

Sommaire

Aller vers

- ☐ Introduction
- ☒ Analyse
- ☐ Datavisualisation
- ☐ Prédiction
- ☐ Conclusion

Projet Feux de Forêt

Partie 1 : Analyse des données

Création des datasets

Le dataset original

	OBJECTID	FOD_ID	FPA_ID	SOURCE_SYSTEM_TYPE	SOURCE_SYSTEM	NWCG_REPORTING_AGENCY	NWCG_REPORTING_UNIT_ID	NWCG_REPORTING_UNIT_NAME	SOURCE_REPORTING_UNIT	SOURCE_REPORTING_UNIT_NAME	LOCAL_FIRE_REPORT
0	1	1	FS-1418826	FED	FS-FIRESTAT	FS	USCAPNF	Plumas National Forest	511	Plumas National Forest	
1	2	2	FS-1418827	FED	FS-FIRESTAT	FS	USCAENF	Eldorado National Forest		503	Eldorado National Forest
2	3	3	FS-1418835	FED	FS-FIRESTAT	FS	USCAENF	Eldorado National Forest		503	Eldorado National Forest
3	4	4	FS-1418845	FED	FS-FIRESTAT	FS	USCAENF	Eldorado National Forest		503	Eldorado National Forest
4	5	5	FS-1418847	FED	FS-FIRESTAT	FS	USCAENF	Eldorado National Forest		503	Eldorado National Forest

Ce [dataset](#) présente les données de 1880465 feux, déclarés sur le territoire américain entre 1992 et 2015.

	Colonnes	Nombre de valeurs manquantes	Pourcentage de valeurs manquantes
0	Unnamed: 0	0	0%
1	OBJECTID	0	0%
2	FOD_ID	0	0%
3	FPA_ID	0	0%
4	SOURCE_SYSTEM_TYPE	0	0%
5	SOURCE_SYSTEM	0	0%
6	NWCG_REPORTING_AGENCY	0	0%
7	NWCG_REPORTING_UNIT_ID	0	0%
8	NWCG_REPORTING_UNIT_NAME	0	0%
9	SOURCE_REPORTING_UNIT	0	0%

Le dataset original contient un grand nombre de valeurs manquantes. Nous le nettoyons en supprimant entre autres les (rares) doublons, afin d'en tirer un dataset propre et plus léger.

Le dataset nettoyé

Le [dataset Kaggle original](#) disposait d'une quarantaine de variables. Nous avons réduit ce chiffre à une dizaine de colonnes, en supprimant les variables redondantes. Pour la suite du processus d'analyse, nous avons choisi de conserver la variable des comtés, malgré son nombre important de valeurs inconnues.

	OBJECTID	FIRE_YEAR	STAT_CAUSE_CODE	STAT_CAUSE_DESCR	FIRE_SIZE	FIRE_SIZE_CLASS	LATITUDE	LONGITUDE	STATE	COUNTY	START_DAY/NIGHT 0/1	CONT_DAY/NIGHT 0/1	START_MONTH	START_DAY_OF_WEEK	SEASON	CAUSE_CATEG
0	1	2,005	9	Miscellaneous	0.1	A	40.0369	-121.0058	CA	63	0	0	2	2	3	3
1	2	2,004	1	Lightning	0.25	A	38.9331	-120.4044	CA	61	0	0	5	2	4	1
2	3	2,004	5	Debris Burning	0.1	A	38.9842	-120.7356	CA	17	1	1	5	0	4	3
3	4	2,004	1	Lightning	0.1	A	38.5592	-119.9133	CA	3	0	0	6	0	1	1
4	5	2,004	1	Lightning	0.1	A	38.5592	-119.9331	CA	3	0	0	6	0	1	1

	Colonnes	Nombre de valeurs manquantes	Pourcentage de valeurs manquantes
0	OBJECTID	0	0%
1	FIRE_YEAR	0	0%
2	STAT_CAUSE_CODE	0	0%
3	STAT_CAUSE_DESCR	0	0%
4	FIRE_SIZE	0	0%
5	FIRE_SIZE_CLASS	0	0%
6	LATITUDE	0	0%
7	LONGITUDE	0	0%
8	STATE	0	0%
9	COUNTY	678148	30%

Le dataset enrichi

Nous avons ajouté à ces données originelles de nouvelles datas, tirées du [National Centers for Environmental Information américain](#). Il s'agit de données météorologiques, contenant l'évolution des températures moyennes et maximales ainsi que celle des précipitations par mois et par comté.

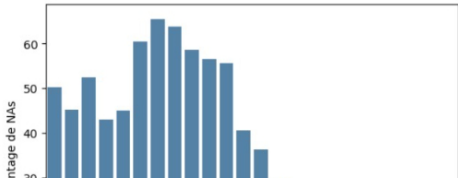
	Année	Etat	Mois	TMAX	TAVG	PCPN	Etat_Comté	Etendue totale des feux	Nombre de feux	Comté
0	1,992	1	1	54.5	44.6	5.62	01-001	0	0	1
1	1,993	1	1	59.3	49.7	6.08	01-001	0	0	1
2	1,994	1	1	52.2	41.7	3.54	01-001	19.6	3	1
3	1,995	1	1	57	46.4	4.95	01-001	2.6	4	1
4	1,996	1	1	56	44	7.04	01-001	0	0	1

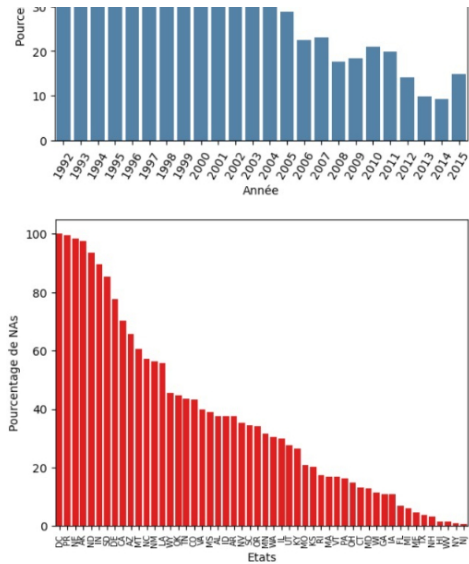
Pour cela, nous n'avons conservé dans ce nouveau DataFrame que les feux dont était précisé le comté, soit 1.2 million environ.

Nous avons également dû supprimer les lignes concernant les feux hawaïens et portoricains, ces deux Etats n'étant pas pris en compte dans les données du NCEI.

Nous gardons en tête que la plupart des lignes supprimées concernaient des feux plus anciens.

Ces graphiques présentent le nombre de feux dont les comtés n'étaient pas renseignés, par année et par état :





## Premières analyses

Nous pouvons d'ores et déjà tirer quelques conclusions de ces données, avant de les modéliser.

*NOTE : Nous utilisons ici les données nettoyées, mais pas enrichies, et travaillons donc sur l'ensemble des feux.*

### Analyse géographique

Les Etats présentant le plus grand nombre de feux sont, dans l'ordre, la Californie, la Géorgie, le Texas, la Caroline du Nord et la Floride.

En revanche, la plupart des mégafeux se déclarent en Alaska : causés souvent par la foudre, ils y sont plus difficiles d'accès pour les pompiers. Ils s'étendent donc sans interruption.

Sélectionnez une période :



Période : (1992, 2015)

☐ Voir les données adaptées aux Mégafeux

STATE	Nombre de feux	Etendue des feux (en km <sup>2</sup> )
CA	189,550	51,602.67
GA	168,867	6,437.08
TX	142,021	39,620.32
NC	111,277	2,957.58
FL	90,261	17,862.54
SC	81,315	2,192.98
NY	80,870	379.64
MS	79,230	4,597.1
AZ	71,586	22,577.66
AL	66,570	3,726.9

### Analyse temporelle

	Année	Nombre de feux	Etendue totale des feux
0	1,992	67,975	8,906.71
1	1,993	61,589	8,873.26
2	1,994	75,955	16,666.7
3	1,995	71,472	8,297.92
4	1,996	75,574	24,312.37
5	1,997	61,450	13,014.72
6	1,998	68,370	8,142.11
7	1,999	89,363	24,616.19
8	2,000	96,416	30,929.14
9	2,001	86,587	15,077.22

Entre 1992 et 2003, 74922 feux de forêt se déclaraient en moyenne annuellement, avec un pic à 96416.

Entre 2004 et 2015, cette moyenne annuelle passe à 81783 feux, avec un pic à 114004.

Cette tendance se répercute sur l'étendue totale brûlée, avec 17044 km<sup>2</sup> incendiés en moyenne chaque année entre 1992 et 2003, contre 30233km<sup>2</sup> entre 2004 et 2015.

La datavizualisation nous permet de creuser ces résultats.