



ANALYSE DES INCENDIES DE FORÊT EN FRANCE : DES DONNÉES À L'ACTION

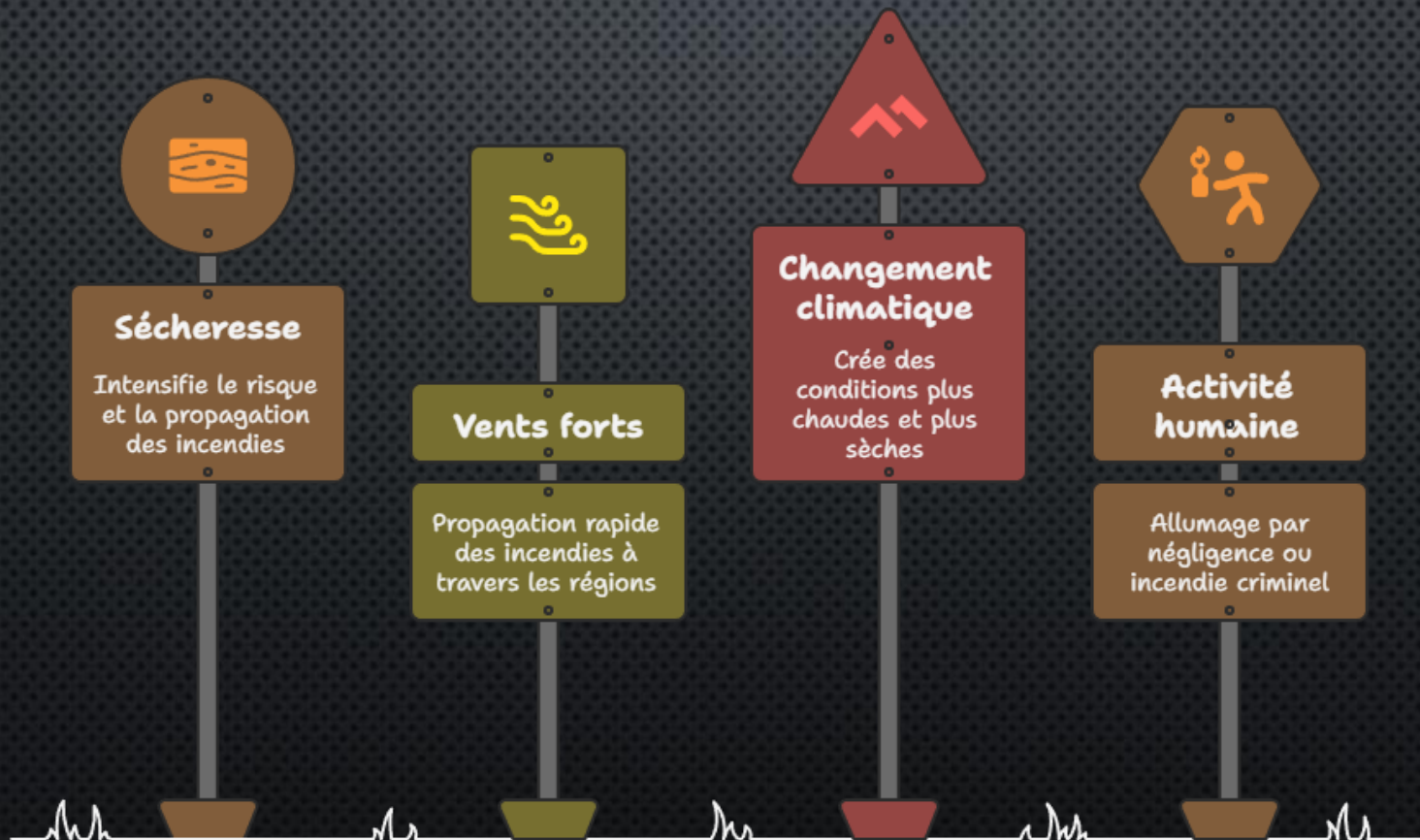
MEMBRE DU PROJET : KAHINA ABBAD

DATE : 16 SEPTEMBRE 2025

PLAN

- INTRODUCTION & CONTEXTE
- ENJEUX DU PROJET
- DONNÉES UTILISÉES
- DÉFIS RENCONTRÉS
- PRÉPARATION & NETTOYAGE
- ANALYSE EXPLORATOIRE (EDA)
- MODÉLISATION PRÉDICTIVE
- RÉSULTATS ET APPORTS
- PISTES D'AMÉLIORATION

INTRODUCTION & CONTEXTE



ENJEUX DU PROJET



DONNÉES UTILISÉES



BDIFF

Base de données sur
les incendies de
forêts de 1973 à
2025.



INSEE

Population, densité
et superficie des
communes.



Référentiels géographiques

Codes INSEE et
coordonnées de
latitude/longitude.



Décision méthodologique

Garder uniquement
la France
métropolitaine et la
Corse, exclure les
DOM-COM.

DÉFIS RENCONTRÉS





PRÉPARATION & NETTOYAGE

TYPAGES DES DONNÉES

```
1 df.info()
```

✓ 0.0s

Avant

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 128916 entries, 0 to 128915  
Data columns (total 18 columns):
```

#	Column	Non-Null Count	Dtype
0	annee	128916 non-null	object
1	numero	128916 non-null	object
2	departement	128916 non-null	object
3	code_insee	128916 non-null	object
4	commune_incendie	128916 non-null	object
5	date_alerte	128916 non-null	object
6	surface_parcourue_m2	128916 non-null	object
7	type_peuplement	128916 non-null	object
8	nature	128916 non-null	object
9	nb_decès	128916 non-null	object
10	nb_batiments_totalelement_detruits	128916 non-null	object
11	nb_batiments_partiellement_detruits	128916 non-null	object
12	commune_ref	128035 non-null	object
13	lat	128035 non-null	object
14	lon	128035 non-null	object
15	population	128035 non-null	object
16	superficie_km2	128035 non-null	object
17	densite	128035 non-null	object

```
dtypes: object(18)  
memory usage: 17.7+ MB
```

```
1 df.info()
```

✓ 0.0s

Après

```
<class 'pandas.core.frame.DataFrame'>
```

```
Index: 127302 entries, 0 to 128915
```

```
Data columns (total 18 columns):
```

#	Column	Non-Null Count	Dtype
0	annee	127302 non-null	object
1	numero_incendie	127302 non-null	int64
2	code_departement	127302 non-null	object
3	code_insee	127302 non-null	object
4	commune_incendie	127302 non-null	object
5	surface_brulee_m2	127302 non-null	int64
6	type_peuplement	127302 non-null	int64
7	nature_incendie	127302 non-null	object
8	commune_ref	127302 non-null	object
9	latitude	127302 non-null	float64
10	longitude	127302 non-null	float64
11	population_commune	127302 non-null	int64
12	superficie_km2	127302 non-null	int64
13	densite_population	127302 non-null	float64
14	mois	127302 non-null	int32
15	jour	127302 non-null	int32
16	heure	127302 non-null	int32
17	type_peuplement_label	127302 non-null	object

```
dtypes: float64(3), int32(3), int64(5), object(7)
```

```
memory usage: 17.0+ MB
```


LES VALEURS MANQUANTES

```
1 detecter_valeurs_manquantes(df)
```

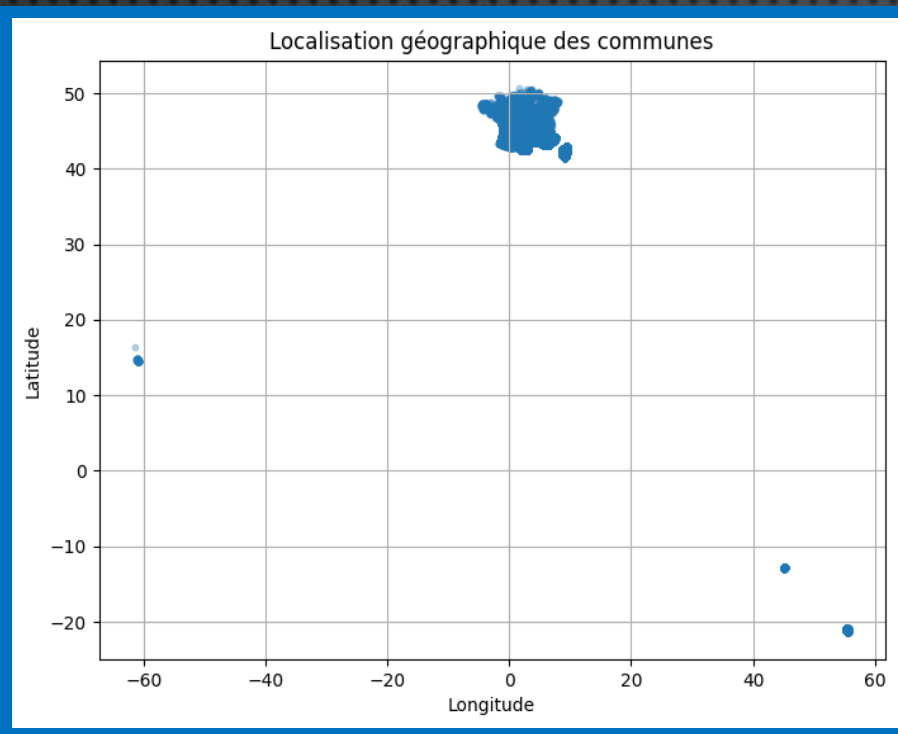
✓ 1.7s

...	# nb_valeurs_manquantes	# pourcentage
nb_batiments_partiellement_detruits	128041	99.32
nb_deces	128041	99.32
nb_batiments_totalement_detruits	128040	99.32
nature		
type_peuplement		
lat		
lon		
population		
superficie_km2		
commune_ref		
densite		
commune_incendie		
numero		
annee		
surface_parcourue_m2		
departement		
code_insee		
date_alerte		

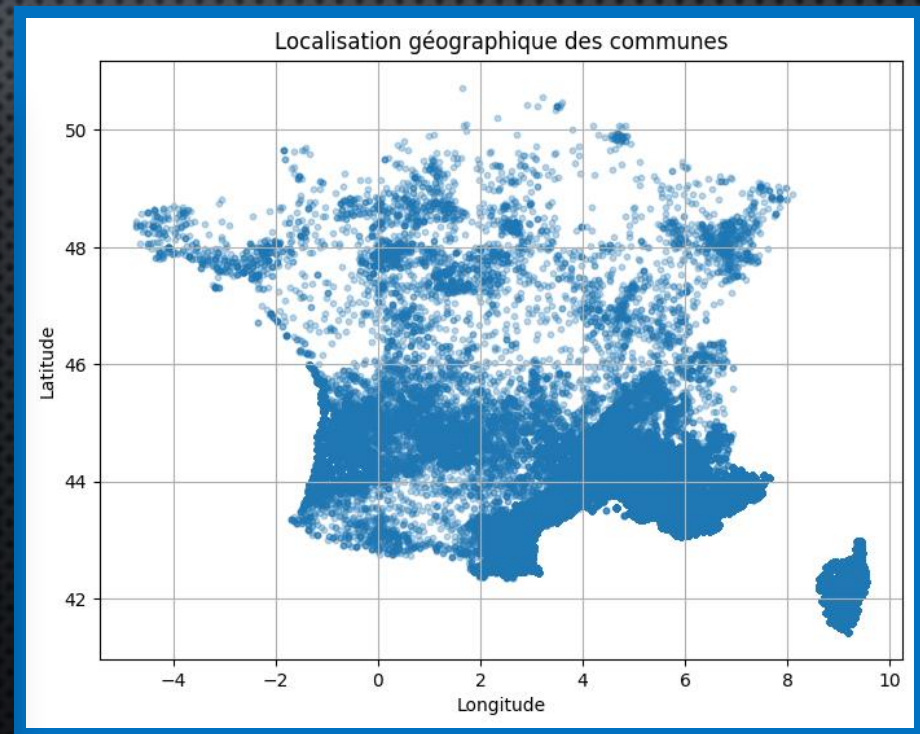
Code	Libellé
1	Landes / Garrigues / Maquis
2	Taillis
3	Futaies feuillues
4	Futaies résineuses
5	Futaies mélangées
6	Régénération / Reboisement

LES VALEURS ABERRANTES

Avant

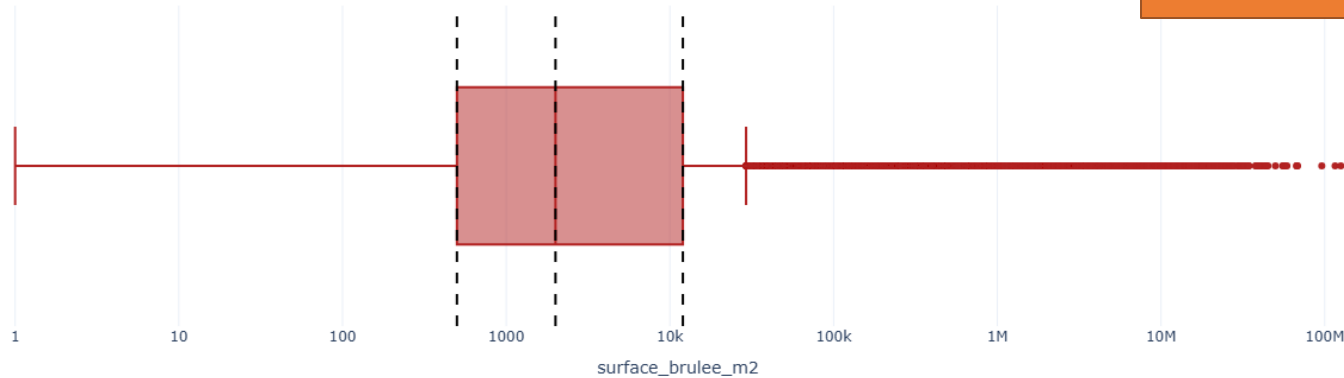


Après

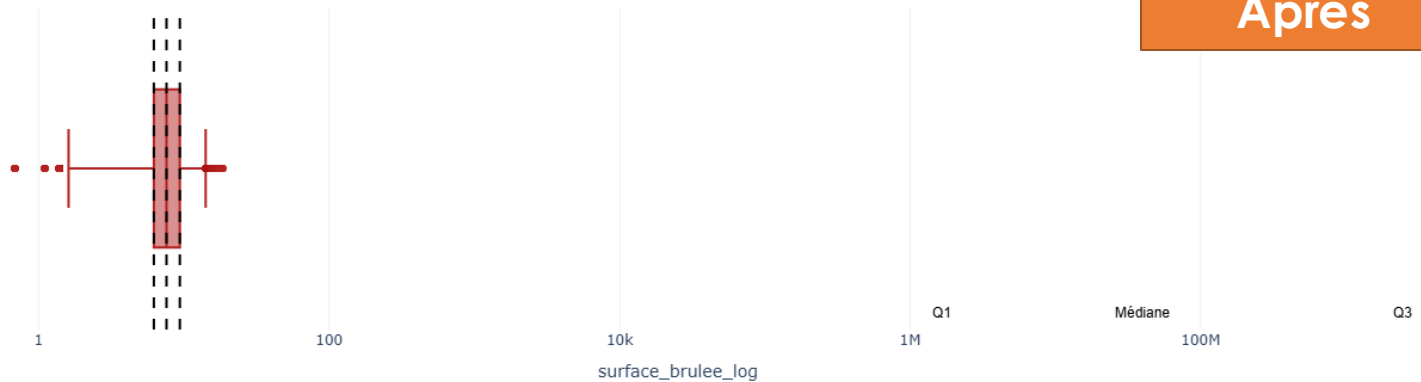


LES VALEURS ABERRANTES

Distribution de surface_brulee_m2 avec quartiles



Distribution de surface_brulee_log avec quartiles



FEATURE ENGINEERING

Nouvelles variables d'incendie



Extraction de la date

Extraction de l'année, du mois et du jour à partir des dates.



Regroupement de la végétation

Regroupement des types de végétation en catégories simplifiées.



Indicateurs binaires

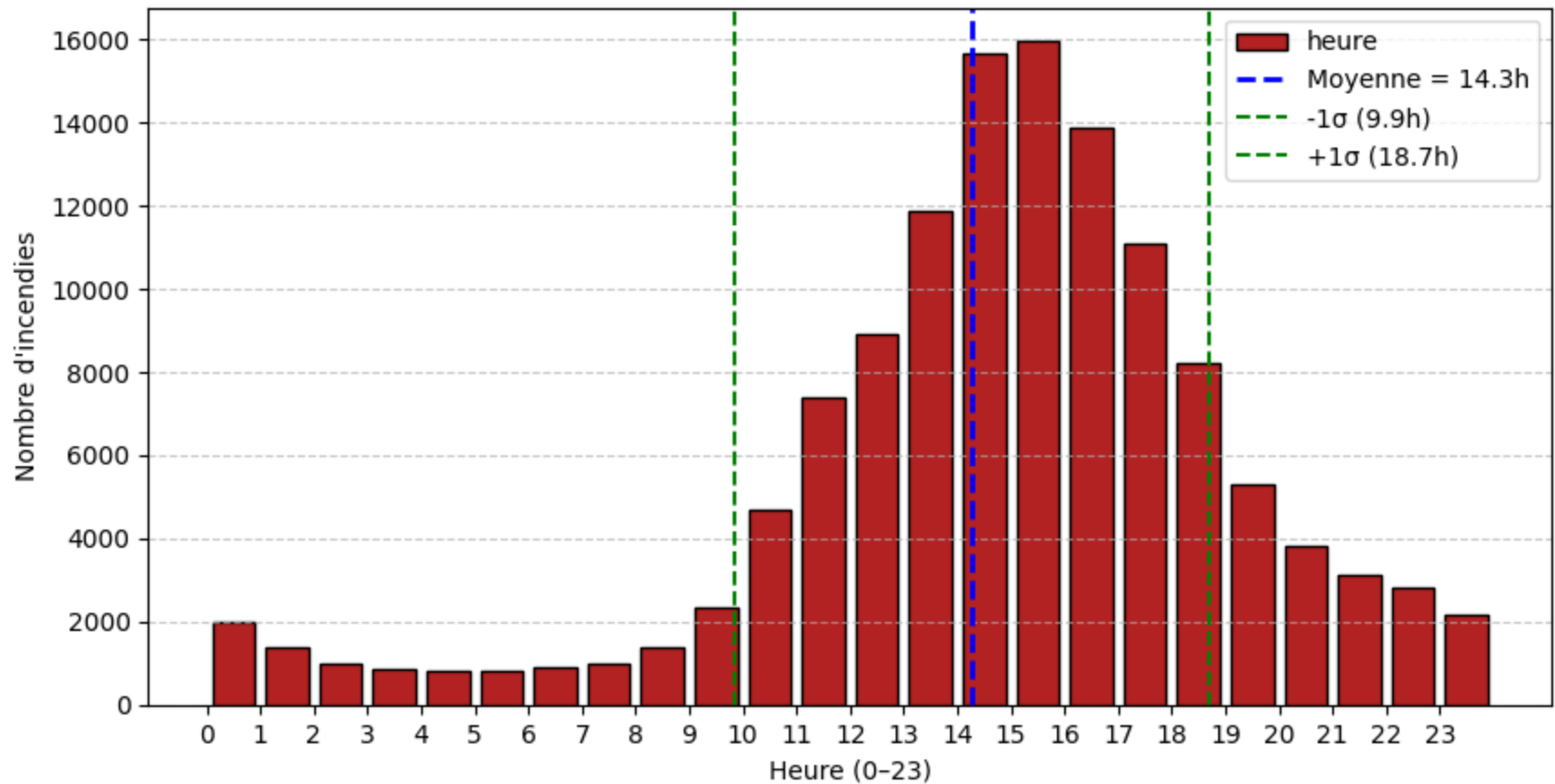
Indicateurs binaires pour identifier si un incendie a touché des habitations.



ANALYSE EXPLORATOIRE (EDA)

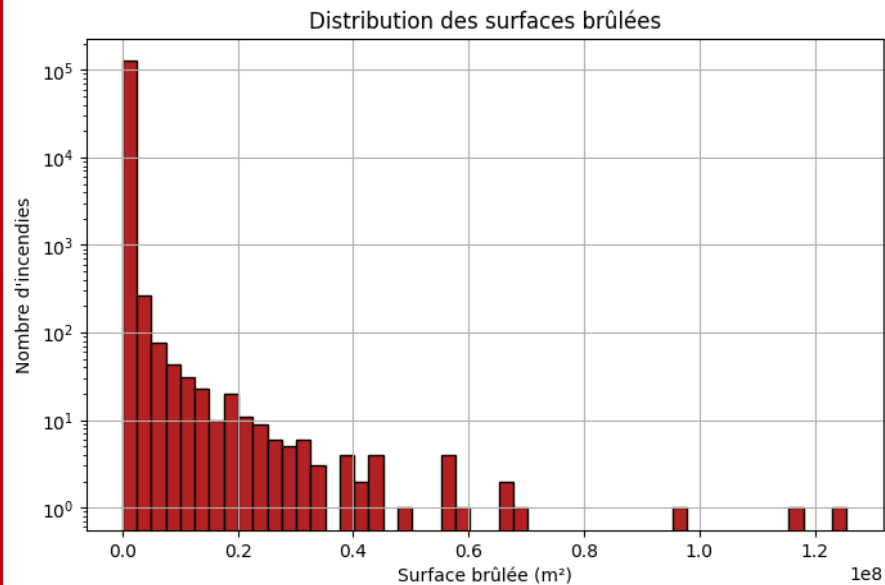
ANALYSE UNIVARIÉE

Répartition des incendies par heure de la journée

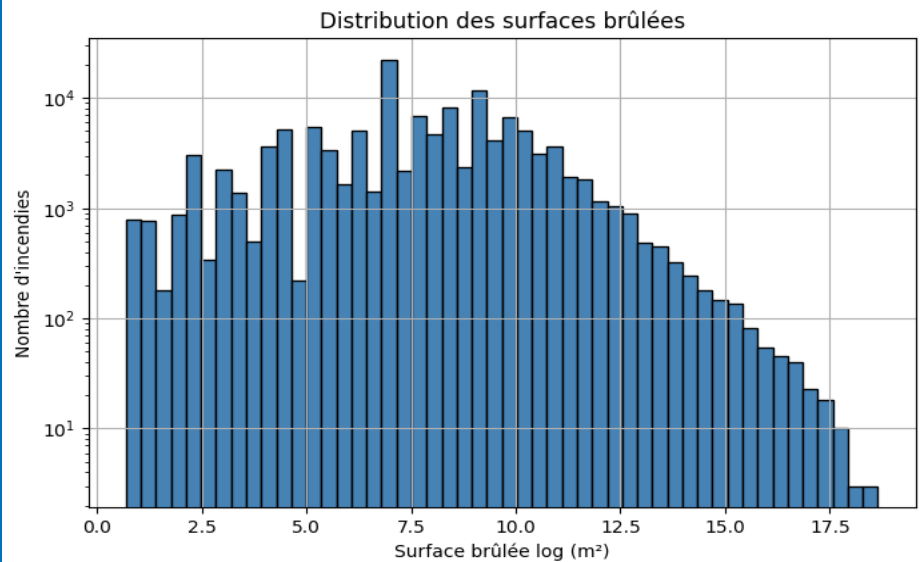


ANALYSE UNIVARIÉE

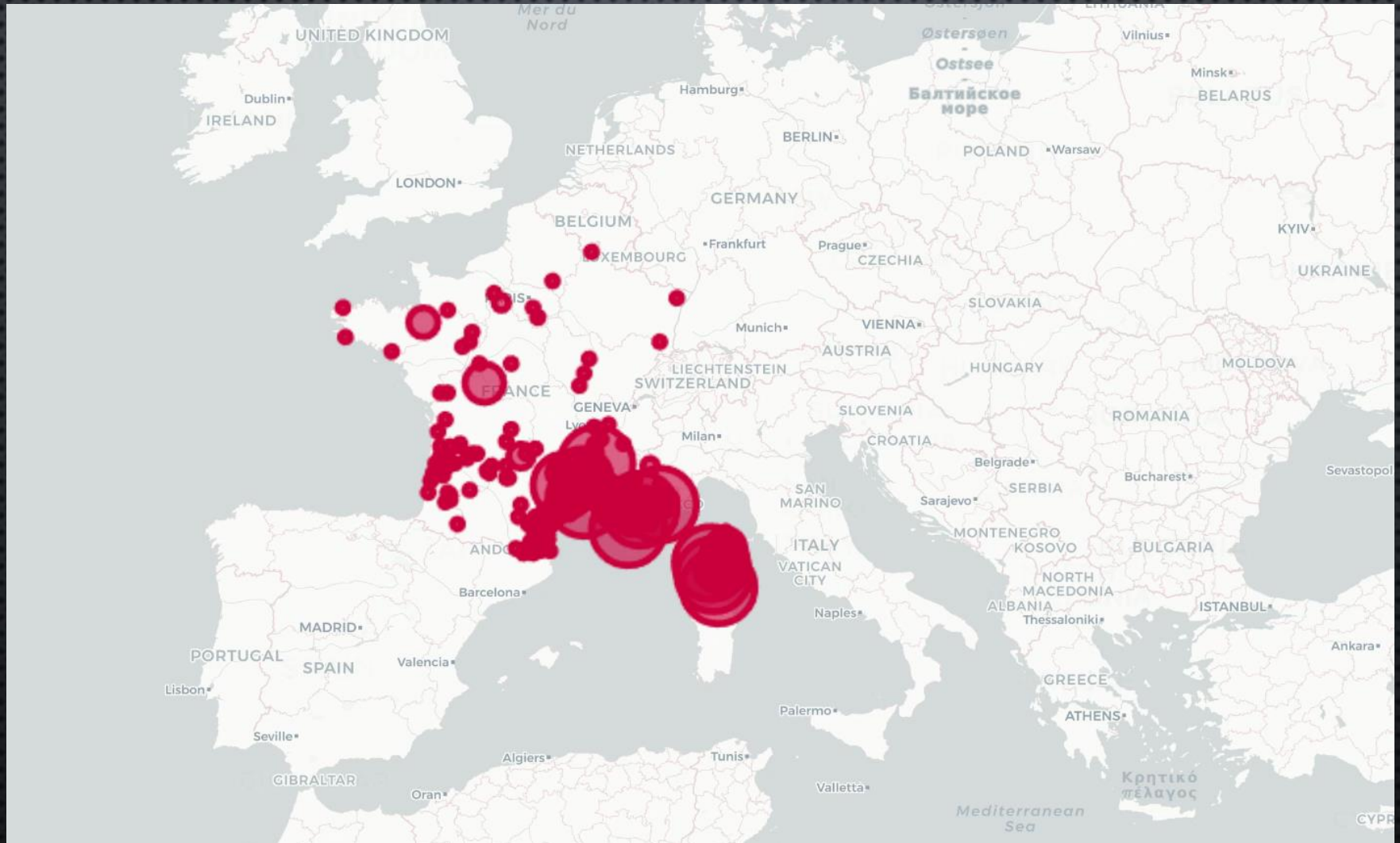
Avant



Avant



ANALYSE MULTIVARIÉE



MODÉLISATION PRÉDICTIVE

L'incendie est-il grave ?



Incendies graves

Le feu couvre une superficie de 100 000 m² ou plus, nécessitant une attention immédiate.

VS

Incendies non graves

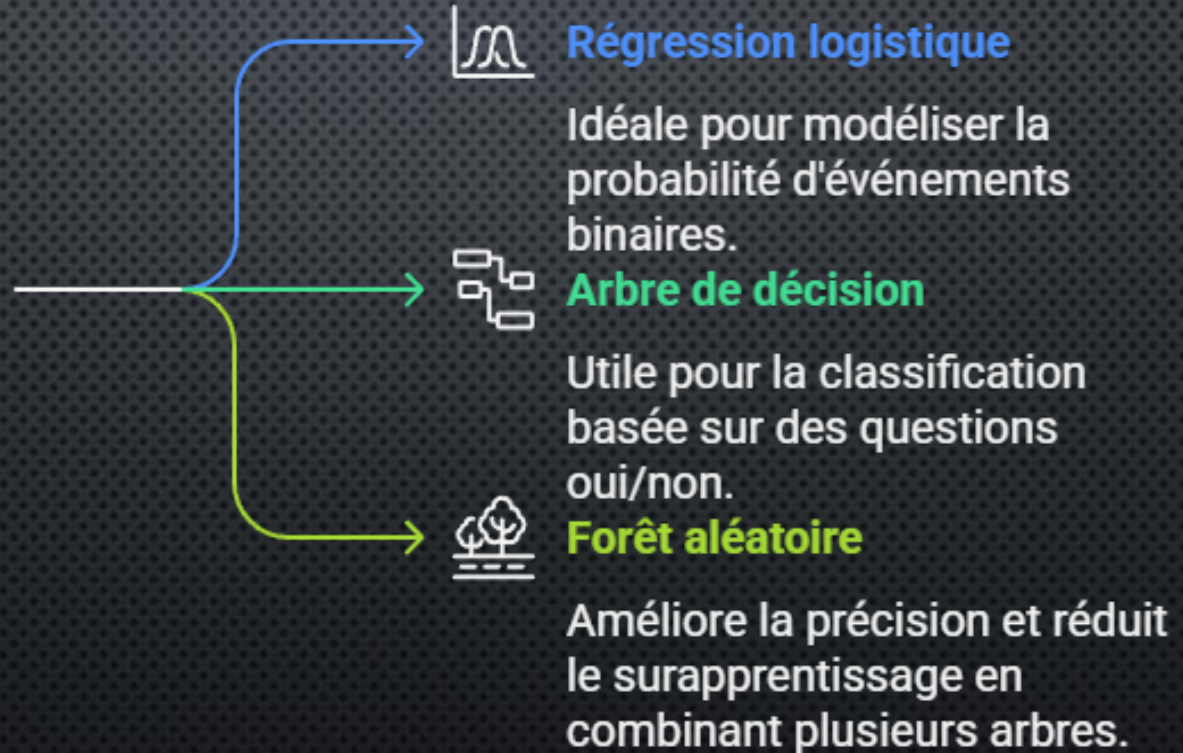
Le feu couvre une superficie inférieure à 100 000 m², nécessitant moins de ressources.



MODÉLISATION PRÉDICTIVE

10010
01100
~~~~~

**Quelle méthode  
statistique  
utiliser pour  
modéliser les  
données  
binaires?**

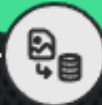




## Pistes d'amélioration

### Enrichissement des données

Ajouter des données météorologiques et de végétation, etc.



### Tester d'autres modèles

Évaluer différents algorithmes d'apprentissage automatique



### Améliorer l'application Streamlit

Développer des onglets d'historique et de prédiction





