

# Projet de fin de semestre

## Description générale

Ce projet consiste à donner des outils d'analyse sur un data set portant sur les morts liées à la consommation de drogue.

Les données recueillies tiennent compte de l'âge et du sexe de la personne, les cause de la mort, le lieu et le type de drogue consommé.

## Provenance des données et méthodes d'obtention

Site officiel [catalog.data.gov](https://catalog.data.gov), open data.

Données de l'état du Connecticut pour la période de 2012 à 2017.

Disponible en téléchargement sur le site :

<https://catalog.data.gov/dataset/accidental-drug-related-deaths-january-2012-sept-2015>

## Structures attendues

- Charger de l'ensemble des données en une base de données mongoDB.
- Créer d'une base de données néo4j avec comme nœuds :
  - le « CaseNumber »
  - l'âge de la victime en créant des catégories (<18,18-22,22-26,26-35,35-45,45-60,>60) ou autres catégories plus pertinentes.
  - Le type de drogue consommée.
  - La ville de résidence « Residence City »
  - tout autre nœud que vous pourriez trouver utiles.

## Livrables attendus

créé de nouveaux liens entre la tranche d'âge et type de drogue dont l'attribut « strength » met en évidence le nombre de victimes.

Créer une fonction python qui pour une ville et un type de drogues affiche tous les cas associés

Créer une fonction qui pour deux affaires (i.e. Case Number) affiche les éléments communs (si oui, ville de résidence, si oui le type de drogue consommé si oui des ages similaire...)

N'hésitez pas à nous montrer toutes autres informations ou fonctions utiles sur ces données.

## Documents attendu :

- Le code complet transférer sur le site github, celui-ci doit marcher en démarrant un docker-compose et en lançant un minimum de commande, voir aucune commande.
- Un Readme clair spécifiant :
  - La structure des bases de données.
  - L'utilité de chaque script et leur fonctionnement.
  - Les commandes à lancer pour exécuter les différentes fonctions pythons.
- Une présentation détaillant :
  - le set de données.

- les bases de données et leur structure.
- les réponses aux livrables.
- d'autres choses intéressantes s'il y en a.