

Projet NoSQL : « Best place to live »

Je vais prochainement déménager à New York City pour mon premier emploi et je cherche à utiliser les données disponibles en open data sur le site de la ville pour trouver un bon endroit où m'installer.

Qu'est-ce qu'un bon endroit où m'installer ?

- 1) Je veux absolument être proche de la Data Science Academy, où j'ai mes contacts.
- 2) Je veux habiter à proximité d'une station de métro.
- 3) Je souhaite avoir un maximum de salles de spectacles à proximité de chez moi.
- 4) Je suis très sensible à la qualité de l'eau, je souhaite donc qu'elle soit bonne dans mon quartier.

Choix de la base de données utilisée :

D'une part j'ai choisi d'utiliser des bases NoSQL car les données qui sont disponibles ne sont pas normalisées d'une manière qui permette un requêtage SQL simple. Comme base NoSQL, j'ai choisi MongoDB car elle permet le requêtage rapide de listes d'objets géolocalisés. En pratique Cassandra ou Redis auraient été des alternatives possibles. En vérité, MongoDB s'est avéré ne pas correspondre vraiment à ma démarche car il n'y est pas possible d'utiliser les résultats des requêtes d'une collection à l'autre.

Descriptif de la méthode/algorithme implémenté pour trouver où vivre :

Comme je veux absolument être proche de la Data Science Academy, j'ai d'abord cherché la liste des stations de métro les plus proches (moins d'un kilomètre). Ces stations sont les lieux candidats à proximité desquels je veux m'installer. A partir de ces points candidats, je cherche plusieurs choses pour les classer :

- nombre de salles de spectacle à moins d'un kilomètre ;
- nombre de plaintes émises à proximité sur la qualité de l'eau,

The best place to live :

Je souhaite habiter à proximité de la station de métro 49th Street ! Mon futur chez moi :



Détail de la démarche :

Point de référence : Data Science Academy (-73.992383, 40.753770).

Stations candidates :

subwayLocations 0.003 sec.			
_id		NAME	location
1	<input type="checkbox"/> ObjectId("58ade3f1aed16427117596e2")	34th St - Penn Station	{ 2 fields }
2	<input type="checkbox"/> ObjectId("58ade3f1aed1642711759693")	34th St - Penn Station	{ 2 fields }
3	<input type="checkbox"/> ObjectId("58ade3f1aed16427117595f2")	Times Sq - 42nd St	{ 2 fields }
4	<input type="checkbox"/> ObjectId("58ade3f1aed1642711759694")	Times Sq - 42nd St	{ 2 fields }
5	<input type="checkbox"/> ObjectId("58ade3f1aed1642711759697")	42nd St - Port Authority Bus Term	{ 2 fields }
6	<input type="checkbox"/> ObjectId("58ade3f1aed164271175957d")	Times Sq - 42nd St	{ 2 fields }
7	<input type="checkbox"/> ObjectId("58ade3f1aed16427117595c0")	Times Sq - 42nd St	{ 2 fields }
8	<input type="checkbox"/> ObjectId("58ade3f1aed164271175969b")	Herald Sq - 34th St	{ 2 fields }
9	<input type="checkbox"/> ObjectId("58ade3f1aed16427117595be")	Herald Sq - 34th St	{ 2 fields }
10	<input type="checkbox"/> ObjectId("58ade3f1aed1642711759569")	42nd St - Bryant Pk	{ 2 fields }
11	<input type="checkbox"/> ObjectId("58ade3f1aed16427117595f4")	28th St	{ 2 fields }
12	<input type="checkbox"/> ObjectId("58ade3f1aed1642711759703")	34th St - Hudson Yards	{ 2 fields }
13	<input type="checkbox"/> ObjectId("58ade3f1aed16427117596ff")	5th Ave - Bryant Pk	{ 2 fields }
14	<input type="checkbox"/> ObjectId("58ade3f1aed164271175968f")	49th St	{ 2 fields }
15	<input type="checkbox"/> ObjectId("58ade3f1aed16427117595bd")	28th St	{ 2 fields }
16	<input type="checkbox"/> ObjectId("58ade3f1aed16427117596e1")	23rd St	{ 2 fields }

Données obtenues sur les stations candidates :

Nom	Distance à la DSA	Nombre de salles de spectacle à moins de 500m	Nombre de plaintes pour la qualité de l'eau (2016-2017)
34th St – Penn Station	185.6m	2	26
34th St – Penn Station	394.3m	2	29
Times Sq – 42nd St	438.9m	27	18
Times Sq – 42nd St	445.5m	29	18
42nd St – Port Authority Bus Term	452.7m	43	21
Times Sq – 42nd St	482.6m	26	18
Times Sq – 42nd St	574.4m	35	19
Herald Sq – 34th St	589.5m	2	26
Herald Sq – 34th St	592.8m	2	29
42nd St – Bryant Pk	658.6m	16	17
28th St	734.3m	3	27
34th St – Hudson Yards	848.3m	0	15
5th Avenue – Bryant Pk	878.7m	5	19
49th St	962.2m	45	20
28th St	976.4m	1	35
23rd St	997m	2	30

Nombre moyen de salles de spectacle dans un rayon de 500m : 15

Nombre moyen de plaintes pour la qualité de l'eau (2016-2017) : 23

Score d'une station candidate : $\text{écart à la moyenne}(\text{nbSalles})^2 / \text{moyenne}(\text{nbSalles}) + \text{écart à la moyenne}(\text{nbPlaintes})^2 / \text{moyenne}(\text{nbPlaintes})$

Station de métro	Score
49th St	152,4386920981
42nd St – Port Authority Bus Term	142,4928247048
Times Sq – 42nd St	97,4050862852
Times Sq – 42nd St	70,1920072661
Times Sq – 42nd St	62,7253405995
Times Sq – 42nd St	59,1920072661
28th St	53,4726612171
23rd St	39,5037238874
34th St – Penn Station	36,9315168029
Herald Sq – 34th St	36,9315168029
28th St	32,3820163488
Herald Sq – 34th St	29,7380563124
42nd St – Bryant Pk	29,6661217075
5th Avenue – Bryant Pk	17,4050862852
34th St – Penn Station	11,6755563124
34th St – Hudson Yards	9,8092643052