

Contexte et objectifs

Le projet Accidentologie-HDS s'inscrit dans le développement d'une application Streamlit dont l'objectif est de faciliter la visualisation et l'analyse des accidents routiers dans le département des Hauts-de-Seine (92). Notre application doit permettre aux utilisateurs de comprendre les tendances temporelles, de localiser précisément les zones à risque, d'étudier la répartition géographique des accidents, et d'identifier les facteurs potentiellement associés à la gravité des collisions. Notre objectif général est de fournir une solution intuitive, robuste et complète, capable de transformer une base de données brute en insights exploitables pour des analyses de sécurité routière. Le travail inclut un important volet de nettoyage, de normalisation et de structuration des données, ainsi qu'un développement d'interfaces claires et fonctionnelles.

2. Données et outils

Les données utilisées proviennent d'un fichier nommé acc.csv, dont l'encodage et le séparateur peuvent varier. Afin d'assurer une lecture fiable, nous avons mis en place une fonction spécialisée de chargement intelligent, capable d'effectuer des tentatives successives selon plusieurs encodages et formats. Les principaux outils que nous avons utilisés incluent Python et ses bibliothèques de manipulation de données (pandas, numpy), Streamlit pour l'interface utilisateur, Plotly pour la visualisation interactive, ainsi que Folium pour les représentations cartographiques plus détaillées. Plusieurs transformations ont été réalisées : normalisation des dates et heures, extraction automatique des coordonnées géographiques depuis la colonne geo_point_2d, et standardisation des intitulés en français pour garantir une meilleure lisibilité.

3. Chronologie détaillée Semaine du 4 au 7 novembre 2025

Recherche, exploration et préparation des données Au début du projet, nous avons consacré plusieurs jours à rechercher les données officielles, à les importer correctement et à effectuer un premier audit qualitatif. Des vérifications ont été réalisées sur les types de colonnes, la cohérence interne des variables, la présence de valeurs aberrantes et les éventuels problèmes de complétion des champs. Une première version du dictionnaire des variables a été créée et une page initiale d'exploration a été ajoutée dans Streamlit, permettant de visualiser les tableaux bruts, de filtrer les données et de consulter les statistiques de base. Plusieurs problèmes ont été identifiés, notamment des lenteurs sur les cartes et la présence de catégories codées sans signification apparente. Nous avons mis en place des corrections telles que la création d'un mapping français des modalités et l'intégration d'un échantillonnage ou d'un clustering pour accélérer l'affichage des points sur la carte.

13 novembre 2025

Mise en place de l'architecture Streamlit et de la navigation Nous avons structuré l'application en plusieurs pages pour clarifier le parcours utilisateur : page d'accueil, analyse temporelle, analyse spatiale, corrélations, insights. Un bandeau de navigation simple et intuitif a été intégré. Sur la page

d'accueil, nous avons ajouté des indicateurs clés (KPI) permettant de visualiser rapidement les statistiques majeures. L'intégration des cartes géographiques avec Plotly et Folium a nécessité plusieurs ajustements. Des erreurs liées aux coordonnées géographiques ou à l'utilisation d'un token Mapbox manquant sont survenues. Après corrections, les cartes proposent désormais différents modes : affichage des points, clustering automatique et heatmap.

17 novembre 2025

Ajout des corrélations et amélioration de la qualité Pour mieux comprendre les relations entre les variables, nous avons intégré plusieurs méthodes de corrélation adaptées à la nature des données : Pearson, Spearman, Cramér V, ratio de corrélation. Les matrices obtenues étant denses, nous avons ajouté un filtrage intelligent : top-k, tooltips et mise en avant des relations significatives. Un regroupement des modalités rares a également été effectué pour améliorer la lisibilité.

1er décembre 2025

Construction des insights et recommandations Nous avons développé une page dédiée aux insights. Elle comprend des KPI, un calcul d'uplift pour les accidents graves, l'identification des facteurs associés au risque et un narratif automatique. L'objectif est de proposer une synthèse claire et accessible, même aux utilisateurs non spécialistes. Nous avons ajouté des mécanismes de gestion d'erreurs pour éviter les interruptions dues à des valeurs manquantes ou des conversions incorrectes. Cette étape a permis d'apporter une dimension plus interprétative et opérationnelle à notre tableau de bord.

4. Points bloquants et solutions apportées

Nous avons rencontré plusieurs difficultés majeures :

- Problèmes d'encodage du fichier, résolus avec une fonction de lecture intelligente.
- Valeurs aberrantes ou mal catégorisées, traitées via un regroupement des modalités rares.
- Absence de latitude et longitude, corrigée grâce à l'extraction depuis `geo_point_2d`.
- Surcharge d'affichage sur les cartes, résolue par du downsampling et du clustering.
- Matrices de corrélation trop complexes, simplifiées grâce au filtrage dynamique.

5. Livrables fournis Nous avons produit les éléments suivants :

- l'application Streamlit multi-pages complète
- le fichier `data_store.py` ;
- le mapping français des variables ;
- la page Insights fonctionnelle ;
- ce journal de bord détaillé retraçant nos étapes de travail.
- projet sur GITHUB