

Rapport Technique

Table des matières

| | |
|----------------------------|----------|
| Contexte | 2 |
| Intégration | 2 |
| Stockage | 2 |
| Visualisation | 2 |
| Kibana | 3 |
| Mapbox | 5 |
| Chart.js | 6 |

Nicolas BONDOUX
Alexandre BAILLY
Julien MORATALLA
Franck DIOMANDE

Contexte

Dans le cadre d'un projet scolaire lié au Big Data, nous avons décidé d'analyser l'utilisation des trottinettes électriques en île de France. Le projet est hébergé sur Github, à l'adresse suivante : <https://github.com/franckdiomande/trottinette>

Le choix des outils sont orientés vers des outils de Big Data au vu de la quantité importante de données qui sera manipulé.

Intégration

Pour l'intégration des données, nous avons recoupés deux sources importantes de données :

- L'API Twitter (<https://developer.twitter.com>)
- L'API Multicycles (<https://developer.multicycles.org/api>)

Pour la récupération des données depuis twitter, nous avons utilisé notre script *nodejs* qui se trouve dans `scripts/twitter_download.js`. Ce script récupère tous les tweets en liaison avec le mot « trottinette » et les marques de trottinette. La requête se fait sur les sept derniers jours. Le résultat de la requête est stocké dans le dossier `data/twitter`.

La deuxième source de données vient Multicycles. Grâce cette API, nous avons pu obtenir la localisation en temps réel, les fournisseurs et les zones géographiques des différentes marques de trottinette. Ces données sont stockées dans le dossier `data/multicycles`.

Stockage

Pour le stockage des données que nous avons récupéré, nous avons utilisé Elasticsearch. Nous avons utilisé le script `scripts/import_to_elasticsearch.js` pour créer nos indexes et importer les données. Ce script crée cinq indexes :

- *twitter* pour les données qui proviennent de Twitter
- *scoot_location* pour les données de localisation des trottinettes et des marques.
- *scoot_provider* pour les données des fournisseurs de trottinette.
- *scoot_zone* pour les données de géolocalisation des trottinettes.

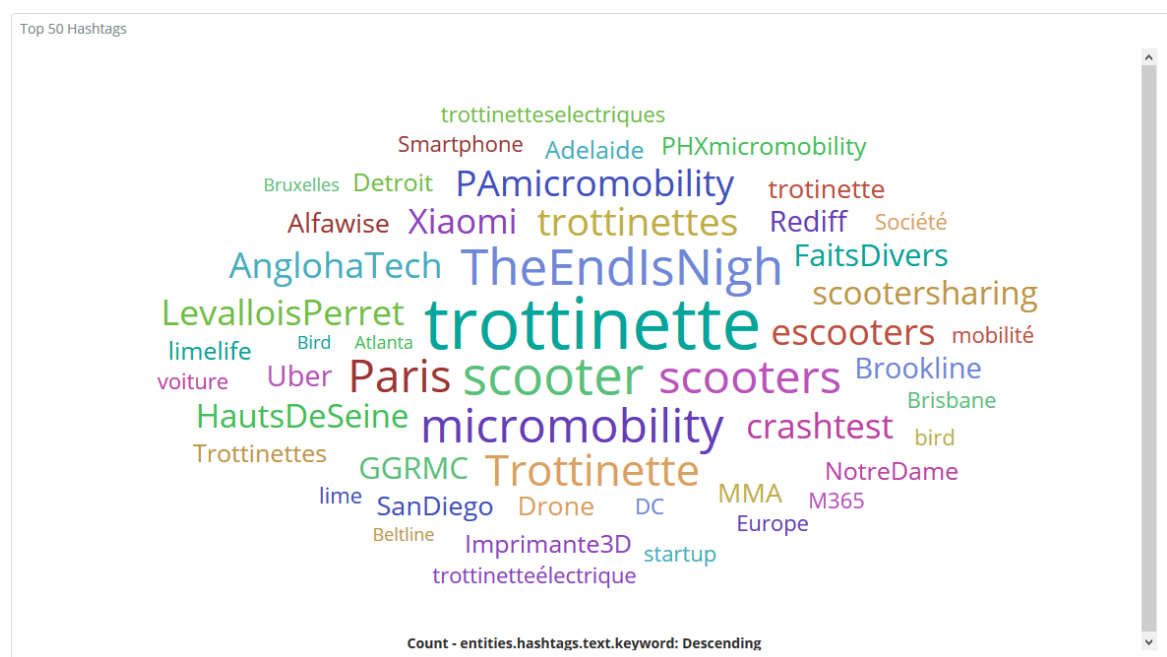
Pour plus de scalabilité, nous avons configuré notre cluster et défini un backup. Nous avons défini trois nœuds primaires dont deux replicas.

Visualisation

Big Data

Pour la visualisation des données, nous avons utilisés plusieurs outils et bibliothèques :

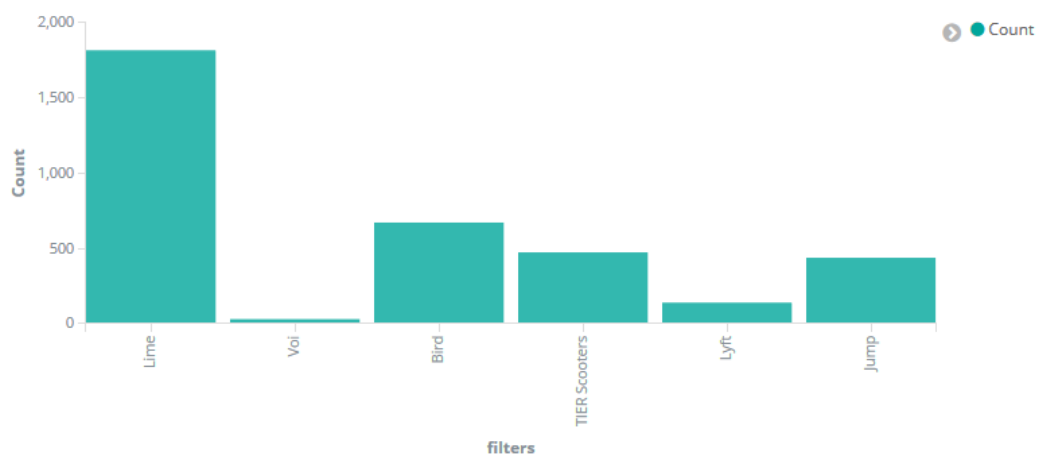
Kibana



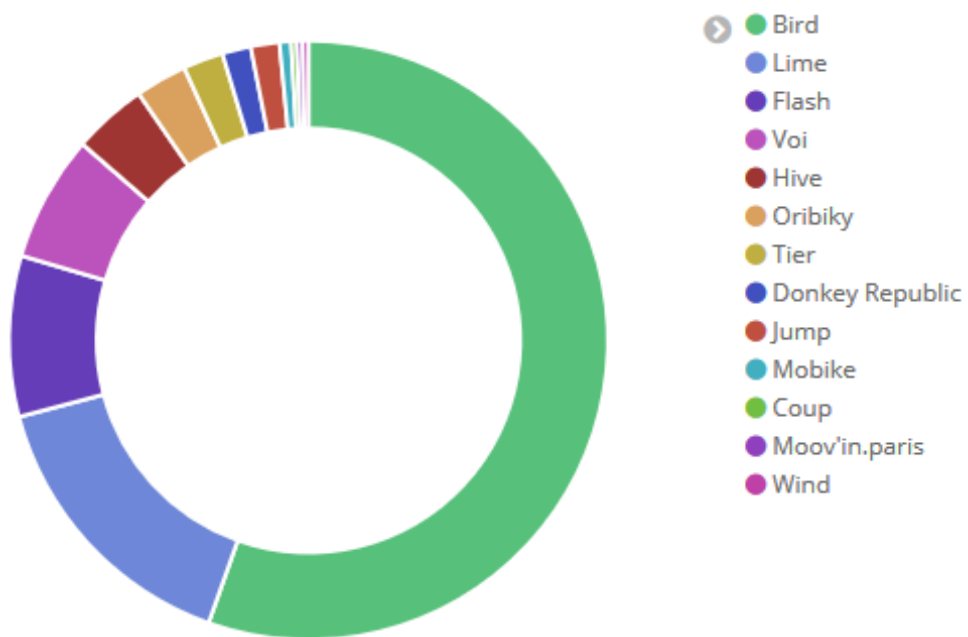
Nicolas BONDOUX
Alexandre BAILLY
Julien MORATALLA
Franck DIOMANDE

Big Data

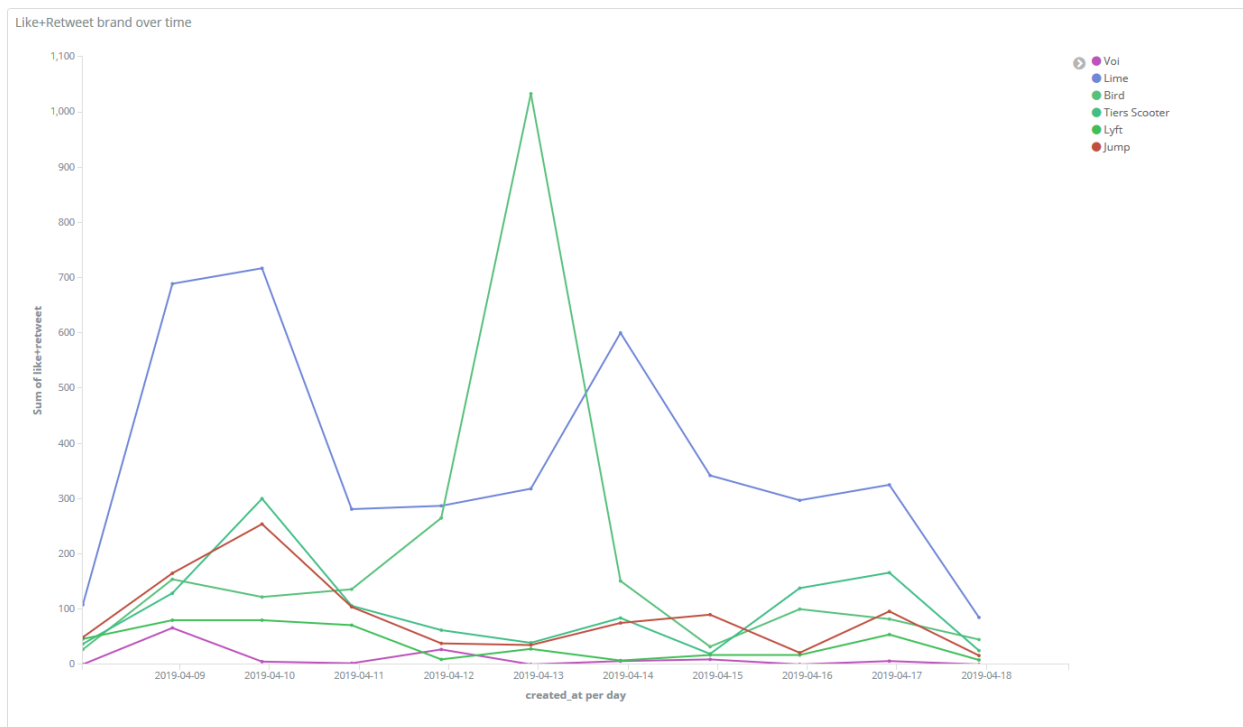
Number Of tweets that mentions a brand name



Scooter distribution

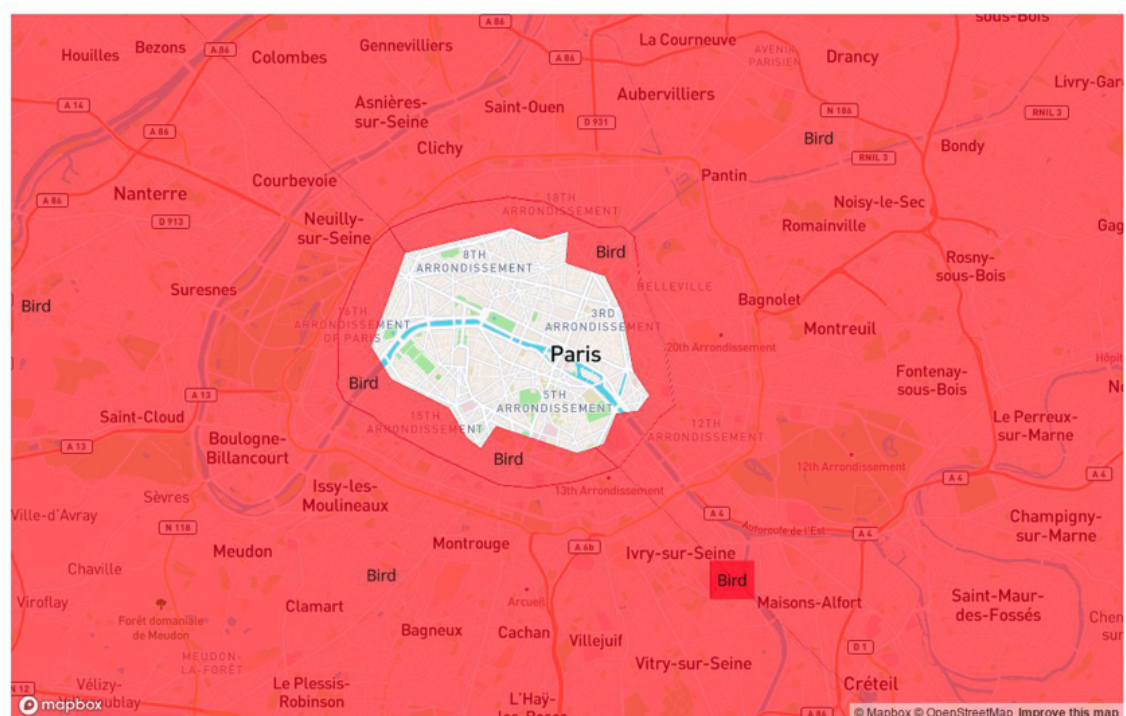


Big Data



Mapbox

Répartition de Bird



Nicolas BONDOUX
Alexandre BAILLY
Julien MORATALLA
Franck DIOMANDE

Répartition de Lime

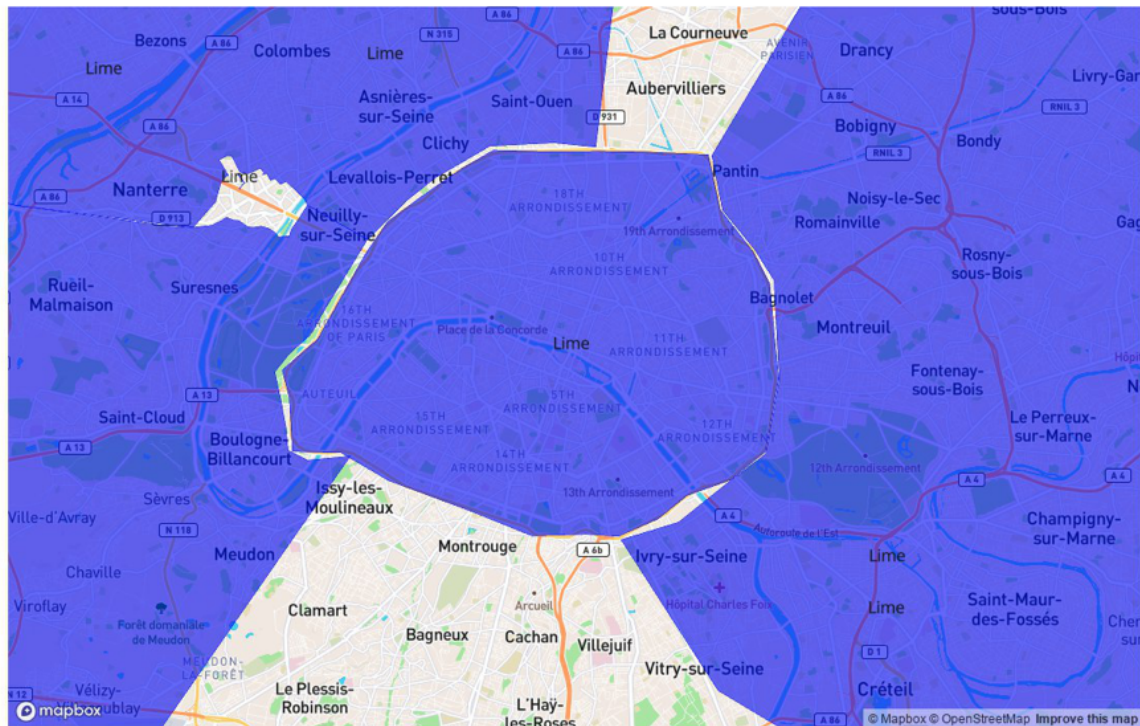
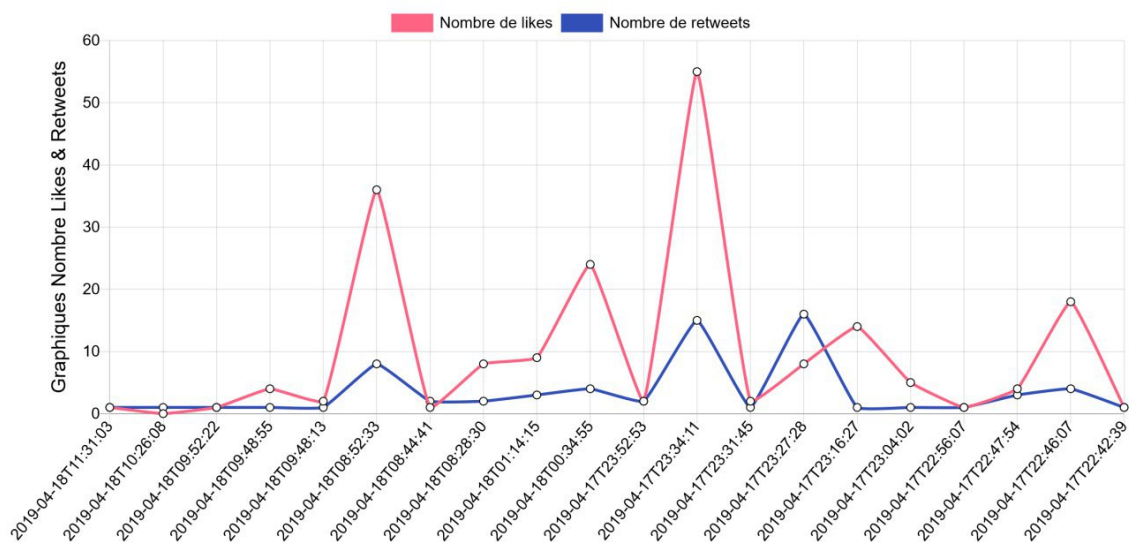


Chart.js



Nicolas BONDOUX
Alexandre BAILLY
Julien MORATALLA
Franck DIOMANDE