



BUT Sciences des données 2022-2023

SAE

Création de reporting à partir de données stockées dans un SGBD relationnel



DIOP Coumba

--

Mr ATIGHEHCHI

Introduction :

Chaque année, le gouvernement britannique recueille et diffuse des données détaillées sur les accidents de la route à travers le pays, dans le cadre de The STATS19. Ces informations incluent des détails tels que les emplacements géographiques, les conditions météorologiques, les types de véhicules impliqués, le nombre de victimes et les manœuvres effectuées. Ces données fournissent une vue d'ensemble précieuse et complète de la situation des accidents de la route au Royaume-Uni, les rendant particulièrement utiles pour l'analyse et la recherche.

La base de données STATS19 comprend quatre tables pour l'année 2020 seulement, chacune contenant des informations différentes : Accident, Vehicle, Casualty et Codebook. La table Accident contient des informations sur chaque accident de la circulation, la table Vehicle contient des informations sur l'implication de chaque véhicule dans un accident, la table Casualty contient des informations sur chaque victime de la route, et la table Codebook est un dictionnaire des codes utilisés dans les autres tables. Ces tables peuvent être reliées grâce à l'attribut commun `accident_reference`.

Ce rapport vise à créer des tableaux de bord pour rendre compte de l'emplacement, des types de véhicules, de l'effet de la limite de vitesse, du type de route, etc. des accidents, ainsi que du nombre de victimes. Les données ont été extraites et analysées à l'aide de requêtes SQL de la base de données STATS19 et des visualisations et analyses plus détaillées et interactives ont été créées avec LOCalc.

Méthodologie :

Pour atteindre les objectifs de ce projet, j'ai fait usage de différents logiciels, comme un tableur via Libre Office Calc pour avoir une vue d'ensemble de la base de données. J'ai également utilisé Libre Office Base pour gérer les bases de données, ainsi que Open Database Connectivity pour connecter Libre Office Base à une source de données externe comme SQLite.

Nombres d'accidents

Tout d'abord, j'ai commencé par des requêtes permettant de comprendre le nombre d'accidents de la route et de véhicules impliqués en 2020.

Pour compter le nombre total d'accidents, j'ai utilisé la requête suivante :

```
SELECT COUNT(*) AS accidents  
From accident
```

Cette requête utilise la commande "SELECT" pour sélectionner des données d'une table spécifique(accident) dans une base de données(stat19). La fonction "COUNT(*)" est utilisée pour compter le nombre total de lignes dans la table accident. La clause "AS" est utilisée pour donner un nom à cette colonne de résultat, dans ce cas "accidents". Enfin, la clause "FROM" spécifie la table à partir de laquelle les données doivent être sélectionnées "accident".

En résumé, cette requête va compter le nombre total de lignes dans la table "accident" et renommer la colonne de résultat en "accidents". Le résultat de cette requête sera un seul nombre entier qui représente le nombre total de lignes dans la table accident.

Ainsi, il y'avait 91199 accidents en 2020. Plusieurs facteurs peuvent expliquer ce nombre élevé d'accidents au Royaume-Uni en 2020. Les conditions météorologiques défavorables, la vitesse excessive et l'utilisation des téléphones cellulaires sont parmi les principaux responsables des accidents de la route. Il y a également un manque de sensibilisation à la sécurité routière qui conduit à une augmentation des comportements à risque sur les routes. La pandémie de Covid-19 a également contribué, car elle a entraîné une augmentation du nombre de personnes en déplacement et à la conduite, ce qui a entraîné une hausse des événements imprévus et des collisions sur la route.

Lieu des accidents

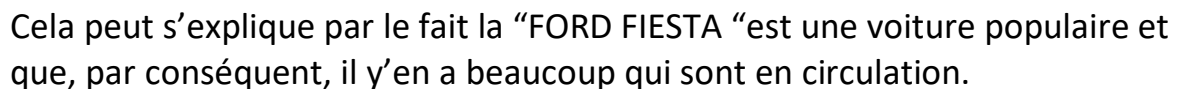
J'ai cherché à déterminer les lieux où les accidents sont plus fréquents.

J'ai utilisé l'attribut "local_authority_district" qui se trouve dans la table "accident" comme la montre la requête ci-dessous :

```
SELECT local_authority_district , COUNT(*)  
FROM accident  
WHERE local_authority_district >0  
GROUP BY local_authority_district
```

En résumé, cette requête va sélectionner le nombre total d'accidents pour chaque district de l'autorité locale, pour les lignes où cette colonne est supérieure à 0. Le résultat de cette requête sera une liste de lignes contenant le district de l'autorité locale et le nombre d'accidents pour ce district.

diagramme du nombre d'accident par modèle de voiture



En examinant les types de véhicules impliqués dans les accidents, J'ai constaté que les véhicules légers (voitures, camionnettes, etc.) étaient les plus présents, suivis par les motocyclettes et les poids lourds. J'ai également remarqué que certains types de véhicules étaient plus souvent impliqués dans des accidents graves, comme les poids lourds et les autobus.

Vitesse limite

J'ai calculé la moyenne de la vitesse limite par la requête qui suit.

```
SELECT AVG (speed_limite)
```

```
FROM accident
```

Cette requête utilise la commande "SELECT" pour sélectionner des données dans la table "accident" de la base de données. Elle utilise également la fonction "AVG" pour calculer la moyenne d'une colonne spécifique dans cette table. Elle calcul la moyenne de la colonne "speed_limit" dans la table "accident". La requête retourne ensuite 38,37 comme valeur moyenne.

En observant les données, j'ai constaté que la vitesse moyenne, qui est de 36,27, n'est pas très élevée malgré un grand nombre d'accidents. Pour mieux comprendre cette tendance, j'ai élaboré une requête qui me permet de visualiser le nombre d'accidents en fonction de la limite de vitesse.

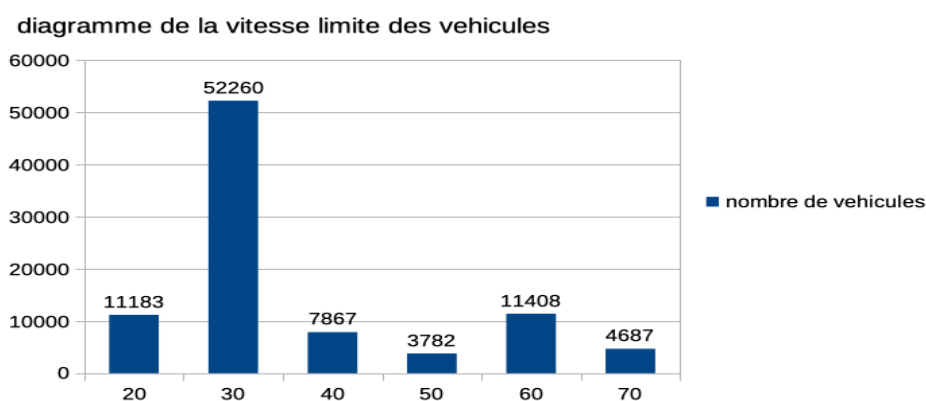
```
SELECT "speed_limit" COUNT(*)
```

```
FROM "accident"
```

```
WHERE "speed_limit" != -1
```

```
GROUP BY "speed_limit"
```

Cette requête utilise la commande S "SELECT" pour sélectionner la colonne "speed_limit" et la fonction "COUNT(*)" pour compter le nombre de lignes dans la table "accident". La clause "WHERE" est utilisée pour filtrer les lignes de la table "accident", dans ce cas, elle filtre les lignes où la "speed_limit" n'est pas égale à -1. La clause "GROUP BY" est utilisée pour regrouper les lignes sélectionnées par la colonne "speed_limit", ce qui donne une table qui montre le nombre d'accidents qui se sont produits à chaque limite de vitesse.



On remarque que les accidents sont beaucoup plus nombreux à basse vitesse. Cela peut être attribué à divers facteurs, y compris une mauvaise application des règles et des lois, une mauvaise conduite, un manque de concentration, un environnement imprudent. D'autres facteurs tels que le mauvais temps, les conditions de route, la fatigue et la distraction peuvent également jouer un rôle. De plus, il est possible que les conducteurs à faible vitesse soient plus susceptibles d'être impliqués dans des accidents en raison du fait qu'ils sont dans des endroits plus fréquentés, comme dans les zones résidentielles ou de petites routes.

Conditions climatiques

J'ai trouvé important d'examiner les conditions climatiques car elles peuvent avoir une influence considérable sur le comportement des automobilistes et les risques d'accidents. Les pluies, les vents et les nuages peuvent affecter la visibilité, ce qui peut être un facteur dans les accidents de la route. J'ai donc fait la requête suivante :

```
SELECT weather_conditions,count(*)  
FROM accident  
WHERE weather_conditions != -1  
GROUP BY weather_conditions
```

Cette requête utilise la commande "SELECT" pour sélectionner la colonne "weather_conditions" et la fonction "COUNT(*)" pour compter le nombre de lignes dans la table "accident". La clause "WHERE" est utilisée pour filtrer les lignes de la table "accident", dans ce cas, elle filtre les lignes où la "weather_conditions" n'est pas égale à -1. La clause "GROUP BY" est utilisée pour regrouper les lignes sélectionnées par la colonne "weather_conditions", ce qui donne une table qui montre le nombre d'accidents qui se sont produits en fonction des conditions météorologiques.

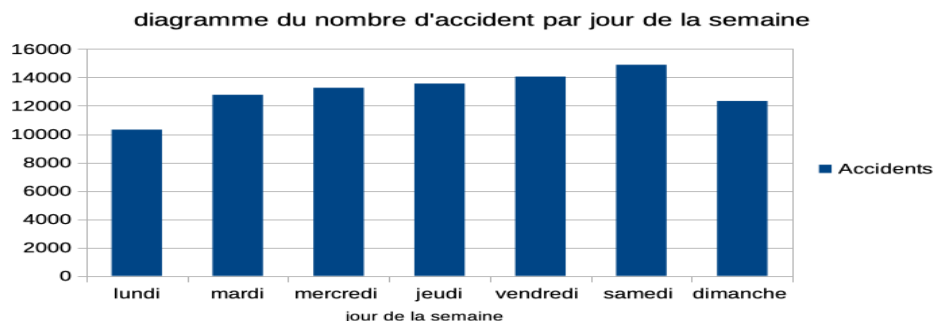
Jours de la semaine

J'ai aussi étudié les nombres d'accidents en fonction des jours de la semaine. Cela peut aider à comprendre comment les facteurs liés à la sécurité routière varient selon le jour de la semaine et à adapter les interventions de façon appropriée. J'ai donc fait la requête suivante :

```
SELECT "day_of_week",COUNT(*)
```

```
FROM "accident"  
WHERE "day_of_week"!= -1  
GROUP BY "day_of_week"
```

Cette requête utilise la commande "SELECT" pour sélectionner la colonne "day_of_week" et la fonction "COUNT(*)" pour compter le nombre de lignes dans la table "accident". La clause "WHERE" est utilisée pour filtrer les lignes de la table "accident", dans ce cas, elle filtre les lignes où la "day_of_week" n'est pas égale à -1. La clause "GROUP BY" est utilisée pour regrouper les lignes sélectionnées par la colonne "day_of_week", ce qui donne une table qui montre le nombre d'accidents qui se sont produits pour chaque jour de la semaine.

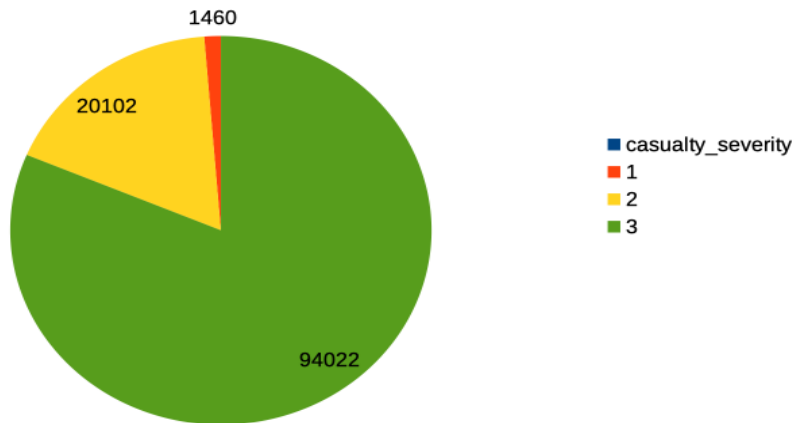


On remarque qu'il y a beaucoup plus d'accidents le samedi. Cela peut s'expliquer par le fait qu'il est jour où les gens font plus de sorties et de courses. Ils sont donc susceptibles de se retrouver sur les routes et d'être impliqués dans les accidents. De plus, ils se reposent généralement le weekend et donc moins fatigués le lundi d'où ils sont susceptibles de faire attention.

Nombres de victimes

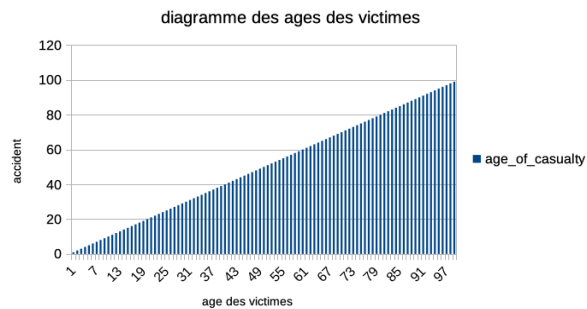
Enfin, en analysant le nombre de victimes, J'ai constaté que la majorité des accidents ne font que des blessés légers ou modérés. Cependant, il y a un nombre significatif d'accidents graves, le diagramme circulaire ci-dessous le montre.

diagramme du nombre de victimes par gravité



J'ai également réalisé une analyse statistique de l'âge des victimes et créé un diagramme pour une meilleure visualisation des données.

	1 colonne	
Moyenne	50	
Erreur type	2,886751346	
Mode	#VALEUR !	
Médiane	50	
Premier quart	25,5	
Troisième qu	74,5	
Variance	825	
Écart type	28,72281323	
Kurtosis	-1,2	
Asymétrie	0	
Plage	98	
Minimum	1	
Maximum	99	
Somme	4950	
Compter	99	



Cela montre que les personnes âgées sont plus susceptibles d'être victimes d'accidents au Royaume-Uni. Cela est dû à une variété de facteurs qui incluent leurs performances physiques moins bonnes, leurs capacités mentales diminuées et leurs mauvaises habitudes de conduites.

CONCLUSION :

En conclusion, ce rapport met en lumière certains des facteurs clés qui contribuent aux accidents de la route au Royaume-Uni. En mettant ces données en perspective avec d'autres informations sur les accidents de la route, comme les conditions météorologiques ou les manœuvres des véhicules, il devient possible de mieux comprendre les causes des accidents de la route et de déterminer des actions concrètes pour prévenir ces accidents à l'avenir.

