# Hélène/Orphée – Codebook

\*Data Analytics 2021\*

**Executive Analysis**

This document describes the dataset used for this project, including general info and variables (data type, units of measurement, etc.) This document is in French.

**1) Notre projet**

Nous souhaitons classifier les accidents de la circulation en fonction du degré de gravité des blessures subies par leurs victimes. Les 4 classes prévues sont les suivantes :

* Personne indemne : 1
* Victime décédée (sur le coup ou dans les trente jours) : 2
* Blessé grave (hospitalisé plus de 24 heures) : 3
* Blessé léger (ayant fait l'objet de soins médicaux mais n'ayant pas été hospitalisé au moins 24 heures) : 4

**2) Jeu de données**

En France, chaque accident de la circulation fait l’objet d’un relevé d’informations par les forces de l’ordre, à condition d’avoir impliqué au moins 1 véhicule et d’avoir fait au moins 1 victime ayant nécessité des soins.

Ce relevé comporte les caractéristiques de l’accident, son lieu, les véhicules impliqués et leurs victimes.

Ces informations alimentent les Bases de données Annuelles des Accidents corporels de la Circulation routière (BAAC).

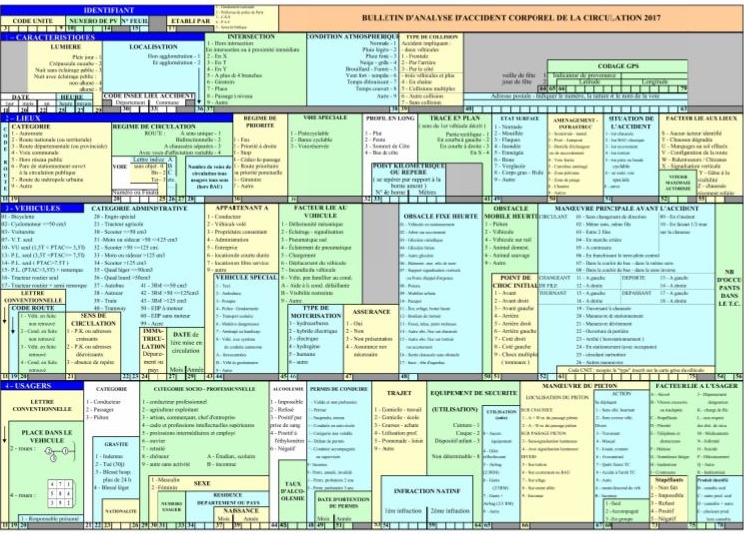
Simplifiées et anonymisées, elles sont mises à disposition du public sur le site [www.data.gouv.fr](http://www.data.gouv.fr) (plateforme internet des données publiques françaises) et sont libres d’utilisation.

Nous étudierons la base de données répertoriant les accidents survenus durant l’année 2019

Elle est composée de 4 fichiers (Caractéristiques – Lieux – Véhicules – Usagers) au format csv, que nous avons fusionnés en un seul fichier ‘victime.csv’.

3) Variables du jeu de données :

X colonnes contentant l'information retenue pour prévoir le degré de gravité des accidents de la route en 2019, en fonction de X indicateurs.



|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Nom de la Variable** | **Description** | **Exemple** | **Unité** | **dtype initial**  **(Objet / int / float)** | **Type de donnée**  **(Catégoriel / Numérique)** | **Influence supposée) sur la variable cible (avant étude)** | **Type de distribution** | **% de valeurs manquantes** | **Outliers éventuels** | **Transformations mathématiques potentiellement intéressantes** | **Feature conservée pour l’étude ?** |
| **Fichier « vict »**  **(victimes de l’accident)** | Num\_Acc. | Numéro d'identifiant de l’accident | 201900000001 | Classement numérique | int / float | C | Néant | Uniforme | 0 |  |  |  |
| Num\_Veh | Identifiant de chaque véhicule impliqué dans un accident spécifique. Repris pour chacun des usagers occupant ce véhicule (y compris les piétons qui sont rattachés aux véhicules qui les ont heurtés) | B01 | Code alphanumérique commun à chaque accident : 1 lettre de l’alphabet + 1 chiffre de 01 à 09 (A01, B01, C01… Z01, A02, B02…) Permet de rattacher aisément chaque victime à l’un des véhicules impliqués dans l’accident en question. | objet | C | Néant | Quasi-binaire | 0 |  |  |  |
| Place | Place occupée dans le véhicule par l'usager au moment de l'accident (voir schéma dans le référentiel officiel) | 2 | Classement numérique | int / float | C | Forte | Pareto | 0 |  |  |  |
| catu | Catégorie de victime (conducteur, passager, piéton) | 2 | Classement numérique | int / float | C | Forte | Pareto | 0 |  |  |  |
| grav | VARIABLE CIBLE  Gravité de blessure de l'usager, les usagers accidentés sont classés en trois catégories de victimes plus les indemnes : 1 – Indemne 2 – Tué 3 – Blessé hospitalisé 4 – Blessé léger | 3 | Classement numérique | int /float | C | CIBLE | Multimodale | 0 |  |  |  |
| sexe | Sexe de la victime | 1 | Classement numérique | int / float | C | Medium | Binaire | 0 |  |  |  |
| an\_nais | Année de naissance de la victime | 1979 | Numéro | int ou float | C | Faible |  | 0 |  |  | NON |
| trajet | Motif du déplacement au moment de l’accident (travail, achats, loisirs…) | 2 | Classement numérique | int ou float | C | Faible | Multimodale | 26.3% |  |  | NON |
| secu1 | Utilisation d’équipement de sécurité (airbag, casque, ceinture…) | 1 | Classement numérique | int ou float | C | Forte | Prédominance d’une valeur | 0,07% |  |  |  |
| secu2 | Utilisation d’équipement de sécurité (airbag, casque, ceinture…) | 1 | Classement numérique | int ou float | C | Forte | Multimodale | 30,2% |  |  | NON |
| secu3 | Utilisation d’équipement de sécurité (airbag, casque, ceinture…) | 1 | Classement numérique | int ou float | C | Forte | Multimodale | 98,7% |  |  | NON |
| locp | Localisation du piéton (passage piéton, chaussée…) | 1 | Classement numérique | int ou float | C | Forte |  | 38% |  |  | NON |
| actp | Action du piéton (traversant, courant…) | 1 | Classement numérique | objet | C | Medium |  | 20,2% |  |  | NON |
| etatp | Précise si le piéton accidenté était seul, accompagné ou en groupe | 3 | Classement numérique | int ou float | C | Faible |  | 91,7% |  |  | NON |
| **Fichier « veh » (véhicule)** | id\_vehicule | Identifiant unique du véhicule. Repris pour chacun des usagers occupant ce véhicule (y compris les piétons qui sont rattachés aux véhicules qui les ont heurtés) | 138 306 524 | Code numérique | objet | C | Néant | ? | 0 |  |  |  |
| senc | Sens de circulation (croissant ou décroissant) | 2 | Classement numérique | int ou float | C | Forte | Multimodal | 0,03% |  |  |  |
| catv | Catégorie du véhicule (voiture, vélo, tracteur…) | 2 | Classement numérique | int ou float | C | Forte | Prédominance d’une valeur | 0 |  |  |  |
| obs | Obstacle fixe heurté (poteau, arbre…) | 3 | Classement numérique | int ou float | C | Forte | Prédominance d’une valeur | 0,03% |  |  |  |
| obsm | Obstacle mobile heurté (piéton, animal…) | 4 | Classement numérique | int ou float | C | Medium | Quasi-binaire | 0,07% |  |  |  |
| choc | Point de choc initial du véhicule (avant, côté…) | 3 | Classement numérique | int ou float | C | Forte | Prédominance d’une valeur | 0,04% |  |  |  |
| manv | Manoeuvre principale avant l’accident (dépassement, marche arrière…) | 5 | Classement numérique | int ou float | C | Medium | Prédominance d’une valeur | 0,03% |  |  |  |
| motor | Type de motorisation du véhicule (hydrocarbures, hybride…) | 1 | Classement numérique | int ou float | C | Faible | Prédominance d’une valeur | 0,3% |  |  | NON |
| occutc | Nombre d’occupants dans le transport en commun | 21 | Nombre d’occupants | int ou float | Numérique | Néant |  | 98,6% |  |  | NON |
| **Fichier « carac »**  **(caractéristiques de l’accident)** | jour | Jour de l’accident | 05 | Numéro du jour | int ou float | C | Néant |  | 0 |  |  | NON |
| mois | Mois de l’accident | 12 | Numéro de mois | int ou float | C | Medium | Uniforme | 0 |  |  |  |
| an | Année de l’accident | 2019 | Numéro de l’année | int ou float | C | Néant | 1 valeur unique | 0 |  |  | NON |
| hrmn | Heure et minutes de l'accident | 17:21 |  | objet | C | Medium |  | 0 |  |  | NON |
| lum | Conditions d’éclairage dans lesquelles l'accident s'est produit (plein jour, aube, nuit…) | 1 | 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 | int ou float | C | Forte | Prédominance d’une valeur | 0 |  |  |  |
| dep | Code INSEE du département | 59 | Code numérique | objet | C | Faible | Multimodal | 1,1% |  |  |  |
| com | Code INSEE de la commune | 59220 | Code numérique | objet | C | Faible |  | 0 |  |  | NON |
| agg | Localisation en agglomération ou hors agglomération | 1 | Classement numérique | int ou float | C | Medium | Binaire | 0 |  |  |  |
| int | Intersection | 2 | Classement numérique. | int ou float | C | Medium | Prédominance d’une valeur | 0 |  |  |  |
| atm | Conditions atmosphériques (vent, brouillard, pluie…) | 3 | Classement numérique | int ou float | C | Forte | Prédominance d’une valeur | 0 |  |  |  |
| col | Type de collision (frontale, latérale…) | 5 | Classement numérique | int ou float | C | Forte | Multimodal | 0 |  |  |  |
| adr | Adresse postale si accident survenu en agglomération | Autoroute A86, rue de Tolbiac… | Nom | objet | C | Faible |  | 0,8% |  |  | NON |
| lat | Latitude | 48,8962100 | Nombre décimal | objet | C | Faible |  | 0 |  |  |  |
| long | Longitude | 2,4701200 | Nombre décimal | objet | C | Faible |  | 0 |  |  |  |
| **Fichier « lieu »** | catr | Catégorie de route (autoroute, route d’agglo, parking…) | 2 | Classement numérique | int ou float | C | Medium | Multimodal | 0 |  |  |  |
| voie | Numéro de la route | 86 | Numéro | objet | C | Néant |  | 4,5% |  |  | NON |
| v1 | Indice numérique du numéro de route (exemple : 2 bis, 3 ter etc.) | 0.0 | Indice numérique | int ou float | C | Néant |  | 18,5% |  |  | NON |
| v2 | Lettre indice alphanumérique de la route | A | Lettre | objet | C | Néant |  | 92,7% |  |  | NON |
| circ | Régime de circulation (à sens unique, bidirectionnelle, à chaussées séparées…) | 3 | Classement numérique. | int ou float | C | Fort | Prédominance d’une valeur | 5,1% |  |  |  |
| nbv | Nombre total de voies de circulation | 10 | Nombre | int ou float | Numérique | Medium | Prédominance d’une valeur | 1% |  |  |  |
| vosp | Existence d’une voie réservée aux cycles, indépendamment du fait que l’accident ait lieu ou non sur cette voie | 0 | Classement numérique | int ou float | C | Faible | Prédominance d’une valeur | 1,1% |  |  |  |
| prof | Déclivité de la route à l'endroit de l'accident | 2 | Classement numérique | int ou float | C | Medium | Prédominance d’une valeur | 0,02% |  |  |  |
| pr | Numéro du point de repère de rattachement (numéro de la borne amont) | 303 | Numéro | objet | C | Néant |  | 0 |  |  | NON |
| pr1 | Distance à la borne amont | 282 | Mètres | objet | Numérique | Néant |  | 0 |  |  | NON |
| plan | Tracé en courbe du lieu | 3 | Classement numérique | int | C | Néant | Prédominance d’une valeur | 0 |  |  | NON |
| lartpc | Largeur du terre-plein central s’il existe (en m) | 50 | |  | | --- | |  | | int ou float | Numérique | Néant |  | 99,6% |  |  | NON |
| larrout | Largeur de la chaussée | 125 | Mètres | int ou float | Numérique | Medium |  | 99,4% |  |  | NON |
| surf | Etat de la surface (flaque, neige…) | 1 | Classement numérique | int ou float | C | Forte | Prédominance d’une valeur | 0,03% |  |  |  |
| infra | Infrastructure éventuelle (pont, voie ferrée…) | 1 | Classement numérique | int ou float | C | Forte | Prédominance d’une valeur | 0,1% |  |  |  |
| situ | Situation de l’accident (sur chaussée, sur trottoir…) | 2 | Classement numérique | int ou float | C | Forte | Prédominance d’une valeur | 0,2% |  |  |  |
| vma | Vitesse maximale autorisée sur le lieu et au moment de l’accident | 120 | Nombre | int ou float | C | Forte | Prédominance d’une valeur | 1,4% | OUI : winsorisées |  |  |