

Master 1 : Données, Apprentissage et Connaissances (DAC)

Rapport projet BIUM

FACTEURS FAVORISANT LES ACCIDENTS EN FRANCE ET PRÉVENTION

Réalisé par :

AKNOUCHE Anis

HADDADI Hacene

Année universitaire : 2020/2021

SOMMAIRE

INTRODUCTION :	3
Présentation de la problématique :	3
Schéma conceptuel :	4
Importation et traitement des données :	4
Exploration et analyse des données :	4
Analyser les caractéristiques d'un accident (Dimension caractéristiques) :	4
Analyser les caractéristiques des usagers impliqués dans un accidents (Dimension Usager) :	6
Analyser les lieux où se sont produit les accidents :	7
Analyser le temps où se produisent les accidents (Dimension Temps) :	9
Analyser les caractéristiques du véhicule impliqué dans l'accident :	11
Modèle de Machine Learning :	11
Synthèse :	12
Facteurs favorisant les accidents :	12
Prévention :	12

INTRODUCTION :

En France, plus de 50 000 accidents routiers sont recensés chaque année par l'Observatoire national interministériel de la sécurité routière "ONISR" et environ 1.8 million de constats amiables sont recensé par les compagnies d'assurance durant la même période.

La lutte contre ce phénomène est prise en charge au niveau national par les pouvoirs publics à travers la délégation à la sécurité routière. Pour chaque accident corporel (soit un accident survenu sur une voie ouverte à la circulation publique, impliquant au moins un véhicule et ayant fait au moins une victime ayant nécessité des soins), des saisies d'information décrivant l'accident sont effectuées par l'unité des forces de l'ordre (police, gendarmerie ...) qui est intervenue sur le lieu de l'accident. Ces saisies sont rassemblées dans une fiche intitulée bulletin d'analyse des accident corporels. L'ensemble de ces fiches constitue le fichier national des accidents corporels de la circulation dit « Fichier BAAC » administré par l'Observatoire national interministériel de la sécurité routière "ONISR".

Ces enquêtes statistiques ont montré que s'agissant des accidents mortels, le comportement humain est mis en cause dans la plupart des cas, mais que la voirie, l'environnement ou le véhicule peuvent aussi avoir une implication significative.

C'est pourquoi, dans ce projet, nous allons utiliser le Data Mining pour analyser les différents facteurs qui pourraient favoriser la survenu d'un accident, pour ainsi connaître les critères essentiels pour une meilleur prévention.

Présentation de la problématique :

Dans ce projet, nous souhaitons étudier les caractéristiques des différents accidents routiers recensés, des informations sur les usagers impliqués, des véhicules accidentés ainsi que le temps et le lieu des accidents. Notre schéma conceptuel doit nous répondre aux besoins suivants :

- Analyser les caractéristiques d'un accident :
 - Le nombre d'accidents selon les conditions d'éclairages
 - Le nombre d'accidents selon le type d'intersection
 - Le nombre d'accidents selon les conditions atmosphériques
 - Le nombre d'accidents en agglomération et en hors agglomération
- Analyser les caractéristiques des usagers impliqués dans un accidents :
 - Le nombre d'accidents selon la gravité et le type de sécurité
 - Le nombre d'accidents selon l'âge du conducteur
 - Le nombre d'accidents selon le sexe
 - Le nombre d'accidents selon la gravité
 - Le nombre d'accidents selon le type de trajet
- Analyser les lieux où se sont produit les accidents :
 - Le nombre d'accidents par département
 - Le nombre d'accidents par catégorie de route
 - Le nombre d'accidents pour chaque route de France
 - Le nombre d'accidents selon le régime de circulation
 - Le nombre d'accidents selon le nombre de voies
 - Le nombre d'accidents selon la surface de la route
 - Le nombre d'accidents selon l'infrastructure
 - Le nombre d'accidents selon la vitesse maximale autorisée
- Analyser le temps où se produisent les accidents :
 - Le nombre d'accidents par heure
 - Le nombre d'accidents par mois
 - Le nombre d'accidents par jour de fête
 - Le nombre d'accidents en jour de fête et hors jour de fête
- Analyser les caractéristiques du véhicule impliqué dans l'accident :
 - Le nombre d'accidents selon la catégorie du véhicule
 - Le nombre d'accidents selon la motorisation
 - Le nombre d'accidents selon la motorisation et la gravité

Schéma conceptuel :

Le schéma conceptuel suivant nous permettra de répondre à tous nos besoins :

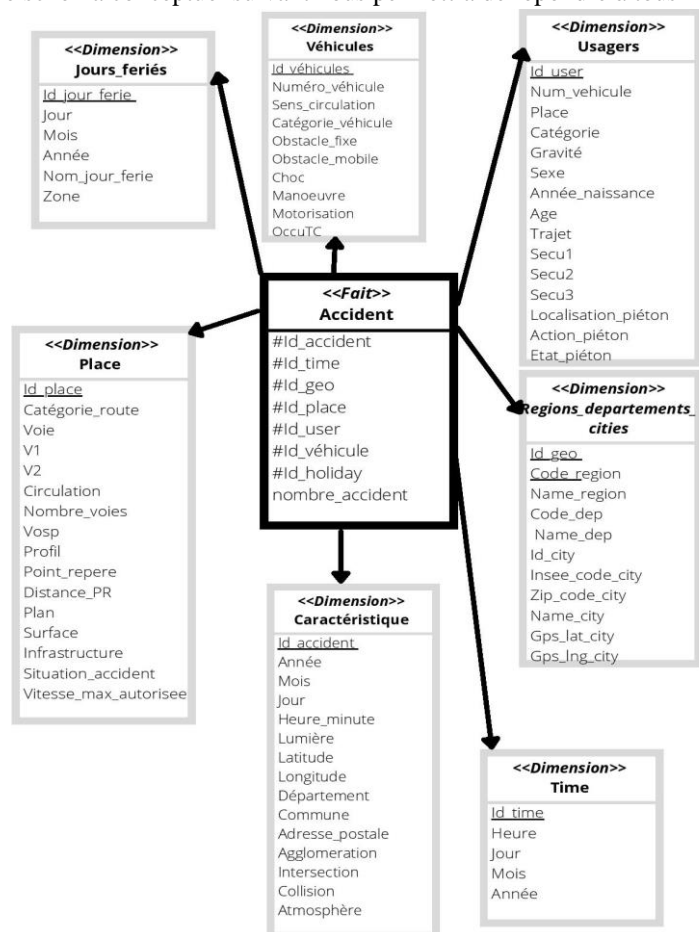


Figure 1 : Schéma en étoile.

Importation et traitement des données :

Pour cette étude, nous allons utiliser les bases de données annuelles des accidents corporels de la circulation routière pour l'année 2019. Ces bases de données sont formées de 4 fichiers (Caractéristiques – Lieux – Véhicules – Usagers) au format csv. On ajoute une base de données des jours de congés en France sous forme de fichier csv, ainsi que 3 fichiers csv (Région - Département - Ville) qui recensent toutes les régions, départements et villes de France.

On commence par importer les fichiers csv sur Dataiku, puis on effectue les pré-traitements :

- Sélection des colonnes pertinentes pour nos analyses, en respectant notre schéma étoile
- Nettoyage des données (Suppression des enregistrements vides, remplacer les valeurs incorrectes ...)
- Corriger les types
- Utiliser des jointures pour préparer les différentes analyses

Exploration et analyse des données :

Analyser les caractéristiques d'un accident (Dimension caractéristiques) :

- Le nombre d'accidents selon les conditions d'éclairages :

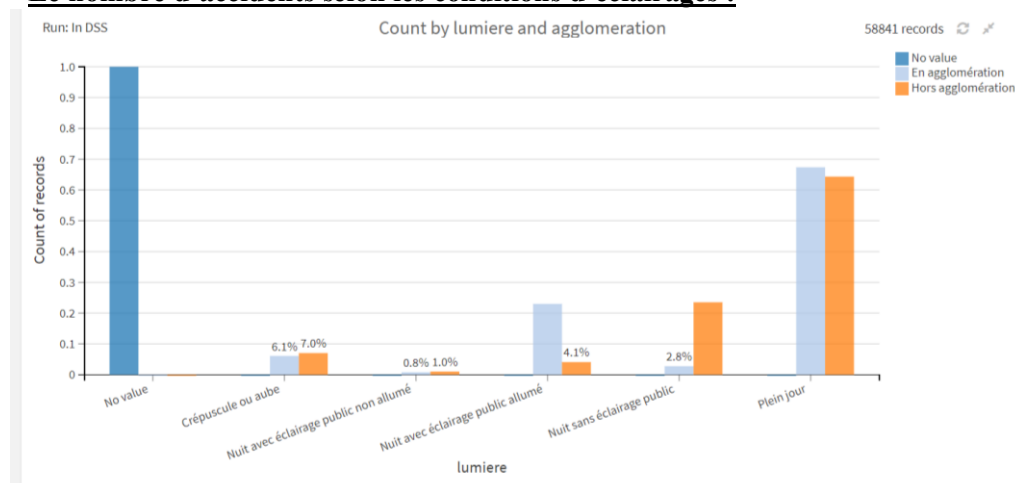


Figure 2 : Nombre d'accidents selon les conditions de lumières et en fonction localisation par rapport à l'agglomération.

- On remarque que la plupart des accidents se sont produit en plein jour (66.28%). Et ceux qui se sont produit en pleine nuit, ils ont eu lieu dans des zones urbaines où l'éclairage public était allumé (16.3%) donc dans un endroit où il y avait de la circulation, mais on observe aussi qu'environ 10% des accidents se sont produit en pleine nuit et sans éclairage.

- Le nombre d'accidents selon le type d'intersection :

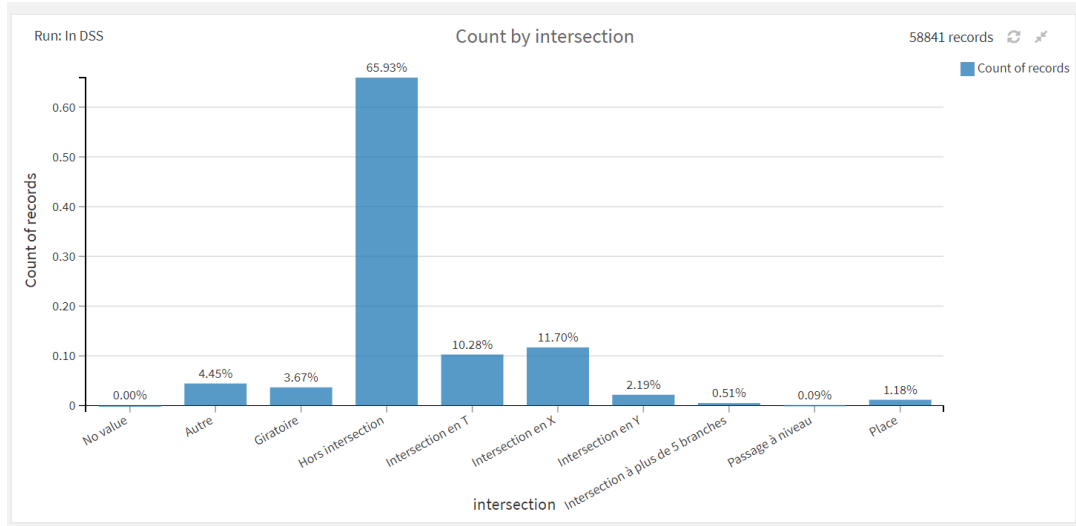


Figure 3 : Pourcentage d'accidents selon le type d'intersection.

- On remarque que la plupart des accidents ont eu lieu hors intersection. Et parmi ceux qui se sont produit dans une intersection on remarque que le type d'intersection en X détient le plus grand nombre d'accidents vu qu'il y a 2 voies qui se croisent, donc la probabilité qu'une collision arrive est plus grande que celle des autres types d'intersections.

- Le nombre d'accidents selon les conditions atmosphériques :

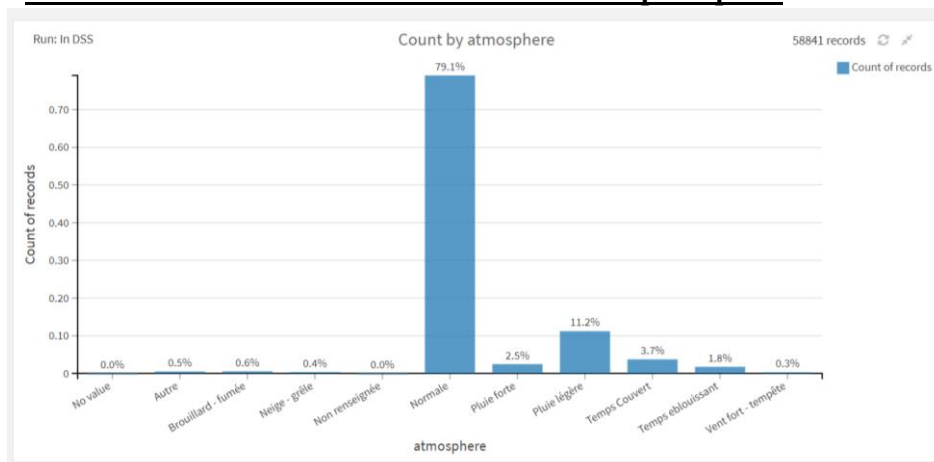


Figure 4 : Pourcentage d'accidents selon les conditions atmosphériques.

- On remarque qu'environ 80% des accidents ont eu lieu dans des conditions atmosphériques normales et environ 11% en temps de pluie légère.

- Le nombre d'accidents en agglomération et hors agglomération :

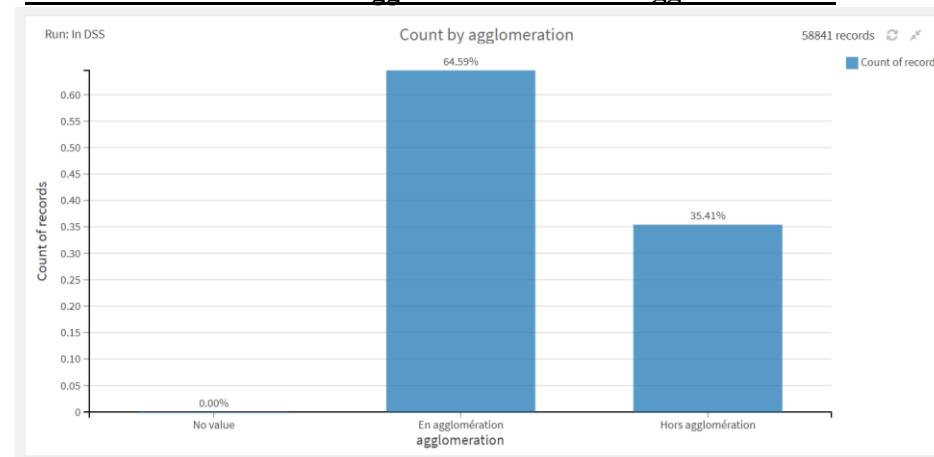


Figure 5 : Pourcentage d'accidents en agglomération et en dehors.

- On remarque que 64% des accidents se produisent en agglomération à cause du fort trafic routier et seulement 35% en dehors des agglomérations.

Analyser les caractéristiques des usagers impliqués dans un accidents (Dimension Usager) :

- Le nombre d'accidents selon la gravité :

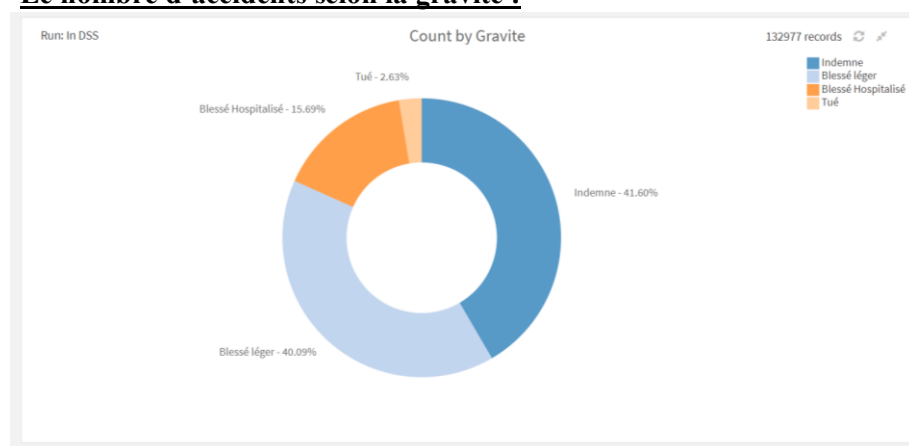


Figure 6 : Pourcentage du nombre d'accidents selon le type de gravité.

- On remarque que pour l'année 2019 environs 80% des accidents sont soit indemne ou blessé léger et moins de 5% ont eu des cas de décès.

- Le nombre d'accidents selon la gravité et le type de sécurité :

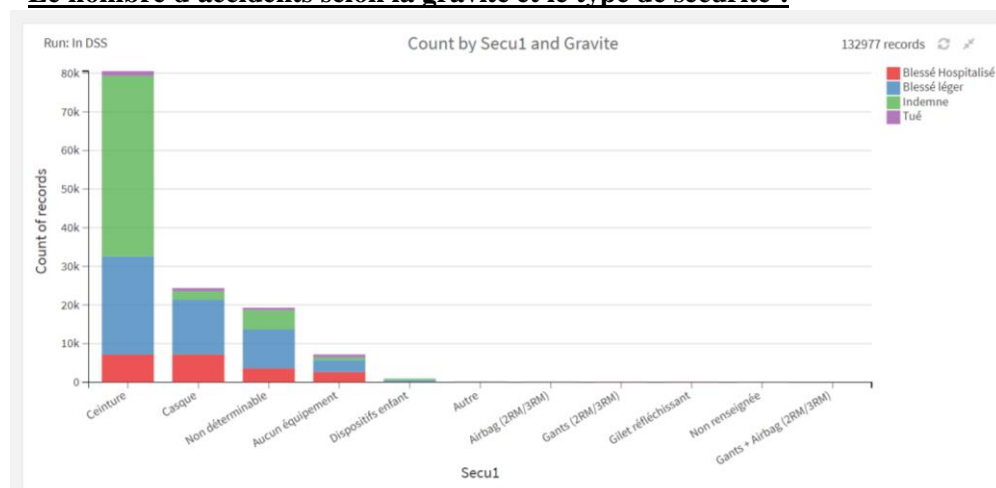


Figure 7 : Pourcentage d'accidents selon la gravité et ayant utilisé une sécurité.

Aucune entrée de table des matières n'a été trouvée.

- Plus de 50% des usagers qui mettent une ceinture sont indemnes et plus de 31% ont des blessures légères.
- Pour les deux roues, plus de 57% de ceux qui mettent un casque ont des blessures légères, et moins de 30% sont hospitalisés.
- Pour ceux qui mettent un casque et des gants, environs 60% sont indemnes et moins de 30% sont hospitalisés et uniquement 3% en sont morts.
- Concernant les dispositifs enfants, plus de 40% sont indemnes et moins de 50% sont hospitalisés.
- Plus de 30% de ceux qui utilisent une ceinture et un airbag comme dispositifs de sécurité sont indemnes, et plus de 45% ont eu des blessures légères et moins de 4% en sont mort.

- Le nombre d'accidents selon l'âge du conducteur :

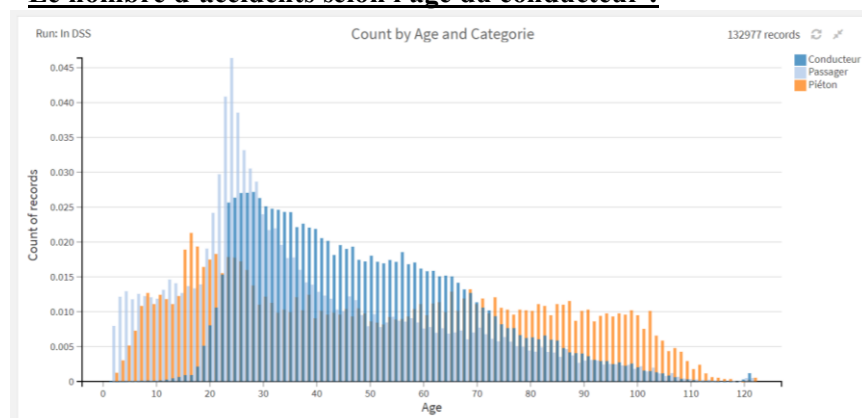


Figure 8 : Nombre d'accidents en fonction de l'âge du conducteur.

- On remarque que la tranche d'âge de plus de 50% des usagers ayant eu un accident se situe entre 63 ans (né en 1958) et 18 ans (né en 2003).

- Le nombre d'accidents selon le sexe :

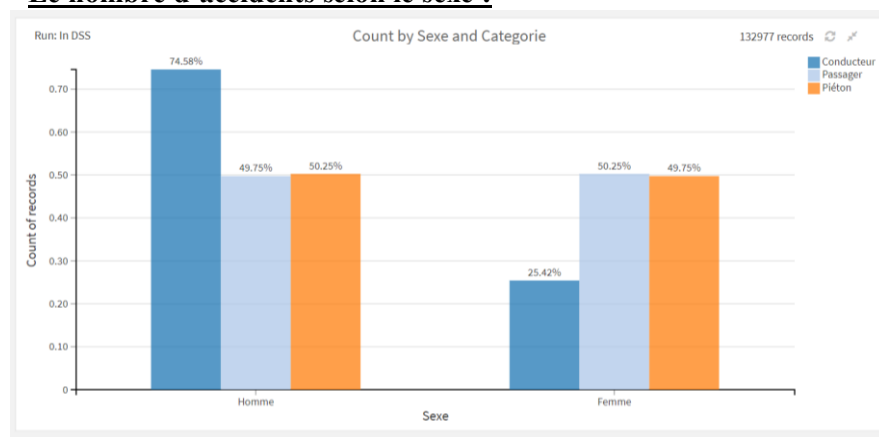


Figure 9 : Nombre d'accidents par sexe.

- On remarque que plus de 70% des conducteurs ayant eu un accident sont de sexe masculin
- On remarque aussi que les piétons et les passagers sont distribués uniformément entre homme et femme.

- Le nombre d'accidents selon le type de trajet :

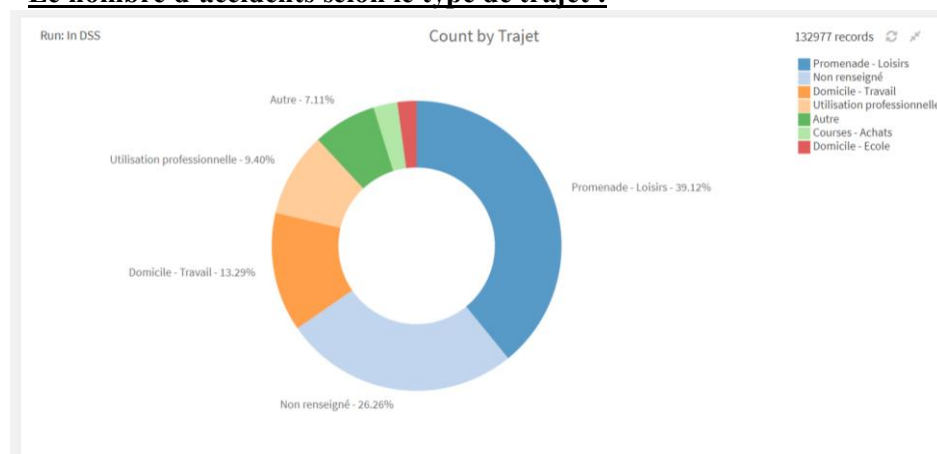


Figure 10 : Nombre d'accidents selon le type de trajet.

- On remarque que 39.12% des accidents arrivent lors de trajets de promenade ou de loisirs. Et environs 13% concerne des trajets entre le domicile et le travail.

Analyser les lieux où se sont produit les accidents :

- Le nombre d'accidents par département :

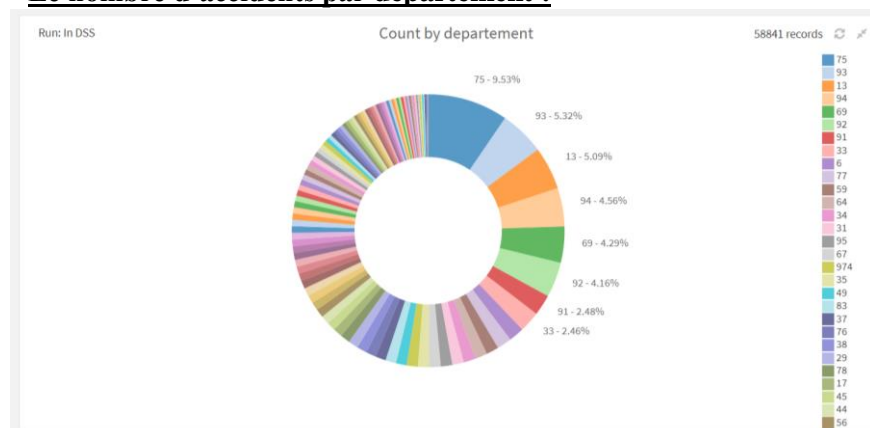


Figure 11 : Nombre d'accidents par département.

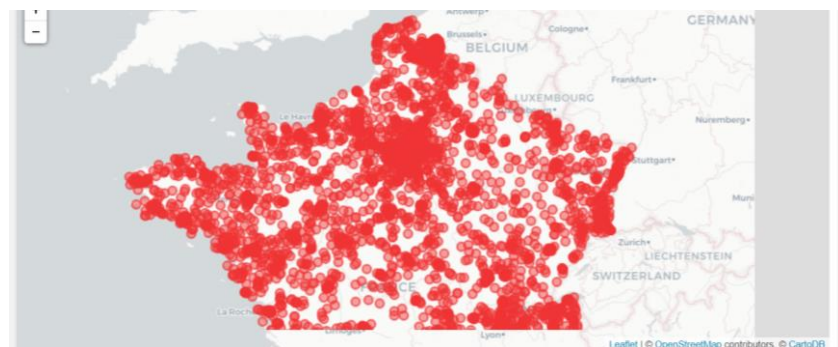


Figure 12 : Localisation des accidents en France.

- On remarque qu'il y a des zones où on a un nombre assez important d'accidents, comme l'Ile de France, au sud le long de la côte d'azur (Monaco, Marseille, Montpellier) et au nord Lille. Ce sont des régions qui ont une forte densité de population et donc beaucoup de trafic routier.

- Le nombre d'accidents par catégorie de route :

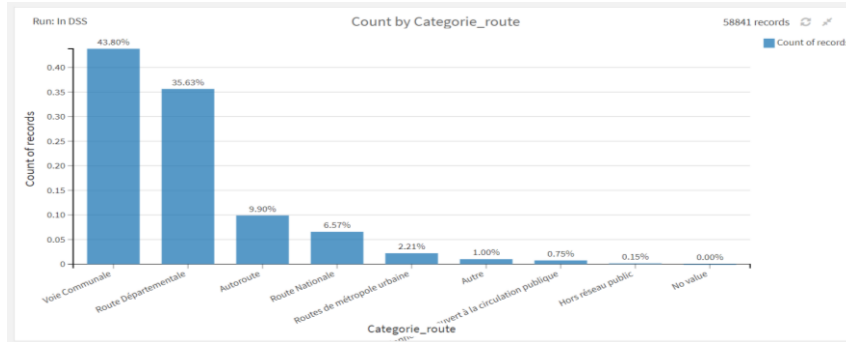


Figure 13 : Nombre d'accidents selon le type de route.

- On remarque que 43.8% des accidents ont eu lieu dans des voies communales, 35.63% dans des routes départementales, 9.9% dans les autoroutes et 6.57% dans les routes nationales.

- Le nombre d'accidents par route en France :

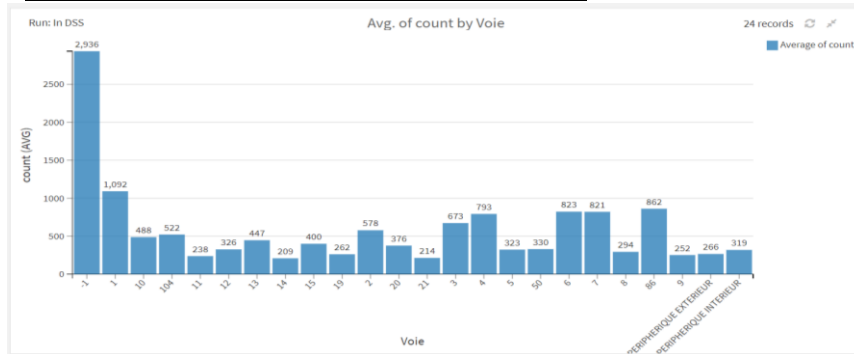


Figure 14 : Nombre d'accidents pour quelques routes en France.

- On remarque un bon nombre d'accidents dont le lieu n'est pas mentionné. ET il y a exactement 23 routes qui compte plus de 200 accidents pour l'année 2019. La route nationale numéro 1 détient le plus grand nombre d'accidents en France.

- Le nombre d'accidents selon le régime de circulation :

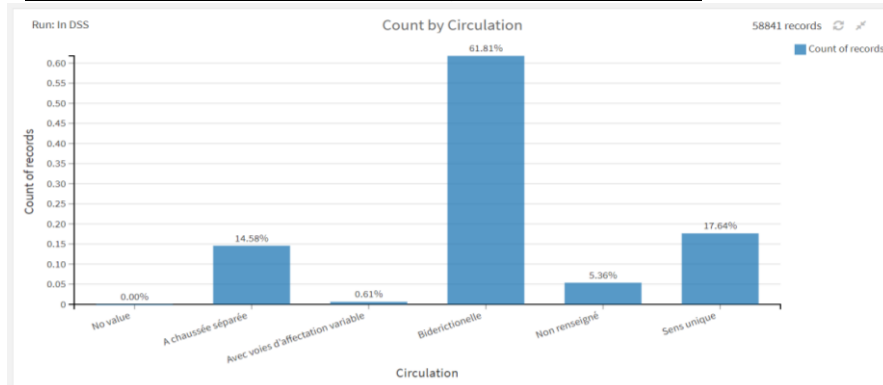


Figure 15 : Nombre d'accidents selon le régime de circulation.

- On remarque que la plupart des accidents se produisent sur les routes bidirectionnelles.

- Le nombre d'accidents selon le nombre de voies :

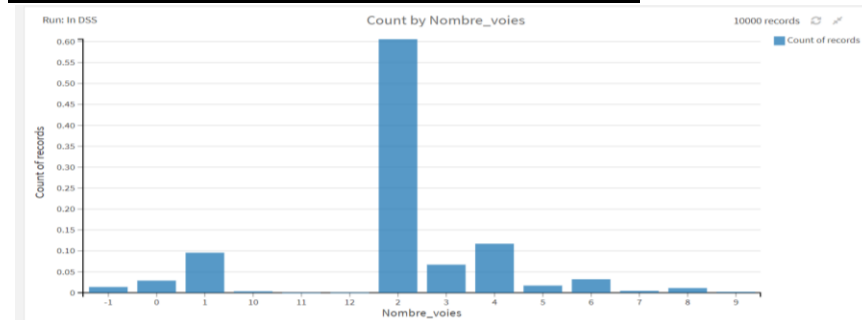


Figure 16 : Nombre de lieux d'accidents selon le nombre de voies.

- On remarque qu'environ 60% des accidents ont eu lieu sur des routes à 2 voies et bidirectionnelles.

- Le nombre d'accidents selon la surface de la route :

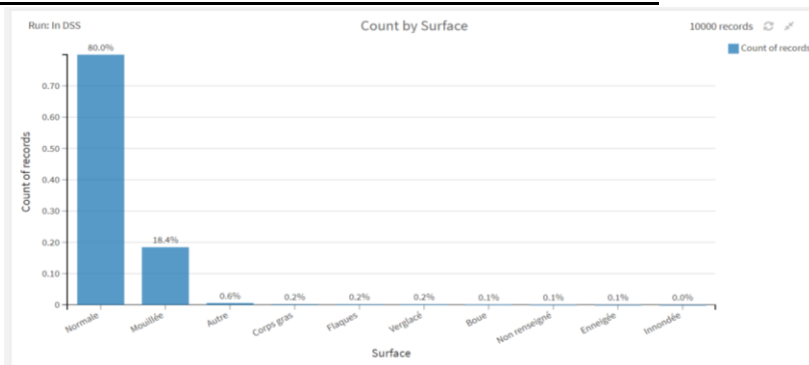


Figure 17 : Nombre d'accident selon le type de surface du lieu de l'accident.

- On remarque que 80% des accidents ont lieu sur des routes avec une surface normale. Et environs 18% des accidents ont lieu sur des routes mouillées.

- Le nombre d'accidents selon l'infrastructure :

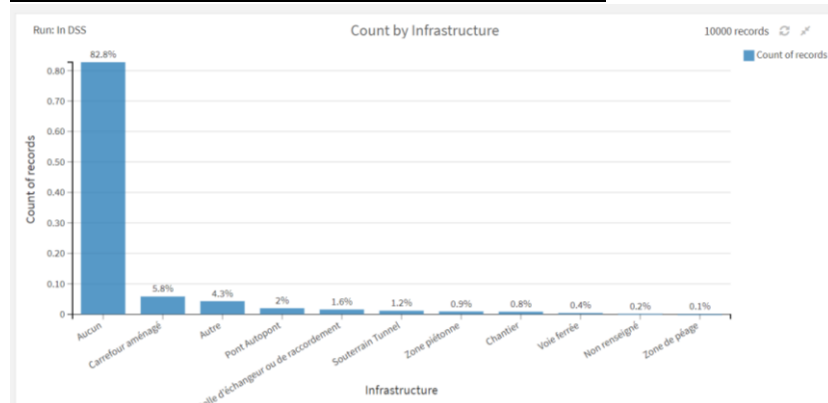


Figure 18 : Nombre d'accident selon le type de l'infrastructure sur le lieu de l'accident.

- On remarque qu'il a 5.8% d'accidents sur les carrefours aménagés et environs 2% sur des ponts. Et plus de 80% des accidents se sont produits sur des routes où il n'y avait pas d'infrastructure particulière.

- Le nombre d'accidents selon la vitesse maximale autorisée :

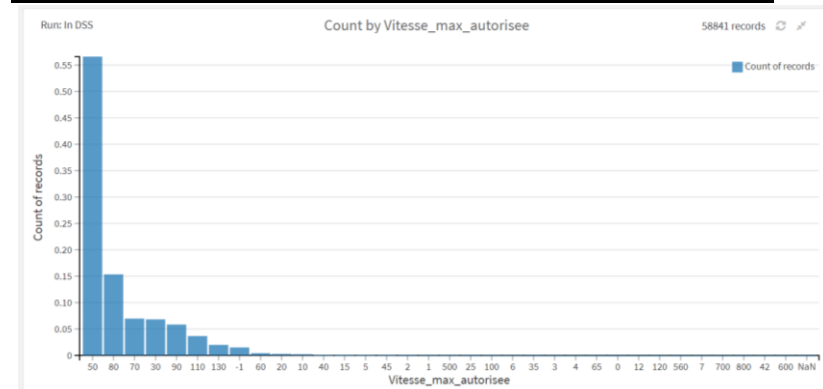


Figure 19 : Nombre d'accidents selon la vitesse maximale autorisée.

- Il y a 55% d'accidents sur les routes où la vitesse est limitée à 50 Km/h.
- Il y a 15% d'accidents sur les routes où la vitesse est limitée à 80 Km/h.
- Il y a environs 7% d'accidents sur les routes où la vitesse est limitée à 70 ou 30 ou 90 Km/h.

Analyser le temps où se produisent les accidents (Dimension Temps) :

- Le nombre d'accidents par heure :

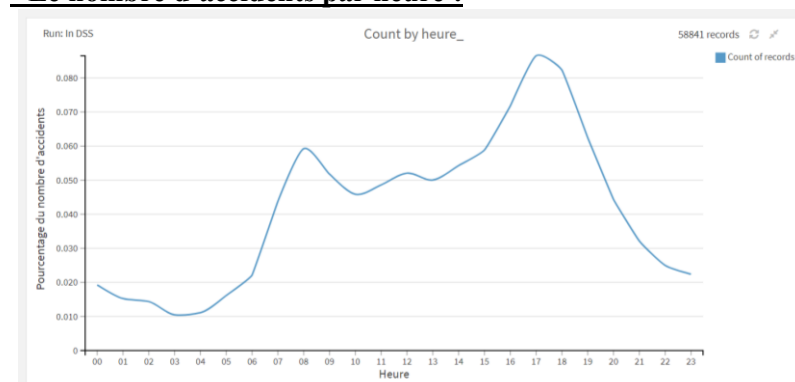


Figure 20 : Nombre d'accidents par heure.

- On remarque qu'il y a deux pics bien visibles. Le premier pic se situe entre 7h et 10h du matin et l'autre pic entre 16h et 19h. Ces deux pics correspondent aux heures de pointes où le trafic est dense.

- Le nombre d'accidents par mois :

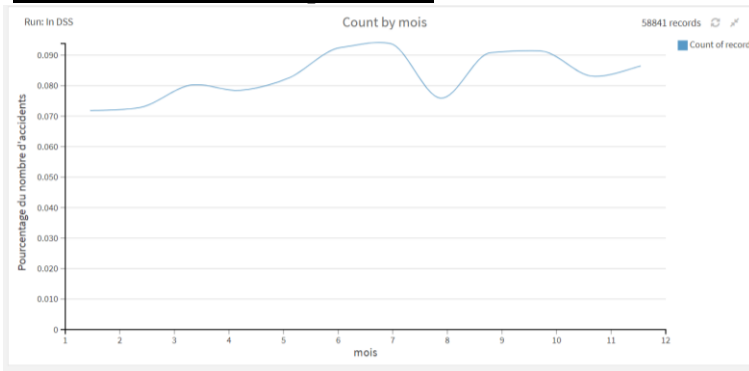


Figure 21 : Nombre d'accidents par mois.

- On remarque un nombre assez important d'accident élevé en été pour les mois de juin et juillet dans les environs de 20% et moins en août. Le nombre d'accident augmente avec la rentrée scolaire en septembre et octobre puis baisse un peu en novembre et remonte vers la fin de l'année en décembre.

- Le nombre d'accidents par jour de fête :

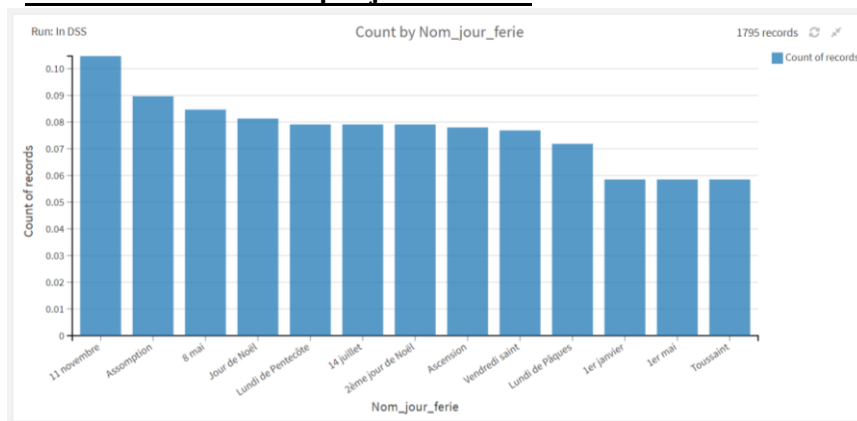


Figure 22 : Nombre d'accidents selon les jours de fête.

- On remarque que le 11 novembre détient le plus grand nombre d'accident comparé aux autres jours de fêtes avec un pourcentage de plus de 10%.
- Et les autres jours de fête ont un pourcentage du nombre d'accident qui est entre 9% et 5%.

- Le nombre d'accidents en jour de fête et hors jour de fête :

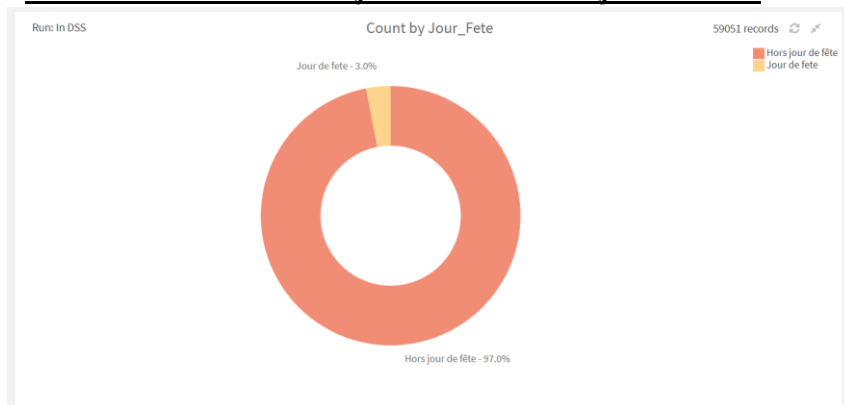
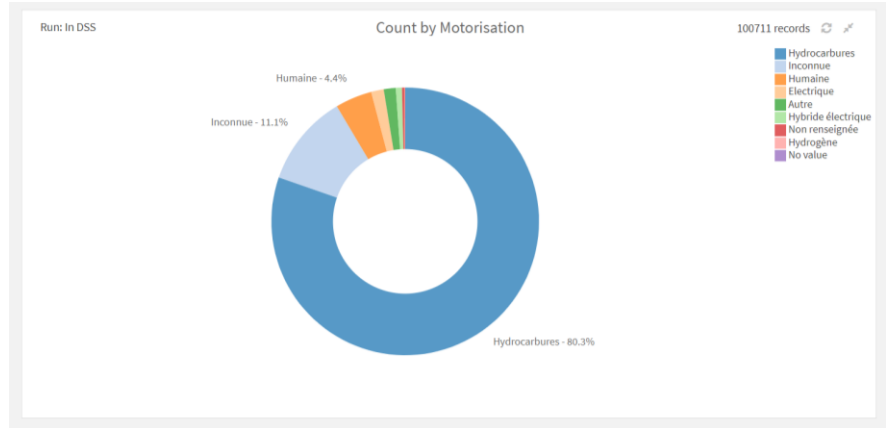


Figure 23 : Nombre d'accidents en jour de fête et hors jour de fête.

- On remarque que 97% des accidents se sont produits hors jour de fête et seulement 3% en jour de fête ce qui est dû à cause de la diminution du trafic routier ces jours-là.

Analyser les caractéristiques du véhicule impliqué dans l'accident :

- Le nombre d'accidents selon la motorisation :



Figure

24 : Nombre de véhicules accidentés selon le type de motorisation.

- 80% des véhicules accidentés roulent à l'hydrocarbure, ce qui représente le type de motorisation le plus commun.
- 1.5% des véhicules accidentés sont Electriques et seulement 0.7% sont hybrides.

- Le nombre d'accidents selon la motorisation et la gravité :

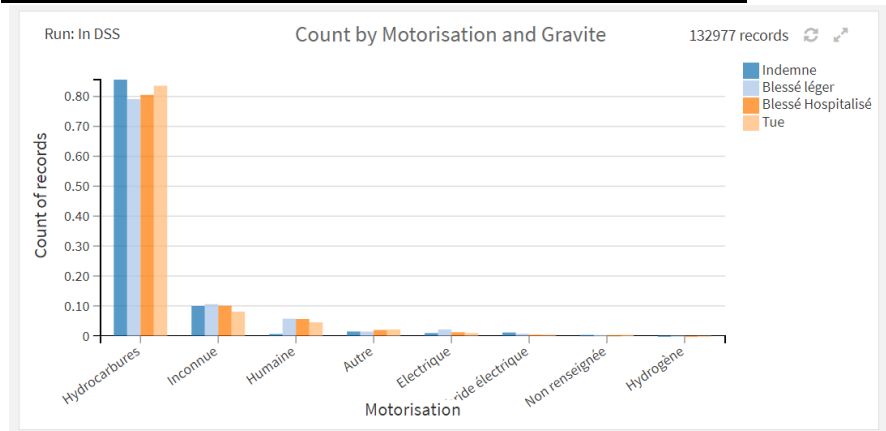


Figure 25 : Nombre d'accidents selon le type de motorisation du véhicule et le niveau de gravité de l'accident.

- On remarque que les niveaux de gravité de l'accident sont distribués uniformément pour le type Hydrocarbures qui sont autour des 80% par rapport à l'ensemble des accidents.
- Pour le type de motorisation électrique on peut remarquer qu'il y a plus de blessés légers dans cette catégorie.

Modèle de Machine Learning :

On va analyser quelques caractéristiques des accidents pour tirer des clusters :

On utilisera les attributs correspondant à l'identifiant de l'accident, le type de lumière, l'agglomération, l'atmosphère, la catégorie et la surface de la route. Ce qui va nous donner la table suivante :

Id_accident	lumiere	agglomeration	atmosphere	Categorie_route	Surface
-------------	---------	---------------	------------	-----------------	---------

Figure 26 : Attributs de la table pour l'analyse des K-moyennes.

On va utiliser l'algorithme des K-moyennes :

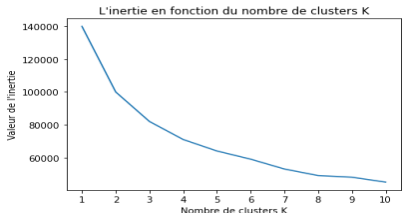


Figure 27 : La méthode du coude en utilisant l'inertie (Elbow methode)

- On remarque qu'à partir de K=4 clusters l'inertie diminue de façon linéaire, donc d'après la méthode d'Elbow le nombre optimal de clusters est 4.

Analyse des Clusters avec K=4 :

Cluster 1 (29.02% sur l'ensemble des données) :

- 100% du cluster a *Hors agglomération* pour la variable **agglomeration** (contre 35.41 % sur l'ensemble des données).
- 52.43% du cluster a *Route Départementale* pour la variable **Categorie_route** (contre 35.63 % sur l'ensemble des données).
- 92.77% du cluster a *Normale* pour la variable **atmosphere** (contre 79.07 % sur l'ensemble des données).

Cluster 2 (19.13% de l'ensemble des données) :

- 74.44% du cluster a *Route Départementale* pour la variable **Categorie_route** (contre 35.63 % sur l'ensemble des données).

- 100% du cluster a **En agglomération** pour la variable **agglomeration** (contre 64.59 % sur l'ensemble des données).
- 96.31% du cluster a **Normale** pour la variable **Surface** (contre 79.79 % sur l'ensemble des données).

Cluster 3 (35.32% de l'ensemble des données) :

- 100% du cluster a **Voie Communale** pour la variable **Categorie_route** (contre 43,80% sur l'ensemble des données)
- 100% du cluster a **En agglomération** pour la variable **agglomération** (contre 64,59% sur l'ensemble des données)
- 95,70% du cluster a **Normale** pour la variable **Surface** (contre 79,79% sur l'ensemble des données)

Cluster 4 (15.60% de l'ensemble des données) :

- 67,64% du cluster a **Pluie légère** pour la variable **atmosphère** (contre 11,21% sur l'ensemble des données)
- 98,69% du cluster a **Mouillée** pour la variable **Surface** (contre 18,62% sur l'ensemble des données)
- 48,82% du pôle a **Plein jour** pour la variable **lumière** (contre 66,28% sur l'ensemble des données)

Outlier Cluster (0.93% de l'ensemble des données) :

- 97,99% du cluster a **Nuit sans éclairage public** pour la variable **lumière** (contre 10,11% sur l'ensemble des données)
- 56,67% du cluster a **Route Départementale** pour la variable **Categorie_route** (contre 35,63% sur l'ensemble des données)
- 43,33% du cluster a **Hors agglomération** pour la variable **agglomération** (contre 35,41% sur l'ensemble des données)

On distingue bien 4 clusters :

- Cluster 1 : Des accidents hors agglomération sur une route départementale avec un temps atmosphérique normal
- Cluster 2 : Des accidents en agglomération sur une route départementale avec une surface de route normale.
- Cluster 3 : Des accidents en agglomération sur une voie communale avec une surface normale.
- Cluster 4 : Des accidents en temps de pluie légère, une route mouillée et en pleine journée.
- Cluster aberrant : Des accidents hors agglomération sur une route départementale en pleine nuit et sans éclairage public.

Synthèse :

Facteurs favorisant les accidents :

- La conduite sur une surface de route mouillée : 18% des accidents.
- La conduite en pleine nuit sans éclairage public : 10% des accidents.
- Intersection en X, T, Y : Environ 10% des accidents.
- La conduite en mauvais temps (Forte pluie, Pluie légère, Temps couvert) : Environ 17% des accidents.
- Trajet de promenade ou de loisir et trajet entre le domicile et le travail : 39% des accidents pour les trajets de promenades et de loisirs et 13% pour les trajets entre le domicile et le travail.
- La forte circulation dans les grandes métropoles (En agglomération) : 64% des accidents.
- La conduite dans une route à double sens (2 Voies au minimum) : Environ 60% des accidents.
- Le non-respect des limitations de vitesses : 55% d'accidents sur les routes où la vitesse est limitée à 50Km/H, 15% d'accidents sur les routes où la vitesse est limitée à 80Km/H et environ 7% d'accidents sur les routes où la vitesse est limitée à 70 ou 30 ou 90Km/H.
- La conduite en heure de pointe : Le premier pic se situe entre 7h et 10h du matin et l'autre pic entre 16h et 19h.

Prévention :

- Plus de 50% des usagers qui mettent une ceinture sont indemnes et plus de 31% ont des blessures légères donc l'utilisation de la ceinture reste une nécessité pour les usagers des voitures en plus de l'airbag et du dispositif de sécurité pour les enfants.
- Environ 60% de ceux qui mettent un casque et des gants sont indemnes et moins de 30% sont hospitalisés et uniquement 3% en sont morts, c'est pourquoi l'utilisation d'un casque et des gants sont une obligation pour les conducteurs de deux-roues.
- Conduire doucement et avec vigilance en pleine nuit ou en mauvais temps.
- Faire attention dans les intersections et les routes à double sens.
- Environ 39% des accidents arrivent lors de trajets de promenade ou de loisirs et c'est souvent dans des endroits nouveaux où l'on ne connaît pas bien la route, c'est pourquoi il faut toujours rester vigilant dans ce genre de situation et même dans le cas de trajet entre le domicile et le travail qui est à priori un itinéraire qu'on connaît très bien, on a recensé environ 13% d'accidents qui concernent ce trajet.
- Respecter les limitations de vitesse.
- Eviter le plus possible la conduite en heure de pointe et les routes à forte circulation.