

De la donnée aux bases de données

Projet final: sécurité routière 2015

Isabella RAEVEL, Anais STORP

13 Mai 2024



Table des matières

1	Introduction	2
2	Dimension spatiale	2
2.1	Sur une année, quelle est la répartition de nos deux indicateurs NA et NT selon les départements?	2
2.2	Peut-on réellement comparer les chiffres brut? Que faire pour y remédier?	2
2.3	Comment se comportent ces indicateurs normalisés dans le temps?	4
2.4	Y a -t-il dans notre jeu de données des éléments qui pourraient expliquer ce classement?	5
3	La dimension du genre	7
3.1	Comment les indicateurs NA2 et NT se répartissent selon le sexe du conducteur?	7
3.2	NA2 étant le nombre de véhicule impliqués dans un accident puisqu'il peut y avoir plusieurs conducteurs impliqué dans le même accident	8
3.3	Est-ce que cela est stable dans le temps et dans l'espace?	8
3.4	Y-a-t-il dans nos données des éléments qui pourraient expliquer cette différence?	10
4	Conclusions et perspectives	10

1 Introduction

Dans le cadre du projet final pour le cours "De la donnée aux bases de données," notre binôme a été chargée d'analyser un jeu de données détaillé concernant la sécurité routière de l'année 2015. L'objectif principal de ce projet est de fournir une analyse approfondie des accidents corporels et des décès survenus sur les routes, en se concentrant spécifiquement sur l'année qui nous a été assignée. Ce rapport PDF, accompagné de notre notebook Jupyter nettoyé, présente notre approche méthodologique, les résultats obtenus, ainsi que des visualisations graphiques pour appuyer nos analyses.

2 Dimension spatiale

2.1 Sur une année, quelle est la répartition de nos deux indicateurs NA et NT selon les départements?

L'analyse des données collectées sur une année révèle une corrélation notable entre le nombre d'accidents corporels (NA) et le nombre de tués (NT) à l'échelle des départements. Cette observation suggère que les départements ayant un taux élevé d'accidents corporels tendent à présenter un nombre significatif de cas mortels.

Afin de préciser l'analyse, il est observé que les départements nommés Bouches-du-Rhône (13), Gironde (33), Oise (60), Rhône (69), Paris (75), Hauts-de-Seine (92), Seine-Saint-Denis (93), et Val-de-Marne (94) présentent une répartition corrélée entre deux indicateurs spécifiques. Cette observation souligne une tendance dans ces zones géographiques particulières.

Il est crucial de poursuivre les analyses pour comprendre les nuances régionales et locales qui pourraient affecter ces indicateurs.

Cette tendance Pourrait être expliquée par plusieurs facteurs:

- *Densité de la circulation* : Les départements avec une densité de circulation plus élevée sont susceptibles d'avoir plus d'accidents et, par conséquent, plus de cas mortels.
- *Infrastructure routière* : Les infrastructures mal conçues ou mal entretenues peuvent augmenter la gravité des accidents.
- *Disponibilité et temps de réponse des services d'urgence* : Un temps de réponse plus long aux accidents peut augmenter le taux de mortalité dans les accidents corporels.

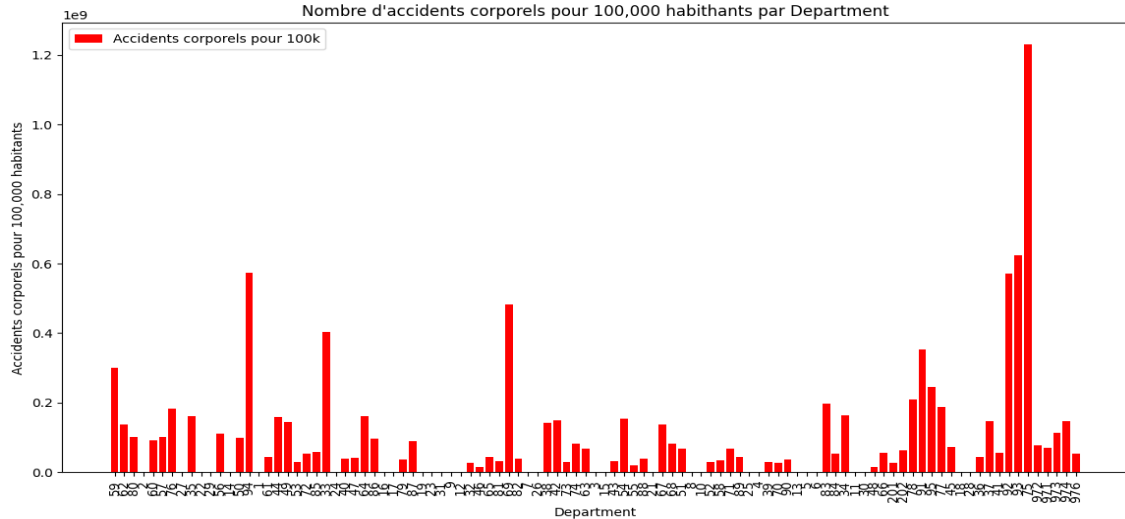
Afin de confirmer ces facteurs hypothétiques, il serait pertinent de calculer les taux de corrélations de ces facteurs. Cependant, dans notre jeu de données ces informations ne sont pas présentes.

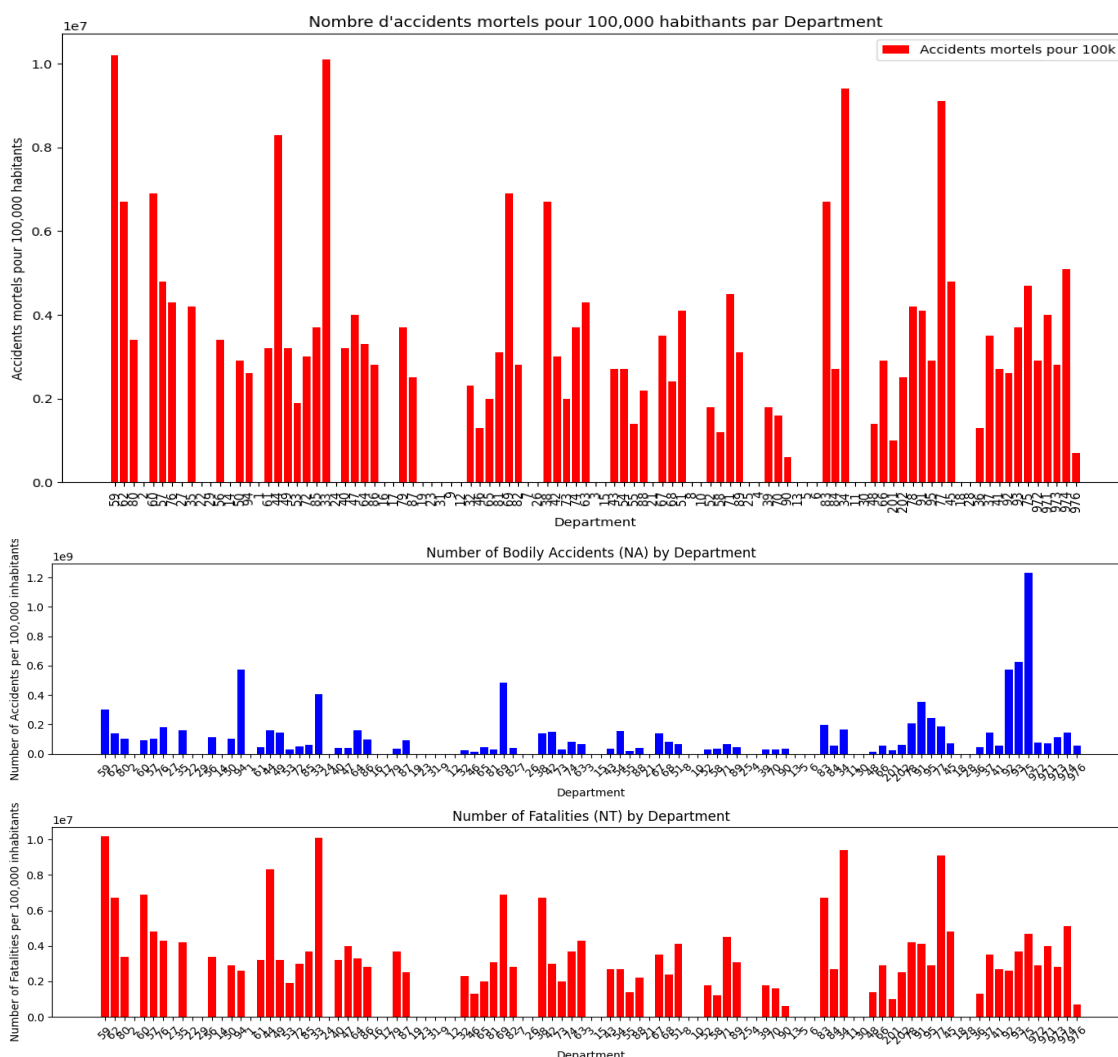
2.2 Peut-on réellement comparer les chiffres brut? Que faire pour y remédier?

Comparer directement les chiffres bruts d'une année précise peut mener à des conclusions erronées, notamment parce que les variations dues à des facteurs externes peuvent influencer les résultats. Pour améliorer la comparaison entre les données de 2015 et 2016, voici quelques stratégies :

1. *Normalisation des Données* : Calculer les taux par habitant ou selon un autre dénominateur pertinent permet de standardiser les données, facilitant ainsi une comparaison équitable entre les années.
2. *Indexation sur une Année de Base* : Choisir une année de base et indexer toutes les données en fonction de cette année peut aussi aider à faire des comparaisons plus équitables, en particulier pour les données économiques ou démographiques.
3. *Analyse de Tendance sur Plusieurs Années* : Examiner les tendances sur un horizon temporel étendu aide à identifier les motifs ou anomalies spécifiques et à lisser les variations annuelles potentiellement atypiques.

Suite à une analyse approfondie des données normalisées, il est possible de constater des corrélations nuancées entre les indicateurs d'accidents non-mortels (NA) et ceux mortels (NT) au sein de certains départements. Plus spécifiquement, les départements du Val-de-Marne (94), de la Gironde (33), du Rhône (69), de Paris (75), des Hauts-de-Seine (92), et de la Seine-Saint-Denis (93) montrent une tendance où le nombre d'accidents non-mortels est proportionnel à celui des accidents mortels. Cette observation indique que, dans ces départements, les facteurs de risque ou les conditions de sécurité routière influencent de manière similaire les incidents entraînant des blessures et ceux résultant en des décès. La normalisation des données, essentielle pour garantir une comparaison équitable entre les départements de tailles démographiques différentes, permet de mettre en évidence ces patterns de manière plus précise et significative.

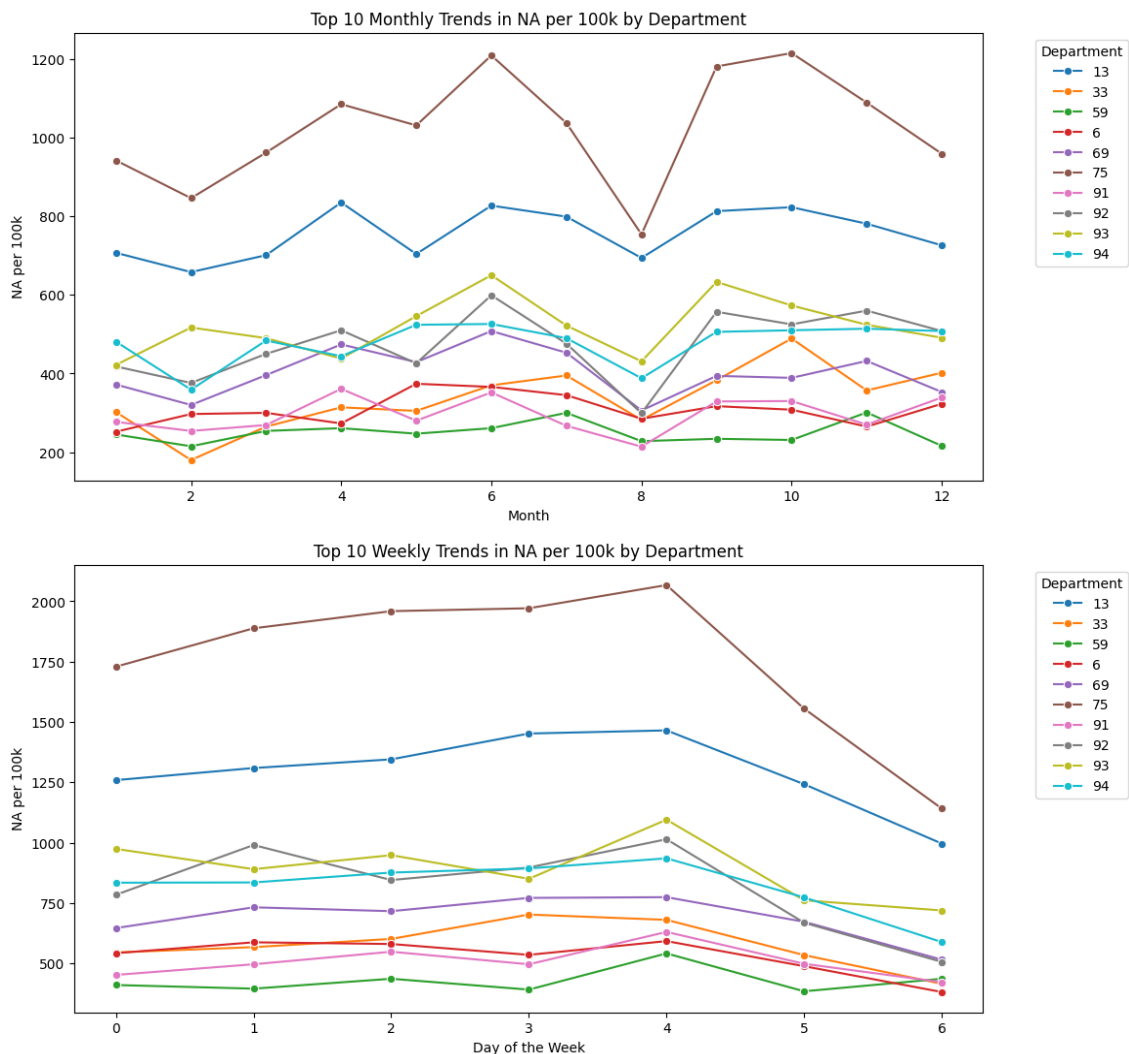




2.3 Comment se comportent ces indicateurs normalisés dans le temps?

En examinant les indicateurs normalisés d'accidents pour les dix départements ayant les meilleurs et les pires taux, nous observons des tendances distinctes en fonction des saisons et des jours de la semaine. Le mois d'août présente généralement moins d'accidents, probablement en raison de la baisse du trafic durant les vacances d'été. À l'opposé, les mois de septembre, octobre et juin enregistrent une hausse des accidents, coïncidant avec la reprise des activités régulières et des conditions météorologiques parfois défavorables.

Sur une base hebdomadaire, le nombre d'accidents augmente jusqu'à atteindre un pic le vendredi, moment où le trafic est dense en raison des déplacements du week-end et de la fin de la semaine de travail. Cette périodicité suggère la nécessité de mesures de sécurité routière renforcées pendant les périodes de forte activité et de flux de trafic, notamment en fin de semaine, pour prévenir efficacement les accidents.



2.4 Y a -t-il dans notre jeu de données des éléments qui pourraient expliquer ce classement?

Dans notre analyse des données relatives aux accidents de la route collectées pour chaque département, plusieurs facteurs sont susceptibles d'expliquer les variations des indicateurs NA et NT. Ces facteurs comprennent, mais ne se limitent pas à, le taux d'utilisation des équipements de sécurité, les caractéristiques démographiques telles que l'âge et le sexe des conducteurs, ainsi que d'autres conditions environnementales et comportementales lors des accidents.

Taux d'Utilisation des Équipements de Sécurité Une corrélation pourrait être observée entre le faible taux d'utilisation des équipements de sécurité (ceintures de sécurité, casques de moto, etc.) et les taux élevés d'accidents corporels et de mortalité. Les départements avec des politiques strictes et une application rigoureuse concernant l'utilisation des équipements de sécurité tendent à avoir des indicateurs NA et NT plus bas.

Caractéristiques Démographiques

L'âge et le sexe des conducteurs impliqués dans les accidents offrent des insights précieux:

- **Âge** : Les jeunes conducteurs, souvent moins expérimentés, pourraient être impliqués dans des accidents plus graves, ce qui influencerait les taux d'accidents

corporels et de mortalité dans certains départements.

- **Sexe** : Les études montrent des différences de comportement au volant entre les sexes, ce qui pourrait également influencer les statistiques d'accidents.

Illustration par des Graphiques

Les données peuvent être illustrées par des graphiques qui montrent la relation entre l'utilisation des équipements de sécurité et les taux d'accidents, ou les distributions des accidents par âge et sexe. Par exemple:

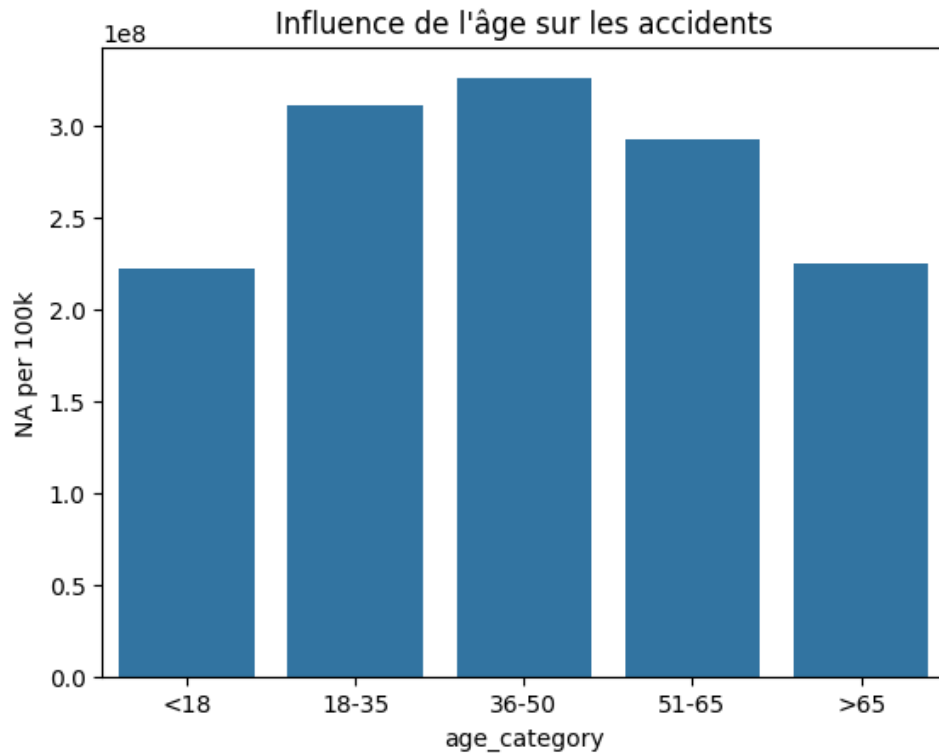


Figure 1: Corrélation entre la tranche d'âge de l'utilisateur et les taux d'accidents

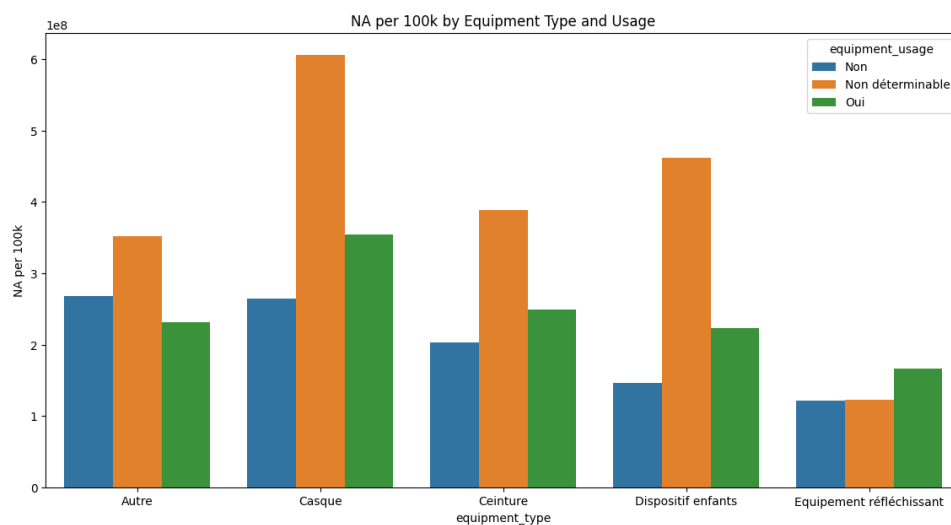


Figure 2: Corrélation entre l'utilisation des équipements de sécurité et les taux d'accidents

En conclusion, bien que ces facteurs offrent des explications potentielles pour les variations observées entre les départements, des études supplémentaires sont nécessaires pour confirmer ces tendances et aider à formuler des politiques de prévention plus efficaces.

3 La dimension du genre

3.1 Comment les indicateurs NA2 et NT se répartissent selon le sexe du conducteur?

La moyenne du nombre de véhicules impliqués dans un accident est plus élevée chez les hommes que chez les femmes. De même, le nombre de morts par accident est plus élevé chez les conducteurs masculins que chez les conductrices féminines.

Illustration des Disparités entre les Sexes

Cette différence peut être illustrée par le graphique suivant, qui montre le nombre moyen de véhicules impliqués par accident et le nombre moyen de décès par accident, ventilés par sexe.

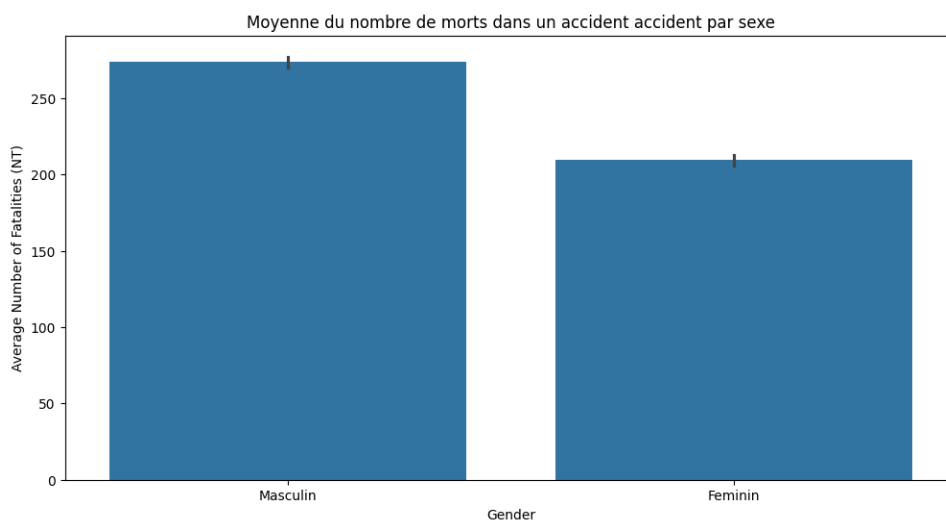


Figure 3: Moyenne du nombre de morts par accident, ventilés par sexe.

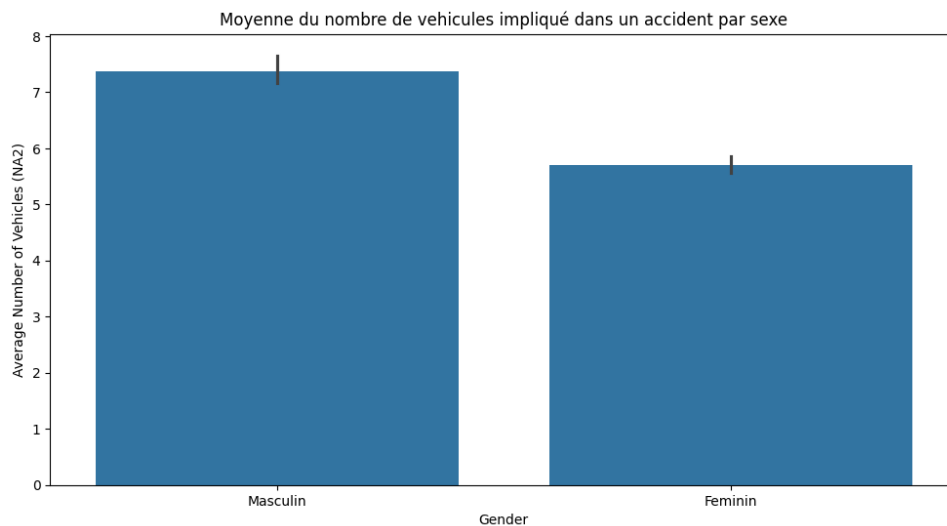


Figure 4: Moyenne du nombre de véhicules impliqués par accident, ventilés par sexe.

Ces différences pourraient s’expliquer par divers facteurs comportementaux et sociaux. Des études ont montré que les hommes ont tendance à adopter des comportements de conduite plus risqués que les femmes, ce qui pourrait contribuer à un taux plus élevé d’implication dans des accidents graves. De plus, les hommes sont statistiquement plus enclins à conduire des véhicules plus grands ou plus puissants, ce qui pourrait aussi influencer la gravité des accidents.

3.2 NA2 étant le nombre de véhicule impliqués dans un accident puisqu’il peut y avoir plusieurs conducteurs impliqué dans le même accident

Dans le cadre de nos analyses sur la sécurité routière, l’indicateur **NA2** est défini comme le nombre de véhicules impliqués dans un seul accident. Cet indicateur est crucial pour comprendre la complexité et la gravité des accidents de la route. Il nous permet d’évaluer non seulement l’étendue des accidents mais aussi d’identifier des modèles dans les accidents multi-véhicules, qui sont souvent plus susceptibles de résulter en blessures graves ou en mortalité.

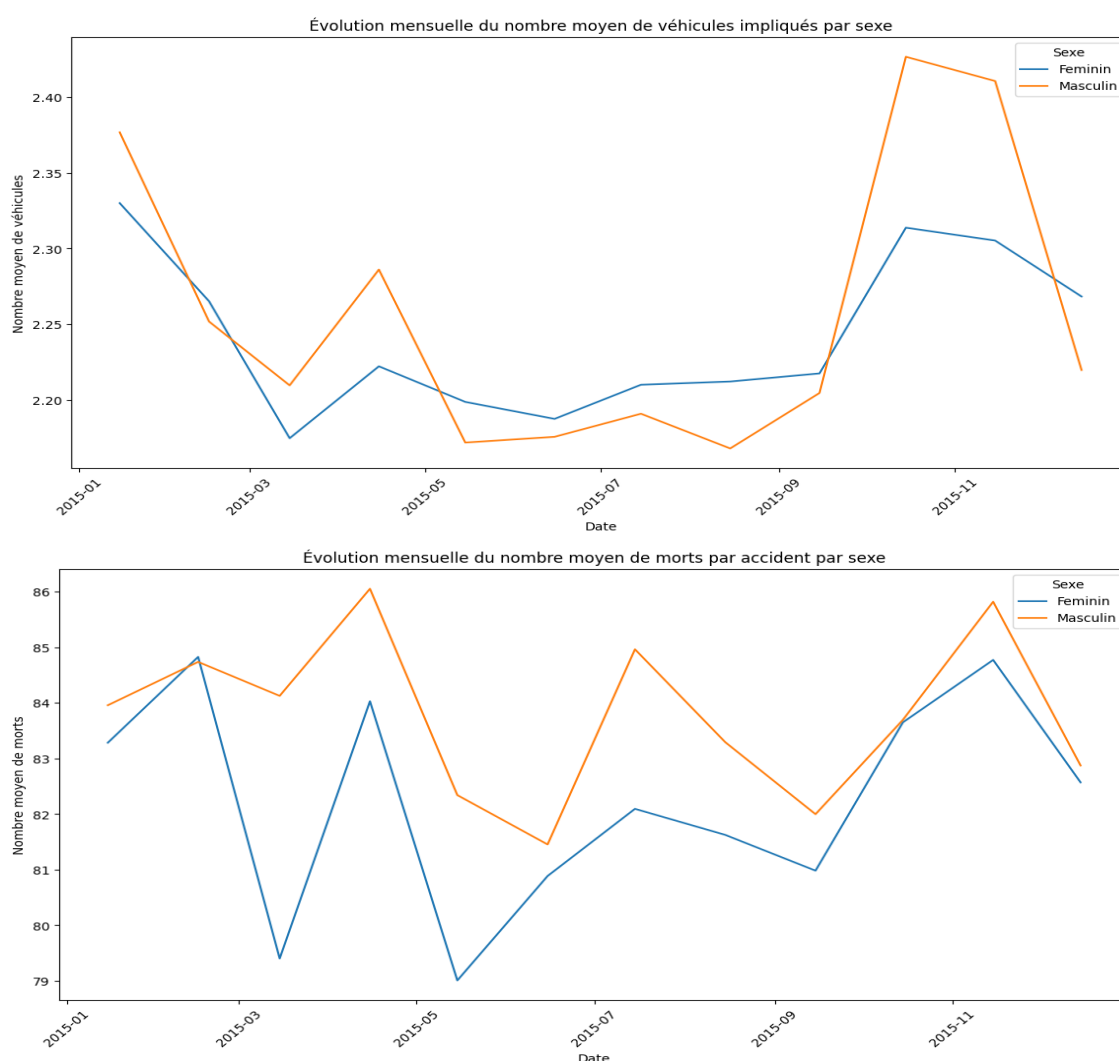
3.3 Est-ce que cela est stable dans le temps et dans l’espace?

L’analyse des données sur les accidents de la route révèle des tendances distinctes qui varient en fonction du sexe des conducteurs tout au long de l’année. Voici un résumé détaillé de ces observations :

- **Janvier à Mars** : Durant cette période, nous observons une baisse générale du nombre d’accidents tant pour les conducteurs masculins que féminins. Cette diminution pourrait être attribuée à des conditions météorologiques plus rigoureuses, qui incitent à une conduite plus prudente.
- **Mi-Avril** : Un pic significatif survient, particulièrement marqué chez les hommes. Cette hausse des accidents peut coïncider avec le début des meilleures conditions

météorologiques, encourageant une augmentation des déplacements et potentiellement une conduite moins prudente.

- **Mai à Août** : Le nombre d'accidents diminue et reste à des niveaux bas pour les deux sexes. Cette période peut correspondre à une acclimatation aux conditions estivales et une attention plus grande portée à la sécurité routière.
- **Septembre** : Nous constatons un pic important, surtout pour les hommes, avec un pic moins prononcé chez les femmes. Cette augmentation pourrait être liée à la rentrée scolaire et à une augmentation globale de l'activité routière.
- **Fin d'Année** : Le nombre d'accidents décroît pour les deux sexes. Cette baisse pourrait être due à une conduite plus prudente durant les mois d'hiver, lorsque les conditions de route sont souvent les plus difficiles.



Ces observations suggèrent que des facteurs saisonniers, tels que les conditions météorologiques et les variations des habitudes de conduite liées à des événements calendaires, jouent un rôle significatif dans la fréquence des accidents de la route. Des campagnes de sensibilisation ciblées pendant les périodes de pic pourraient aider à réduire le nombre d'accidents.

3.4 Y-a-t-il dans nos données des éléments qui pourraient expliquer cette différence?

Dans l'analyse des données relatives aux accidents de la route, plusieurs facteurs peuvent être considérés pour expliquer les variations des indicateurs NA (Nombre d'Accidents corporels) et NT (Nombre de Tués) entre différents départements ou périodes. Nous examinons ci-dessous quelques-uns de ces éléments clés :

Caractéristiques des Infrastructures Routières

Les différences dans la qualité et le type d'infrastructures routières entre les régions peuvent influencer significativement le nombre et la gravité des accidents. Par exemple, des routes mal entretenues ou des zones à forte circulation sans aménagements adéquats pour les piétons peuvent augmenter le risque d'accidents graves.

Conditions Météorologiques et Environnementales

Les conditions météorologiques extrêmes, comme le brouillard, la neige, ou la pluie intense, peuvent augmenter la probabilité d'accidents dans certaines régions. De même, la période de l'année (saisons) influence souvent le nombre d'accidents en raison de la variation des conditions de conduite.

Démographie des Conducteurs

L'âge, le sexe, et l'expérience des conducteurs sont des facteurs démographiques qui peuvent affecter le comportement au volant et donc les taux d'accidents. Les jeunes conducteurs, par exemple, sont souvent plus susceptibles d'être impliqués dans des accidents en raison de leur manque d'expérience.

Ces facteurs peuvent être analysés à l'aide de méthodes statistiques avancées pour déterminer leur impact potentiel sur les taux d'accidents et de mortalité observés dans nos données. Des études supplémentaires pourraient inclure des analyses de régression, des études de corrélation, ou des modèles prédictifs pour quantifier ces effets de manière plus précise.

4 Conclusions et perspectives

Au terme de notre étude sur la sécurité routière, nous avons identifié plusieurs facteurs influençant les taux d'accidents et de mortalité sur les routes. L'analyse a mis en lumière l'impact des conditions saisonnières, notamment comment les variations météorologiques affectent la fréquence des accidents. De plus, les données indiquent que les caractéristiques démographiques telles que l'âge et le sexe jouent un rôle crucial dans les dynamiques d'accidents.

Cependant, notre étude comporte des limites notables. Elle repose sur des données déclaratives qui peuvent souffrir de problèmes de précision, incluant des biais de sous-déclaration ou de sur-déclaration. En outre, notre analyse n'a pas pu prendre en compte des variables psychologiques et comportementales profondes, qui nécessitent des approches qualitatives pour une meilleure compréhension du comportement des conducteurs face aux risques routiers.

Pour approfondir notre compréhension, il est essentiel d'accéder à des données supplémentaires. Des informations détaillées sur les comportements de conduite à risque, obtenues via des dispositifs de suivi en véhicule ou des enquêtes post-accident, enrichiraient notre base de données. De même, des études sur les attitudes et perceptions des

conducteurs vis-à-vis des règles de sécurité routière pourraient éclairer les motivations derrière les comportements dangereux. De plus, des données horaires et des informations sur l'efficacité des campagnes de sensibilisation permettraient de cerner plus précisément les moments et les méthodes d'intervention les plus efficaces.