

Mini Projet

Machine Learning

Kilian Longepe

Le but de cette étude est de déterminer si une corrélation existe entre la magnitude d'un séisme sur l'échelle de Richter et des éléments tels que la durée du séisme, la profondeur de celui-ci, la date ou encore la magnitude maximum des différentes ondes du tremblement.

Dans ce but j'ai étudié un dataset sur les séismes en Turquie de 1910 à 2017 (<https://www.kaggle.com/caganseval/earthquake>).

Exploration du dataset

In [56]: earthquake.head()

	id	date	time	lat	long	country	city	area	direction	dist	depth	xm	md	richter	mw	ms	mb
0	2.000000e+13	2003.05.20	12:17:44 AM	39.04	40.38	turkey	bingol	baiklicay	west	0.1	10.0	4.1	4.1	0.0	NaN	0.0	0.0
1	2.010000e+13	2007.08.01	12:03:08 AM	40.79	30.09	turkey	kocaeli	bayraktar_izmit	west	0.1	5.2	4.0	3.8	4.0	NaN	0.0	0.0
2	1.980000e+13	1978.05.07	12:41:37 AM	38.58	27.61	turkey	manisa	hamzabeyli	south_west	0.1	0.0	3.7	0.0	0.0	NaN	0.0	3.7
3	2.000000e+13	1997.03.22	12:31:45 AM	39.47	36.44	turkey	sivas	kahvepinar_sarkisla	south_west	0.1	10.0	3.5	3.5	0.0	NaN	0.0	0.0
4	2.000000e+13	2000.04.02	12:57:38 AM	40.80	30.24	turkey	sakarya	meseli_serdivan	south_west	0.1	7.0	4.3	4.3	0.0	NaN	0.0	0.0

In [57]: earthquake.tail()

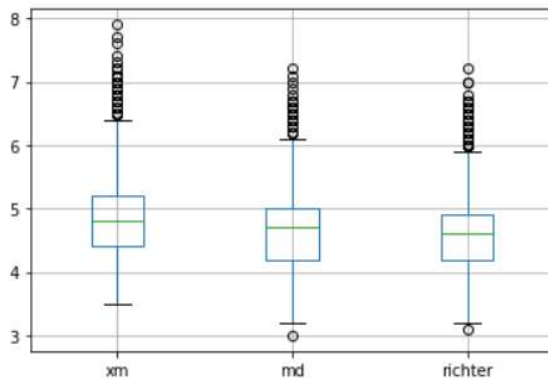
	id	date	time	lat	long	country	city	area	direction	dist	depth	xm	md	richter	mw	ms	mb
24002	2.020000e+13	2017.03.10	12:23:42 AM	37.50	38.48	turkey	adiyaman	kirmacik_samsat	NaN	NaN	13.0	4.2	0.0	4.2	4.0	0.0	0.0
24003	2.020000e+13	2017.03.10	12:42:44 AM	40.47	35.84	turkey	amasya	kizoglu	NaN	NaN	5.6	3.6	0.0	3.5	3.6	0.0	0.0
24004	2.020000e+13	2017.04.05	12:08:11 AM	38.63	27.57	turkey	manisa	cavusoglu	NaN	NaN	10.6	4.1	0.0	4.0	4.1	0.0	0.0
24005	2.020000e+13	2017.04.05	12:35:09 AM	38.65	27.55	turkey	manisa	cavusoglu	NaN	NaN	8.7	3.8	0.0	3.6	3.8	0.0	0.0
24006	2.020000e+13	2017.04.05	12:25:25 AM	38.63	27.58	turkey	manisa	cavusoglu	NaN	NaN	9.0	3.5	0.0	3.4	3.5	0.0	0.0

Dans un 1er temps j'ai fait une brève analyse du dataset afin de sélectionner les colonnes à garder soit ici la date, la profondeur(depth), la magnitude maximum des différents types ondes(xm), la durée du séisme(md) et la magnitude (richter).

	date	depth	xm	md	richter
1	2007.08.01	5.2	4.0	3.8	4.0
20	1966.08.19	14.0	5.0	4.7	4.7
27	1966.08.19	50.0	5.0	4.7	4.7
28	1966.12.30	31.0	4.5	4.3	4.3
29	1967.05.22	54.0	4.9	4.7	4.6

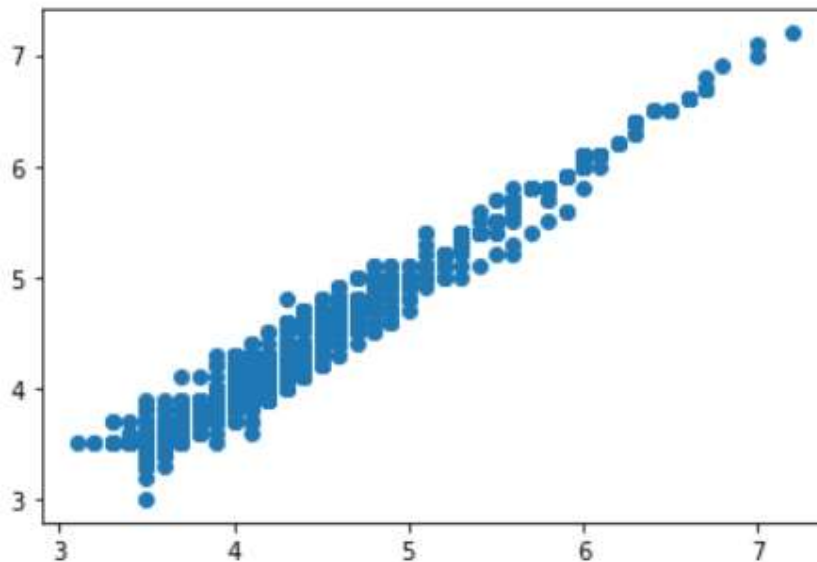
Ensuite j'ai clean toutes les données inutilisable soit toutes les lignes ou l'une de ces 5 informations manquaient, avec ça j'en ai déduis les valeurs min/max, la moyenne et la médiane de chacune des colonnes.

	depth	xm	md	richter
count	3557.000000	3557.000000	3557.000000	3557.000000
mean	31.003430	4.794968	4.606466	4.595193
std	27.231561	0.632224	0.571780	0.554410
min	1.000000	3.500000	3.000000	3.100000
25%	10.000000	4.400000	4.200000	4.200000
50%	26.000000	4.800000	4.700000	4.600000
75%	40.000000	5.200000	5.000000	4.900000
max	225.000000	7.900000	7.200000	7.200000



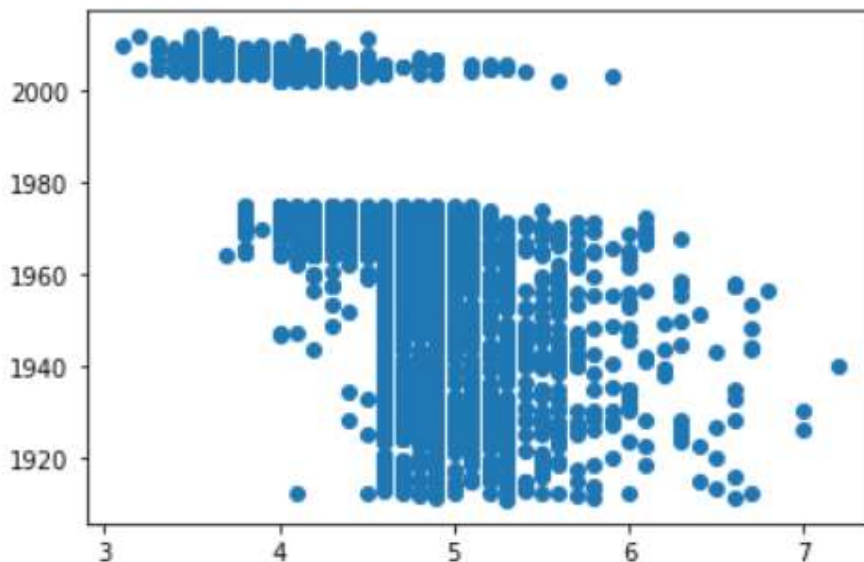
Avec les boîtes à moustaches j'ai pu déterminer si certaine valeur sortait de la plage des valeur possible, ce qui ici n'a pas été le cas.

Pour entrer dans le vif du sujet une fois que toutes les données sont valide j'ai cherché si une corrélation pouvait exister entre les différentes colonnes:

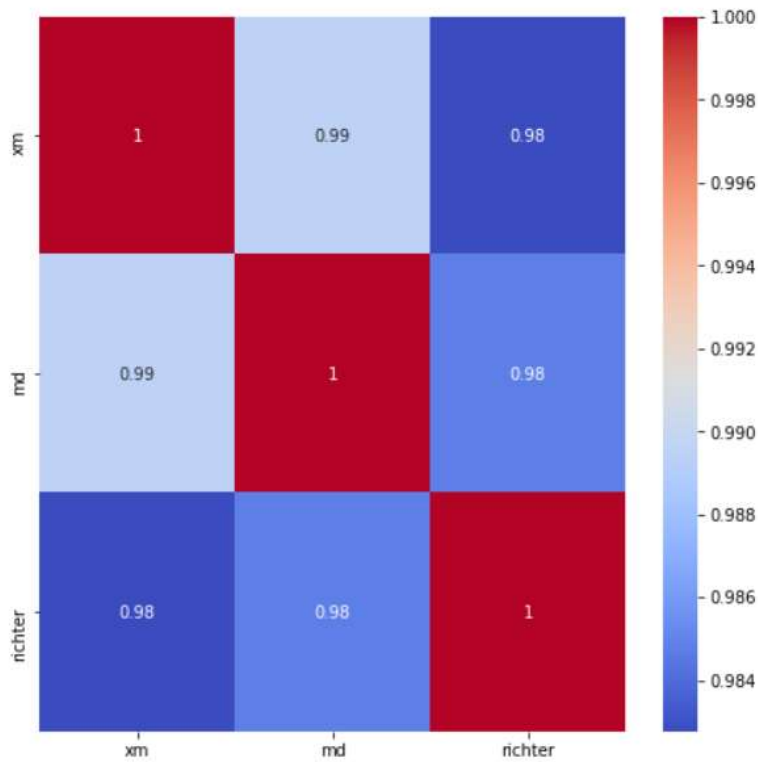


Une superbe corrélation entre xm, md et Richter c'est dévoiler, j'ai donc pu en déduire que ces 3 valeurs étaient intrinsèquement liées.

Puis une corrélation plus légère entre la date et la magnitude (Richter) est aussi apparue alors qu'aucune corrélation n'a été trouvée avec la profondeur des séismes que je n'ai donc pas gardé.



Ici bien que des données autour de l'an 2000 manquent, on peut clairement voir que les séismes sont moins violents d'année en année.



Grâce à une heatmap et au coefficient de régression linéaire (ici 97.26%) j'ai donc pu conclure que la durée des séismes ainsi que la puissance max des différents types d'ondes impactent directement la puissance sur l'échelle de Richter des tremblements de terre en Turquie.