Projet Data-scientiste

Analyse des bases de données tirer de l’observatoire national interministériel de la sécurité routière

Jouët Jérôme / Dahlem Romain

# Description des bases de données annuelles des accidents corporels de la circulation routière

Pour chaque accident corporel (soit un accident survenu sur une voie ouverte à la circulation publique, impliquant au moins un véhicule et ayant fait au moins une victime ayant nécessité des soins), des saisies d’information décrivant l’accident sont effectuées par l’unité des forces de l’ordre (police, gendarmerie, etc.) qui est intervenue sur le lieu de l’accident. Ces saisies sont rassemblées dans une fiche intitulée bulletin d’analyse des accidents corporels. L’ensemble de ces fiches constitue le fichier national des accidents corporels de la circulation dit " Fichier BAAC1" administré par l’Observatoire national interministériel de la sécurité routière "ONISR".

## Définition Accident corporel au sein de la base de données

Un accident corporel (mortel et non mortel) de la circulation routière :

* Implique au moins une victime,
* Survient sur une voie publique ou privée, ouverte à la circulation publique,
* Implique au moins un véhicule.

Un accident corporel implique un certain nombre d’usagers. Parmi ceux-ci, on distingue :

* Les personnes indemnes : impliquées non décédées et dont l’état ne nécessite aucun soin médical du fait de l’accident.
* Les victimes : impliquées non indemnes.

Parmi les victimes, on distingue :

* Les personnes tuées : personnes qui décèdent du fait de l’accident, sur le coup ou dans les trente jours qui suivent l’accident,
* Les personnes blessées : victimes non tuées.

Parmi les personnes blessées, il convient de différencier :

* Les blessés dits « hospitalisés » : victimes hospitalisées plus de 24 heures,
* Les blessés légers : victimes ayant fait l'objet de soins médicaux mais n'ayant pas été admises comme patients à l'hôpital plus de 24 heures.

## Spécifications de la base

La base de données des accidents corporels de la circulation récupéré sur le site data.gouv.fr est répartie en 4 rubriques par année données

1. La rubrique CARACTERISTIQUES qui décrit les circonstances générales de l’accident
2. La rubrique LIEUX qui décrit le lieu principal de l’accident même si celui-ci s’est déroulé à une intersection
3. La rubrique VEHICULES impliqués
4. La rubrique USAGERS impliqués

Le n° d'identifiant de l’accident (Cf. "Num\_Acc") présent dans ces 4 rubriques permet d'établir un lien entre toutes les variables qui décrivent un accident. Quand un accident comporte plusieurs véhicules, il faut aussi pouvoir relier chaque véhicule à ses occupants. Ce lien est fait par la variable Num\_veh.

## Liste complète des champs avec le détail de leur contenu pour chaque fichier

### Le fichier CARACTERISTIQUES

Num\_Acc : Numéro d'identifiant de l’accident

Lum : Lumière : conditions d’éclairage dans lesquelles l'accident s'est produit

* 1 – Plein jour
* 2 – Crépuscule ou aube
* 3 – Nuit sans éclairage public
* 4 - Nuit avec éclairage public non allumé
* 5 – Nuit avec éclairage public allumé

Dep : Département : Code INSEE (Institut National de la Statistique et des Etudes Economiques) du département suivi d'un 0 (201 Corse-du-Sud - 202 Haute-Corse)

Com : Commune : Le numéro de commune est un code donné par l‘INSEE. Le code comporte 3 chiffres calés à droite.

Agg : Localisation :

* 1 – Hors agglomération
* 2 – En agglomération

Int : Intersection :

* 1 – Hors intersection
* 2 – Intersection en X
* 3 – Intersection en T
* 4 – Intersection en Y
* 5 - Intersection à plus de 4 branches
* 6 - Giratoire
* 7 - Place
* 8 – Passage à niveau
* 9 – Autre intersection

Atm : Conditions atmosphériques :

* 1 – Normale
* 2 – Pluie légère
* 3 – Pluie forte
* 4 – Neige - grêle
* 5 – Brouillard - fumée
* 6 – Vent fort - tempête
* 7 – Temps éblouissant
* 8 – Temps couvert
* 9 – Autre

col : Type de collision :

* 1 – Deux véhicules - frontale
* 2 – Deux véhicules – par l’arrière
* 3 – Deux véhicules – par le coté
* 4 – Trois véhicules et plus – en chaîne
* 5 – Trois véhicules et plus - collisions multiples
* 6 – Autre collision
* 7 – Sans collision

adr : Adresse postale : variable renseignée pour les accidents survenus en agglomération

gps : Codage GPS :1 caractère indicateur de provenance :

* M = Métropole
* A = Antilles (Martinique ou Guadeloupe)
* G = Guyane
* R = Réunion
* Y = Mayotte

lat : Latitude

long : Longitude

### Le fichier LIEUX

Num\_Acc : Identifiant de l’accident identique à celui du fichier "rubrique CARACTERISTIQUES" repris du fichier csv Accident

catr : Catégorie de route :

* 1 - Autoroute
* 2 - Route Nationale
* 3 - Route Départementale
* 4 - Voie Communale
* 5 - Hors réseau public
* 6 - Parc de stationnement ouvert à la circulation publique
* 9 – autre

voie : Numéro de la route

V1 : Indice numérique du numéro de route (exemple : 2 bis, 3 ter etc.)

V2 : Lettre indice alphanumérique de la route

circ : Régime de circulation :

* 1 – A sens unique
* 2 – Bidirectionnelle
* 3 – A chaussées séparées
* 4 – Avec voies d’affectation variable

nbv : Nombre total de voies de circulation

vosp : Signale l’existence d’une voie réservée, indépendamment du fait que l’accident ait lieu ou non sur cette voie.

* 1 – Piste cyclable
* 2 – Banque cyclable
* 3 – Voie réservée

prof : Profil en long décrit la déclivité de la route à l'endroit de l'accident

* 1 - Plat
* 2 - Pente
* 3 - Sommet de côte
* 4- Bas de côte

pr : Numéro du PR de rattachement (numéro de la borne amont)

pr1 : Distance en mètres au PR (par rapport à la borne amont)

plan : Tracé en plan :

* 1 – Partie rectiligne
* 2 – En courbe à gauche
* 3 – En courbe à droite
* 4 – En « S »

lartpc : Largeur du terre plein central (TPC) s'il existe

larrout : Largeur de la chaussée affectée à la circulation des véhicules ne sont pas compris les bandes d'arrêt d'urgence, les TPC et les places de stationnement

surf : Etat de la surface

* 1 - normale
* 2 - mouillée
* 3 - flaques
* 4 - inondée
* 5 - enneigée
* 6 - boue
* 7 - verglacée
* 8 - corps gras - huile
* 9 - autre

infra : Aménagement - Infrastructure :

* 1 – Souterrain - tunnel
* 2 – Pont - autopont
* 3 – Bretelle d’échangeur ou de raccordement
* 4 - Voie ferrée
* 5 – Carrefour aménagé
* 6 – Zone piétonne
* 7 – Zone de péage

situ : Situation de l’accident :

* 1 – Sur chaussée
* 2 – Sur bande d’arrêt d’urgence
* 3 – Sur accotement
* 4 – Sur trottoir
* 5 – Sur piste cyclable

env1 : point école : proximité d'une école

### Le fichier VÉHICULES

Num\_Acc : Identifiant de l’accident identique à celui du fichier "rubrique CARACTERISTIQUES" repris pour chacun des véhicules décrits impliqués dans l’accident

Num\_Veh : Identifiant du véhicule repris pour chacun des usagers occupant ce véhicule (y compris les piétons qui sont rattachés aux véhicules qui les ont heurtés) – Code alphanumérique

senc : Sens de circulation :

* 1 – PK ou PR ou numéro d’adresse postale croissant
* 2 – PK ou PR ou numéro d’adresse postale décroissant

catv : Catégorie du véhicule :

* 01 - Bicyclette
* 02 - Cyclomoteur <50cm3
* 03 - Voiturette (Quadricycle à moteur carrossé) (anciennement "voiturette ou tricycle à moteur")
* 04 - Référence plus utilisée depuis 2006 (scooter immatriculé)
* 05 - Référence plus utilisée depuis 2006 (motocyclette)
* 06 - Référence plus utilisée depuis 2006 (side-car)
* 07 - VL seul
* 08 - Catégorie plus utilisée (VL + caravane)
* 09 - Catégorie plus utilisée (VL + remorque)
* 10 - VU seul 1,5T <= PTAC <= 3,5T avec ou sans remorque (anciennement VU seul 1,5T <= PTAC <=
* 3,5T)
* 11 - Référence plus utilisée depuis 2006 (VU (10) + caravane)
* 12 - Référence plus utilisée depuis 2006 (VU (10) + remorque)
* 13 - PL seul 3,5T <PTCA <= 7,5T
* 14 - PL seul > 7,5T
* 15 - PL > 3,5T + remorque
* 16 - Tracteur routier seul
* 17 - Tracteur routier + semi-remorque
* 18 - Référence plus utilisée depuis 2006 (transport en commun)
* 19 - Référence plus utilisée depuis 2006 (tramway)
* 20 - Engin spécial
* 21 - Tracteur agricole
* 30 - Scooter < 50 cm3
* 31 - Motocyclette > 50 cm3 et <= 125 cm3
* 32 - Scooter > 50 cm3 et <= 125 cm3
* 33 - Motocyclette > 125 cm3
* 34 - Scooter > 125 cm3
* 35 - Quad léger <= 50 cm3 (Quadricycle à moteur non carrossé)
* 36 - Quad lourd > 50 cm3 (Quadricycle à moteur non carrossé)
* 37 - Autobus
* 38 - Autocar
* 39 - Train
* 40 - Tramway
* 99 - Autre véhicule (dont piéton en roller ou en trottinette à partir de l’année 2018 requalifié en "engin de déplacement personnel")

obs : Obstacle fixe heurté :

* 1 – Véhicule en stationnement
* 2 – Arbre
* 3 – Glissière métallique
* 4 – Glissière béton
* 5 – Autre glissière
* 6 – Bâtiment, mur, pile de pont
* 7 – Support de signalisation verticale ou poste d’appel d’urgence
* 8 – Poteau
* 9 – Mobilier urbain
* 10 – Parapet
* 11 – Ilot, refuge, borne haute
* 12 – Bordure de trottoir
* 13 – Fossé, talus, paroi rocheuse
* 14 – Autre obstacle fixe sur chaussée
* 15 – Autre obstacle fixe sur trottoir ou accotement
* 16 – Sortie de chaussée sans obstacle

obsm : Obstacle mobile heurté :

* 1 – Piéton
* 2 – Véhicule
* 4 – Véhicule sur rail
* 5 – Animal domestique
* 6 – Animal sauvage
* 9 – Autre

choc : Point de choc initial :

* 1 - Avant
* 2 – Avant droit
* 3 – Avant gauche
* 4 – Arrière
* 5 – Arrière droit
* 6 – Arrière gauche
* 7 – Côté droit
* 8 – Côté gauche
* 9 – Chocs multiples (tonneaux)

manv : Manœuvre principale avant l’accident :

* 1 – Sans changement de direction
* 2 – Même sens, même file
* 3 – Entre 2 files
* 4 – En marche arrière
* 5 – A contresens
* 6 – En franchissant le terre-plein central
* 7 – Dans le couloir bus, dans le même sens
* 8 – Dans le couloir bus, dans le sens inverse
* 9 – En s’insérant
* 10 – En faisant demi-tour sur la chaussée

**Changeant de file**

* 11 – A gauche
* 12 – A droite

**Déporté**

* 13 – A gauche
* 14 – A droite
* Tournant
* 15 – A gauche
* 16 – A droite

**Dépassant**

* 17 – A gauche
* 18 – A droite

**Divers**

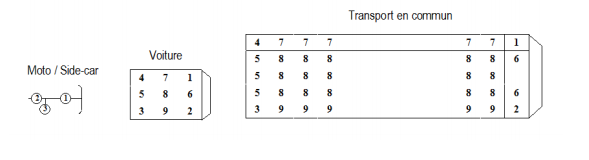
* 19 – Traversant la chaussée
* 20 – Manœuvre de stationnement
* 21 – Manœuvre d’évitement
* 22 – Ouverture de porte
* 23 – Arrêté (hors stationnement)
* 24 – En stationnement (avec occupants)

occutc : Nombre d’occupants dans le transport en commun

### Le fichier USAGERS

Num\_Acc : Identifiant de l’accident identique à celui du fichier "rubrique CARACTERISTIQUES" repris pour chacun des usagers décrits impliqués dans l’accident

Num\_Veh : Identifiant du véhicule repris pour chacun des usagers occupant ce véhicule (y compris les piétons qui sont rattachés aux véhicules qui les ont heurtés) – Code alphanumérique

place : Permet de situer la place occupée dans le véhicule par l'usager au moment de l'accident catu : Catégorie d'usager :

* 1 - Conducteur
* 2 - Passager
* 3 - Piéton
* 4 - Piéton en roller ou en trottinette (catégorie déplacée, à partir de l’année 2018, vers le fichier "Véhicules" Catégorie du véhicule :
* 99 - Autre véhicule. Cette catégorie est désormais considérée comme un véhicule : engin de déplacement personnel)

grav : Gravité de l'accident : Les usagers accidentés sont classés en trois catégories de victimes plus les indemnes :

* 1 - Indemne
* 2 - Tué
* 3 - Blessé hospitalisé
* 4 - Blessé léger

sexe : Sexe de l'usager

* 1 - Masculin
* 2 – Féminin

An\_nais : Année de naissance de l'usager

trajet : Motif du déplacement au moment de l’accident :

* 1 – Domicile – travail
* 2 – Domicile – école
* 3 – Courses – achats
* 4 – Utilisation professionnelle
* 5 – Promenade – loisirs
* 9 – Autre

secu : sur 2 caractères : le premier concerne l’existence d’un Équipement de sécurité

* 1 – Ceinture
* 2 – Casque
* 3 – Dispositif enfants
* 4 – Equipement réfléchissant
* 9 – Autre

Le second concerne l’utilisation de l’Équipement de sécurité

* 1 – Oui
* 2 – Non
* 3 – Non déterminable

locp : Localisation du piéton :

**Sur chaussée :**

* 1 – A + 50 m du passage piéton
* 2 – A – 50 m du passage piéton

**Sur passage piéton :**

* 3 – Sans signalisation lumineuse
* 4 – Avec signalisation lumineuse

**Divers :**

* 5 – Sur trottoir
* 6 – Sur accotement
* 7 – Sur refuge ou BAU
* 8 – Sur contre allée

actp : Action du piéton :

**Se déplaçant**

* 0 - non renseigné ou sans objet
* 1 - Sens véhicule heurtant
* 2 - Sens inverse du véhicule

**Divers**

* 3 - Traversant
* 4 - Masqué
* 5 - Jouant – courant
* 6 - Avec animal
* 9 - Autre

etatp : Cette variable permet de préciser si le piéton accidenté était seul ou non

* 1 – Seul
* 2 – Accompagné
* 3 – En groupe

# Travail réalisé

Nous avons décidé d’utiliser les bases de données de 2015 à 2018, il a donc fallu intégrer un total de 16 bases de données CSV (4 par année) au projet

A l’initialisation du projet nous avons souhaité pouvoir analyser les données le plus facilement possible, pour cela il nous a été nécessaire de normaliser la totalité des bases de données (de transformer les id avec leur chaine de caractère correspondant), les opérations sont présentes dans le fichier ***Normalize\_Carac.py***, grâce a ce fichier nous pouvons prendre n’importe quelle base de données comportant les données similaires

Suite à cela il a été nécessaires de merger les 4 base d’une année les unes avec les autres afin d’avoir toutes les informations des accidents, les opérations qui servent à cela sont présente dans le fichier ***DataAccident.py***

Nous avons choisi de créer des fichier python pour ses opérations afin de pouvoir facilement les appelés depuis notre fichier V***isualisation.ipynb***

## Visualisation.ipynb

Ce fichier regroupe la totalité des opérations visant à offrir une visibilité a l’analyse des données des différente bases de données.

Nous avons tout d’abord importé les lib nécessaire vue au cours de la formation en veillant à bien respecter les normes de nommage :

import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
import DataAccident as data

import seaborn as sns

Ainsi que l’importe de DataAccident qui permet la récupération des bases de données merger

df2018 = data.getData2018Merge()  
df2017 = data.getData2017Merge()  
df2016 = data.getData2016Merge()  
df2015 = data.getData2015Merge()

## Véhicule Accidentée par catégorie de 2015 à 2018

On peut voir dans ce diagramme que les véhicules qui ont subi le plus d’accident au cours de l’année sont les véhicules de type voiture suivie par les 2 roues, étant donné que ce sont les véhicules les plus présent sur la voie publique, il n’est pas surprenant de les retrouver en 1ere et 2nd position.

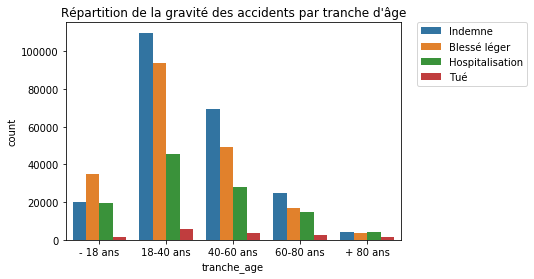
## Répartition par tranche d’âge de la gravité des accidents de 2015 à 2018

## 

## Répartition par gravité des accidents par année

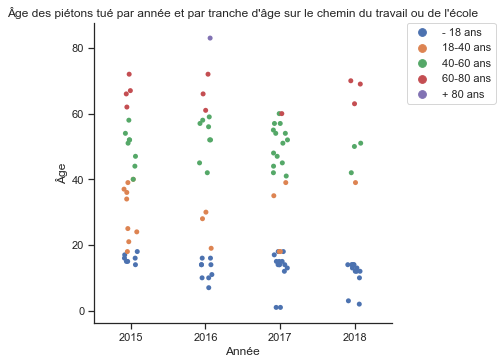
Nous pouvons voir dans ce diagramme qu’il y à sensiblement le même nombre de personne accidentées peu importe la gravité des victimes tous les ans. Cependant il y à une différence flagrante entre le nombre de personne Hospitalisé entre 2017 et 2018, il est possible que ce sois grâce au changement de vitesse autorisée du 1er juillet 2018 qui est passé de 90 à 80 sur les routes à forte mortalité

## Répartition par gravité des par tranche d’âge



Ici nous pouvons déduire que les conducteurs entre 18 et 40 ans ainsi que ceux entre 40 et 60 ans sont les plus touché dans les accidents. C’est chiffre sont tout à fait logique du au fait que les personnes les plus actives dans la vie de tous les jours se situe dans celle-ci.

## Âge des piétons tués par année et par tranche d'âge sur le chemin du travail ou de l'école

Nous avons ici la représentation graphique en points des piétons tué sur le chemin du travail ou école par tranche d’âge.

Ici nous pouvons voir que les plus touché sont malheureusement les enfants, malgré le fait qu’il y a une impression de régression du nombre de tué par années

## Répartitions des accidents survenue en et hors agglomération

Ici nous pouvons apercevoir qu’il y a plus d’accident qui surviennent en agglomération que hors agglomération, il est normal d’avoir plus d’accident en agglomération que hors agglomération à cause du nombre de personne en ville.

## Répartitions de la gravité des accidents en fonction de l’agglomération

Ici nous pouvons apercevoir qu’il y a plus de pourcentage de personne tué et moins de personne indemne ou blessé légèrement hors agglomération, étant donné que la vitesse étant plus importante hors agglomération on peut donc affirmer que la vitesse de l’accident rend les séquelles subit par les victimes plus importante.

## Répartition de la gravité des accidents par mois

On peut apercevoir ici qu’il y a au mois d’aout moins d’accident en général, on peut aussi voir qu’il y a moins d’accident au mois de février ce qui peut être expliquer par un mois moins long en termes de jour, nous avions personnellement pensé qu’il y aurait plus d’accidents sur les mois de juillet et aout a cause des départs en vacances qui surcharge les routes.

## Répartition de la gravité des accidents par heure

On peut apercevoir sur ce graphique que les heure ou il y à le plus d’accidents sont les heure comprise entre 16h et 20h, ce qui n’est pas anodin car ce sont les heure ou les usagers de la route rentre du travail possiblement fatigué de leur journée.

## Les cartographie

Nous avons généré des page html pour représenter une cartographie des accidents de 2018 et 2017

Pour cela nous avons utilisé les librairies *folium* et *markerCluster* de folium, vous trouverez ces cartes en lançant le fichier ***Cartographie\_Personne\_Morte\_2017-2018.html***

Le marker rouge représente les personnes tuées sur la route en 2017 et les bleus ceux tuées en 2018