

Mini Projet NoSQL

Seydou KEITA

Février 2018

1 Introduction

Le but du projet est de déterminer le meilleur quartier/arrondissement/rue où s'installer à New York, ce à l'aide d'un **SGBD NoSQL** et le portail [Open-Data de New York](#).

Pour ce faire, je suis parti de la table de données **IPIS (Integrated Property Information System)** contenant des informations sur l'ensemble des logements sociaux de la ville de New York. Comme SGBD, mon choix s'est porté sur **MongoDB**.

2 Pourquoi MongoDB ?

Le choix de MongoDB s'explique par plusieurs facteurs. Tout d'abord, *le format de stockage*, MongoDB est une base de données *orientée document*. Ce type de stockage est très intéressant de nos jours car il offre une bonne gestion des données non structurées et une grande flexibilité dans la programmation.

Par ailleurs, la force de MongoDB dans le cadre de ce projet réside dans le fait qu'il dispose des *fonctions pré-implémentées permettant d'effectuer opérations sur des coordonnées géo-locales* telle que le calcul de distance où encore la recherche des documents(cordonnées) les plus proches d'un document donné dans un rayon bien défini.

3 Description de l'algorithme implémenté

Pour répondre à la question, j'ai téléchargé 7 jeux de données du portail Open Data :

1. [IPIS \(Integrated Property Information System\)](#) : la liste des logements sociaux de la ville de New York ;
2. [City Subway Stations](#) : la liste des stations métro ;

3. [City Bike Live Station Feed](#) : le flux de données sur les stations de vélos libre service indiquant le nombre de vélos disponibles par station à un instant donné. Il va nous permettre de connaître le nombre de stations non loin d'un logement donné ;
4. [Legally Operating Businesses](#) : jeux de données regroupant les commerçants et les entreprises disposant d'une licence. Il va nous permettre d'avoir une idée sur le nombre d'entreprises à proximité du logement.
5. [NYC Wi-Fi Hotspot Locations](#) : listes des emplacements de points d'accès Wi-Fi publics à travers la ville ;
6. [New York City Museums](#) : liste des musées de la ville avec leur coordonnées géo-locales ;
7. [Restaurants dataset](#) : la liste des restaurants de la ville de New York. Cependant, ce jeux de données provient du site officiel de MongoDB.

Le logement idéal, à mon sens, est celui qui à proximité des différentes commodités (transports, entreprises, restaurants, points d'accès Wi-Fi etc.).

Ainsi, à partir des différents jeux de données, j'ai réalisé une association entre les logements et les commodités qui les entourent. Elle consiste à associer à chaque logement :

- le nombre d'entreprises se trouvant dans un rayon de 10 km ;
- le nombre de stations métro se trouvant dans un rayon de 500 m ;
- le nombre de stations de vélos se trouvant dans un rayon d'1 km ;
- le nombre de restaurants se trouvant dans rayon de 10 km ;
- le nombre de points d'accès Wi-Fi publics se trouvant dans rayon d'1 km

Finalement, nous trouvons seulement **11 logements** qui sont répondent à tous les critères. Voici leurs principales caractéristiques :

	Arrondissement\ Quartier	Adresse	Longitude	Latitude	Entreprises	Subway Station	Restaurants	Museums	Nb Wifi	Bike Stations	
1	Manhattan	Midtown-Midtown South	336-50 5 AVENUE	-73.98512	40.747683	53.0	4.0	65.0	27.0	1.0	1.0
2	Manhattan	Gramercy	219-53 3 AVENUE	-73.985058	40.736194	31.0	5.0	20.0	22.0	1.0	1.0
3	Manhattan	Gramercy	207 3 AVENUE	-73.985256	40.735919	30.0	5.0	23.0	22.0	1.0	1.0
4	Brooklyn	Greenpoint	107 NORMAN AVENUE	-73.950756	40.72596	24.0	1.0	15.0	0.0	1.0	1.0
5	Manhattan	Midtown-Midtown South	282 7 AVENUE	-73.994168	40.746147	22.0	5.0	30.0	21.0	1.0	1.0
6	Brooklyn	Park Slope-Gowanus	25 FOURTH AVENUE	-73.978814	40.683666	20.0	6.0	8.0	5.0	1.0	1.0
7	Manhattan	East Harlem South	1659 MADISON AVENUE	-73.9474	40.796474	12.0	2.0	4.0	9.0	1.0	1.0
8	Manhattan	East Harlem South	53 EAST 110 STREET	-73.947328	40.796032	12.0	1.0	5.0	10.0	1.0	1.0
9	Manhattan	Central Harlem North-Polo Grounds	2405 8 AVENUE	-73.949392	40.812619	8.0	1.0	2.0	4.0	1.0	1.0
10	Manhattan	SoHo-TriBeCa-Civic Center-Little Italy	125 WORTH STREET	-74.002673	40.715428	6.0	7.0	4.0	24.0	1.0	1.0
11	Manhattan	SoHo-TriBeCa-Civic Center-Little Italy	68 VARICK STREET	-74.006267	40.722751	4.0	4.0	8.0	21.0	1.0	1.0

FIGURE 1 – Top 11 des logements sociaux de New York

3.1 Environnement de travail

- Linux Ubuntu : 17.04
- MongoDB Version 3.6.2
- Robot Mongo 3T - 1.1