

Rapport de projet SAS

Groupe : Louis Coste, Romain Dudoit, Franck DORONZO

Table des matières

| | |
|---|----|
| I) Introduction | 2 |
| II) Analyse de la sévérité des accidents par pays..... | 4 |
| III) Analyse de la sévérité des accidents en fonction de l'âge du conducteur..... | 6 |
| IV) Sévérité victime – sévérité accident | 11 |
| V) Conclusion | 11 |

I) Introduction

La sécurité routière est un sujet de plus en plus présent dans les discours des gouvernements et dans la pensée collective. Cette évolution est amenée par la constatation du nombre de morts sur les routes. En 2018, 3488 personnes ont perdu la vie sur les routes de France soit une baisse de 5.3% par rapport à 2017. Si cette progression est rassurante, le nombre de mort sur les routes reste beaucoup trop élevé. De fait, nos voisins d'outre-manche ont quant à eux beaucoup moins de mort sur les routes. En effet, les Anglais ne comptaient que 1710 victimes sur leurs routes. A l'instar de la France, ce chiffre est en baisse de 5% par rapport à l'année précédente et de 47.8% entre 2001 et 2016. Le nombre de morts peut notamment s'expliquer par l'amélioration des performances des véhicules en circulation. Ceux-ci allant de plus en plus vite sur des routes par toujours entretenu notamment en campagne. D'autres facteurs entre aussi en compte comme l'alcool au volant, la consommation de drogues ou encore la fatigue.

Mais outre le nombre de morts, beaucoup d'accidents ne font heureusement que des blessés. Cependant ceux-ci entrent aussi dans les statistiques fournis chaque année par les différents organismes gouvernementaux en charge de la sécurité routière. Ceux-ci fournissent également des jeux de données anonymisés chaque année sur les accidents de la route dans leurs pays respectifs. En regardant les datasets et les statistiques fournis par l'organisme anglais de la sécurité routière, on a pu se rendre compte que la plupart sont calculés à l'échelle nationale. En effet, nous n'avons pas réussi à trouver un rapport regroupant les statistiques pour chaque région de la Grande Bretagne. De même, les statistiques avancées par ces organismes prennent en compte le nombre d'accidents et leurs sévérités uniquement. Ils n'ont pas, à notre connaissance, croisé ces données avec le type de véhicule, piéton ou cycliste présent dans l'accident. Pour résumer, les statistiques avancées restent assez vagues sur les circonstances des accidents survenus sur le territoire anglosaxon. Si effectivement il serait ambitieux de notre part de vouloir effectuer des statistiques sur l'ensemble des critères du jeu de données nous allons tout de même essayer d'effectuer des analyses en corrélant des variables qui nous semblent cohérentes à rassembler.

Pour effectuer notre analyse, nous avons importé 9 fichiers de *Department of Transport of the United Kingdom Government*. Ces fichiers représentent les données de 2017 à 2019 chacun découpé en trois fichiers :

- ACCIDENT_20XX
- CASUALTIES_20XX
- VEHICLES_20XX

Chaque fichier possède un nombre précis de colonne précisé dans le tableau ci-dessous.

| Accident Circumstances | Vehicle | Casualty |
|--|----------------------------------|------------------------------------|
| Accident Index | Accident Index | Accident Index |
| Police Force | Vehicle Reference | Vehicle Reference |
| Accident Severity | Vehicle Type | Casualty Reference |
| Number of Vehicles | Towing and Articulation | Casualty Class |
| Number of Casualties | Vehicle Manoeuvre | Sex of Casualty |
| Date | Vehicle Location-Restricted Lane | Age of Casualty |
| Day of Week | Junction Location | Age Band of Casualty |
| Time | Skidding and Overturning | Casualty Severity |
| Location Easting OSGR | Hit Object in Carriageway | Pedestrian Location |
| Location Northing OSGR | Vehicle Leaving Carriageway | Pedestrian Movement |
| Longitude | Hit Object off Carriageway | Car Passenger |
| Latitude | 1st Point of Impact | Bus or Coach Passenger |
| Local Authority | Was Vehicle Left Hand Drive | Pedestrian Road Maintenance Worker |
| Local Authority | Journey Purpose of Driver | Casualty Type |
| 1st Road Class | Sex of Driver | Casualty IMD Decile |
| 1st Road Number | Age of Driver | Casualty Home Area Type |
| Road Type | Age Band of Driver | |
| Speed limit | Engine Capacity | |
| Junction Detail | Vehicle Propulsion Code | |
| Junction Control | Age of Vehicle | |
| 2nd Road Class | Driver IMD Decile | |
| 2nd Road Number | Driver Home Area Type | |
| Pedestrian Crossing-Human Control | | |
| Pedestrian Crossing-Physical Facilities | | |
| Light Conditions | | |
| Weather Conditions | | |
| Road Surface Conditions | | |
| Special Conditions at Site | | |
| Carriageway Hazards | | |
| Urban or Rural Area | | |
| Did Police Officer Attend Scene of Accident | | |
| Lower Super Output Area of Accident_Location | | |

II) Analyse de la sévérité des accidents par pays

Pour analyser la sévérité des accidents il faut tout d'abord regarder la répartition de ceux-ci. La sévérité des accidents est identifiée par 3 indices énoncé ci-dessous :

1. Fatal
2. Sérieux
3. Léger

Le graph ci-dessous nous offre une première image de cette répartition sur l'ensemble du dataset (2017-2019).

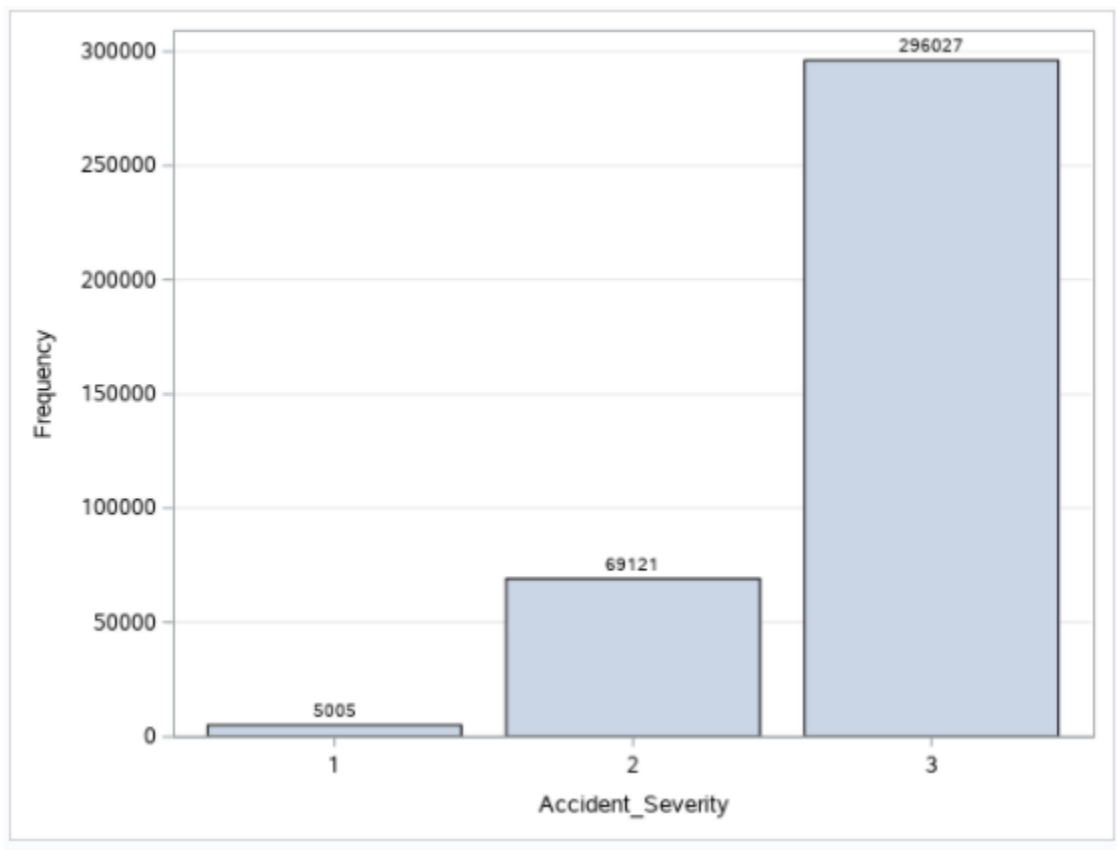


Figure 1. Répartition des accidents par rapport à leur sévérité

| Accident_Severity | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|-------------------|-----------|---------|----------------------|--------------------|
| 1 | 5005 | 1.35 | 5005 | 1.35 |
| 2 | 69121 | 18.67 | 74126 | 20.03 |
| 3 | 296027 | 79.97 | 370153 | 100.00 |

Figure 2. Tableau de fréquence et pourcentage

De ce graphique on s'aperçoit directement que la majorité des accidents sont légers 296027 observations soit 79.97% des accidents recensés. Les accidents sérieux représentent 69121 observations soit 18.67% et les accidents fatals représentent quant à eux 5005 observations soit 1.35% des accidents.

L'ensemble des accidents présenté dans le graph et le tableau ci-dessus ont été recensé dans toute la Grande Bretagne (Angleterre, Pays de Galles et Ecosse). On va donc dans un premier temps regarder la répartition des accidents par pays.

| Country | Sum |
|----------|--------|
| England | 337886 |
| Scotland | 19202 |
| Wales | 13085 |

Figure 3. Répartition des accidents par pays

On peut s'apercevoir la l'inégalité de répartition entre les trois pays de Grande Bretagne. L'Angleterre étant le pays qui recense le plus d'accidents. Cela peut être due au nombre supérieur de régions en Angleterre par rapport à l'Ecosse et au Pays de Galles. De même, on peut aisément imaginer que cela est due au nombre d'habitant par pays. L'Angleterre possède 66 Millions contre 5.4 Millions pour l'Ecosse et 3.1 millions pour le Pays de Galles. Il semble donc logique que l'Angleterre ai plus d'accidents que ses voisins car il possède sans aucun doute plus de monde sur ses routes.

Malgré cela, il serait tout de même intéressant de regarder quelle est la répartition de la sévérité des accidents par pays. On va donc regarder le nombre d'accidents par pays en fonction de leur sévérité.

| Country | n_1 | pourcentage_n1 | n_2 | pourcentage_n2 | n_3 | pourcentage_n3 |
|----------|------|----------------|-------|----------------|--------|----------------|
| England | 4267 | 0.006465 | 61983 | 0.093914 | 271616 | 0.411539 |
| Scotland | 448 | 0.014452 | 4455 | 0.14371 | 14299 | 0.461258 |
| Wales | 290 | 0.00537 | 2683 | 0.049685 | 10112 | 0.187259 |

Figure 4. Nombre d'accidents par pays en fonction de la sévérité et du nombre d'habitant par pays

Les données intéressantes de ce tableau sont les pourcentages. En effet, ils représentent le pourcentage d'accidents par pays en fonction de leurs sévérités et du nombre d'habitant du pays en question. On peut tout de suite remarquer que si l'Ecosse à mon d'accidents en nombre par rapport à l'Angleterre, le pourcentage est quant à lui bien plus élevé. Avec 448 morts pour 5.4 millions d'habitant (0.01%) contre 4267 pour 66 millions (0.006%). Il en va de même pour les accidents de type 2 et 3.

La deuxième analyse que l'ont pu faire est que le Pays de Galles est le pays qui possède les meilleurs statistiques en termes de sécurité routière en fonction du nombre d'habitant avec 0.005% d'accidents fatal, 0.04% d'accidents grave et 0.18% d'accidents léger.

Finalement, de ces données nous allons essayer de générer un score de sécurité routière par pays et ainsi définir quel pays de Grande Bretagne apparait comme étant le plus sûr en termes d'accidents. Pour cela, nous allons utiliser le nombre d'accidents par pays (NA), la sévérité des accidents (SA) et la population totale du pays (NP). La formule de calcul du score sera la suivante :

$$Score = \frac{\sum_{i=1}^{NA} SA}{NP}$$

Cette équation sera donc la somme de la sévérité de tous les accidents d'un pays ramené à la population. Cependant, la pondération de la sévérité des accidents sera inversée. En effet, il nous semble plus intéressant d'appuyer sur les accidents fatals pour évaluer la dangerosité des routes d'un pays. Ainsi les accidents fatals auront une pondération de 5, les accidents grave auront une pondération de 3 et les accidents légers auront une pondération de 1. Les scores calculés se placeront dans une fourchette allant

de 0 à 5. Il est cependant très improbable que ces scores atteignent 5. En effet, Cela signifierait que l'ensemble de la population d'un pays soit mort sur les routes.

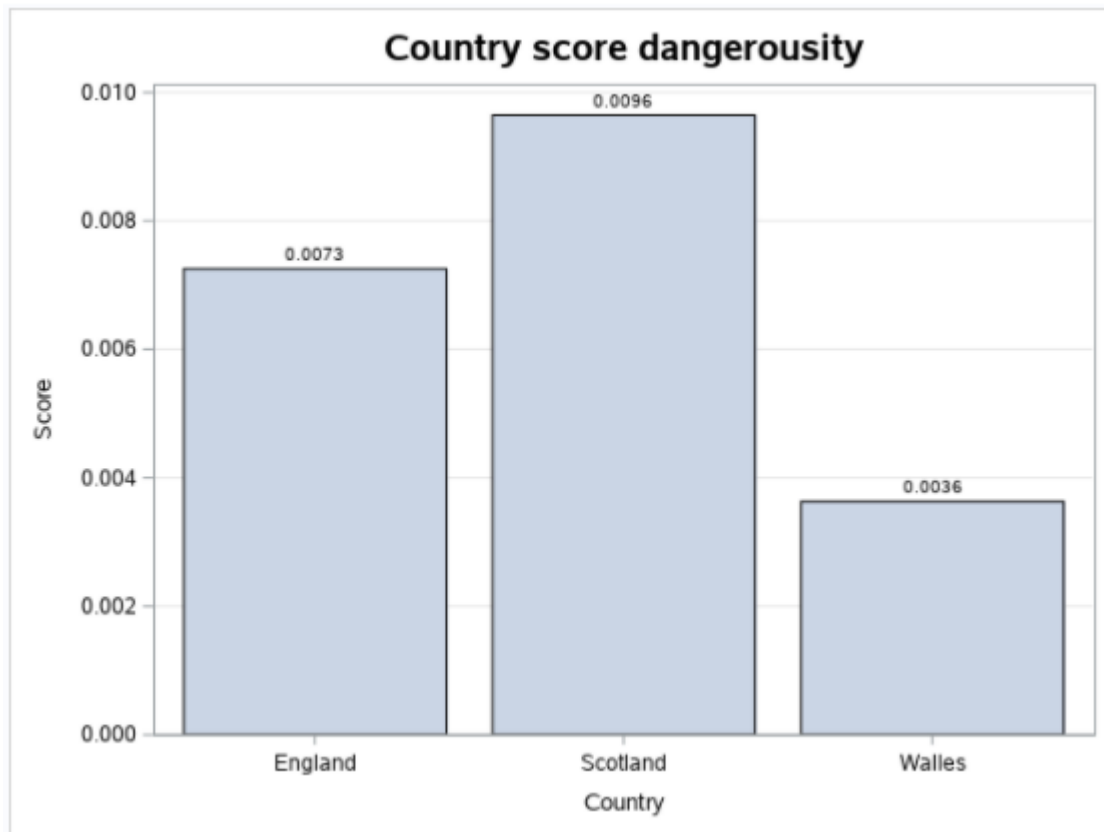


Figure 5. Graphique des Scores de dangerosité des routes par pays

Comme on pouvait s'y attendre après la précédente analyse, le pays le plus dangereux sur les routes de Grande Bretagne est bien l'Ecosse (0.0096), suivi par l'Angleterre (0.0073) et du Pays de Galles (0.0036). On peut donc en conclure qu'il est bien plus sûr de rouler sur les routes du Pays de Galles que sur celle de l'Angleterre et de l'Ecosse. Cependant, il convient de mesurer nos propos. En effet, comme montré sur les précédents tableaux, le taux d'accidents dans chaque pays ne dépasse pas 1% de la population sur l'ensemble des trois années. Ainsi, si cette analyse nous permet de comparer les trois pays de Grande Bretagne, il convient de dire que les routes d'outre-manche reste tout de même globalement sécurisé.

III) Analyse de la sévérité des accidents en fonction de l'âge du conducteur

Une idée très répandue dit que les jeunes et les personnes âgées sont les personnes les plus dangereuses sur la route. Notre objectif va être d'essayer de voir si cette idée est vraie. Pour cela, nous allons utiliser la colonne *Age_of_Driver* et *Age_Band_of_Driver* du dataset Véhicules. Nous allons tout d'abord commencer par observer l'âge moyen des conducteurs et la répartition des accidents par tranche d'âge.

| Analysis Variable : Age_of_Driver | |
|-----------------------------------|------------|
| | Mean |
| | 40.7009594 |

Figure 6. Moyenne d'âge des conducteurs

On peut remarquer que l'âge moyen des conducteurs est de 40 ans. Cela nous donne déjà une première image de la répartition des accidents en fonction de l'âge. Cependant, il nous faut analyser la répartition dans les tranches d'âge pour avoir une image plus pertinente de ces données.

| label | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---------|-----------|---------|----------------------|--------------------|
| 0 - 5 | 128 | 0.02 | 128 | 0.02 |
| 11 - 15 | 5243 | 0.77 | 5371 | 0.79 |
| 16 - 20 | 45018 | 6.60 | 50387 | 7.39 |
| 21 - 25 | 67767 | 9.94 | 118154 | 17.33 |
| 26 - 35 | 145218 | 21.30 | 263370 | 38.63 |
| 36 - 45 | 113311 | 16.62 | 376681 | 55.25 |
| 46 - 55 | 104036 | 15.26 | 480717 | 70.52 |
| 56 - 65 | 62620 | 9.19 | 543337 | 79.70 |
| 6 - 10 | 1196 | 0.18 | 544533 | 79.88 |
| 66 - 75 | 31180 | 4.57 | 575713 | 84.45 |
| Over 75 | 19457 | 2.85 | 595170 | 87.30 |
| Unknown | 86546 | 12.70 | 681716 | 100.00 |

Figure 7. Répartition des tranches d'âge

| Age_of_Driver | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---------------|-----------|---------|----------------------|--------------------|
| 1 | 6 | 0.00 | 6 | 0.00 |
| 2 | 3 | 0.00 | 9 | 0.00 |
| 3 | 19 | 0.00 | 28 | 0.00 |
| 4 | 31 | 0.01 | 59 | 0.01 |
| 5 | 69 | 0.01 | 128 | 0.02 |
| 6 | 108 | 0.02 | 236 | 0.04 |
| 7 | 165 | 0.03 | 401 | 0.07 |
| 8 | 214 | 0.04 | 615 | 0.10 |
| 9 | 281 | 0.05 | 896 | 0.15 |
| 10 | 428 | 0.07 | 1324 | 0.22 |
| 11 | 708 | 0.12 | 2032 | 0.34 |
| 12 | 973 | 0.16 | 3005 | 0.50 |
| 13 | 1093 | 0.18 | 4098 | 0.69 |
| 14 | 1218 | 0.20 | 5316 | 0.89 |
| 15 | 1251 | 0.21 | 6567 | 1.10 |
| 16 | 2586 | 0.43 | 9153 | 1.54 |
| 17 | 6852 | 1.15 | 16005 | 2.69 |
| 18 | 11213 | 1.88 | 27218 | 4.57 |
| 19 | 11807 | 1.98 | 39025 | 6.56 |
| 20 | 12558 | 2.11 | 51583 | 8.67 |

Figure 8. Répartition des accidents sur les conducteurs âgés de moins de 20 ans

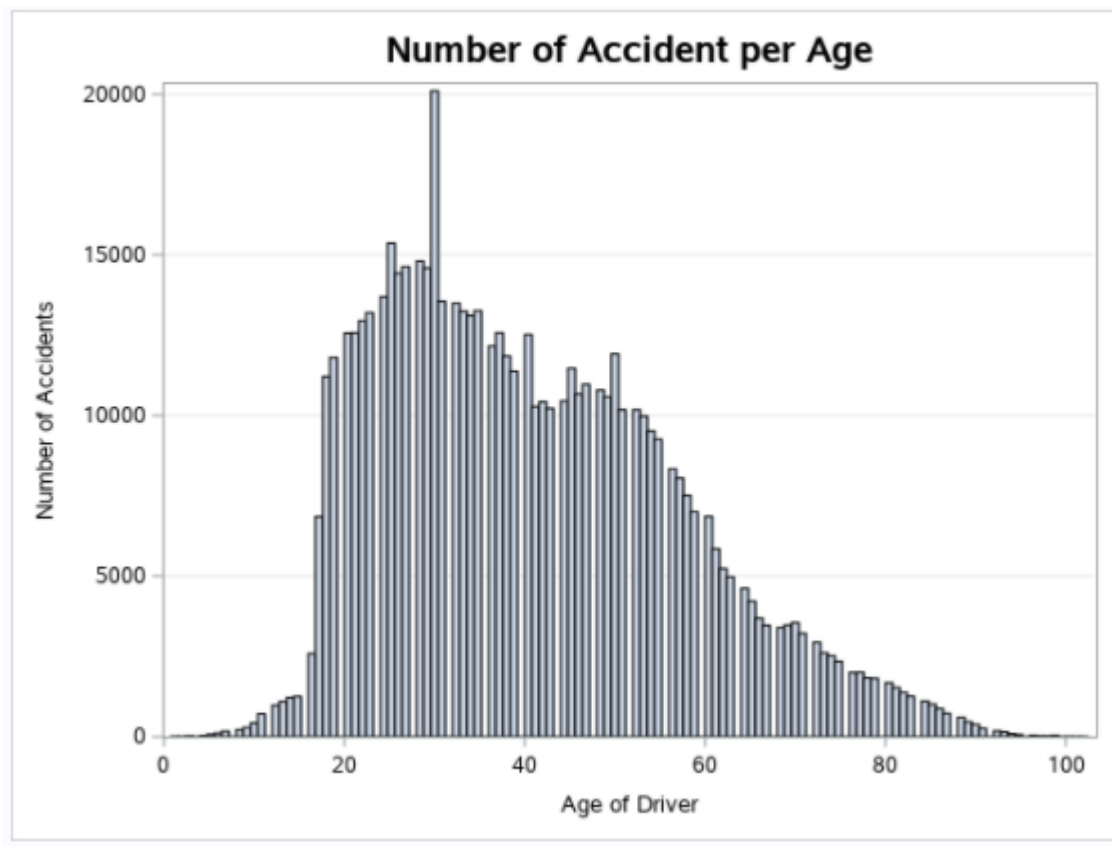


Figure 7. Histogramme du nombre d'accidents en fonction de l'âge

A l'aide du tableau ci-dessus et de l'histogramme, on peut remarquer deux choses. La première qui est tout à fait surprenante est que 16005 accidents ont été recensés par des conducteurs en-dessous de 18 ans. Il y a même 128 accidents dont le conducteur serait âgé entre 0 et 5 ans. Cependant, le manque d'informations quant à cette colonne du dataset ne nous permet de faire une analyse de ces résultats aussi surprenant soient-ils.

La deuxième analyse que l'on peut faire grâce à l'histogramme est que le nombre d'accidents est le plus élevé dans la période de 25 à 35 ans avec un pic à 30 ans. De même, on peut aussi constater que les

données suivent une loi normale. On va donc effectuer un test de student pour voir si l'âge est corrélé au nombre d'accident avec :

- H0 : L'âge n'est pas corrélé au nombre d'accidents
- H1 : L'âge est corrélé au nombre d'accidents

On va tout d'abord effectuer un test de normalité sur les données pour être sûr que celles-ci suivent bien une loi normale avant d'effectuer le test de student.

Variable: Age_of_Driver

| Tests for Normality | | | | |
|---------------------|-----------|----------|-----------|---------|
| Test | Statistic | | p Value | |
| Kolmogorov-Smirnov | D | 0.080885 | Pr > D | <0.0100 |
| Cramer-von Mises | W-Sq | 910.8933 | Pr > W-Sq | <0.0050 |
| Anderson-Darling | A-Sq | 5907.097 | Pr > A-Sq | <0.0050 |

Variable: Age_of_Driver

| N | Mean | Std Dev | Std Err | Minimum | Maximum |
|--------|---------|---------|---------|---------|---------|
| 595170 | 40.7010 | 18.4226 | 0.0213 | 1.0000 | 102.0 |

| Mean | 95% CL Mean | Std Dev | 95% CL Std Dev |
|---------|-----------------|---------|-----------------|
| 40.7010 | 40.6592 40.7427 | 18.4226 | 18.3931 18.4521 |

| DF | t Value | Pr > t |
|--------|---------|---------|
| 595169 | 1911.98 | <.0001 |

Figure 8. Test de normalité et de Student

On peut voir que la p-value pour le test de normalité est inférieur à 0.05 (5%) donc les données suivent bien une loi normale. De même, on peut voir que la p-value du test de Student est inférieur à 0.05 (5%). On peut donc rejeter l'hypothèse H0. Ainsi, l'âge du conducteur est bien corrélé au nombre d'accidents.

Si l'âge du conducteur est corrélé avec le nombre d'accidents qu'en est-il pour la sévérité des accidents ? Pour cela nous allons effectuer un test de Pearson pour vérifier la corrélation des deux variables. Mais avant, nous allons étudier la répartition de la sévérité des accidents avec les tranches d'âge. En effet, si cette étude était faite avec l'âge, cela donnerai des résultats illisibles. Les tranches d'âge nous donneront déjà une bonne idée de cette répartition.

| Table of label by Accident_Severity | | | | |
|-------------------------------------|-------------------|--------|--------|--------|
| label | Accident_Severity | | | Total |
| | 1 | 2 | 3 | |
| 0 - 5 | 2 | 38 | 88 | 128 |
| 11 - 15 | 21 | 1258 | 3984 | 5243 |
| 16 - 20 | 480 | 10865 | 33891 | 45016 |
| 21 - 25 | 816 | 15324 | 51627 | 67767 |
| 26 - 35 | 1646 | 31980 | 111590 | 145216 |
| 36 - 45 | 1329 | 24519 | 87483 | 113311 |
| 46 - 55 | 1436 | 24387 | 78213 | 104036 |
| 56 - 65 | 1059 | 15851 | 45710 | 62620 |
| 6 - 10 | 2 | 293 | 901 | 1196 |
| 66 - 75 | 608 | 8181 | 22391 | 31180 |
| Over 75 | 477 | 5481 | 13499 | 19457 |
| Total | 7856 | 138177 | 449137 | 595170 |
| Frequency Missing = 14732 | | | | |

Figure 9. Tableau de répartition de la sévérité des accidents par tranche d'âge

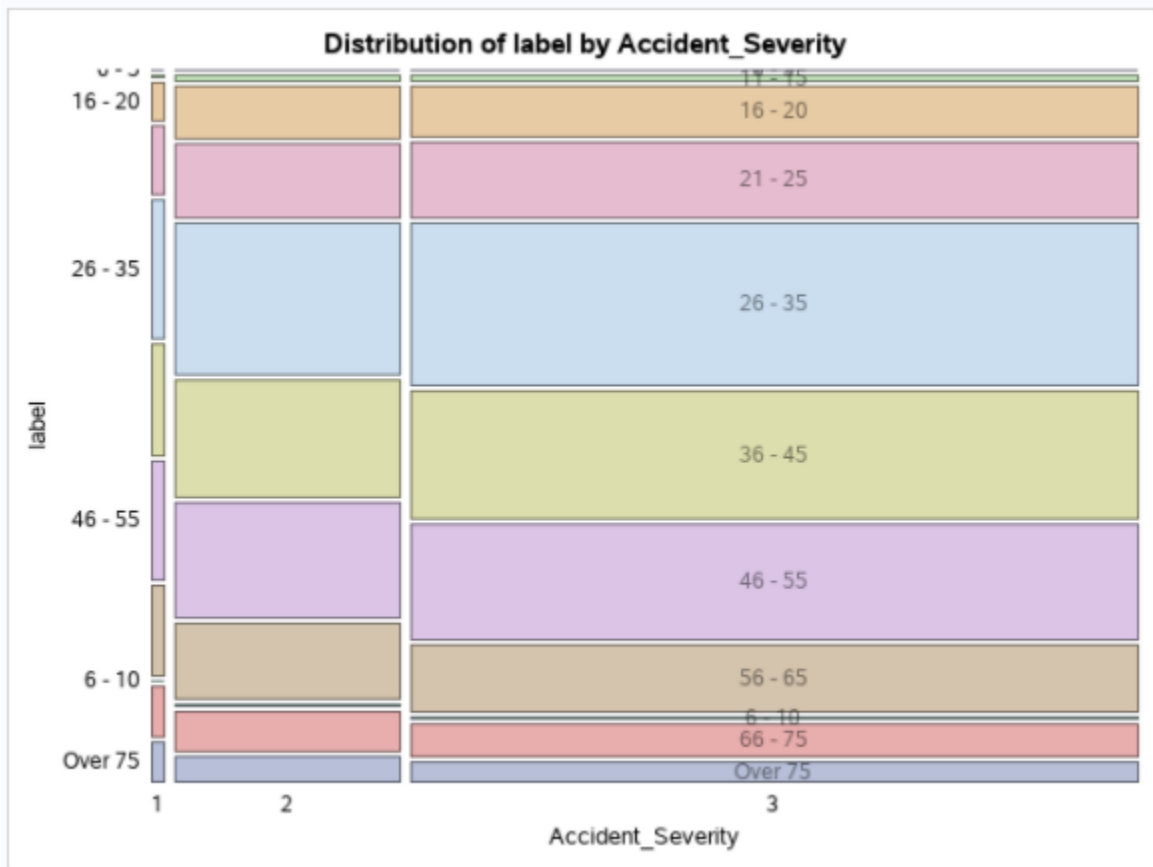


Figure 10. Graph mosaïque de la répartition de la sévérité des accidents avec les tranches d'âge

| | |
|-------------------|-------------------|
| 1 With Variables: | Accident_Severity |
| 1 Variables: | Age_of_Driver |

| Pearson Correlation Coefficients Number of Observations | |
|--|--------------------|
| | Age_of_Driver |
| Accident_Severity | -0.03897 595170 |

Figure 11. Test de corrélation entre l'âge du conducteur et la sévérité de l'accident

Le test de Pearson ne nous donne pas de résultats probants. Ainsi, on peut affirmer que l'âge du conducteur n'influence pas la sévérité de l'accident.

On peut donc conclure que les jeunes conducteurs sont plus sujets aux accidents mais qu'en aucun cas ils ont plus de chance d'avoir des accidents plus sévères que les autres. Cependant, une fois de plus ces analyses sont à utiliser avec prudence par le fait que certaines données sont aberrantes comme le montre les premiers tableaux. En effet, le nombre de conducteurs en dessous de 18 ans peuvent influencer les résultats présentés ci-dessus même si cela est étonnant au vue la provenance des données traités (Ministère des transports de Grande Bretagne).

IV) Sévérité victime – sévérité accident

V) Conclusion

| | |
|---|---|
| Figure 1. Répartition des accidents par rapport à leur sévérité | 4 |
| Figure 2. Tableau de fréquence et pourcentage | 4 |
| Figure 3. Répartition des accidents par pays | 5 |
| Figure 4. Nombre d'accidents par pays en fonction de la sévérité et du nombre d'habitant par pays | 5 |
| Figure 5. Graphique des Scores de dangerosité des routes par pays..... | 6 |
| Figure 6. Moyenne d'âge des conducteurs..... | 7 |
| Figure 7. Répartition des tranches d'âge | 7 |
| Figure 8. Histogramme du nombre d'accidents en fonction de l'âge | 8 |