



Rapport du projet KITE

BIG DATA ET FLUX ROUTIERS: ACCIDENTALITÉ

Auteurs :

Cyrine ABDESSAMAD
Emna KAMOUN
Maxence LECANU
Mohamed Ali ABDELWAHED
Sameh RAHMOUNI

Encadrants :

CNE Cédric ROGER
Mme Cécile FAUCONNET
Mme Célia ZYLA
M. François PESSAUX
LCL Michael FUMERY
CNE Marjory CANONNE

Année universitaire 2017-2018

Table des matières

1	Présentation du projet	5
2	Travail élaboré	5
2.1	Bases de données utilisées	5
2.1.1	Accidentalité	5
2.1.2	Météo	6
2.1.3	Trafic	6
2.2	Modélisation	7
2.2.1	Modèle explicatif	8
2.2.2	Modèles prédictifs	9
2.3	Analyse descriptive des données	14
2.4	Prise en compte des phénomènes météorologiques	23
2.4.1	Température	24
2.4.2	Humidité	25
2.4.3	Gravité des accidents par rapport à la météo	27
2.5	Effets des événements calendaires et culturels	29
2.5.1	Jours fériés	29
2.5.2	Vacances scolaires	29
2.5.3	Changement d'heure	30
2.6	Résultats des modèles	30
2.6.1	Modèle explicatif	30
2.6.2	Modèles prédictifs	39
2.7	Cartographie des accidents	42
3	Perspectives	42
4	Difficultés rencontrées	42
	Conclusion	44

Table des figures

1	Exemple démonstratif du principe des arbres de décision . . .	10
2	Exemple d'une classification SVM multinomiale	12
3	Exemple d'une classification KNN binomiale avec $k=3$	13
4	Exemple d'une classification KNN binomiale avec $k=3$ (suite)	14
5	Evolution du nombre d'accidents par année	15
6	Comparaison entre les évolutions du nombre d'accidents pour chaque niveau de gravité par année	16
7	Nombre d'accidents par mois en 2009	17
8	Comparaison des nombres d'accidents moyens par jour entre février, janvier et décembre	17
9	Répartition des accidents par jour de la semaine	18
10	Nombres d'accidents en agglomération/hors agglomération par catégorie de l'usager	19
11	Catégorie de l'usager par gravité de l'accident	19
12	Gravité des accidents par sexe	20
13	Évolution du nombre de conducteurs impliqués dans un acci- dent par sexe	21
14	Répartition des accidents par tranche d'âge	21
15	Nombre d'accidents par type de trajet	22
16	Evolution du nombre d'accidents avec la température	24
17	Evolution de la fréquence des accidents avec la température . .	24
18	Evolution de la fréquence des accidents avec l'humidité	25
19	Evolution du nombre d'accidents avec l'humidité	26
20	Graphes pour les victimes indemnes	27
21	Graphes pour les victimes légèrement blessées	27
22	Graphes pour les victimes hospitalisées	28
23	Graphes pour les victimes tuées	28
24	Résultats des prédictions dans le département 75	39
25	Résultats des prédictions dans le département 45	40

26	Résultats des prédictions dans le département 59	40
27	Résultats des prédictions dans le département 976	41

Remerciements

Nous tenons à remercier vivement le CNE Cédric ROGER, le LCL Michael FUMERY et la CNE Marjory CANONNE pour cette opportunité, pour la confiance qu'ils nous ont accordée et pour leur soutien tout au long du projet.

Nous adressons également nos remerciements à Mme Cécile FAUCONNET, doctorante à l'ENSTA ParisTech, Mme Celia ZYLA, coordinatrice de la formation KITE, et M. François PESSAUX, enseignant-chercheur à l'Unité d'Informatique et d'Ingénierie des Systèmes, pour la qualité de leur encadrement, leurs conseils précieux, leur encouragement et leur aide inconditionnelle.

Finalement, nous remercions toute personne ayant aidé, de près ou de loin, à la concrétisation de nos travaux tout au long du projet.

1 Présentation du projet

Le projet s'intitule **Big Data et flux routiers**. Le sujet est vaste et peut être traité sous différents angles. Une première étape consiste à faire un point sur les travaux déjà engagés sur ce sujet et à réaliser un benchmark international sur cette thématique. Suite à ces recherches, une discussion avec les encadrants du projet a eu lieu. Divers sujets ont été abordés :

- congestion : apprendre, à partir des bases de données, les endroits et les heures qui présentent un problème de congestion et gérer la circulation.
- accidentalité : étudier le taux d'accidentalité à des endroits particuliers (stades, points de vente d'alcool) et à des événements spéciaux (noël, réveillon, ...) et prévoir des mesures préventives.
- se concentrer sur une zone, modéliser les flux routiers et les prévoir à long terme. Ceci peut être exploité dans les projets d'aménagement des villes (amélioration du plan routier), de réduction de pollution, etc.

En réalisant un point sur les données disponibles en France, nous nous sommes fixés le thème de l'accidentalité, la problématique étant la suivante :

- quelles sont les causes qui expliquent l'accidentalité en France ?
- comment prédire l'accidentalité en France ?

Le but de ce projet est d'étudier les possibilités que peuvent offrir les nouvelles technologies, et la manière dont elles peuvent augmenter l'efficacité des missions de la gendarmerie

Ce projet présente un intérêt particulier pour la mission numérique de la gendarmerie nationale et la sous-direction de la sécurité publique et de la sécurité routière.

2 Travail élaboré

2.1 Bases de données utilisées

2.1.1 Accidentalité

Cette base de données, extraite du fichier BAAC¹, regroupe les accidents corporels de la circulation entre 2005 et 2016. Elle contient des informations récoltées par l'unité des forces de l'ordre et publiée sur le site *data.gouv.fr*.

1. Bulletin d'Analyse d'Accident Corporel de la circulation

Dans cette base de données, on trouve quatre rubriques précisées dans le tableau 1 :

Rubrique	Variables retenues
Caractéristiques	date, condition d'éclairage, localisation, agglomération/ hors agglomération, conditions atmosphériques
Usagers	catégorie, gravité, sexe, année de naissance, trajet
Lieux	catégorie, profil et état de la route, infrastructure
Véhicules	catégorie, obstacle fixe/mobile heurté

TABLE 1 – Catégories et variables retenues de la base de données accidentalité

Une liste détaillée des variables et leurs significations se trouve dans l'annexe.

2.1.2 Météo

Cette base de données, trouvée sur *donneespubliques.meteofrance.fr*, contient les conditions météorologiques de 2005 à 2016 par jour. Une description détaillée des variables de la base peut être trouvée sur le lien suivant : https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=90&id_rubrique=32.

On a retenu 4 variables de cette base de données qui sont :

- t : Température en Kelvin
- u : Humidité en pourcentage
- rr3 : Précipitations dans les 3 dernières heures en millimètre
- ssfrai : Hauteur de la neige fraîche en mètre

2.1.3 Trafic

Cette base de données contient les données trafic du département du Loiret entre 2005 et 2016. Elle provient du lien suivant : <http://open.isogeo.com/s/bf2c4a53b8e34b448ff9c3f40d198c1f/g0nJ08Ck0y2RMdnLdjuEH6Nu6UeU0> Pour le travail effectué, on a utilisé la variable suivante :

- MJA : Moyenne journalière annuelle du trafic

2.2 Modélisation

Au départ, nous avions l'idée de prédire si, sous des conditions précisées, il va y avoir un accident ou pas. On considèrerait donc une variable *accident* qui prend 0 si ce n'est pas un accident et 1 sinon. Cette approche n'est pas faisable avec les données dont nous disposons (base BAAC) puisqu'on ne dispose que des cas là où il y a eu un accident, ce sont les 1 de la variable *accident* à prédire.

Comme les accidents décrits dans cette base admettent forcément au moins une victime blessée ou tuée, nous avons pensé qu'en ayant les accidents où il n'y a eu aucune victime (toutes les victimes sont indemnes), nous pourrions considérer ces cas comme étant les 0 de la variable *accident*, c'est-à-dire il n'y a pas d'accident. Pour cela, nous avons tenté d'avoir les données des assurances, où on trouve tous types d'accidents. Ceci n'était pas possible car ce sont des données confidentielles.

Par conséquent, une autre approche a été adoptée. Etant donnée qu'un accident est décrit essentiellement par son intensité de gravité, une alternative consiste donc à prédire la gravité d'un accident, celle-ci est décrite par l'état des victimes. La variable que nous allons expliquer et puis prédire prend 4 valeurs :

$$\text{Gravité d'un accident} = \begin{cases} 0 & \text{si la victime est indemne} \\ 1 & \text{si elle blessée légèrement} \\ 2 & \text{si elle est hospitalisée} \\ 3 & \text{si elle est tuée} \end{cases}$$

C'est la variable dépendante. Comme on peut le voir, les valeurs qu'elle prend sont ordonnées par degré d'agressivité.

Les données que nous utilisons pour prédire la gravité d'un accident sont principalement celles de la rubrique **usagers** auxquelles nous rajoutons des variables des autres rubriques (**caractéristiques** et **véhicules**) et de la base météo. Ce qu'on rajoute doit être mis au bon format. En effet, une observation (une ligne) de la rubrique **usagers** correspond à une victime d'un accident identifié par un numéro unique. Un accident à 4 victimes est décrit en 4 lignes, une victime par ligne, accompagnée par le numéro de l'accident. Toute nouvelle donnée intégrée à cette base (voir annexe) a été mise sous cette forme, dupliquée pour toutes les victimes d'un même accident.

Afin d'expliquer la variable *Gravité d'un accident*, nous adoptons un modèle probabiliste.

2.2.1 Modèle explicatif

Revenons sur l'écriture mathématique de la variable à expliquer *Gravité d'un accident* qu'on note dans cette partie Y . Y_i prend les valeurs 0,1,2 ou 3 selon l'état de la victime i . Il existe une variable latente Y_i^* (transformée de Y_i) qui se décompose en éléments déterministes et aléatoire. Nous notons X_i le vecteur des variables indépendantes qui résument les caractéristiques de la victime et des conditions de l'accident associé et β_j le vecteur des paramètres associés. Le modèle prend la forme :

$$Y_i^* = \beta' X_i + \epsilon_i$$

où ϵ_i sont les résidus de ce modèle posé.

On a donc :

$$Y_i = \begin{cases} 0 & Y_i^* \leq \alpha_1 \\ 1 & \alpha_1 \leq Y_i^* \leq \alpha_2 \\ 2 & \alpha_2 \leq Y_i^* \leq \alpha_3 \\ 3 & \alpha_3 \leq Y_i^* \end{cases}$$

Ce qui permet d'avoir :

$$\begin{cases} \mathbb{P}(Y_i = 0) &= \mathbb{P}(\epsilon_i \leq \alpha_1 - \beta' X_i) \\ \mathbb{P}(Y_i = 1) &= \mathbb{P}(\alpha_1 - \beta' X_i \leq \epsilon_i \leq \alpha_2 - \beta' X_i) \\ \mathbb{P}(Y_i = 2) &= \mathbb{P}(\alpha_2 - \beta' X_i \leq \epsilon_i \leq \alpha_3 - \beta' X_i) \\ \mathbb{P}(Y_i = 3) &= \mathbb{P}(\alpha_3 - \beta' X_i \leq \epsilon_i) \end{cases}$$

Si $F(\cdot)$ définit la fonction de répartition de ϵ :

$$\begin{cases} \mathbb{P}(Y_i = 0) &= F(\alpha_1 - \beta' X_i) \\ \mathbb{P}(Y_i = 1) &= F(\alpha_1 - \beta' X_i) - F(\alpha_2 - \beta' X_i) \\ \mathbb{P}(Y_i = 2) &= F(\alpha_2 - \beta' X_i) - F(\alpha_3 - \beta' X_i) \\ \mathbb{P}(Y_i = 3) &= F(\alpha_3 - \beta' X_i) \end{cases}$$

Le choix d'une distribution normale conduit à un modèle probit ordonné, alors qu'en choisissant une distribution logistique nous serons amenés à considérer un modèle logit ordonné. Les données dont nous disposons sont positives, il est donc judicieux de faire l'hypothèse de gaussianité. Par conséquent, nous privilégions dans ce qui suit le probit ordonné.

De plus, il faut tenir en compte que les victimes appartiennent à un même accident. Ceci est fait grâce à un modèle à effet fixe selon le numéro de l'accident.

Somme toute, le modèle adopté est le probit ordonné à effet fixe.

En enchaînant une analyse descriptive des données, nous avons identifié certains facteurs influents sur l'intensité de la gravité d'un accident. Nous les intégrons dans le modèle de probabilité adopté. Ce sont :

- weekend/reste de la semaine
- luminosité
- pluviométrie
- humidité
- agglomération/hors agglomération
- catégorie de la victime : piéton/conducteur
- âge de la personne impliquée
- son sexe
- type du trajet effectué
- catégorie de véhicules

Avec un tel modèle, le but principal est d'expliquer le phénomène. Il estime un coefficient pour chaque variable explicative. La valeur numérique des paramètres estimés n'a pas vraiment d'intérêt en soi. C'est ainsi que les coefficients des variables explicatives ne sont pas directement interprétables : la seule information utilisable est le signe des paramètres dans la mesure où il indique qu'une augmentation de la variable explicative associée se traduit par une réduction de l'événement ou par une augmentation de la probabilité qu'un accident soit grave.

D'autres modèles sont mis en place pour la prédiction.

2.2.2 Modèles prédictifs

Dans cette partie, nous allons prédire la gravité des accidents dans différents départements. Pour ce faire, nous nous appuyons sur plusieurs modèles d'apprentissage automatique supervisé.

Étant donné que la variable cible prend 4 valeurs, les algorithmes de classification multinomiale ou en classes multiples sont les plus adaptés à notre problématique.

Les variables prédictives utilisées par tous les modèles sont :

- *mois* : mois de l'accident
- *hrmn* : heure et minutes de l'accident
- *long, lat* : coordonnées géographiques de l'accident
- *lum* : conditions d'éclairage dans lesquelles l'accident s'est produit
- *atm* : conditions atmosphériques
- *agg* : hors ou en agglomération

- *sexe* : sexe de l'utilisateur
- *age* : âge de l'utilisateur
- *catu* : catégorie de l'utilisateur : conducteur, passager, piéton, piéton en roller ou en trottinette
- *trajet* : motif du déplacement au moment de l'accident

Modèle 1 : Arbre de décision

Le classificateur d'arbre de décision est une approche systématique pour la classification multi-classe. Il pose un ensemble de questions à l'ensemble de données (liées à ses attributs / caractéristiques). L'algorithme de classification de l'arbre de décision peut être visualisé sur un arbre binaire. Sur la racine et chacun des nœuds internes, une question est posée et les données sur ce nœud sont divisées en enregistrements séparés qui ont des caractéristiques différentes. Les feuilles de l'arbre se réfèrent aux classes dans lesquelles l'ensemble de données est divisé.

Ce modèle a l'avantage d'être lisible et rapide à exécuter.

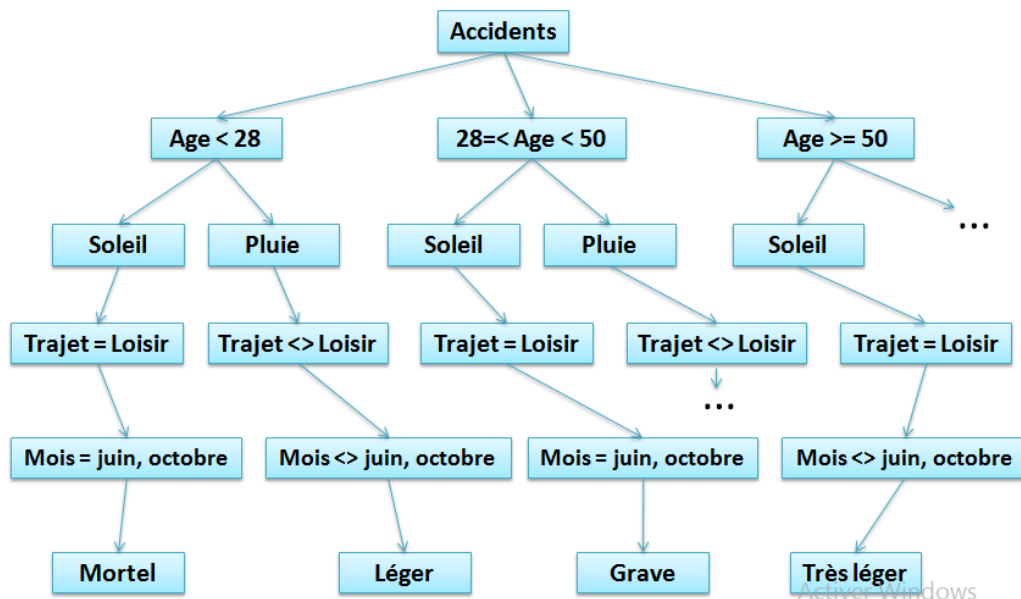


FIGURE 1 – Exemple démonstratif du principe des arbres de décision

Modèle 2 : Régression logistique multinomiale

En statistiques, la régression logistique multinomiale est une méthode de classification qui généralise la régression logistique aux problèmes multi-

classes, c'est-à-dire avec plus de deux résultats discrets possibles. Autrement dit, c'est un modèle qui est utilisé pour prédire les probabilités des différents résultats possibles d'une variable dépendante distribuée catégoriquement, étant donné un ensemble de variables indépendantes (qui peuvent être réelles, binaires, catégorielles, etc.).

La forme générale de la fonction prédictive F est :

$$F(X) = \alpha x_1 + \beta x_2 + \gamma x_3 + \dots + \omega x_n + \epsilon$$

- ϵ : est une constante
- $\alpha, \beta, \gamma \dots$: représente les coefficients de la fonction prédictive F(X)
- X : est un vecteur/tableau de variables prédictives
- x_i : représente la ième variable prédictive

Modèle 3 : Classificateur SVM

Les machines à vecteurs de support ou séparateurs à vaste sont un ensemble de techniques d'apprentissage supervisé destinées à résoudre des problèmes de discrimination et de régression.

C'est une méthode de classification efficace lorsque le vecteur de caractéristiques est de grande dimension.

La résolution des problèmes de discrimination, c'est à dire décider à quelle classe appartient un échantillon, passe par la construction d'une fonction h qui à un vecteur d'entrée x fait correspondre une sortie y : $y=h(x)$

Le but de l'algorithme est d'apprendre la fonction h(x) par le biais d'un ensemble d'apprentissage et de trouver la ou les frontières de décision (appelées aussi hyperplans séparateurs) optimales.

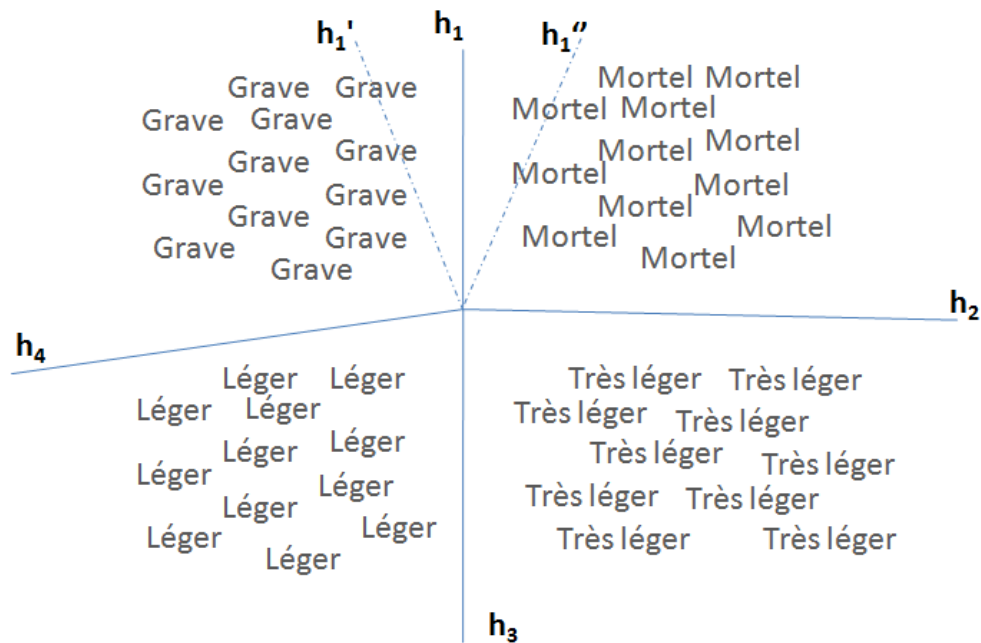


FIGURE 2 – Exemple d’une classification SVM multinomiale

Modèle 4 : Classificateur Naïve Bayes

La classification naïve bayésienne est un type de classification Bayésienne probabiliste simple basée sur le théorème de Bayes avec une forte indépendance (dite naïve) des hypothèses. Elle met en œuvre un classifieur bayésien naïf, ou classifieur naïf de Bayes, appartenant à la famille des classifieurs linéaires.

En termes simples, un classifieur bayésien naïf suppose que l’existence d’une caractéristique pour une classe, est indépendante de l’existence d’autres caractéristiques. Un fruit peut être considéré comme une pomme s’il est rouge, arrondi, et fait une dizaine de centimètres. Même si ces caractéristiques sont liées dans la réalité, un classifieur bayésien naïf déterminera que le fruit est une pomme en considérant indépendamment ces caractéristiques de couleur, de forme et de taille.

Malgré leur modèle de conception « naïf » et ses hypothèses de base extrêmement simplistes, les classifieurs bayésiens naïfs ont fait preuve d’une efficacité plus que suffisante dans beaucoup de situations réelles complexes.

L’avantage du classifieur bayésien naïf est qu’il requiert relativement peu de données d’entraînement pour estimer les paramètres nécessaires à la classification.

Modèle 5 : Classificateur KNN

En intelligence artificielle, la méthode des k plus proches voisins, en anglais 'k-nearest neighbors', est une méthode d'apprentissage supervisé utilisée dans la régression et dans la classification

Dans ce cadre, on dispose d'une base de données d'apprentissage constituée de N couples « entrée-sortie ». Pour estimer la sortie associée à une nouvelle entrée x , la méthode des k plus proches voisins consiste à prendre en compte (de façon identique) les k échantillons d'apprentissage dont l'entrée est la plus proche de la nouvelle entrée x , selon une distance à définir.

Un objet est classé par un vote majoritaire de ses voisins, l'objet étant assigné à la classe la plus commune parmi ses k plus proches voisins (k est un entier positif, typiquement petit). Par exemple, si $k = 1$, alors l'objet est simplement assigné à la classe de ce voisin le plus proche.

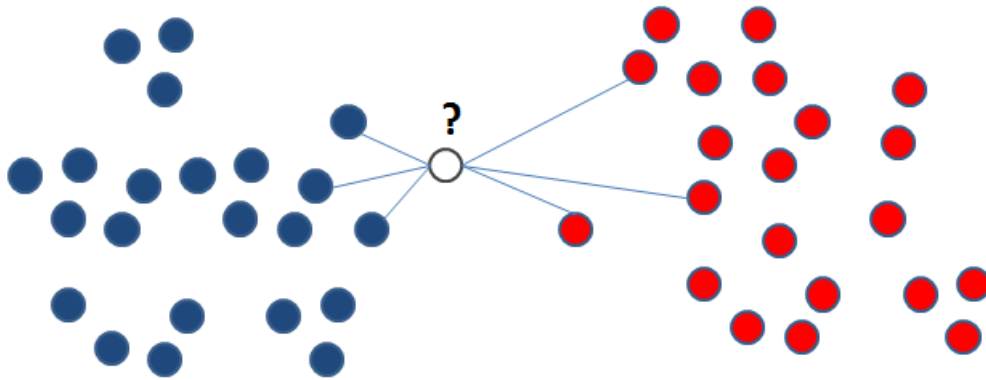


FIGURE 3 – Exemple d'une classification KNN binomiale avec $k=3$

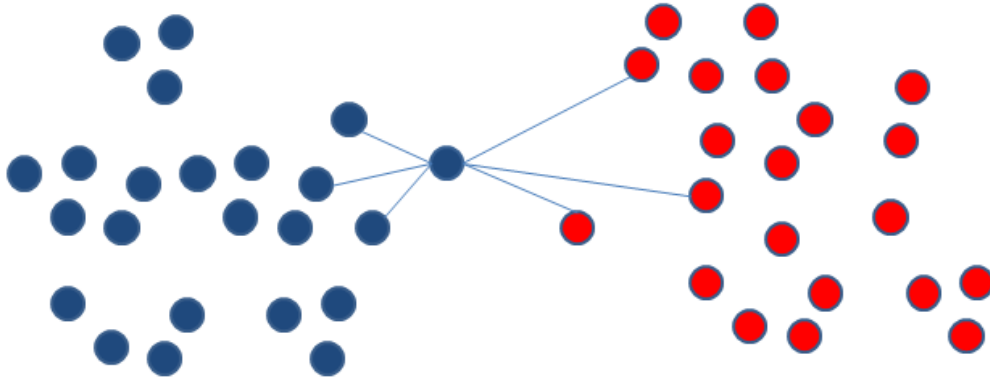


FIGURE 4 – Exemple d’une classification KNN binomiale avec $k=3$ (suite)

2.3 Analyse descriptive des données

Cette partie est consacrée à l’analyse descriptive des données afin de voir quels facteurs sont susceptibles d’influencer le nombre d’accidents ainsi que leurs gravités.

Cette analyse a été faite en premier lieu pour la France métropolitaine puis pour le département 45. Les courbes et les résultats énoncés dans cette parties sont ceux de la France métropolitaine cependant ce sont presque les mêmes résultats obtenus pour le département 45. Il n’y a que la grandeur des chiffres qui diffère.

Avant toute chose, nous avons commencé par étudier l’évolution du nombre d’accidents au fil des années. Bien que ce nombre a globalement diminué entre les années 2005 et 2016, on remarque tout de même qu’entre les années 2013 et 2016, il y a une certaine augmentation. La courbe ci-dessous, illustre ce résultat.

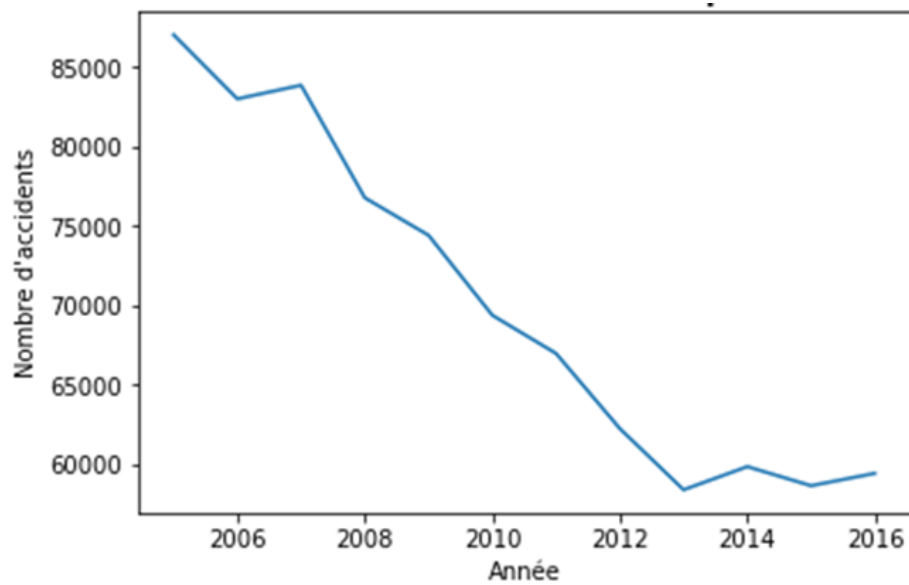


FIGURE 5 – Evolution du nombre d'accidents par année

Pour analyser davantage ce résultat, nous avons étudié l'évolution du nombre d'accidents pour chaque niveau de gravité et ce en divisant les victimes en groupes selon la gravité de leurs états (tué, hospitalisé, blessé léger, indemne). On remarque qu'on a presque la même évolution pour toutes les gravités par contre le nombre de victimes diffère d'un niveau de gravité à un autre.



FIGURE 6 – Comparaison entre les évolutions du nombre d'accidents pour chaque niveau de gravité par année

Ensuite, nous nous sommes intéressés à l'évolution du nombre d'accidents par mois, par jour de semaine ainsi que par heure. Plusieurs résultats ont été dégagés suite à cette analyse. En effet, en se focalisant sur les mois, on a trouvé que juin et octobre sont les mois les plus dangereux de l'année contrairement à février qui est le mois avec le plus petit nombre d'accidents. La figure 3 montre ce résultat pour l'année 2009. Ce même résultat a été trouvé presque pour toutes les années entre 2005 et 2016 en ce qui concerne la France métropolitaine. Cependant, pour le département 45, ce n'est pas le cas pour le département 45 où il est présent dans certaines années mais pas dans d'autres.

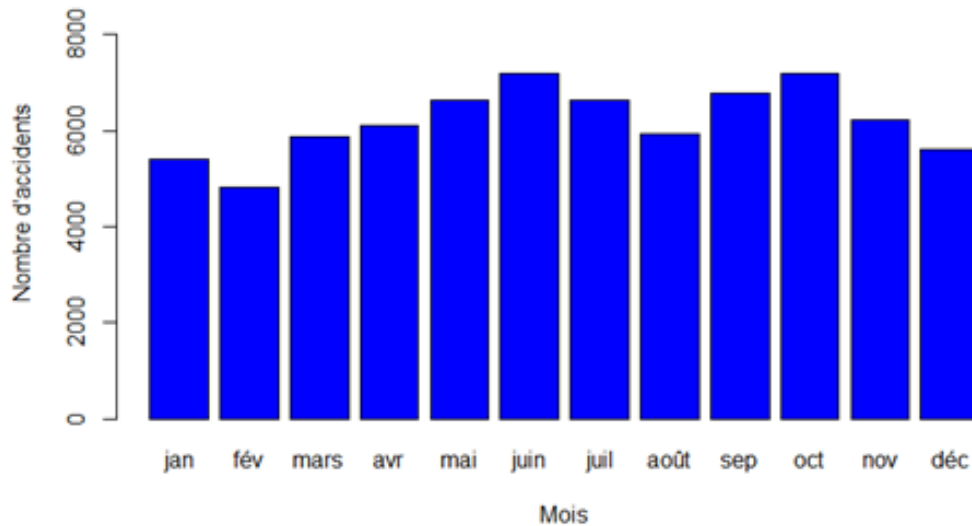


FIGURE 7 – Nombre d'accidents par mois en 2009

Ces résultats peuvent être expliqués par le fait que juin coïncide avec le départ en vacances et octobre coïncide avec les vacances de la Toussaint, considérée le weekend le plus meurtrier. En ce qui concerne février, deux suppositions ont été faites : La première étant que cette chute est due au fait que c'est le mois qui contient de moins de jours. Or la figure ci-dessous démontre que cette hypothèse est fausse :

Mois	Février	Janvier	Décembre
Nombre d'accidents moyen par jour	189.0714	204.0323	239.9032

FIGURE 8 – Comparaison des nombres d'accidents moyens par jour entre février, janvier et décembre

D'après la figure, on voit bien que même la moyenne journalière du nombre d'accidents de février est inférieure aux autres mois. La deuxième hypothèse est que ce résultat vient probablement du fait que février est le mois où il neige le plus. Les résultats du modèle qui seront détaillés plus tard démontreront que cette hypothèse tient debout.

En ce qui concerne la répartition des accidents par jour de semaine, on a

trouvé qu'en moyenne sur l'ensemble des années, dimanche était le jour avec le moins d'accidents contrairement à vendredi qui était celui avec le plus d'accidents. Cependant, les pourcentages d'accidents par jour de semaine sont assez proches les uns aux autres comme le montre la figure suivante :

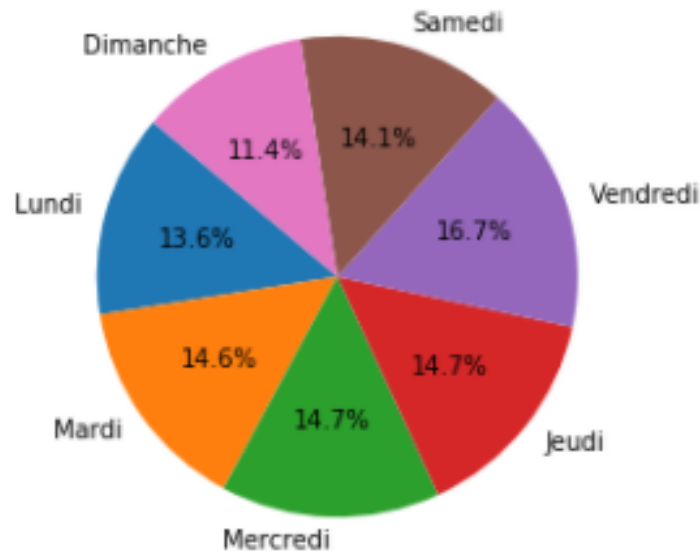


FIGURE 9 – Répartition des accidents par jour de la semaine

La deuxième partie de cette analyse descriptive est relative aux caractéristiques des usagers, à noter leurs catégories, leurs sexes ainsi que leurs âges.

Nous avons commencé une comparaison entre les nombres d'accidents en agglomération et hors agglomération par catégorie de l'utilisateur. Le but de cette analyse est de vérifier une hypothèse qu'on avait faite suite à la comparaison du nombre d'accidents en agglomération et hors agglomération. En effet, on a trouvé que le nombre d'accidents dans les agglomérations sont plus que le double de ceux en dehors des agglomérations. L'hypothèse qu'on avait faite est que ceci est dû aux piétons qui sont plus présents aux agglomérations. Mais en analysant plus nos données comme le montre la figure 6, nous avons trouvé que cette différence est dû essentiellement à la catégorie conducteur. Donc il y a beaucoup plus de risque pour un conducteur de heurter une autre voiture ou un obstacle en agglomération.

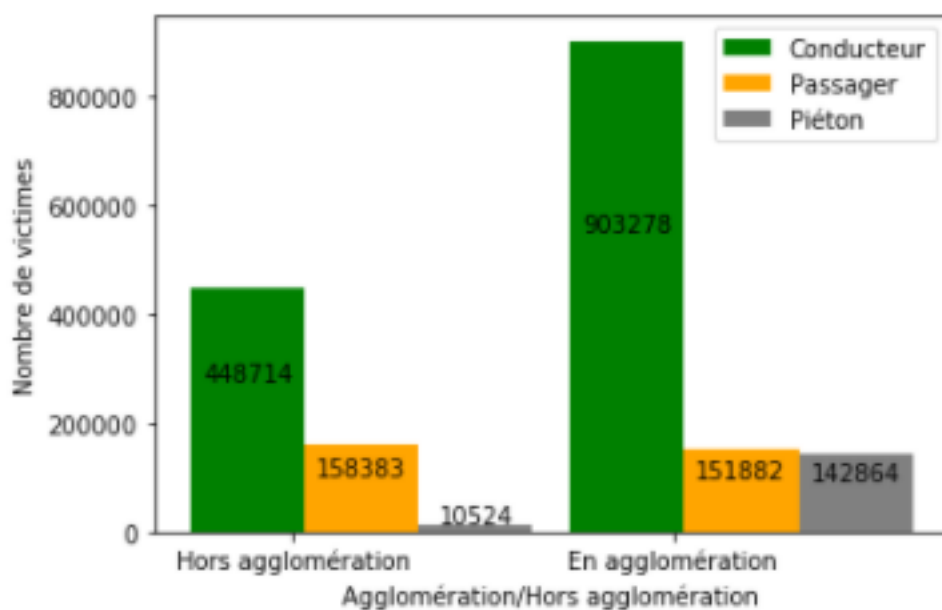


FIGURE 10 – Nombres d'accidents en agglomération/hors agglomération par catégorie de l'utilisateur

Ensuite, on s'est concentré sur la gravité des accidents par catégorie de l'utilisateur ainsi que par sexe. Les deux figures qui suivent illustrent les résultats obtenus :

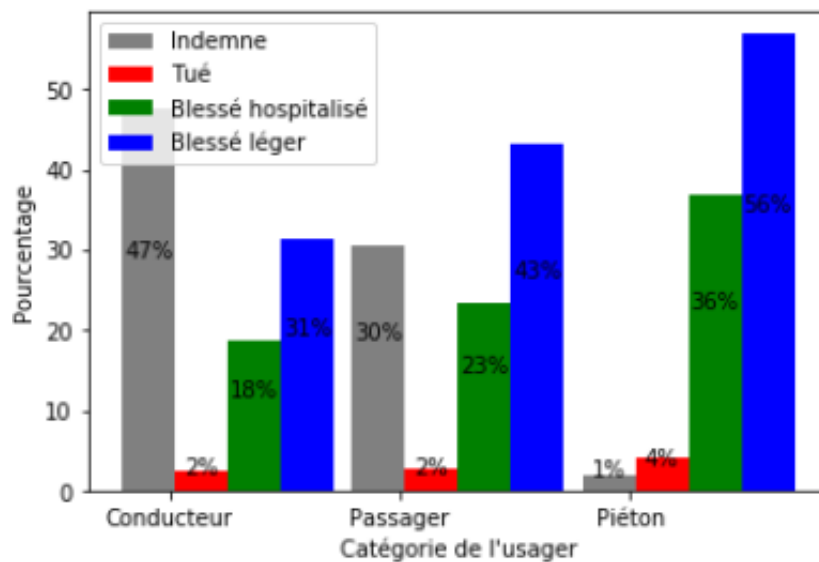


FIGURE 11 – Catégorie de l'utilisateur par gravité de l'accident

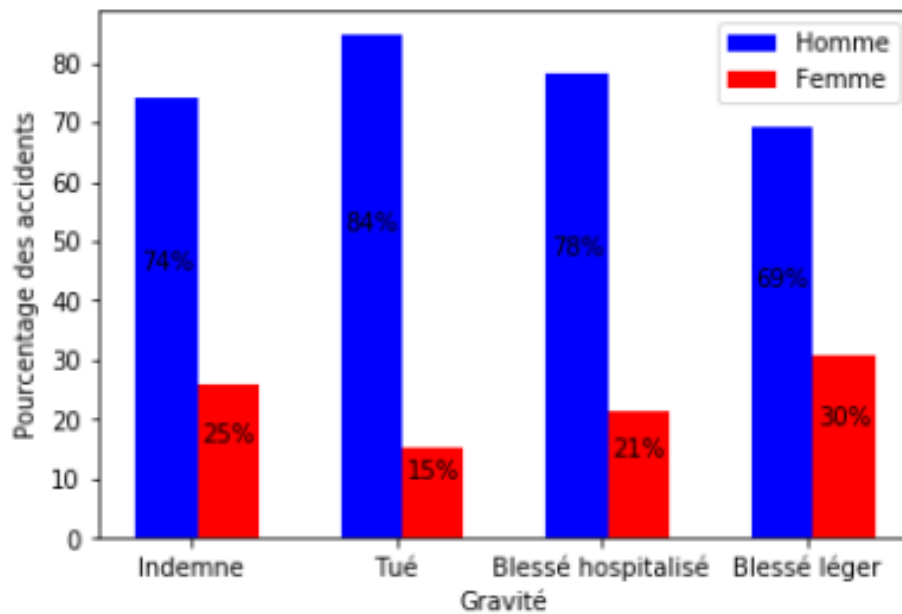


FIGURE 12 – Gravité des accidents par sexe

A priori, on pourrait dire que le sexe et la catégorie de l'utilisateur ont de l'influence sur la gravité de l'accident. Il faudrait voir les résultats du modèle fait, pour pouvoir affirmer ou infirmer cette hypothèse.

Les autres analyses faites dans cette partie sont l'évolution du nombre de conducteurs impliqués dans un accident par sexe et la répartition des accidents par tranche d'âge. On obtient les résultats suivants :

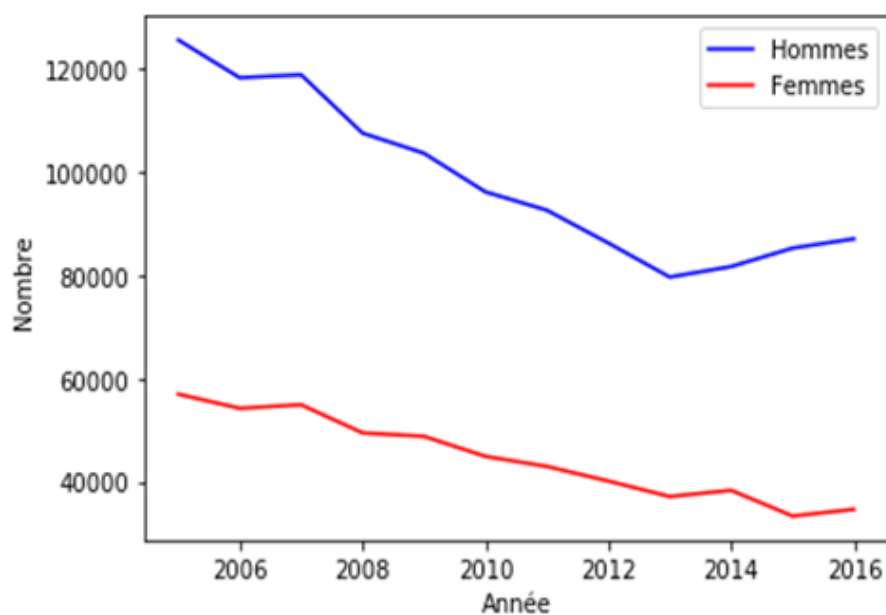


FIGURE 13 – Évolution du nombre de conducteurs impliqués dans un accident par sexe

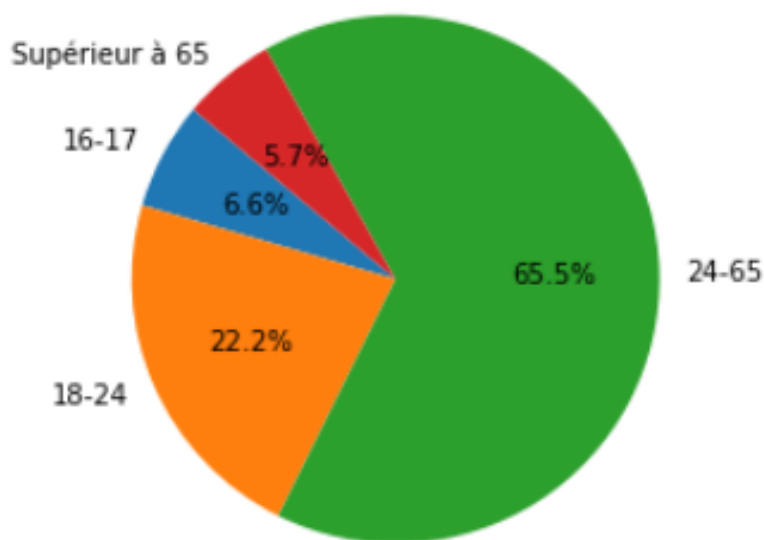


FIGURE 14 – Répartition des accidents par tranche d'âge

Cependant, les résultats de ces graphes-là ne sont pas vraiment significatifs. Ils faudrait disposer des nombres de conducteurs hommes et femmes ainsi

que des âges des différents conducteurs pour pouvoir tirer des conclusions de cette analyse.

De plus, nous avons étudié le nombre d'accidents par type de trajets. On a trouvé que la plupart des accidents sont commis dans des trajets de loisirs comme le montre la figure suivante :

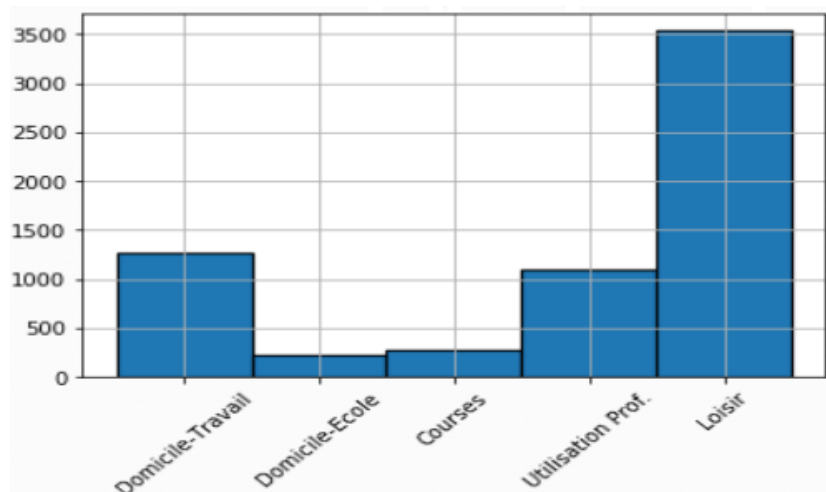


FIGURE 15 – Nombre d'accidents par type de trajet

Pour finir, nous avons regardé le pourcentage d'accidents par type de véhicule. Le tableau qui suit montre que les véhicules légers sont les plus impliqués dans des accidents. C'est logique puisqu'ils sont les plus présents sur la route. Toutefois, ce ne sont pas eux qui commettent les accidents les plus graves. On montre que presque 24% des accidents impliquant des poids lourds ou des moyens de transport en commun sont mortels.

	Tous les départements	Loiret	
Catégories	Pourcentage des accidents impliquant au moins un véhicule de type	Pourcentage des accidents impliquant au moins un véhicule de type	Pourcentage des accidents mortels parmi les accidents impliquant au moins un véhicule de
Vélos	6,85%	6,91%	5,19%
Cyclos et assimilés	17.07%	16,22%	3%
VL et assimilés	85.84%	86,89%	9,90%
PL et TC	6.29%	7,60%	23,87%
2RM et assimilés	16.68%	14,18%	11,40%
Autres : tracteur, tram, train ..	1.56%	1,68%	21.27%

TABLE 2 – Pourcentage des accidents par catégorie de véhicule

2.4 Prise en compte des phénomènes météorologiques

Afin d'établir la relation entre les accidents et les conditions météorologiques, nous avons utilisé les données de *Météo France* présentées dans la section 2.1.2. La fusion de ces données avec la base accidentaire prend énormément de temps, ce qui fait que nous avons fait ceci uniquement pour le Loiret.

Étant donné que la fréquence d'apparition des différents types de conditions météorologiques peut varier énormément il a été choisi de raisonner avec des grandeurs intensives. Il y a beaucoup plus de jours de pluie, ou sans pluie, que de jours où il neige par exemple, et donc on trouvera naturellement plus d'accidents lorsqu'il pleut que lorsqu'il neige.

Ainsi, pour un type de condition météo, on a représenté le nombre d'accident par jour où la météo correspond à cette condition. Le but ici est d'avoir une idée des conditions météorologiques favorisant particulièrement l'accidentalité, ou alors accentuant la gravité de ces accidents.

2.4.1 Température

La première donnée prise en compte est la température. On peut s'attendre à voir une hausse du nombre d'accidents lors des périodes de grand froid, à cause de phénomènes comme le verglas ou la neige qui y sont liés.

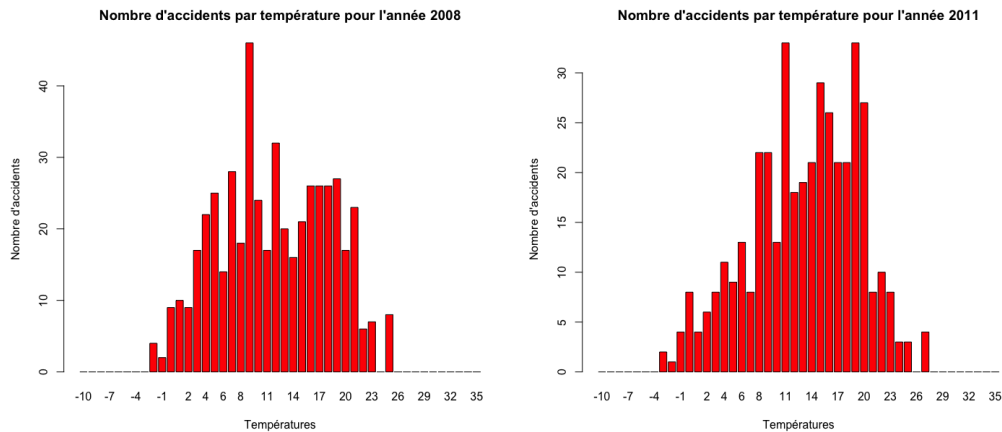


FIGURE 16 – Evolution du nombre d'accidents avec la température

On peut voir ici qu'il y a effectivement plus d'accidents associés aux températures "communes" (entre 10 et 20 degrés). Il y a également un pic à 20 degrés sur l'année 2011. Les différentes pistes permettant de l'explorer sont les départs en vacances et le retour des beaux jours.

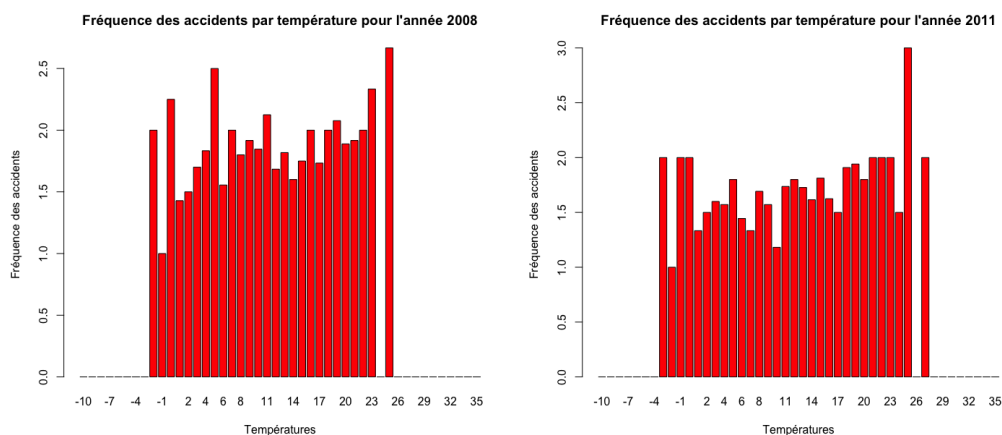


FIGURE 17 – Evolution de la fréquence des accidents avec la température

On observe sur ces deux graphes que les fréquence d'accidents ont plutôt tendance à augmenter lorsque les température sont très élevéesou très basses. Les périodes de températures en dessous de 5 degrés et au dessus de 20 degrés semblent particulièrement accidentogènes.

2.4.2 Humidité

Une humidité élevée correspondant à des conditions particulièrement exécrable, on s'attend à ce que ce soit les conditions les plus accidentogènes.

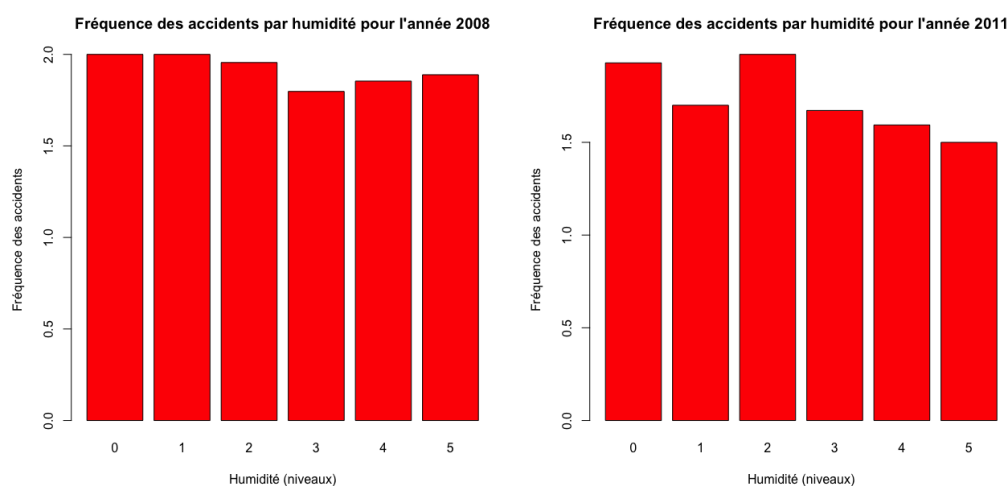


FIGURE 18 – Evolution de la fréquence des accidents avec l'humidité

Description des niveaux :

- 0-50% = niveau 0
- 50-60% = niveau 1
- 60-70% = niveau 2
- 70-80% = niveau 3
- 80-90% = niveau 4
- 90-100% = niveau 5

Les jours avec l'humiditié entre 60 et 90% sont ceux qui présentent le plus grand nombre d'accidents. Comme pour la température, ce sont aussi les conditions les plus fréquentes. On obtient le graphique suivant en considérant la fréquence.

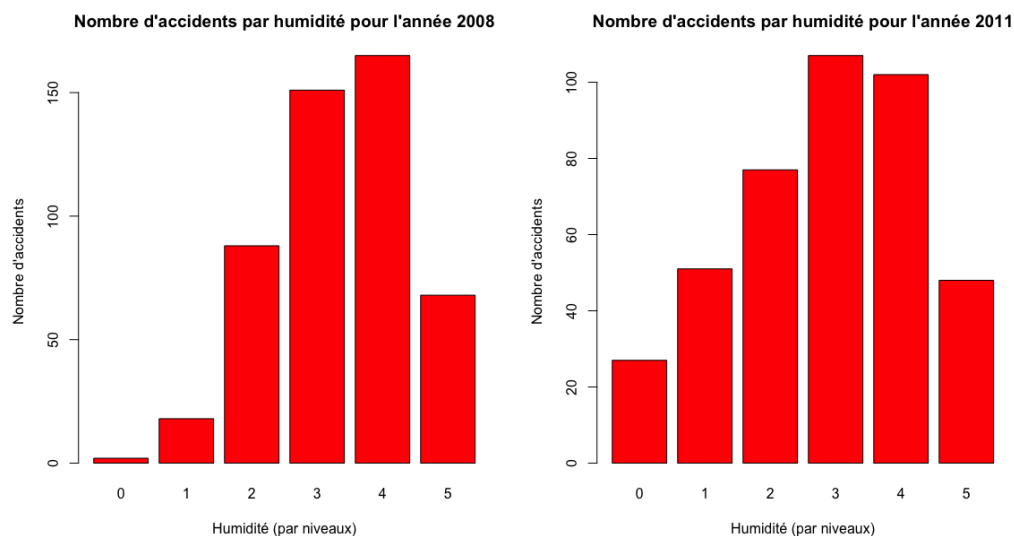


FIGURE 19 – Evolution du nombre d'accidents avec l'humidité

On peut voir sur ce graphique que l'humidité n'a pas d'influence particulière sur le nombre d'accidents à priori. Il n'y a pas de niveau provoquant significativement plus d'accident, et pour 2011 on peut même noter une légère diminution de la fréquence des accidents lorsque les conditions empirent. Cela peut s'expliquer que lors de mauvaises conditions météo, l'attention et la prudence sont accrues, ou alors les véhicules ne sont pas utilisés.

2.4.3 Gravité des accidents par rapport à la météo

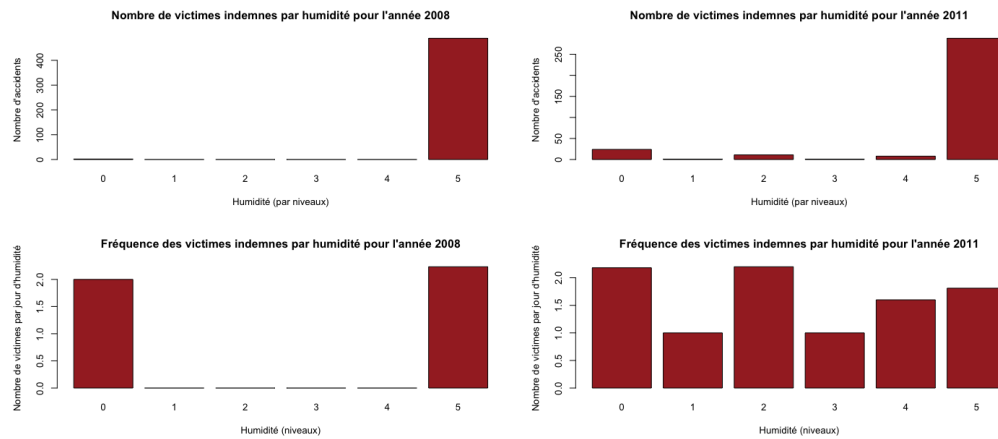


FIGURE 20 – Graphes pour les victimes indemnes

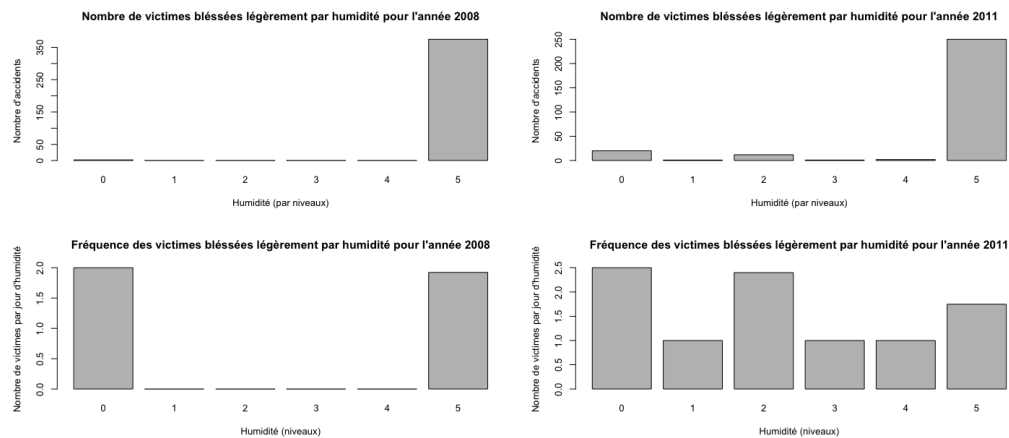


FIGURE 21 – Graphes pour les victimes légèrement blessées

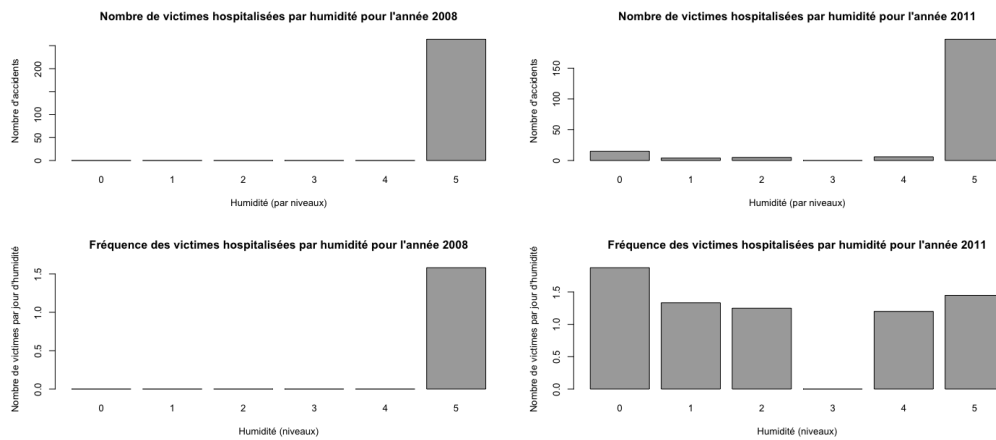


FIGURE 22 – Graphes pour les victimes hospitalisées

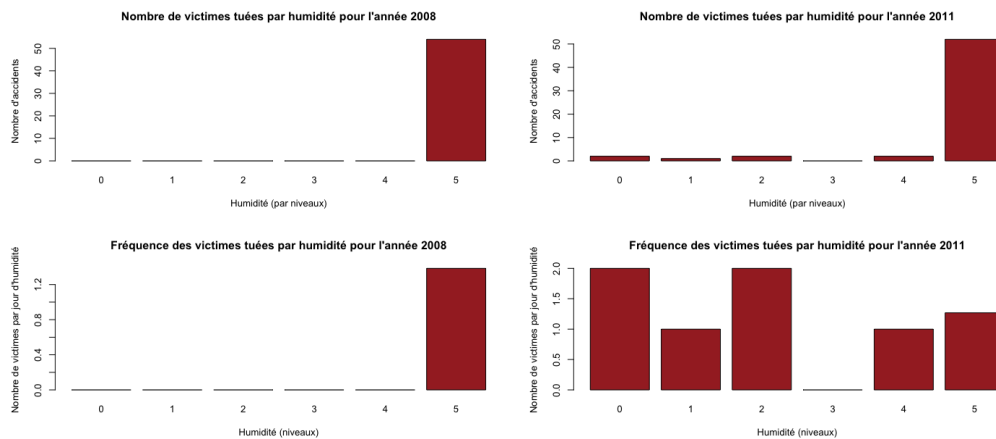


FIGURE 23 – Graphes pour les victimes tuées

De l'étude des ces graphiques on peut déjà déduire qu'il va être difficile d'étudier l'année 2008 car on manque d'échantillons. D'un point de vue pragmatique, cela est une bonne chose, on ne souhaite pas la multiplication des accidents pour les étudier, au contraire. En revanche cela va constituer un problème pour étudier les relations entre la météo et la gravité des accidents, puisque pour l'humidité, en 2008, la plupart des accidents ont été enregistrés avec une humidité entre 90 et 100%.

2.5 Effets des événements calendaires et culturels

2.5.1 Jours fériés

Le but de cette partie est de voir l'impact des jours fériés sur le taux de l'accidentalité, pour la France et le Loiret. Nous avons comparé le nombre d'accidents commis dans des jours fériés aux nombres d'accidents commis dans d'autres jours. Nous n'avons obtenu aucun résultat qui mérite d'être énoncé. On ne peut pas dire qu'il y a un lien particulier entre ces jours-là et les accidents.

2.5.2 Vacances scolaires

Dans cette partie nous voulons étudier l'effet des vacances scolaires sur le nombre d'accidents dans le Loiret. Pour ce faire, nous avons calculé le nombre d'accidents par jour pendant les vacances scolaires de 2010 à 2016. Le tableau suivant contient les résultats trouvés.

Vacances/Année	2010	2011	2012	2013	2014	2015	2016
Hiver	1,33	1,13	0,67	0,375	0,89	1,19	1,33
Printemps	1,4	1	0,56	1,06	0,94	1,06	1,47
Été	1,21	1,08	0,73	0,65	0,72	1,03	0,78
Noël	0,87	0,875	0,56	0,69	0,31	0,69	0,8
Toussaint	0,82	1,09	1,27	0,75	0,625	1,06	1

TABLE 3 – Nombre d'accidents par jour durant les vacances scolaires de 2010 à 2016

Si on lit le tableau année par année, on remarque que le maximum de nombre d'accidents est atteint aux vacances d'hiver, de printemps ou de la Toussaint. Mais, on ne peut pas tirer une conclusion précise sur l'effet des vacances sans avoir le nombre d'accidents moyen par jour en chaque année.

Année	2010	2011	2012	2013	2014	2015	2016
Nombre d'accidents par jour	1,39	1,14	0,93	0,88	0,92	0,86	0,73

TABLE 4 – Nombre d'accidents moyen par jour de 2010 à 2016

D'après le tableau 4, la moyenne est dépassée pendant les vacances de

printemps en 2010 et à partir de 2013. Donc, on peut dire que les vacances de printemps en un impact négatif sur le nombre d'accidents. En ce qui concerne les autres vacances, on ne peut pas conclure.

2.5.3 Changement d'heure

Pour évaluer l'impact du changement d'heure sur le taux de l'accidentalité au Loiret, nous avons comparé le nombre des accidents dans la semaine qui suit cet événement et le nombre d'accidents dans une semaine normale. Nous avons eu une différence de 27 accidents par jours en moyenne. Mais, nous ne sommes pas sûrs totalement que le changement est la seule raison derrière ce petit gap. En fait, le passage à l'heure d'hiver coïncide avec les vacances de la Toussaint connu par le weekend le plus meurtrier. En suivant le même raisonnement, le passage à l'heure d'été peut coïncider à son tour avec des événements que nous ignorons jusqu'à présent.

2.6 Résultats des modèles

2.6.1 Modèle explicatif

2.6.1.1 Modèle global Un premier modèle que nous avons lancé sur les données contient les variables suivantes :

- *Wekeend* : indique si l'accident s'est passé le weekend
- *Luminosité* : décrit l'intensité de la lumière : pas d'éclairage, nuit avec éclairage, crépuscule ou aube, plein jour
- *Pluviométrie* : décrit les conditions météo : conditions normales, pluie faible, pluie forte et neige
- *Agglomération* : indique si l'accident s'est passé en agglomération ou hors agglomération
- *Catégorie de l'usager* : indique si la personne impliquée dans l'accident est piéton ou conducteur
- *Âge de la victime*
- *Sexe de la victime* : homme ou femme
- *Trajet1* : indique si la personne impliquée dans l'accident faisait le trajet domicile-travail ou pas
- *Trajet2* : indique si la personne impliquée dans l'accident faisait le trajet domicile-école ou pas
- *Trajet3* : indique si la personne impliquée dans l'accident conduisait pour faire les courses

- *Trajet4* : indique si la personne impliquée dans l'accident conduisait pour une utilisation professionnelle (chauffeur, transporteur ..)
- *Trajet5* : indique si la personne impliquée dans l'accident conduisait pour une promenade

Nous avons fait tourné ce modèle sur les données de toute la France, après élimination de la région Parisienne et de la Corse. En effet, Paris ne fait pas partie des zones d'intervention de la gendarmerie et la Corse présente un comportement particulier : il y a peu de gens qui y habitent et comme c'est une destination touristique, il y a un certain déséquilibre entre les périodes de l'année en terme d'accidents.

Nous avons clustérisé par département et avons pris les moyennes des effets sur chaque département. C'est ainsi que nous obtenons une vue globale de l'importance de chaque variable. Le tableau suivant résume les résultats de ce modèle. Les astérisques indiquent que la variable est significative dans le modèle, c'est-à-dire qu'elle impacte la probabilité que l'accident soit grave. Dans ce cas, le signe du coefficient correspondant est interprétable.

Variables	Description	Coefficients
<i>Weekend</i>	reste de la semaine/ weekend	0.0780*
<i>Luminosité</i>	variable numérique : pas d'éclairage/ nuit avec éclairage/ crépuscule ou aube/ plein jour	0.0633*
<i>Pluviométrie</i>	variable numérique : conditions normales/ pluie légère/ pluie forte/ neige	-0.0314*
<i>Agglomération</i>	hors agglomération/ agglomération	-0.3135*
<i>Catégorie de l'usager</i>	piéton/conducteur	-1.0924*
<i>Âge</i>	variable numérique	-0.0044*
<i>Sexe</i>	homme/femme	-0.0977*
<i>Trajet1</i>	domicile-travail	0.1601*
<i>Trajet2</i>	domicile-école	0.0656*
<i>Trajet3</i>	courses	0.0758*
<i>Trajet4</i>	utilisation professionnelle	-0.4775*
<i>Trajet5</i>	promenade	0.1226*

TABLE 5 – Résultats du modèle global lancé sur tous les départements sauf la région Parisienne et la Corse, nombre d'observations = 723063

Toutes les variables ont des contributions significatives dans le modèle. Commençons par la variable *Weekend*, le coefficient estimé qui lui est associé est positif, ce qui se traduit par le fait que les accidents le weekend sont plus graves que ceux au cours de la semaine. Généralement les soirées sont organisées pendant le weekend, la possibilité qu'on se met au volant en état d'ivresse ou de distraction est importante, ce qui fait que les accidents mortels sont plus probables. *Luminosité* par contre a une influence négative sur la probabilité d'avoir un accidents grave. Plus il y a de l'éclairage sur la route, moins d'accidents graves se produisent. Ce qui est logique. La négativité du coefficient de *Pluviométrie* est synonyme de l'effet suivant : plus le temps est mauvais, plus on fait attention en route, moins d'accidents graves se produisent. *Agglomération* ayant un coefficient négatif a une influence qui tend à diminuer la probabilité d'avoir un accident grave : on peut expliquer ceci par le fait qu'on conduit plus prudemment en agglomération, ce qui évite d'avoir des accidents mortels.

Passons maintenant aux caractéristiques liées aux usagers. Le fait d'être conducteur ou piéton influe sur la probabilité d'être en état grave. En effet, les piétons ont plus de risque d'être victime d'accidents graves. Ce qui fait que dans un accident entre un conducteur et un piéton, le piéton a une probabilité plus importante que celle du conducteur d'être tué. En considérant l'âge de la personne impliquée dans l'accident comme variable continue, on constate que moins on est jeune, moins on est impliqué dans des accidents graves, sous-entendu que les jeunes sont plus probables de commettre/subir des accidents graves. Analysons maintenant l'impact du type de trajet, les personnes qui sont sur la route pour des fins professionnelles (chauffeurs transporteurs ..) font preuve de vigilance et de professionnalisme, ils sont impliqués dans des accidents moins graves. Par contre, les autres trajets, domicile-travail, domicile école, courses et promenade font augmenter la probabilité qu'il y ait un accident plus grave : en allant au travail on est pressé et on ne veut pas être en retard. Pour emmener les enfants à l'école aussi. On fait les courses généralement le soir après une longue journée de travail, on a envie de rentrer le plus tôt possible. Les promenades sont souvent liées à une distraction, une consommation de l'alcool, ce qui fait qu'on est moins concentrés.

Nous insistons sur l'aspect global de ces résultats, puisqu'ils sont obtenus sur l'ensemble de la base.

Pour avoir une vue locale de ce sujet, on applique, séparément, le même modèle à plusieurs départements.

2.6.1.2 Application du modèle à certains départements Nous appliquons le modèle explicatifs à d'autres départements. Les résultats sont résumés dans le tableau suivant :

Variables	Description	Dép. 13	Dép. 59	Dép 05
<i>Weekend</i>	reste de la semaine/ weekend	0.0672*	-0.0004*	0.1116
<i>Luminosité</i>	variable numérique : pas d'éclairage/ nuit avec éclairage/ crépuscule ou aube/ plein jour	-0.0508*	-0.0880*	-0.0540*
<i>Pluviométrie</i>	variable numérique : conditions normales/ pluie légère/ pluie forte/ neige	-0.0009	-0.0612*	-0.0469
<i>Agglomération</i>	hors agglomération/ agglomération	-0.1968*	-0.6780*	-0.2657*
<i>Catégorie de l'utilisateur</i>	piéton/conducteur	-1.0500*	-1.0365*	-1.2243*
<i>Âge</i>	variable numérique	-0.0062*	0.0001*	-0.0042
<i>Sexe</i>	homme/femme	-0.0092	-0.1014*	-0.0810
<i>Trajet1</i>	domicile-travail	0.1617*	0.3491*	0.1192
<i>Trajet2</i>	domicile-école	0.0044	-0.0823*	0.1563
<i>Trajet3</i>	courses	-0.0299	0.0159	0.0864
<i>Trajet4</i>	utilisation professionnelle	-0.3724*	-0.5555*	-0.4346*
<i>Trajet5</i>	promenade	0.0413	0.1454*	0.1113
Nombre d'observations		73150	33424	1956

TABLE 6 – Résultats du modèle pour quelques exemples de départements

Nous avons choisi le département de Bouches de Rhône (13) car il présente le plus grand nombre d'accidents après Paris. La significativité et l'influence des variables diffèrent. la pluviométrie n'est plus significative, ce qui peut être la conséquence d'un climat spécifique à ce département. Le sexe aussi est non significatif, ainsi que les trajet domicile école et promenade.

Au Nord (59), *Weekend* a un effet négatif : au weekend on a tendance à avoir des accidents moins graves, ceci peut être expliqué par le nombre peu important des habitants de cette région. De même pour *Âge* : plus on est vieux plus on est appliqué dans des accidents graves.

Pour le département des Hautes-Alpes (05), seules la lumière, l'agglomération, le trajet professionnel et la catégorie de la victime sont significatives. Elles ont le même effet que le modèle général.

2.6.1.3 Modèle du Loiret (45) En partant du modèle précédent, nous avons rajouté les variables suivantes :

- *Humidité*
- *Véhicule1* indique si l'accident implique au moins un véhicule de type vélos
- *Véhicule2* indique si l'accident implique au moins un véhicule de type cyclos et assimilés
- *Véhicule3* indique si l'accident implique au moins un véhicule de type VL et assimilés
- *Véhicule4* indique si l'accident implique au moins un véhicule de type PL et TC
- *Véhicule5* indique si l'accident implique au moins un véhicule de type 2RM et assimilés

Afin d'interpréter l'impact de chaque variable explicative sur le variable dépendante, nous utilisons, en plus du signe du coefficient correspond, l'effet marginal de chacune. En effet, pour mesurer la sensibilité de la probabilité d'avoir un accident de gravité $i=0,1,2,3$ par rapport aux variables explicatives, on calcule la variation de celle-ci induite par une « faible » perturbation d'une variable explicative X_i . Lorsque les variables explicatives sont continues, l'effet marginal est égal à la dérivée de la probabilité estimée p par rapport aux composantes de X_i : $\frac{\partial p}{\partial X_i}$.

Par contre, lorsque les variables explicatives sont qualitatives, l'effet marginal se calcule selon le procédé suivant : soit X_i une variable explicative factorielle qui prend deux valeurs 0 ou 1. Le modèle simplifié de la probabilité p s'écrit

$$p_i = \Phi(\beta_0 + \dots + \beta_i X_i + \dots)$$

Pour évaluer l'impact de X_i sur la probabilité de l'événement $Y=i$, $i=0,1,2,3$, (Y est la gravité de l'accident), on calcule d'une part la probabilité p_{i1} associée à $X_i = 1$ et d'autre part la probabilité p_{i0} associée à $X_i = 0$, à toute autre variable égale. L'effet de X_i est donc mesuré par l'écart entre les deux probabilités p_{i0} et p_{i1} .

Seuls les effets marginaux d'une variable significative sont interprétables, en signe et en valeur également. L'interprétation du signe de l'effet marginal d'une variable se fait de la même façon que le coefficient du modèle.

Le tableau suivant présente les résultats, coefficients et effets marginaux, du modèle appliqué au département du Loiret.

Variables	Description	Coefficients	Effets marginaux			
			sur $\mathbb{P}(Y=0)$	sur $\mathbb{P}(Y=1)$	sur $\mathbb{P}(Y=2)$	sur $\mathbb{P}(Y=3)$
<i>Weekend</i>	reste de la semaine/ weekend	0.0246	-0.0094	0.0008	0.0065	0.0020
<i>Luminosité</i>	variable numérique : pas d'éclairage/ nuit avec éclairage/ crépuscule ou aube/ plein jour	-0.0679*	0.0255*	-0.0023*	-0.0178*	-0.0055*
<i>Pluviométrie</i>	variable numérique : conditions normales/ pluie légère/ pluie forte/ neige	-0.0500*	0.0188*	-0.0017*	-0.0131*	-0.0040*
<i>Humidité</i>	variable numérique	-0.0007	0.0004	-0.0001	-0.0003	-0.0001
<i>Agglomération</i>	hors agglomération/ agglomération	-0.5621*	0.2063*	-0.0079*	-0.1460*	-0.0523*
<i>Catégorie de l'usager</i>	piéton/conducteur	-0.8754*	0.2690*	0.0654*	-0.2*	-0.1345*
<i>Âge</i>	variable numérique	-0.0024*	0.0009*	-0.0001*	-0.0006*	-0.0002*
<i>Sexe</i>	homme/femme	-0.0708*	0.0271*	-0.0026*	-0.0188*	-0.0057*
<i>Trajet1</i>	domicile-travail	0.1883*	-0.0693*	0.002*	0.0499*	0.0175*
<i>Trajet2</i>	domicile-école	0.2288*	-0.0825*	-0.0001	0.0601*	0.0224*
<i>Trajet3</i>	courses	0.0695	-0.0239	0.0015	0.0169	0.0055
<i>Trajet4</i>	utilisation professionnelle	-0.5533*	0.2169*	-0.0525*	-0.1338*	-0.0306*
<i>Trajet5</i>	promenade	0.2190*	-0.0818*	0.0043*	0.0581*	0.0195*
<i>Véhicule1</i>	vélos	-0.0422	0.0164	-0.0017	-0.0114	-0.0034
<i>Véhicule2</i>	cyclos et assimilés	-0.0202	0.0089	-0.0008	-0.0062	-0.0019
<i>Véhicule3</i>	VL et assimilés	0.0107	-0.0011	0.0001	0.0007	0.0002
<i>Véhicule4</i>	PL et TC	0.0490	-0.0173	0.0013	0.0122	0.0039
<i>Véhicule5</i>	2RM et assimilés	0.0315	-0.0109	0.0009	0.0076	0.0024

TABLE 7 – Résultats du modèle pour le Loiret, nombre d'observations = 12706

Le fait d'être le weekend était significatif pour toute la France (sauf la région parisienne et la Corse) alors qu'il ne l'est pas pour le Loiret. Le signe du coefficient montre l'intensité de la lumière a un effet similaire à celui décrit précédemment. La présence de la lumière fait diminuer la probabilité d'avoir un accident grave. Regardons maintenant les effets marginaux. Quand il y a plus d'éclairage, la probabilité d'avoir une victime indemne $\mathbb{P}(Y = 0)$ augmente de 2,55%, alors que la probabilité qu'un usager dans un accident soit hospitalisé $\mathbb{P}(Y = 2)$ ou tué $\mathbb{P}(Y = 3)$ diminue, respectivement, de 1.78% et 0.5%. La même interprétation est faite pour *Pluviométrie*. Cependant, *Humidité* n'est pas significatif, ce qui fait que ses effets marginaux ne le sont pas également. De plus, un accident en agglomération a plus de chance d'être moins grave, c'est un effet qu'on a déjà vu pour toute la France. La probabilité que la victime soit indemne augmente de 20.63% et celle qu'elle soit tuée diminue de 5.23%.

Passons maintenant aux caractéristiques de l'usager. Le fait d'être conducteur fait diminuer le risque corporel par rapport au piéton. Un conducteur a 26.9% plus de chance de sortir intacte d'un accident. La probabilité d'être blessé légèrement augmente de 6,54%. Son hospitalisation et sa mort sont moins probables : diminuent respectivement de 20% et 13.45%, ce qui est énorme. L'âge de la victime est significatif mais son effet est faible : les valeurs des effets sont proches et petites. Il garde la même influence décrit précédemment. En outre, les femmes impliquées dans un accident ont moins de risque. Par rapport au fait d'être un homme, une victime féminine a plus de chance de survivre : 2.7% de plus d'être indemne, 1.88% de moins d'être hospitalisée et 0.5% de moins d'être tuée. Ce n'est pas énorme, mais c'est un plus qu'ont les femmes par rapport aux hommes. Observons maintenant les différents types de trajets, ils ont les mêmes effets vus précédemment. Les effets marginaux montrent que le trajet domicile-école est le plus significatif en terme d'influence sur les probabilités de gravité. Une victime faisant ce trajet a une probabilité de 8.25% de plus d'être touchée : 6.01% de plus d'être hospitalisée et 2.24% d'être tuée. Cependant, c'est l'utilisation professionnelle qui est plus sûre : 21.69% de chance de plus que la victime soit indemne, 5% de moins qu'elle soit blessée, 13% de moins qu'elle soit hospitalisée et 3% de moins qu'elle soit tuée. Finalement, les catégories de véhicule s'avèrent non significatives.

La cartographie des accidents (voir section 2.6) montre deux zones d'accumulation des accidents. Il s'agit d'Orléans et de Montargis.

On applique le même modèle décrit précédemment pour ces deux zones. Les deux tableaux suivants présentent les résultats obtenus.

Variables	Description	Coefficients	Effets marginaux			
			sur $\mathbb{P}(\mathbf{Y}=0)$	sur $\mathbb{P}(\mathbf{Y}=1)$	sur $\mathbb{P}(\mathbf{Y}=2)$	sur $\mathbb{P}(\mathbf{Y}=3)$
<i>Weekend</i>	reste de la semaine/ weekend	0.0940*	-0.0365*	0.0113*	0.0221*	0.0030
<i>Luminosité</i>	variable numérique : pas d'éclairage/ nuit avec éclairage/ crépuscule ou aube/ plein jour	-0.0674*	0.02632*	-0.0087*	-0.0156*	-0.0021*
<i>Pluviométrie</i>	variable numérique : conditions normales/ pluie légère/ pluie forte/ neige	-0.0331	0.0129	-0.0043	-0.0077	-0.001
<i>Humidité</i>	variable numérique	-0.0001	0.00004	-0.00002	-0.00003	0
<i>Agglomération</i>	hors agglomération/ agglomération	-0.3372*	0.1264*	-0.0275*	-0.0849*	-0.0141*
<i>Catégorie de l'usager</i>	piéton/conducteur	-1.0970*	0.3460*	0.0291*	-0.2844*	-0.0908*
<i>Âge</i>	variable numérique	-0.0048*	0.0019*	-0.0006*	-0.0011*	-0.0001*
<i>Sexe</i>	homme/femme	-0.0571	0.0223	-0.0075	-0.0131	-0.0017
<i>Trajet1</i>	domicile-travail	0.1372	-0.0528	0.015*	0.0331	0.0048
<i>Trajet2</i>	domicile-école	0.1828	-0.0698	0.0180	0.0449	0.0068
<i>Trajet3</i>	courses	0.0121	-0.0047	0.0015	0.0028	0.0004
<i>Trajet4</i>	utilisation professionnelle	-0.4489*	0.1775*	-0.0785*	-0.0895*	-0.0095*
<i>Trajet5</i>	promenade	0.1575*	-0.0607*	0.0176*	0.0378*	0.0054*
<i>Véhicule1</i>	vélos	0.0114	-0.0045	0.0015	0.0027	0.0004
<i>Véhicule2</i>	cyclos et assimilés	0.0539	-0.0209	0.0066	0.0127	0.0017
<i>Véhicule3</i>	VL et assimilés	0.0248	-0.0097	0.0066	0.0057	0.0007
<i>Véhicule4</i>	PL et TC	0.0653	-0.0253	0.0078	0.0154	0.0021
<i>Véhicule5</i>	2RM et assimilés	0.0225	-0.0088	0.0028	0.0052	0.0007

TABLE 8 – Résultats du modèle pour Orléans, nombre d'observations = 3443

Variables	Description	Coefficients	Effets marginaux			
			sur $\mathbb{P}(Y=0)$	sur $\mathbb{P}(Y=1)$	sur $\mathbb{P}(Y=2)$	sur $\mathbb{P}(Y=3)$
<i>Weekend</i>	reste de la semaine/ weekend	0.0701	-0.0272	0.012	0.0131	0.0021
<i>Luminosité</i>	variable numérique : pas d'éclairage/ nuit avec éclairage/ crépuscule ou aube/ plein jour	-0.0898*	0.035*	-0.0159*	-0.0166*	-0.0026*
<i>Pluviométrie</i>	variable numérique : conditions normales/ pluie légère/ pluie forte/ neige	-0.0972	0.0379	-0.0172	-0.0179	-0.0028
<i>Humidité</i>	variable numérique	-0.0041	0.0016	-0.0007	-0.0007	-0.0001
<i>Agglomération</i>	hors agglomération/ agglomération	-0.4583*	0.1684*	-0.0515*	-0.0968*	-0.0202*
<i>Catégorie de l'usager</i>	piéton/conducteur	-1.008*	0.3215*	-0.0138	-0.2307*	-0.077*
<i>Âge</i>	variable numérique	-0.0074*	0.0029*	-0.0013*	-0.0014*	-0.0002*
<i>Sexe</i>	homme/femme	-0.0481	0.0188	-0.0086	-0.0088	-0.0014
<i>Trajet1</i>	domicile-travail	0.0725	0.0284	-0.0135	-0.013	-0.002
<i>Trajet2</i>	domicile-école	0.1013	-0.0391	0.0164	0.0195	0.0033
<i>Trajet3</i>	courses	-0.1260	0.0497	-0.0245	-0.022	-0.0032
<i>Trajet4</i>	utilisation professionnelle	-0.5811*	0.2284*	-0.1367*	-0.0819*	-0.01*
<i>Trajet5</i>	promenade	0.0541	-0.0211	0.0095	0.01	0.0016
<i>Véhicule1</i>	vélos	-0.0158	0.0062	-0.0028	-0.0029	-0.0005
<i>Véhicule2</i>	cyclos et assimilés	-0.0033	0.0013	-0.0006	-0.0006	-0.0001
<i>Véhicule3</i>	VL et assimilés	0.0422	-0.0165	0.0077	0.0077	0.0011
<i>Véhicule4</i>	PL et TC	0.0956	-0.0369	0.0156	0.0183	0.003
<i>Véhicule5</i>	2RM et assimilés	-0.052	0.0204	-0.0096	-0.0094	-0.0014

TABLE 9 – Résultats du modèle pour Montargis, nombre d'observations=1225

On retrouve globalement les mêmes effets pour les différents facteurs, avec quelques différences. *Weekend* n'est pas significatif sur Le Loiret en globalité alors qu'il l'est pour Orléans, c'est logique : en ville la vie des citoyens fait que cette différenciation entre le weekend et le reste de la semaine est belle et bien présente et claire. *Pluviométrie* n'est pas significatif non plus pour ces deux zones, ce qui peut s'expliquer par un climat spécifique à ces zones. *Sexe*, *Trajet1* et *Trajet2* perdent leur significativité également.

2.6.2 Modèles prédictifs

Performances des modèles sur différents départements

- Paris (75) :

Observations	175 706		13065 (sans données manquantes)	
	Précision %	Temps d'entraînement (s)	Précision %	Temps d'entraînement (s)
Arbre de décision	65	0.17	63	0.02
Régression Logistique	62	5.99	60	0.45
Classificateur SVM	62	548.34	60	2.49
Classificateur NB	61	36.57	6	0.005
Classificateur KNN	60	0.07	59	0.13

FIGURE 24 – Résultats des prédictions dans le département 75

Comparons la performances des différents modèles en terme de précision et de temps d'entraînement :

- le modèle AD est le plus performant
- le modèle RL est moins performant et plus lent par rapport au premier. Mais, par rapport aux autres, il est évidemment le meilleur.
- le reste des modèles manque de rapidité ou bien de précision.

On remarque aussi que la diminution du nombre des observations réduit un peu la précision pour les modèles AD, RL, SVM et KNN. Mais, celle du modèle NB est ramenée à 10% de l'ancienne valeur.

- Loiret (45) :

Observations	11260		7886 (sans données manquantes)	
	Précision %	Temps d'entraînement (s)	Précision %	Temps d'entraînement (s)
Arbre de décision	48	0.012	48	0.011
Régression Logistique	46	0.166	45	0.159
Classificateur SVM	43	2.189	43	1.633
Classificateur NB	45	0.003	44	0.003
Classificateur KNN	45	0.055	43	0.03

FIGURE 25 – Résultats des prédictions dans le département 45

- les modèles NB et KNN détrônent le modèle RL de sa deuxième place vu qu'ils atteignent presque la même précision mais beaucoup plus rapidement.

- Département du Nord (59) :

Observations	51314		7357 (sans données manquantes)	
590	Précision %	Temps d'entraînement (s)	Précision %	Temps d'entraînement (s)
Arbre de décision	52	0.053	49	0.012
Régression Logistique	50	1.019	47	0.128
Classificateur SVM	49	34.799	48	1.004
Classificateur NB	50	0.017	46	0.003
Classificateur KNN	49	1.117	46	0.013

FIGURE 26 – Résultats des prédictions dans le département 59

Les résultats du département 59 (ainsi que ceux d'autres départements) confirment le fait que la diminution du nombre d'observations engendre une perte de précision.

Ils confirment aussi que le modèle AD est le plus adéquat pour faire la prédiction de la gravité des accidents dans la plupart des départements.

- Département (976) (exception) :

Observations	910		633 (sans données manquantes)	
976	Précision %	Temps d'entraînement (s)	Précision %	Temps d'entraînement (s)
Arbre de décision	52	0.001	48	0.003
Régression Logistique	57	0.228	60	0.028
Classificateur SVM	55	0.395	61	0.007
Classificateur NB	4	0.002	6	0.256
Classificateur KNN	49	0.002	50	0.001

FIGURE 27 – Résultats des prédictions dans le département 976

On remarque qu'il y a eu beaucoup de changements avec ce département exceptionnel.

- Le modèle AD est détrôné de sa première place
- les classificateurs SVM et RL gagne beaucoup en précision et en rapidité
- Le classificateur NB perd énormément de sa précision

Remarques

- Les données manquantes sont parmi les raisons principales qui dégradent la qualité des modèles prédictifs et leurs robustesses.
- Vu que chaque département a ses propres caractéristiques, il est difficile de construire un modèle qui convient parfaitement à tous les départements. Donc, il faut s'approfondir dans l'étude de chaque département pour identifier chaque fois les variables prédictives prépondérantes, le modèle le plus adéquat et les paramètres les plus adaptés.
- Un travail de 'features engineering' permettra probablement d'améliorer la performance de ces modèles : discrétisation des variables continues comme les coordonnées géographiques et les heures des accidents, construction et introduction de nouvelles variables prédictives comme le changement d'heure et les événements calendaires, regroupement des départements qui se ressemblent dans l'étape de l'entraînement du modèle, etc.

2.7 Cartographie des accidents

Cette partie englobe une grande partie du travail qui a été fait auparavant allant de l'analyse descriptive jusqu'aux effets marginaux. En effet, nous avons créé deux cartes. La première sert à visualiser les accidents ainsi que leurs localisations. Elle contient des filtres qui permettent de raffiner les critères de visualisation tout comme le sexe, l'âge, les conditions météorologiques... On peut aussi y visualiser le trafic pour l'année 2010. La deuxième carte est consacrée aux effets marginaux. Elle permet de voir pour un niveau de gravité choisi l'effet de chaque variable sur ce dernier.

3 Perspectives

- Intégration de la rubrique lieux de la base d'accidentalité : ceci sera intéressant car on pourrait étudier l'impact de chaque structure de route (pont, tunnel ..)
- Étude des événements culturels dans une zone spécifique : matches, fêtes, booms, ..
- Impact de l'infrastructure
- Plusieurs variables dans le fichier relatif au trafic sont à exploiter
- Cartographier le trafic de toutes les années
- Illustrer sur la carte les effets marginaux des zones d'accumulation des accidents
- Prédiction du nombre d'accidents et visualisation sur une carte
- Étude des possibilités relatives à l'acquisition et au traitement des données au sein de la Gendarmerie, y compris au niveau européen. remédier au problème des données manquantes
- améliorer la précision des algorithmes prédictifs : rajouter d'autres données qui captent des spécificités du Loiret,

4 Difficultés rencontrées

- Tout au long de notre travail nous avons fait face à diverses difficultés :
- Fixation d'un plan de travail : au début, le sujet était très vaste et pouvait être abordé sous plusieurs angles, des choix étaient donc faits après une discussion entre les personnes impliquées dans le projet.
 - Collecte et préparation des données : cette étape est primordiale dans notre projet mais pas du tout évidente. En effet, certains fichiers

contiennent des données non organisées. Par exemple, dans le fichier du trafic en 2016 on trouve des données d'autres années.

- Données manquantes : dans la base des accidents corporels, certaines adresses et coordonnées GPS sont inexistantes. Pour cela, nous avons supprimé les accidents sans adresses puis nous avons converti les adresses en coordonnées GPS.
- Gestion du travail en groupe : comme notre équipe est hétérogène, nous n'avons pas les mêmes connaissances et nous n'utilisons pas les mêmes outils. Bien que cette diversité soit enrichissante, dans certains cas elle rendait le travail un peu plus compliqué.
- Gestion du temps : comme dans tout autre projet, la gestion du temps est une tâche difficile surtout à cause de la différence entre les emplois du temps des membres de l'équipe.

Conclusion

Ce projet était pour nous une opportunité pour travailler sur des problèmes concrets de la gendarmerie. Le but est d'aider les gendarmes dans la gestion du trafic et d'améliorer l'efficacité de cette entité en terme de sécurité routière.

Nous avons puisé dans nos connaissances techniques diverses pour mener ce projet. Explication des accidents, modélisation, prédiction et cartographie, ce qu'on a fait n'est qu'un pas vers une révolution potentielle dans le monde numérique de la gendarmerie.

Tout ce travail sera surement complété l'année prochaine par des ingénieurs de l'ENSTA, toujours compétents et dynamiques.

Codebook des variables de la base accidentalité

La rubrique CARACTERISTIQUES

Num_Acc : Numéro d'identifiant de l'accident

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

jour : Jour de l'accident

mois : Mois de l'accident

an : Année de l'accident

hrmn : Heure et minutes de l'accident

lum : 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 selon le 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 selon un code donné par l'INSEE. Le code comporte 3 chiffres calés à droite.

agg : Localisation de l'accident

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 côté
- 4 – Trois véhicules et plus – en chaîne
- 5 – Trois véhicules et plus - collisions multiples
- 6 – Autre collision
- 7 – Sans collision

adr : Adresse postale renseignée pour les accidents survenus en agglomération

gps : un caractère indicateur de provenance

M = Métropole

A = Antilles (Martinique ou Guadeloupe)

G = Guyane

R = Réunion

Y = Mayotte

Coordonnées géographiques en degrés décimaux :

lat : Latitude

long : Longitude

La rubrique LIEUX

Num_Acc : Identifiant de l’accident identique à celui du fichier "rubrique CARACTERISTIQUES" repris dans l’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
larout : 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

La rubrique 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
- 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
- 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** : Manoeuvre 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 – Manoeuvre de stationnement
 21 – Manoeuvre d'évitement
 22 – Ouverture de porte
 23 – Arrêté (hors stationnement)
 24 – En stationnement (avec occupants)
occutc : Nombre d'occupants dans le transport en commun

La rubrique USAGERS

Num_Acc : Identifiant de l'accident identique à celui du fichier "rubrique CARACTERISTIQUES" repris pour chacun des usagers décrits impliqués dans l'accident

place : Permet de situer la place occupée dans le véhicule par l'utilisateur au moment de l'accident

catu : Catégorie d'utilisateur :

- 1 - Conducteur
- 2 - Passager
- 3 - Piéton
- 4 - Piéton en roller ou en trottinette

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'utilisateur

1 - Masculin

2 - Féminin

An_nais : Année de naissance de l'utilisateur

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 - Équipement 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

Variables construites

Weekend : pour un accident donné, elle prend 1 s'il s'est passé le weekend (samedi + dimanche), et 0 sinon *Luminosité* : c'est une variable numérique qui correspond à une échelle de mesure de l'intensité de la lumière :

- 0 pas d'éclairage
- 1 nuit avec éclairage
- 2 crépuscule ou aube
- 3 plein jour

Cette variable a été construite à partir de lum de la base accidentalité.

Pluviométrie : c'est une variable numérique qui correspond à une échelle de mesure des conditions météorologiques :

- 0 conditions normales
- 1 pluie légère
- 2 pluie forte
- 3 neige

Cette variable a été construite à partir de atm de la base accidentalité.

Catégorie de l'usager : a été construite à partir de catu de la base accidentalité.

- 0 si piéton (inclut piéton avec trottinette)
- 1 si conducteur

Trajet1 : correspond au trajet domicile-travail. Elle prend 1 si la victime fait ce trajet et 0 sinon.

Trajet2 : correspond au trajet domicile-école. Elle prend 1 si la victime fait

ce trajet et 0 sinon.

Trajet3 : correspond au trajet courses. Elle prend 1 si la victime fait ce trajet et 0 sinon.

Trajet4 : correspond au trajet utilisatio professionnelle. Elle prend 1 si la victime fait ce trajet et 0 sinon.

Trajet5 : correspond au trajet promenade. Elle prend 1 si la victime fait ce trajet et 0 sinon.

Ces 5 variables trajet ont été contruite à partir de traj de la base accidentalité.

Véhicule1 : correspond aux véhicules de type vélos. Si la victime est impliquée dans un accident comportant au moins un véhicule de ce type, elle prend 1 sinon 0

Véhicule2 : correspond aux véhicules de type cyclos et assimilés. Si la victime est impliquée dans un accident comportant au moins un véhicule de ce type, elle prend 1 sinon 0

Véhicule3 : correspond aux véhicules de type VL et assimilés. Si la victime est impliquée dans un accident comportant au moins un véhicule de ce type, elle prend 1 sinon 0

Véhicule4 : correspond aux véhicules de type PL et TC. Si la victime est impliquée dans un accident comportant au moins un véhicule de ce type, elle prend 1 sinon 0

Véhicule5 : correspond aux véhicules de type 2RM et assimilés. Si la victime est impliquée dans un accident comportant au moins un véhicule de ce type, elle prend 1 sinon 0

Ces 5 variables véhicule ont été contruite à partir de catv de la base accidentalité.

Langages de programmation utilisés

- Python
- R
- Stata
- Java