

Università degli Studi di Milano - Bicocca

Corso di Laurea Magistrale di Data Science, a.a. 2020/2021



Le reazioni emotive degli utenti di Twitter alle notizie covid-19 in Italia

Report di Data Visualization, Giugno 2021

Daniele Quattrocchi – matricola 825418

Vittorio Menardo – matricola 812341

Raffaele Moretti – matricola 794537

Indice

1 Introduzione	3
2 Valutazione euristica	4
3 Task e risultati	5
4 Questionario psicometrico	8
5 Conclusioni	11

1 Introduzione

La pandemia causata dal Covid – 19, con le sue inevitabili conseguenze, ha cambiato le abitudini di tutti noi. Abbiamo deciso di analizzare questa situazione, studiando le eventuali correlazioni tra i sentimenti espressi dalle persone attraverso la piattaforma Twitter Italia e l'andamento della pandemia in Italia, durante il periodo 24 Febbraio 2020 – 2 Maggio 2021. Il nostro progetto si basa sulla creazione di due visualizzazioni, prodotte al termine di tutta la fase di raccolta e gestione dei dati necessari alla loro creazione, pubblicate sulla piattaforma Tableau Public.

Prima visualizzazione

La prima visualizzazione del nostro progetto ha l'obiettivo di mostrare l'andamento dei tweet contenenti le parole "paura" e "covid" e l'andamento dei nuovi positivi. Entrambi i dati sono stati presi su base giornaliera e poi messi in media mobile a 7 giorni (settimanale). Per effettuare la visualizzazione è stato quindi utilizzato un doppio asse, dove su un asse viene riportata la misura dei nuovi positivi giornalieri, mentre sull'altro asse vengono riportati i tweet giornalieri contenenti le parole sopra descritte. Sull'asse x viene invece messo il tempo, ovviamente su scala giornaliera. Entrambe le scale degli assi sono lineari. L'arco temporale del nostro lavoro va dal 24 Febbraio 2020, il giorno del primo bollettino nazionale della Protezione Civile riguardo al Covid-19, fino al 2 Maggio 2021.

Link Tableau Public:

<https://public.tableau.com/app/profile/vittorio5755/viz/Seriestorica/Dashboard2>

Seconda visualizzazione

La seconda visualizzazione è stata pensata con l'intento di far vedere quali sono state le emozioni più utilizzate durante la pandemia; per fare questo sono stati presi tutti i tweets contenenti la parola "covid" + una parola chiave. Le parole chiave, e quindi le emozioni, che abbiamo voluto analizzare sono: paura, schifo, vergogna, shock e odio. Per raggiungere questo scopo è stata effettuata una race chart tra queste emozioni, in modo da vedere il loro andamento nel tempo, su base mensile.

Link Tableau Public:

https://public.tableau.com/app/profile/vittorio5755/viz/Racechart_16238603585210/Foglio2

Qualità delle visualizzazioni

Una volta create le visualizzazioni, sono stati fatti dei controlli per verificare se effettivamente le visualizzazioni che abbiamo prodotto comunicavano efficacemente il point della visualizzazione al lettore, con anche l'obiettivo di correggere eventuali errori presenti in esse. Il controllo della qualità è stato effettuato tramite: valutazione euristica, esecuzione dei task e un questionario psicometrico.

2 Valutazione euristica

Per valutare le visualizzazioni è stato chiesto a 5 persone di eseguire una valutazione euristica per evidenziare eventuali criticità. Lo scopo è quello di garantire una migliore usabilità delle visualizzazioni sviluppate. L'approccio alla valutazione delle visualizzazioni consisteva in un'interazione olistica in cui gli utenti sono stati lasciati liberi di esplorare le visualizzazioni create.

2.1 Prima visualizzazione

Gli utenti hanno riscontrato la seguente criticità: Nella serie storica gli assi verticali dovrebbero richiamare a livello di colore le linee rappresentate in modo da rendere più facile l'associazione dei valori all'asse di riferimento. A questo problema non siamo riusciti a trovare una soluzione in quanto il software di riferimento Tableau non consente di colorare gli assi e il nome degli assi.

2.2 Seconda visualizzazione

Gli utenti hanno riscontrato le seguenti criticità:

1 Il race chart aveva dei colori troppo accesi che gli utenti non ritenevano adeguati rispetto alle emozioni che dovevano richiamare.

2 Il race chart su Tableau Public impiegava troppo tempo per passare dall'inizio della serie temporale alla fine a causa della lentezza del server. La granularità era fissata per giorno.

Soluzioni adottate:

1 Il colore dei chart è stato cambiato in una scala di blu che ha reso la visualizzazione più formale e apprezzata.

2 La granularità del grafico è stata cambiata da giorni a mesi per non stancare il lettore e per non far durare la visualizzazione troppo tempo.

3 Task e risultati

In questa sezione si andranno ad elencare i risultati ottenuti attraverso lo user test. Sono stati proposti 3 task eterogenei ad un campione di 12 persone, questo è avvenuto per entrambe le visualizzazioni. Il sondaggio è servito per comprendere i punti deboli delle visualizzazioni e misurare l'efficacia e l'efficienza delle stesse, su una popolazione più o meno esperta. Abbiamo misurato l'efficacia delle visualizzazioni sottoponendo il campione a delle domande e segnando il numero di risposte corrette e quello di risposte sbagliate, rappresentandole su un bar chart a due colori e inserendo al suo interno un margine d'errore dovuto al campionamento. Per misurare l'efficienza delle visualizzazioni sono stati misurati i tempi di risposta degli individui e sono stati confrontati con la media delle tempistiche ottimali dei 3 componenti del gruppo. Il tutto è stato rappresentato in un violin plot con all'interno un box plot al quale è stato associato il valore della mediana e del range interquartile.

3.1 Prima visualizzazione

Sono stati proposti i seguenti task:

- 1) Indica approssimativamente qual è stato il valore massimo di nuovi positivi e in che intervallo di tempo si è verificato
- 2) Indica approssimativamente qual è stato il valore massimo di tweets sulla parola paura e in che intervallo di tempo si è verificato
- 3) Indica approssimativamente il periodo temporale in cui l'andamento di tendenza dei nuovi positivi e dei tweets in questione è stato opposto

Di seguito i risultati:

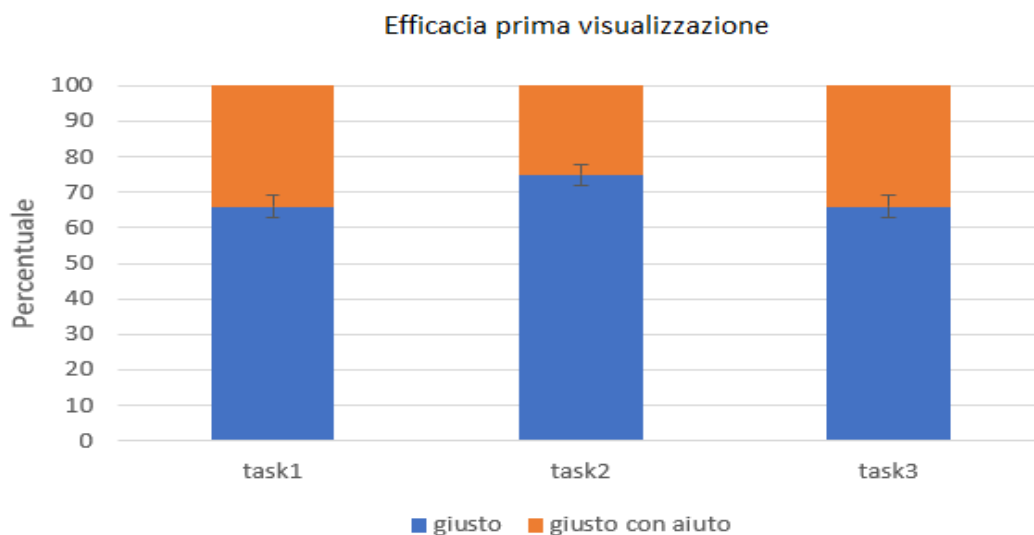


Figura 1: Bar chart errori task

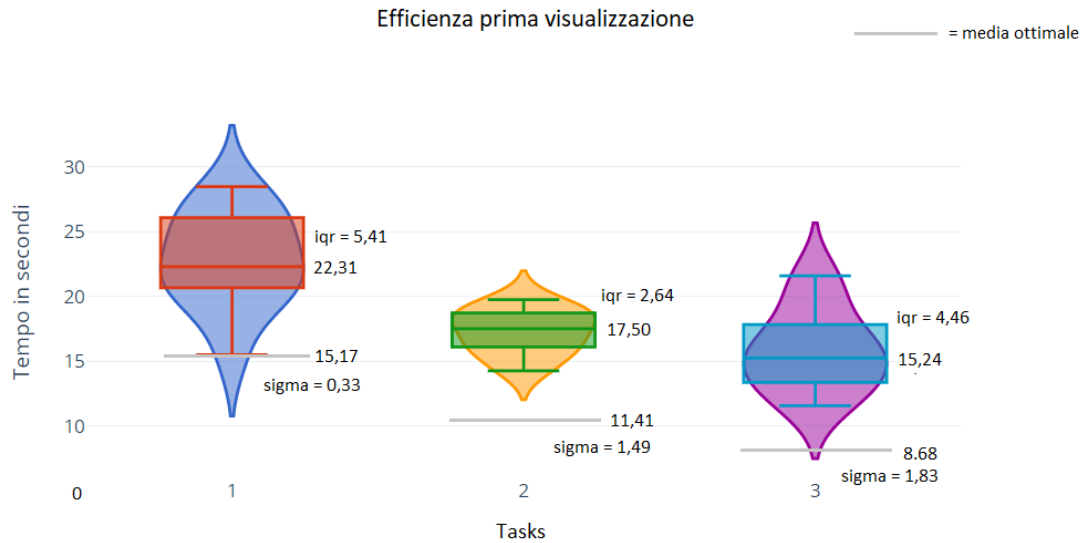


Figura 2: Violin plot tempi di esecuzione

Inizialmente gli errori sono stati causati dalla confusione dovuta al doppio asse e nel secondo task a seguito della spiegazione la percentuale di errore è diminuita, inaspettati sono stati invece gli errori sul terzo task a causa dell'interpretazione delle parole: 'andamento opposto'.

3.2 Seconda visualizzazione

Sono stati proposti i seguenti task:

- 1) Individua le prime 2 emozioni preponderanti all'inizio e alla fine dell'interazione
- 2) Individua l'emozione meno espressa sulla piattaforma a fine visualizzazione
- 3) Individua il numero totale di tweets sulla vergogna al marzo 2020

Di seguito i risultati:

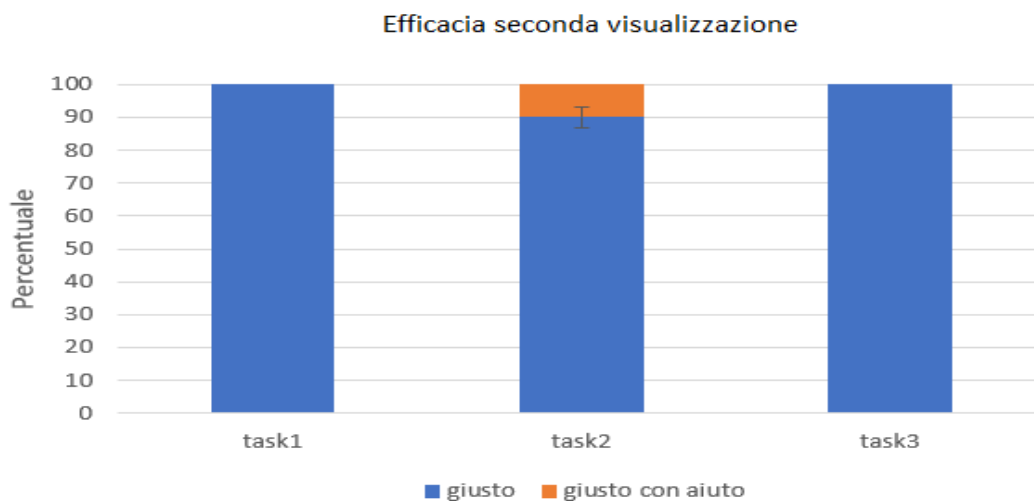


Figura 3: Bar chart errori task

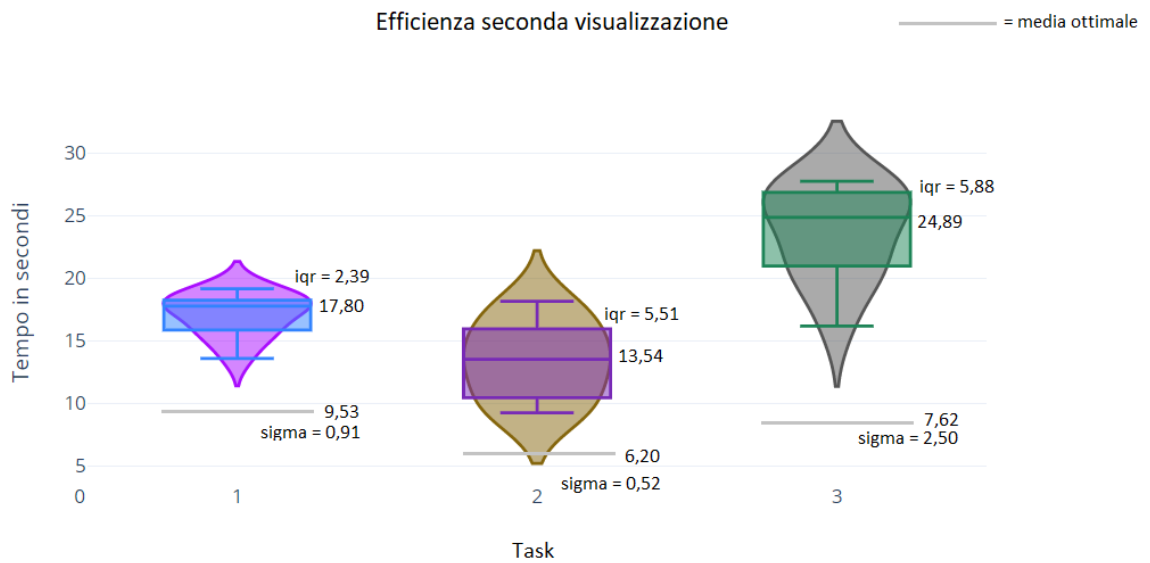


Figura 4: Violin plot tempi di esecuzione

La seconda visualizzazione come atteso essendo più intuitiva ha riscontrato dei risultati migliori in termini di efficacia in quanto è più familiare anche grazie all'uso delle emoticon. Il terzo task ha richiesto più tempo a causa della poca praticità del campione di impostare la data precisa richiesta dal task.

4 Questionario psicometrico

Sempre per valutare la qualità delle nostre visualizzazioni ci siamo affidati ad un questionario psicometrico attraverso la scala Cabitza – Locoro; essa si basa su una valutazione da 1 (pochissimo) a 6 (tantissimo), che l'utente deve esprimere sulla visualizzazione, per ognuno dei seguenti aggettivi:

- Utile
- Chiara
- Bella
- Informativa

Il questionario è stato sottoposto a 24 utenti. Una volta presi tutti i risultati è stata effettuato un raggruppamento dei risultati su due fasce:

1 - 3: Valutazione bassa

4 - 6: Valutazione alta

La visualizzazione dei risultati del questionario psicometrico viene infatti realizzata su questi due tipi di valutazione, mettendoli in confronto su uno stacked bar chart. Nello specifico, su quest'ultimo viene riportata, per ogni aggettivo, la percentuale di utenti che hanno dato un'alta valutazione contrapposta alla percentuale di utenti che hanno dato una bassa valutazione.

Prima visualizzazione

Nella prima visualizzazione la caratteristica più apprezzata è l'informativa, grazie ai vari indicatori che segnalano le date più importanti direttamente sul grafico. Le caratteristiche meno apprezzate sono invece quelle della chiarezza e della bellezza della visualizzazione, anche a causa della presenza del doppio asse e delle linee temporali di non immediato riscontro.

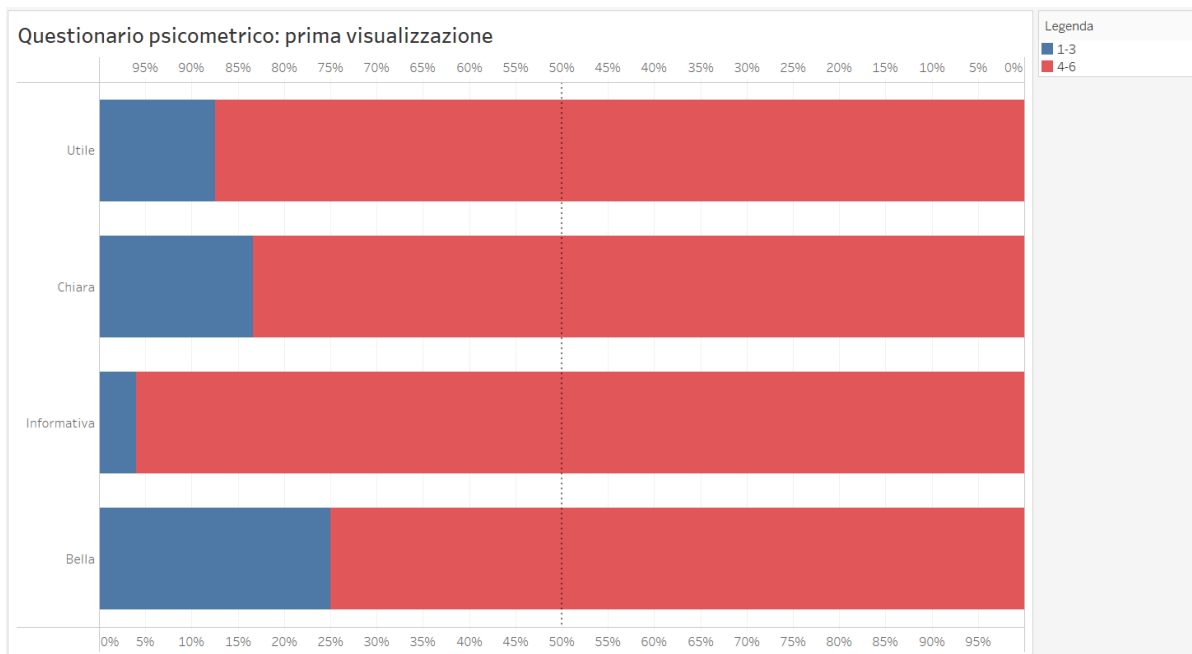


Figura 5: Stacked bar chart

Il correlogramma indica una buona correlazione tra gli aggettivi in generale ad eccezione degli aggettivi “bella” e “chiara”, che potrebbe indicare che gli intervistati che non hanno giudicato chiara la visualizzazione non l’hanno giudicata bella.

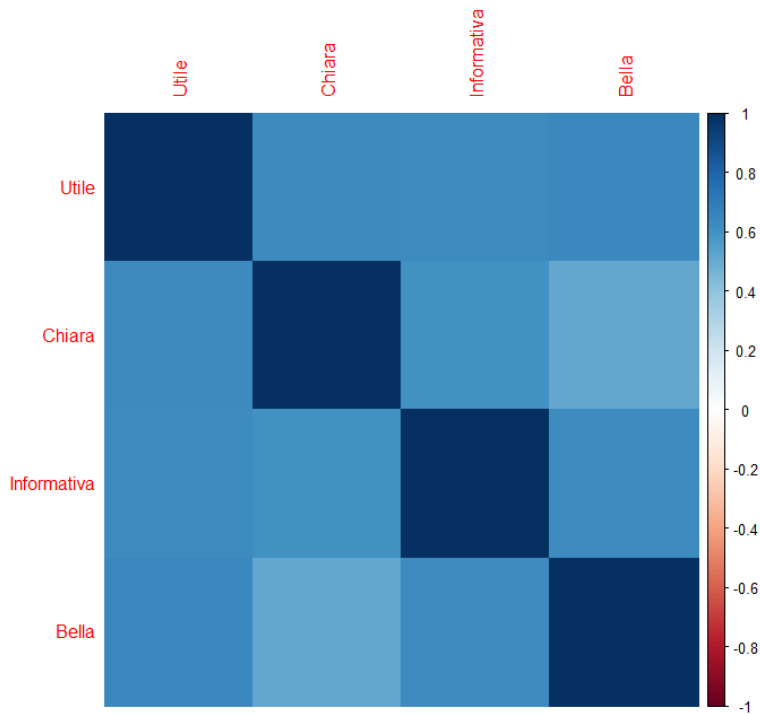


Figura 6: Correlogramma

Seconda visualizzazione

Nella race chart, a differenza della serie storica, viene apprezzata maggiormente la bellezza della visualizzazione, oltre alla sua chiarezza: questo è dato sicuramente dal fatto che la race chart, combinata con gli emoticon che rappresentano emozioni diverse, risultano di impatto immediato e danno un tocco di originalità alla visualizzazione. Di contro, le caratteristiche meno apprezzate sono la sua utilità e informativa, vista la sua capacità ridotta di informare le persone su eventi più dettagliati, a causa della velocità e del tempo su base mensile della visualizzazione.

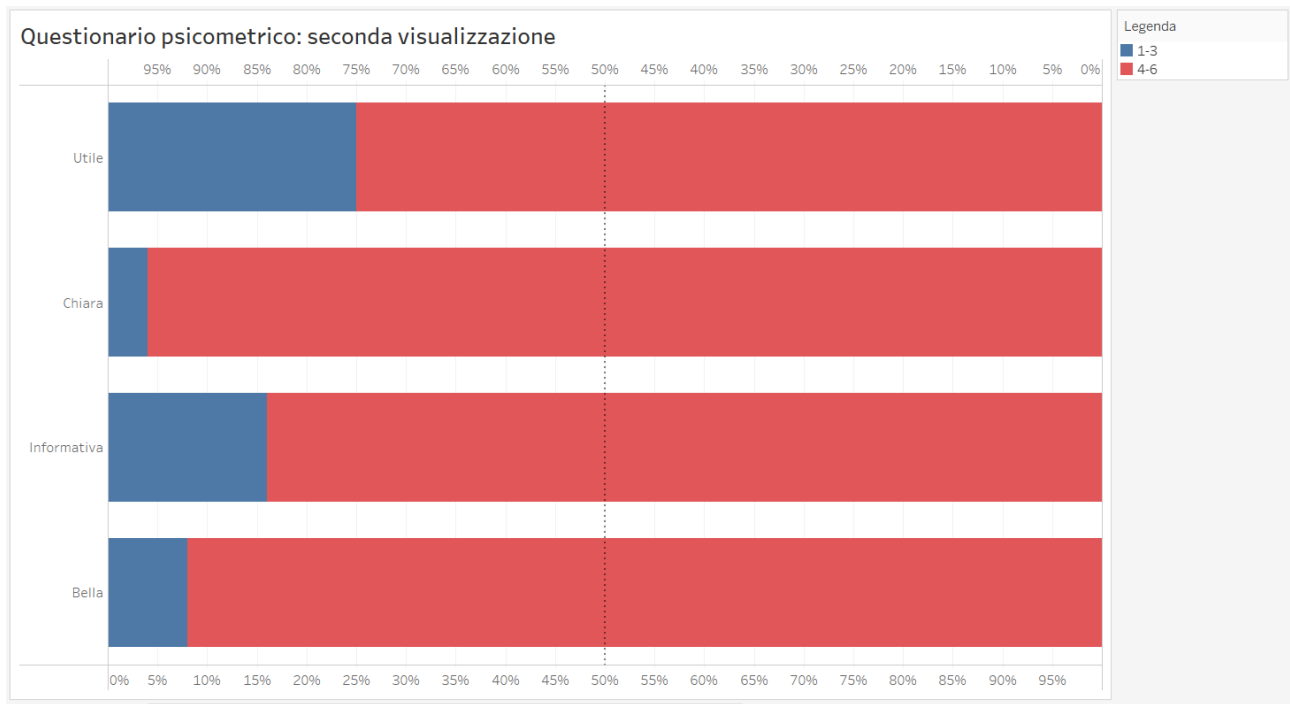


Figura 7: Stacked bar chart

Il correlogramma della seconda visualizzazione indica una correlazione maggiore tra gli aggettivi “bella” e “chiara”, che come abbiamo visto sono anche le caratteristiche più apprezzate. Gli aggettivi meno correlati sono invece “bella” e “informativa”, dato il loro significato diverso che viene attribuito alla visualizzazione in questione.

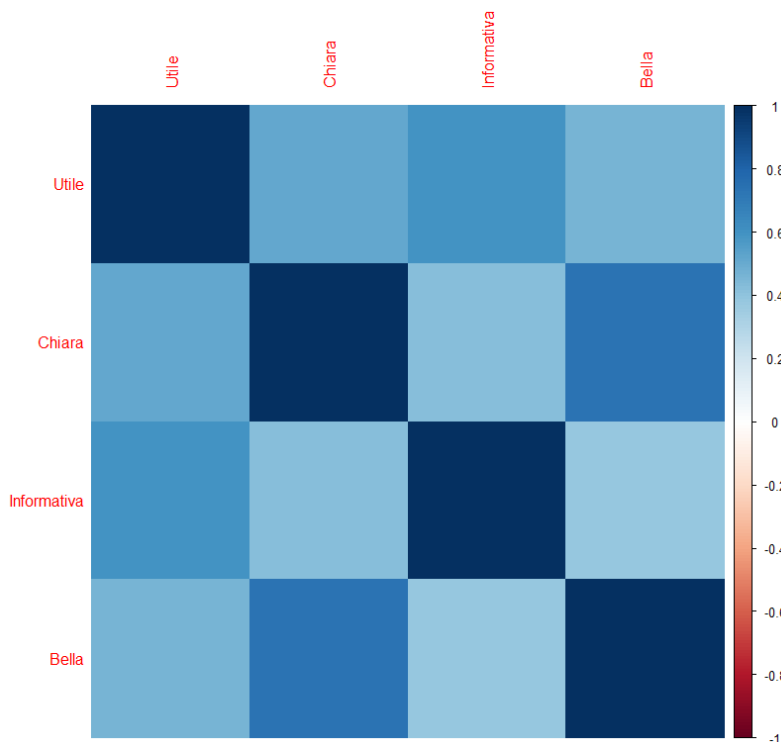


Figura 8: Correlogramma

5 Conclusioni

Nella prima visualizzazione abbiamo ricostruito l'andamento della paura, espressa dagli utenti di Twitter in Italia, con l'andamento dei contagi giornalieri; in particolare abbiamo notato che gli utenti della piattaforma esprimono più paura nei momenti in cui i contagi aumentano. Questo è dovuto anche al fatto che nei momenti in cui aumentano i contagi si parla anche di più del covid, avendo come conseguenza una maggior circolazione di notizie e una maggior paura al riguardo. In questo caso il risultato della ricerca è stato quello atteso.

Inoltre nella seconda visualizzazione abbiamo messo a confronto il numero totale dei tweet contenenti diversi sentimenti (con la parola covid) e abbiamo quindi mostrato l'andamento mensile degli stessi, per capirne meglio l'entità durante la pandemia. Abbiamo notato che i sentimenti preponderanti sono stati la paura e la vergogna, risultato che ci ha un po' sorpreso perché ci saremmo aspettati valori più alto di odio, mentre esso rimane invece tra le ultime posizioni.