# Humour, Tragedy, and Timing: An Empirical Study of Meme Uptake Across Public Events

Candidate Number: 1092913

*Wordcount: 3,907*

**Abstract**

This paper investigates whether memes about tragic events emerge later than those about non-tragic events. Drawing from Reddit data across ten tragic and ten non-tragic viral events, the study analyses over 90,000 filtered comments and posts to explore timing and patterns in meme engagement. Using time-series plots, heatmaps, boxplots, and violin plots, and statistical testing, the analysis finds evidence supporting the hypothesis that tragic events show delayed comedic uptake. However, exceptions and distribution variability highlight the nuance of humour in digital public discourse.

## 1 Introduction

The creation of comedy out of tragedy is a social phenomenon that can be traced throughout human history. From medieval plague ballads such as *Danse Macabre*[1], to the "gallows humour"[2] of persecuted Christians, soldiers, paramedics, and mourners, social groups have long turned to humour in the face of suffering.

Philosophers, psychologists, and literary theorists have all sought to explain this paradoxical but widely observed phenomenon. For Nietzsche, laughter is a release born of suffering (Nietzsche, 1967); for Camus, it is an act of defiance in the face of meaninglessness – a rebellion that insists, even in the bleakest conditions, that "one must imagine Sisyphus happy" (Camus, 1955). In literature, Aristotle's concept of catharsis in tragedy suggests a dialectic relationship between tragedy and comedy, where tragedy is "resolved" with the "release" of laughter (Janko, 1984), and psychologists such as Freud and Vaillant classify humour as a "mature" coping mechanism, allowing us to process what might otherwise overwhelm us (Vaillant, 1977).

In the digital age, "memes" have become a significant medium for this social process of transformation. Disturbing or tragic events are frequently coopted by the "memescape" and filtered back into reality through the lens of humour (Yalcinkaya, 2019). Many studies have examined the use of memes as a response to tragedy within online environments, seeking to answer the question of why people make memes about traumatic events. In an analysis of memes following the Manchester Arena bombing, Merrill and Lindgren (2021) suggest that memes provided

[1] Translating as the 'Dance of Death,' *Danse Macabre* is an artistic genre, frequently depicting skeletons dancing to their graves. The genre emerged out of the Black Death in 14th- and 15th-century art and has been analysed as a comedic response to deadly epidemics (DesOrmeaux, 2007).

[2] Defined as "a humorous response that appears to be inappropriate or illogical in the face of hopeless situations" by Napp (2023), citing Wormer and Davis (1997).

"post-terror togetherness," building social solidarity; Akram et al. (2021) argue that memes circulating during the COVID-19 pandemic functioned as collective coping strategies; Tokariuk (2023) shows how Ukrainian civilians used memes to boost morale, subvert Russian propaganda, and build a sense of collective identity online; and Kuipers (2005), studying dark humour meme responses to mass shootings and 9/11, suggest that memes serve to process collective trauma.

While there are many reasons why people might create memes after tragic events, including to soothe collective suffering, to bind social groups, or to "play" with transgressive or taboo ideas in creative environments (Poplar, 2022), it is reasonable to expect that this transition from tragedy to humour does not happen immediately. It seems plausible that a pause after a tragic event will take place, where jokes are "too soon," too distasteful, or the event too shocking. This expectation forms the central hypothesis driving my research: that a pause exists between tragedy and humour. As such, this study asks:

> **RQ:** Is there a difference in the timing of meme uptake between tragic, emotionally charged events and non-tragic events?

It is important to note that this paper does not examine whether memes about tragic events are more viral than those about other types of events. Rather, it investigates whether there is a difference in the *timing* and *pattern* of meme uptake between tragic and non-tragic events. To answer this question, this paper proceeds as follows. I will firstly outline my experimental design, data, and key limitations of this investigation before presenting and analysing my results.

## 2 Experimental Setup

### i) Event Selection Criteria

My first step was to select a sample of both tragic and non-tragic events for analysis. All events were filtered using the following criteria: (i) they had to be widely publicised in U.S. media to ensure high visibility and engagement from Reddit's primarily U.S.-based community; (ii) they needed a clear start date to allow for unambiguous tracking of meme uptake over time; and (iii) they had to have generated viral memes, ensuring there was sufficient content to analyse patterns of spread. Finally, their subjective categorisation as tragic and non-tragic had to leave little to no room for disagreement. As such, I selected from wars, high-profile deaths, terrorist attacks, and natural disasters for tragic events. In contrast, non-tragic events were light-hearted or absurd in tone. These included viral political memes, celebrity scandals, and other culturally significant humorous incidents. To mitigate bias, the events were selected at the beginning according to the above criteria and did not change over the course of the experiment. Ten events were selected to ensure a large data source. Brief descriptions of the more obscure events are given in Appendix A.

Table 1: **Tragic and Non-Tragic Events Used in the Study**

| **Tragic Events** | **Date** |
|---|---|
| October 7th Attack | October 7, 2023 |
| *Oceangate** | June 18, 2023 |
| Ukraine War | February 24, 2022 |
| Queen Elizabeth II Death | September 8, 2022 |
| Las Vegas Shooting | October 1, 2017 |
| George Floyd's Death | May 25, 2020 |
| Kobe Bryant's Death | January 26, 2020 |
| Hurricane Dorian | August 24, 2019 |
| Notre Dame Fire | April 15, 2019 |
| *Harambe's Death** | May 28, 2016 |
| **Non-Tragic Events** | |
| *Trump Mugshot** | August 24, 2023 |
| Elon Buys Twitter | October 28, 2022 |
| *Chris Rock Slap** | March 28, 2022 |
| *Suez Canal Boat** | March 23, 2021 |
| *Biden Fall** | March 20, 2021 |
| Harry & Meghan Oprah Interview | March 7, 2021 |
| *Bernie 'Mittens'** | January 21, 2021 |
| *Will Smith Genie** | February 11, 2019 |
| *Elon on Joe Rogan** | September 6, 2018 |
| Leonardo DiCaprio Wins Oscar | February 26, 2016 |

### ii) Subreddit Selection

To collect a robust and representative sample of meme activity, I selected the five largest meme-focused subreddits: `r/memes` (35.4m members), `r/dankmemes` (5.94m), `r/meme` (2.7m), `r/Memes_Of_The_Dank` (1.0m), and `r/PoliticalMemes` (43k). All subreddits were created before the earliest event above. Larger, but thematically narrow subreddits, such as `r/HistoryMemes` (13m members), whose content focusses on "events over 20 years ago," were excluded. Reddit was chosen as the platform of focus because it is both highly active and readily scrapable, offering structured metadata (e.g., timestamps, posts, comments) that allows for precise temporal analysis. While further research could examine a broader range of niche or culturally specific subreddits to compare subcultural differences in humour uptake, this paper focuses on mainstream meme environments, with the aim of capturing large-scale patterns of when humour emerges in response to public tragedies.

### iii) Reddit Scraping

The first challenge I faced was in accessing the large amount of subreddit data. Initial attempts to extract data using the Reddit API, via the Python wrapper `praw`, proved too limited. Although the API allowed for real-time querying of subreddit submissions and comments, strict rate limits, and incomplete data coverage significantly constrained the utility of this method.

To overcome these limitations, I used compressed Reddit data archives retrieved through Pushshift and hosted on Academic Torrents (Watchful1, 2025). Using qBittorrent, I downloaded the relevant `.zst` files for the five subreddits containing posts and comments. Each file included either post or comment data, with full metadata and timestamps extending through the end of 2024.

To scrape relevant data, I wrote two Python scripts:

- `events.py` defines a structured dictionary of events, each with a specific start date and associated keyword groups for filtering relevant content. These include both single keywords (e.g., `‘harambe’`) and multi-word groups (e.g., `[‘kobe’, ‘crash’]`) associated with each event. (See Appendix B)
- `extract.py` processes the `.zst` files line by line, matching posts and comments to events based on the keyword dictionary within a time window spanning two days before to fourteen days after each event date. (See Appendix C)

Once filtered, each match is tagged with a day offset (e.g., Day 0 = event day, Day 1 = the day after) and classified as either a post or a comment. These daily counts are compiled into a structured pandas data frame and exported as a `.csv` file. This file served as the basis for all subsequent analysis of meme uptake patterns. Given that engagement with the events quickly drops off across the Reddit meme pages, only the following 14 days after the event are included in the data frame. Filtering by short time windows further massively reduced the run time of `extract.py`, allowing for alterations to the dictionary in `events.py`. Including the two days prior to each event allowed me to iteratively refine my dictionary filters to reduce false positives adding noise to the data extracted.

I applied a quality quota that fewer than 5% of the total posts and comments filtered could occur in the two days prior to the event start date written into `events.py`, ensuring that extracted content reflected direct responses to the event itself. The table below (Table 2) shows the percentage of filtered posts and comments that occurred in the two days prior to each event. The 5% threshold was more difficult to meet for certain events. For example, memes about the Oceangate submersible began circulating before the official implosion was confirmed; speculation about Queen Elizabeth II’s death began after her hospitalisation (but before her death); and Leonardo DiCaprio’s potential Oscar win generated meme posts and many comments prior to the award ceremony. In such cases, the filtering process became a balancing act between precision and over-exclusion.

These edge cases illustrate the temporal fuzziness of meme emergence in fast-moving or highly anticipated scenarios. They also complicate efforts to draw a strict line between pre- and post-event humour—a distinction easier to draw for shock events such as Kobe Bryant’s death—especially when meme culture blends satire, speculation, and reaction in real time. Despite this limitation, a 5% false positive rate quota [3] provides a reasonable degree of confidence that the filters are capturing meaningful post-event meme activity.

[3] This pre-event match rate serves as a proxy for the false positive rate—that is, the proportion of posts or comments incorrectly captured by the keyword filter prior to the actual occurrence of the event.

The table below shows the proportion of posts and comments made in the two days prior to the event that were picked up by the filter. To gather these proportions, I used `error_rate.py` (see codebook). False positive rates above 4%, while still minimal, are shown in dark orange, above 3% in pale orange, and below 2% in green for clear visualisation.

Table 2: **False Positive Rate**

| **Category** | **Event** | **Posts** | **Comments** |
|---|---|---|---|
| Tragic | October 7th Attack | 0.000 | 0.000 |
| Tragic | Oceangate | 0.042 | 0.021 |
| Tragic | Ukraine War | 0.019 | 0.022 |
| Tragic | Queen Elizabeth II Death | 0.043 | 0.037 |
| Tragic | Las Vegas Shooting | 0.000 | 0.000 |
| Tragic | George Floyd's Death | 0.000 | 0.001 |
| Tragic | Kobe Bryant's Death | 0.000 | 0.000 |
| Tragic | Hurricane Dorian | 0.000 | 0.003 |
| Tragic | Notre Dame Fire | 0.000 | 0.000 |
| Tragic | Harambe Death | 0.000 | 0.001 |
| Non-Tragic | Trump Mugshot | 0.013 | 0.048 |
| Non-Tragic | Elon Buys Twitter | 0.047 | 0.021 |
| Non-Tragic | Chris Rock Slap | 0.002 | 0.047 |
| Non-Tragic | Suez Canal Boat | 0.000 | 0.032 |
| Non-Tragic | Biden Fall | 0.028 | 0.007 |
| Non-Tragic | Harry & Meghan Oprah Interview | 0.027 | 0.003 |
| Non-Tragic | Bernie 'Mittens' | 0.013 | 0.008 |
| Non-Tragic | Will Smith Genie | 0.000 | 0.000 |
| Non-Tragic | Elon Weed Podcast | 0.000 | 0.041 |
| Non-Tragic | Leonardo DiCaprio Wins Oscar | 0.048 | 0.039 |

## 3 Results and Discussion

When visualising and interpreting my results, the first consideration was how to effectively compare meme uptake patterns. Figure 1 shows the raw counts of posts made on each day for all the events, calculated in `raw_ts.py`. This figure highlights the high variance in virality between events. Without normalisation, events with higher engagement dominate the results, making it difficult to compare the timing and pattern of meme uptake across events.

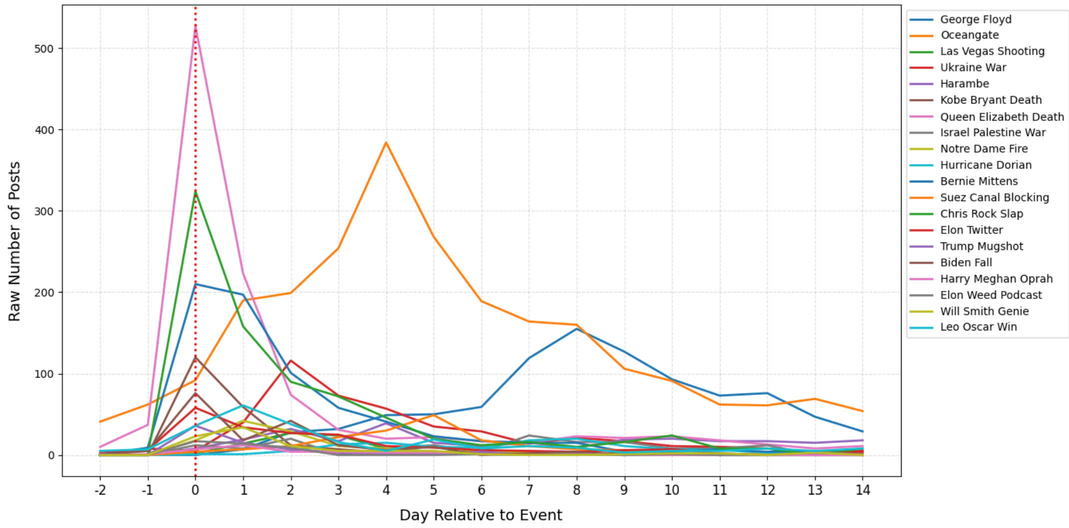


Figure 1: **Time-series of raw post counts by event**

We therefore needed to normalise the data to compare trends of uptake rather than virality. This process is shown in normalised_ts.py where I divide the daily values of posts and comments by the maximum value observed for each event. I then created an average normalised value by day relative to the start dates to compare tragic and non-tragic events, both by posts and comments.

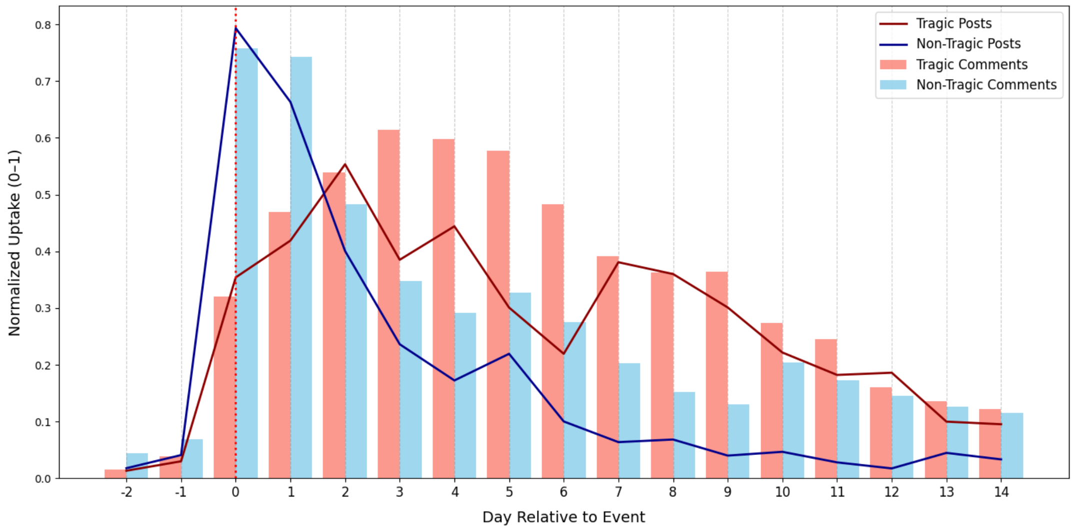


Figure 2: **Time-series of normalised post and comment counts**

Immediately, a difference in uptake distribution is visible. While, by mean average, the non-tragic posts and comments peak on the day of the event and then gradually decline over the 14-day period, tragic events do not peak in posts until 3 days after the event, and 4 days after for comments. This time-series serves to support our underlying hypothesis that a time-delay will be observable for meme subreddit engagement after tragic events compared to non-tragic events.

Interestingly, although beyond the scope of this paper, we also see a difference in the relationship between comments and posts between the two categories; while the comment distribution closely follows post distribution for non-tragic events, comment distribution remains high 3-6 days after the tragic events as post counts start to drop off. As a side note, we further observed that the ratio of comments to posts was far higher for tragic events (11.89:1) than non-tragic events (4.62:1). Further studies might consider not only the differences in uptake distribution patterns, but differences in the types of engagement between event types.

The next phase of the paper examines the distribution patterns of individual events to assess whether these broader trends held consistently across each specific case. In doing so, we also considered the potential for Simpson's Paradox, where aggregated trends might obscure or reverse the relationships observed within subgroups (Weinberger, 2021).

To enable direct comparisons between posts and comments, we needed to develop a single metric for uptake. Given that our dictionary filtered a total of 82,066 comments and 8,475 posts (roughly 10:1), we assigned a weight of 0.1 to the comments. This weighting was chosen to reflect the fact that comments, while still important for engagement, generally constitute a secondary form of interaction compared to posts. For all subsequent visualisations created, we use the following formula:

$$\text{Engagement} = \text{Posts} + 0.1 \times \text{Comments}$$

The first use of this formula was in heatmap.py, shown in the codebook, and visualised below:

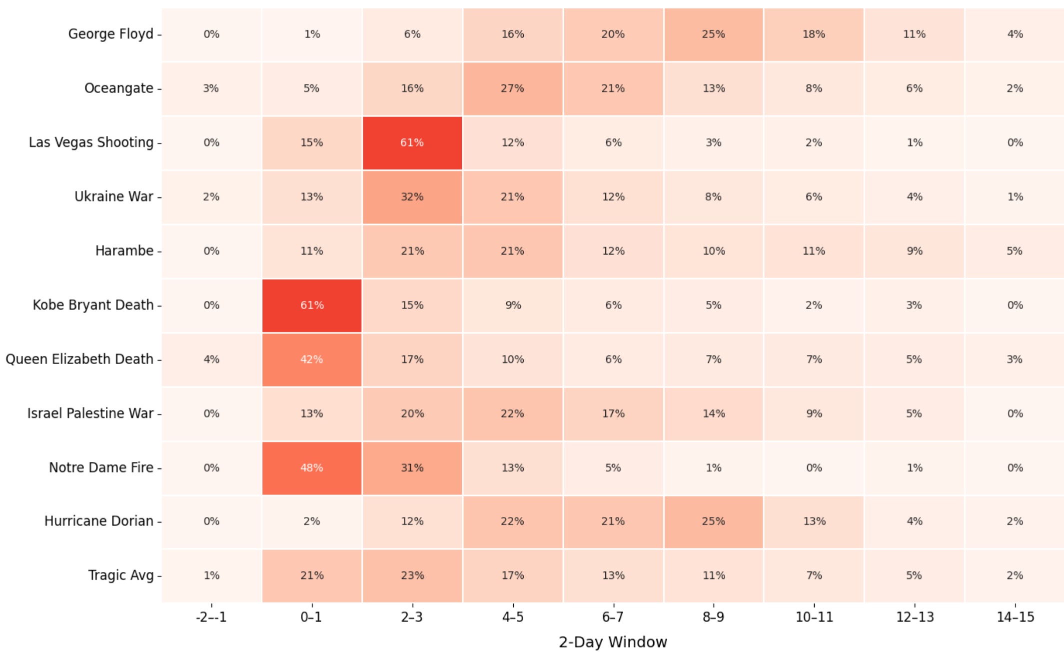


Figure 3: **Heatmap of meme engagement for tragic events**

Focusing on the mean average of normalized engagement distributions between tragic and non-tragic events, the heatmaps corroborate the patterns observed in the time-series. Figure 3 shows that 21% of engagement for tragic events occurs on days 0 and 1, compared to 45% for non-tragic events in Figure 4, suggesting a delay in meme uptake for tragic events. Additionally, the engagement distribution for non-tragic events has a longer tail, with more sustained engagement levels from Days 2 to 13, indicating prolonged periods of engagement after initial, generally shallower peaks.

The heatmaps, however, reveals many exceptions to these pattern differences. Most notably, within the tragic events, Kobe Bryant's death, Queen Elizabeth II's death, and the Notre Dame Fire all show a higher distribution of engagement on days 0 and 1 than the non-tragic events' average. Further, the high concentration of meme engagement on days 2-3 after the Las Vegas shooting, while a delay in engagement compared to the non-tragic event average, deviates from the general distribution of tragic events which generally exhibit higher dispersion with a slight early skew and tapering decline over time .

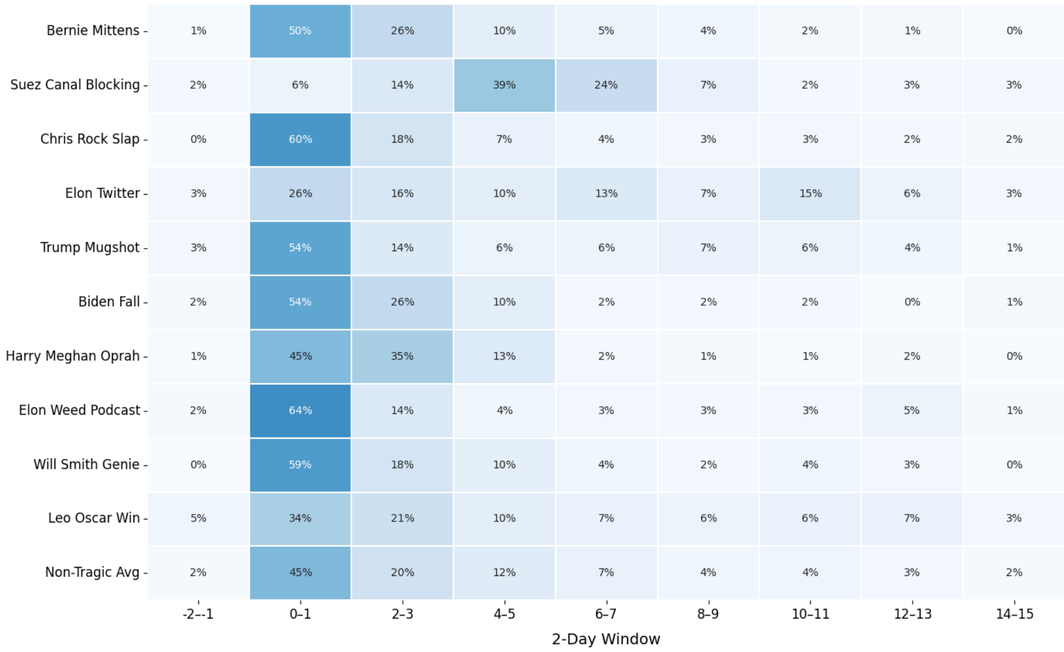


Figure 4: **Heatmap of meme engagement for non-tragic events**

To better visualise the distribution of meme engagement across individual events, we created a horizontal boxplot (see box_plots.py in codebook). Each box represents the interquartile range (IQR), spanning the middle 50% of sampled engagement days, with the median marked by a vertical line and the mean overlaid as a black diamond. Whiskers extend to the most extreme values within $1.5 \times$ IQR of the box. This convention, while not equivalent to a 95% confidence interval, is a robust way to visualise spread and identify patterns without distortion from extreme values. Engagement was calculated as posts + $0.1 \times$ comments, and 1,000 engagement-weighted days were sampled per event using probabilities derived from this score. We used pandas for data handling, numpy for sampling, seaborn for statistical plotting, and matplotlib for display. Ordering events by their simple mean engagement day enabled clearer visual comparisons across tragic and non-tragic categories.

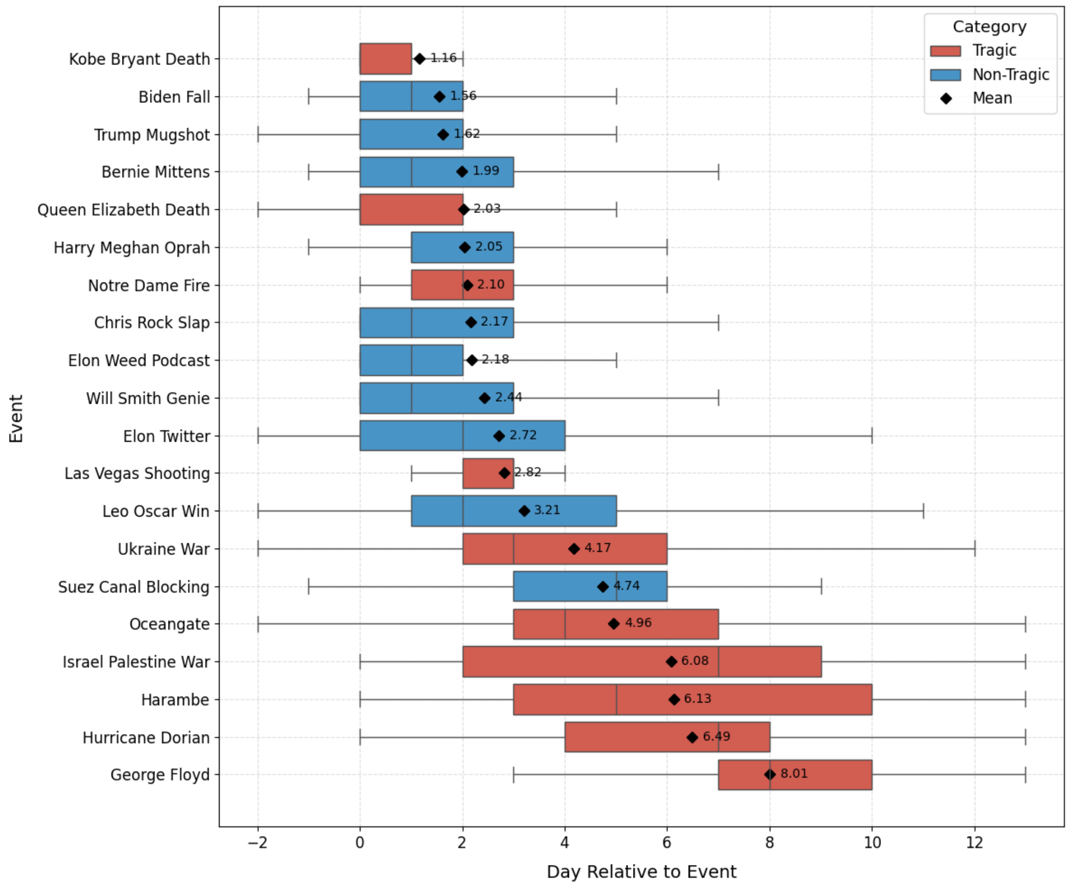


Figure 5: **Box plots with average uptake days across events**

From the horizontal box plots, a general pattern difference is visible: tragic events tend to have a later mean engagement day than non-tragic events, with seven out of the ten highest means belonging to tragic events. This supports the hypothesis that emotionally charged events experience a relative delay in meme uptake. However, notable exceptions complicate this trend. Despite being a tragic event, Kobe Bryant's death shows the earliest mean engagement day overall (1.11), with a sharply left-skewed and tightly clustered distribution around days 0–1. Conversely, the viral, non-tragic Suez Canal blockage (4.84) displays a relatively late mean engagement day and a right-skewed distribution, indicating sustained meme activity well after the event's onset.

These exceptions highlight that while trauma may often delay collective comedic engagement, the nature of an event—its suddenness, media framing, or inherent memeability—can override this effect. This suggests that emotional valence alone does not fully determine meme uptake dynamics. A larger and more diverse sample of events may help account for these anomalies. For example, although widely publicised in U.S. media, the Notre Dame fire and the death of Queen Elizabeth II may have felt more distant to Reddit's predominantly U.S.-based userbase, potentially softening their emotional impact and accelerating meme engagement. Moreover, a key limitation of this study is that it measures engagement volume, not its content or tone. Early reactions to Kobe Bryant's death, for instance, may have been expressions of grief rather than humour. Future research could incorporate sentiment analysis of post content or visual analysis of image-based memes to better distinguish between satirical, supportive, or critical modes of engagement.

To further investigate overall differences in meme uptake timing between event types, we created a violin plot (violin.py in codebook). While the boxplot visualised the distribution of individual engagement days per event, the violin plot focuses instead on the distribution of mean uptake days *across* events—treating each event's mean as a single data point. This approach allowed us to directly compare the average engagement timing between tragic and non-tragic categories. Using pandas and numpy, we calculated the engagement for each day as posts + 0.1 × comments, then computed a weighted mean uptake day per event by multiplying each day by its relative engagement and normalising by total engagement. The resulting values were visualised using seaborn and matplotlib, with a violin plot showing kernel density estimates. Vertical dashed lines mark the average uptake day within each category.

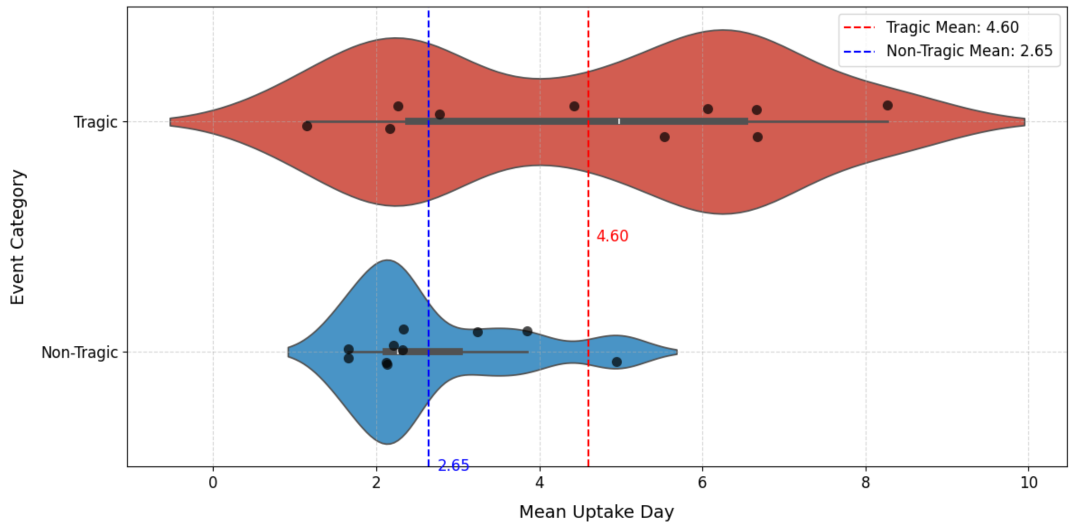


Figure 6: **Violin plot of mean uptake days within event types**

The violin plot reveals a notable difference in distribution shape and dispersion between the two categories. Non-tragic events exhibit a relatively tight distribution, with most events clustering around a mean uptake day of approximately 2.82. While not strictly unimodal, the distribution shows a clear concentration near the 1.5–2.5 day mark, with only a few events—such as *Leo's Oscar win* (3.34), *Elon Musk buying Twitter* (3.74), and the *Suez Canal blockage* (4.84)—pulling the distribution slightly rightward. This indicates low dispersion and a broadly consistent pattern of rapid meme uptake. By contrast, tragic events have a significantly higher mean of 4.60 and display far greater variability, with values ranging from *Kobe Bryant's death* (1.11) to *George Floyd's death* (8.27). Notably, four tragic events—*Kobe Bryant*, *Queen Elizabeth II* (2.19), *Notre Dame* (2.28), and the *Las Vegas shooting* (2.83)—fall below the non-tragic mean and are similarly clustered. While the mean differences are striking, the tragic distribution is far more dispersed, bimodal, and positively skewed, reinforcing that meme responses to tragedy follow no single temporal pattern.

To test whether meme uptake patterns differed significantly between tragic and non-tragic events, we applied both Welch's t-test and the non-parametric Mann–Whitney U test. The Welch's t-test indicated a statistically significant difference in mean uptake days between the two groups ($T = 2.36$, $p = 0.036$), supporting the hypothesis that tragic events experience later meme engagement. However, the Mann–Whitney U test, which compares ranked distributions without assuming normality (a more suitable test given the distributions) returned a non-significant result ($U = 75.0$, $p = 0.064$). This suggests that while differences in central tendency exist, they may not reflect a broader, consistent shift across the full distribution. Taken together, the tests provide cautious support for the hypothesis, while also highlighting the complexity and variability of meme uptake around tragic events.

# 4 Conclusion

This study set out to explore whether tragic events exhibit delayed meme engagement compared to non-tragic events. Through a combination of time-series analysis, heatmaps, boxplots, violin plots, and statistical testing, the results offer cautious support for the central hypothesis. On average, memes about tragic events emerge later than those about non-tragic events. However, while non-tragic events tend to cluster around a fast, consistent meme uptake window, tragic events have wider variability and greater dispersion across cases.

Several limitations, however, must be acknowledged. Due to the size of the .zst files and the runtime of the extract.py, the number of events that could be feasibly analysed was constrained. While the selected cases span a mixture of tragic and non-tragic moments, a larger sample would allow for finer-grained comparisons — for example, between celebrity deaths, natural disasters, wars, and terrorist attacks within the tragic category, or between political scandals, internet culture, and light-hearted memes in the non-tragic category. This would help clarify which kinds of events are more "memeable", how cultural distance or proximity shapes humour, and whether certain types of tragedy elicit more cautious or delayed humour uptake Second, the study did not control for the overall volume of Reddit activity during each event window, meaning that meme frequency was not adjusted relative to baseline posting behaviour on the subreddits. During moments of global attention or when multiple events occurred simultaneously, meme activity may appear dampened or inflated due to competition effects or surges in unrelated posting. Third, while the engagement metric offers a useful proxy for attention, it does not capture the tone of posts — a distinction crucial when assessing humour. For instance, early engagement with tragic events may reflect grief or outrage rather than satire.

Future research could address these gaps in several ways. A larger and more diverse event pool could enable typological analysis within each category, revealing which forms of tragedy elicit slower or faster comedic responses. Incorporating sentiment analysis or natural language processing techniques would allow researchers to distinguish between humorous and non-humorous content, offering a more refined view of meme intent. Further, image-based meme classification could help identify patterns specific to visual humour formats, which often spread differently than text-based content (Merrill & Lindgren, 2021)Finally, examining meme uptake across other platforms — such as TikTok, Instagram, or X — would enable a more robust analysis of popular meme culture.

To conclude, this study contributes to a growing body of research on digital humour and collective trauma by offering a computational framework for analysing the timing of meme emergence. The findings indicate that tragic events are, on average, associated with delayed meme uptake — yet notable exceptions highlight the potential cultural, emotional, and contextual factors that otherwise shape virality. In navigating the delicate space between tragedy and comedy, meme culture reveals both its adaptability and its moral ambiguity — turning grief into absurdity, solemnity into satire, and silence into punchlines, sometimes before we're ready. Ultimately, these dynamics reflect the role of meme culture subreddits as not just as a site of entertainment, but as a platform for negotiating collective pain, where humour becomes both a shield and a mirror in times of crisis.

*Wordcount: 3,907*

## Bibliography

Akram, U., Irvine, K., Allen, S. F., Stevenson, J. C., Ellis, J. G., & Drabble, J. (2021). Internet memes related to the covid-19 pandemic as a potential coping mechanism for anxiety. *Scientific Reports*, *11*, Article 22305. doi: 10.1038/s41598-021-00857-8

Camus, A. (1955). *The myth of sisyphus.* Vintage Books. (J. O'Brien, Trans. (Original work published 1942))

DesOrmeaux, A. L. (2007). *The black death and its effect on fourteenth- and fifteenth-century art* (Unpublished master's thesis). Louisiana State University and Agricultural and Mechanical College. (LSU Digital Commons. `https://repository.lsu.edu/gradschool_theses/1641`)

Janko, R. (1984). *Aristotle on comedy: Towards a reconstruction of poetics ii.* University of California Press.

Kuipers, G. (2005). Where was king kong when we needed him? public discourse, digital disaster jokes, and the functions of laughter after 9/11. *The Journal of American Culture*, *28*(1), 70–84. doi: 10.1111/j.1542-734X.2005.00155.x

Merrill, S., & Lindgren, S. (2021). Memes, brands and the politics of post-terror togetherness: Following the manchester bee after the 2017 manchester arena bombing. *Information, Communication Society*, *24*(16), 2403–2421. doi: 10.1080/1369118X.2021.1942957

Napp, T. (2023). *The medicine of memes: An exploration of humor styles, memes, and dark humor* (Unpublished master's thesis). Abilene Christian University. (`https://digitalcommons.acu.edu/etd/619`)

Nietzsche, F. (1967). *The birth of tragedy.* Vintage Books. (W. Kaufmann, Trans. (Original work published 1872))

Poplar, D. (2022). Framing the ethical boundaries of humor. *The Philosophy of Humor Yearbook*, *3*(1), 153–178. doi: 10.1515/phhumyb-2022-005

Tokariuk, O. (2023). *From memes to morale: Decoding ukraine's comedy arsenal against disinformation.* Reuters Institute for the Study of Journalism. Retrieved from https://reutersinstitute.politics.ox.ac.uk/memes-morale-decoding-ukraines-comedy-arsenal-against-disinformation

Vaillant, G. E. (1977). *Adaptation to life.* Little, Brown.

Watchful1. (2025). *Subreddit comments/submissions 2005–06 to 2024–12 [dataset].* https://www.reddit.com/r/pushshift/comments/1itme1k/separate_dump_files_for_the_top_40k_subreddits/.

Weinberger, N. (2021). *Simpson's paradox.* In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Summer 2021 Edition). Retrieved from https://plato.stanford.edu/archives/sum2021/entries/paradox-simpson/

Wormer, K. S. V., & Davis, D. R. (1997). *Addiction treatment: A strengths perspective.* Brooks/Cole.

Yalcinkaya, G. (2019). *This decade, memes became a coping mechanism for our chaotic times.* Dazed Digital. Retrieved from https://www.dazeddigital.com/life-culture/article/47271/1/this-decade-memes-became-a-coping-mechanism-for-our-chaotic-times

## Appendix A: Event Descriptions

**Oceangate (June 18, 2023).** The Titan submersible, operated by OceanGate Expeditions, imploded during a descent to the Titanic wreckage site in the North Atlantic, killing all five passengers. The incident gained viral attention.

**Harambe's Death (May 28, 2016).** A 17-year-old gorilla named Harambe was fatally shot at the Cincinnati Zoo after a child entered his enclosure. The zoo's decision sparked controversy and global debate on animal ethics, later evolving into an enduring meme.

**Trump Mugshot (August 24, 2023).** Former U.S. President Donald Trump's booking photo was released following his indictment in Georgia for attempting to overturn the 2020 election. The mugshot rapidly became a viral cultural artifact.

**Chris Rock Slap (March 28, 2022).** During the Academy Awards, actor Will Smith slapped comedian Chris Rock on stage after a joke referencing Jada Pinkett Smith. The unscripted moment generated international media coverage and instantly spawned thousands of memes.

**Bernie Mittens (January 21, 2021).** A photo of Senator Bernie Sanders sitting alone in a chair, wearing handmade mittens at Joe Biden's inauguration, went viral. The image was humorously inserted into a wide range of settings, symbolizing practicality, grumpiness, and political disaffection.

**Will Smith Genie (February 11, 2019).** The release of Disney's *Aladdin* live-action trailer featured Will Smith as a blue CGI genie, prompting mixed reactions and online ridicule.

**Elon Weed Podcast (September 6, 2018).** Elon Musk smoked marijuana during a live interview on *The Joe Rogan Experience* podcast, triggering stock dips for Tesla and public reactions ranging from concern to amusement. The moment became a meme.

## Appendix B: events.py

```python
from datetime import datetime, timezone

events = {
#     # === TRAGIC EVENTS ===
    "george_floyd": {
        "start_date": datetime(2020, 5, 25, tzinfo=timezone.utc),
        "keywords": [['black', 'lives', 'matter'], ['george', 'floyd'], ['police',
'brutality'], ['protests', 'george']]
    },
    "oceangate": {
        "start_date": datetime(2023, 6, 18, tzinfo=timezone.utc),
        "keywords": [['titan', 'dead'], ['titan', 'death'], ['oceangate',
'implosion'], ['oceangate', 'meme'], ['oceangate', 'dead'], ['titan', 'imploded'],
['sub', 'crushed'], ['sub', 'implosion'], 'implosion']
    },
    "las_vegas_shooting": {
        "start_date": datetime(2017, 10, 1, tzinfo=timezone.utc),
        "keywords": [['vegas', 'shooting'], ['vegas', 'dead'], ['mass', 'shooting']]
    },
    "ukraine_war": {
        "start_date": datetime(2022, 2, 24, tzinfo=timezone.utc),
        "keywords": [['war', 'ukraine'], ['russian', 'invasion'], ["ukraine",
"putin"], ["russia", "invasion"], ['snake island'], "East Ukraine", ['ghost of',
'kyiv'], ]
    },
    "harambe": {
        "start_date": datetime(2016, 5, 28, tzinfo=timezone.utc),
        "keywords": ['harambe', 'ripharambe', ['justice', 'harambe'], ['harambe',
'shooting'], ['harambe', 'killed'], ['harambe', 'memes'], ['rip', 'harambe']]
    },
    "kobe_bryant_death": {
        "start_date": datetime(2020, 1, 26, tzinfo=timezone.utc),
        "keywords": [['kobe', 'crash'], ['helicopter', 'crash'], ['kobe', 'crash',],
['kobe', 'bryant']]
    },
    "queen_elizabeth_death": {
        "start_date": datetime(2022, 9, 8, tzinfo=timezone.utc),
        "keywords": [['queen', 'elizabeth'], 'rip queen', ['queen', 'died'], ['queen',
'dead']]
    },
    "israel_palestine_war": {
        "start_date": datetime(2023, 10, 7, tzinfo=timezone.utc),
        "keywords": [['hamas', 'attack'], ['hamas', 'paraglider'], ['israel', 'gaza'],
['October', '7'], ['free', 'gaza'], ['gaza', 'children'], ['free', 'palestine'],
['ghost,' 'gaza'], ['idf', 'tiktok'], ['netanyahu', 'meme'], ['hamas', 'war'],
['rocket', 'attack'], ['attack', 'israel'], ]
    },
    "notre_dame_fire": {
        "start_date": datetime(2019, 4, 15, tzinfo=timezone.utc),
        "keywords": [['notre dame', 'fire'], ['cathedral', 'fire']]
    },
    "hurricane_dorian": {
    "start_date": datetime(2019, 8, 24, tzinfo=timezone.utc),  # Formed: Aug 24,
Bahamas hit: Sept 1
    "keywords": [
        ['hurricane', 'dorian'], ['sharpie', 'trump'], ['dorian', 'florida'],
['dorian', 'bahamas'],['dorian', 'storm'], ['dorian', 'hovering'], ['trump',
'dorian'], ['category', 'dorian'], ['dorian', 'map'], ['florida', 'hurricane'],
['dorian', 'florida', 'man']
    ]
},
```

```
# #     # # === NON-TRAGIC EVENTS ===
    "bernie_mittens": {
        "start_date": datetime(2021, 1, 21, tzinfo=timezone.utc),
        "keywords": [['bernie', 'mittens'], ['bernie', 'sitting'], ["bernie", 'inauguration'],
['bernie', 'meme']]
    },
    "suez_canal_blocking": {
        "start_date": datetime(2021, 3, 23, tzinfo=timezone.utc),
        "keywords": [['evergiven'], ['ever', 'given'], ['suez', 'blocked'], ['suez', 'ever'], ['suez',
'memes'], ['suez', 'sideways'], ['container', 'ship', 'suez']]
    },
    "chris_rock_slap": {
        "start_date": datetime(2022, 3, 28, tzinfo=timezone.utc),
        "keywords": [['chris rock', 'slap'], ['will smith', 'smacked'], ['will smith', 'smacks'],
['keep', 'wife'], ['rock', 'slap'], ['smith', 'slap']]
    },
    "elon_twitter": {
        "start_date": datetime(2022, 10, 28, tzinfo=timezone.utc),
        "keywords": [['elon', 'twitter', 'buy'], ['elon', 'twitter', 'takeover'], ['musk', 'buys',
'twitter'], ['twitter', 'ownership'], ['twitter', 'x'], ['elon', 'twitter'], ['elon', 'buys']]
    },
    "trump_mugshot": {
        "start_date": datetime(2023, 8, 24, tzinfo=timezone.utc),
        "keywords": [['trump', 'mugshot'], ['trump', 'arrest'], ['trump', 'charged'], ['trump',
'jail'], ['trump', 'mugshot'], ['trump', 'photo'], ['georgia', 'trump']]
    },
    "biden_fall": {
        "start_date": datetime(2021, 3, 20, tzinfo=timezone.utc),
        "keywords": [['biden', 'fall'], ['biden', 'fell'], ['biden','tripped'], ['biden', 'stairs'],
['biden', 'air force'], 'AF1']
    },
    "harry_meghan_oprah": {
        "start_date": datetime(2021, 3, 7, tzinfo=timezone.utc),
        "keywords": [['meghan', 'oprah'], ['harry', 'oprah'], ['meghan', 'interview'], ['royal',
'interview'],['oprah', 'interview']]
    },
    "elon_weed_podcast": {
        "start_date": datetime(2018, 9, 6, tzinfo=timezone.utc),
        "keywords": [['elon', 'rogan'], ['elon', 'podcast'], ['elon', 'joint'], ['elon', 'smoking'],
['elon', 'blunt'], ['elon', 'joe'], ['elon', 'high'], ['joe', 'rogan']]
    },
    "will_smith_genie": {
    "start_date": datetime(2019, 2, 11, tzinfo=timezone.utc),
    "keywords": [['will', 'smith', 'genie'], ['blue', 'genie'],  ['will', 'smith', 'blue'], ['genie',
'nightmare']]
    },
    "leo_oscar_win": {
    "start_date": datetime(2016, 2, 26, tzinfo=timezone.utc),
    "keywords": ['dicaprio',['leonardo', 'dicaprio', 'wins', 'oscar'], ['oscar', 'winner', 'leo'],
['leonardo', 'dicaprio', 'winning'], ['leo', 'won', 'oscar'], ['leonardo', 'dicpario', 'won'], ['leo',
'oscar', 'memes'], ['oscar', 'goes', 'leo'],]
    },

}
```

## Appendix C: extract.py

```python
import zstandard
import json
from datetime import datetime, timezone, timedelta
import io
import pandas as pd
from collections import defaultdict
from events import events
import time
start_time = time.time()


# === SUBREDDITS ===
subreddits = {
    "dankmemes": {
        "submissions": "dankmemes_submissions.zst",
        "comments": "dankmemes_comments.zst"
    },
     "meme": {
         "submissions": "meme_submissions.zst",
        "comments": "meme_comments.zst"
    },
     "memes": {
         "submissions": "memes_submissions.zst",
        "comments": "memes_comments.zst"
    },
    "Memes_Of_The_Dank": {
        "submissions": "Memes_Of_The_Dank_submissions.zst",
        "comments": "Memes_Of_The_Dank_comments.zst"
    },
     "PoliticalMemes": {
         "submissions": "PoliticalMemes_submissions.zst",
        "comments": "PoliticalMemes_comments.zst"
     }
}

# === UTILS ===
def read_lines_zst(filepath):
    with open(filepath, 'rb') as fh:
        dctx = zstandard.ZstdDecompressor(max_window_size=2**31)
        reader = dctx.stream_reader(fh)
        text_stream = io.TextIOWrapper(reader, encoding='utf-8')
        for line in text_stream:
            yield line.strip()

def text_matches(text, keyword_groups):
    text = text.lower()
    return any(all(word in text for word in group) for group in keyword_groups)

# === PRE-INDEX EVENTS BY DATE ===
date_to_events = defaultdict(list)

for event_name, data in events.items():
    start_date = data["start_date"].date()
    for offset in range(-2, 15):  # -2 to 14 days after event
        active_date = start_date + timedelta(days=offset)
        date_to_events[active_date].append((event_name, offset, data["keywords"]))

# === AGGREGATION ===
daily_counts = defaultdict(int)
```

```python
# === MAIN LOOP ===
for subreddit, paths in subreddits.items():
     for content_type in ["submissions", "comments"]:
        file_path = paths[content_type]
        print(f"🔍 Reading {file_path}...")

        for line in read_lines_zst(file_path):
            try:
                obj = json.loads(line)
                created = datetime.fromtimestamp(int(obj['created_utc']), timezone.utc).date()
                text = (obj.get('title', '') + ' ' + obj.get('selftext', '') + ' ' + obj.get('body', '')).lower()

                if created in date_to_events:
                    for event_name, offset, keyword_groups in date_to_events[created]:
                        if text_matches(text, keyword_groups):
                            daily_counts[(event_name, offset, content_type)] += 1
            except Exception:
                continue

# === BUILD OUTPUT TABLE ===
rows = []
for day in range(-2, 15):
    row = {"Day": day}
    for event_name in events:
        row[f"{event_name}_posts"] = daily_counts.get((event_name, day, "submissions"), 0)
        row[f"{event_name}_comments"] = daily_counts.get((event_name, day, "comments"), 0)
    rows.append(row)

df = pd.DataFrame(rows)
df = df.sort_values(by="Day")
df.to_csv("big_run_4.csv", index=False)
print(df.head())
print("\n✅ Done! Output saved to: big_run_4.csv")

elapsed = time.time() - start_time
print(f"\n⏱️ Total runtime: {elapsed:.2f} seconds")
```