

# Gravité des Accidents de vélo

Cette présentation vise à analyser la gravité des accidents de vélo. Nous allons explorer les données disponibles. Nous identifierons les facteurs clés influençant la gravité. Enfin, nous proposerons des solutions pour améliorer la sécurité.



# Introduction: Comprendre la Gravité

## Contexte

Des milliers d'accidents graves surviennent chaque année. La compréhension des facteurs est primordiale.

## Problématique

Peut-on prédire la gravité d'un accident ? Quels sont les facteurs les plus déterminants ?

# Données Utilisées: Base de Données

## Source

Base de données des accidents corporels (data.gouv.fr).

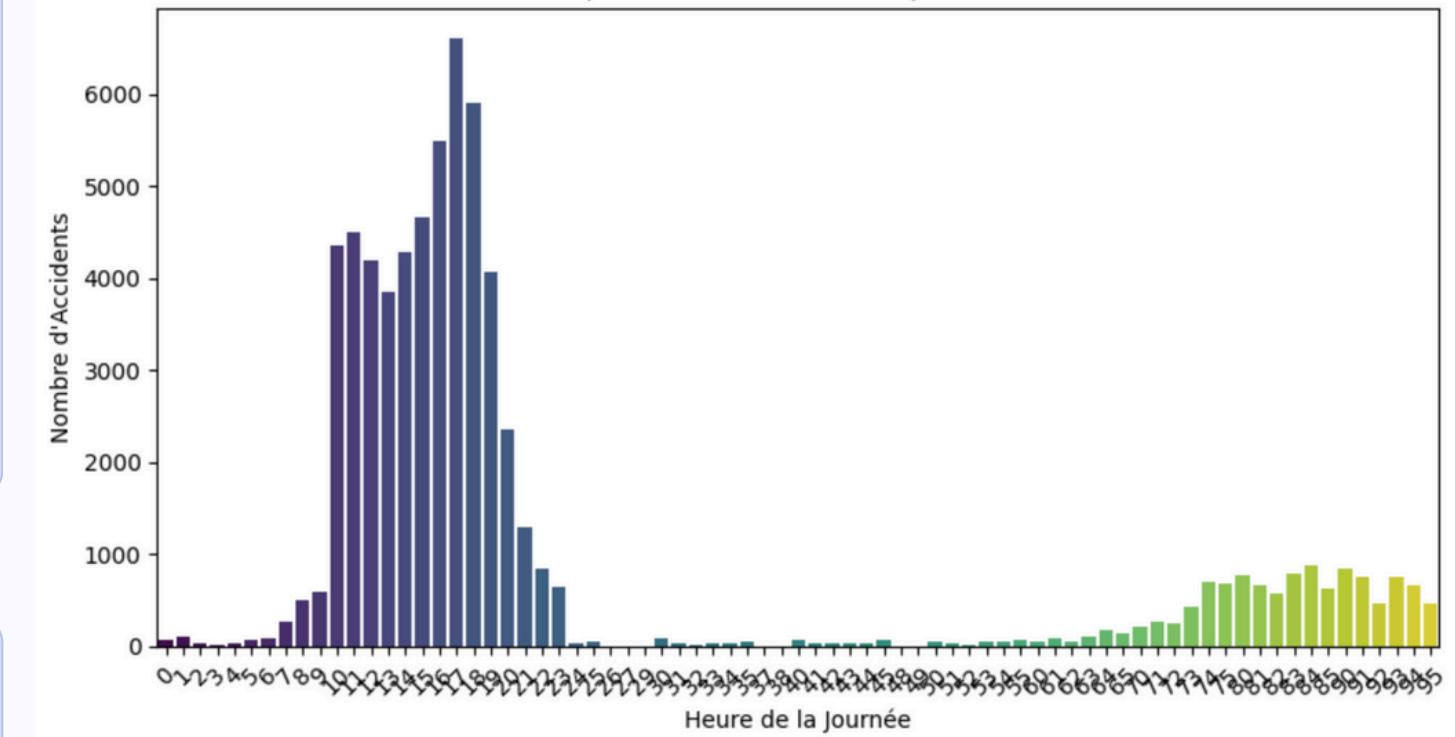
## Variables

Localisation, heure, conditions météo, infrastructure.

## Variable Cible

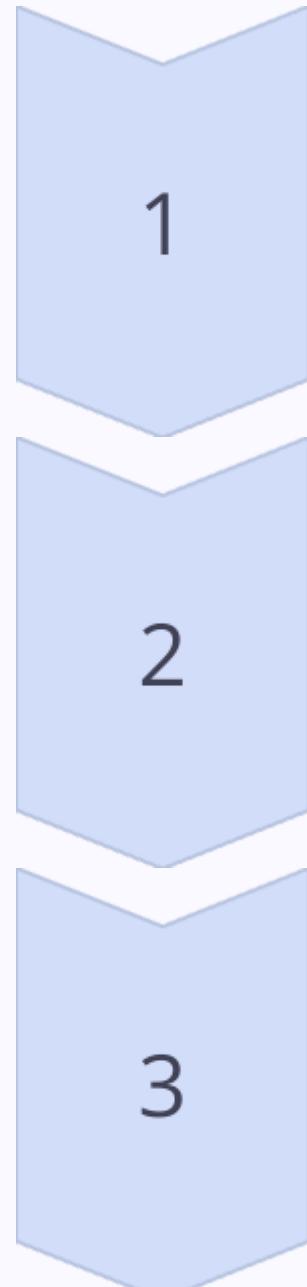
grav (Gravité de l'accident).

Répartition des Accidents par Heure



Num_Acc	date	an	mois	jour	hrmn	dep	com	lat	long	...	age	trajet	secuexist	equipement	obs	obsm	choc	manv	vehiculeid	heure
200500000030	2005-01-13	2005	janvier	jeudi	19:45	62	62331	50.300	2.840	...	58.0	5.0	3	0	0.0	2.0	8.0	11.0	200500000030B02	19
200500000034	2005-01-19	2005	janvier	mercredi	10:45	62	62022	0.000	0.000	...	20.0	5.0	3	0	0.0	2.0	1.0	1.0	200500000034B02	10
200500000078	2005-01-26	2005	janvier	mercredi	13:15	02	02173	0.000	0.000	...	71.0	5.0	2	2	0.0	2.0	1.0	1.0	200500000078B02	13
200500000093	2005-01-03	2005	janvier	lundi	13:30	02	02810	49.255	3.094	...	51.0	4.0	3	0	0.0	2.0	3.0	21.0	200500000093B02	13
200500000170	2005-01-29	2005	janvier	samedi	18:30	76	76196	0.000	0.000	...	74.0	5.0	1	9	0.0	2.0	4.0	2.0	200500000170A01	18

# Prétraitement des Données



## Nettoyage

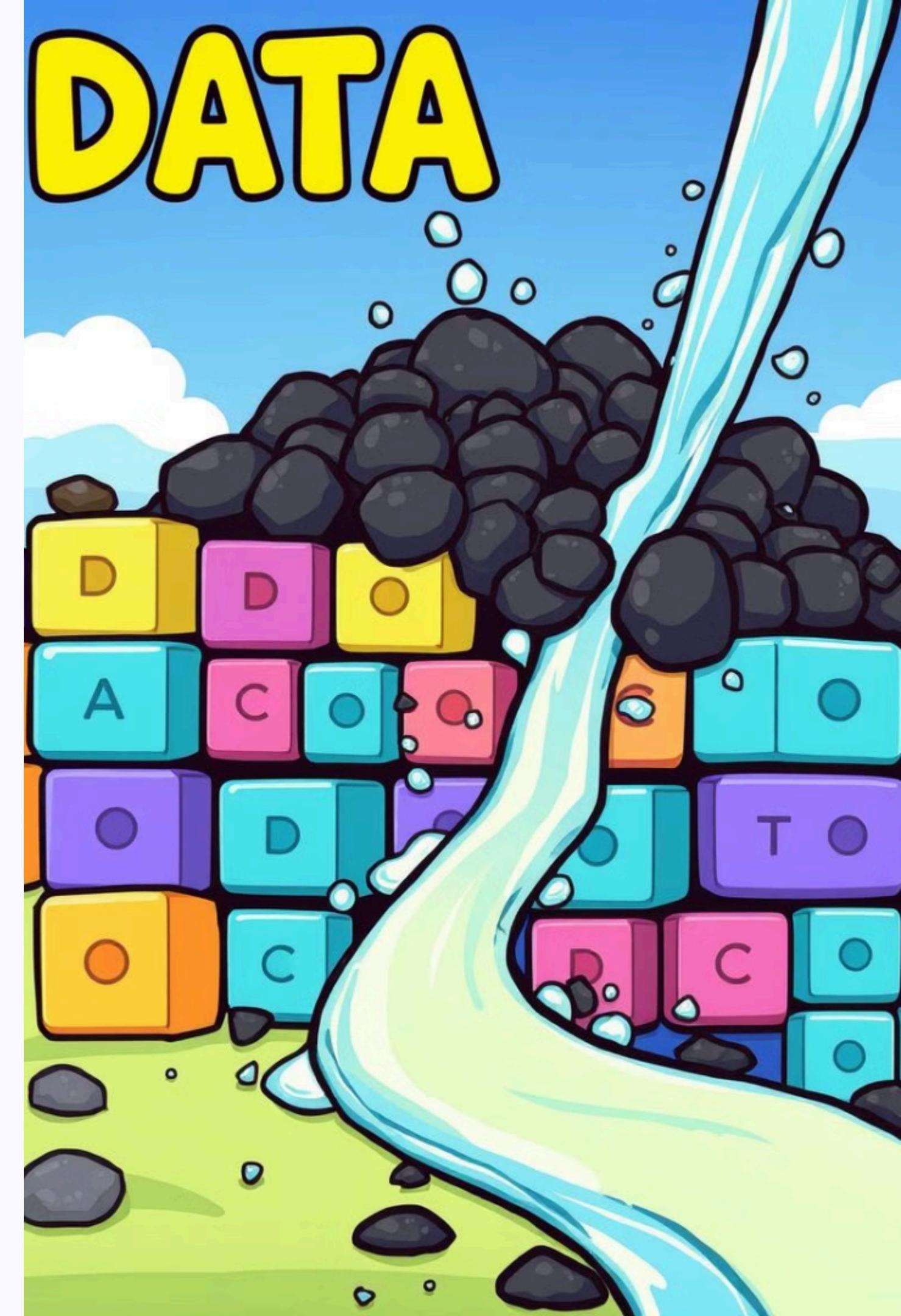
Suppression des accidents de faible gravité. Suppression des colonnes avec trop de valeurs manquantes.

## Correction

Correction du format des données (longitude, latitude, âge).

## Gestion

Médiane pour les variables numériques. Mode pour les variables catégoriques.



# Analyse Exploratoire

## 1 Distribution

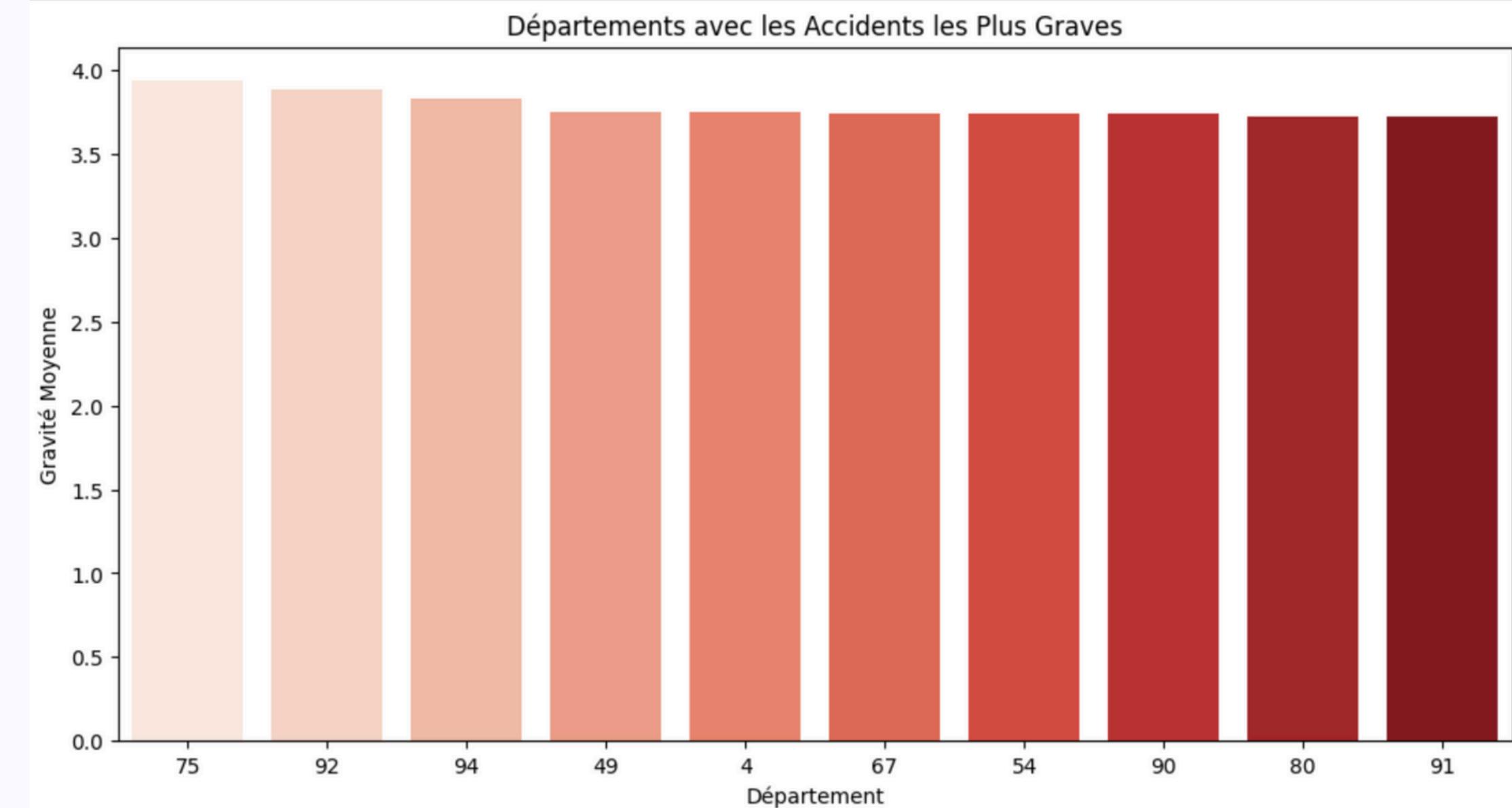
Majorité d'accidents  
moyennement graves.

## 2 Répartition

Pics d'accidents aux heures  
de pointe.

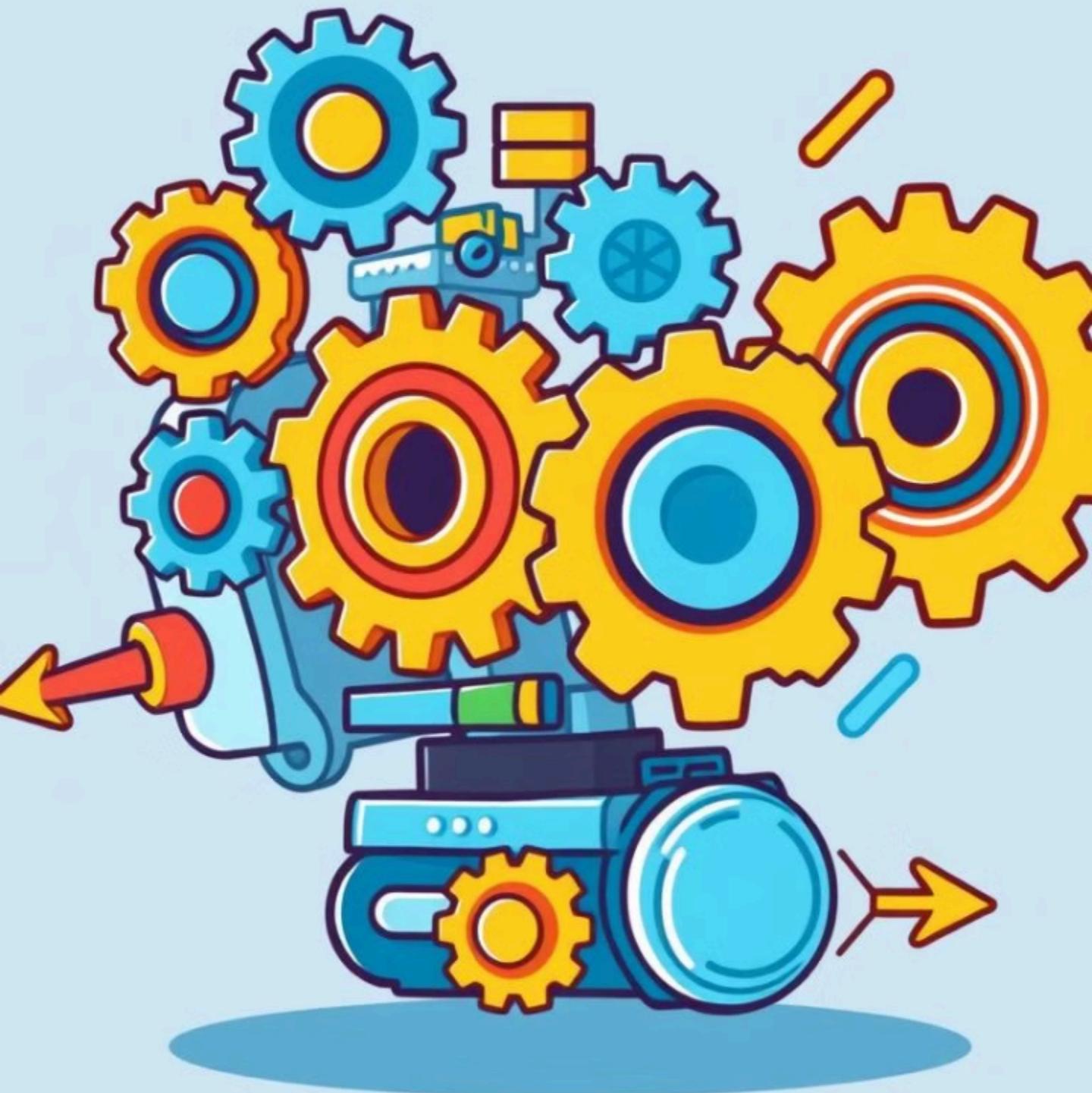
## 3 Carte

Concentration en zones urbaines.



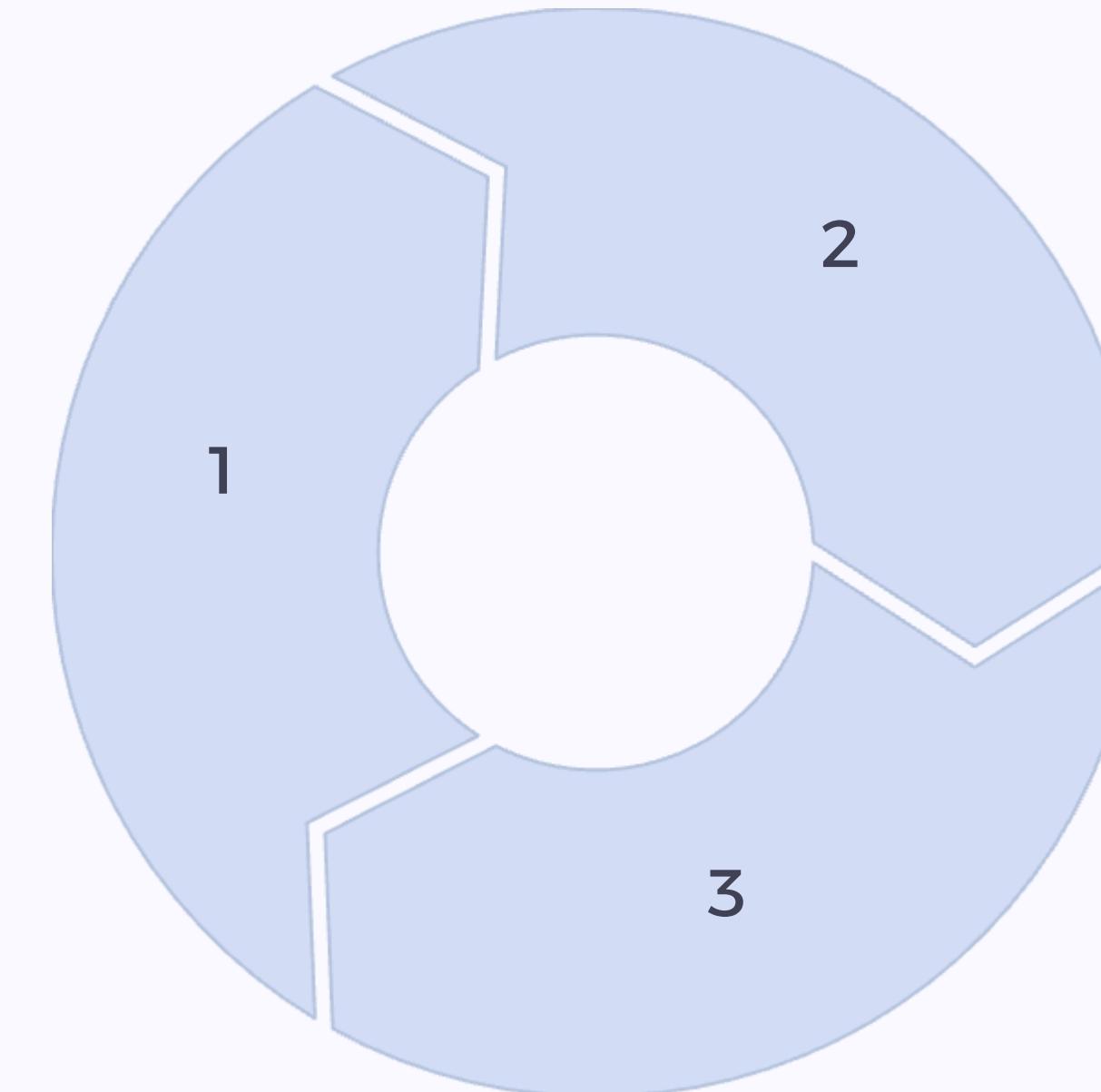
# Feature Engineering

- 1 Facteurs retenus  
Heure, localisation, conditions routières.
- 2 Extraction  
Nouvelles variables : heure de l'accident.
- 3 Normalisation  
Pour une meilleure comparabilité.



# Modélisation: Choix des Algorithmes

Régression Logistique  
Baseline simple.



k-Nearest Neighbors  
Prise en compte des similarités.

Random Forest  
Modèle puissant.

# Résultats: Comparaison des Modèles

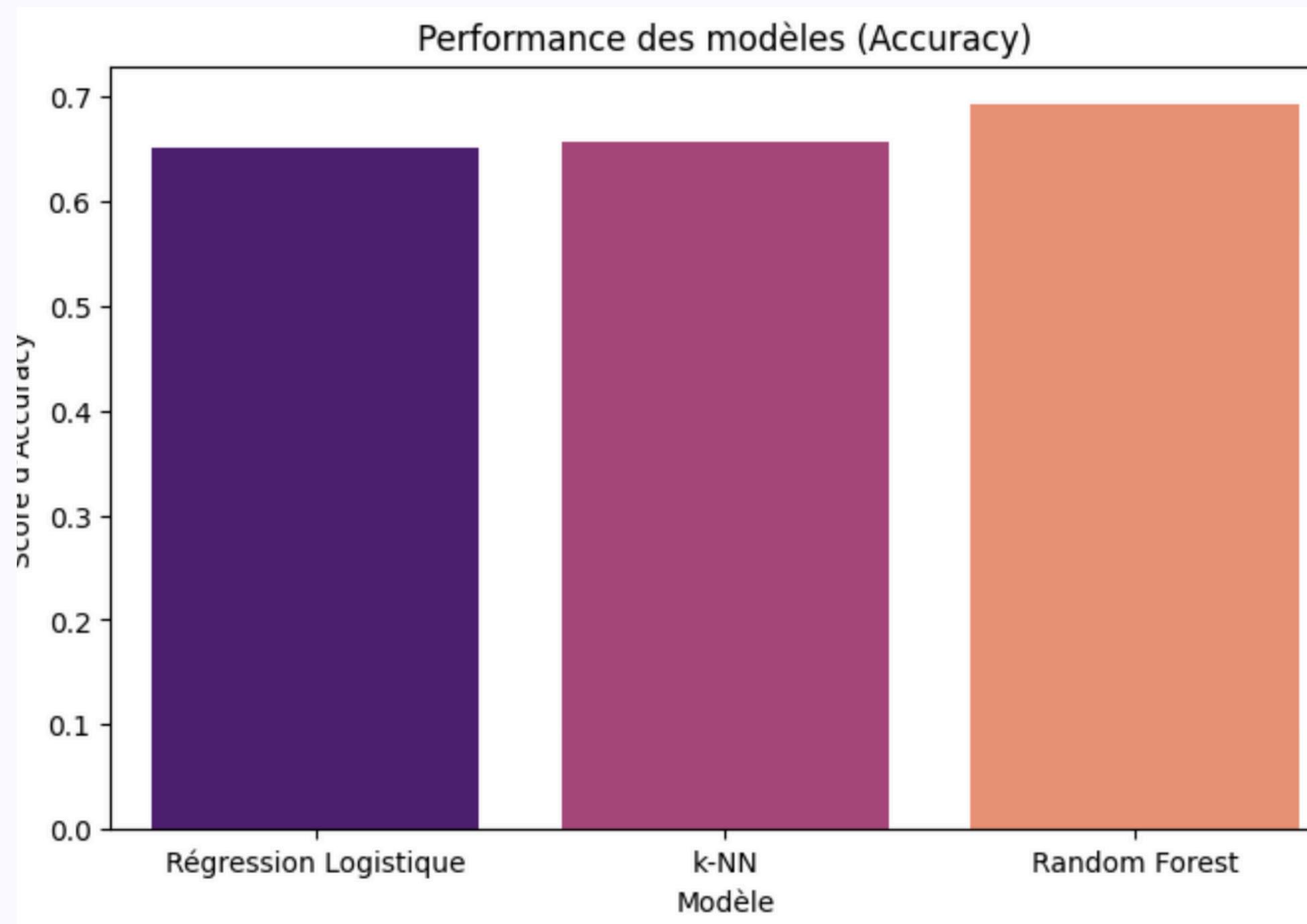


Régression Logistique: 72%.

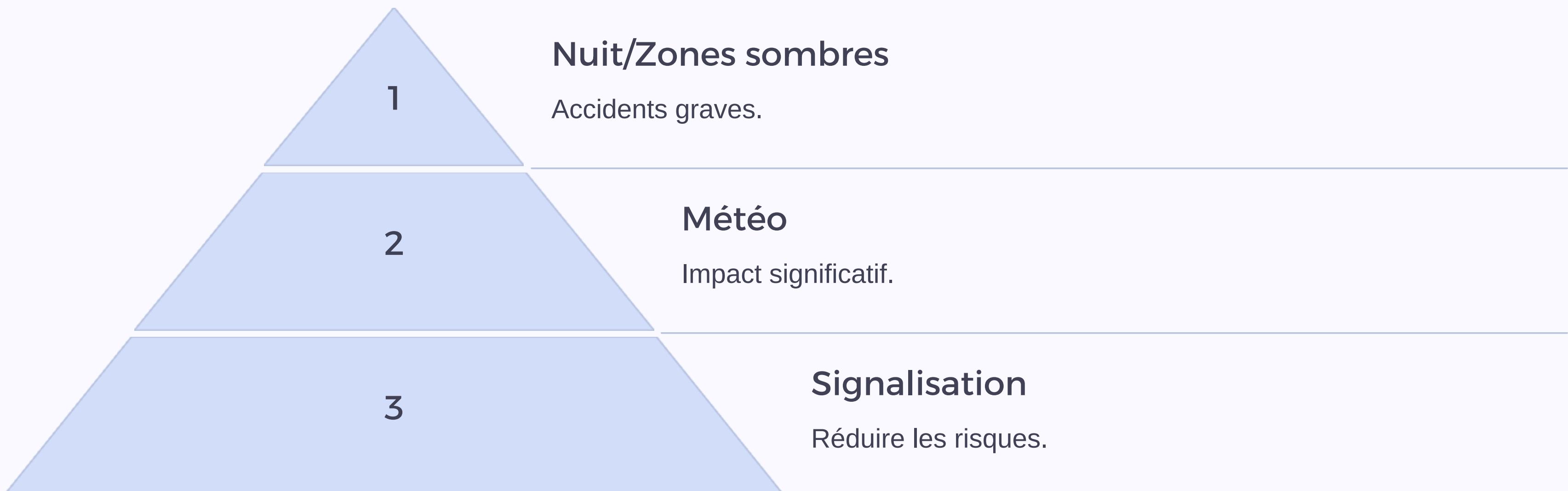


kNN: 74%.

Random Forest: **79%**.



# Analyse Business: Insights Clés





# Conclusion et Perspectives

## Synthèse

Modèle fiable. Prédiction de la gravité.

## Prévention

Insights exploitables pour améliorer.

## Amélioration

Données météo et trafic en temps réel.