

# Projet de fin de semestre

## Description générale

Ce projet consiste à donner des outils d'analyse sur un data set portant sur les infractions routières. Les données recueillies tiennent compte du type d'infraction de la couleur et de la marque du véhicule et d'information sur le conducteur.

## Provenance des données et méthodes d'obtention

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

Données de la région de Montgomery dans le Maryland aux États-Unis.

Disponible en téléchargement sur le site :

<https://catalog.data.gov/dataset/traffic-violations-56dda>

## Structures attendues

- Charger de l'ensemble des données en une base de données mongoDB.
- Créer d'une base de données néo4j avec comme nœuds :
  - l'id de l'infraction
  - La marque de la voiture,
  - La couleur de la voiture,
  - Le type d'infraction,
  - La ville du conducteur.
  - tout autre nœud que vous pourriez trouver utiles.

## Livrables attendus

Écrire une fonction python qui trouve tous les accidents dans une zone géographique délimité par des coordonnées gps.

Rajouter des liens entre les villes et le type d'infraction comptant le nombre d'infraction sur cette ville.

Écrire une fonction python qui pour deux infractions affiche toutes les similitudes s'il y en a.

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.