

# Data Mining

## Projet 2019-2020

---

### Choix du sujet

Nous avons choisi de travailler sur les courses de taxis dans la ville de New York. L'objectif était de regrouper les courses selon leurs points de départ/d'arrivée pour trouver les lieux les plus souvent desservis par les taxis. Ces lieux seraient alors des lieux remarquables comme des endroits touristiques ou encore des gares et des aéroports. Nous pourrions donc identifier les lieux les plus visités. Cela représente plusieurs intérêts pour quelqu'un qui est nouveau à New York :

- Savoir quels lieux sont probablement intéressants à visiter.
- Savoir d'où il est probablement facile de trouver un taxi.

### Transformations

Nous avons dans un premier temps nettoyé les données, c'est-à-dire supprimer les enregistrements erronés ainsi que ceux ne nous intéressant pas (coordonnées invalides, nombre de passager égal à zéro, ...). Puis nous nous sommes focalisés sur trois transformations autour des coordonnées géographiques:

- Réaliser un clustering des points de départs des courses
- Réaliser un clustering des points de dépôt
- Associer une zone de départ à une zone d'arrivée grâce à un biclustering départ/arrivée

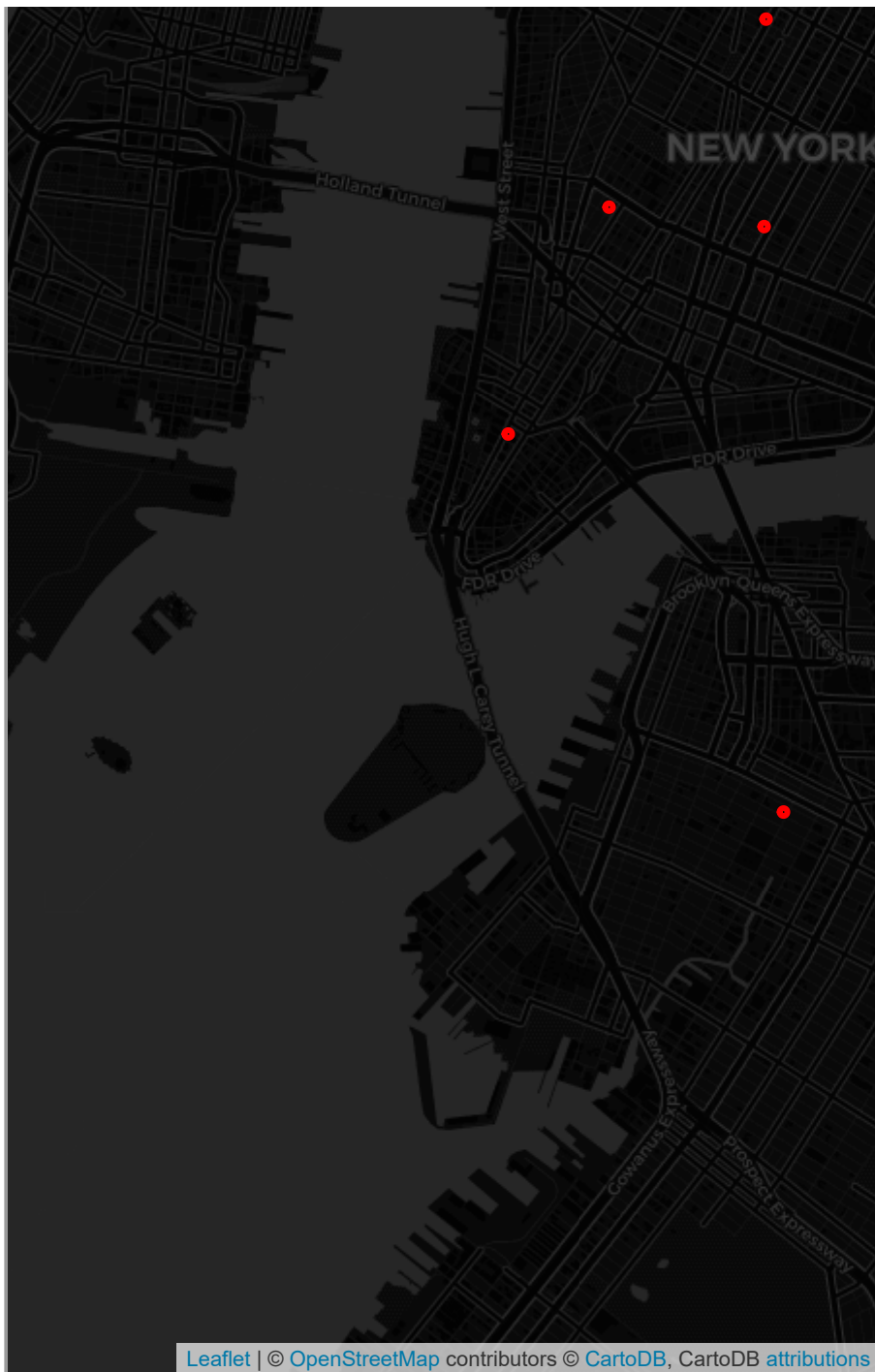
### Résultats

#### Clustering points de départ



665 Lexington Avenue, New York, New York 10022, United States  
Passenger Pick-up 2, Terminal 5, New York, New York 11430, United States  
Kimono House, 131 Thompson St, New York, New York 10012, United States  
Grand Central Parkway, Queens, New York 11371, United States  
Dolce Nail, 1595 3rd Ave, New York, New York 10128, United States  
267 Roosevelt Drive, Seymour, Connecticut 06483, United States  
35 Longley Road, Nazareth, Pennsylvania 18064, United States  
55 East 21st Street, New York, New York 10010, United States  
TDF (Theatre Development Fund), 520 8th Ave, New York, New York 10018, United States  
170 West 76th Street, New York, New York 10023, United States  
197 Windsor Way, Hillside, New Jersey 07205, United States  
383 Ellisdale Road, Chesterfield, New Jersey 08515, United States  
2237 62nd Street, Brooklyn, New York 11204, United States  
468 Rodney Street, Brooklyn, New York 11211, United States  
333 West 16th Street, New York, New York 10011, United States  
247 West 138th Street, New York, New York 10030, United States  
128 East 4th Street, New York, New York 10003, United States  
74 Overland Avenue, Amityville, New York 11701, United States  
146 Fulton Street, New York, New York 10007, United States  
ZogSports - Edward A. Reynolds HS (M506), 140 West 102nd Street btw Amsterdam & Columbus, New York, New York 10025, United States  
35-01 38th Street, Queens, New York 11101, United States  
213 East 73rd Street, New York, New York 10021, United States  
38 Beaver Run Road, Hamburg, New Jersey 07419, United States  
193 Dean Street, Brooklyn, New York 11217, United States  
254 West 54th Street, New York, New York 10019, United States  
21 Deerhill Drive, Ho Ho Kus, New Jersey 07423, United States





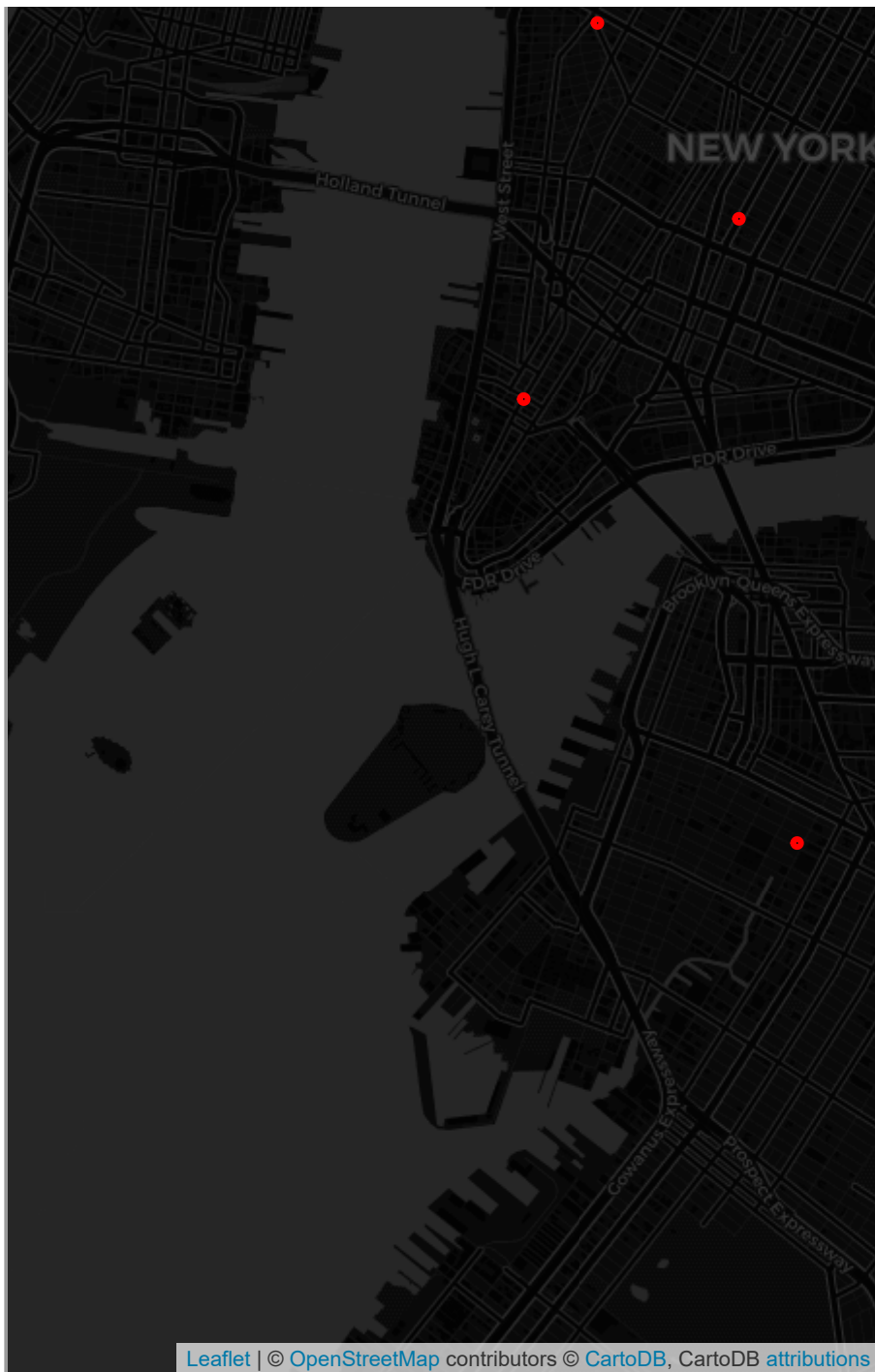
Leaflet | © OpenStreetMap contributors © CartoDB, CartoDB attributions

Clustering points d'arrivée



239 Wyckoff Street, Brooklyn, New York 11217, United States  
Martino Hall, 45 Columbus Ave, New York, New York 10023, United States  
The Office Center at Mitchel Field, 377 Oak Street, Garden City, New York 11530, United States  
120 Greenwich Avenue, New York, New York 10011, United States  
19 Groton Street, Queens, New York 11375, United States  
Raquets And Tennis, 370 Park Ave, New York, New York 10152, United States  
11 High Ridge Road, Seymour, Connecticut 06483, United States  
149 West 121st Street, New York, New York 10027, United States  
Park South Hotel, 124 E 28th St, New York, New York 10016, United States  
410 Burlington-Bordentown Road, Bordentown, New Jersey 08505, United States  
Agua De La Vida Day Spa, 561 Metropolitan Ave, New York, New York 11211, United States  
215 East 81st Street, New York, New York 10075, United States  
Magnolia Way, Bronx, New York 10467, United States  
2431 Southmoore Drive, Bath, Pennsylvania 18014, United States  
6602 Wallaston Court, Brooklyn, New York 11204, United States  
Enterprise Rent-A-Car, 25 Newark Airport, Bldg 25, Newark, New Jersey 07114, United States  
Global Entry Kiosk, Terminal 7 JFK Airport, New York, New York 11430, United States  
55 2nd Avenue, New York, New York 10003, United States  
101-12 23rd Avenue, Queens, New York 11369, United States  
1028 Lincoln Place, Brooklyn, New York 11213, United States  
Trans-Manhattan Expressway, New York, New York 10033, United States  
10 Dartmouth Drive, Bay Shore, New York 11706, United States  
5 Brink Road, Wantage, New Jersey 07461, United States  
120 Church Street, New York, New York 10007, United States  
61-00 Francis Lewis Boulevard, Queens, New York 11364, United States  
175 West 90th Street, New York, New York 10024, United States  
Rubin Singer Studio, New York, New York 10018, United States  
15 Golden Pond Road, West Harrison, New York 10604, United States  
32-56 41st Street, Queens, New York 11103, United States





## Biclustering

T.Bourg & T.Buttez - M2IA - UCBL