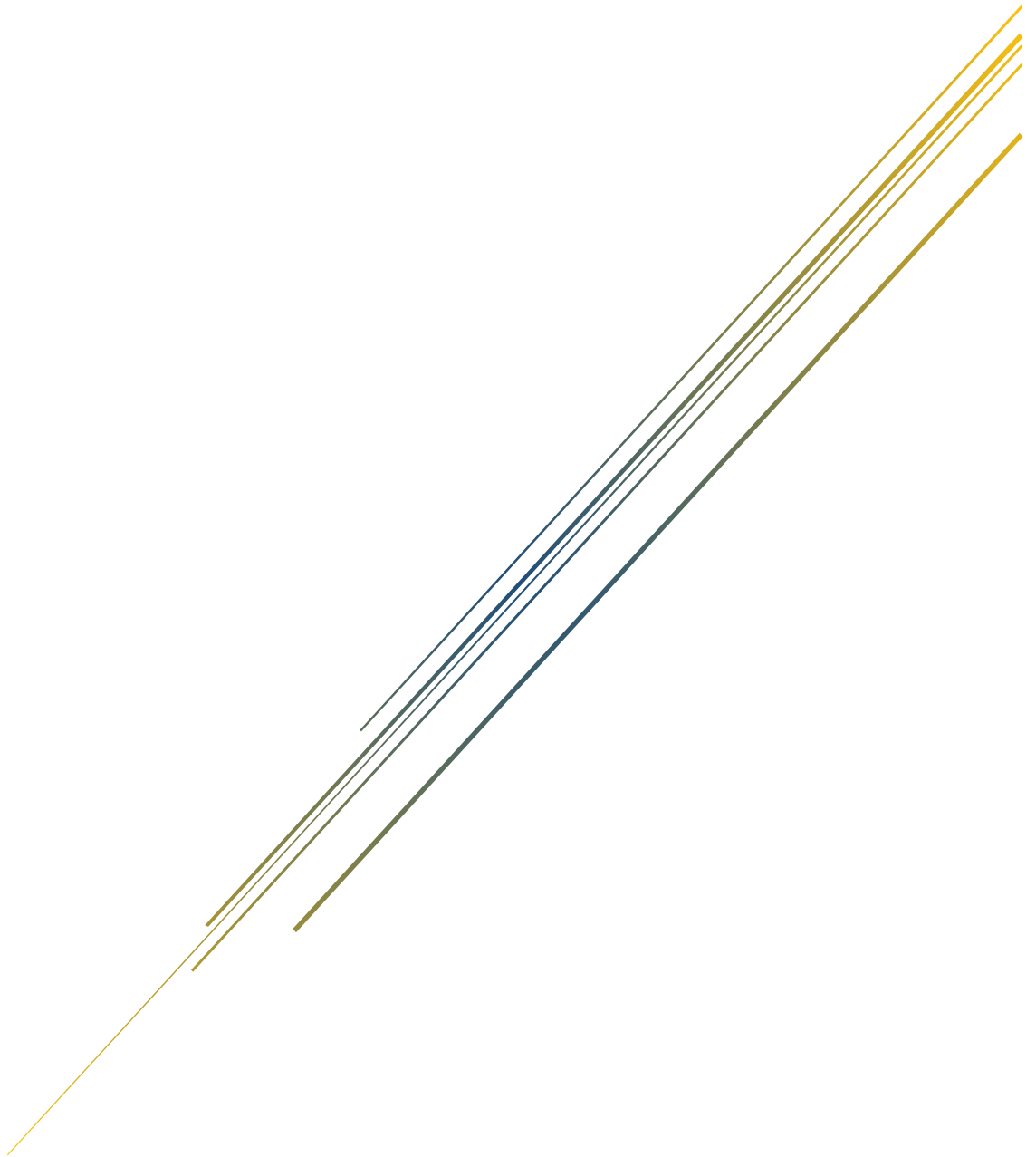


## S2.04 - EXPLOITATION D'UNE BASE DE DONNEES



## Table des matières

Etape 1 : Prise en main de la base et réflexion statistique .....	2
➤ Tableau individus-variables.....	2
➤ Qui sont les individus du tableau individus-variables ? .....	3
➤ Liste des variables pouvant être analysées .....	3
➤ Problématique imaginé.....	5
Etape 2 : Production statistique (4 points).....	8

## Table des illustrations

Figure 1 : Capture d'écrans du concepteur de la base de données .....	2
Graphe 1 : Gravité des accidents en fonction de leurs causes.....	5
Graphe 2 : Pourcentage d'accidents dû à l'ivresse par rapport à son total selon le type de luminosité produite entre 1984 et 1998 .....	6
Graphe 3 : Pourcentage d'accidents par type de personne impliqué liées à l'ivresse.....	6
Graphe 4 : Comparaison du nombre d'accidents avec et sans alcool par année .....	7

## Etape 1 : Prise en main de la base et réflexion statistique

### ➤ Conception de la base de données :

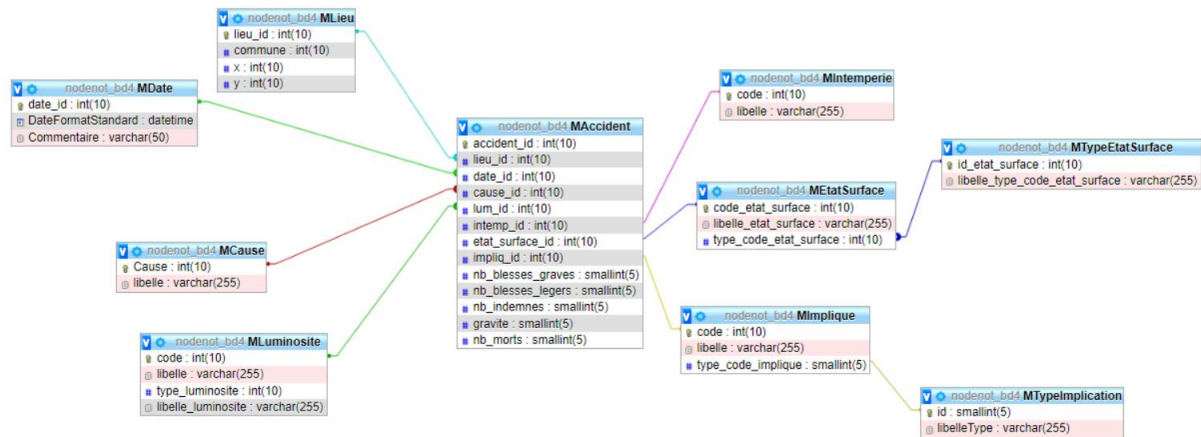


Figure 1 : Capture d'écrans du concepteur de la base de données

### ➤ Tableau individus-variables

Un tableau individus-variables est un tableau disposant d'autant de ligne qu'il contient d'accidents, et ayant autant de colonne qu'il y a d'information lié à un accident. C'est-à-dire que pour un accident, nous pouvons connaître où a eu lieu l'accident, dans quelle certaine zone, à quelle date/heure etc....

- **Sous Excel**, pour organiser les données il faut faire correspondre pour chaque accident leurs numéro ainsi que sa clef étrangère. Ils seront effectués grâce à la RechercheV() disponible sous Excel.

*Exemple* : Pour lier chaque individu avec le libellé de la cause de l'accident, on tape =RechercheV(valeurChercher ; MCause ; 2 ; Tue) où "valeurChercher" correspond à une valeur de la colonne Cause\_id de la feuille MAccident.

- **Sous Python**, on utilise la bibliothèque Pandas. Elle permet de manipuler des données, et permet de gérer des DataFrame.

On commence par utiliser ReadTable pour lire les données « texte » et importer les données dans un DataFrame. Ensuite on utilise la fonction merge permettant de faire une jointure entre deux DataFrame.

*Exemple* : Pour lier chaque individu avec la date de l'accident, nous faisons merge le fichier accident (fic\_accident) avec le fichier date (fic\_date) avec une jointure LEFT sur la date\_id, ce qui donne :  
accident\_date = fic\_accident.merge( fic\_date, how = « left », on= « date\_id » )

Une autre manière de le faire est de faire correspondre l'index de MAccident avec la Cause ce qui donne :  
accident\_ivresse = accident\_ivresse.merge(fic\_cause, left\_on='cause\_id', right\_on='Cause', left\_index = False, right\_index = True)

➤ Qui sont les individus du tableau individus-variables ?

Chaque ligne de ce tableau correspond à un individu et dans notre cas l'individu est un accident.

➤ Liste des variables pouvant être analysées

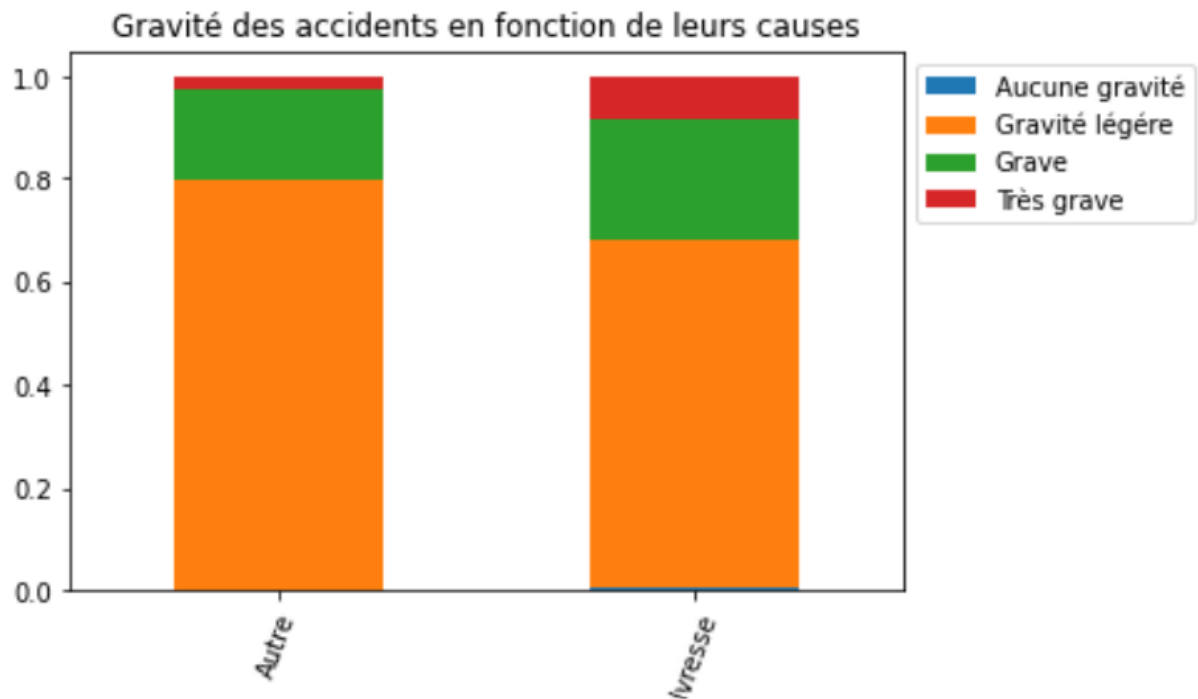
Nom	Type	Liste des modalités
nb_blesses_graves	Quantitative discrète	Min : 0 Max : 6
nb_blesses_legers	Quantitative continue	Min : 0 Max : 24
nb_indemnes	Quantitative discrète	Min : 0 Max : 9
gravite	Quantitative discrète	Min : 0 Max : 3
nb_morts	Quantitative discrète	Min : 0 Max : 2
labelle_luminosite	Qualitative ordinales	Jour, demi-jour, nuit éclairée, nuit éclairée insuffisant, nuit sans éclairage
cause	Qualitative nominale	Répertoire 58 causes d'accident
date	Qualitative ordinales	Va de 1er février 1984 au 30 avril 1998
etat_surface	Qualitative nominale	Humide, Mouillée, Enneigée, Verglacée, Gras boueux, Gravillons, Sec Normal
impliquer	Qualitative nominale	Autre, Piéton moins 10 ans, Piéton plus 10 ans, Bicyclette, Cyclomoteur, Moto, Voiture légère, Poids

intempérie	Qualitative nominale	lourd, Bus, Tram, Engin  Beau temps, Pluie forte, Pluie légère, Neige, Grêle, Brouillard, Vent fort tempête, Inconnu
commune	Qualitative ordinales	Numéro de la commune
luminosité	Qualitative ordinales	Jour, Nuit
implication	Qualitative nominale	Autre raison, Humain, Deux roues, Véhicule
Type_etat_surface	Qualitative ordinales	Normal, Dangereux, T Dangereux
Description	Qualitative ordinales	Jour férié : Armistice, lundi de Pâque, Noel, Assomption, Lundi de Pentecôte, Fête nationale, Fête du travail, Victoire des Alliés, Jour de l'an, Ascension

## ➤ Problématique imaginé

Nous souhaitons nous intéresser aux accidents impliquants de l'alcool :

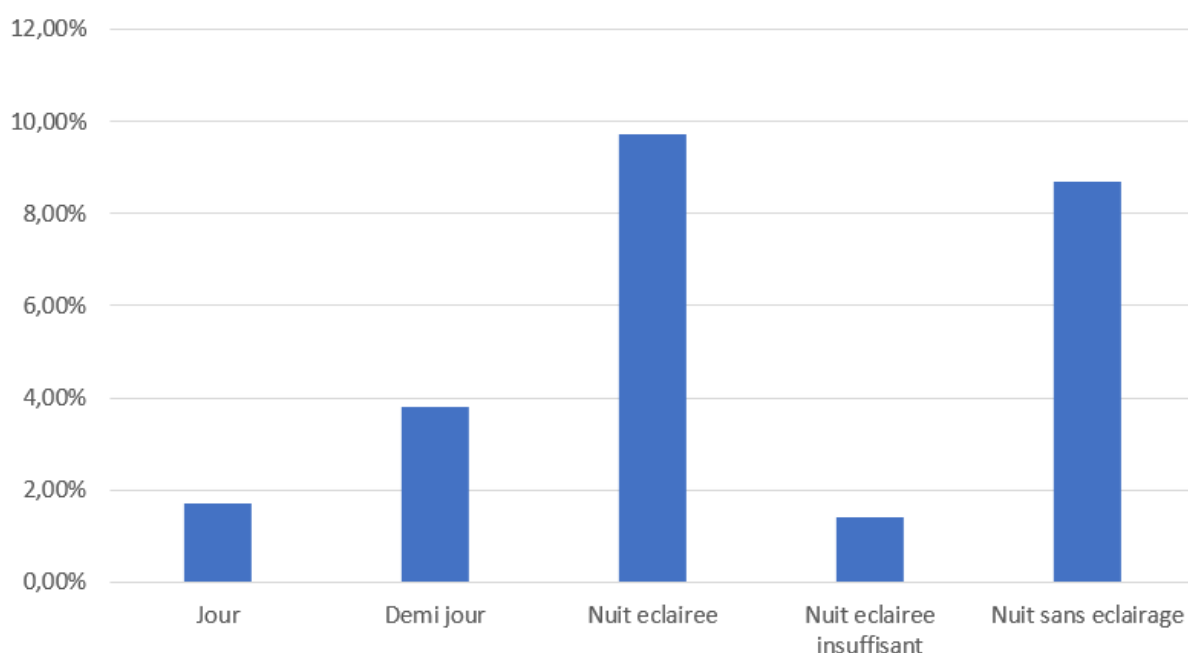
Sont-ils plus graves que les autres types d'accidents ?



Graph 1 : Gravité des accidents en fonction de leurs causes

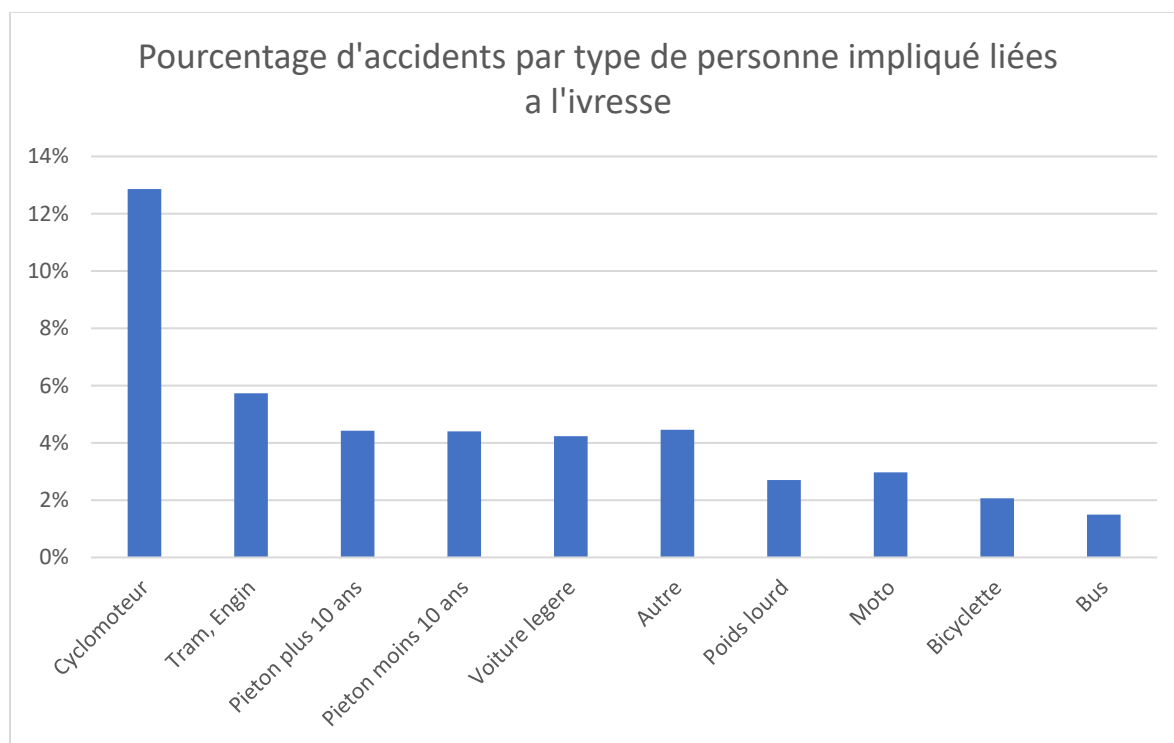
Y a-t-il un moment pour laquelle les accidents dû à l'alcool sont plus fréquent ?

Pourcentage d'accidents dû à l'ivresse par rapport à son total selon le type de luminosité produite entre 1984 et 1998.



Graph 2 : Pourcentage d'accidents dû à l'ivresse par rapport à son total selon le type de luminosité produite entre 1984 et 1998

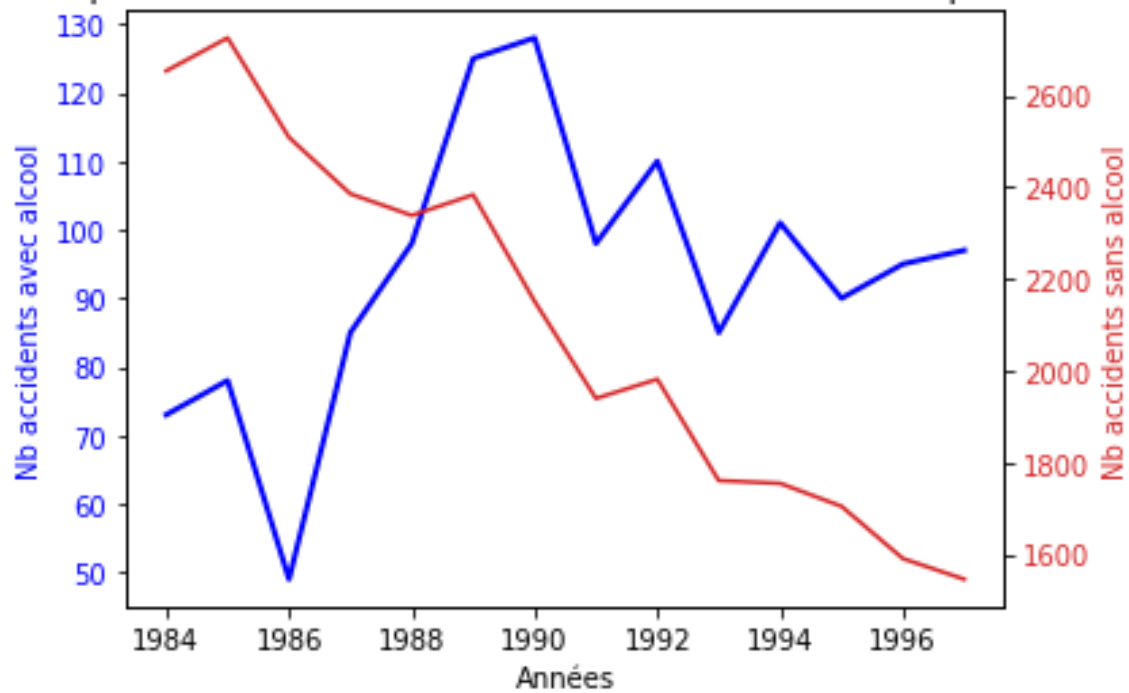
Quels types de véhicule est le plus fréquent lorsqu'il y a un accident impliquant de l'alcool ?



Graph 3 : Pourcentage d'accidents par type de personne impliquée liées à l'ivresse

Le nombre d'accident impliquant l'alcool a-t-il augmenter au fur et à mesure des années ?

Comparaison du nombre d'accidents avec et sans alcool par année



Graphe 4 : Comparaison du nombre d'accidents avec et sans alcool par année



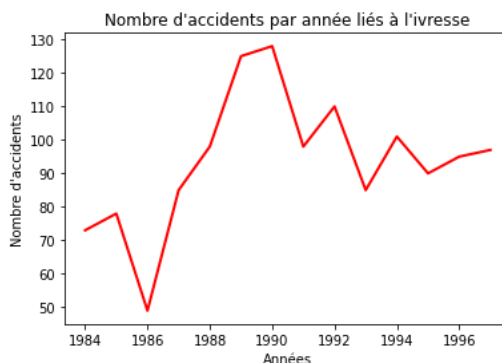
## Etape 2 : Production statistique (4 points)

- Vous expliquerez, dans un paragraphe, quelle a été votre démarche à partir des questions que vous vous êtes posées à la fin de l'étape 1. Quels graphiques vous avez réalisés ? Quels graphiques vous avez finalement choisis ? Pour quelle raison ?  
*Vous pourrez donner en exemple les graphiques que vous n'avez pas gardés. Contrairement à ceux de l'affiche/tableau de bord, il n'est pas nécessaire que ces graphiques (titres, légende) soient "parfaits".*

On s'est réparti les questions pour nous permettre d'avoir une charge de travail équivalente, l'un a produit les 2 premières questions, et l'autre les 2 dernières.

Pour chaque question, nous avons regroupé les variables à utiliser dans le graphique, puis nous les avons exploitées.

Premièrement, nous avons créé des graphiques ne liant que l'ivresse à elle-même, compliquant l'analyse. C'est donc dans un deuxième temps que nous avons inclus les autres types d'accidents permettant de voir sa proportionnalité.



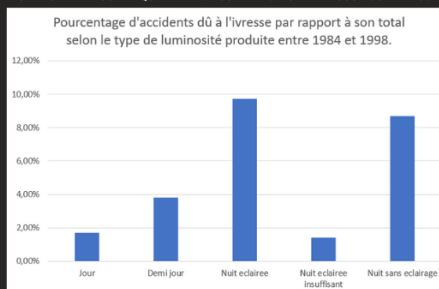
Par exemple, nous n'avons pas gardé ce graphique car il n'était pas assez représentatif. Nous l'avons donc amélioré afin d'avoir une autre courbe représentant le nombre d'accidents par année et qui ne sont pas liés à l'ivresse. Quand nous l'avons fait l'échelle n'était pas non plus adapté, le nombre d'accidents liés à l'ivresse comparé aux autres était faible. Nous avons donc fait un graphique à deux échelles qui est exploitable et très représentatif. (cf. [Graph 4 : Comparaison du nombre d'accidents avec et sans alcool par année](#))

- Vous donnerez l'affiche/tableau de bord.

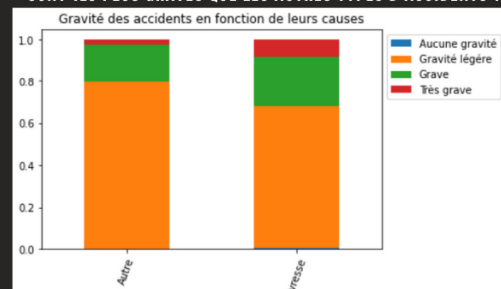
**LA SOLUTION NE SE TROUVE PAS  
AU FOND D'UNE BOUTEILLE.**

# UN MONDE SANS L'ALCOOLISME AU VOLANT

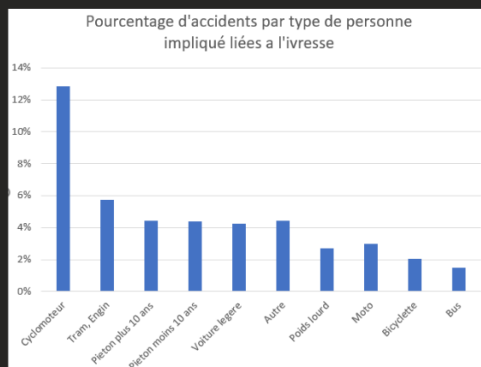
Y A-T-IL UN MOMENT POUR LEQUEL LES ACCIDENTS DÙ À L'IVRESSE SONT PLUS FRÉQUENT ?



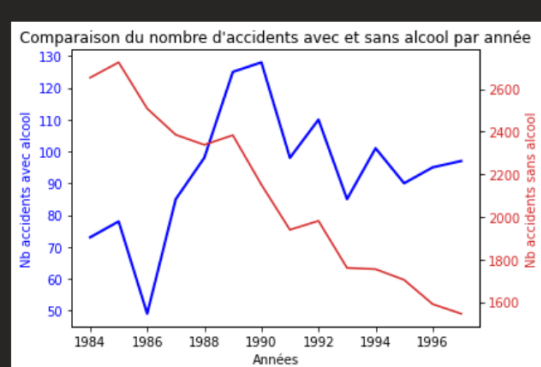
SONT-ILS PLUS GRAVES QUE LES AUTRES TYPES D'ACCIDENTS ?



QUELS TYPE DE VÉHICULE EST LE PLUS FRÉQUENT LORSQU'IL Y A UN ACCIDENT IMPLIQUANT DE L'ALCOOL ?



LE NOMBRE D'ACCIDENT IMPLIQUANT L'ALCOOL A-T-IL AUGMENTÉ AU FUR ET À MESURE DES ANNÉES ?



LES ACCIDENTS SONT EN BAISSE, PAS CEUX CONCERNANT L'IVRESSE

**CELUI QUI CONDUIT C'EST CELUI  
QUI NE BOIT PAS**

- Dans un petit paragraphe, vous ferez un bilan des compétences que vous pensez avoir acquises et des qualités que vous pensez avoir mis en œuvre dans cette première partie de la SAE. Même si celles-ci varient d'un étudiant à l'autre, il est important que vous en discutiez au sein de votre groupe et que vous réfléchissiez aussi aux "traces" permettant de les justifier.

Dans cette partie, nous avons appris à :

- Créer un graphique complet sous Python, de manière qu'il puisse être analysable.
  - Manipuler les DataFrame de la bibliothèque Pandas
  - Manipuler les jointures de DataFrame avec le merge
  - Créer des graphiques à deux échelles
- Créer un graphique complet sous Excel
  - Manipuler les données : Mettre en pourcentage en fonction du total, moyenne
  - Manipuler les fonctions : RechercheV et autre

On a su être :

- Efficace dans la production de la partie 2
- Autonome pour créer les graphiques que nous nous étions assignés
- Collaboratif pour réaliser l'affiche A3 grâce à l'application web Canva ou nous avons pu partager notre projet afin de mettre chacun nos graphiques et notre touche de créativité.