

ECOLE NATIONALE DE LA STATISTIQUE ET DE L'ANALYSE DE  
L'INFORMATION

---

# Projet NoSQL

## Où vivre à New York ?

---

*Auteur :*  
Pierre LAFFITTE

11 février 2018



# Chapitre 1

## Définition du meilleur endroit où vivre à New York

Le but de ce projet est de déterminer l'endroit idéal où vivre à New York.

Avant de pouvoir donner une adresse précise, dans un premier temps, je vais affiner ma zone de recherche, en déterminant d'abord un quartier (Borough). Une fois avoir trouvé mon quartier idéal, je vais essayer d'affiner d'essayer de trouver un quartier ou une rue.

Je vais vous donner ma définition de l'endroit idéal :

Dans un premier temps, j'essaierai de m'installer dans un quartier dynamique, avec beaucoup d'activités culturelles et dans lequel je pourrais m'installer avec ma famille. Par ailleurs être proche des musées et des cinémas serait appréciable.

Une fois la zone de recherche, j'essairai de trouver une adresse, proche d'une station de métro pour pouvoir bouger facilement. Un dernier critère et pas des moindres, étant fan de la NBA, mon logement idéal devrait me permettre d'accéder rapidement à la salle des Knicks de New York. Il faudrait que mon logement soit à proximité d'une des lignes de métro me permettant d'aller directement au Madison Square Garden, sans avoir besoin de changer de métro.

# Chapitre 2

## Les bases de données

### 2.1 Description des jeux de données

Les données viennent du site de l'open data de la ville de New York. J'ai choisi d'importer 4 bases de données.

#### 2.1.1 Commodités

Ces bases vont me permettre d'affiner ma recherche dans un quartier, pour m'installer sur du long terme :  
Les hôpitaux : la présence de services de santé, en cas de pépin serait très utile.  
marchés  
Les marchés : la présence de marchés, bio qui plus est un plus.

#### 2.1.2 Activités culturelles

Les musées : cette base va essayer de me déterminer dans quelle rue du quartier considéré j'aimerais me trouver.

#### 2.1.3 Stations de métro

Les stations de métro : cette base me permettra de définir plus précisément la rue.

### 2.2 Comment charger mes données

Créer une machine virtuelle ubuntu 17.10. Après avoir télécharger les fichiers sur mon github, exécuter la ligne de commande suivante dans un terminal : `bash script.sh`.  
Ce script shell va télécharger mongoDB, les bases de données utilisées dans ce projet et les modules python pour remplir la base de données.

# Chapitre 3

## Les requêtes

### 3.1 Comment exécuter mes requêtes

Une fois les installations terminées, se placer dans le dossier des requêtes. Ouvrir un terminal dans ce dossier. Exécuter la ligne de commande : "mongo". Il va ouvrir mongo dans la console et exécuter la commande : "use datas".

```
projet@projet-VirtualBox:~/Downloads/NoSQL-master/requetes$ mongo
MongoDB shell version v3.6.2
connecting to: mongodb://127.0.0.1:27017
MongoDB server version: 3.6.2
Server has startup warnings:
2018-02-11T13:02:42.398+0100 I STORAGE [initandlisten]
2018-02-11T13:02:42.398+0100 I STORAGE [initandlisten] ** WARNING: Using the XFS
filesystem is strongly recommended with the WiredTiger storage engine
2018-02-11T13:02:42.398+0100 I STORAGE [initandlisten] ** See http://d
ochub.mongodb.org/core/prodnotes-filesystem
2018-02-11T13:02:43.065+0100 I CONTROL [initandlisten]
2018-02-11T13:02:43.065+0100 I CONTROL [initandlisten] ** WARNING: Access contr
ol is not enabled for the database.
2018-02-11T13:02:43.065+0100 I CONTROL [initandlisten] ** Read and wri
te access to data and configuration is unrestricted.
2018-02-11T13:02:43.065+0100 I CONTROL [initandlisten]
> use datas
switched to db datas
> 
```

Charger les requêtes dans le shell avec la structure suivante :  
load("requetei.js") avec i allant de 1 à 10

## 3.2 Ma recherche

Je commence mes recherches par le comptage des marchés et hôpitaux dans New York.

Borough	Brooklyn	Bronx	Manhattan	Queens	Staten Island
Marchés	52	28	48	22	6
Hopitaux	68	36	72	36	2

Tableau 1 - le nombre d'infrastructures par quartier

Les Borough les plus fournis sont Manhattan et Brooklyn en marchés et hôpitaux.

Borough	Child Health Center	Diagnostic & Treatment Center	Acute Care Hospital	Nursing Home
Brooklyn	32	4	6	6
Manhattan	40	4	6	2

Tableau 2 - répartition des infrastructures médicales dans Manhattan et Brooklyn.

Je regarde la composition des structures de santé dans ces deux quartiers, mais là encore les deux quartiers sont équivalents. Je vais concentrer ma recherche dans ces deux quartiers.

Par ailleurs, souhaitant aller régulièrement au Madison Square Garden, pour aller voir les matches des Knicks de New York, j'aimerais habiter sur une des lignes de métro.

```
{ "_id" : "Acute Care Hospital", "count" : 3 }
{ "_id" : "Child Health Center", "count" : 20 }
true
> load('requete5.js')
{
  "_id" : ObjectId("5a800c0862bd3f21b0193f4e"),
  "type" : "Feature",
  "properties" : {
    "name" : "34th St - Penn Station",
    "url" : "http://web.mta.info/nyct/service/",
    "line" : "1-2-3",
    "objectid" : "358",
    "notes" : "1,2,3-all times"
  },
  "geometry" : {
    "type" : "Point",
    "coordinates" : [
      -73.99105699913983,
      40.75037300003949
    ]
  }
}
true
>
```

Sortie 1 - La station de métro la plus proche du Madison Square Garden

Il s'agit de la station de la 34 rue, à la Station Penn où passent les lignes 1, 2 et 3. J'enregistre les stations correspondantes à ces trois lignes dans une nouvelle base : mes\_stations.

Pour les musées, je regarde les zip code les plus fournis en musées.

```

load('requete7.js')
{"_id" : "10011.0", "count" : 9 }
{"_id" : "10021.0", "count" : 6 }
{"_id" : "10128.0", "count" : 5 }
{"_id" : "10012.0", "count" : 5 }
{"_id" : "0.0", "count" : 5 }
{"_id" : "10013.0", "count" : 5 }
{"_id" : "10032.0", "count" : 5 }
{"_id" : "10019.0", "count" : 4 }
{"_id" : "10024.0", "count" : 4 }
{"_id" : "10301.0", "count" : 4 }
{"_id" : "10004.0", "count" : 3 }
{"_id" : "10022.0", "count" : 3 }
{"_id" : "10036.0", "count" : 3 }
{"_id" : "10001.0", "count" : 3 }
{"_id" : "11101.0", "count" : 3 }
{"_id" : "10028.0", "count" : 3 }
{"_id" : "10003.0", "count" : 3 }
{"_id" : "10002.0", "count" : 2 }

```

Sortie 2 - Nombre de musées par zip code

Le coin de New York le plus fourni est le "11011". Il s'avère que ce zipcode correspond à un coin de Manhattan. Je regarde la répartition des musées dans ce zipcode

```

> load('requete10.js')
[
  {
    "_id" : ObjectId("5a800c0862bd3f21b0193eb3"),
    "properties" : {
      "name" : "18th St",
      "objectid" : "203"
    },
    "geometry" : {
      "type" : "Point",
      "coordinates" : [
        -73.99787100060406,
        40.741039999802105
      ]
    },
    "nb_stations" : [
      "1",
      "2"
    ],
    "dist" : {
      "calculated" : 482.1009175056605,
      "location" : {
        "type" : "Point",
        "coordinates" : [
          -73.99787100060406,
          40.741039999802105
        ]
      }
    }
  }
]

```

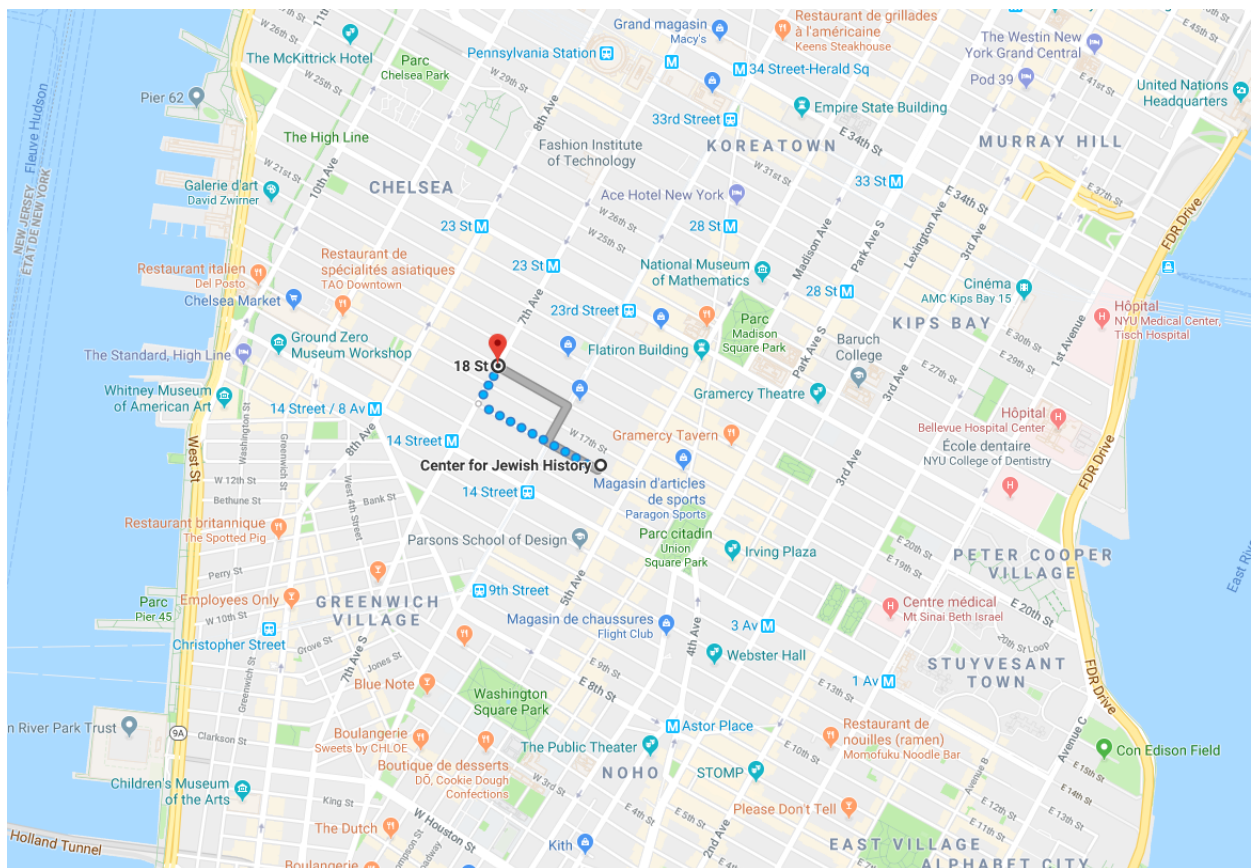
Sortie 3 - Nombre de musées par rue

Il s'avère qu'il en a 6 dans la même rue : le 15 W. 16th W street. Comme on a 6 musées à la même adresse, je décide de me trouver un logement à proximité. Maintenant je regarde quelle station (sélectionnée précédemment) est proche de cette rue : il s'agit de celle la 18eme rue. Elle se trouve à moins de 500 mètres des musées.

```
> load('requete8.js')
{ "_id" : "15 W. 16th St.", "count" : 6 }
{ "_id" : "945 Madison Ave.", "count" : 1 }
{ "_id" : "222 East 6th Street", "count" : 1 }
{ "_id" : "28 East 20th Street", "count" : 1 }
{ "_id" : "792 Eastern Pkwy", "count" : 1 }
{ "_id" : "104 E. 126th Street", "count" : 1 }
{ "_id" : "144 West 125th Street", "count" : 1 }
{ "_id" : "75 Stuyvesant Pl.", "count" : 1 }
{ "_id" : "12 Fulton Street", "count" : 1 }
{ "_id" : "550 Madison Ave", "count" : 1 }
{ "_id" : "1071 Fifth Avenue", "count" : 1 }
{ "_id" : "440 Seguin Avenue", "count" : 1 }
{ "_id" : "209 E. 23rd St.", "count" : 1 }
{ "_id" : "515 Malcolm X Boulevard", "count" : 1 }
{ "_id" : "1538 Woodrow Road", "count" : 1 }
```

Sortie 4 - Station de métro la plus proche du 15 W. 16th street

Je décide de me prendre un logement entre la 16eme et la 18eme rue à New York. Il se trouve qu'à pied, je suis aussi à proximité du Madison Square Garden.



Sortie 5 - Mon endroit idéal pour vivre à New York