

Analyse et Visualisation des Accidents de la Circulation Routière

Projet Data Sciences

ESPRIT 2024

Projet

Développer une solution complète pour analyser et visualiser des accidents de la circulation routière.



Développer une solution complète pour analyser et visualiser des accidents de la circulation routière.



Les accidents de la circulation routière

sont des événements soudains et non intentionnels qui se produisent sur la route. Ils peuvent être causés par un certain nombre de facteurs, notamment la vitesse excessive, la conduite sous l'influence de l'alcool ou de drogues, l'inattention ou l'utilisation de la technologie dans la voiture.

Statistiques sur le nombre de véhicules dans le monde

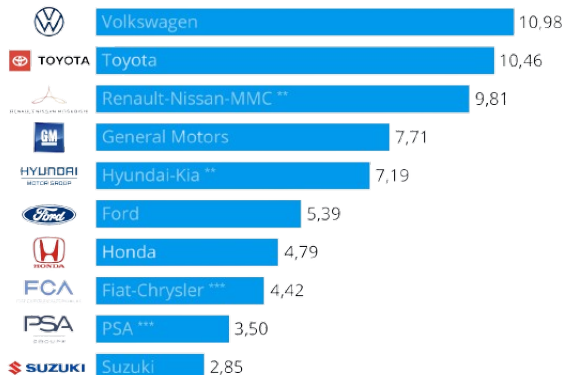
En 2023, il y avait environ 1,4 milliard de véhicules à moteur en circulation dans le monde. Ce nombre devrait augmenter à 1,7 milliard d'ici 2030.

Les États-Unis sont le pays avec le plus grand nombre de véhicules à moteur, avec environ 278 millions de véhicules. La Chine est en deuxième position, avec environ 260 millions de véhicules.

Les voitures particulières représentent la majorité des véhicules à moteur dans le monde, avec environ 947 millions de véhicules en circulation. Les camions et les autobus représentent environ 400 millions de véhicules, et les motos et les scooters représentent environ 30 millions de véhicules.

Les plus grands constructeurs automobiles au monde

Principaux groupes automobiles selon le nombre de véhicules vendus dans le monde en 2019, en millions



* Les ventes de Toyota, Nissan, Mitsubishi, Honda et Suzuki se réfèrent à l'année fiscale se terminant le 31 mars 2020.

** Constructeurs liés au sein d'une alliance stratégique ou d'un conglomérat.

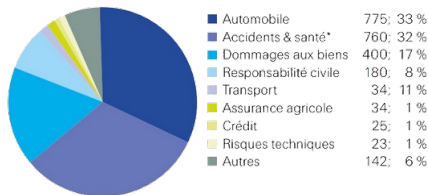
*** Fusion entre PSA et Fiat-Chrysler (groupe Stellantis) prévue début 2021.

Figure – source : Statista

Assurance Non-vie

En 2018 le montant des primes non-vie collectées par les assureurs a franchi pour la première fois des 3 373 milliards de dollars au niveau mondial.

(en milliards USD et en %), 2018E



Les accidents de la route dans le monde

En 2023, il y a eu environ 1,35 million de décès dans le monde dus à des accidents de la route (une moyenne d'une personne toutes les 21 secondes).

Les accidents de la route sont :

- ▶ la 1ère cause de décès chez les jeunes âgés de 15 à 29 ans.
- ▶ la 5ième cause de décès dans le monde, tous âges confondus.



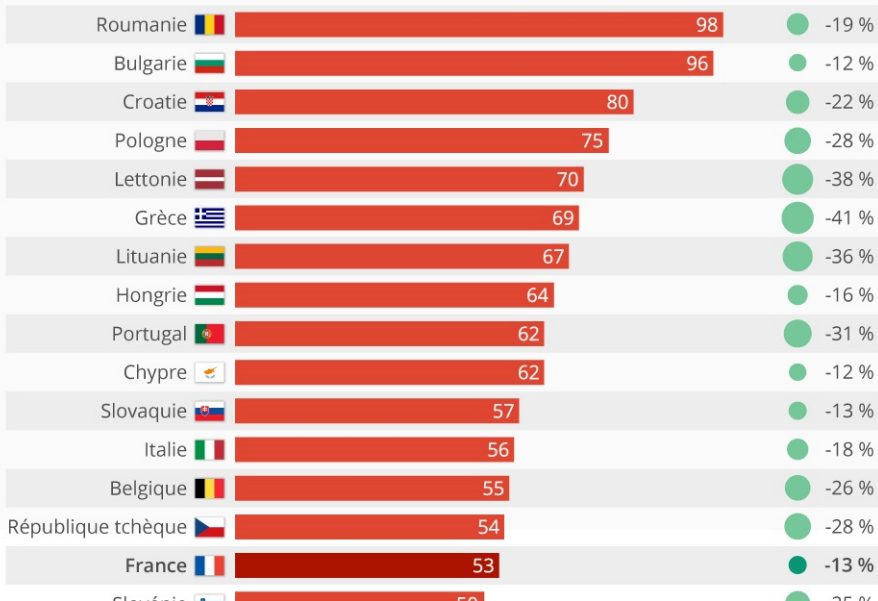
Des facteurs qui contribuent aux accidents de la route

- ▶ La vitesse excessive
- ▶ La conduite sous l'influence de l'alcool ou de drogues
- ▶ L'inattention
- ▶ L'utilisation de la technologie dans la voiture
- ▶ ...

Mortalité routière : Les pays qui déplorent le plus de victimes

Décès dus aux accidents de la route par million d'habitants en 2017

Evolution 2010-2017

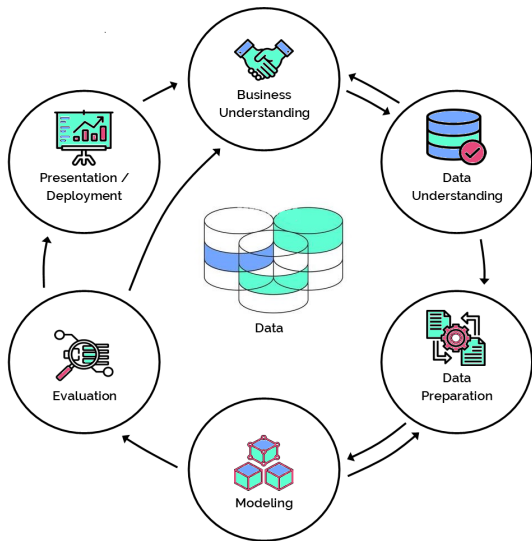


Travail demandé

- **Objectif** : Développer une solution complète pour analyser et visualiser les accidents de la route.
- **Approches** : Data science, analyse de données, visualisation.
- **Implique** : Développement d'une interface utilisateur, déploiement du modèle prédictif, validation continue.



Méthode de travail : Cross-Industry Standard Process for Data Mining (CRISP-DM)



La méthode CRISP-DM est itérative, ce qui signifie que les phases peuvent être revisitées en fonction des résultats obtenus

planning

Sprint 1 : Data Scraping

Durée : Semaines 1-2

Etape 1

Technologies :

BeautifulSoup



planning

Sprint 1 : Data Scraping

Durée : Semaines 1-2

Etape 1

Technologies :

BeautifulSoup



Scrapy
Python Library



Etape 2

- ▶ Utiliser SQL server integration service (SSIS) pour automatiser le processus de collecte de données.
- ▶ Mise en place de flux de données pour extraire, transformer et charger (ETL) les informations dans la base de données.



Microsoft®
SQL Server®
Integration Services

Durée : Semaines 3-4

Il y a plusieurs techniques statistiques et visuelles peuvent être utilisées pour comprendre la distribution des données et identifier des caractéristiques significatives

Technologies :

matplotlib



seaborn

pandas

Planning

Sprint 3 : Modélisation

Durée : Semaines 5-6

création, entraînement et évaluation de modèles prédictifs basés sur les données disponibles...

Technologies :



Planning

Sprint 4 : Analyse de Sentiment

Durée : Semaine 7

Compréhension approfondie des nuances émotionnelles présentes dans les accidents de la route.

Utilisation de techniques de traitement du langage naturel ...



POSITIVE



NEUTRAL



NEGATIVE

Planning

Sprint 5 : Développement de l'Interface Utilisateur

Durée : Semaines 8-9

Conception et implémentation...

Intégration de fonctionnalités interactives...

Technologies :

The Django logo, featuring the word "django" in a bold, dark green, lowercase sans-serif font.The Flask logo, which includes a black silhouette of a flask on the left and the word "Flask" in a large, black, serif font on the right. Below "Flask" is the tagline "web development, one drop at a time" in a smaller, black, sans-serif font.The Plotly Dash logo, consisting of a blue square icon with white vertical bars and dots on the left, followed by the text "plotly | Dash" in a blue, sans-serif font.The Streamlit logo, featuring a red, stylized crown icon on the left and the word "Streamlit" in a large, dark blue, sans-serif font on the right.

Planning

Sprint 6 : Déploiement sur une Plateforme

Durée : Semaine 10



Livrables

- ▶ Rapport de data scraping avec les données collectées.
- ▶ Rapport d'analyse exploratoire des données.
- ▶ Modèle de prédiction des fake jobs avec un rapport d'évaluation.
- ▶ Intégration de l'analyse de sentiment dans le modèle.
- ▶ Interface utilisateur fonctionnelle.
- ▶ Modèle déployé sur une plateforme avec documentation.