

Analyse des accidents corporels de la circulation

Cette étude présente une analyse exploratoire du jeu de données public intitulé « **Base de données des accidents corporels de la circulation** » disponible sur la plateforme data.gouv.fr. Le fichier recense **454 372 accidents** décrits par 39 variables (lieu, gravité, nombre d'usagers impliqués, etc.). L'objectif de ce travail est d'extraire des tendances générales et de montrer comment appliquer les techniques de data science vues en cours : chargement et nettoyage avec pandas, calcul de statistiques descriptives et visualisations interactives avec Plotly.

Méthodologie

1. **Environnement.** Après téléchargement du fichier CSV, un environnement virtuel a été créé et les dépendances installées :

```
python -m venv .venv
source .venv/bin/activate # ou .venv\Scripts\activate sur Windows
pip install -r requirements.txt
```
2. **Exécution du script.** Le script `main.py` charge le CSV encodé en latin-1, convertit les colonnes numériques (ttue, tbg, tbl, tindm, nbimplique), supprime les valeurs manquantes et enregistre un fichier nettoyé. Trois graphiques interactifs sont produits et stockés dans le dossier `outputs`.
`python main.py`
3. **Versionnement Git et mise en ligne.** Le projet a été versionné puis publié sur GitHub :
`git init
git add .
git commit -m "Analyse des accidents de la route"
git remote add origin https://github.com/.../projet-accidents.git
git push -u origin main`

Nettoyage des données

Le fichier original est volumineux (plus de 60 Mo) et comporte des colonnes à types mixtes. L'encodage latin-1 et le séparateur ; ont été précisés lors du chargement. Les colonnes numériques ont été converties en entiers ou flottants et les lignes avec des informations essentielles manquantes ont été exclues. Le jeu de données nettoyé est enregistré dans `clean_accidents_insee.csv` pour faciliter des analyses ultérieures.

Résultats

La base finale comporte **454 372 accidents** décrits par 39 variables. Les statistiques descriptives révèlent un taux de mortalité très faible : le nombre moyen de personnes tuées par accident est d'environ **0,06**, et la médiane est nulle, ce qui signifie que la majorité des accidents n'occasionnent aucun défunt. Le nombre moyen de blessés graves est de **0,47** et celui des blessés légers de **0,79**, avec des maximums respectifs de 42 et 53. Les accidents impliquent en moyenne

2,2 véhicules (maximum de 86).

Analyse par département

En agrégeant les accidents par département, la concentration se fait sentir dans les zones urbaines. Le top 5 est dominé par **Paris (750)** avec 47 278 accidents, suivi des **Bouches-du-Rhône (130)** avec 26 634 accidents. **Seine-Saint-Denis (930)**, **Oise (60)** et **Hauts-de-Seine (920)** complètent le classement avec respectivement 17 108, 16 611 et 15 231 accidents. Ces chiffres doivent être mis en perspective avec la population et le trafic local pour interpréter correctement la sinistralité.

Visualisations

Trois graphiques interactifs ont été générés :

- Un **diagramme en barres** présentant les 15 départements les plus accidentogènes.
- Un **histogramme** montrant la distribution du nombre de tués par accident, confirmant que la majorité des accidents ne provoquent aucun décès.
- Un **nuage de points** localisant chaque accident selon sa latitude et sa longitude, offrant une vue d'ensemble de la répartition géographique.

Ces visualisations sont consultables dans le dossier outputs du projet et permettent d'explorer les tendances de manière interactive.

Conclusion

Cette analyse met en évidence un taux de mortalité par accident très bas et une forte concentration des accidents dans les départements urbains. Le travail illustre l'intérêt des techniques de data science pour valoriser des données publiques. En publiant le code et les résultats sur GitHub, la traçabilité et la reproductibilité sont assurées. Des analyses futures pourraient compléter cette étude en intégrant des variables comme la météo, l'état des infrastructures ou le comportement des usagers afin de mieux comprendre les facteurs de risque et d'orienter les politiques de prévention.