

Projet NoSQL

Rapport final



Préambule

1.1 Choix du quartier

Ma première préoccupation concernant le choix du quartier va être de choisir un endroit qui soit le moins pollué possible. Vivre et travailler dans une grande ville comme New York peut être quelque peu stressant, je vais donc chercher à me rapprocher d'une zone avec beaucoup d'espaces verts et de parcs, sans pour autant m'isoler du centre ville. N'ayant pas de voiture, je rechercherai donc un quartier proche d'une station de métro de manière à faciliter mes déplacements. Enfin, étant passionné de natation s'il me reste plusieurs choix d'emplacements, j'opterai pour celui le plus proche d'une piscine.

1.2 Choix des données et du système de stockage

1.2.1 Les données

Pour réaliser ce projet, je vais donc utiliser quatre bases de données disponibles sur le site <https://nycopendata.socrata.com/>.

- [Air quality](#) , pour rendre compte de la qualité de l'air à New York. Cette base comporte 270 observations, cependant je ne vais prendre en compte que les données concernant les particules fines qui sont très néfastes pour la santé.
- [Open spaces](#). Cette base comporte 12 491 observations pour décrire les parcs, espaces verts, terrains de sport de New York.
- [Subway stations](#) est la base contenant la positions des 473 stations de métro de New York.
- [Indoor swimming pools](#) est la base contenant l'emplacement des 13 piscines intérieurs de New York.

1.2.2 La base de données

J'ai choisi d'utiliser une base de données MongoDB pour sa rapidité, sa simplicité d'utilisation ainsi que sa structure de type documents. Cela me permettra d'insérer facilement mes données, de types mixtes, formatées en Json pour la plupart. Enfin MongoDB permet de créer des bases pouvant facilement évoluer ce qui me permettra de supprimer et de rajouter des champs dans les documents.

1.2.3 Importation

Pour importer les données, il faut lancer mongoDb en local puis exécuter le script python *noSql.py* disponible dans le dépôt git. Suivant votre environnement, il vous faudra peut-être installer *requests* et *pymongo* dans l'environnement python avec les commandes : *pip install requests* et *pip install pymongo*.

Les données sont insérées en base directement depuis les API fournis par NYC Open Data. Les fichiers ne sont pas téléchargés sur le pc. Certaines données sont insérées directement dans la base comme celle concernant la qualité de l'air et d'autres comme parc et subway sont insérées document par document dans la base en supprimant certains champs inutiles et en organisant les données de localisation sous la forme d'un champ latitude et d'un autre longitude.

Les données concernant les piscines sont traitées différemment puisque le fichier est récupérer depuis NYC Open Data sous la forme d'un fichier xml. On utilise alors *minidom* pour parser ce fichier et on parcourt les noeuds tout en insérant dans la base les données nécessaires (nom de la piscine, latitude et longitude).

Après l'exécution du script, nos données sont disponibles dans la base de données mydb dans mongo sous la forme de quatre collections : air, park, sub et pool.

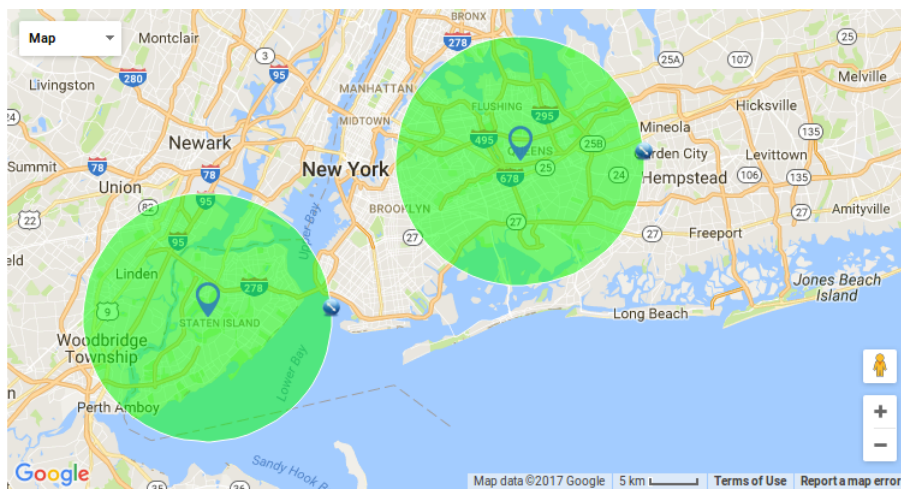
Choix du lieu de résidence

Tous les résultats énoncés dans la suite de ce document s'appuient sur des requêtes Mongo disponibles et commentées dans le fichier *requetesMongo* de ce dépôt git.

2.1 Étude de la qualité de l'air

La première idée est de regarder la qualité de l'air dans les différents quartiers de New York. La base de données traitant de la qualité de l'air à New York mesure différents composants de l'air et notamment le taux de particules fines dans les différents quartiers.

Les deux quartiers possédant le taux le plus bas de particules fines sont Staten Island et le Queens. Nous venons donc de réduire notre zone de recherche à ces deux quartiers.



En utilisant quelques outils externes on peut maintenant réduire la zone de recherche à deux cercles de 12km centrés sur Staten Island et le Queens. Tous les calculs de distance à partir de coordonnées de type latitude, longitude sont effectués selon la formule définie dans le document suivant. [1].

2.2 Espaces verts et parcs

Ma deuxième préoccupation avant de choisir mon prochain lieu d'habitation est la présence de parcs et d'espaces verts dans les environs. Après avoir récupéré la base de données des parcs de New York et l'avoir filtrée pour ne garder que les parcs tel qu'on le conçoit (suppression de terrains de sports, des gymnases et simples rues vertes...), on s'aperçoit que le Queens possède 807 parcs et Staten Island 337.

En terme de superficie d'espaces verts le Queens est aussi plus intéressant. Avec plus de 224 hectares d'espaces verts, le Queens devient donc mon futur quartier de prédilection pour mon installation à venir.

2.3 Transports et loisirs

Maintenant que le cadre est fixé, il s'agit de rechercher un emplacement à une distance convenable d'une station de métro afin de se déplacer facilement. Pour moi une distance convenable correspond à une distance assez faible pour être rapidement à la station mais pas trop petite non plus pour réduire au maximum les nuisances occasionnées par les bruits du métro ainsi que de la "foule" présente aux abords de la station aux heures de pointes.

Passionné par la natation, mon dernier critère de recherche sera la présence d'une piscine couverte pour pouvoir aller nager par tous temps, dans le environ.

Une cinquantaine de stations de métro se trouvent dans le Queens. En revanche il n'y a que deux piscines couvertes. Je vais donc chercher la station de métro la plus proche pour chaque piscine et choisir mon emplacement à mi-chemin entre la piscine et la station de métro.

- La station la plus proche de Roy Wilkins Recreation Center Pool est celle de Jamaica - 179th St située à environ 3km.
- La station la plus proche de Flushing Meadows Corona Park Pool & Rink est celle de Mets - Willets Point située à moins de 800m

Avec ces résultats je décide d'aller m'installer dans le Queens, à mi-chemin entre le Flushing Meadows Corona Park Pool & Rink et la station de métro de Mets - Willets à l'angle de Roosevelt Street et de la 114^{eme}.

Bibliographie

[1] http://geodesie.ign.fr/contenu/fichiers/Distance_longitude_latitude.pdf