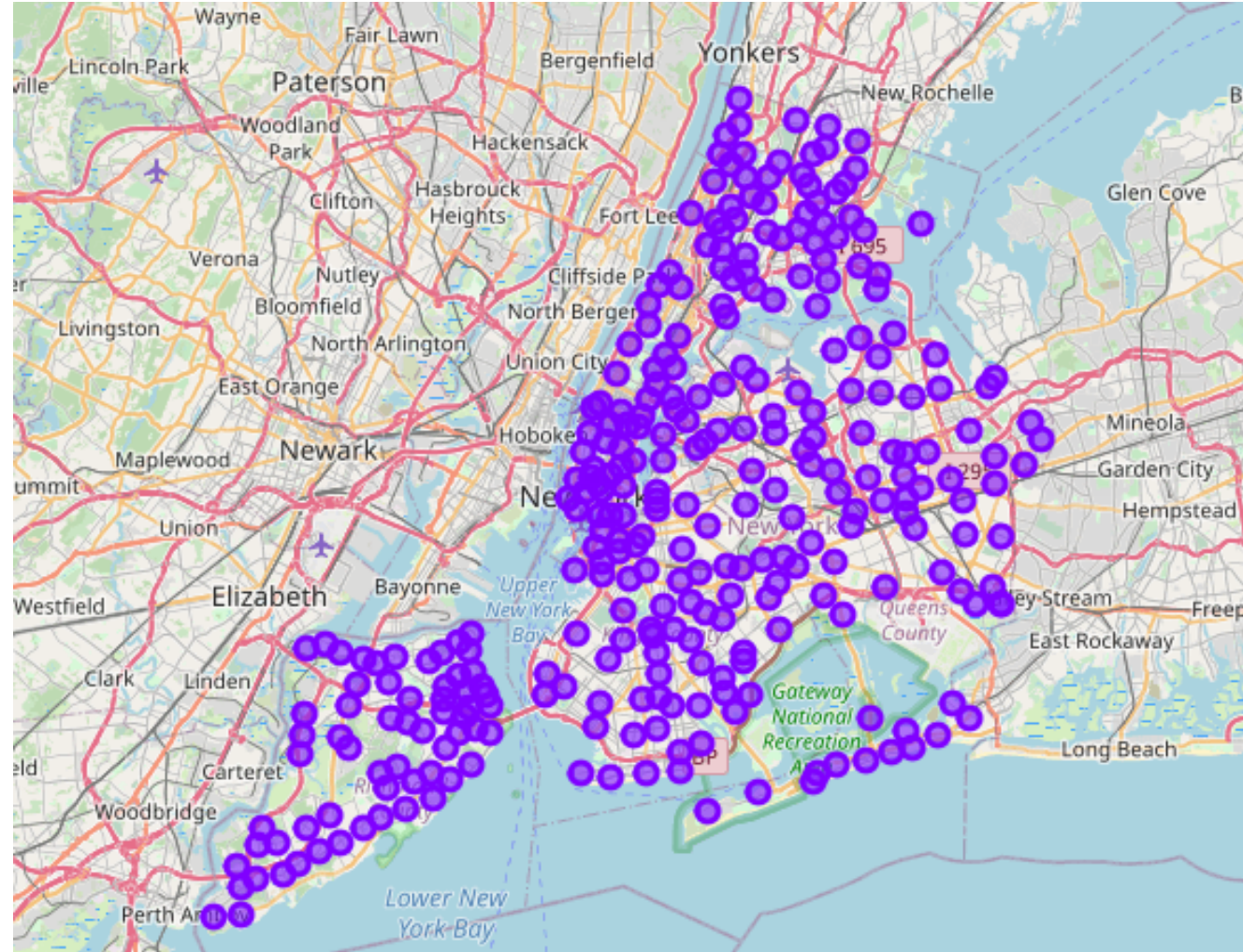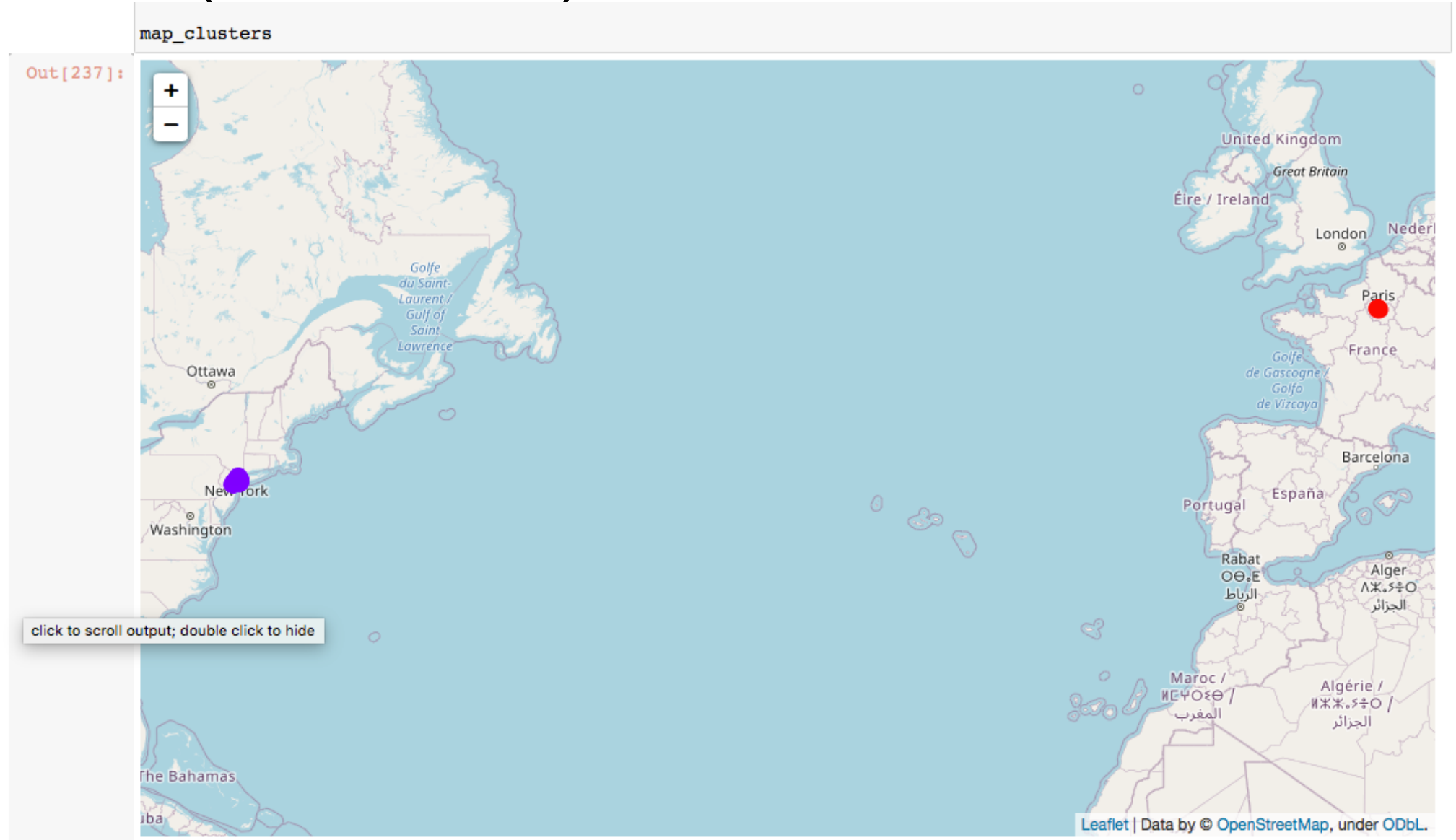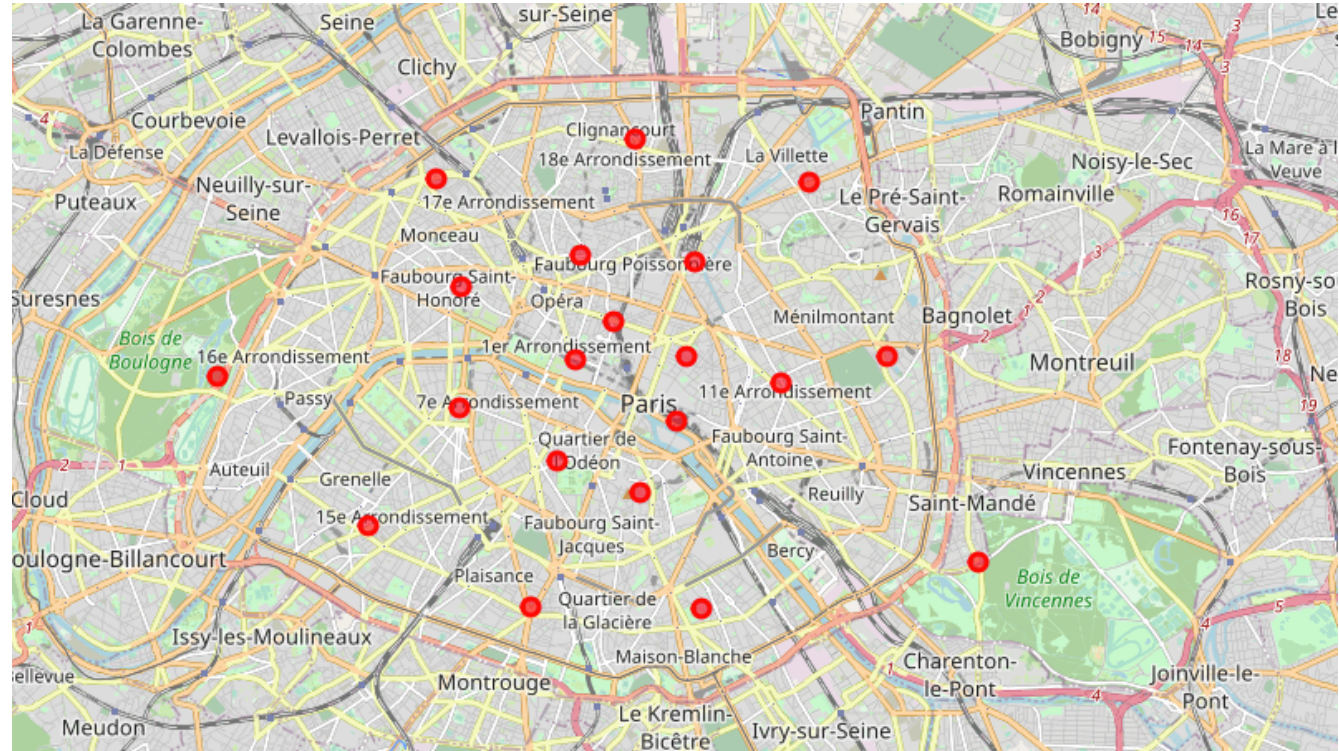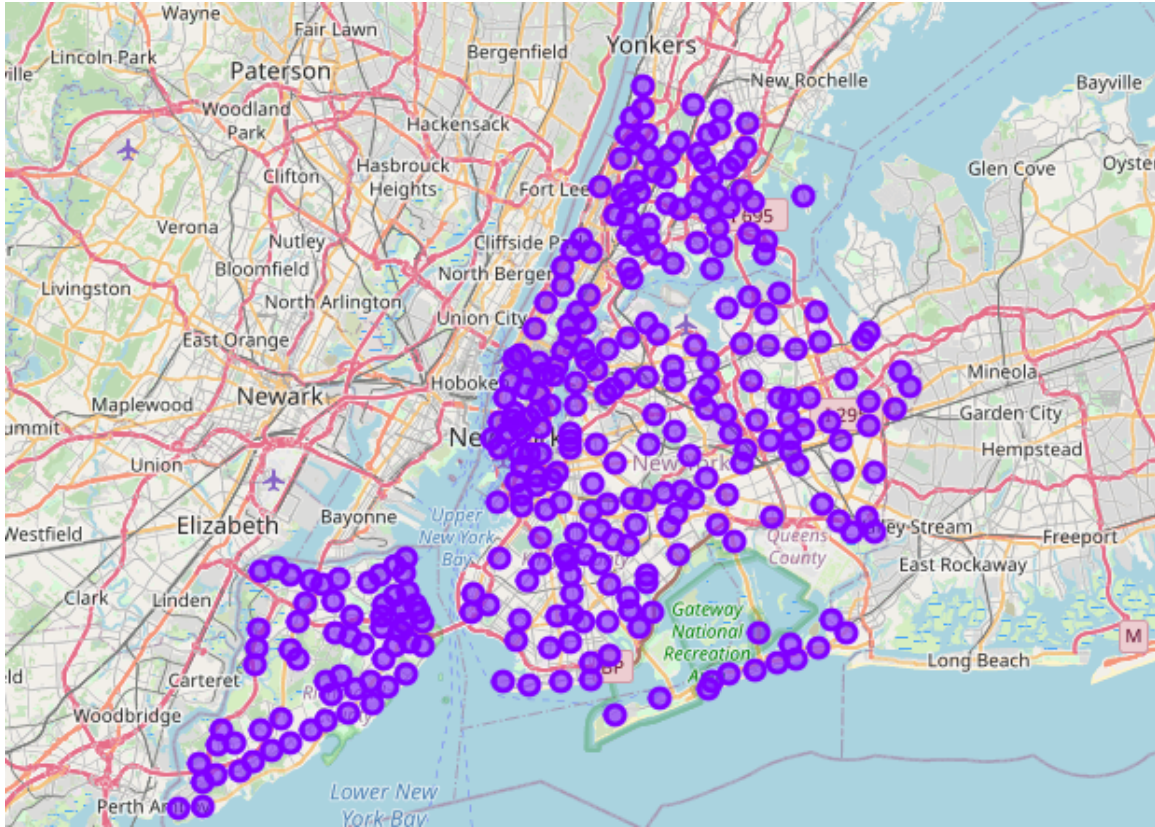# New York city vs Toronto: similar in one segment (one color)

New York city vs Toronto (Cont'd): similar in one segment (enlarged view)
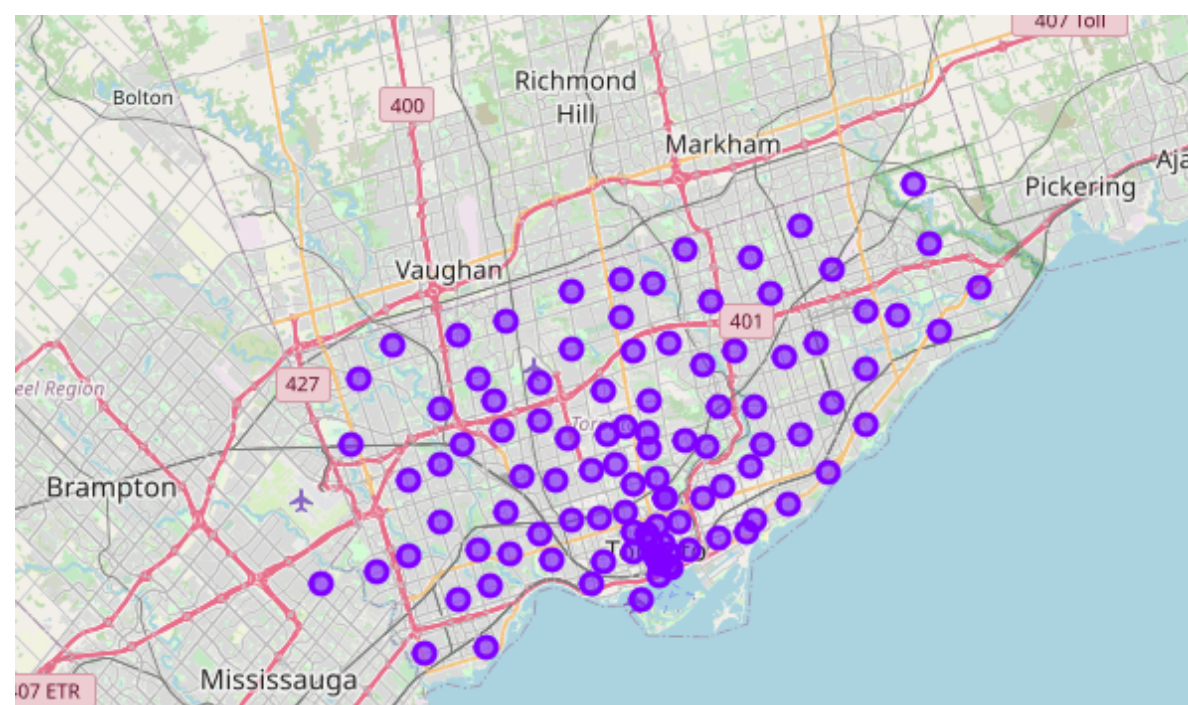
New York city vs Paris: dissimilar into two segment (see colors)

New York city vs Paris (Cont'd): dissimilar into two segment (enlarged view)

Compare difference of category between the cities

Compare the number of catagories and similar catagories between the three cities

```
In [148]: manhattan_grouped_columns_names = list(manhattan_grouped.columns[:])
          print('size of catagories for New York City:',len(manhattan_grouped_columns_names))

          size of catagories for New York City: 22
```

```
In [149]: toronto_grouped_columns_names = list(toronto_grouped.columns[:])
          print('size of catagories for Toronto:', len(toronto_grouped_columns_names))

          size of catagories for Toronto: 22
```

```
In [150]: paris_grouped_columns_names = list(paris_grouped.columns[:])
          print('size of catagories for Paris:', len(paris_grouped_columns_names))

          size of catagories for Paris: 52
```

Compare catagories difference between the cities:

```
In [151]: paris_grouped_columns_names_set = set(paris_grouped_columns_names)
          manhattan_grouped_columns_names_set = set(manhattan_grouped_columns_names)
          toronto_grouped_columns_names_set = set(toronto_grouped_columns_names)
```

```
In [152]: common_catagories_NewYork_Toronto_set = manhattan_grouped_columns_names_set.intersection(toronto_grouped_columns_names_
```

```
In [153]: print('Does New York City has the same catagories as Toronto? Answer is:', common_catagories_NewYork_Toronto_set == man

          Does New York City has the same catagories as Toronto? Answer is: True
```

```
In [154]: common_catagories_three_cities_set = paris_grouped_columns_names_set.intersection(manhattan_grouped_columns_names_set)
          print('How many of similar catagories between New York city or Toronto and Paris? the number is:', len(common_catagorie

          How many of similar catagories between New York city or Toronto and Paris? the number is: 7
```

```
In [155]: common_catagories_three_cities = list(common_catagories_three_cities_set)
          print('the common catagories between the three cities are:', common_catagories_three_cities)

          the common catagories between the three cities are: ['Sandwich Place', 'Pizza Place', 'Deli / Bodega', 'Seafood Resta
          urant', 'Neighborhood', 'Ice Cream Shop', 'Coffee Shop']
```

Conclusion section where you conclude the report

- New York city and Toronto neighborhoods are similar
- New York city and Paris neighborhoods are quite dissimilar
- Paris has more categories to explore