The COVID-19 Effect:

From Panic to Burnout

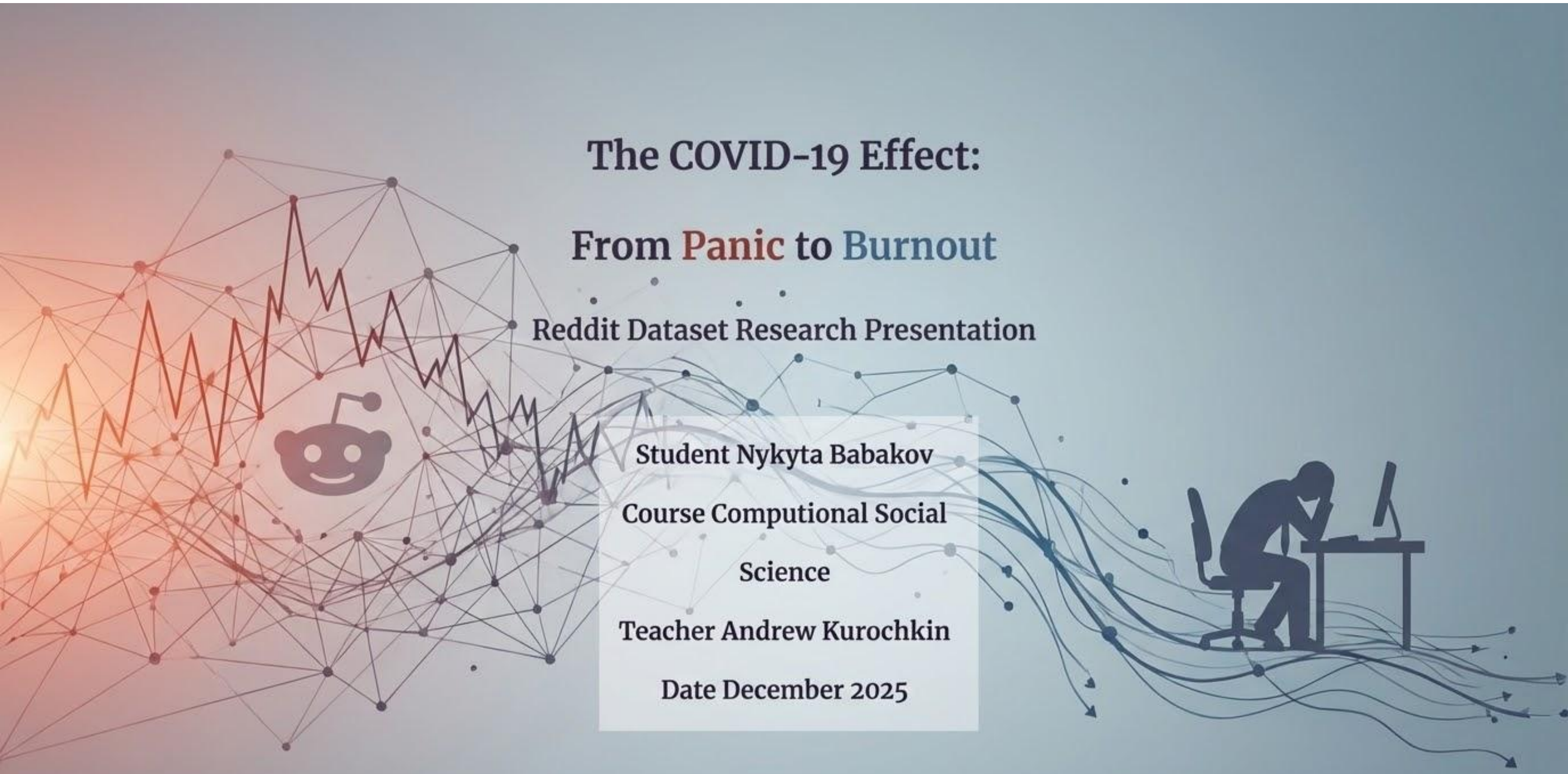
Reddit Dataset Research Presentation

Student Nykyta Babakov

Course Computional Social Science

Teacher Andrew Kurochkin

Date December 2025



Agenda

- Data Collection & Preparation
- EDA Workflow
- Final Insights
- Conclusion

Methodology & Data Collection

Dataset

- **Size:** 1,4 GB
- **Number:** ~3M posts/comments
- **Timeframe:** January-April 2020
- **Subreddits:** r/Depression, r/SuicideWatch, r/MentalHealth, r/Anxiety, r/Coronavirus, r/COVID-19_support

Tools Used

- **PushshiftDumps:** decompress .zst files, filter by topic or subreddit, export to CSV
- **TextBlob:** Polarity (-1 to +1), Subjectivity (0 to 1)
- **Seaborn/Matplotlib:** Visualizing patterns

Data Cleaning & Preparation

Missing Values

Check for NaN in key columns like Title, Text/Body, Score and remove.

Duplicates

Remove duplicate rows, especially in Title, Text/Body, Score.

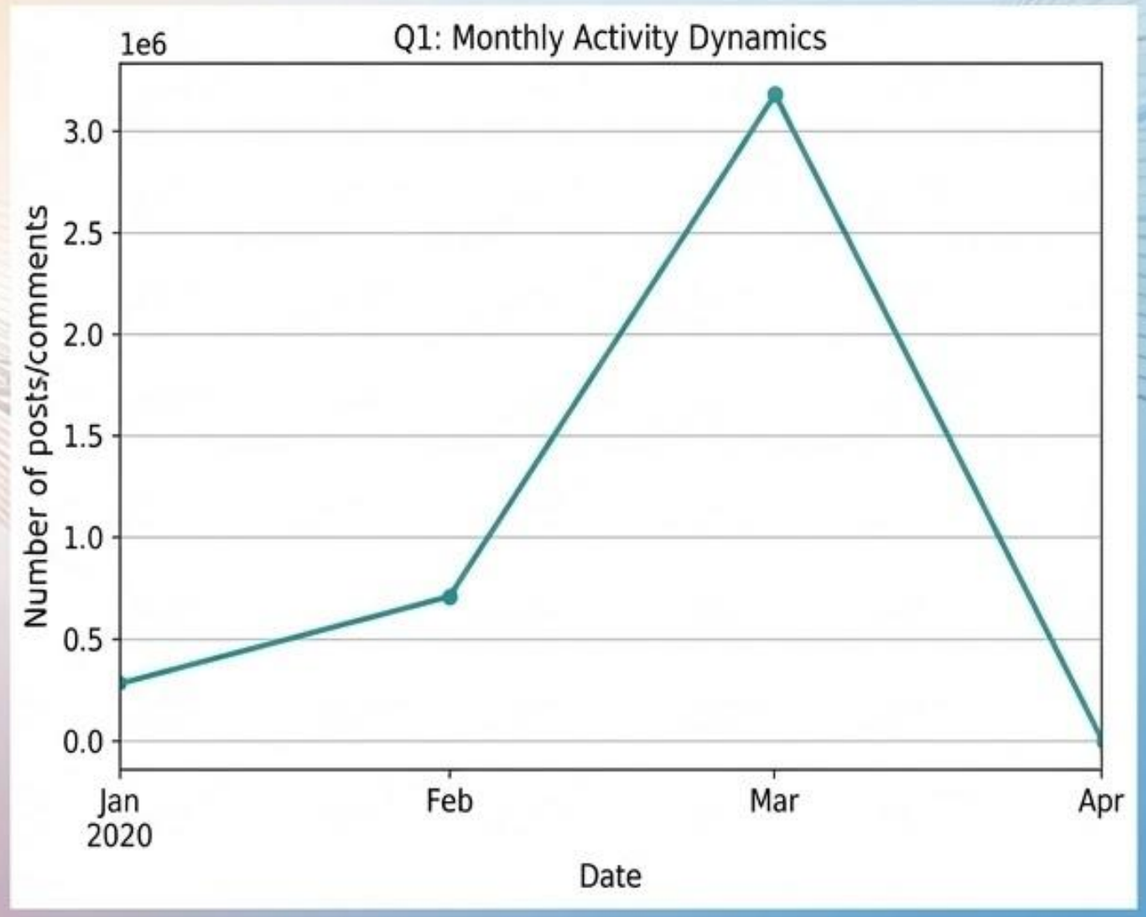
Data Types

Ensure numbers are numeric, dates are date objects.

Activity Dynamics: Monthly Trends

Q1: How has user activity changed over time?

Sharp peak in March 2020 (over 3 million entries) correlates with COVID-19 onset and quarantine restrictions.



Activity by Day of the Week

Q2: On which days do people post most frequently?

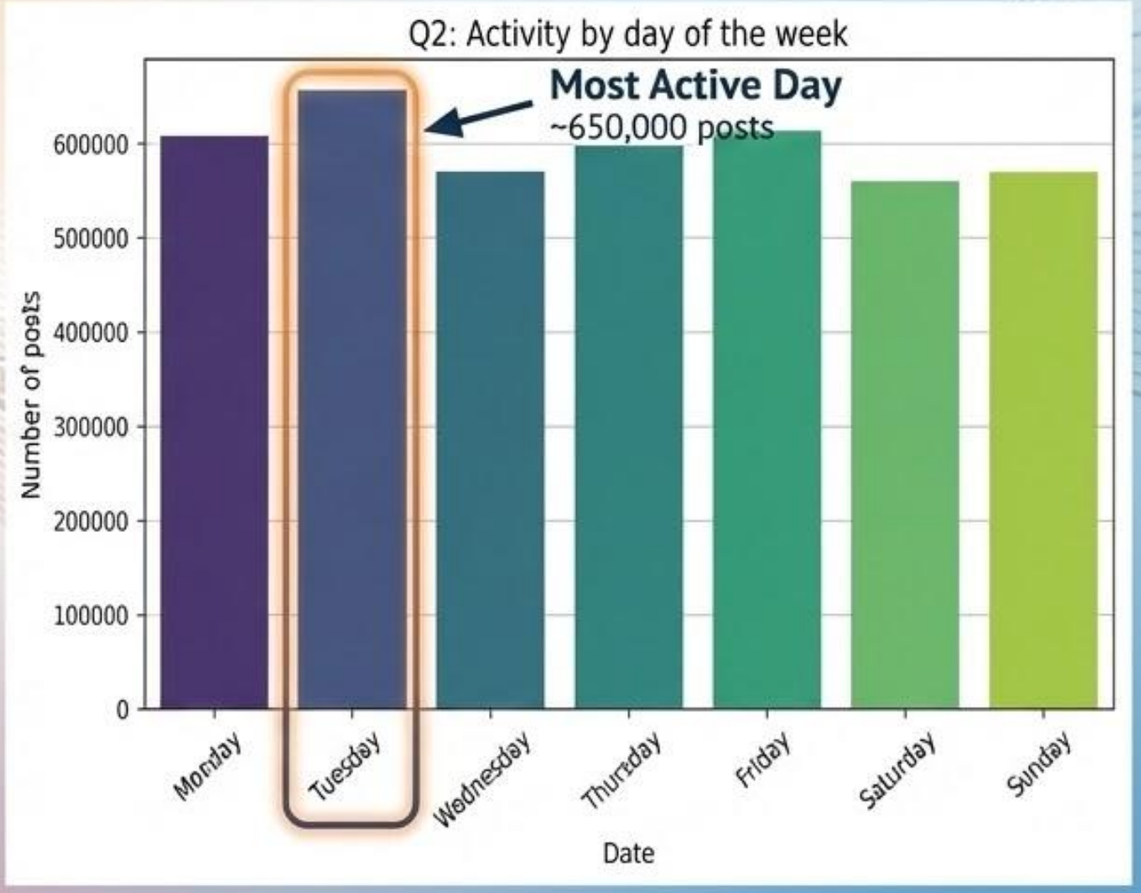
Hypothesis:

Are people lonelier (more active) on weekends, or are they busier?

Findings:

Hypothesis disproven: Distribution is even, with slight dip on Saturday/Sunday.

Tuesday is most active.

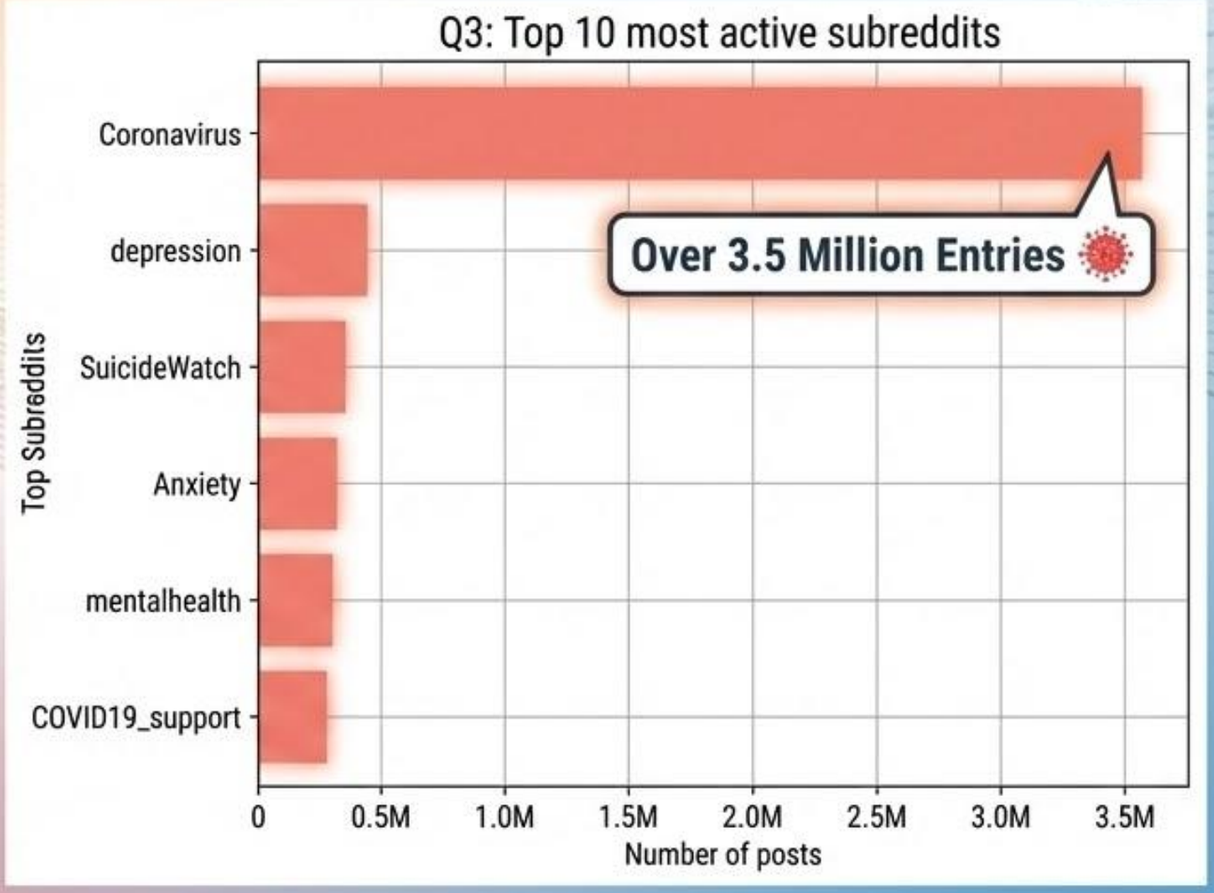


Top 10 Subreddits

Q3: Which subreddits are most active?

The Absolute Leader:

r/Coronavirus is the absolute leader (**over 3.5 million entries**), significantly ahead of mental health topics. Logical for a new viral term.



Top Authors: Anonymity & Bots

Q4: Who are the most active authors?

Top Authors by Post Count

u/[deleted] - 618334
u/AutoModerator - 76958
u/DeadlyKitt4 - 6724
u/ThisToWillPass - 4868
u/AnotherTrowaway12 - 4520
u/barber5 - 3660
u/IReadTheWholeArticle - 3582
u/Viewfromthe31stfloor - 3526
u/KorgRue - 3248
u/PsecretPseudonym - 3202

Deleted Posts

Absolute leader, especially in mental health subreddits. Users delete traces.

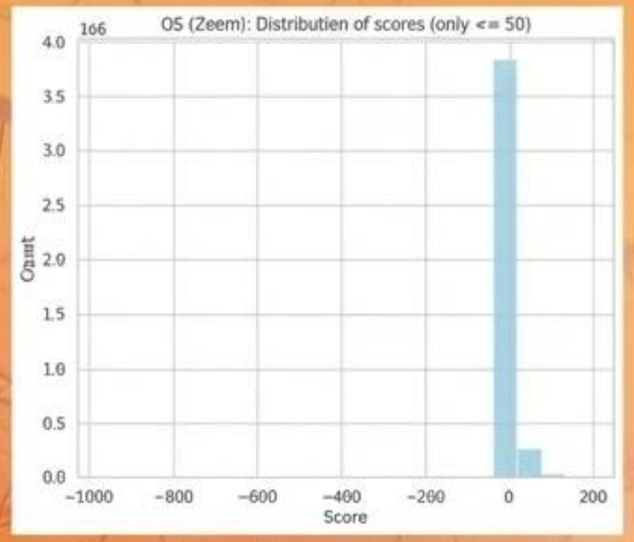
Auto-Moderation

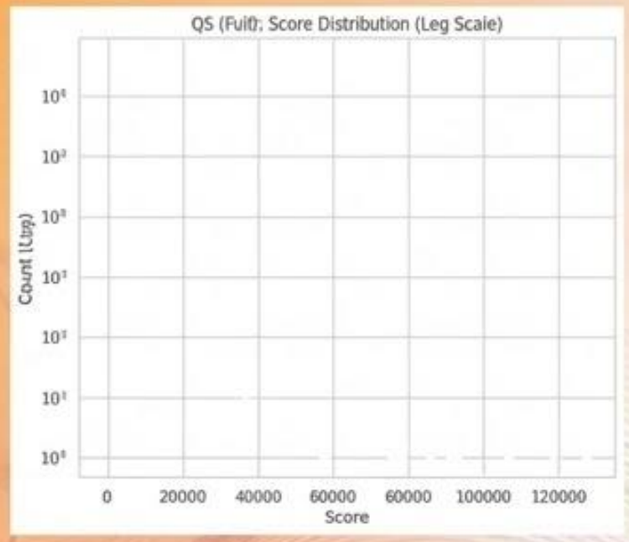
Second place, generating automated messages (community rules, warnings).

Live Users

Start at 3rd place, 100x fewer messages than 'deleted' category, indicating high audience turnover.

Score Distribution & Average Rating





Score Distribution

Most posts have a score from -1 to 1.
Few posts get high scores, confirming skewed distribution.

Top subreddits by average post rating:

Subreddit Coronavirus 15.401117
COVID19_support 5.450614
Anxiety 5.045908
depression 4.843577
SuicideWatch 3.240182
mentalhealth 2.991188

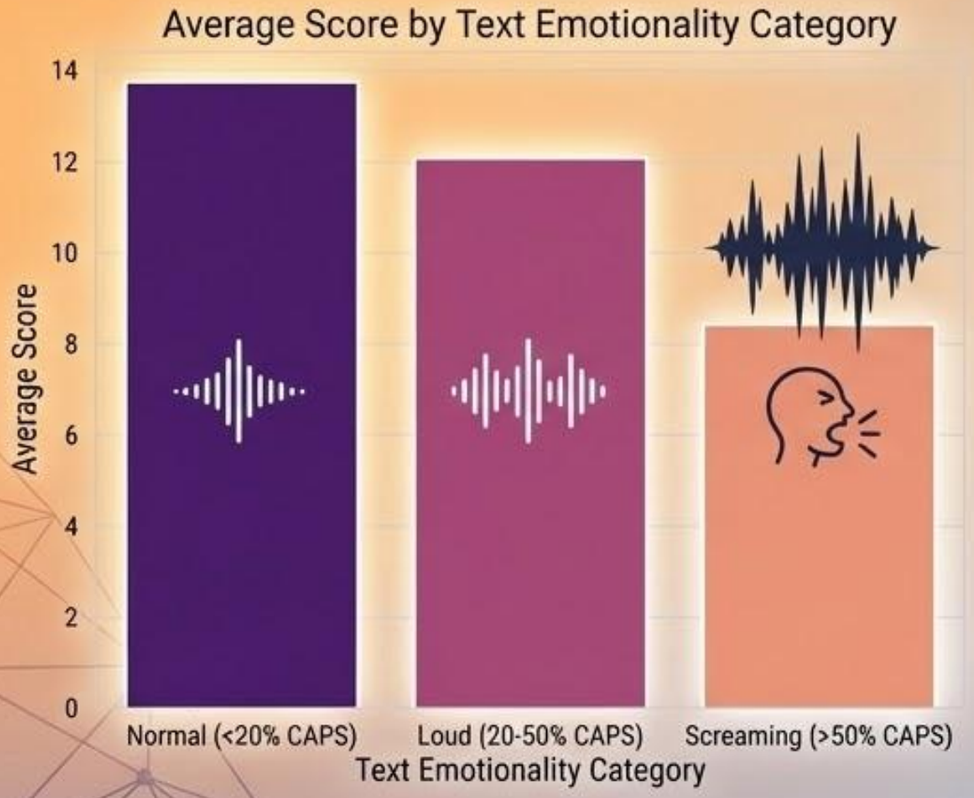
Top Subreddits by Average Rating

r/Coronavirus (15.4 points) leads due to news virality. Support subreddits (3.0-5.5 points) have lower average scores.

- Coronavirus (15.4 points)
- depression 4.843577
- COVID19_support 5.450614
- SuicideWatch 3.240182
- Anxiety 5.045908
- mentalhealth 2.991188

Scream Factor: The CAPS LOCK Effect

Q15: Do posts with CAPS LOCK get higher ratings?



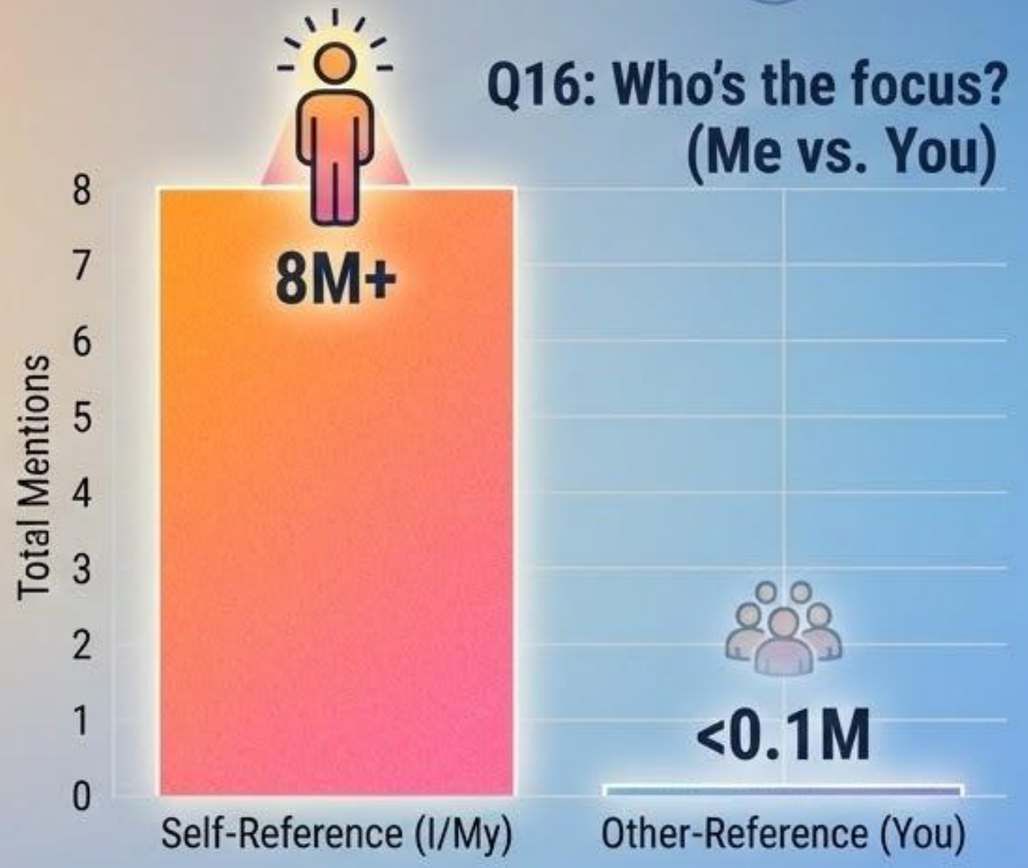
Does using CAPS LOCK affect a post's rating? Pure proof that aggressive style doesn't work.

Posts where more than 50% of the text is written in capital letters ("Screaming") receive the lowest average rating (about 8).

"Normal" and "Loud" posts are rated much higher (13-14 points).

Me vs You

Huge difference. Almost all texts (8 million instances) are focused **on the self** (I/My), while addressing **the other (You) is almost absent**. This is typical of support communities where people share their experiences or concerns.

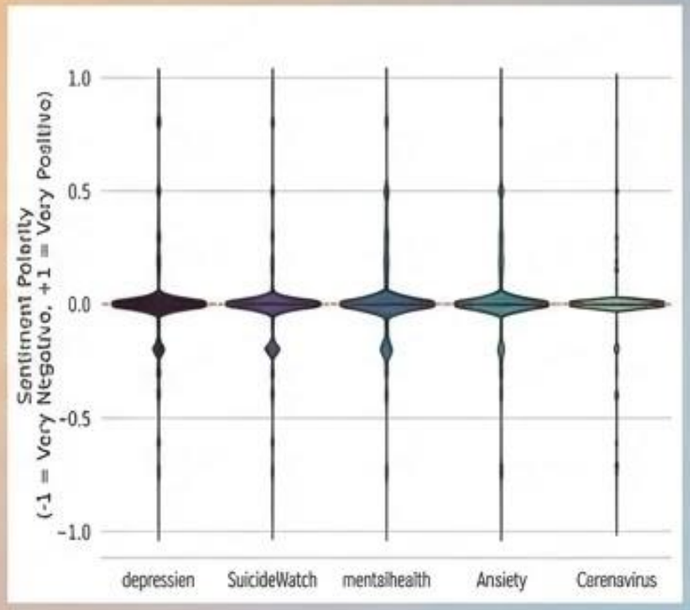


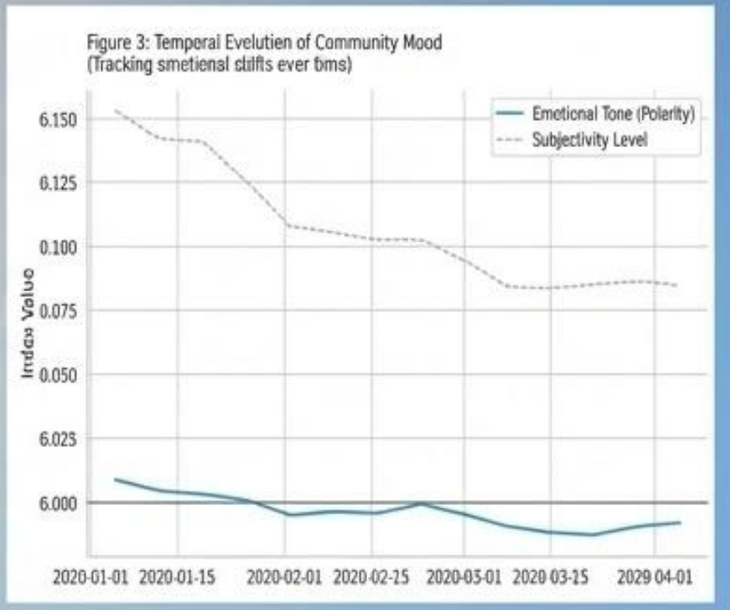
Community Mood: Emotional Desensitization

Subjectivity Level dropped sharply from February to March, indicating "community burnout."

Users shifted from personal emotions to "doom-scrolling" objective events. Less about feeling, more about reacting to facts.

All communities showed a "depressed bias," even r/Coronavirus, highlighting the universal mental toll.

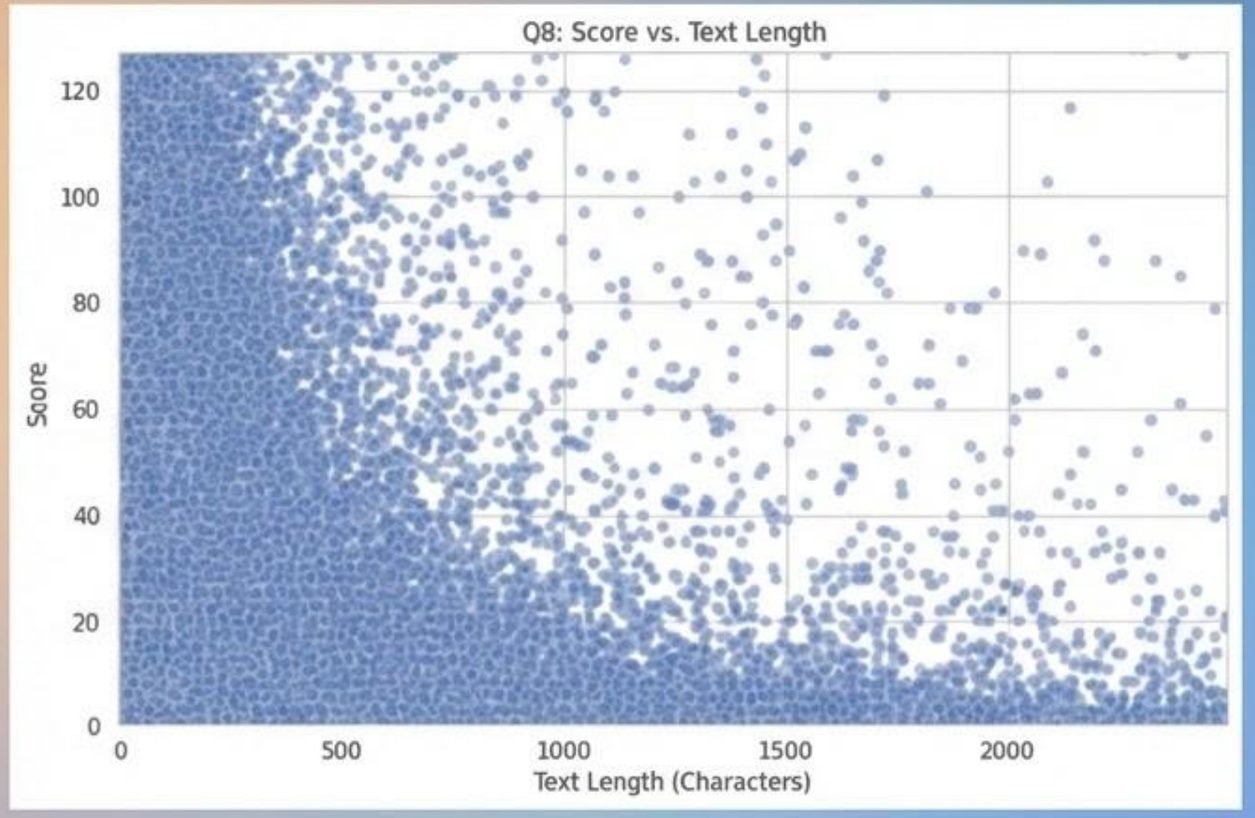




Score vs Text Length

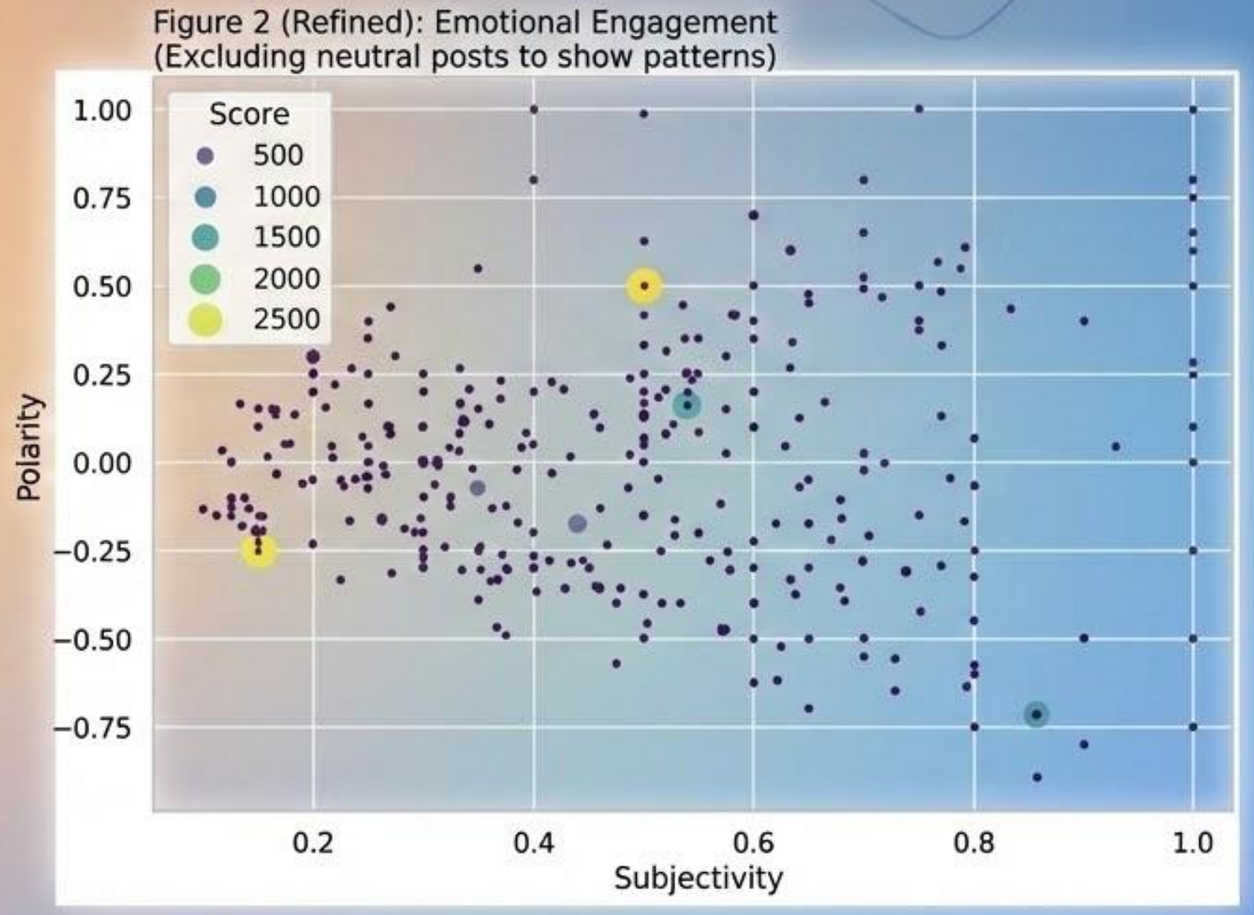
Longer texts (1000+ characters) for venting saw a **rapid drop** in engagement.

Do people like to read long posts? Not so.



Engagement Dynamics: Empathy vs. Information

Higher subjectivity correlated with higher polarity and engagement, exception may be some medical articles or researches.



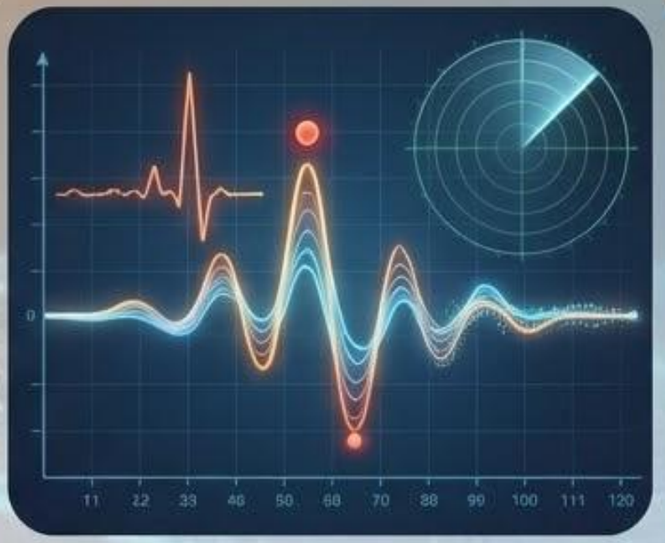
Future Work

Additional Text Analysis

Explore advanced NLP models to uncover new patterns, sentiment trends, and linguistic markers in the community's discourse.

User Segmentation

Segment users based on language patterns, activity levels, and emotional expression to understand distinct community subgroups.



Anomaly Detection

Identify specific prominent dates, events, and outliers that cause significant shifts in community behavior and sentiment.

My GitHub Repository

https://github.com/NykytaBabakov/CSS-2025

Thank you for your attention!