Università di Pisa

Scuola di Ingegneria

Corso di Laurea in Ingegneria Informatica

Tesi di Laurea

Reddit come Indicatore Sociale: un'Analisi *Data-Driven* sulla Variazione di Comportamento degli Utenti dopo l'Omicidio di George Floyd

---

Candidato:
Iacopo Stracca

Relatori:
Prof. Marco Avvenuti
Ing. Lorenzo Cima

Anno Accademico 2022-2023

Abstract

Questa tesi esplora l'impatto dei social media come specchio della società, focalizzandosi sull'omicidio di George Floyd del 25 maggio 2020. Utilizzando Reddit come piattaforma di studio, si analizzerà il comportamento degli utenti prima e dopo l'evento, esaminandone i commenti e cercando inoltre correlazioni tra il loro punteggio e la tossicità. L'obiettivo è comprendere come gli eventi del mondo reale influenzino il comportamento delle persone online.

Indice

1 Introduzione

1.1 Panoramica

Nel corso dell'ultimo decennio, i social media sono stati il principale mezzo di scambio di informazioni e opinioni personali, costituendo uno specchio della società reale, arrivando talvolta perfino a influenzare gli eventi del mondo esterno. Il problema principale che questa tesi si pone è quello di stabilire, attraverso l'analisi di un caso specifico come l'omicidio di George Floyd del 25 maggio 2020, se è possibile riscontrare, attraverso analisi eseguite sui commenti pubblicati dagli utenti su un social network, variazioni di comportamento in conseguenza a questo tipo di avvenimenti.

1.2 Reddit

Reddit è una piattaforma social che si caratterizza per la presenza di comunità (*subreddit*) inerenti a qualche argomento (politica, musica, sport...), all'interno delle quali gli utenti possono interagire tra di loro.

La diffusione negli Stati Uniti è notevole, ed è caratterizzata da una continua crescita di anno in anno[1]. Eventi e discussioni rilevanti a livello nazionale, come le elezioni, le proteste sociali e gli scandali, spesso generano un aumento di utenti e attività su Reddit, contribuendo ulteriormente alla sua diffusione e alla sua influenza nell'opinione pubblica.

1.3 Gli obiettivi

1.3.1 Data mining

Non disponendo della potenza di calcolo necessaria per riuscire ad analizzare tutti i commenti di Reddit, il primo obiettivo di questo lavoro è quello di ottenere un dataset di commenti molto più ristretto, ma sufficientemente rappresentativo, su cui poter svolgere le successive analisi.

1.3.2 Comportamento degli utenti

Attraverso l'analisi dei commenti selezionati, in particolare sfruttando la libreria *Detoxify*, si cercherà di analizzare la variazione del comportamento degli utenti in relazione all'omicidio di Floyd, tenendo conto dell'estensione temporale dell'evento: non solo l'omicidio in sé, ma anche le proteste (in alcuni casi violente) che ne sono scaturite nelle settimane successive.

1.3.3 Analisi statistica

Ulteriore oggetto di studio sarà l'esistenza di una correlazione tra il *punteggio* di un commento e la sua *tossicità*; infine si analizzeranno statisticamente le distribuzioni di *tossicità* e *punteggio* dei commenti per ogni utente, facendo distinzione tra i periodi *pre* e *post* omicidio, al fine di osservare qualche risultato interessante anche in questo ambito.

2 Related Works

2.1 Letteratura

Sull'analisi del comportamento degli utenti sui social media esistono sono già stati effettuati diversi studi, tra cui:

- **Grimmelmann, 2015, The Virtues of Moderation**[2]: tratta dell'importanza della moderazione nelle comunità online. Esamina come questa possa contribuire a mantenere discussioni civili, prevenire abusi e promuovere comportamenti positivi nelle piattaforme digitali.
- **Francois, 2021, Measuring coordinated versus spontaneus activity in online social movements**[3]: si concentra sull'analisi delle attività online, cercando di distinguere tra azioni *coordinate* e *spontanee* all'interno di tali contesti, con un'attenzione particolare alle dinamiche di mobilitazione online.
- **Jhaver, 2021, Evaluating the Effectiveness of Deplatforming as a Moderation Strategy on Twitter**[4]: esamina l'efficacia del *deplatforming* come strategia di moderazione su Twitter. Esplora come l'eliminazione o la restrizione dei contenuti o degli utenti controversi influisca sulla qualità e sull'ambiente sociale della piattaforma.
- **Trujillo M, 2021, Examining the Use of Community-Specific Language Post-Subreddit Ban**[5]: analizza le reazioni degli utenti al *ban* di subreddit specifici. Esplora come l'eliminazione di tali comunità influenzi il linguaggio e il comportamento degli stessi utenti in altri subreddit della piattaforma.
- **Habib, 2022, Are Proactive Interventions for Reddit Communities Feasible?**[6]: esplora la possibilità di applicare interventi *proattivi* per gestire le comunità su Reddit, cercando di migliorare l'ambiente sociale della piattaforma. Analizza la fattibilità e l'efficacia di queste strategie nell'affrontare i problemi all'interno delle comunità online.
- **Singhal, 2022, SoK- Content Moderation in Social Media, from Guidelines to Enforcement, and Research to Practice**[7]: offre un'analisi approfondita sulla moderazione dei contenuti nei social media. Esamina il passaggio dalle linee guida alla loro applicazione pratica, esplorando come la ricerca influenzi l'implementazione delle strategie di moderazione.
- **Trujillo, 2022, Make Reddit Great Again: Assessing Community Effects of Moderation Interventions on r/The_Donald**[8]: uno studio sugli utenti del subreddit *The_Donald*, volto a misurare l'efficacia dei vari livelli di moderazioni attuati da Reddit: *Quarantine*, *Restriction* e *Ban*.

3 Dataset

3.1 Strategia di costruzione del dataset

Illustriamo l'idea che è stata seguita per ottenere un dataset di commenti non troppo ingente ma su cui, al tempo stesso, sia possibile uno studio della variazione di comportamento degli utenti.

3.1.1 Scelta dell'arco temporale

Il dataset dei commenti di Reddit è disponibile pubblicamente[9] in file compressi raggruppati per mese. Essendo l'omicidio di Floyd avvenuto il 25 maggio 2020, scegliamo di prendere in esame i commenti da aprile a luglio 2020.

3.1.2 Quali commenti selezionare?

Dobbiamo decidere quali di questi commenti inserire nel dataset che andremo a utilizzare per le successive analisi. L'idea è cercare di ottenere un primo insieme di commenti sul tema George Floyd, selezionare gli utenti con più commenti in questo gruppo e analizzare il loro comportamento su tutti i commenti da loro pubblicati nel periodo aprile-luglio.

3.2 Estrazione

I files sono di dimensioni troppo grandi per poter essere estratti nella loro interezza. Si è fatto uso di uno script Python[10] in grado di prendere in input un file compresso, contenente dunque tutti i commenti di un certo mese, esaminare i commenti uno ad uno decidendo se tenerli o scartarli in base a una certa regola e restituire in output un altro file compresso contenente i commenti non scartati. Il file ottenuto, a seconda del criterio scelto, può essere di dimensione nettamente minore rispetto all'originale, dunque è possibile decomprimerlo e ottenere un .csv. Per ogni commento il dataset include diversi campi, di cui quelli di nostro interesse sono:

- **author**: l'username dell'utente che ha pubblicato il commento;
- **body**: il testo del commento;
- **subreddit**: il subreddit nel quale il commento è stato diffuso;
- **timestamp**: data e ora di caricamento del commento;
- **score**: una sorta di "punteggio" associato al commento. Aprrofondiremo il concetto nella sezione 5.1.

3.2.1 Criteri di estrazione

Come primo criterio di selezione è stato deciso di imporre la presenza della parola *Floyd*. Applicando unicamente questo criterio, tuttavia, si è notato come molti commenti "superstiti" fossero riferiti ai *Pink Floyd*, a *Floyd Mayweather Jr.*, pugile professionista, o a *Leonard Floyd*, giocatore di football americano. Si è deciso dunque di imporre ulteriori criteri di selezione sul *body* dei commenti:

- L'assenza della parola *pink*;

- l'assenza della parola *boxe*;
- la presenza di almeno una parola tra *black*, *George* e *kill*.

Sono stati aggiunti inoltre dei criteri, non più sul *body*, bensì sul nome del *subreddit* di appartenenza del commento analizzato:

- L'assenza della parola *guitar*;
- l'assenza della parola *mma*;
- l'assenza della parola *boxing*;
- l'assenza della parola *nfl*.

In tabella sono riportati i numeri dei commenti per ogni mese ottenuti con il filtraggio più elementare (solo presenza di "floyd") e con il filtraggio appena esposto.

Mese	Primo filtraggio	Filtraggio completo	Differenza
Aprile 2020	15,357	205	-98.6%
Maggio 2020	81,941	53,180	-35.1%
Giugno 2020	208,871	159,190	-23.8%
Luglio 2020	55,020	36,093	-34.4%

Tabella 1: Numero di commenti ottenuti con i due livelli di filtraggio effettuati

Notiamo che il filtro ha eliminato quasi tutti i commenti di aprile, mentre il mese meno intaccato è stato giugno, nel quale presumibilmente la vicenda Floyd è stata maggiormente oggetto di discussione. Il fatto che i commenti di aprile non siano zero, ci fa capire che il filtro non è perfetto e potrebbe essere ulteriormente affinato, ma il loro numero risulta abbastanza ridotto da renderci complessivamente soddisfatti del risultato.

3.3 Eliminazione dei commenti di bot

Ciò che si nota provando a leggere un campione dei commenti al passo precedente ottenuti, è la frequente alternanza tra commenti di persone *reali* e commenti pubblicati da *bot*. Effettuiamo, con uno script Python, una selezione per cercare di ridurre al minimo la quantità di commenti provenienti da bot.

3.3.1 Criteri di eliminazione

Per prima cosa stabiliamo di escludere i commenti provenienti da utenti che contengono nel proprio *username* le parole *auto* o *bot*. Nonostante il gran numero di commenti eliminati da tale criterio, è possibile effettuare un'ulteriore scrematura prendendo in esame i commenti "superstiti" e raggruppandoli per *user* e *body*, contando per ciascuna coppia il numero di occorrenze. Un comportamento tipico dei bot è infatti quello di commentare un gran numero di volte con lo stesso *body*.

Scegliamo di considerare come bot tutti quelli utenti che abbiano commentato, nell'arco dei 4 mesi, almeno 30 volte con le stesse identiche parole.

3.4 Selezione degli utenti da analizzare

Una volta eliminati i bot, possiamo stilare una classifica degli utenti con più commenti "superstiti" all'attivo.

3.4.1 Ultima eliminazione (manuale)

Dando un'occhiata alla suddetta classifica si nota come ci sia qualche utente in testa decisamente "staccato" dagli altri. Uno di questi, l'utente *deleted*, non rappresenta un utente vero e proprio, bensì l'insieme degli utenti cancellati. Non essendo i relativi commenti riconducibili a un unico utente, ma a una moltitudine indefinita, decidiamo di non prenderli in considerazione.

Altri due utenti che occupano le prime posizioni, dopo un'ispezione diretta dei commenti, si sono rivelati essere bot sfuggiti ai precedenti filtraggi, e sono stati eliminati dalla classifica.

Riportiamo in tabella i numeri dei commenti, mese per mese, prima e dopo l'eliminazione dei bot.

Mese	Filtraggio completo	Eliminazione Bot	Differenza
Aprile 2020	205	188	-8.29%
Maggio 2020	53,180	48,308	-9.16%
Giugno 2020	159,190	140,276	-11.86%
Luglio 2020	36,093	32,851	-8.98%

Tabella 2: Numero di commenti ottenuti prima e dopo l'eliminazione dei bot

In questo caso notiamo una differenza in percentuale abbastanza uniforme nei vari mesi.

3.4.2 Scelta degli utenti

Dalla classifica così ottenuta decidiamo di prendere in considerazione gli utenti con almeno 35 commenti, ovverosia i primi 46. La scelta è stata fatta tenendo conto delle limitate capacità di calcolo a disposizione e dell'impraticabilità, quindi, di selezionare un numero troppo elevato di commenti.

3.5 Ottenimento del dataset di commenti da analizzare

Una volta selezionati gli utenti, prendiamo nuovamente in esame i file compressi contenenti tutti i commenti raggruppati per mese. Tramite una rivisitazione dello script Python utilizzato nella sezione 3.2, estraiamo ed esportiamo in .csv unicamente i commenti provenienti da utenti che siano tra i 46 selezionati.

Ricapitolando, i commenti che compongono il dataset che si è usato per le successive analisi, sono tutti i commenti pubblicati da aprile a luglio 2020 da un sottoinsieme di utenti, ovvero i 46 più attivi sul tema George Floyd. Il numero di commenti selezionati è di circa 125,000.

4 Analisi sulla Toxicity

4.1 La libreria *Detoxify*

L'obiettivo dell'analisi è ora quello di associare, ad ogni commento, un valore che ne indichi la "tossicità". Per risolvere questo problema, piuttosto complesso, possiamo fare uso della libreria Python *Detoxify* nella sua versione *original*, pubblicata nel 2018, che ha dimostrato un'accuratezza del 98.64%. Oltre al valore di *Toxicity*, la funzione fornita dalla libreria calcola, per ogni commento, anche un valore di *Severe_ toxicity*, oltre ad altri valori che non sono di interesse per l'analisi, quali *Obscene*, *Threat*, *Insult* ed *Identity_ attack*, che vanno a indicare l'ambito nel quale, nello specifico, il commento risulta più "tossico". Ciascuno dei suddetti indici è rappresentato da un valore decimale compreso tra 0 e 1, dove 0 indica l'assenza della caratteristica descritta e 1 la totale adesione del commento a tale caratteristica.

4.2 La scelta tra *Toxicity* e *Severe_ toxicity*

Al contrario degli altri indici, che individuano tipologie specifiche di commenti malevoli, la scelta tra *Toxicity* e *Severe_ toxicity* non è, a primo impatto, così scontata.

Per capire quale indice convenga utilizzare nell'analisi, raccogliamo i valori di *Toxicity* e di *Severe_ toxicity* di tutti i commenti componenti il dataset, e andiamo a raffiguare un violin plot per ciascuno dei due indici, per studiarne la distribuzione.

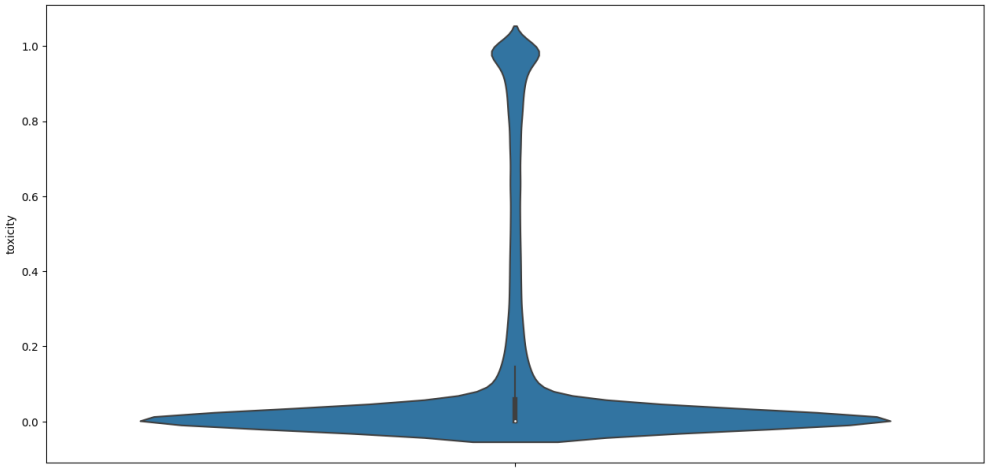

Figura 1: Violin plot della Toxicity

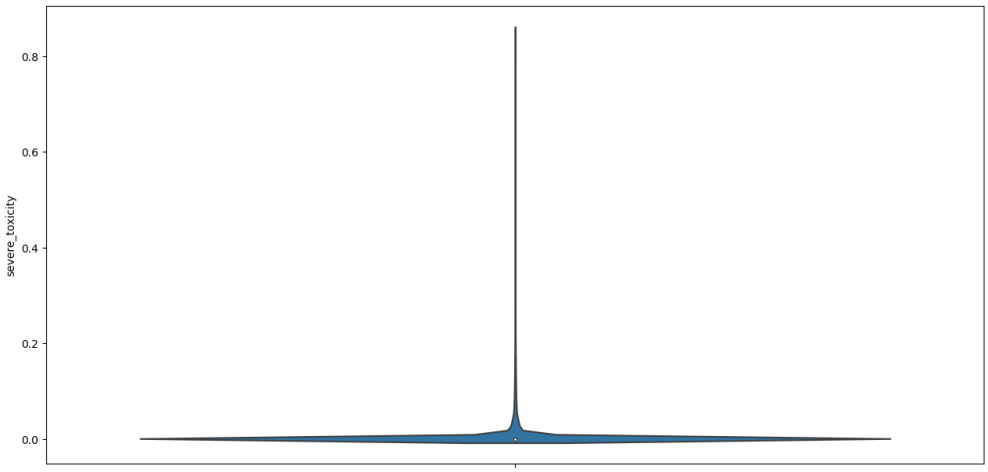

Figura 2: Violin plot della Severe_toxicity

Come si può notare, entrambi i grafici presentano un "rigonfiamento" in corrispondenza di valori prossimi a zero. Questo perché la maggior parte dei commenti del dataset sono inoffensivi, e privi dunque di tossicità. La differenza tra i due violin plots è tuttavia evidente nella parte superiore: mentre in quello della *Toxicity* si nota che è presente un certo numero di commenti per ogni valore della scala, con un picco in corrispondenza dei valori vicini ad 1, nel grafico riferito alla *Severe_ toxicity* emerge che sono pochissimi i commenti con valori non prossimi allo zero, e dunque risulta molto difficile fare un'analisi, perché la distribuzione è troppo "schiacciata" verso l'estremo inferiore.

Come indice di riferimento per quantificare la tossicità dei commenti useremo quindi, nelle analisi seguenti, la *Toxicity*.

4.3 Analisi sulla Toxicity media

Come prima analisi decidiamo di rappresentare su un grafico, giorno per giorno, il valore medio della Toxicity dei commenti pubblicati. Ci si aspetterebbe che, sia nei giorni immediatamente successivi all'omicidio, che, ancor di più, una volta scoppiata la discussione in merito alle proteste violente, la Toxicity media abbia subito un aumento.

In ognuno dei grafici che seguono, è stata evidenziata in arancione la data dell'omicidio di Floyd; in verde la data in cui hanno avuto inizio le proteste violente.

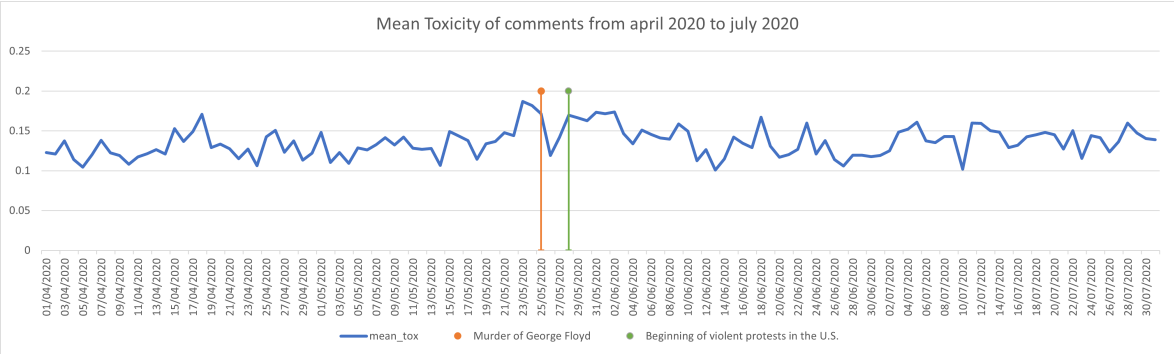

Figura 3: Toxicity media giorno per giorno

Dal grafico appare evidente che le aspettative non sono state rispettate. Sembra addirittura che sia stato registrato un picco negativo di Toxicity media il 26 maggio 2020, giorno immediatamente successivo all'omicidio. Per provare a spiegare questo risultato ricorriamo a un altro grafico, che mostra invece com'è cambiato il numero di commenti giorno per giorno.

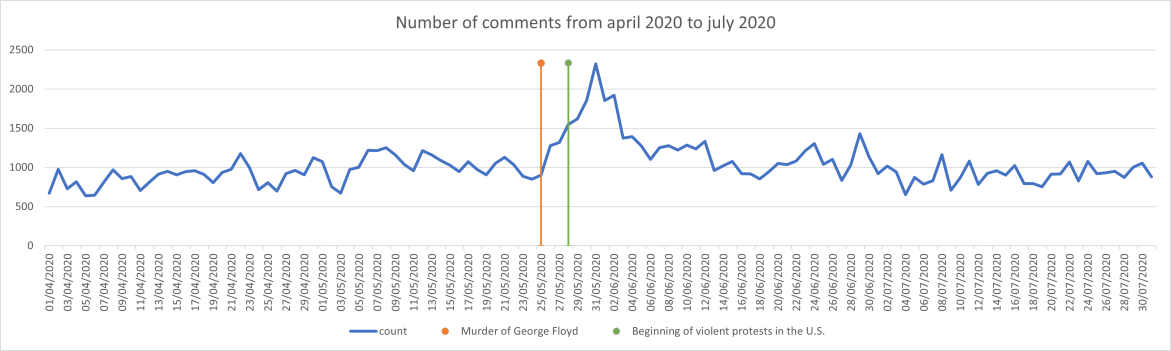

Figura 4: Numero di commenti giorno per giorno

Dalla lettura di questo grafico si evince che il numero di commenti ha subito un sensibile aumento in seguito all'omicidio di Floyd, e ancor di più nei giorni delle proteste. L'invarianza della Toxicity media è dunque dovuta alla grande crescita del numero di commenti totali, ossia il denominatore nel calcolo della media.

4.4 Analisi sulla Toxicity con soglia

Possiamo svolgere una seconda analisi seguendo un diverso approccio. Fissata una soglia, che chiamiamo *value*, rappresentiamo, giorno per giorno, il numero di commenti tali che toxicity(*commento*) $>$ *value*, per *value*=$\{0.5,0.7,0.9\}$.

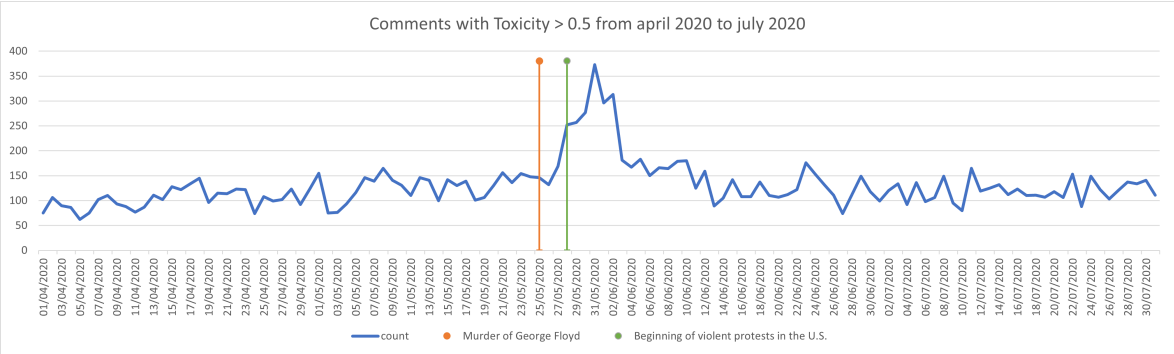

Figura 5: Numero di commenti, giorno per giorno, con Toxicity > 0.5

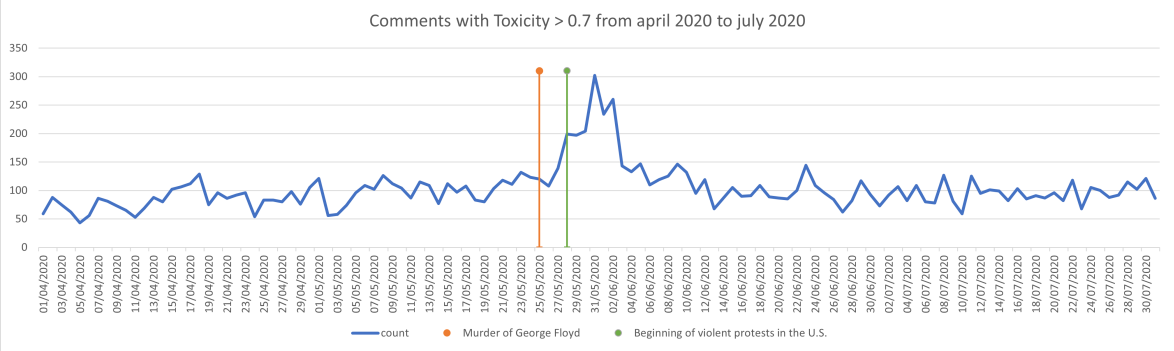

Figura 6: Numero di commenti, giorno per giorno, con Toxicity > 0.7

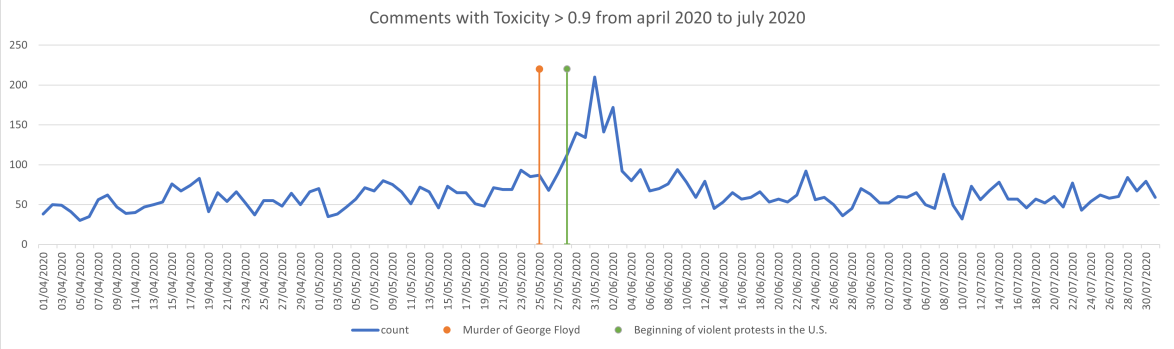

Figura 7: Numero di commenti, giorno per giorno, con Toxicity > 0.9

Svolgendo questo tipo di analisi, il picco immediatamente successivo all'inizio delle proteste risulta evidente. In questo caso, possiamo dunque affermare che è possibile riscontrare la reazione innescata dall'avvenimento nell'analisi dei commenti.

4.5 Analisi in percentuale

Proviamo a unire i due approcci utilizzati in precedenza: abbiamo capito che può essere conveniente tenere in considerazione il numero di commenti con Toxicity maggiore di un certo *value* di soglia, ma vorremmo comunque tener conto del numero totale di commenti.

Decidiamo quindi di tracciare un grafico, raggruppando i commenti non più per giorno ma per settimana, in cui raffiguriamo, per i diversi valori di *value* già usati precedentemente (0.5, 0.7, 0.9), la percentuale dei commenti con toxicity(*commento*) $>$ *value* sul numero totale di commenti.

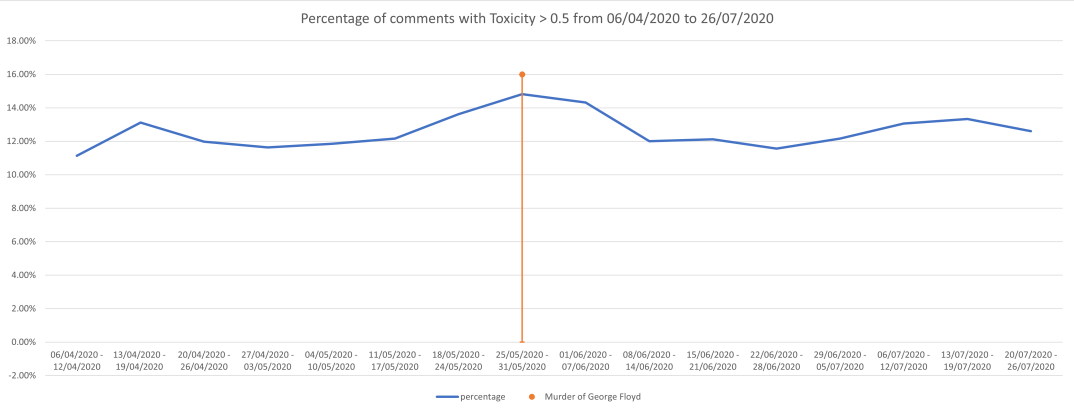

Figura 8: Percentuale di commenti con Toxicity $>$ 0.5 sul totale dei commenti, settimana per settimana

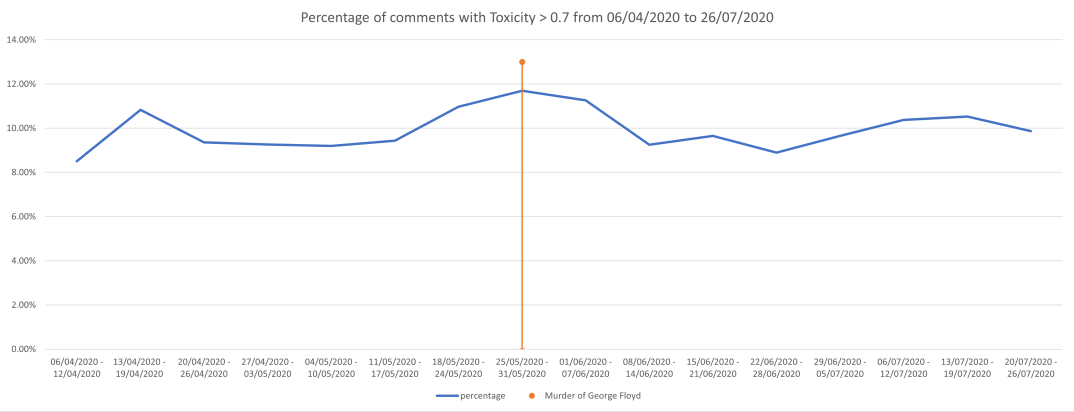

Figura 9: Percentuale di commenti con Toxicity $>$ 0.7 sul totale dei commenti, settimana per settimana

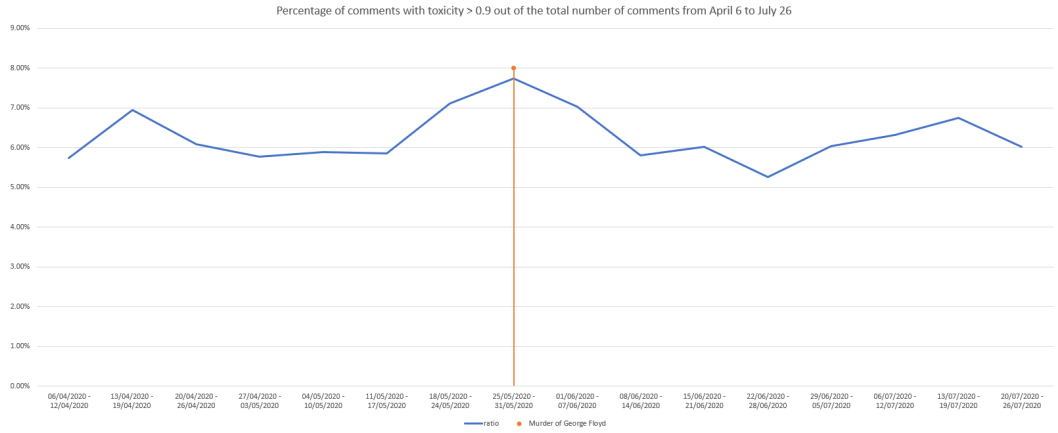

Figura 10: Percentuale di commenti con Toxicity > 0.9 sul totale dei commenti, settimana per settimana

Notiamo dei picchi in corrispondenza della settimana 25/05-31/05, non pronunciati come nell'analisi a soglia precedentemente svolta, ma comunque evidenti, al contrario di quanto osservato nella analisi svolta sulla Toxicity media (4.3).

5 Analisi della correlazione tra Toxicity e Score dei commenti

5.1 *Score* di un commento

Lo *Score* di un commento è la differenza tra *upvotes* e *downvotes* che il commento stesso riceve sulla piattaforma. Se, ad esempio, un commento viene valutato positivamente da 14 utenti e valutato negativamente da altri 4, il suo score è uguale a 14-4=10.

5.2 Correlazione tra Toxicity e Score

Andiamo a studiare se esiste qualche tipo di correlazione tra la Toxicity di un commento e lo Score da esso totalizzato. Per farlo, tracciamo uno scatter plot, rappresentando sull'asse orizzontale la Toxicity e su quello verticale lo Score. Ogni "puntino" rappresenta un commento.

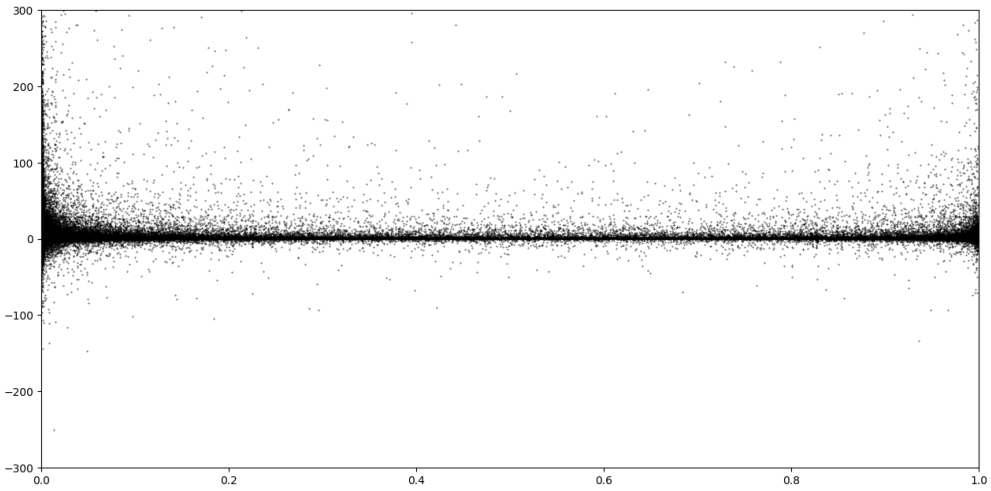

Figura 11: Toxicity e Score per ciascun commento

Essendo il dataset composto da oltre 125.000 commenti, risulta impensabile comprendere qualcosa affidandosi unicamente alla rappresentazione mediante scatter plot. Proviamo quindi a sovrapporre ad esso un kernel density plot, che raffigura con diversi colori zone a diverse densità di "puntini".

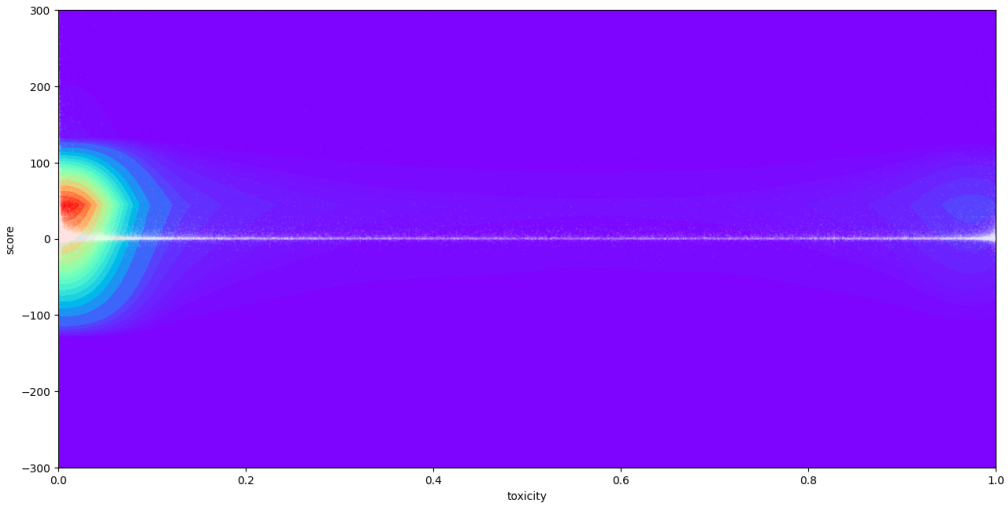

Figura 12: Toxicity e Score per ciascun commento con kernel density plot

La grande concentrazione di commenti a bassa tossicità non lascia interpretare al meglio il grafico. Proviamo dunque a ripetere il plot su due dataset separati: uno contenente i commenti aventi Toxicity < 0.5, l'altro i commenti con Toxicity ≥ 0.5

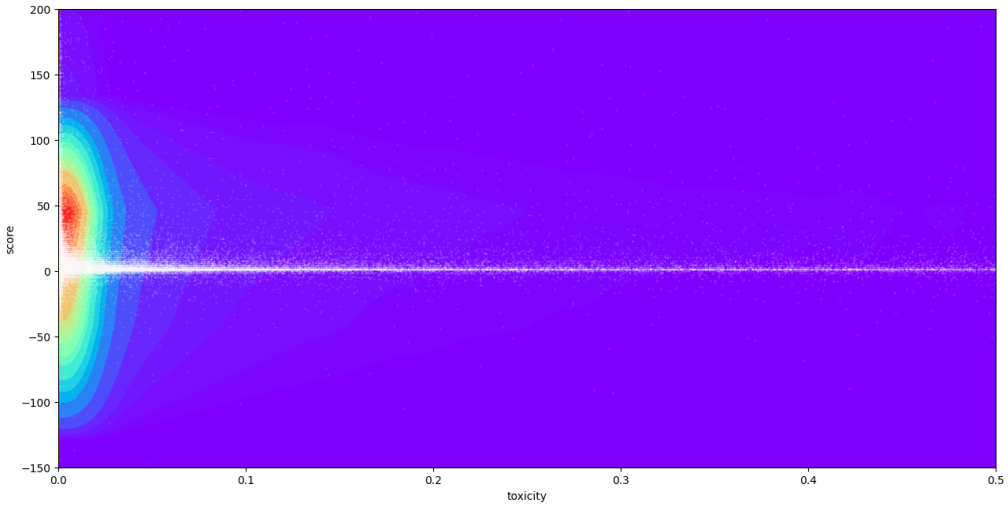

Figura 13: Toxicity e Score per ciascun commento (avemte Toxicity$<$0.5) con kernel density plot

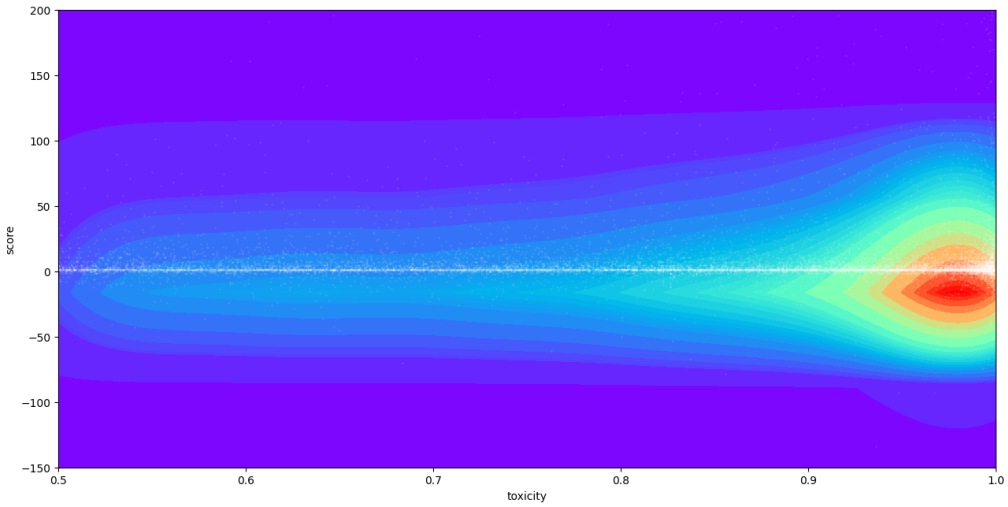

Figura 14: Toxicity e Score per ciascun commento (avente Toxicity$\geq$0.5) con kernel density plot

È ora possibile arrivare alla conclusione che i commenti con Toxicity < 0.5 (primo grafico) tendono ad avere uno score positivo, dunque sono stati più apprezzati dagli utenti; quelli con Toxicity > 0.5 (secondo grafico) tendono ad avere uno score minore di zero, quindi sono stati meno apprezzati.

6 Analisi statistica sulla distribuzione della Toxicity

Dall'analisi svolta sulla media della Toxicity (Sezione 4.3) non eravamo riusciti ad osservare variazioni apprezzabili nei giorni successivi all'evento. Decidiamo quindi di analizzare i valori di Toxicity dei vari commenti in maniera più approfondita: consideriamo ogni utente singolarmente e studiamo, per ognuno, la distribuzione di Toxicity dei suoi commenti nei periodi *pre* e *post* omicidio. Per studiare le distribuzioni di probabilità ci avvaliamo di due indici statistici: la *skewness* e la *kurtosis*.

6.1 Skewness

Data una variabile aleatoria, una misura dell'asimmetria della sua distribuzione di probabilità è data dalla *skewness*. Questo indice può assumere valori sia positivi che negativi: un valore positivo di skewness indica che la *moda* della variabile aleatoria è maggiore della *media* e della *mediana* della stessa; un valore negativo indica che la *media* è maggiore della *moda* e della *mediana*.

6.1.1 Divisione del dataset per utenti

L'analisi che segue consiste nel:

- dividere i commenti del dataset per utente;
- considerare ognuno dei 46 insiemi di commenti (uno per ogni utente);
- calcolare per ognuno di essi la skewness sui valori di Toxicity dei commenti.

6.1.2 Violin plot su tutto il periodo

Per rappresentare la distribuzione dei valori di skewness ottenuti facciamo uso di un violin plot.

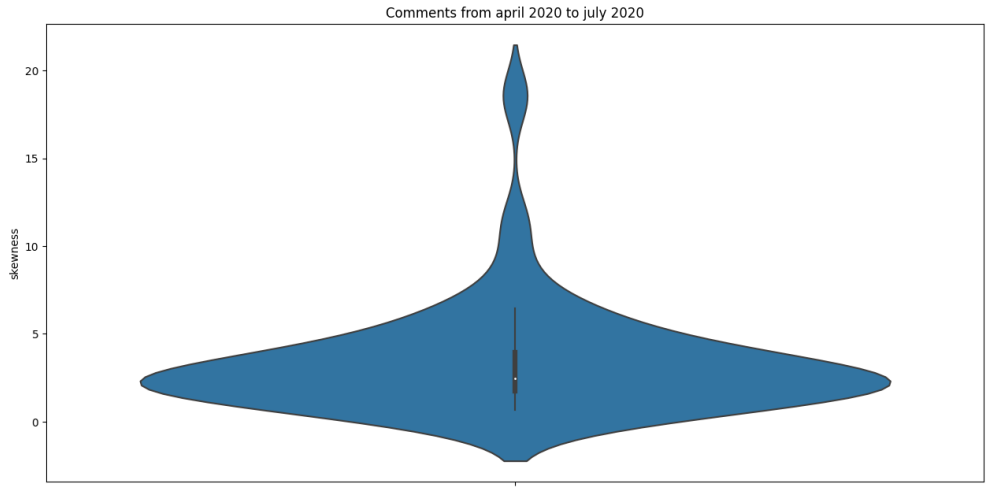

Figura 15: Violin plot dei valori di skewness calcolata sulla Toxicity dei commenti per ogni utente

Si nota come la quasi totalità degli utenti abbia una skewness maggiore di zero. Questo implica una distribuzione di Toxicity asimmetrica, con il picco (la *moda*) spostata verso sinistra. Più è alto il valore di skewness, più è marcata l'asimmetria nella distribuzione.

6.1.3 Violin plot diviso per periodi

Dividiamo ora il dataset dei commenti da aprile a luglio in due dataset distinti: il primo contenente i commenti dal 1° aprile al 25 maggio, data dell'omicidio di George Floyd; il secondo contenente i rimanenti, quindi quelli pubblicati dal 26 maggio al 31 luglio.

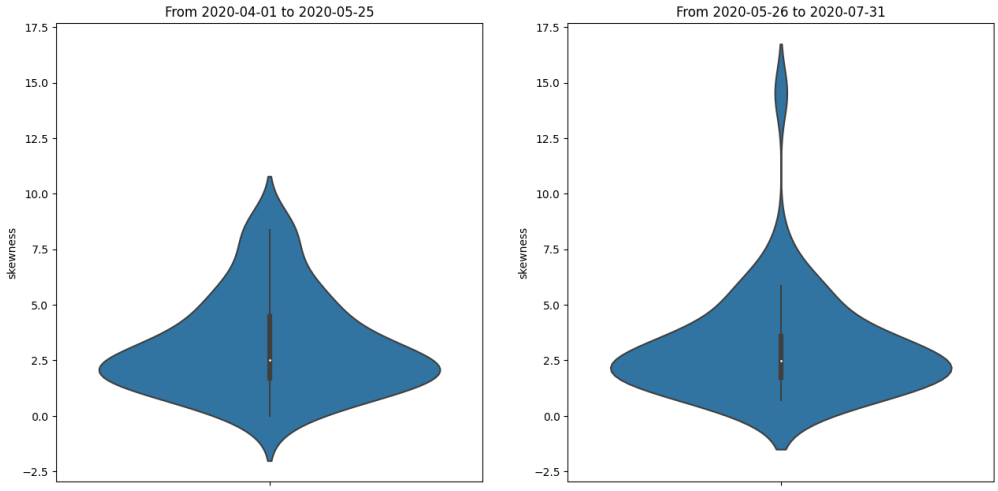

Figura 16: Violin plot dei valori di skewness calcolata sulla Toxicity dei commenti per ogni utente prima e dopo l'omicidio di Floyd

Non notiamo un cambiamento drastico, ma una tendenza a valori di skewness più alti, ad indicare una più elevata asimmetria della distribuzione di probabilità.

6.2 Kurtosis

Data una variabile aleatoria, la forma della sua distribuzione di probabilità può essere valutata mediante la *kurtosis*. Questo valore indica quanto le code della distribuzione siano più "pesanti" o "leggere" rispetto a una distribuzione normale. Un valore di kurtosis maggiore di zero indica code più "pesanti" (una distribuzione più "appuntita" al centro); un valore minore di zero indica code più "leggere" (una distribuzione più "piatta" al centro).

6.2.1 Divisione del dataset per utenti

Ripetiamo un'analisi analoga a quella appena svolta sulla skewness, che prevede quindi di:

- dividere i commenti del dataset per utente;

- considerare ognuno dei 46 insiemi di commenti (uno per ogni utente);
- calcolare per ognuno di essi la kurtosis sui valori di Toxicity dei commenti.

6.2.2 Violin plot su tutto il periodo

Possiamo rappresentare nuovamente i valori ottenuti mediante un violin plot.

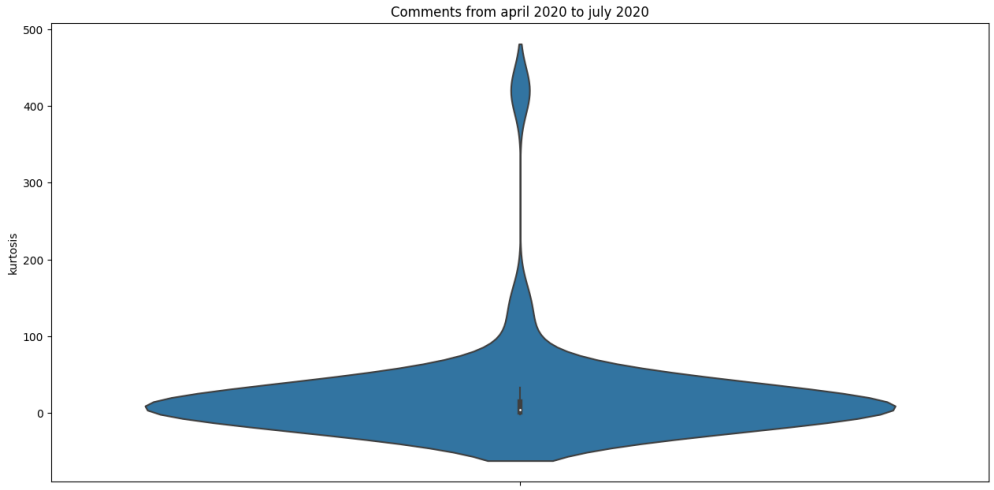

Figura 17: Violin plot dei valori di kurtosis calcolata sulla Toxicity dei commenti per ogni utente

Analogamente a quanto osservato per la skewness, anche in questo caso alla quasi totalità degli utenti è associata una kurtosis maggiore di zero, che vuol dire che la distribuzione della Toxicity è più "appuntita" della distribuzione normale. La maggior parte degli utenti ha una kurtosis minore di 100, mentre è ben visibile la presenza di qualche utente con valori molto elevati.

6.2.3 Violin plot diviso per periodi

Dividiamo nuovamente il dataset dei commenti da aprile a luglio in due dataset distinti: il primo contenente i commenti dal 1° aprile al 25 maggio, data dell'omicidio di George Floyd; il secondo contenente i rimanenti, quindi quelli pubblicati dal 26 maggio al 31 luglio.

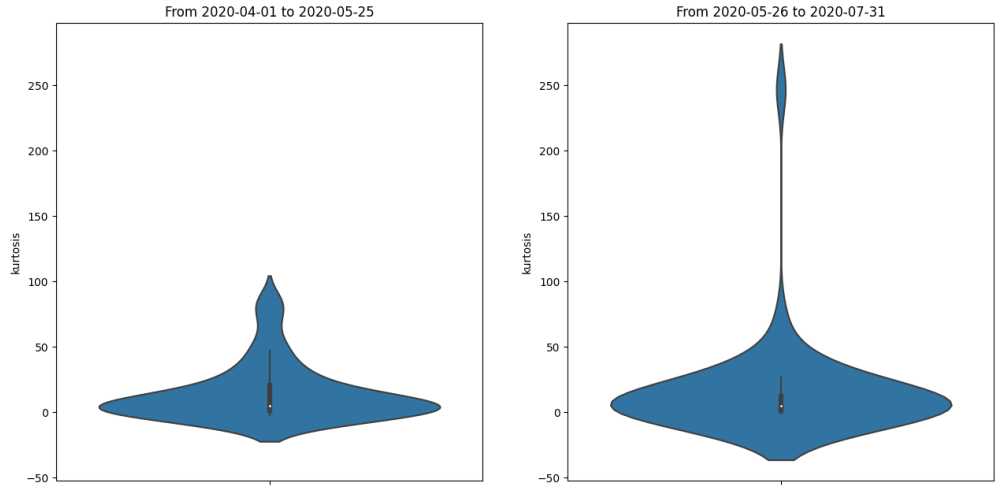

Figura 18: Violin plot dei valori di kurtosis calcolata sulla Toxicity dei commenti per ogni utente prima e dopo l'omicidio di Floyd

Anche in questo caso notiamo la presenza di valori di kurtosis molto alti nel secondo periodo, rappresentativi di distribuzioni di probabilità più "appuntite" in corrispondenza del picco. Tuttavia, fatta eccezione per questi valori molto elevati, i due violin plots si somigliano abbastanza.

6.3 Correlazione con la Toxicity media

Cerchiamo ora di individuare un'eventuale correlazione tra i valori di skewness e kurtosis trovati e i valori di Toxicity media calcolati per ogni utente. Una volta ottenuta, per ogni utente, la tripla {toxicity_media, skewness, kurtosis}, sono stati tracciati due grafici separati: uno per quanto concerne la skewness; l'altro per la kurtosis.

6.3.1 Correlazione tra skewness e Toxicity media

Per studiare la correlazione tra i due indici realizziamo uno scatter plot, avente sulle ascisse i valori di Toxicity media e sulle ordinate i valori di skewness. Ogni punto rappresenta un utente.

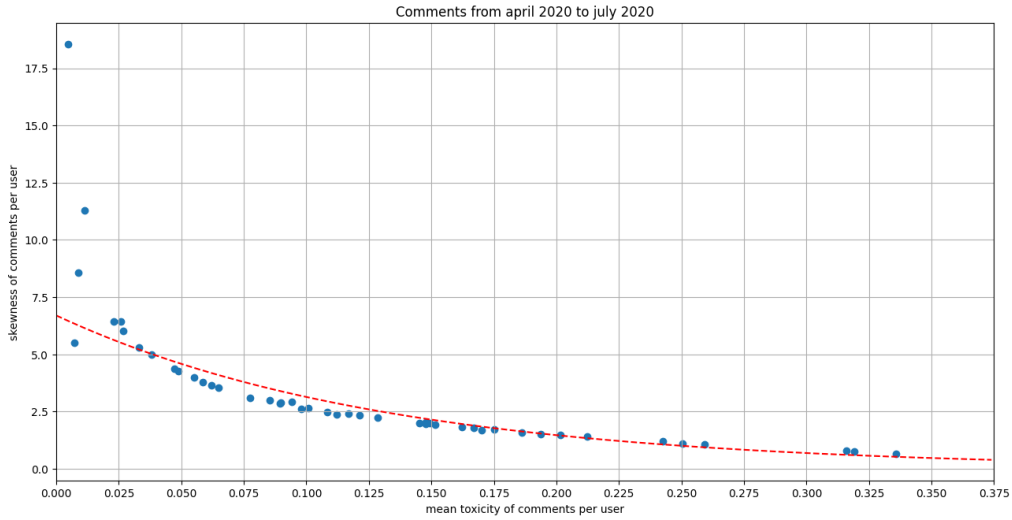

Figura 19: Scatter plot di Toxicity media e skewness, per ogni utente. In rosso trendline esponenziale

Si nota una marcata correlazione tra i due valori, vicina ad essere di tipo esponenziale (la trendline esponenziale è rappresentata tratteggiata in rosso). Dal grafico è evidente come gli utenti caratterizzati da una più alta Toxicity media dei propri commenti, tendono ad avere una più bassa skewness della distribuzione della Toxicity dei vari commenti. Dunque chi ha commentato, in media, in maniera più tossica, ha una distribuzione meno asimmetrica; chi ha commentato, in media, in maniera meno tossica, ha una distribuzione più asimmetrica. Questo ci torna se ci ricordiamo di come è distribuita la Toxicity (Figura 1): essendo la maggior parte dei commenti caratterizzati da una Toxicity bassa, è ipotizzabile che gli utenti con una distribuzione più asimmetrica siano quelli abituati a commentare in maniera inoffensiva; quelli con una distribuzione meno asimmetrica sono quelli che, pur commentando spesso in maniera tranquilla, si lasciano andare qualche volta a commenti più irriverenti.

Ci chiediamo ora se siano presenti differenze apprezzabili nel tipo di correlazione tra questi due indici tra il periodo precedente e quello successivo all'omicidio di Floyd. Segue quindi la stessa rappresentazione, divisa per periodo.

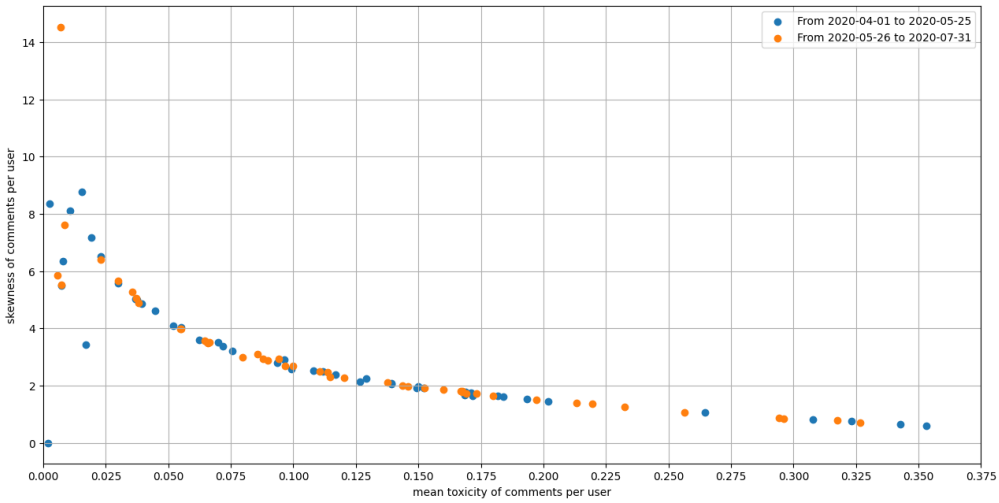

Figura 20: Scatter plot di Toxicity media e skewness, per ogni utente, prima e dopo l'omicidio di George Floyd

Non sono apprezzabili differenze significative. Il che ci induce a pensare che la correlazione tra skewness e media, calcolate sulla Toxicity, persista a prescindere dal periodo e quindi del set di commenti preso in considerazione.

6.3.2 Correlazione tra kurtosis e Toxicity media

Anche in questo caso è stato tracciato uno scatter plot, rappresentando sulle ascisse i valori di Toxicity media e sulle ordinate i valori di kurtosis. Ogni punto rappresenta un utente.

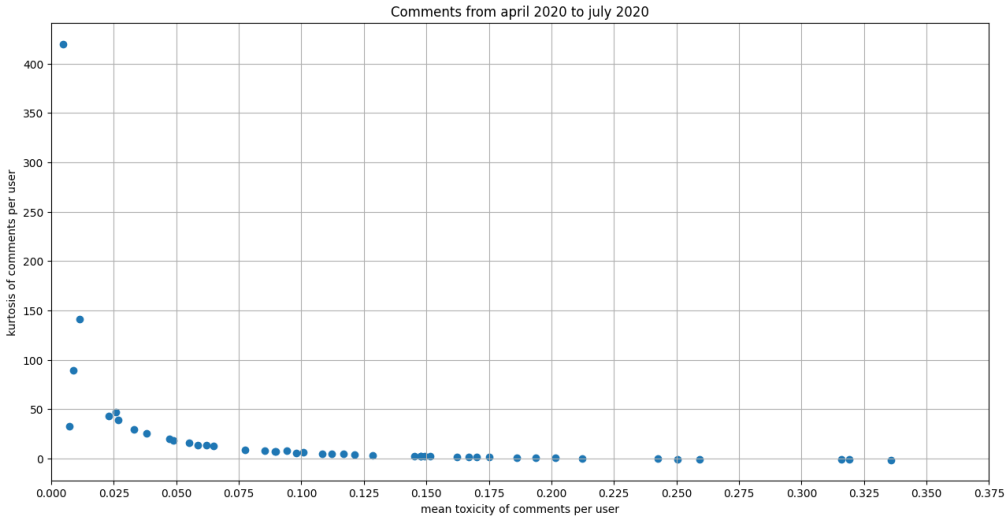

Figura 21: Scatter plot di Toxicity media e kurtosis, per ogni utente

Anche in questo caso è rilevabile una correlazione tra i due indici, simile a quella osservata precedentemente (Sezione 6.3.1). La curva appare più "schiacciata" a causa di un utente con un valore molto alto di kurtosis. Si può comunque notare come gli utenti che hanno registrato una più alta Toxicity media dei propri commenti, tendono ad avere una più bassa kurtosis della distribuzione della Toxicity dei vari commenti. Dunque chi ha commentato, in media, in maniera più tossica, ha una distribuzione meno "appuntita"; chi ha commentato, in media, in maniera meno tossica, ha una distribuzione più "appuntita".

Analogamente a quanto fatto nell'analisi precedente, si è provato ad osservare eventuali differenze nella correlazione tra questi due indici tra il periodo precedente e quello successivo all'omicidio di Floyd. Segue quindi la stessa rappresentazione, divisa per periodo.

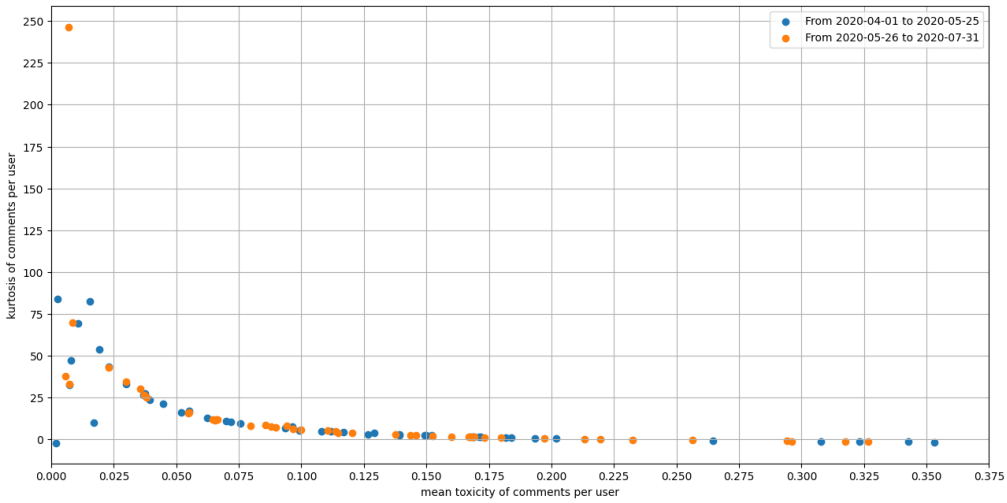

Figura 22: Scatter plot di Toxicity media e kurtosis, per ogni utente, prima e dopo l'omicidio di George Floyd

Come nel caso della skewness, non sono visibili differenze significative. Il che ci suggerisce che la correlazione tra kurtosis e media, della distribuzione di Toxicity, persiste a prescindere dal periodo e quindi del set di commenti preso in considerazione.

6.4 Confronto con la distribuzione di probabilità dello Score

Andiamo ora a confrontare i valori di skewness e kurtosis precedentemente calcolati (sulla Toxicity) con i valori di skewness e kurtosis che andiamo a calcolare sullo Score.

6.4.1 Confronto tra skewness della Toxicity e skewness dello Score

Tracciamo uno scatter plot, rappresentando sulle ascisse i valori di skewness calcolati sulla Toxicity; sulle ordinate i valori di skewness calcolati sullo Score.

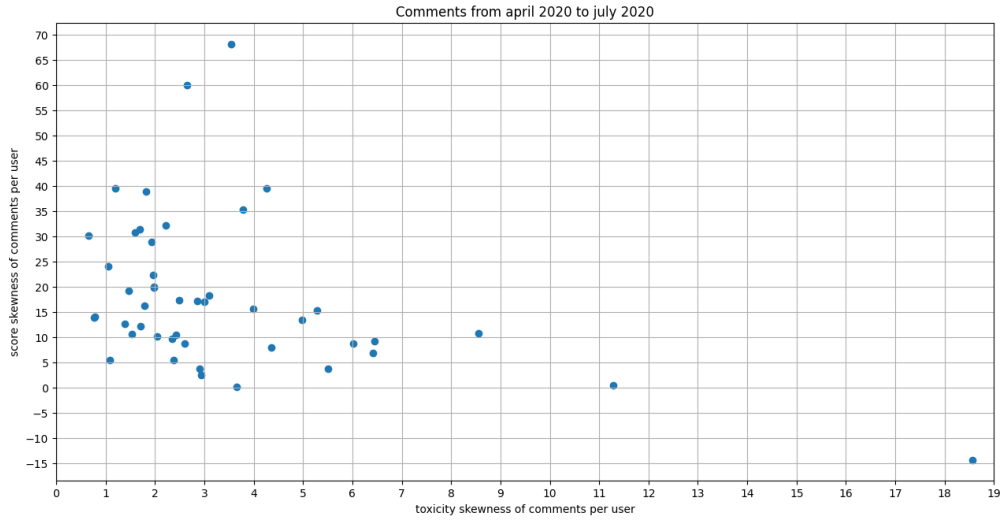

Figura 23: Scatter plot di skewness calcolata sui valori di Toxicity e skewness calcolata sui valori di score

Anche se non appare evidente, si intravede una tendenza: chi ha una skewness sulla Toxicity più elevata tende ad avere una skewness sullo Score più bassa, e viceversa. Ricordandoci di cosa rappresenta la skewness, ciò equivale a dire che chi ha una distribuzione di Toxicity più asimmetrica verso sinistra (quindi è solito commentare in maniera inoffensiva) tende ad avere una distribuzione di Score più asimmetrica verso destra (totalizza spesso score alti). Questo costituisce una conferma del risultato osservato nella Sezione 5.2.

Rappresentiamo ora lo stesso grafico suddiviso per i due sottoperiodi già precedentemente utilizzati, quello antecedente e quello successivo all'omicidio di Floyd.

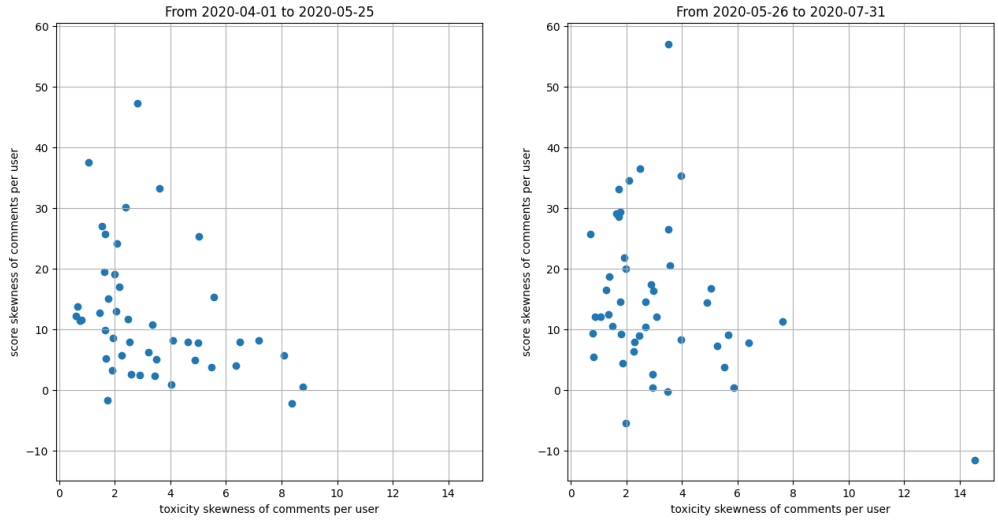

Figura 24: Scatter plot di skewness calcolata sui valori di Toxicity e skewness calcolata sui valori di Score, prima e dopo l'omicidio di George Floyd

I due grafici non presentano molte differenze. Si può notare, ad esempio, che nel secondo periodo i valori di skewness calcolati sulla Toxicity, fatta eccezione per un outlier, siano più schiacciati verso lo zero, dunque in media più bassi. Questo, per il concetto di skewness, ci dice che l'asimmetria della distribuzione di Toxicity verso i valori più bassi è ancora presente ma meno accentuata. Ci fornisce quindi, ricordandoci anche della correlazione tra skewness della Toxicity e Toxicity media trovata nella Sezione 6.3.1, una prova della crescita complessiva di Toxicity nel secondo periodo.

6.4.2 Confronto tra kurtosis della Toxicity e kurtosis dello Score

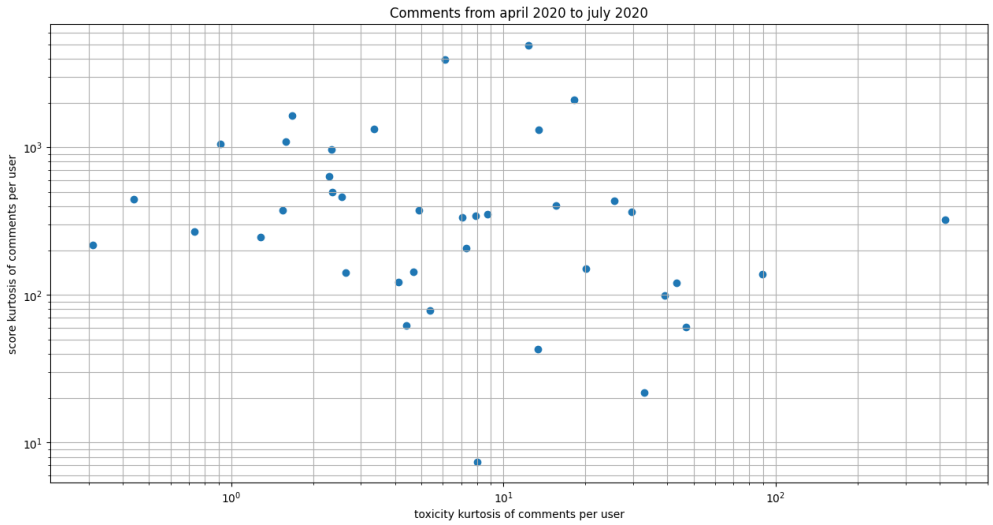

Figura 25: Scatter plot di kurtosis calcolata sui valori di Toxicity e kurtosis calcolata sui valori di Score

In questo caso, sfruttando l'assenza di valori di kurtosis negativi, si è fatto uso della scala semi-logaritmica, che semplifica di molto la lettura del grafico. Anche così però, rispetto allo studio precedente effettuato sulla skewness, appare meno chiara una correlazione tra i due indici. Si nota una tendenza, seppur meno evidente, analoga alla precedente: a valori più alti di kurtosis della Toxicity corrispondono valori più bassi di kurtosis dello Score, e viceversa. Questo significa che chi ha una distribuzione di Toxicity più "appuntita" tende ad avere una distribuzione di Score più "piatta".

Rappresentiamo anche in questo caso lo stesso grafico suddiviso per i due periodi, precedente e successivo all'omicidio di Floyd.

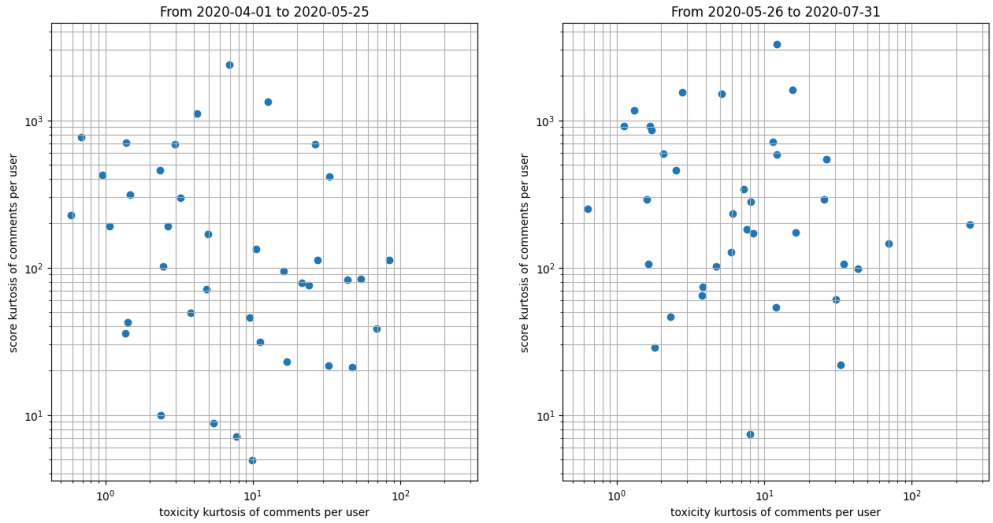

Figura 26: Scatter plot di kurtosis calcolata sui valori di Toxicity e kurtosis calcolata sui valori di score, prima e dopo l'omicidio di George Floyd

Analogamente a quanto osservato eseguendo la stessa analisi sulla skewness, non si riesce a notare una differenza nella correlazione tra le due variabili. Si può invece fare qualche osservazione sui valori della kurtosis della Toxicity. Nel primo periodo il numero di utenti con tale indice compreso tra 1 e 10 è circa uguale a quello di utenti con il suddetto indice compreso tra 10 e 100; nel secondo periodo si ha invece una netta predominanza degli utenti con kurtosis della Toxicity compresa tra 1 e 10. Questo "appiattimento" dei valori verso il basso testimonia la presenza di distribuzioni di Toxicity più "piatte", e dunque, per quanto trovato nella sottosezione 6.3.2, una generale crescita di Toxicity.

7 Variazione della Toxicity media per utente

Nella pagina che segue sono stati riportati, per ogni utente, i valori di Toxicity media, oltre al numero di commenti, nei periodi antecedente e successivo all'omicidio di Floyd. Viene espressa inoltre, nelle ultime due colonne, la differenza di Toxicity tra i due periodi, in valore assoluto e in valore percentuale.

Si nota come per 24 utenti su 46 la Toxicity media abbia subito un incremento superiore al 10%, per 13 utenti un incremento o decremento in modulo minore del 10%, e solo per 9 utenti il decremento sia stato maggiore del 10%.

Gli utenti che hanno registrato aumenti di Toxicity molto alti in percentuale sono spesso caratterizzati dall'esiguità, prima dell'omicidio, sia del valore di Toxicity media che soprattutto del numero di commenti. È per questo che nell'analisi svolta sulla Toxicity media (Sezione 4.3), senza suddivisione per utente, non abbiamo riscontrato una grande variazione.

Utente	Comm. prima	Tox. media prima	Comm. dopo	Tox. media dopo	Diff.	Diff. in %
kiwasabi	2	0.0019	656	0.1459	0.1440	7777.18
lllllbbbbb	15	0.0171	78	0.1146	0.0975	569.04
lucidludic	192	0.0078	350	0.0357	0.0279	357.62
sadirichardss	362	0.0024	407	0.0070	0.0046	194.53
shad0wtig3r	406	0.1686	728	0.2962	0.1277	75.72
coldminnesotan	286	0.0191	674	0.0300	0.0110	57.60
aquarain	1675	0.0369	3377	0.0548	0.0179	48.63
oispa	2442	0.0376	3016	0.0551	0.0175	46.55
semper_veritatem	2027	0.0447	2220	0.0644	0.0197	43.99
richardeid	156	0.1081	1353	0.1525	0.0444	41.10
ssjb788	239	0.1293	227	0.1675	0.0382	29.57
nothinbuttherain	342	0.0755	671	0.0967	0.0212	28.11
gordonv	3033	0.0300	1821	0.0381	0.0080	26.73
eremite00	738	0.0520	741	0.0654	0.0134	25.71
segvcore	398	0.0717	1220	0.0898	0.0181	25.21
moak0	863	0.1391	1856	0.1732	0.0341	24.48
bout_that_action	1052	0.0993	1703	0.1204	0.0211	21.24
tommygunz007	3382	0.1391	2564	0.1676	0.0286	20.53
rtechie1	509	0.0551	1008	0.0658	0.0107	19.42
ral365	3028	0.0937	2260	0.1106	0.0169	18.02
xxoites	3408	0.1169	4232	0.1377	0.0208	17.82
whatisliquidity	1203	0.1839	1854	0.2134	0.0295	16.04
sand313man	192	0.1710	1228	0.1973	0.0264	15.43
theodore-nyc	114	0.0699	458	0.0796	0.0097	13.91
dylang92	1328	0.2020	1802	0.2196	0.0176	8.71
yaosio	2403	0.0624	3399	0.0665	0.0041	6.54
tethercat	875	0.0229	1065	0.0231	0.0002	0.74
denverprotestsdaily	135	0.0072	136	0.0072	0.0000	-0.65
ilikerelish	848	0.1717	1252	0.1693	-0.0024	-1.40
xarnzul	1031	0.3231	2477	0.3177	-0.0054	-1.67
lovemerightsuhodobak	40	0.0965	41	0.0942	-0.0023	-2.42
luv_tummy	1685	0.1499	156	0.1436	-0.0063	-4.17
jimjomjimmy	1010	0.1690	257	0.1600	-0.0090	-5.34
casualphilosopher1	3151	0.0394	5272	0.0372	-0.0023	-5.74
twss416	1769	0.1935	2024	0.1800	-0.0135	-6.99
outoftowner2	488	0.3532	907	0.3267	-0.0265	-7.50
markmywords1347	3760	0.1817	2894	0.1670	-0.0147	-8.07
icomeforthereaper	1980	0.2645	4385	0.2326	-0.0320	-12.09
murrlogic	3583	0.3426	4415	0.2943	-0.0484	-14.11
big_red_meatstick	148	0.3077	2363	0.2564	-0.0513	-16.68
adzling	317	0.1267	734	0.0997	-0.0269	-21.27
dr_gonzo	321	0.1118	3626	0.0880	-0.0238	-21.30
newsspotter	141	0.0107	758	0.0084	-0.0023	-21.77
smolsmoller	66	0.1494	232	0.1137	-0.0357	-23.90
vanulovesyou	445	0.1521	497	0.0856	-0.0665	-43.70
dreamolli	220	0.0155	169	0.0056	-0.0098	-63.55

8 Conclusioni

Sono state svolte analisi di diverso tipo, volte a identificare le variazioni nel comportamento degli utenti di Reddit in seguito all'omicidio di George Floyd. I risultati ottenuti hanno rivelato una serie di tendenze e correlazioni significative:

- l'aumento dei commenti "tossici" nei giorni successivi all'omicidio e alle proteste (4.4) è la prova del notevole impatto avuto dall'evento sulla piattaforma;
- la correlazione negativa tra la Toxicity media per utente e il suo Score medio (5.2) dimostra che gli utenti sono più propensi ad apprezzare commenti privi di tossicità;
- la correlazione negativa tra la Toxicity media per utente e le misure di skewness (6.3.1) e kurtosis (6.3.2) delle distribuzioni di probabilità della Toxicity dei suoi commenti, ci ha suggerito l'identificazione di due tipologie di utenti: quelli che hanno una Toxicity media dei loro commenti molto bassa, che tendono ad avere una distribuzione decisamente asimmetrica e "appuntita"; quelli che hanno una Toxicity media dei commenti un po' più alta, che tendono ad averla più simmetrica e meno "appuntita". I primi sono gli utenti che sono soliti commentare in maniera pacifica; i secondi commentano comunque solitamente con toni non offensivi, ma si lasciano andare di frequente a commenti più irrispettosi;
- il fatto che un maggior numero di utenti abbia registrato un aumento della Toxicity media rispetto a coloro la cui tossicità è diminuita (7), è un'ulteriore conferma di come l'avvenimento abbia contribuito a "incattivire", almeno nel complesso, gli utenti su Reddit, e quindi, di riflesso, nella società.

Per quanto riguarda le future ricerche a partire da questo lavoro, ci sono diverse direzioni promettenti. Alcuni possibili studi futuri sono:

- ripetere le analisi cambiando l'evento oggetto di studio, dunque esaminare come i valori di Toxicity dei commenti, o più in generale i comportamenti degli utenti, cambino nel tempo in risposta ad eventi di attualità di diverso tipo;
- a partire da un aumento osservato di commenti "tossici", identificare la causa scatenante (l'approccio opposto a quello adottato in questo studio);
- sviluppare algoritmi di moderazione avanzati per identificare più velocemente gli utenti "tossici", o per prevenire focolai di tossicità in conseguenza ad eventi come quello studiato;
- esplorare altre piattaforme social per valutare se i comportamenti degli utenti rispetto allo stesso evento siano differenti su diversi social networks.

9 Bibliografia e Sitografia

[1] *Reddit User Base & Growth Statistics: How Many People Use Reddit? (Sep 2023)*. URL: https://www.bankmycell.com/blog/number-of-reddit-users/.

[2] James Grimmelmann. «The Virtues of Moderation». In: *Yale JL & Tech.* 17 (2015), p. 42.

[3] Camille Francois, Vladimir Barash e John Kelly. «Measuring Coordinated versus Spontaneous Activity in Online Social Movements». In: *New Media & Society* (2021).

[4] Shagun Jhaver et al. «Evaluating the Effectiveness of Deplatforming as a Moderation Strategy on Twitter». In: *Proceedings of the ACM on Human-Computer Interaction* 5.CSCW2 (2021), pp. 1–30.

[5] Milo Z Trujillo et al. «When the Echo Chamber Shatters: Examining the Use of Community-specific Language post-Subreddit Ban». In: *arXiv preprint arXiv:2106.16207* (2021).

[6] Hussam Habib et al. «Are Proactive Interventions for Reddit Communities Feasible?» In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 16. 2022, pp. 264–274.

[7] Mohit Singhal et al. «SoK: Content Moderation in Social Media, from Guidelines to Enforcement, and Research to Practice». In: *2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P)*. IEEE. 2023, pp. 868–895.

[8] Amaury Trujillo e Stefano Cresci. «Make Reddit Great Again: Assessing Community Effects of Moderation Interventions on r/The_Donald». In: *Proceedings of the ACM on Human-Computer Interaction* 6.CSCW2 (2022), pp. 1–28.

[9] *Reddit comments/submissions 2005-06 to 2022-12*. URL: https://academictorrents.com/details/7c0645c94321311bb05bd879ddee4d0eba08aaee.

[10] *combine_folder_multiprocess.py*. URL: https://github.com/Watchful1/PushshiftDumps/blob/master/scripts/combine_folder_multiprocess.py.