

## CASO ASTRAZENECA: EU VS GB

### ABSTRACT

In questo progetto si è valutato il sentimento e l'opinione pubblica sul vaccino AstraZeneca e sull'Unione Europea durante il mese di Marzo 2021, collezionando più di 200 000 messaggi da Twitter – il più grande servizio online di microblogging al mondo.

Si è eseguita una sentiment analysis utilizzando la libreria Vader sul software Python e attraverso tale tecnica, è stato possibile quantificare il sentimento pubblico quotidiano sia nel Regno Unito, sia nei paesi dell'Unione Europea. In particolare, ci si è concentrati sul giorno 15 Marzo, data in cui l'Unione Europea ha sospeso le somministrazioni del vaccino AstraZeneca. Infine, si è svolta anche una emotion analysis tramite la libreria NRCLEX in modo da ottenere insight più specifici e granulari

### INTRODUZIONE

Twitter, la piattaforma di microblogging, con oltre 187 milioni di utenti attivi giornalieri, viene considerata come uno strumento robusto per comprendere meglio la percezione pubblica su larga scala del vaccino COVID-19. Con quasi 400 milioni di menzioni, #COVID19 è stato l'hashtag più usato su Twitter nel 2020. I social media sono diventati sempre più riconosciuti per la loro estrema diffusione di informazioni (accurate o meno) e la dispersione del sentimento che attraversa rapidamente i confini geografici e sociali. L'analisi del testo dei social media può informare in tempo reale i cambiamenti e l'evoluzione degli atteggiamenti a livello di popolazione. Come è evidente con l'ascesa dell'"infodemia" durante la pandemia di COVID-19, Twitter è diventato una fonte di dati particolarmente utile per la salute pubblica e la ricerca sanitaria ed è stato ripetutamente utilizzato per studiare il sentimento pubblico e comprendere le tendenze durante la pandemia di COVID-19.

Esaminando i sentimenti nei confronti del vaccino contro l'influenza A H1N1 nel 2009, è emerso che i tassi di vaccinazione previsti sulla base del sentiment di Twitter erano simili ai tassi di vaccinazione stimati dalle indagini telefoniche tradizionali utilizzate dai Centers for Disease Control and Prevention (CDC). Uno studio precedente ha osservato che l'esposizione alle informazioni su Twitter può spiegare le differenze nell'adozione del vaccino contro il papillomavirus umano (HPV) che non sono rappresentate da fattori socioeconomici come l'istruzione, l'assicurazione o la povertà.

Per tali ragioni si è deciso in questo progetto di studiare l'andamento del sentimento e delle emozioni nel corso del mese di marzo 2021 con particolare attenzione al lunedì 15, data di sospensione del vaccino AstraZeneca in Unione Europea. Il focus non si limita solo sul giudizio del vaccino in sé, ma anche sulle relazioni e sulle implicazioni che intercorrono tra Regno Unito e Unione Europea come soggetti geopolitici.

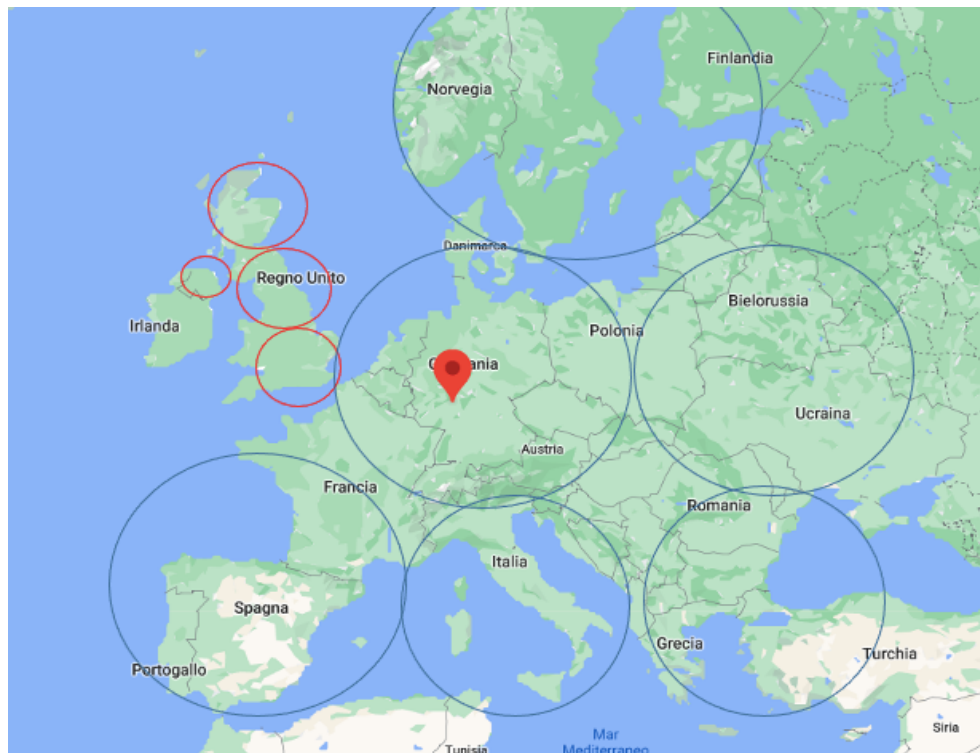
### DATA ACQUISITION

Sono stati estratti dati dalla piattaforma social Twitter durante il mese di Marzo 2021 attraverso la libreria *bsi\_sentiment.twitter* per Python.

Tale libreria, a differenza di Tweepy, (libreria che utilizza l'API ufficiale del social network) permette la raccolta di dati anche per periodi superiori alla settimana antecedente (questo perché non utilizza un API ma è uno strumento di web scraping). Purtroppo, tale vantaggio è compensato dal fatto di dover selezionare "a mano" le zone geografiche di interesse tramite la definizione di una città di partenza e un raggio entro il quale estrarre i tweet.

Nello specifico, si è deciso di filtrare i dati per tweet di lingua inglese e per località geografica. Infatti, dato l'obiettivo di confrontare Unione europea e Regno Unito sono state selezionate le zone evidenziate nella mappa sottostante ed escludendo successivamente

eventuali duplicati. Per le due aree oggetto di studio sono stati estratti dati in due tranche con le due diverse parole chiave di ricerca: AstraZeneca ed EU.



## DATA PREPROCESSING

La libreria *bsi\_sentiment.twitter* fornisce svariati metadati relativi a un tweet. Tuttavia, per questo progetto si necessitava solo del testo e dell'ora del tweet, quindi il resto delle colonne sono state rimosse. Il passo successivo è stato convertire il testo in minuscolo che impedisce al modello di considerare ad esempio "They" e "they" come diversi. Poi, le stopwords— come "a", "her", "his" - e i segni di punteggiatura sono stati rimossi usando il corpus di parole di arresto di NLTK. In seguito, il testo è stato tokenizzato - che è il processo di divisione del testo in frasi e frasi in parole. Questo è stato seguito dalla lemmatizzazione in cui le parole di terza persona sono convertite in prima persona e i tempi futuri sono convertiti in tempo presente. Grazie a questa preelaborazione del testo, ci si assicura che l'algoritmo di sentiment e di emotion analysis funzioni correttamente.

## SENTIMENT & EMOTION ANALYSIS

E' stata eseguita l'analisi del sentimento sul set di dati utilizzando *VADER* che è uno strumento di analisi del sentimento basato su lessico e regole specificamente adatti allo studio dei testi nei social media. Vader fornisce un punteggio composto che assegna un punteggio di polarità al testo complessivo. Questo punteggio varia da -1 a +1. Se il punteggio è compreso tra -0,05 e +0,05, allora è classificato come neutro. Se è maggiore di 0,05, allora è considerato positivo e se è inferiore a -0,05, allora è considerato negativo. Infine, si è utilizzato la libreria NRClex per etichettare le parole all'interno di ogni tweet con i corrispondenti affetti emotivi (cioè, la ruota delle emozioni di Plutchik che include rabbia, anticipazione, paura, disgusto, gioia, tristezza, sorpresa e fiducia) basati sul lessico degli affetti del National Research Council Canada (NRC).

## RISULTATI

Come precedentemente illustrato nell'abstract, si è svolta l'estrazione e la successiva esecuzione della sentiment<sup>1</sup> e della emotion<sup>2</sup> analysis prendendo come parole chiave **AstraZeneca** ed **EU** per le due diverse regione geografiche in questione: Unione Europea e Regno Unito.

Considerando il tag AstraZeneca si sono potuti constatare valori di sentiment neutri e non significativi per entrambe le zone in tutto il mese di marzo ad eccezione del 15 e del 16 (il 15 Marzo l'Unione Europea sospende la somministrazione del vaccino AstraZeneca) in cui si è verificata una netta differenza di score: valori positivi per il Regno Unito ed estremamente negativi per l'UE.

Per quanto riguarda il tag EU come nel caso precedente i valori non sono significativi nel mese di marzo a parte i giorni dal 13 al 15 Marzo, in cui nel vecchio continente si assiste ad un incremento notevole di punteggio verso il segno negativo, raggiungendo il picco di -0,44 il 15 Marzo. In Regno Unito non si registra un trend costante e nemmeno un ciclo ma i valori oscillano per tutto il mese su valori neutri.

Per l'emotion analysis si è deciso di suddividere i dati sempre per regioni geografiche ma ci si è concentrati su tre differenti periodi temporali: il giorno significativo del 15 marzo, le due settimane precedenti e le due settimane successive.

Considerando il tag AstraZeneca sia nel periodo precedente al 15 sia successivamente, non si notano differenze significative di emotion tra le due aree geografiche tranne per un maggior punteggio di *positive* nel Regno Unito. In data 15 marzo le differenze sono nettissime con punteggi molto alti di *fear disgust* e *anger* in Unione Europea, mentre tali valori in Gran Bretagna si assestano molto bassi e l'emotion *positive* è sette volte quella rilevata in UE.

Per quanto riguarda il tag EU, come nel caso appena discusso, sia nel periodo precedente che successivo al 15 marzo non ci sono grosse differenze tra le due aree, tranne che per un lieve valore medio più alto dell'emotion *negative* in Regno Unito in entrambi i periodi. In data 15 marzo cambia lo scenario. Infatti, in Europa esplodono i valori per *disgust anger sadness* e *negative* e si dimezza il valore di *positive*, mentre in Regno Unito i valori rimangono simili rispetto agli altri giorni del mese e il valore di *positive* è in tale data il doppio di quello in UE.

## CONCLUSIONI E SVILUPPI FUTURI

Questo studio si è prefissato di analizzare i sentimenti verso il vaccino AstraZeneca durante il mese di marzo 2021 con particolare attenzione a lunedì 15, data in cui l'Unione Europea sospende la somministrazione del vaccino per alcune morti sospette dopo l'assunzione, mentre nel Regno Unito si è proceduto senza problemi al suo utilizzo.

Per tale regione si è deciso di integrare anche la sentiment analysis rispetto al tag Unione Europea e in particolare come gli utenti hanno reagito alla decisione politica del blocco di AstraZeneca (prodotto da un'azienda anglo-svedese) e se tale reazione fosse dissimile tra le due aree geografiche. Dalle analisi già commentate nella sezione *risultati*, le differenze tra gli utenti del Regno Unito ed Unione Europea non sono significative nei trenta giorni in esame, tranne per la data chiave del 15 in cui i risultati e valori sono opposti. Infatti, si osserva una netta contrapposizione sul tag AstraZeneca per cui in Europa si raggiunge il valore più negativo di tutto il mese (-0,51) mentre nel Regno Unito il valore è decisamente positivo (0,29). Tali risultati si riscontrano anche nella emotion analysis in cui sempre rispetto al 15 marzo i sentimenti positivi sono dominanti in Regno Unito mentre dominano quelli negativi in UE.

Il risultato più interessante ed emblematico riguarda lo studio del sentimento sul tag EU. Infatti, nonostante le metriche e valori siano simili per quanto già detto nel caso AstraZeneca si è osservato un sentimento nel Regno Unito che ha sempre oscillato tra il neutro e il negativo.

---

<sup>1</sup> [https://public.tableau.com/profile/alessandro.pontini#!/vizhome/FINALE\\_1/AstrazenecainEuropa](https://public.tableau.com/profile/alessandro.pontini#!/vizhome/FINALE_1/AstrazenecainEuropa)

<sup>2</sup> <https://public.tableau.com/profile/alessandro.pontini#!/vizhome/Istagrammi/EmotionAnalysis?publish=yes>

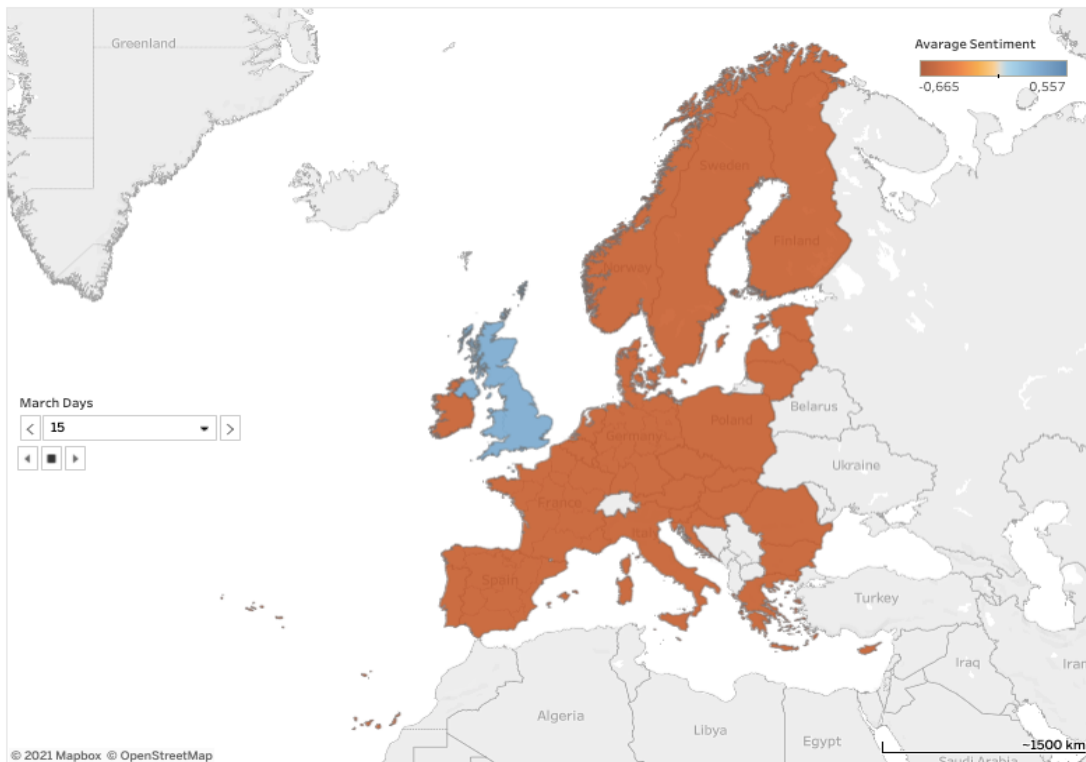
Una possibile spiegazione di ciò potrebbe essere da imputare all'uso strumentale del successo della campagna vaccinale inglese rispetto a quello europeo da parte di politici e sostenitori della Brexit. Lo scontro geo-politico avvenuto tra l'uscente Regno Unito e l'Unione Europea, svoltosi per molti mesi sui tavoli di Bruxelles, si sarebbe spostato sui social utilizzati come strumento di propaganda dai votanti del "leave".

Per quanto riguarda le limitazioni dello studio è innegabile che l'utilizzo di tweet di sola lingua inglese abbia portato un bias, soprattutto sui dati raccolti in Unione Europea. Utilizzare e parlare l'inglese è una buona proxy di istruzione e cultura elevata e tale popolazione è solitamente considerata meno populista, meno ingannabile e plasmabile da propaganda e fake news e in particolare più vicina al progetto europeista. Per tale ragione con software e strutture di database più potenti si potrebbero estrarre e analizzare tutti i tweet dei paesi europei con le rispettive lingue ufficiali.

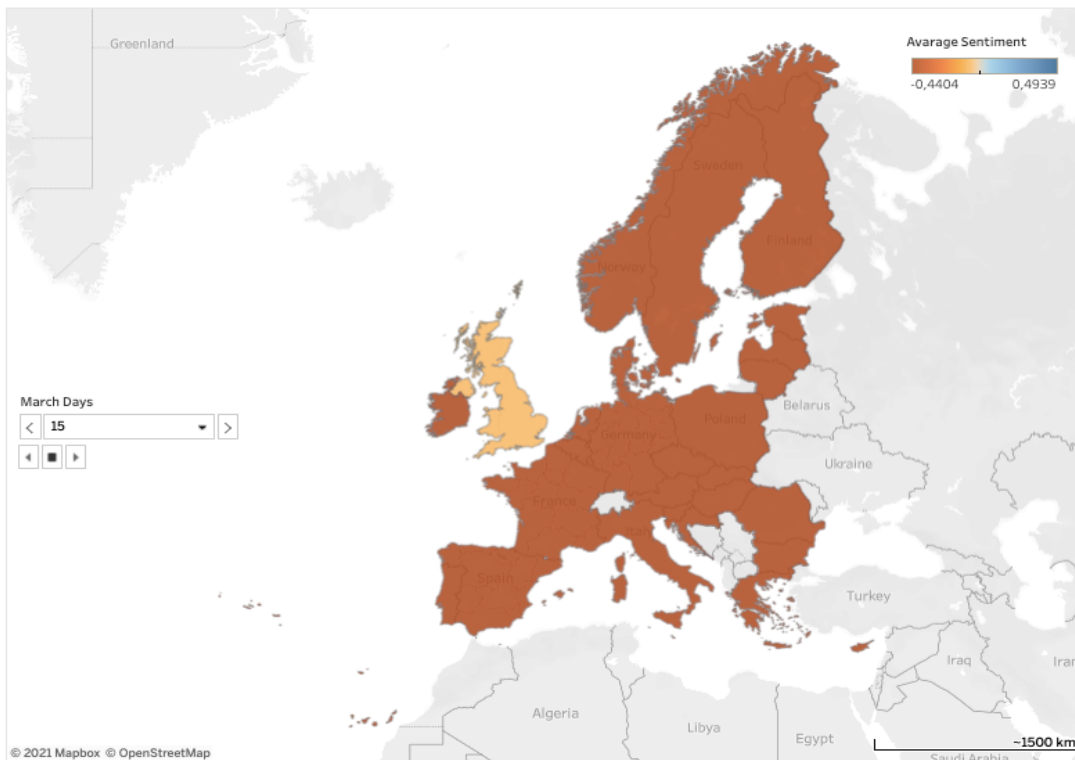
Inoltre, è innegabile che aver utilizzato solo Twitter come fonte dati porti con sé del bias. In letteratura infatti, solitamente si associa Facebook come social di aggregazione per teorie complottiste, gruppi di destra populistici e dove le fake news nascono e si sviluppano con ritmo ed efficacia maggiore. Purtroppo, non è possibile estrarre dati da Facebook tramite API o librerie, dato che l'azienda richiede contratti di partnership con università ed aziende private attentamente selezionate.

## APPENDICE: VISUALIZZAZIONI

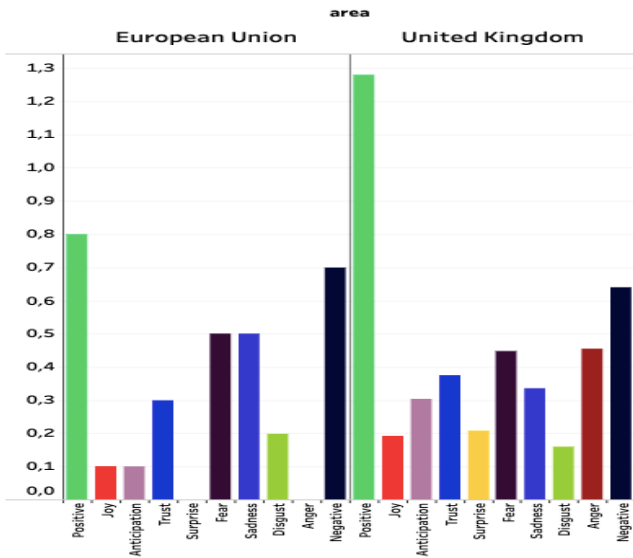
#ASTRAZENECA 01-03-2021 / 31-03-2021



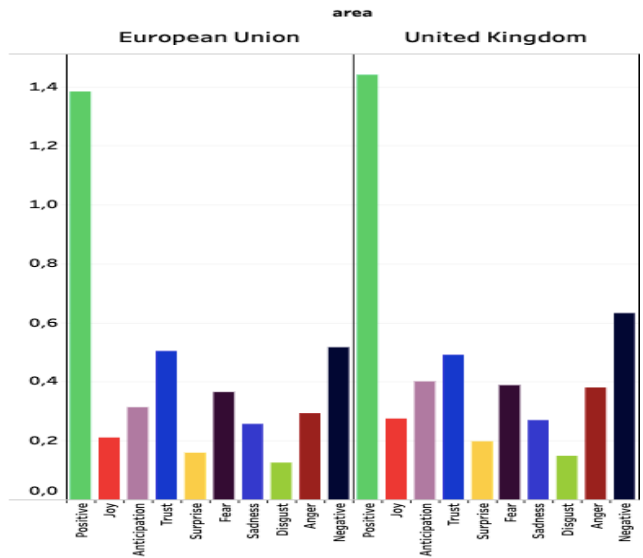
#EU 01-03-2021 / 31-03-2021



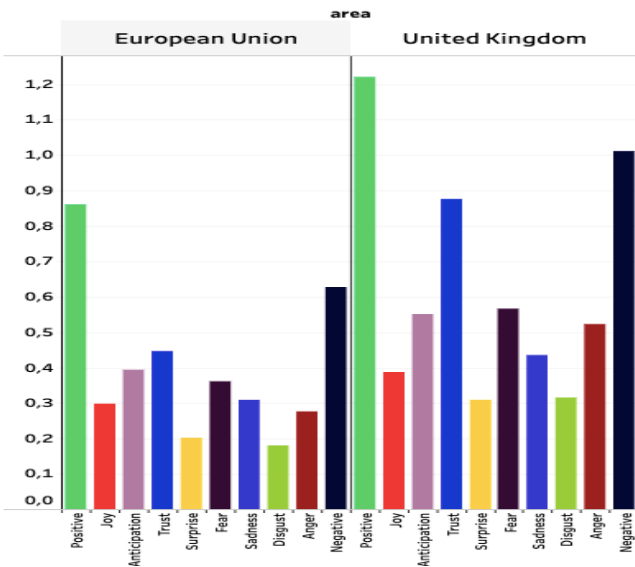
#Astrazeneca before of 15st March



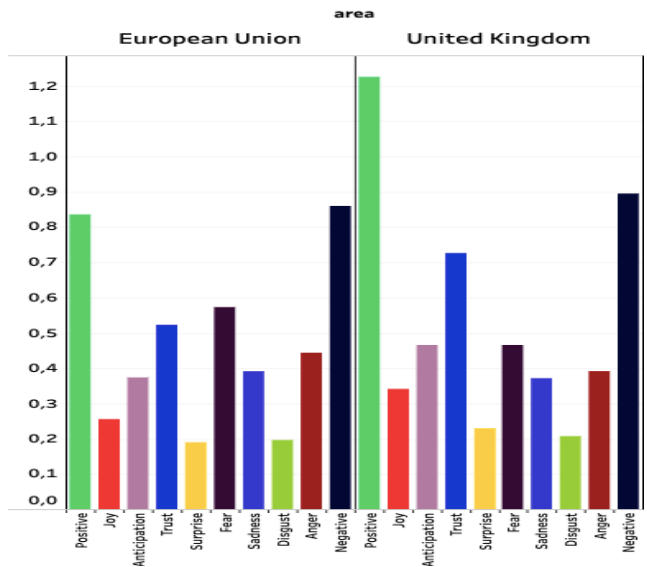
#Astrazeneca after of 15st March



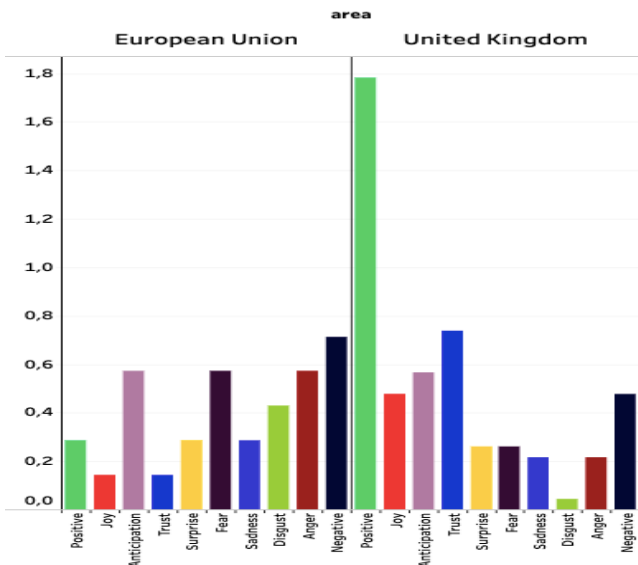
#European Union before of 15st March



#European Union after of 15st March



#Astrazeneca 15st March



#European Union 15st March

