

IBM Applied Data Science Capstone Project

The battle of the neighborhoods

(By Stefan Siegert, May 2020)

IBM Applied Data Science Capstone Project

Business case:

- Kathy is a headhunter and international staff recruiter in Toronto, Canada
- For her candidates she likes to offer a list of their favorite neighborhoods



IBM Applied Data Science Capstone Project

4 different kinds of candidates:

- (1) Hipsters, who prefer cafés and coffee shops in their neighborhood
- (2) Family people like parks and playgrounds
- (3) Sportive candidates need gyms and fitness centers near their home
- (4) Others just want to go shopping

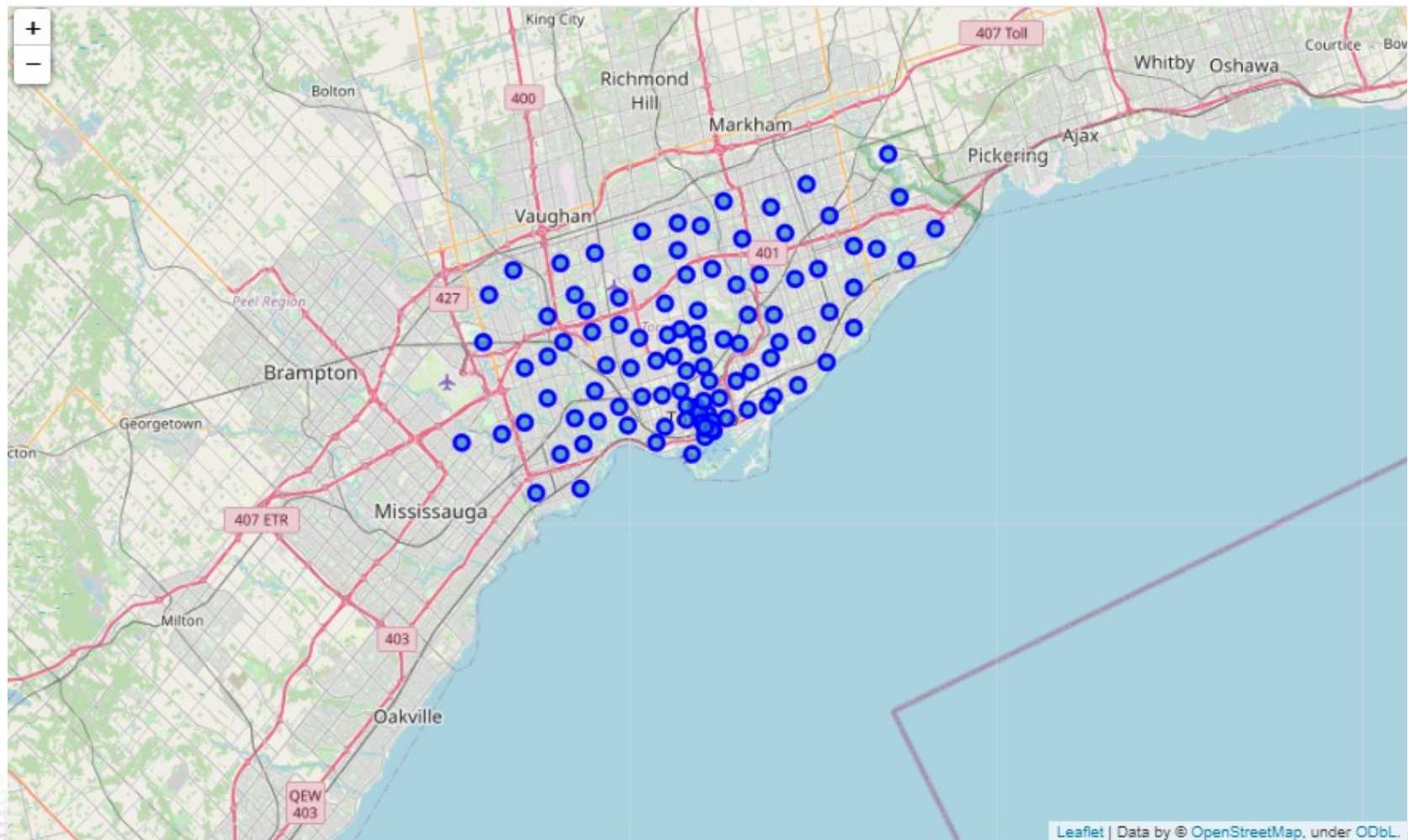
IBM Applied Data Science Capstone Project

Local neighborhood data of Toronto:

- Source 1: Wikipedia list with postal codes
- Source 2: *.csv file with latitude and longitude data of each of Toronto's neighborhoods
- Source 3: Foursquare API to get data of venues of each of Toronto's neighborhoods
- Source 4: Openstreetmap for the maps

IBM Applied Data Science Capstone Project

Map of all neighborhoods in Toronto:



IBM Applied Data Science Capstone Project

Definition of cluster #1 (Hipster neighborhoods):

- A café or a coffee shop ranking in the 3 most common venues around
- There must be sushi places or Japanese restaurants in the top 10 ranking

```
# Check for cluster #01: hipster's happiness
if (row_list_top3.count('Coffee Shop') == 1) & (row_list_top3.count('Café') == 1) & ((row_list.count('Sushi Restaurant') == 1) | (row_list.count('Japanese Restaurant') == 1)):
    df_toronto_areastyles.at[index, 'Cluster Labels'] = '0'
    newcluster = True
```

IBM Applied Data Science Capstone Project

Definition of cluster #2 (family people):

- Parks and playgrounds are somewhere in the top 10 ranking

```
# Check for cluster #02: family friendly areas  
if (row_list.count('Park') == 1) & (row_list.count('Playground') == 1):  
    df_toronto_areastyles.at[index, 'Cluster Labels'] = '1'  
    newcluster = True
```


IBM Applied Data Science Capstone Project

Definition of cluster #3 (sportive candidates):

- Gyms or fitness centers ranking somewhere in the top 6 of most common venues in a neighborhood

```
# Check for cluster #03: workout areas  
if ((row_list_top6.count('Gym') == 1) | (row_list_top6.count('Gym / Fitness Center') == 1)):  
    df_toronto_areastyles.at[index, 'Cluster Labels'] = '2'  
    newcluster = True
```


IBM Applied Data Science Capstone Project

Definition of cluster #4 (shopping people):

- 3 or more most common venues in a neighborhood include one of the the expressions 'Store', 'store', 'Shop', 'shop', 'Mall' or 'Supermarket' in the first 6 top ranking positions.

```
# Check for cluster #04: shopping to the max
a = 0
for b in range(0,10):
    venue = str(row_list_top6[b])
    if ('Store' in venue) | ('Shop' in venue) | ('shop' in venue) | ('Mall' in venue) | ('store' in venue) | ('Supermarket'
in venue):
        if venue != 'Coffee Shop':
            a = a+1
if a>2:
    df_toronto_areastyles.at[index, 'Cluster Labels'] = '3'
    newcluster = True
```

IBM Applied Data Science Capstone Project

Resulting neighborhoods as a list:

- **Hipster candidates** (cluster #1): Garden District, Ryerson, Central Bay Street, Toronto Dominion Centre, Design Exchange, Runnymede and Swansea.
- **Family candidates** (cluster #2): Moore Park, Summerhill East, Milliken, Agincourt North, Steeles East, L'Amo or Rosedale.
- **Workout and gym candidates** (cluster #3) she can recommend Don Mills, Richmond, Adelaide, King, Commerce Court, Victoria Hotel, Davisville North, Canada Post Gateway Processing Centre, First Canadian Place or Underground city.
- Candidates who prefer **shopping areas around their home** (cluster #4): neighborhoods Weston, Downsview, Lawrence Manor, Lawrence Heights or Parkwoods..

IBM Applied Data Science Capstone Project

Resulting neighborhoods as a map:

