

Analyse Data science des Accidents de la Route









Sommaire



O1 Introduction au projet

O2 Source de données

03 Méthodologie d'extraction

04

Technologies utilisées



Introduction

Notre projet vise à transformer une vaste collection de données sur les accidents en insights actionnables et en connaissances approfondies.

En analysant des données détaillées, notre objectif est d'identifier les causes, les tendances et les facteurs contributifs des accidents routiers, dans le but ultime de renforcer la sécurité routière et de sauver des vies.

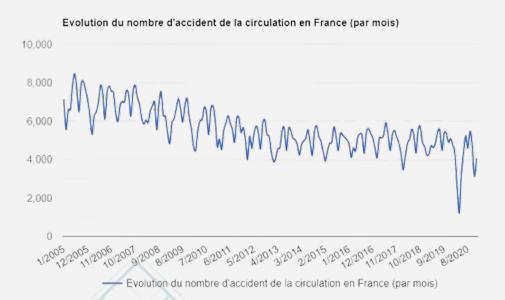
Source de données

Les données que nous utilisons pour notre analyse sont extraites du site **data.gouv.fr**, qui propose un ensemble complet d'informations sur les accidents de la route en France, couvrant la période de **2005 à 2022.**

Ces ensembles de données sont organisés annuellement en fichiers CSV, comprenant des catégories détaillées telles que les caractéristiques de l'accident qui décrit les circonstances générales de l'accident, le lieu qui décrit le lieu principal de l'accident même si celui-ci s'est déroulé à une intersection, les véhicules impliqués et les informations sur les usagers impliqués.

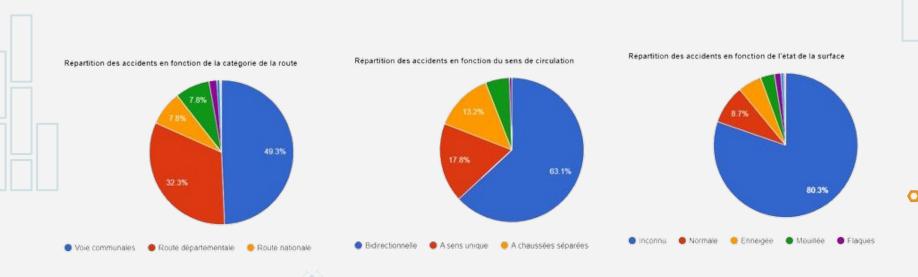
Source de données

Ce graphique linéaire montre l'évolution mensuelle du nombre d'accidents de la circulation en France sur une période étendue, s'étalant de janvier 2005 jusqu'à décembre 2022.



Source de données

Les trois graphiques représentent une analyse statistique des accidents de la route, montrant la répartition des accidents en fonction de la catégorie de la route, du sens de circulation et de l'état de la surface.



Notre processus d'extraction de données s'appuie sur un script Python automatisé utilisant la bibliothèque requests pour interagir avec l'API du site data.gouv.fr.

Voici les étapes clés de notre méthode :

1-Requête API: Le script initie une requête GET vers l'endpoint API spécifié, qui est conçu pour lister les ressources disponibles dans le dataset ciblé, avec un paramètre de pagination et de filtrage par mot-clé.

```
import requests
import os

# Replace 'YOUR_API_ENDPOINT' with the actual API endpoint you want to use
api_endpoint = 'https://www.data.gouv.fr/api/2/datasets/53698f4ca3a729239d2036df/resources/?page=1&type=main&page_size=999&q='
response = requests.get(api_endpoint)
```

2-Sélection des Données : Nous définissons les catégories de données à télécharger, telles que 'véhicules', 'lieux', 'usagers', et 'caractéristiques', représentant les différentes dimensions des informations sur les accidents de la route.

3-Organisation des Fichiers : Pour chaque catégorie, le script prépare un dossier de destination dans le répertoire spécifié sur le disque local, créant un nouveau dossier si nécessaire.

4-Téléchargement : En parcourant la réponse JSON de l'API, le script identifie et téléchargez les fichiers pertinents à partir de leurs URL respectives, en les sauvegardant dans les dossiers correspondants.

5-Gestion des Erreurs: Tout au long du processus, le script vérifie les codes de statut des réponses HTTP pour s'assurer que la récupération des données et le téléchargement des fichiers se déroulent sans erreur.

```
dataToDownload = ['vehicules','lieux','usagers','caracteristiques']
basePath = 'C:/Esprit/dataScienceProject/webscrapping/'
if response.status code == 200:
    data = response.json()
   for fileData in dataToDownload:
        download folder = basePath+fileData+"/"
        if not os.path.exists(download folder):
            os.makedirs(download folder)
        # Assuming the API response contains a list of elements
        for element in data['data']:
            if 'url' in element and fileData in element['url']:
                file url = element['url']
                # Download the file
                file response = requests.get(file url)
                if file response.status code == 200:
                    # Extract the filename from the URL
                    filename = os.path.basename(file url)
                    # Specify the full path to save the file in the download folder
                    file path = os.path.join(download folder, filename)
                    with open(file path, 'wb') as file:
                        file.write(file response.content)
                    print(f"File '{filename}' downloaded successfully to '{download folder}'.")
                    print(f"Failed to download file from {file url}. Status code: {file response.status code}")
    print(f"Failed to retrieve data from the API. Status code: {response.status code}")
```

Après l'extraction des données, une étape critique a été la récupération des clés correspondant aux différentes catégories d'informations présentes dans les fichiers CSV. À partir du document descriptif fourni par l'ONISR, nous avons pu établir un fichier JSON qui contient toutes les clés et leurs valeurs possibles.

0

lum

Lumière : conditions d'éclairage dans lesquelles l'accident s'est produit :

- 1 Plein jour
- 2 Crépuscule ou aube
- 3 Nuit sans éclairage public
- 4 Nuit avec éclairage public non allumé
- 5 Nuit avec éclairage public allumé

Description a partir du fichier description-des-bases-de-données-annuelles-2022.pdf

```
"lum": [

"Plein jour",

"Cr\u00e9puscule ou aube",

"Nuit sans \u00e9clairage public",

"Nuit avec \u00e9clairage public non allum\u00e9",

"Nuit avec \u00e9clairage public allum\u00e9"
],
```

Clé et les valeurs possible dans le fichier keyValueResult.json

0

Ce script a utilisé le fichier JSON, qui agit comme une carte de référence, pour parcourir chaque CSV et remplacer les clés numériques par leurs significations textuelles correspondantes.

Nettoyage des Chaînes de Caractères : Une fonction clean_string est définie pour nettoyer les chaînes de caractères en éliminant les accents et caractères spéciaux grâce à la bibliothèque unidecode.

```
def clean_string(s):

""" Nettoie la chaîne de caractères en remplaçant les caractères spéciaux """

return unidecode.unidecode(s)
```

•

Ce script a utilisé le fichier JSON, qui agit comme une carte de référence, pour parcourir chaque CSV et remplacer les clés numériques par leurs significations textuelles correspondantes.

Détection du Séparateur CSV : La fonction detect_separator lit la première ligne d'un fichier CSV pour déterminer si le séparateur des données est une virgule ou un point-virgule.

```
def detect_separator(csv_file):
    """ Detecte le separateur dans un fichier CSV """
    with open(csv_file, 'r', encoding='ISO-8859-1') as file:
        first_line = file.readline()
        if ';' in first_line:
            return ';'
        else:
            return ','
```

Remplacement des Valeurs : La fonction replace_values_in_csv charge le fichier JSON de mappage, qui contient les correspondances entre les clés numériques et leurs valeurs descriptives.

```
def replace values in csv(json mappings file, csv files):
    with open(json mappings file, 'r', encoding='utf-8') as file:
        mappings = json.load(file)
    for key in mappings:
        if isinstance(mappings[key], list):
            mappings[key] = [clean string(val) for val in mappings[key]]
    modified files = []
    for csv file in csv files:
        try:
           sep = detect separator(csv file)
           df = pd.read csv(csv file, encoding='ISO-8859-1', sep=sep, quotechar='"')
            modified = False
            for col in df.columns:
                if col in mappings and isinstance(mappings[col], list):
                    df[col] = df[col].apply(lambda x: safe apply(x, mappings[col]))
                    modified = True
            if modified:
                modified csv file = csv file.replace('.csv', ' modified.csv')
                df.to csv(modified csv file, index=False)
                modified files.append(modified csv file)
        except pd.errors.ParserError as e:
            print(f"Error processing file {csv file}: {e}")
    return modified files
```

Technologies utilisées



Langage de programmation



Editeur de code



4	А	В	C		D	E	F	G	Н	1	J	K	L	M	N	0	Р	Q	R
1	Num_Acc	an	mois	jc	our	hrmn	lum	agg	int	atm	col	com	adr	gps	lat	long	dep		
2	2.01E+11	1	5	1	12	1900	Nuit sans	En agglom	Hors inters	Non rense	Deux vehic	11	CD41B	M	5051500	294400	590		
3	2.01E+11	į.	5	1	21	1600	Plein jour	En agglom	Hors inters	Non rense	Non rense	51	rue de Lil	l∈ M	5053700	280200	590		
4	2.01E+11		5	1	21	1845	Nuit sans	Hors agglo	Hors inters	Normale	Non rense	51		M	5054600	280000	590		
5	2.01E+11		5	1	4	1615	Plein jour	Hors agglo	Hors inters	Non rense	Trois vehic	82		M	5098700	240800	590		
6	2.01E+11		5	1	10	1945	Nuit sans	Hors agglo	Hors inters	Pluie leger	Trois vehic	478		M	5096400	247500	590		
7	2.01E+11	1	5	1	28	1415	Plein jour	Hors agglo	Hors inters	Temps ebl	Trois vehic	82		M	5099500	239700	590		
8	2.01E+11		5	1	3	1530	Plein jour	Hors agglo	Hors inters	Non rense	Deux vehic	130		M	0	0	590		
9	2.01E+11		5	1	18	2115	Nuit avec	En agglom	Hors inters	Normale	Trois vehic	404	le hamea	u M	5105200	253300	590		
10	2.01E+11		5	1	25	1715	Plein jour	Hors agglo	Intersection	Temps ebl	Deux vehic	309		M	5097300	257100	590		
11	2.01E+11		5	1	29	2015	Nuit sans	Hors agglo	Hors inters	Normale	Trois vehic	260		M	5106500	252400	590		
12	2.01E+11		5	1	23	230	Nuit sans	Hors agglo	Hors inters	Non rense	Deux vehic	401		M	0	0	590		
13	2.01E+11		5	1	11	615	Nuit avec	En agglom	Hors inters	Non rense	Trois vehic	212	Rue de Li	ll M	5064700	273300	590		
14	2.01E+11		5	1	1	800	Plein jour	Hors agglo	Hors inters	Non rense	Trois vehic	282		M	5081700	248300	590		
15	2.01E+11		5	1	30	1930	Nuit avec	En agglom	Hors inters	Normale	Trois vehic	202		M	5067800	283400	590		
16	2.01E+11		5	1	28	1930	Nuit avec	En agglom	Hors inters	Non rense	Trois vehic	44	rue de W	il <mark> M</mark>	5061500	325200	590		
17	2.01E+11	3	5	1	19	2030	Nuit sans	Hors agglo	Hors inters	Non rense	Deux vehic	144		M	5048800	348700	590		
18	2.01E+11		5	1	9	2000	Nuit avec	En agglom	Hors inters	Normale	Deux vehic	36	41 Rue d'	E M	5015000	387000	590		
19	2.01E+11		5	1	3	730	Nuit sans	Hors agglo	Hors inters	Temps cou	Non rense	531		M	0	0	590		
20	2.01E+11	3	5	1	9	1630	Plein jour	En agglom	Hors inters	Non rense	Non rense	139	RUE H.BA	AF M	5012300	340500	590		
21	2.01E+11		5	1	30	2030	Nuit avec	En agglom	Hors inters	Normale	Trois vehic	139	rue de Pa	ri M	5012300	341500	590		
22	2.01E+11		5	1	31		-	En agglom					RUE Pierr	e M	5007200	333800	590		
23	2.01E+11	1	5	1	18	1930	Nuit sans	Hors agglo	Hors inters	Non rense	Deux vehic	10		M	5016500	313500	590		
24	2.01E+11		5	1	12	1800	Nuit avec	En agglom	Hors inters	Non rense	Deux vehic	449	RUE JEAN	M	0	0	590		
25	2.01E+11		5	1	15	1100	Plein jour	Hors agglo	Hors inters	Non rense	Non rense	186		M	5020100	409600	590		
26	2.01E+11		5	1	13	900	Plein jour	En agglom	Hors inters	Non rense	Trois vehic	421	RUE NEU	VM	5011700	308700	620		
27	2.01E+11		5	1	11	1030	Plein jour	En agglom	Intersection	Normale	Trois vehic	817	4 route N	a M	0	0	620		
20	~ ~~~		-			4400	- ·												

A1		· : >	< _/	fx	Nun	n_Acc																	
4	А	В	С		D	E	F		G	F	ł	1	j	K	L	М		N	0	Р	Q	R	9
1	Num_Acc	catr	voie	v1		v2	circ	r	nbv	pr		pr1	vosp	prof	plan	lartpc	ları	rout	surf	infra	situ	env1	
2	2.01E+11	Route Dep	41		0	В	A sens	unic	2		1	430	0	Non rense	Non rense		0	63	Non rense	. (Non rense	C)
3	2.01E+11	Route nati	41		0			0	2		0	0	Non rense	Non rense	Non rense		0	100	Non rense	(Sur accote	C)
4	2.01E+11	Route nati	41		0			0	0	i e	0	0	Non rense	Non rense	Non rense		0	C	Normale	(Sur accote	C)
5	2.01E+11	Route Dep	916		0		A sens	unic	2		0	0	0	Non rense	Non rense		0	C	Non rense	(Non rense	C)
6	2.01E+11	Route Dep	110		0		A sens	unic	2		24	630	0	Non rense	En courbe		0	59	Normale	(Sur chauss	C)
7	2.01E+11	Route Dep	916		0		A sens	unic	2	9	47	0	0	Non rense	En courbe		0	70	Normale	(Sur accote	C)
8	2.01E+11	Route Dep	600		0		A sens	unic	0	ĺ	6	638	0	Non rense	Non rense		0	78	Non rense	. (Non rense	C)
9	2.01E+11	Route Dep	947		0		A sens	unic	2		53	900	0	Non rense	En courbe		0	69	Non rense	. (Sur chauss	C)
10	2.01E+11	Route Dep	55		0		A sens	unic	2		3	200	0	0	Non rense		0	67	Non rense	. (Non rense	C)
11	2.01E+11	Route nati	1		0			0	2		32	13	0	Non rense	Non rense		0	70	Normale	(Non rense	C)
12	2.01E+11	Route nati	42		0		A sens	unic	2		1	860	0	Non rense	Non rense		0	75	0	(Non rense	C)
13	2.01E+11	Route Dep	945		0		A sens	unic	2		3	600	0	Non rense	Non rense		0	90	Normale	(Sur bande	C)
14	2.01E+11	Route Dep	916		0		A sens	unic	2		25	500	0	Non rense	En courbe		0	71	Non rense	(Sur chauss	C)
15	2.01E+11	Route Dep	945		0		A sens	unic	0	1	0	0	0	0	Non rense		0	59	Normale	(Non rense	C)
16	2.01E+11	Route Dep	2		0		A sens	unic	2		0	0	0	Non rense	Non rense		0	106	Normale	(Non rense	C)
17	2.01E+11	Route Dep	368		0		A sens	unic	2	9	1	800	0	Non rense	Non rense		0	57	Corps gras	s (Non rense	C)
18	2.01E+11	Route Dep	124		0		A sens	unic	0	i .	14	200	0	Plat	Non rense		0	58	Normale	(Non rense	C)
19	2.01E+11	Route Dep	21		0		A sens	unic	2		0	0	0	Plat	En courbe		0	65	Boue	(Non rense	C)
20	2.01E+11	Voie Comr	0		0			0	1		0	0	0	Non rense	Non rense		0	C	Non rense	Bretelle d	Non rense	C)
21	2.01E+11	Voie Com	0		0		A sens	unic	2	9	0	0	0	Non rense	En courbe		0	0	Normale	(Non rense	C)
22	2.01E+11	Route Dep	16		0		A sens	unic	0	i	0	0	0	0	Non rense		0	69	Non rense	. (Non rense	C)
23	2.01E+11	Route nati	30		0		A sens	unic	2		4	700	0	Non rense	Non rense		0	C	Non rense	(Non rense	C)
24	2.01E+11	Route Dep	158		0	Α		0	0	i e	0	0	0	Non rense	Non rense		0	62	Non rense	. (Non rense	C)
25	2.01E+11	Route Dep	963		0		A sens	unic	2		28	660	0	Non rense	Partie rect		0	55	Boue	(Non rense	C)
26	2.01E+11	Route Dep	15		0		A sens	unic	0		4	450	0	0	Non rense		0	75	Non rense	. (Non rense	C)
27	2.01E+11	Route nati	39		0		Bidirect	tior	2		0	0	Non rense	Non rense	Non rense		44	75	Normale	(Sur accote	C)
20	2045.44	ь	47	100	_						_			ь.				70				,	

)	POSSIBLE	DATA LOSS	Some fea	f _x		n_Acc	you save this wo	orkbook in	the comma-	delimited (.c	sv) format. To	o preserve th	ese features	s, save it in a	an Excel file f	ormat.	Don't show	again	Save As
	Α	В	С		D	E	F	G	Н	1 1	J	K	L	M	N	0	Р	Q	R
r	lum_Acc	catr	voie	v1		v2	circ	nbv	pr	pr1	vosp	prof	plan	lartpc	larrout	surf	infra	situ	env1
	2.01E+11	Route Dep	41		0	В	A sens unio		2	430	0	Non rense	Non rense	(63	Non rense	0	Non rense	
	2.01E+11	Route nati	41		0		0		2 (0	Non rense	Non rense	Non rense	(100	Non rense	0	Sur accote	
	2.01E+11	Route nati	41		0		0	i i	0 (0	Non rense	Non rense	Non rense	(0 0	Normale	0	Sur accote	
	2.01E+11	Route Dep	916		0		A sens unio		2 (0	0	Non rense	Non rense	(0	Non rense	0	Non rense	
	2.01E+11	Route Dep	110	i	0		A sens unio		2 24	630	0	Non rense	En courbe	(59	Normale	0	Sur chauss	
	2.01E+11	Route Dep	916		0		A sens unio		2 47	7 0	0	Non rense	En courbe	(70	Normale	0	Sur accote	
	2.01E+11	Route Dep	600	i	0		A sens unio	i i	0 6	638	0	Non rense	Non rense	(78	Non rense	0	Non rense	
	2.01E+11	Route Dep	947		0		A sens unio		2 53	900	0	Non rense	En courbe	(69	Non rense	0	Sur chauss	
	2.01E+11	Route Dep	55		0		A sens unio		2 3	200	0	0	Non rense	(67	Non rense	0	Non rense	
1	2.01E+11	Route nati	1		0		0		2 32	2 13	0	Non rense	Non rense	(70	Normale	0	Non rense	
	2.01E+11	Route nati	42		0		A sens unio		2 :	860	0	Non rense	Non rense	(75	0	0	Non rense	
	2.01E+11	Route Dep	945		0		A sens unio		2 3	600	0	Non rense	Non rense	(90	Normale	0	Sur bande	
	2.01E+11	Route Dep	916		0		A sens unio		2 25	500	0	Non rense	En courbe	(71	Non rense	0	Sur chauss	
	2.01E+11	Route Dep	945		0		A sens unio		0 (0	0	0	Non rense	(59	Normale	0	Non rense	
	2.01E+11	Route Dep	2		0		A sens unio		2 (0	0	Non rense	Non rense	(106	Normale	0	Non rense	
	2.01E+11	Route Dep	368		0		A sens unio		2 :	800	0	Non rense	Non rense	(57	Corps gras	0	Non rense	
	2.01E+11	Route Dep	124		0		A sens unio	i i	0 14	200	0	Plat	Non rense	(58	Normale	0	Non rense	
	2.01E+11	Route Dep	21		0		A sens unio		2 (0	0	Plat	En courbe	(65	Boue	0	Non rense	
	2.01E+11	Voie Comr	0	1	0		0		1 (0	0	Non rense	Non rense	(0	Non rense	Bretelle d'	Non rense	
	2.01E+11	Voie Comr			0		A sens unio		2 (0	0	Non rense	En courbe	(0 0	Normale	0	Non rense	
	2.01E+11	Route Dep	16		0		A sens unio		0 (0	0	0	Non rense	(69	Non rense	0	Non rense	
	2.01E+11	Route nati	30	i .	0		A sens unio		2 4	700	0	Non rense	Non rense	(0 0	Non rense	0	Non rense	
	2.01E+11	Route Dep				Α	0		0 (_	Non rense	Non rense	(Non rense	0	Non rense	
	2.01E+11	Route Dep	963		0		A sens unio		2 28	660	0	Non rense	Partie rect	. (55	Boue	0	Non rense	
	2.01E+11	Route Dep			0		A sens unio		0 4	450	0	0	Non rense	(75	Non rense	0	Non rense	
-		Route nati			0		Bidirection		2 (0	Non rense	Non rense	Non rense	44	1 75	Normale	0	Sur accote	
11 3	2045.44	12 0 0	47				20 2	193	,	7 2 2	1	n .	04/4	100	72			A.1	

4	Α	В	С	D	Е	F	G	Н	1 1	J
1	Num Acc	17	catv	occutc	obs	obsm	choc	many	num veh	,
2	2.01E+11		VU seul 1,5	0	0	Aucun	Non rense	Non rense		
3	2.01E+11		VU seul 1,5		0	Aucun	Arriere gau	Dans le co	B02	
4	2.01E+11	0	VU seul 1,5	0	0	Aucun		Deporte a	2000	
5	2.01E+11	0	Bicyclette	0	0	Aucun	Non rense	Non rense	B02	
6	2.01E+11	0	Bicyclette	0	0	Aucun	Non rense	Non rense	A01	
7	2.01E+11	0	Bicyclette	0	0	Aucun	Non rense	Non rense	B02	
8	2.01E+11	0	VU seul 1,5	0	0	Aucun	Non rense	Changeant	A01	
9	2.01E+11	0	VU seul 1,5	0	0	Aucun	0	Non rense	B02	
10	2.01E+11	0	VU seul 1,5	0	0	Aucun	Arriere gau	Non rense	C03	
11	2.01E+11	0	VU seul 1,5	0	Glissiere b	0	Arriere gau	Non rense	A01	
12	2.01E+11	0	VU seul 1,5	0	Batiment,	0	Arriere dro	Non rense	A01	
13	2.01E+11	0	PL seul > 7	0	0	Aucun	Avant	Depassant	A01	
14	2.01E+11	0	VU seul 1,5	0	0	Aucun	Aucun	Non rense	B02	
15	2.01E+11	0	VU seul 1,5	0	Batiment,	0	Arriere gau	Changeant	A01	
16	2.01E+11	0	VU seul 1,5	0	0	Aucun	Non rense	Non rense	A01	
17	2.01E+11	0	VU seul 1,5	0	0	Aucun	Arriere gau	Depassant	B02	
18	2.01E+11	0	VU seul 1,5	0	0	0	Aucun	Non rense	A01	
19	2.01E+11	0	VU seul 1,5	0	0	Aucun	Aucun	Tournant a	A01	
20	2.01E+11	0	VU seul 1,5	0	0	0	Arriere gau	Non rense	B02	
21	2.01E+11	0	VU seul 1,5	0	Glissiere b	0	Non rense	0	A01	
22	2.01E+11	0	VU seul 1,5	0	Sans objet	0	Avant	Non rense	A01	
23	2.01E+11	0	VU seul 1,5	0	0	Non rense	Non rense	Non rense	A01	
24	2.01E+11	Non rense	VU seul 1,5	0	0	Non rense		Inconnue		
25	2.01E+11	0	VU seul 1,5	0	0	Aucun	Arriere dro	Non rense	A01	
26	2.01E+11	0	VU seul 1,5	0	0	Aucun	Non rense	Non rense	B02	
27	2.01E+11	0	VU seul 1,5		0	Aucun	Aucun	Deporte a	A01	
20	2045.44		n (0		A.1		D02	

Merci pour votre attention

Samar Neji Mohamed Aziz Njaimi Maroua Jbeli Malik nairi Hamza Ben Torkia Khalil Khaled