

NoSQL

Maxence Ecole

Choix des logements

Je me suis placé dans le cas où je disposais des coordonnées de plusieurs logements disponible à l'achat ou la location. Pour chaque disponibilité on calcule un score qui dépend de paramètres que l'on associe à nos préférences. Une fois ce score évalué il suffit de trier les différentes habitations selon cette dernière variable.

Choix des données

A partir du site <http://opendata.cityofnewyork.us/> j'ai choisi 3 bases de données :

<https://data.cityofnewyork.us/Health/NYC-Health-and-Hospitals-Corporation-Facilities/ymhw-9cz9> : Etant très malade, je souhaite avoir un centre hospitalier à coter de chez moi. On peut affiner la recherche en précisant un type d'établissement ou bien sélectionner quelques centres hospitaliers.

<https://data.cityofnewyork.us/Recreation/New-York-City-Art-Galleries/tgyc-r5jh> : Je suis passionné d'art et je veux donc le plus de galeries possible dans mon voisinage.

<https://data.cityofnewyork.us/Transportation/Subway-Entrances/drex-xx56> : N'ayant pas de voiture il me faut une station de métro à côté de chez moi. Je peux affiner ma recherche en sélectionnant la/les lignes de métro.

Je vais utiliser une base de données MongoDB pour stocker ces données. J'ai choisi MongoDB pour sa rapidité, sa simplicité à rajouter et enlever des champs à nos bases. De plus Je peux facilement importer mes données avec un script python.

Importation des données

Après avoir lancé mongo en local on peut exécuter le script python noSQL.py pour importer les données. On garde seulement les informations qui nous intéressent. Ce script génère également une base home contenant les coordonnées des 12 localisations que l'on va comparer.

Calcul du score

Le score aurait pu être déterminé depuis python, mais j'ai choisi de déterminer le score de chaque logement depuis mongo. Pour chaque logement on cherche le centre hospitalier le plus proche. Bien sûr si on le souhaite on peut spécifier un type ou une liste d'établissement hospitalier. Pour les établissements hospitaliers, on ne retient que la distance du plus proche.

Ensuite on cherche une forte densité de galerie d'art. J'ai créé avec succès une table «inter» qui contient pour chaque habitation un tableau des distances de l'habitation à chaque galerie d'art. Je n'ai malheureusement pas réussi à compter le nombre de fois où cette distance est inférieure à un 1(km) par exemple. A défaut d'avoir le nombre de galerie dans un cercle de 1km autour du logement je me contente de la galerie la plus proche.

Enfin à New York city le métro est le mode de transport le plus commode. En fonction du lieu de son travail, on peut chercher la station de métro menant à ce dernier la plus proche de chacun des logements.

Une fois ces trois mesures obtenues on les pondère comme on le souhaite par exemple 0.2 pour le centre hospitalier, 0.1 pour les galeries d'art et 0.7 pour la station de métro. On détermine le score et on trouve le logement qui nous convient le plus.