

# **Projet noSQL**

*Juliette Choudjaye*

Le but de ce projet était de déterminer, grâce à des requêtes effectuées sur les données de la ville de New York, de déterminer le meilleur endroit où vivre dans New York.

## 1) Le meilleur endroit

Pour moi, plusieurs critères sont importants, je suis en effet sensible à l'environnement, mais aussi aux restaurants, bibliothèques et la facilité d'accès aux produits locaux.

## 2) La base de données

J'ai choisi 4 datasets pour m'aider à trouver le meilleur endroit :

- « Public recycling bins » qui recense les poubelles de recyclage à New York
- « New York Public Library (NYPL) Branch Services from 7-2010 to 6-2011 » qui, même s'il est ancien permet d'avoir une idée de la présence de bibliothèques
- « New York City Farmers Market » qui recense les magasins vendant des produits locaux
- DOHMH New York City Restaurant Inspection Results qui présente les résultats des inspections sanitaires des restaurants

J'ai choisi ces quatre jeux de données car ils permettent de prendre en compte les principaux aspects de mon « meilleur endroit »

## 3) La méthode implémentée

Tout d'abord, j'ai réalisé l'installation de neo4j sur ma machine (la marche à suivre est dans le script neo4j).

Ensuite, j'ai téléchargé les données (commandes dans le script telechargement\_data).

Après cela, j'ai enlevé les données manquantes, puis les colonnes sur lesquelles je ne comptais pas réaliser des requêtes, afin que l'import dans neo4j soit le plus aisé possible (le script r exécuté avec le script preparation\_data).

Pour ce qui est des requêtes, j'ai commencé par importer les données.

Ensuite, j'ai sélectionné deux « boroughs » sur les 5 de New York en choisissant les deux qui contenaient le plus de poubelles de recyclage (donc respectueux de l'environnement), j'ai donc retenu Manhattan et le Queens.

Après cela, j'ai départagé les deux grâce à la table sur les bibliothèques : en effet, aucune des bibliothèques de la table n'est dans le Queens, j'ai donc choisi Manhattan.

Enfin, je me suis intéressée au zipcode : j'ai sélectionné les 5 zipcodes dans lesquels il y avait le plus de farmers market, j'ai obtenu : 10029, 10031, 10004 , 10019, 10027, 10009, 10003, 10026, 10011 (le premier a 3 farmers market et les suivants en ont 2).

J'ai finalement choisi le zipcode 10004 car c'est dans celui-ci qu'il y a eu le moins d'inspections de restaurants chinois ou français (donc il y a moins de problèmes d'hygiène).

#### 4) La réponse

