

Università degli Studi di Milano - Bicocca
Corso di Laurea Magistrale di Data Science
Anno accademico 2020/2021



**Tottenham-Manchester City, analisi del sentiment in relazione agli
eventi del match**

Report Data Visualization

Settembre 2021

Cristiano Ruttico mat 809360

Francesco Martinelli mat 873685

Enrico Ragusa mat 862702

Introduzione

La Premier League, uno dei maggiori campionati di calcio, ha avuto inizio il 13 agosto 2021. Con il ritorno di persone allo stadio, abbiamo deciso di analizzare i Tweet della partita Tottenham-Manchester City, poiché rappresentava una sfida importante anche alla luce delle voci di mercato riguardanti il calciatore Harry Kane (Tottenham) e del suo probabile passaggio al Manchester City.

Prima infografica

Link:

https://public.tableau.com/app/profile/cristiano6823/viz/Primavisualizzazione_16306757953910/Dashboard1

Nella prima infografica vengono riportate due visualizzazioni: la prima riguarda il numero di tweet per minuto di partita, disambiguabili tramite l'utilizzo dei filtri per tifoseria di appartenenza; è inoltre possibile visualizzare gli eventi accaduti minuto per minuto semplicemente passando con il cursore del mouse sul grafico. L'asse x presenta dei valori di tipo stringa che indicano i minuti del match ed eventuali minuti di recupero, l'asse y, rappresenta il numero di tweet. Sono inoltre presenti dei filtri che permettono di scegliere di visualizzare eventi del Tottenham o del Manchester City.

La seconda visualizzazione della prima infografica rappresenta invece il sentiment medio dei tifosi minuto per minuto; in particolare sull'asse x abbiamo i minuti della partita e sull'asse y il compound (misura di sintesi del sentiment fornita dalla libreria Vader) medio per minuto; anche in questo grafico è possibile disambiguare per tifoseria così da scegliere di visualizzare il sentiment di una squadra alla volta.

Seconda infografica

Link:

<https://public.tableau.com/app/profile/cristiano6823/viz/Secondavisualizzazione/Dashboard2>

Nella seconda infografica vengono riportate tre visualizzazioni: la prima riguarda un wordcloud; in particolare vengono presentate le parole più citate all'interno dei tweet: si nota come Harry Kane, sebbene non abbia giocato la partita, sia stato al centro della discussione dei tifosi. Per quanto riguarda la seconda visualizzazione abbiamo scelto di riportare i primi 10 giocatori più citati all'interno dei tweet, in particolare abbiamo scelto di usare un lollipop chart poiché la lunghezza delle barre era molto simile in alcuni casi; è stato scelto di ribaltare gli assi al fine di avere le etichette dei giocatori ben leggibili, infatti sull'asse delle y troviamo le etichette dei giocatori, invece sulle x troviamo il numero di citazioni all'interno dei tweet. La terza visualizzazione è uno stacked bar chart; esso presenta sull'asse x i 9 hashtag più usati durante la partita, sull'asse y troviamo invece il numero di volte in cui sono stati citati; tramite i filtri è inoltre possibile selezionare delle soglie per escludere un certo numero di hashtag, ed è inoltre possibile selezionare la squadra di cui si vuole visualizzare gli hashtag.

Terza infografica:

Link:

<https://public.tableau.com/app/profile/cristiano6823/viz/Terzavisualizzazione/Statiche>

Nella terza infografica vengono riportate due visualizzazioni: la prima visualizzazione riguarda le statistiche dei passaggi dei giocatori, in particolare vengono evidenziate le partecipazioni alla fase di costruzione dell'azione; in particolare l'asse x rappresenta il numero di passaggi tentati, e l'asse y invece rappresenta la percentuale di passaggi riusciti. La grandezza della bolla, come precedentemente introdotto rappresenta le azioni create. Per quanto riguarda la seconda visualizzazione è stato rappresentato un bubble chart indicante, tramite la grandezza della bolla, il numero di goal attesi per giocatore. Per quanto riguarda gli goal attesi si intende una statistica che fornisce la quantità di goal che ci si aspetterebbe da quel giocatore; è inoltre possibile filtrare per numero di tiri effettuati e di mettere in evidenza i giocatori di una squadra oppure dell'altra.

Valutazione euristica

Per valutare le visualizzazioni è stato chiesto a 6 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 tale per cui gli utenti sono stati lasciati liberi di esplorare le infografiche create.

Riportiamo di seguito i risultati:

Prima infografica: La valutazione euristica è stata effettuata solo sulla prima infografica la quale comprende 2 visualizzazioni; in particolare per la prima visualizzazione: non ha presentato grandi problemi nella comprensione dei dati presentati se non per l'utilizzo dei filtri, i quali non sono stati molto considerati. Per quanto riguarda la seconda visualizzazione non sono stati incontrati particolari problemi.

Test Utente

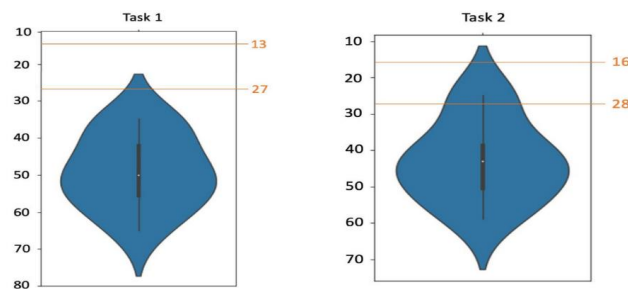
In questa sezione si andranno ad elencare i risultati ottenuti attraverso lo user test. Sono stati proposti 2 task eterogenei al campione di riferimento di 12 persone, questo è avvenuto per 2 visualizzazioni, una appartenente alla seconda infografica e l'altra appartenente alla terza. Il sondaggio è servito per comprendere i punti deboli delle visualizzazioni e misurare l'efficacia e l'efficienza delle stesse. I task proposti erano relativamente semplici:

- task 1) visualizzare quali fossero i 2 hashtag più citati per ogni squadra (**terza visualizzazione, seconda infografica**)
- task 2) visualizzare i giocatori con almeno 3 tiri, con un valore di goal attesi maggiore di 0.20, e di questi saper dire quanti fossero i tiri in porta. (**seconda visualizzazione, terza infografica**)

Abbiamo misurato l'**efficacia** delle visualizzazioni misurando la proporzione di task svolti correttamente sul totale; i risultati dell'efficacia non sono riportati graficamente poiché non si sono registrati errori nello svolgimento. Per misurare l'**efficienza** delle visualizzazioni sono stati misurati

i tempi di risposta degli individui e sono stati confrontati con la media delle tempistiche dei 3 componenti del gruppo. Il tutto è stato rappresentato in un violin plot con all'interno un box plot.

Di seguito riportiamo i violin plots raffiguranti anche i nostri tempi(veloce e lento) di svolgimento dei task, identificabili dalle linee arancioni.



Questionario psicometrico

Inoltre, al fine di valutare la qualità delle visualizzazioni ci siamo affidati ad un questionario psicometrico attraverso la scala **Cabitzza – 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. Ottenuti i risultati si è provveduto a fare un raggruppamento:

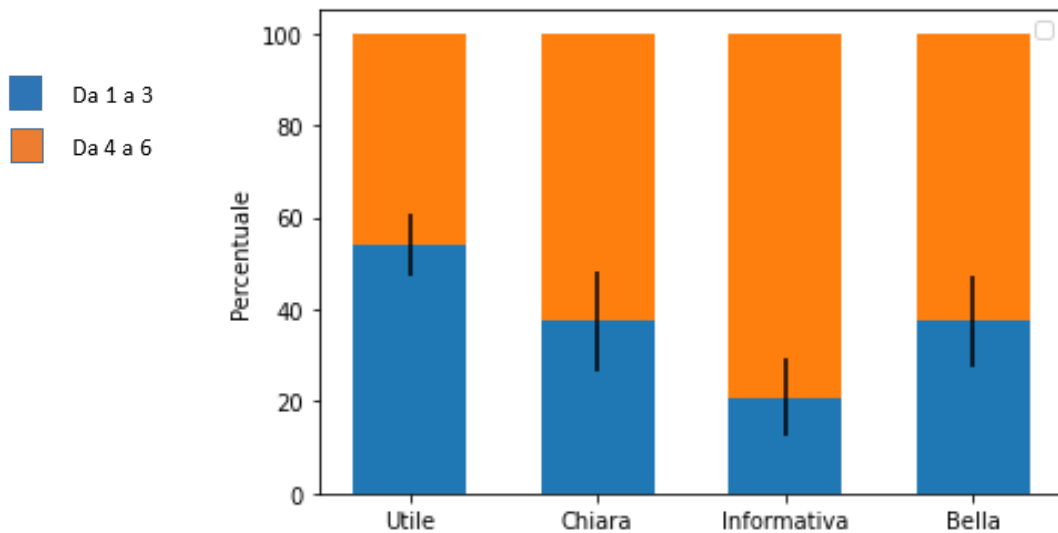
Da 1 a 3: Valutazione bassa ■

Da 4 a 6: Valutazione alta ■

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

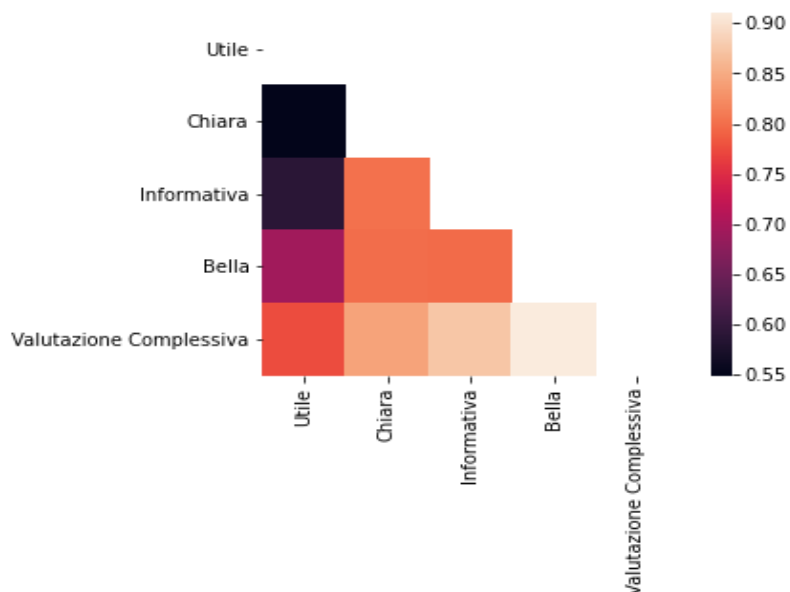
Valutazione psicometrica prima visualizzazione della prima infografica

Nella prima visualizzazione le caratteristiche più apprezzate sono l'informatività e la bellezza; le caratteristiche meno apprezzate sono invece quelle della chiarezza e della utilità.



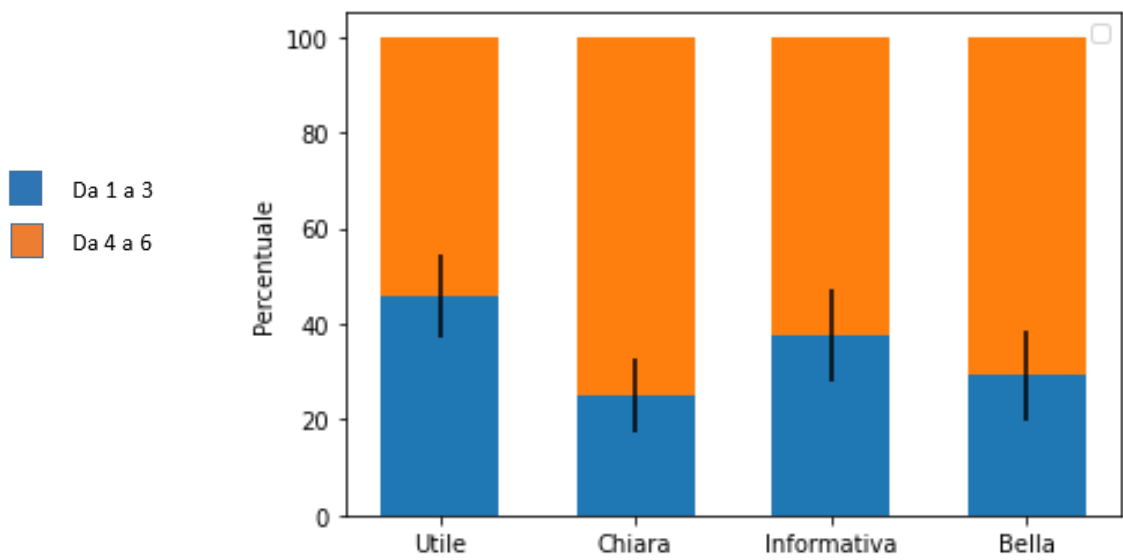
Correlogramma prima visualizzazione della prima infografica

Il correlogramma indica una buona correlazione tra gli aggettivi in generale ad eccezione degli aggettivi "utile" e "chiara" e "utile" e "informativa".



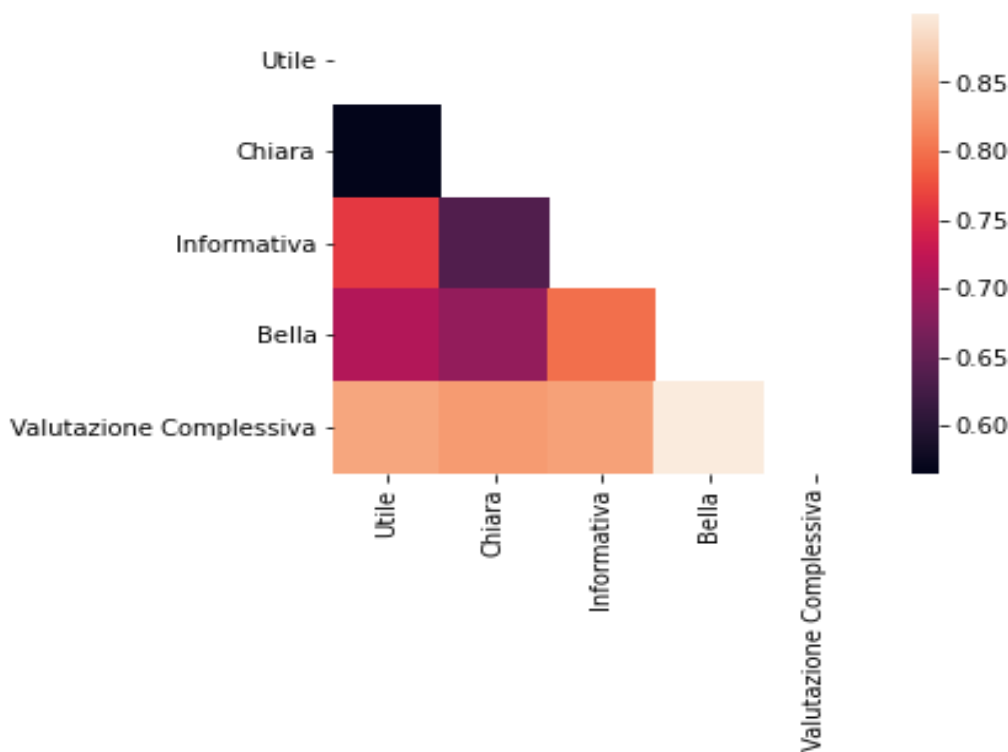
Valutazione psicometrica seconda visualizzazione della seconda infografica

In questa visualizzazione le caratteristiche più apprezzate sono la chiarezza e la bellezza. Le caratteristiche meno apprezzate sono invece quelle dell’utilità e informatività.



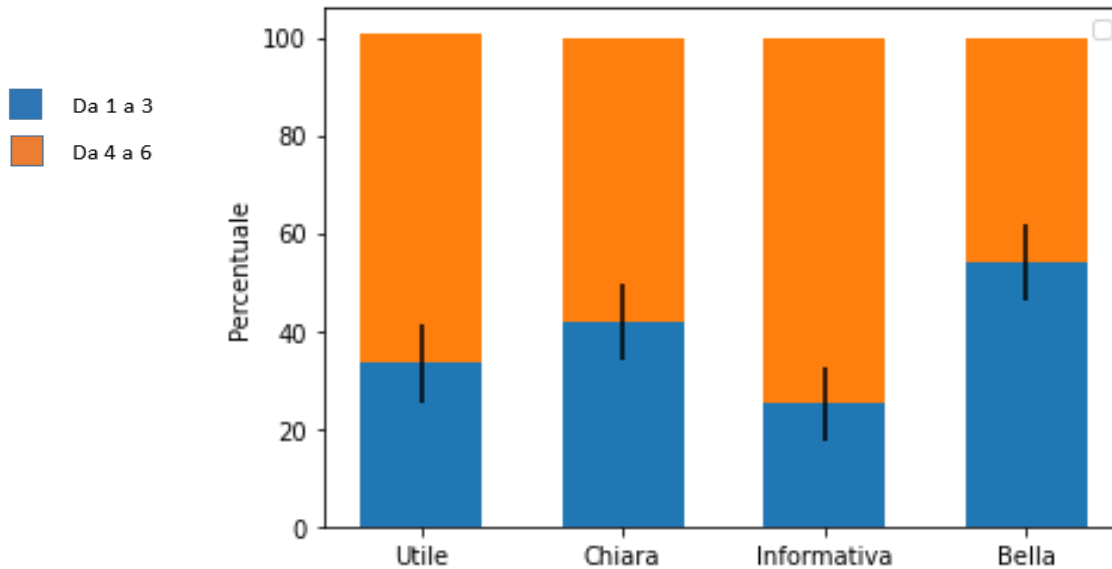
Correlogramma seconda visualizzazione della seconda infografica

Il correlogramma indica una buona correlazione tra gli aggettivi in generale ad eccezione degli aggettivi “utile” e “chiara” e “utile” e “bella”, “chiara” e “informativa”.



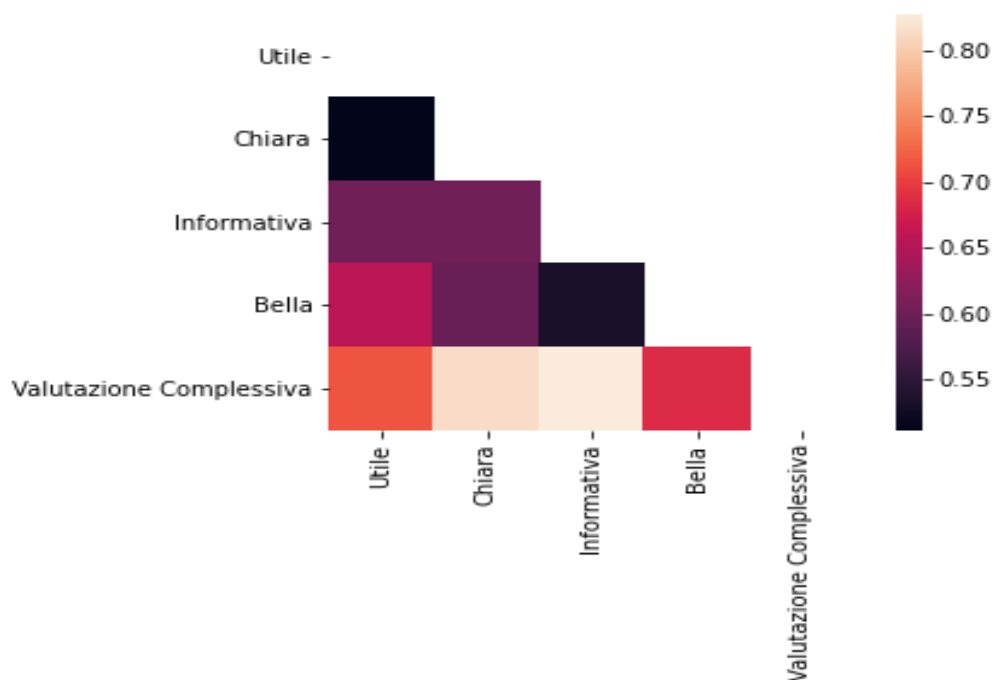
Valutazione psicometrica terza visualizzazione terza infografica

In questa visualizzazione le caratteristiche più apprezzate sono l'informatività e l'utilità. Le caratteristiche meno apprezzate sono invece quelle della bellezza e della chiarezza.



Correlogramma terza visualizzazione della terza infografica

Il correlogramma indica una buona correlazione tra gli aggettivi in generale ad eccezione degli aggettivi "utile" e "chiara" e "utile" e "bella", "chiara" e "bella", "informativa" e "bella"



Conclusioni

Per quanto riguarda la **prima infografica**: sia la prima visualizzazione che la seconda riteniamo siano calzanti nel semplicemente mostrare, con l'ausilio del context, l'andamento generale della partita; soprattutto il gol di Son ha rappresentato per entrambe un punto di svolta grafico sia a livello di numero di tweet, per quanto riguarda la prima visualizzazione, che per sentiment negativo, per quelli che sono i tifosi del City nella seconda visualizzazione.

Per quanto riguarda la **seconda infografica**: si ritiene che il wordcloud dia risultati abbastanza significativi in quanto mostra come tra le parole più twittate sia presente 'Harry Kane' il quale, sebbene non abbia preso parte al match, era al centro delle voci di mercato tra le due squadre. Per la seconda visualizzazione vediamo chiaramente come il fatto che Son abbia segnato porti ad un consistente distacco tra il numero di tweet in cui il suo nome è presente e quelli in cui sono presenti gli altri giocatori. Per quanto riguarda la terza visualizzazione notiamo che (si suppone grazie alla vittoria del Tottenham) l'hashtag più twittato sia #coys, ovvero l'hashtag ufficiale della squadra di Londra.

Sulla **terza infografica**: per quello che concerne la prima visualizzazione, si nota come i giocatori più pericolosi ed abili nella costruzione del gioco siano stati quelli del City poiché presentano le percentuali più alte in termini di passaggi riusciti e, contemporaneamente, un alto numero di occasioni create. Infine, la seconda visualizzazione rappresenta, con la grandezza delle aree, i gol attesi; in questo modo si può avere a primo impatto visivo l'informazione senza dover necessariamente utilizzare i filtri.