Is there a correlation between twitter usage and the occurrence of world events.

James Zoryk,
Nikolina Micek,
Sanne Peter

Research Question

Is there a correlation between social media posts and events

Our hypothesis: significant increase in the density of social media posts when notable news or world events occur

Analyzing the density of tweets created in time to see if we can observe a significant trend in the data

Reasearch question

Social media platform we choose is Twitter:

- all tweets are in public domain
- twitter serves as an online discussion platform
- each tweet contains potential meta data that can be used

Specific languages can lead to more localized events

Getting the data

A non-profit organization called Archive Team runs a web-scrapper that collects data from all tweets created on twitter.

Whole month of raw tweet data was downloaded.

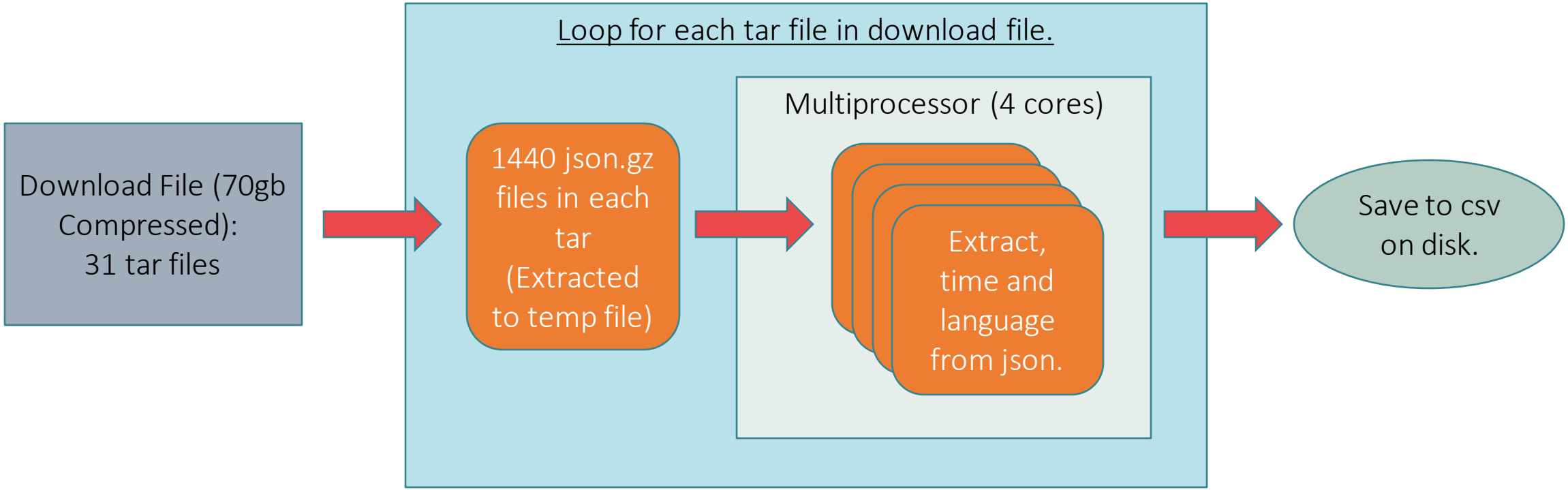
However, this is still rather large in file size, since it contains a lot of metadata, such as 'message', 'filter', etc.

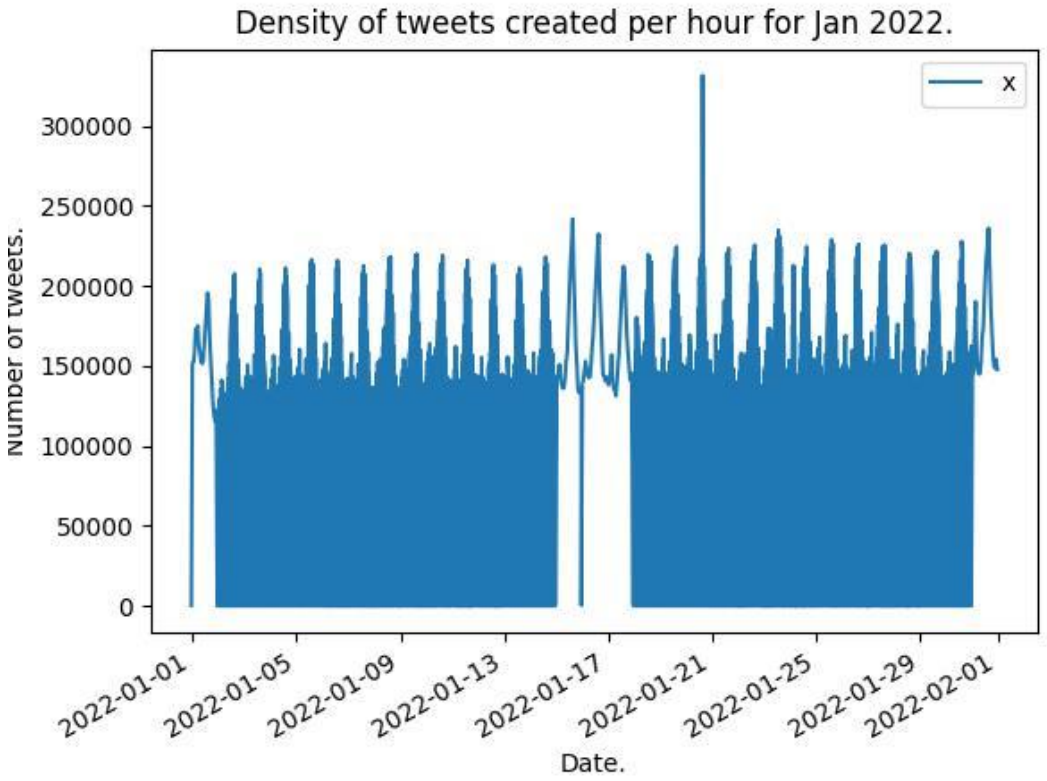
We only need language and time created.

Process this raw data into smaller usable data.

Getting the data: Processing the raw data

Wrote a python function that processes the data as follows:

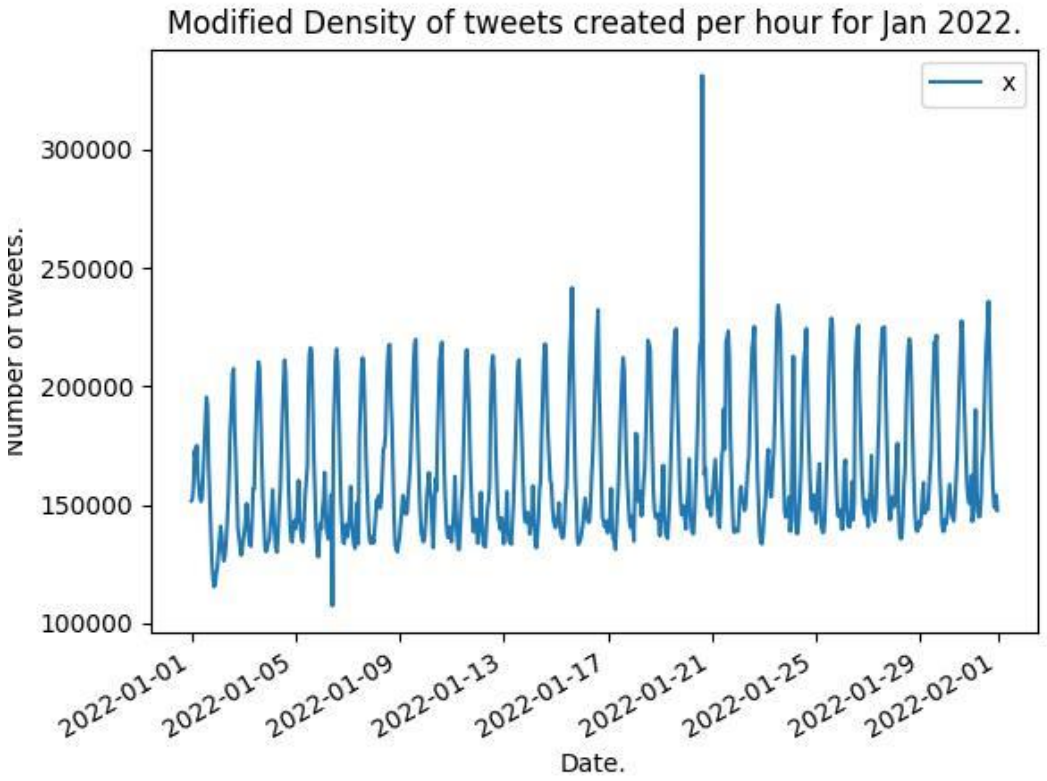




Visualising the data.

Plotting the raw data in 1 Hour intervals, (to see the data)

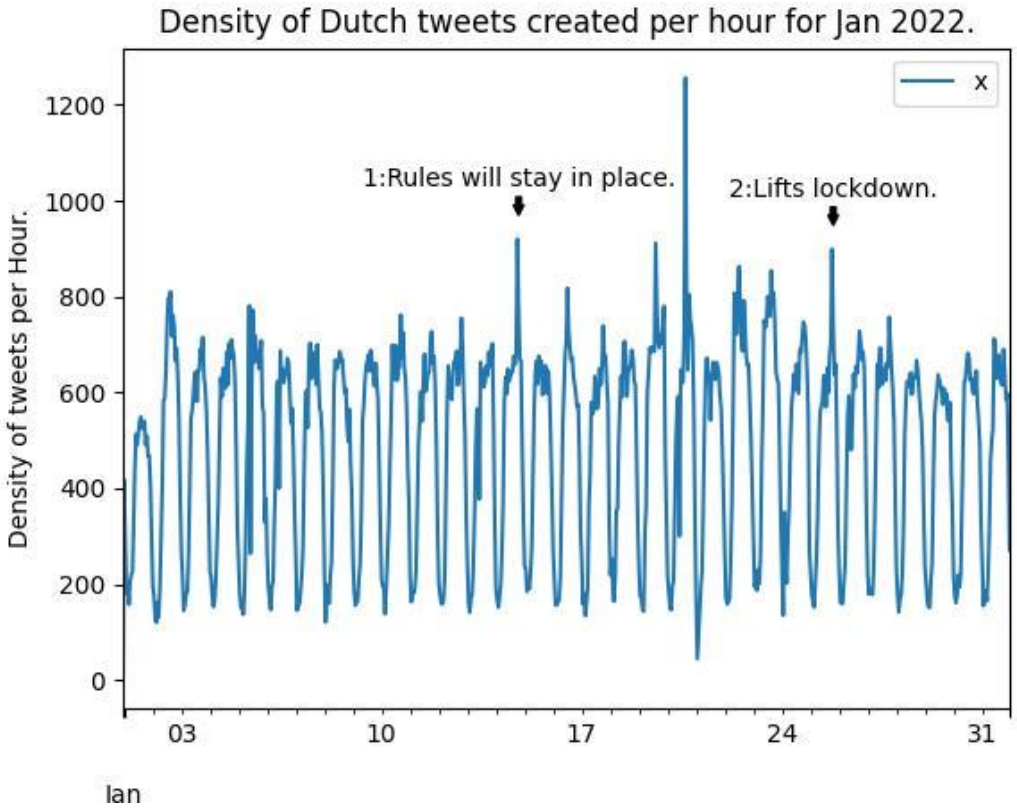
Observing the outliers.



Visualising the data.

Plotting the raw data in 1 Hour intervals, with removing outliers.

Outliers when we have no data in the time intervals.

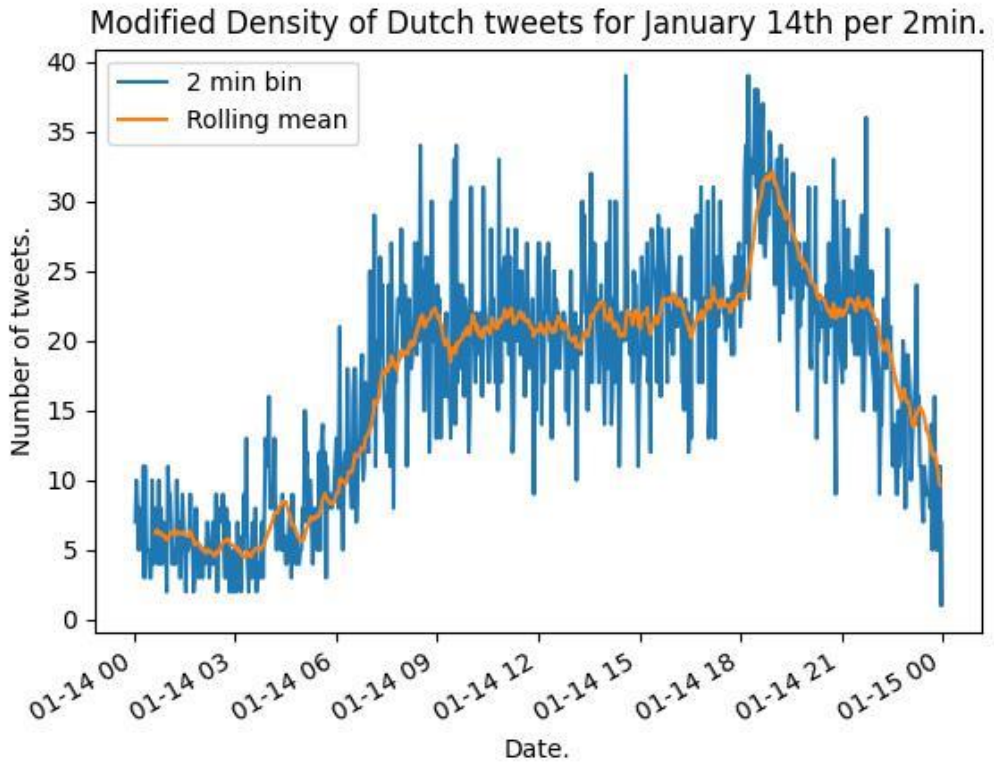


Visualising the data.

Plotting Dutch tweets for January 1 hour intervals.

Adding markers for both press conferences.

Observe increase in both days, not formal yet.

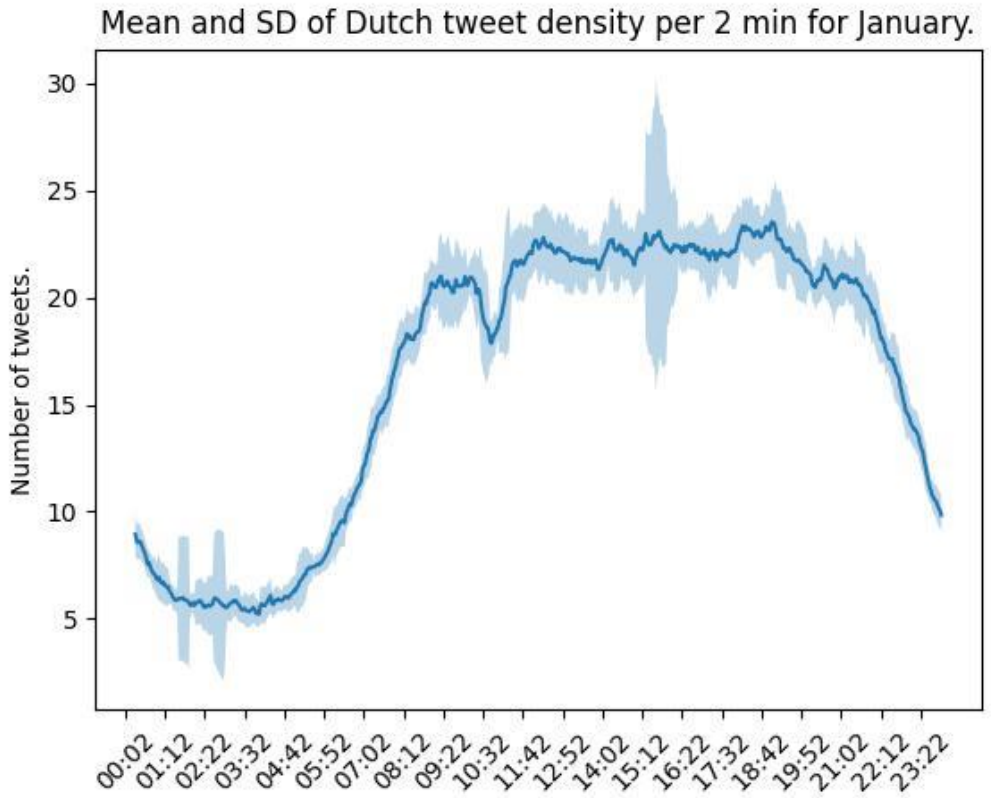


Visualising the data.

Increasing the resolution of the time intervals: 2min

Looking at period of 24 Hours.

Using the rolling mean with window size 20 (40mins) makes it more readable.

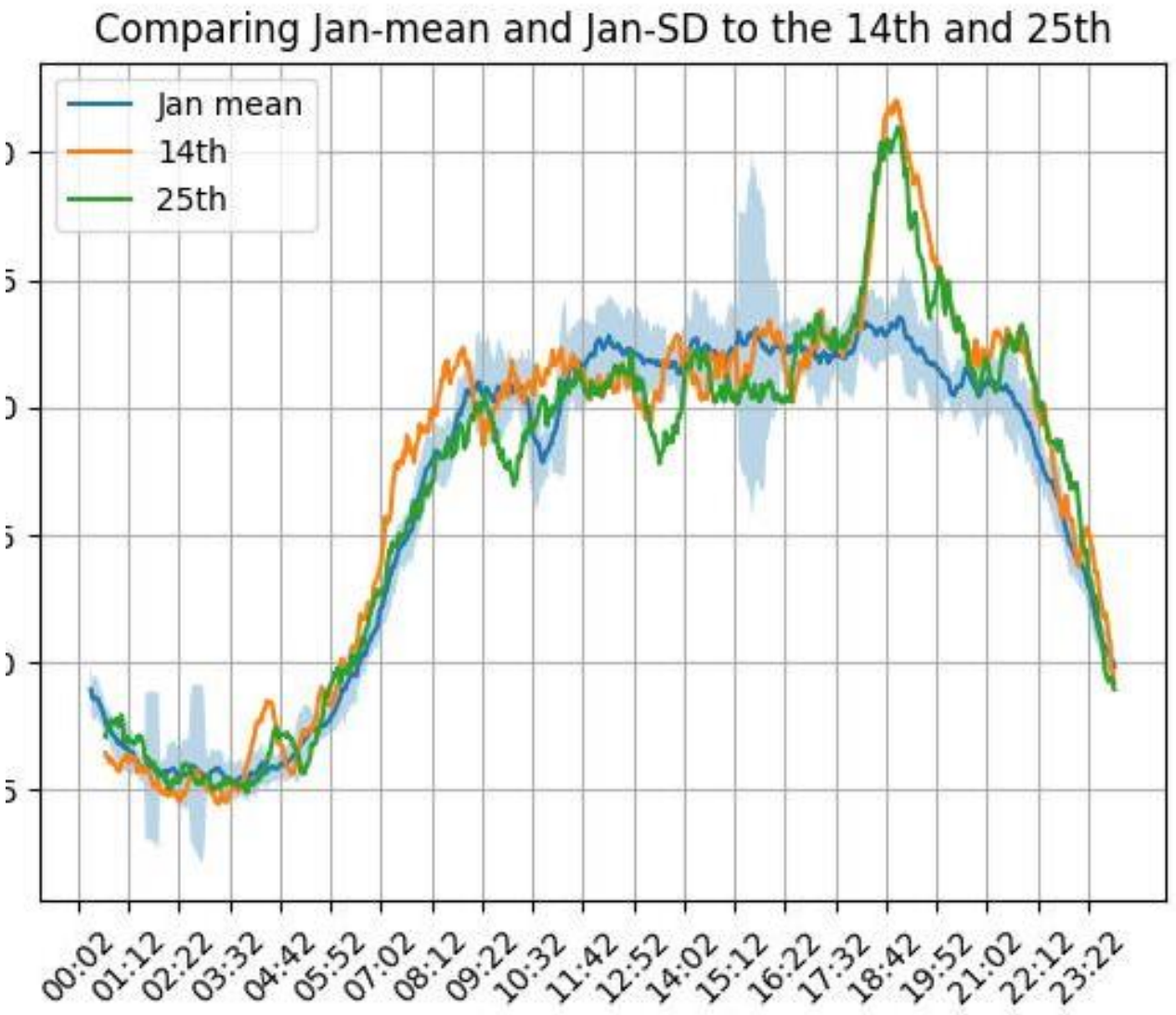


Visualising the data.

Increasing the resolution of the time intervals: 2min.

Taking the mean and Standard Deviation for each day in January.

Plot of the rolling mean and a fill interval of 2*SD.



Conclusion and findings

We plot the rolling mean of both the 14th and 25th and compare to the month rolling mean along with 2*SD.

Indeed, we see peaks larger than 2*SD for the time of both press conferences, 19:00 CET.

While the rest follows the general trend.

Conclusion

Hence in our limited observation, we do see that there is a correlation between the density of tweets created per time interval and world events.

The scope of this report was limited to software, hardware and time to process the raw data.

Further investigations:

- The peak around the 21st January 2022 – Unsure meatloaf
- Using more of the metadata from each tweets, such as searching the message for keywords and hashtags.