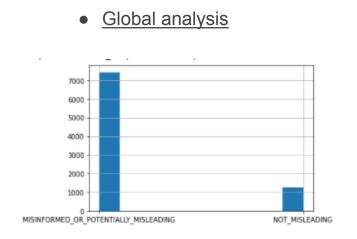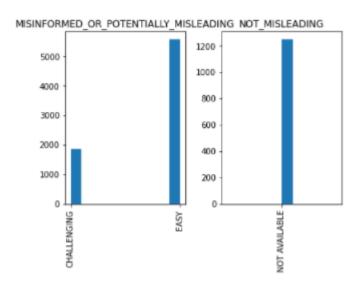## III- Tweet analysis

In order to have a better understanding of the data, I did some analysis on it :
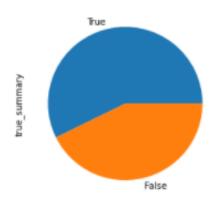
- ● <u>Global analysis</u>

Distribution of the number of misleading/not misleading tweets

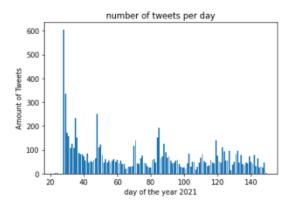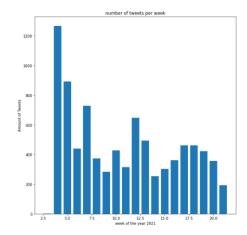Distribution on the difficulty to classify a tweet

True claims :

We consider a confirmation tweet true if there is a link to a source in the summary field. I added a column "true_summary" to the dataframe.

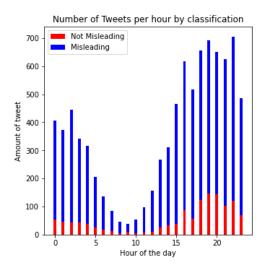*Distribution of summary considered as true or false*

● <u>Time analysis</u>

Number of tweets per day from 01/01/2021 to 29/05/2021

number of tweets noted per week since the 1st January 2021

Specific Tweets analysis

In total there are 6886 tweets, 958 tweets that have been labelled by more than one participant and 52 tweets labelled by more than 5 participants
Here are the tweets the most noted :

Tweet	Number of votes	majority classified	percentages of users who labeled the tweet ([click to see the pie chart](#))
I am happy to work with Republicans on this issue where there's common ground, but you almost had me murdered 3 weeks ago so you can sit this one out. Happy to work w/ almost any other GOP that aren't trying to get me killed.In the meantime if you want to help, you can resign. https://t.co/4mVREbaqqm	61	MISINFORMED_OR_POTENTIALLY_MISLEADING with 57 votes	UnverifiedClaimAsFact : 37% FactualError : 30.3% MissingImportantContext : 24%
This is the latest manipulative take on the right.They are manipulating the fact that most people don't know the layout the Capitol complex.We were all on the Capitol complex - the attack wasn't just on the dome. The bombs Trump supporters planted surrounded our offices too. https://t.co/jI18e0XRrd	22	MISINFORMED_OR_POTENTIALLY_MISLEADING with 22 votes	UnverifiedClaimAsFact : 35.5% FactualError : 33.3% MissingImportantContext : 30.8%
On a day in which #AlexandriaOcasioSmollett is trending, please never forget the time that @AOC staged a photo shoot dressed in all white at a parking lot to spread lies about immigrant children in cages. Faking her own attempted murder was the next logical step. https://t.co/y8mNNowMGy	15	MISINFORMED_OR_POTENTIALLY_MISLEADING with 14 votes	FactualError : 34.4% UnverifiedClaimAsFact : 28.1 % MissingImportantContext : 21.9 %
I don't think this even matters at this point. Time magazine just came out said that a cabal of elites rigged the election I'm sorry they said they didn't rig the election they "fortified" it, by changing the rules and laws as well as manipulating the flow of information https://t.co/1RO1tjbifv	12	NOT_MISLEADING with 11 votes	notMisleadingFactuallyCorrect : 71% notMisleadingOutdatedButNotWhenWritten : 7.1% notMisleadingOther : 7.1%

● Participant analysis

There are 1703 participants.

participantId	number of notes	Summary considered as		Classification of Tweets
		True	False	Not misleading (orange) / Misinformed or potentially misleading (blue)
8B903B14E301B4BFF4E2E41F05146CDD	697	697	0	100.0% / 0.0%
785F608A49A333D59696479B96A0FE49	226	138	86	93.4% / 6.6%
7FC8E5C1761E00B8B7C27C7249834934	209	208	1	98.1% / 1.9%

12

06C7B658C5B6735445616F768A6A8528	164	160	4	

- Ratings analysis

IV- XLM classifier

The goal of the task is to automatically recognise the statistical sentences (ie statistical fact about COVID-19) and check if the claim is true or not. I used a classifier (BERT and then XLM) in order to retrieve the statistical sentences from all the Tweets available.
The code implemented is available here.

The results are not really conclusive as most of the sentences are actually not statistical.

1. Global analysis
   - number of statistical sentences collected : 5552 (out of 5585 sentences given in input)
   - For comparison, the first classifier gave 1828 outputs.
   - All the sentences have claim=True , which means the classifier estimates the sentences is True
2. Ambiguous fields analysis

I looked at the values taken by the column 'ambiguous fields'. It represents the fields where the classifier is not sure what data to use.
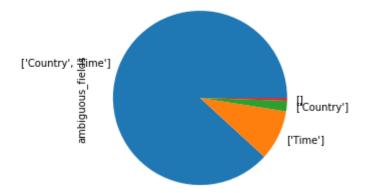
```
['Country', 'Time']    4896
['Time']                519
['Country']             108
[]                       29
Name: ambiguous_fields, dtype: int64
```


We can notice that there are some sentences without ambiguous fields, let's have a look at some of these sentences.

It's Wednesday, March 31, 2021, and Donald J Trump is the REAL &
LEGITIMATE President of the United States of America!
#Treason #TrumpWon https://t.co/umlCGZvn8K

U.S. Military has everything it needs, will not wait for completion
of audits. Trump won. They know it, will expose it and all involved.
No one will live in The White House again. Trump will be back. May,
June, July. D.C. Military operations underway now. Trigger.
Gen. Flynn https://t.co/UdpByoUQf8

President Donald J. Trump won the US Presidential election on
11/3/2020.
Never forget.

1/ The data are clearer and clearer: the early phase of mass #Covid
vaccination campaigns leads to a surge in cases and deaths.
The two European countries that have vaccinated most aggressively
since early February are Hungary and Serbia. Guess what's happened?
https://t.co/UYGYl1RgAU

JUST IN: President Biden says the United States will have enough
doses for every adult American by the end of May.

```
The timeline speeds up by two months the administration's previous
goal of having enough shots by the end of July.
https://t.co/oY22y5CphE
```

⇒ We can notice that indeed, in each sentence there is a country and a date, so the classifier has precise data for these fields. However , most of the sentences are not related to COVID.

3. Sentences related to COVID

In this section, I tried to look at the sentence dealing with covid. To do that, I took the sentences in which the words "Covid" or "Corona" were present. There are 414 of them. Let's display some of them

```
CDC does not currently recommend the use of facemasks to help prevent
novel #coronavirus. Take everyday preventive actions, like staying
home when you are sick and washing hands with soap and water, to help
slow the spread of respiratory illness. #COVID19
https://t.co/uArGZTJhXj https://t.co/yzWTSgt2IV


Why was GAM started in 2021? Because museums need to survive.
Visitors have declined rapidly due to COVID. Museums need $ for
upkeep of paintings and to pay staff and curators. They have no other
sources of revenue except governments, wealthy donors, and art
aficionados. #NFT


My 7-part series on who created COVID-19 can be found here.
You don't want to miss it. https://t.co/5EejBQwVuV

The Biden administration has moved Gitmo detainees to the front of
the line for COVID-19 vaccines. We've gone from America First to
Terrorists First in the most literal of ways.

If every willing person in America is vaccinated for #COVID19 by May,
as POTUS has said, why put our lives on hold till July the 4th?
```

```
Today Texas set a new one-day record by giving 270,523 COVID vaccine
shots to Texans. More than 55% of seniors have received a shot and
more than 28% of Texans between 50-64 have received a shot. COVID
hospitalizations continue to decline &amp; are at the lowest level in
5 months.
────────────────────────────────────────────────────────────────
100% of people who get the COVID vaccine die.

100% of people who do not get the COVID vaccine die.

100% of people who get COVID die.

100% of people who do not get COVID die.

Let this inform your understanding of public health statistics.

────────────────────────────────────────────────────────────────
@RudyGiuliani I urge you to look at TOTAL deaths year by year for
past ten years. 2020 is lower than 8 out of 10 previous years. Where
are all the "COVID" deaths??? It is one big HOAX...
```

4. Sentences with covid and no ambiguous fields

There is only one sentence that correspond to this :

	sentences	sols	choices	ambiguous_ fields	claim	confidence	probas	covi d
1274	1/ The data are clearer and clearer: the early phase of mass #Covid vaccination campaigns leads to a surge in cases and deaths.\n\nThe two European countries that have vaccinated most aggressively since early February are Hungary and Serbia. Guess what's happened? https://t.co/UYGYl1RgAU	[]	{'formula': array(['No nStat'], dtype='<U 7'), 'table': array(['No nStat'], dtype='<U 7'), 'countries' : ['Hungary- Serbia'], 'month': ['February']}	[]	1.0	True	[0.9976131, 0.9976131]	True

Even this sentence is not very statistical as there are no numbers given.