



Data Visualization Report

# **Come l'andamento del bitcoin modella il sentiment sulla piattaforma Twitter**

**Marco Branciforti<sup>1</sup>, Emanuele Marnati<sup>2</sup>, Valerio Schips<sup>3</sup>**

---

<sup>1</sup> B.Sc. Statistica e gestione delle informazioni (UniMib), M.Sc. Data Science, mat 796670

<sup>2</sup> B.Sc. Economia e Commercio (UniMib), M.Sc. Data Science, mat 812503

<sup>3</sup> B.Sc. Ingegneria informatica (PoliMi), M.Sc. Data Science, mat 872954

## Indice

Data Visualization Report .....	1
Indice .....	2
Elenco Figure.....	2
1. Introduzione.....	3
2. Infografiche.....	4
2.1 Infografica 1 .....	4
2.2 Infografica 2 .....	5
2.3 Infografica 3 .....	6
2.4 Infografica 4&5.....	7
3. Valutazione Euristica Delle Visualizzazioni.....	9
3.1 V.E. Prima infografica.....	9
3.2 V.E. Seconda infografica.....	9
3.3 V.E. Terza infografica.....	10
3.4 V.E. Quarta e quinta infografica.....	10
4. Questionario Psicometrico ....	11
4.1 Prima infografica.....	12
4.2 Seconda infografica.....	13
4.3 Terza Infografica .....	14
4.4 Quarta e Quinta Infografica ..	15
5. Valutazione della fruizione degli utenti.....	16
5.1 Prima infografica.....	16
5.2 Seconda infografica.....	17
5.3 Terza infografica.....	18
5.4 Quarta e quinta infografica ..	19
6. Link a Tableau Public .....	21

## Elenco Figure

Figura 1 .....	4
Figura 2 .....	5
Figura 3 .....	6
Figura 4 .....	7
Figura 5 .....	8
Figura 6 .....	12
Figura 7 .....	12
Figura 8 .....	13
Figura 9 .....	13
Figura 10 .....	14
Figura 11 .....	14
Figura 12 .....	15
Figura 13 .....	15
Figura 14 .....	16
Figura 15 .....	17
Figura 16 .....	18
Figura 17 .....	19
Figura 18 .....	20

## 1. Introduzione

L'idea del progetto nasce con l'intento di capire come l'andamento di alcuni valori tecnico finanziari dell'asset bitcoin abbia influito sul sentimento sviluppatosi nello stesso arco temporale sul social network Twitter. Il progetto consiste in un applicativo in tempo reale in grado di raccogliere diversi valori e misure dell'asset bitcoin, la più celebre e capitalizzata criptovaluta, e di raccogliere contemporaneamente l'umore che gli utenti mostrano sul social network più idoneo a tale scopo: Twitter. Per il mese di novembre sono stati raccolti e memorizzati circa 600.000 tweet con rispettiva sentiment e hashtag. Sono stati integrati i valori OCLH (open, close, high, low, volume) dell'andamento del bitcoin (44.000 valori) e inoltre si è deciso di integrare questi dati con un valore ancor più tecnico che è il transaction per second (TPS), il quale mostra il quantitativo di transazioni validate con risoluzione al secondo (valore medio dei secondi in un'ora) sempre nello stesso arco. Si è proceduto poi con un'analisi statistica e visuale dei dati raccolti tenendo come linea guida la chiarezza, dato che l'argomento è complesso ed estremamente strutturato. L'ultimo approccio è stato quello di implementare un modello di machine learning. Per ottenere la visualizzazione dei dati si è adoperato il Software Tableau (versione 2020.3) dopo aver effettuato alcune manipolazioni con python ai fini di una corretta struttura per le visualizzazioni. Infine, si è scelto di adoperare R per analizzare il prodotto finito secondo i canoni della valutazione euristica, della psicomетria e della effettiva fruizione che compie un campione di utenza che non ha alcun tipo di cultura finanziaria o del mondo delle criptovalute. A tal fine sono state realizzate 5 infografiche statiche di cui 2 (le ultime) sono parte di una visione speculare che può essere intesa come un prodotto unico.

## 2. Infografiche

Nelle analisi che seguono si farà spesso riferimento a come Pay Pal, noto servizio di pagamento ideato tra gli altri da Elon Musk, abbia segnato profondamente gli andamenti del mese divenuti Bullish<sup>4</sup> per via di dichiarazioni ufficiali -e non- rilasciate dal CEO della medesima azienda Dan Schulman e dai suoi funzionari. L'obiettivo della ricerca è capire se e come "i sentimenti escano dalla stanza dei mercati".

### 2.1 INFOGRAFICA 1

#### **Analisi dell'andamento del bitcoin (con un grafico a candela) rispetto all'andamento medio del valore di sentiment nel mese di novembre.**

Il candel-stick, ovvero grafico a candela è uno metodo noto ed efficace che permette di analizzare, tramite il corpo della candela, come il prezzo giornaliero sia variato e, tramite le linee verticali sottostanti, quale sia stato il massimo ed il minimo. Come è facile intuire in blu si riporta un andamento giornaliero crescente, mentre in rosso decrescente.

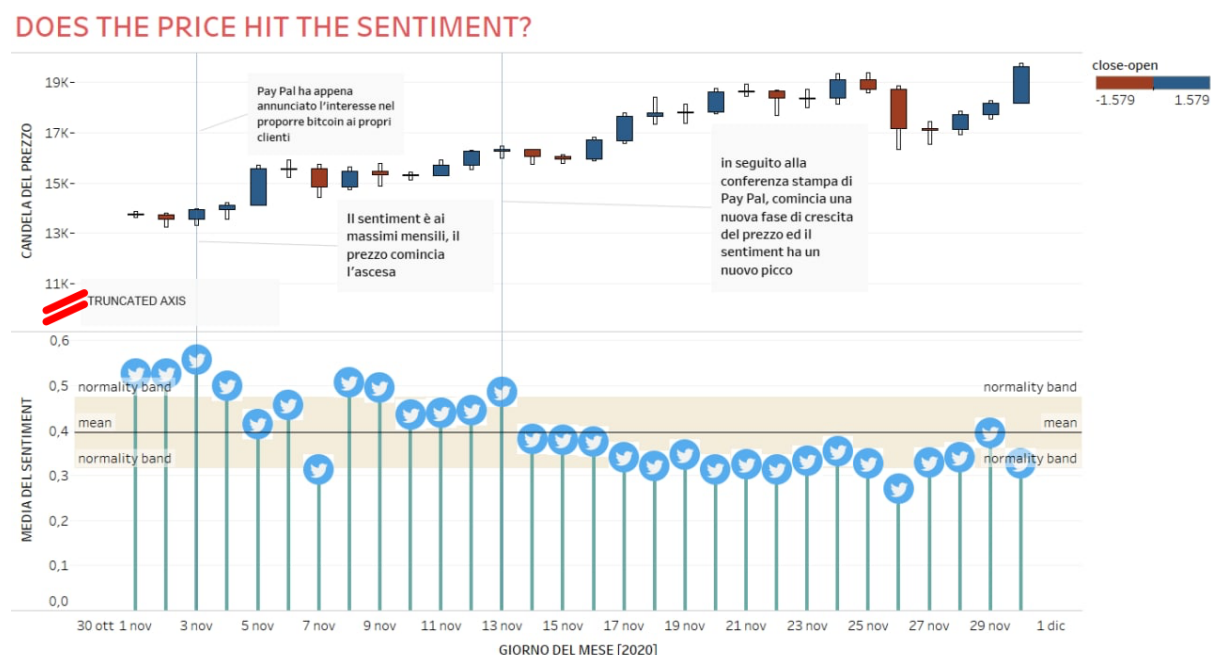


Figura 1

<sup>4</sup> Termine per indicare come il mercato si stia muovendo al rialzo in modo aggressivo (contr. Bearish)

## 2.2 INFOGRAFICA 2

### Correlazione del volume rispetto all'andamento del prezzo massimo giornaliero di bitcoin nel mese di novembre.

Il grafico mostra una generale correlazione delle due grandezze che racchiudono molteplici significati i quali gli autori debbono necessariamente semplificare. Novembre è stato un mese molto importante per bitcoin, questo scatter plot mette in luce come, in una situazione estremamente Bullish come questa, il volume cresca assieme al prezzo e viceversa. Il fenomeno è reso evidente anche dalla linea netta verticale che separa il grafico, una linea naturale che indica la metà (naturalmente formata) del mese. È evidente un crescendo del fenomeno anche in relazione dalla dimensione temporale.

#### NOVEMBER VOLUME-PRICE SCATTER PLOT

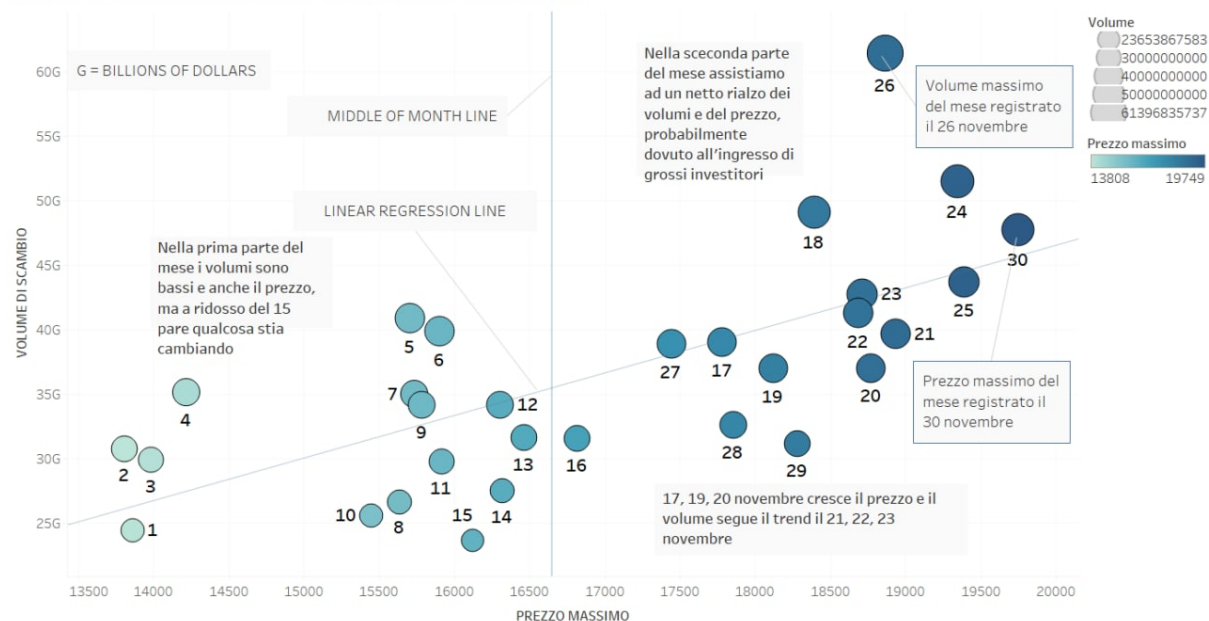


Figura 2

## 2.3 INFOGRAFICA 3

### Analisi di tre particolari hashtag indicanti lo stesso asset.

Questa infografica è stata pensata tenendo bene in mente il concetto di **rasoio di Hanlon**<sup>5</sup> applicato all'utenza Twitter. La comunità, lettori compresi, conosce la differenza tra "bitcoin", "Bitcoin" e "BTC"? Analizzando il fenomeno (opportunamente ponderato per tutti gli hashtag contenuti nella base dati di novembre: la popolazione di riferimento), si è ipotizzato che l'utenza prenda la scelta sulla base della mera pigrizia e non sappia distinguere protocollo Bitcoin da asset bitcoin!

Per una fruizione ottimale, si tenga a mente l'andamento, generalmente crescente, del prezzo mostrato nella prima infografica.

#### DO THE PEOPLE KNOW?

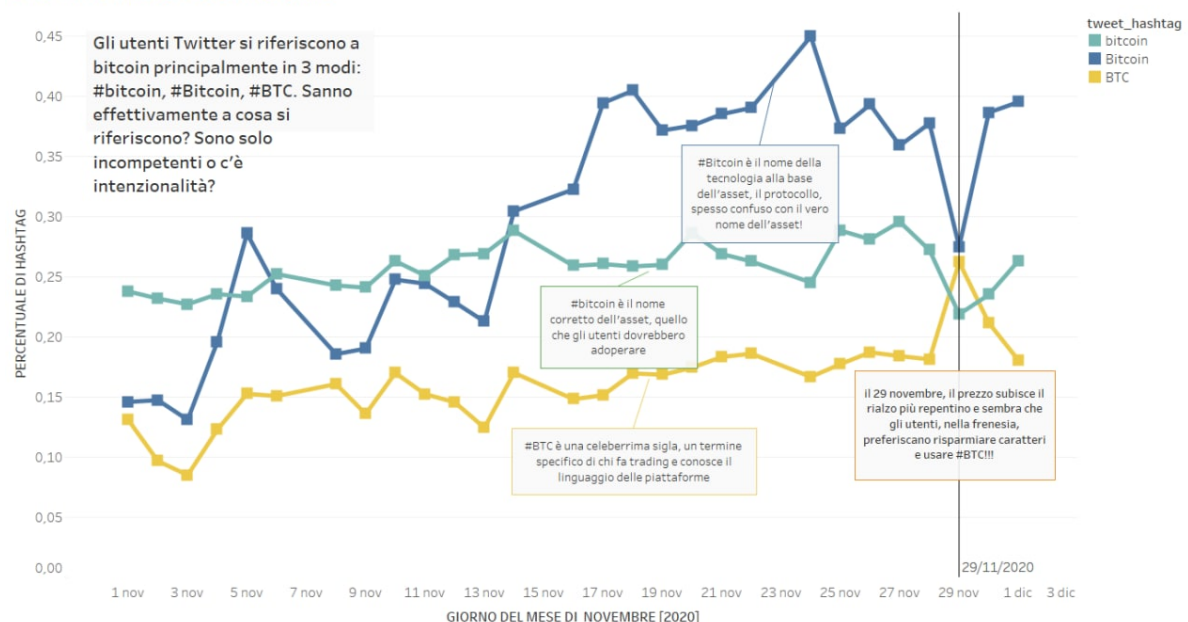


Figura 3

<sup>5</sup> Mai attribuire a malafede quel che si può adeguatamente spiegare con la stupidità

## 2.4 INFOGRAFICA 4&5

Le ultime due infografiche analizzano i fenomeni già trattati nelle prime visualizzazioni raffigurando però la percentuale di utenti che si sono espressi favorevolmente o sfavorevolmente all'andamento. Si cerca di cogliere a colpo d'occhio se effettivamente l'andamento delle transazioni al secondo e del volume influenza l'utenza.

Si può fruire delle seguenti tenendo in considerazione un altro aspetto degno di nota: le masse e gli investitori capitalizzati. Questo scontro, salito agli arbori della cronaca sul finire del gennaio 2021 con la vicenda di Reddit e Game Stop<sup>6</sup>, ha radici antiche. Il popolo di retailers (piccoli investitori) ha sempre avuto in odio i big della finanza e questo grafico vede gli utenti di Twitter come i piccoli investitori e il volume (specie nei picchi) rappresentante degli squali (talvolta definiti balene) della finanza speculativa.

### INFOGRAFICA 4

#### Percentuale di utenti positivi rispetto all'andamento delle transazioni al secondo e del volume.

##### DO TPS & VOLUME HIT THE POSITIVE TWITTER USERS IN NOVEMBER?

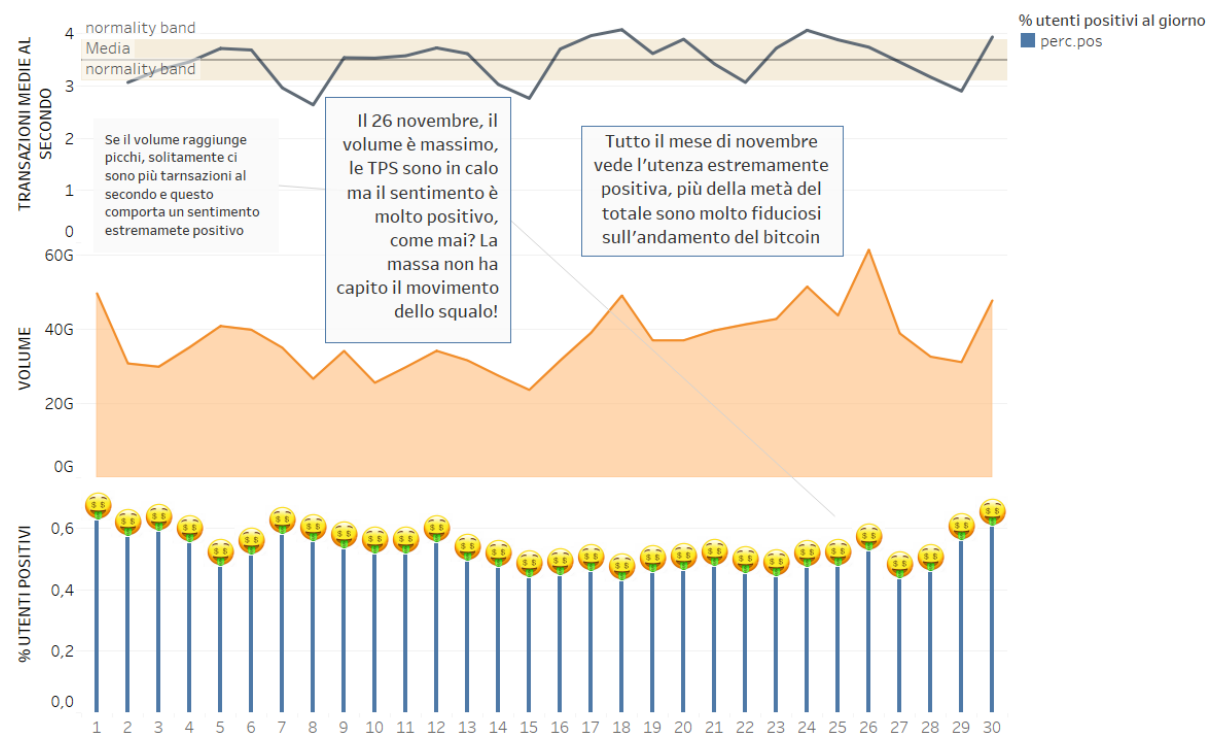


Figura 4

<sup>6</sup> [Non solo GameStop: così la rivolta dei trader sta scuotendo le Borse - 24+ \(ilsole24ore.com\)](https://ilsole24ore.com)

## INFOGRAFICA 5

### Percentuale di utenti negativi rispetto all'andamento delle transazioni al secondo e del volume.

#### DO TPS & VOLUME HIT THE NEGATIVE TWITTER USERS IN NOVEMBER?

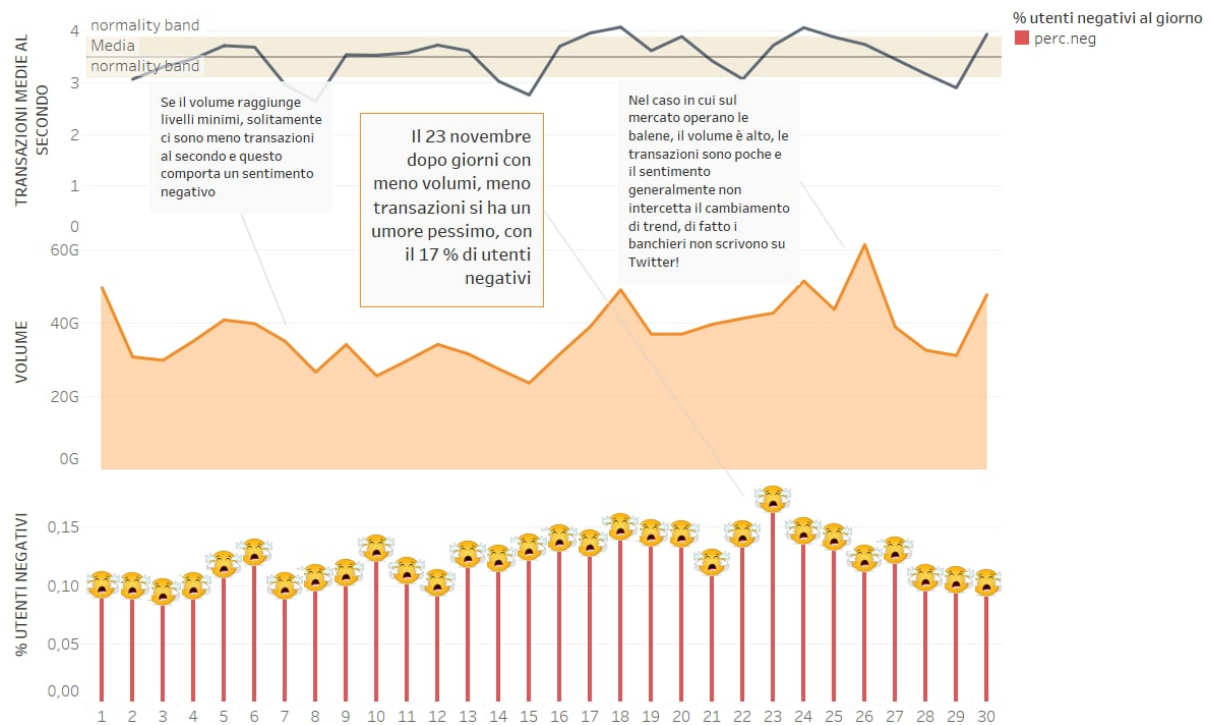


Figura 5



### 3. Valutazione Euristica Delle Visualizzazioni

La prima analisi psicometrica delle visualizzazioni è stata consumata all'interno del medesimo gruppo di lavoro, in quanto l'opinione in perenne contrasto di un economista, di un ingegnere informatico e di uno statistico, è stata la base di partenza dell'analisi semantica ed euristica del prodotto finito. Dopo aver trovato una soluzione che fosse accettata da tutte le parti sono state intervistate 4 persone per avallare la solidità concettuale del lavoro, interessante notare come tra questi individui ci fosse una persona affetta da daltonismo, che ha permesso un riscontro fedele delle scelte cromatiche.

#### 3.1 V.E. PRIMA INFOGRAFICA

La principale critica mossa verso la prima visualizzazione è stata la complessità nel capire nel dettaglio le candele del bitcoin, infatti gli utenti giovani (due dei quattro) hanno in primo luogo mostrato di non aver compreso i candel-stick. Si può attribuire tale comportamento all'idea che gli utenti, giovani e costantemente bombardati da pubblicità che mostrano grafici finanziari, abbiano con tracotanza ammesso di capire ma in secondo luogo non fossero in grado di entrare nel merito delle specifiche di questa comparazione visiva. Gli utenti hanno male interpretato la banda di normalità indicata con il colore giallo, immaginando che fosse una distorsione coloristica e non un'ulteriore informazione; si è per questo cercato di aumentare la chiarezza della visualizzazione ingrandendo le componenti del grafico e aggiustando i contorni del corpo delle candele e i fonts.

#### 3.2 V.E. SECONDA INFOGRAFICA

La prima impressione alla vista di questa infografica è stata un generale interesse, più elevato rispetto a quello riservato alla prima. Gli individui adulti (due dei quattro) hanno riscontrato difficoltà nel capire tutto il context della visualizzazione, forse per via delle scritte in piccolo. Si riscontra come la presenza di molto context sia stata funzionale all'esplicazione (gli utenti giovani hanno immediatamente colto il trend). Nonostante la presunta complessità gli utenti hanno in generale capito l'intenzione della infografica.

### **3.3 V.E. TERZA INFOGRAFICA**

L'utenza necessita del tempo per capire quale sia la dinamica del grafico ma pare essere più attenta ai dettagli dopo le prime due visualizzazioni. Non sono stati riscontranti particolari problemi con questa visualizzazione.

### **3.4 V.E. QUARTA E QUINTA INFOGRAFICA**

Riguardo queste non ci sono stati particolari problemi nel comprendere il point delle due visualizzazioni grazie principalmente alla presenza delle icone emoji. Tuttavia, alcuni utenti affermano di non comprendere appieno il collegamento tra le tre grandezze. Come biasimare questa obiezione, il fenomeno in analisi è complesso e semplificare comporta omissione di informazione.

## 4. Questionario Psicometrico

Ad un campione di 26 individui è stato somministrato un questionario psicometrico, i risultati ottenuti utilizzando la scala Cabitza-Locoro sono stati analizzati tramite il software R adoperando uno “stacked barplot ” ed un “correlogramma” per una visione grafica immediata.

Le domande del questionario sono state le seguenti:

- I. Quanto ritieni Utile l’infografica?
- II. Quanto ritieni Intuitiva l’infografica?
- III. Quanto ritieni Chiara l’infografica?
- IV. Quanto ritieni Informativa l’infografica?
- V. Quanto ritieni Bella l’infografica?
- VI. Qual’ è la valutazione complessiva che dai all’infografica?

## 4.1 PRIMA INFOGRAFICA

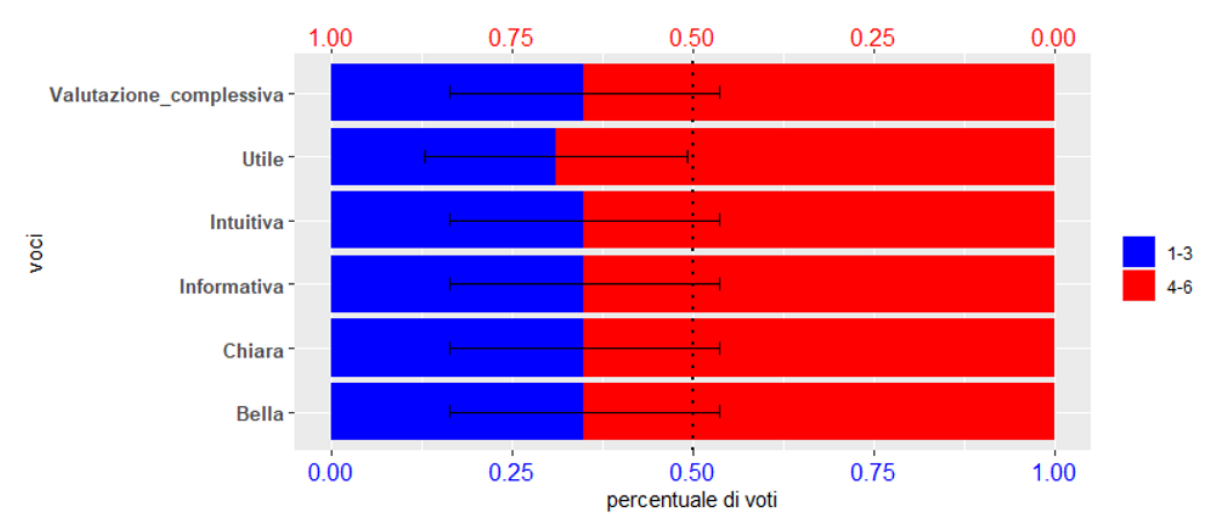


Figura 6

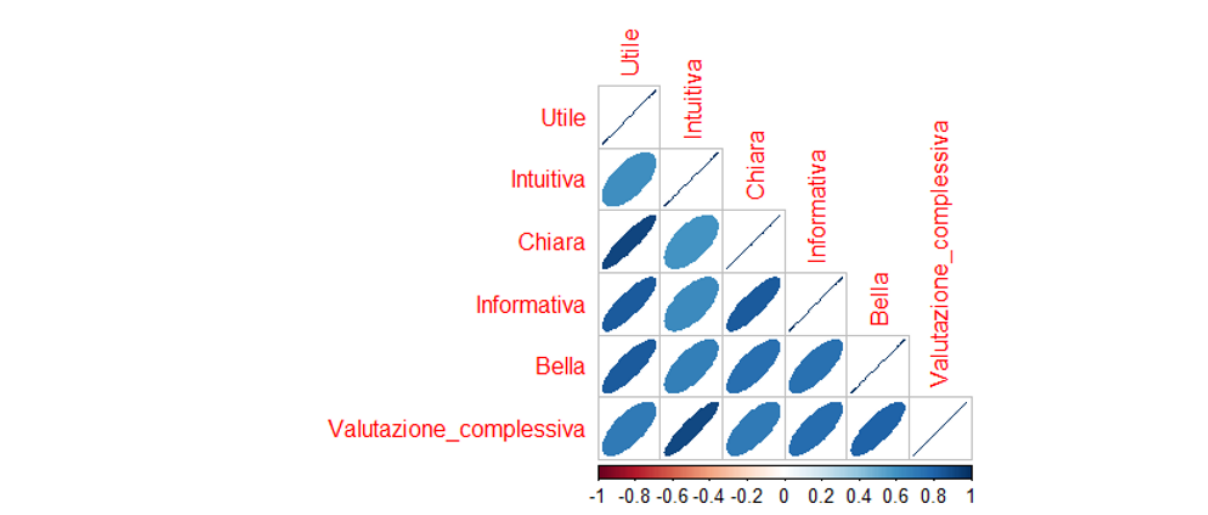


Figura 7

L'infografica ha riscontrato un particolare successo, specie nella utilità. Infatti questo è l'unico indicatore che ha una proporzione di voti negativi significativamente inferiori al 50%. La correlazione tra intuitività e valutazione complessiva risulta particolarmente elevata, un chiaro segnale di come questa visualizzazione sia immediata.

## 4.2 SECONDA INFOGRAFICA

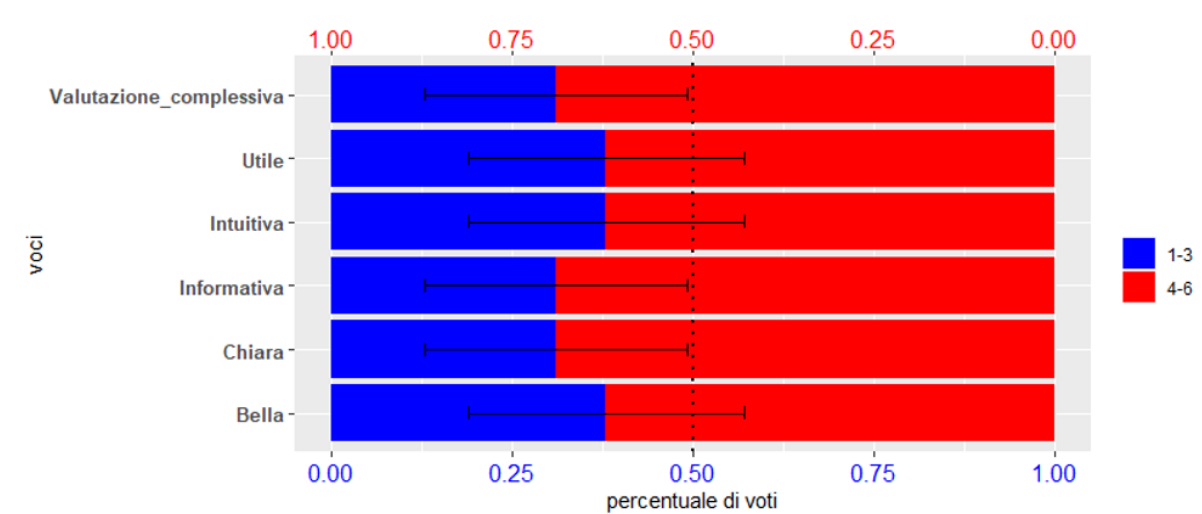


Figura 8

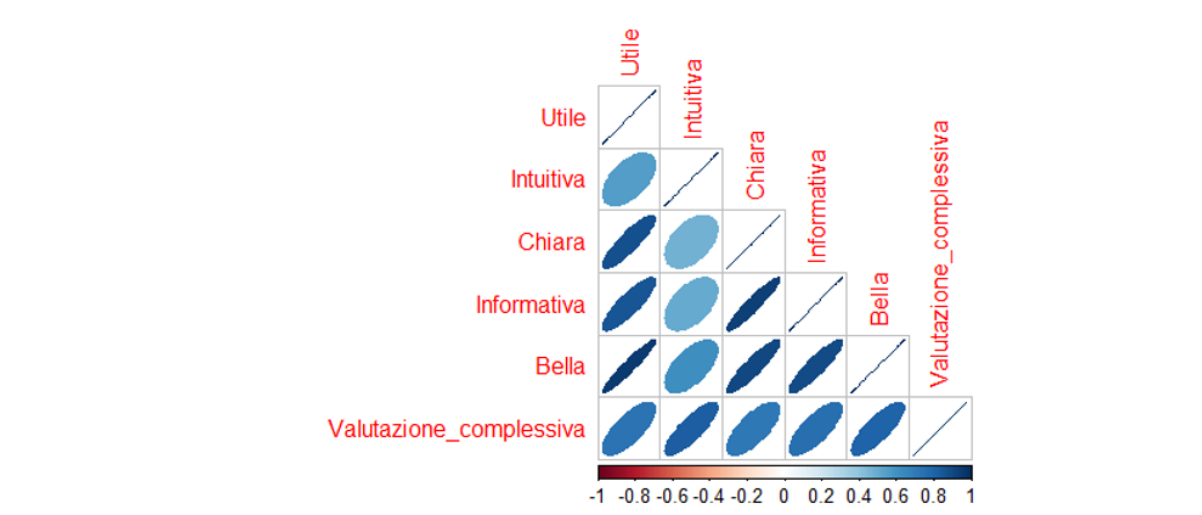


Figura 9

L'infografica presenta un leggero equilibrio tra gli indicatori. Le correlazioni sono tutte positive, le più elevate riguardano bellezza, informatività e chiarezza, un chiaro segnale di come questa visualizzazione sia piacevole e informativa.

### 4.3 TERZA INFOGRAFICA

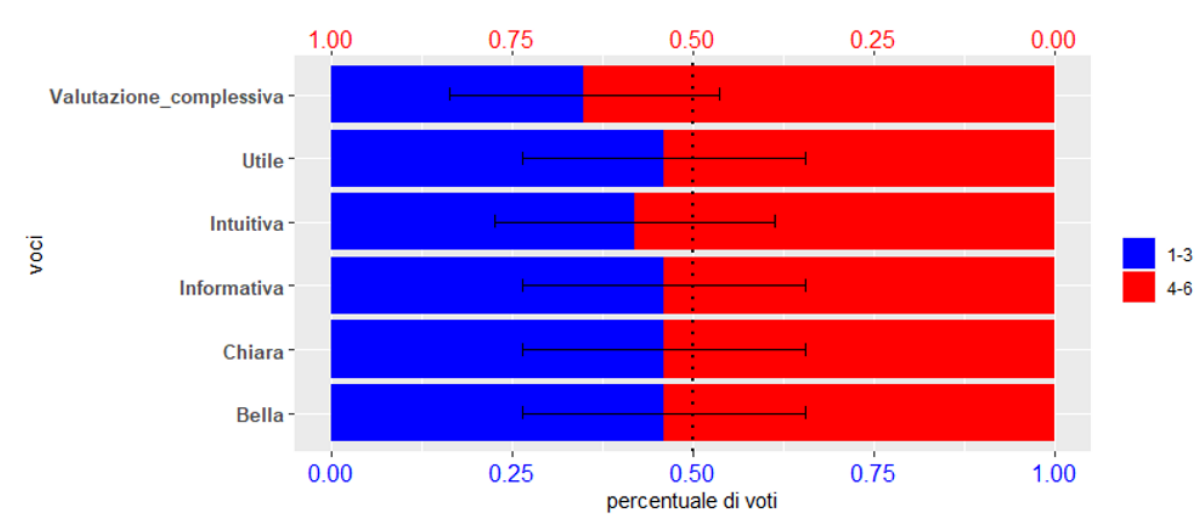


Figura 10

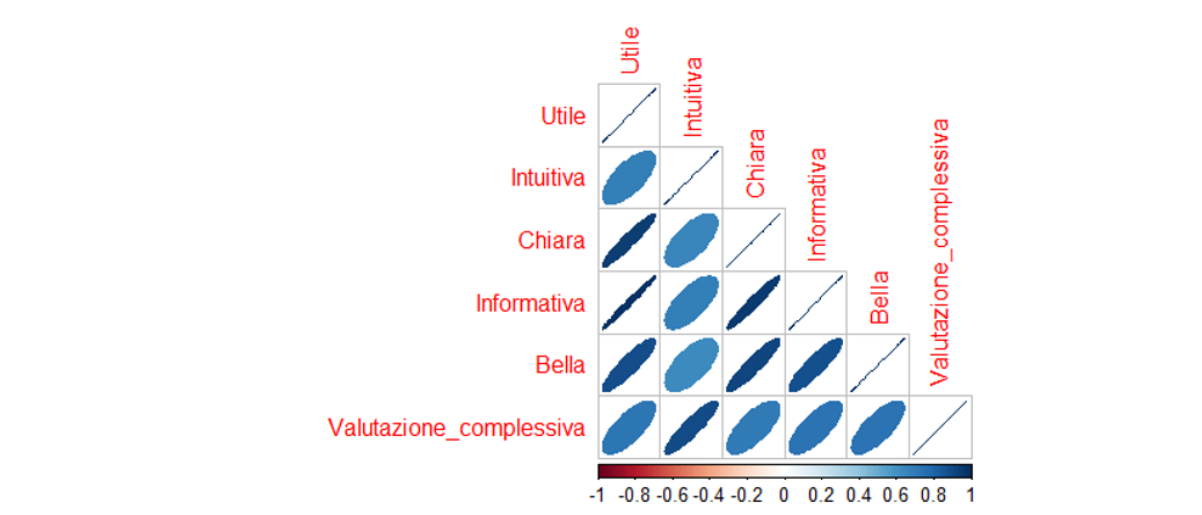


Figura 11

L'infografica non ha riscontrato un successo particolarmente schiacciante, ad eccezione della sua valutazione complessiva. Infatti nessuno degli indicatori risulta avere una proporzione di positivi significativamente superiore del 50% dei voti. Le correlazioni più rilevanti coinvolgono gli indicatori: informativa, utile e chiara.

## 4.4 QUARTA E QUINTA INFOGRAFICA

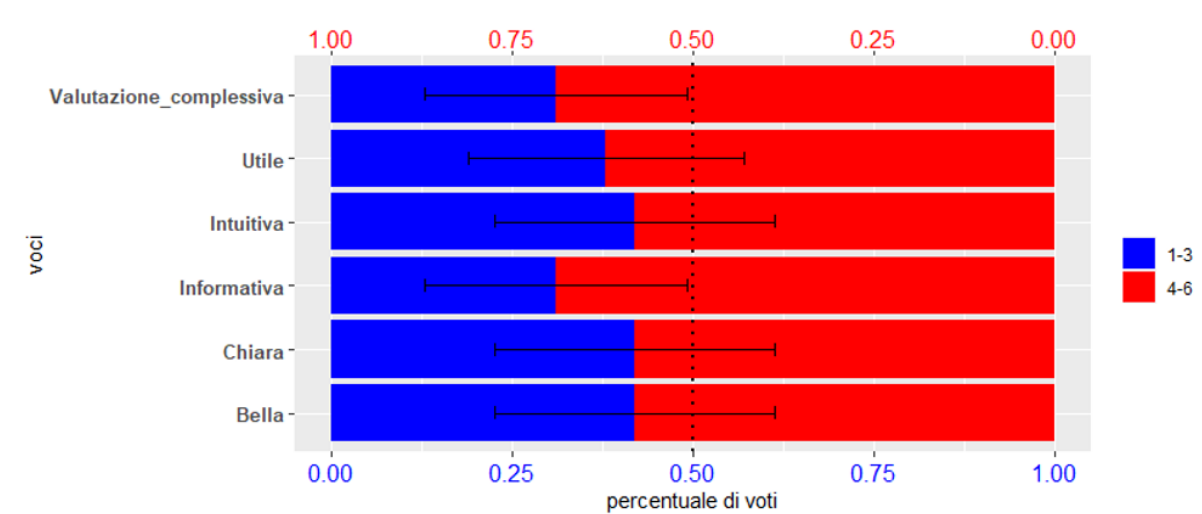


Figura 12

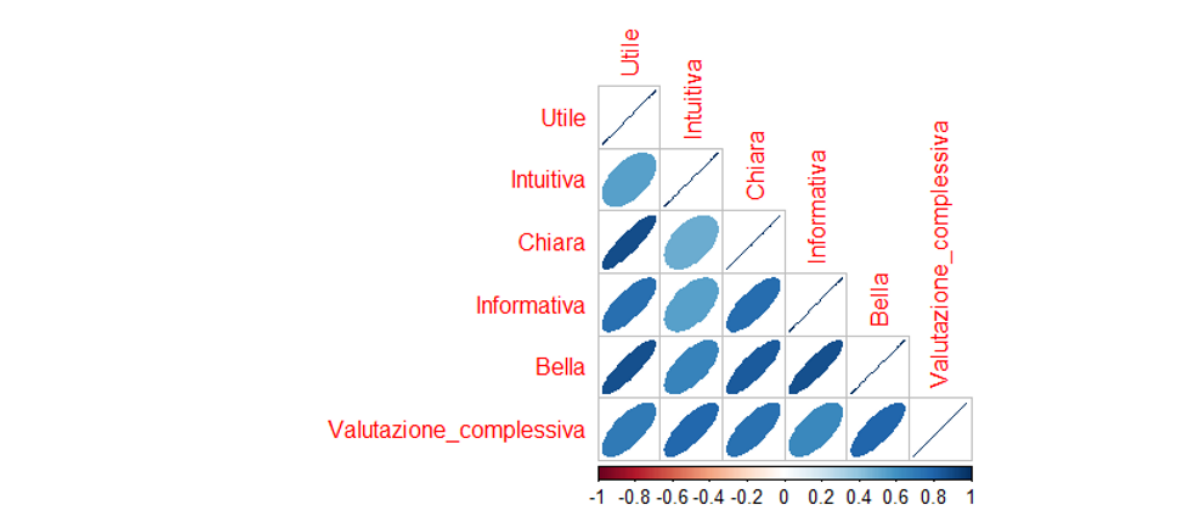


Figura 13

I risultati dei voti hanno proporzioni lievemente differenti, tra le quali si possono defilare la valutazione complessiva e l'aspetto informativo con una proporzione di negativi significativamente inferiore alla metà dei voti. La correlazione chiarezza-utilità, bellezza-utilità e bellezza-informatività risultano essere le più elevate.

## 5. Valutazione della fruizione degli utenti

Nell'ottica di effettuare una analisi della fruizione e della chiarezza all'insegna della comprensione generale del fenomeno, sono state poste domande molto generali che tuttavia hanno sortito ottimi risultati sullo stesso campione (del test psicometrico). È stato somministrato questo breve questionario composto da 8 domande. Di seguito verranno analizzate le effettive interazioni con le visualizzazioni e, per avallare determinante ipotesi che stimino anche il tempo di risposta e non solo la correttezza di essa, sono stati misurati gli intervalli tra l'istante in cui si pone la domanda e la effettiva risposta, sia essa errata o corretta. Il tratto orizzontale nero posto presso ogni violin plot sta ad indicare il tempo ottimale di risposta che è stato calcolato come il tempo medio impiegato dai tre componenti del gruppo. Questo tempo ottimale deve essere interpretato come un riferimento rispetto ai tempi impiegati da altri fruitori "esterni" al progetto.

### 5.1 PRIMA INFOGRAFICA

- Quando pensi indicativamente il prezzo ha raggiunto il massimo?
- Quando pensi indicativamente che il sentimento sia stato positivo?

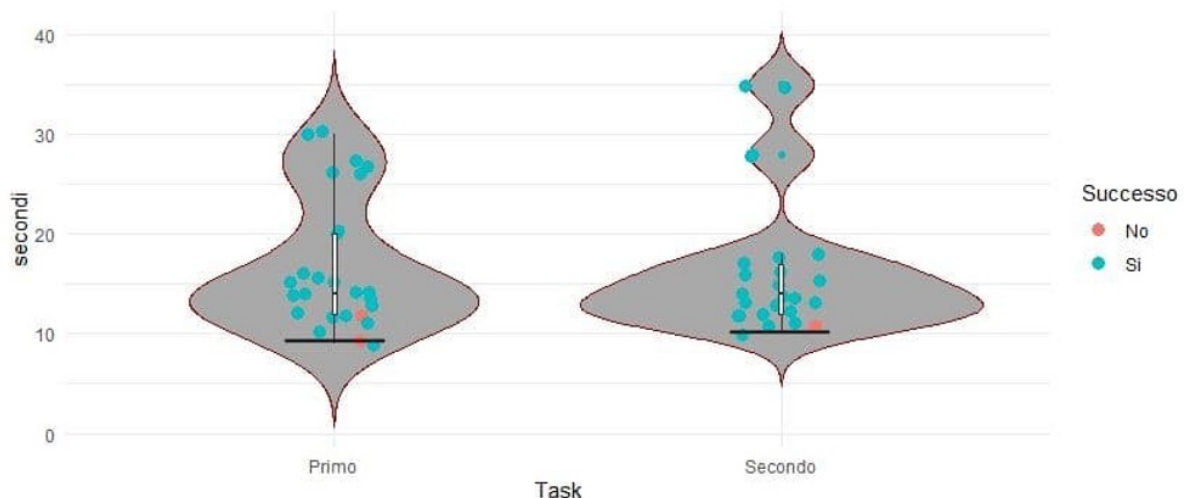


Figura 14

Una grande percentuale dell'utenza impiega circa 15 secondi per rispondere. Specie la seconda domanda, che ha, per sua natura, tempi di ricerca più elevati e sortisce una certa insicurezza nell'utente che temporeggia, il secondo task ottiene un esito migliore e l'errore è 2 per la prima interrogazione e 1 per la seconda.



## 5.2 SECONDA INFOGRAFICA

- Quando pensi che il volume ha raggiunto il suo massimo?
- Quando pensi che il prezzo ha raggiunto il suo massimo?

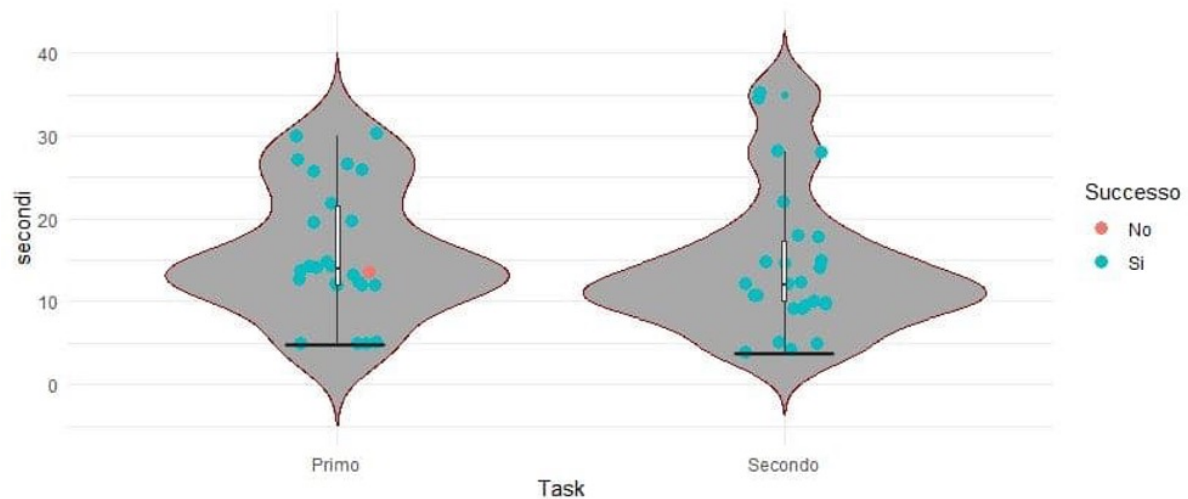


Figura 15

La grande maggioranza dell'utenza impiega circa 12 secondi per rispondere. La comprensione delle domande relative ai due task richiede intervallo di tempo differenti: la seconda domanda comporta anche ragionamento e capire quale sia l'asse giusto da guardare, l'utente rimane titubante sulla porzione del grafico che è fitto di informazioni, tuttavia l'errore è inferiore rispetto alla prima valutazione e si attesta a 1 e 0 rispettivamente.

### 5.3 TERZA INFOGRAFICA

- Quale ritieni, nel mese di novembre, sia stato l'hashtag più popolare?
- Quando ritieni che l'hashtag BTC sia quasi popolare quanto l'hashtag bitcoin?

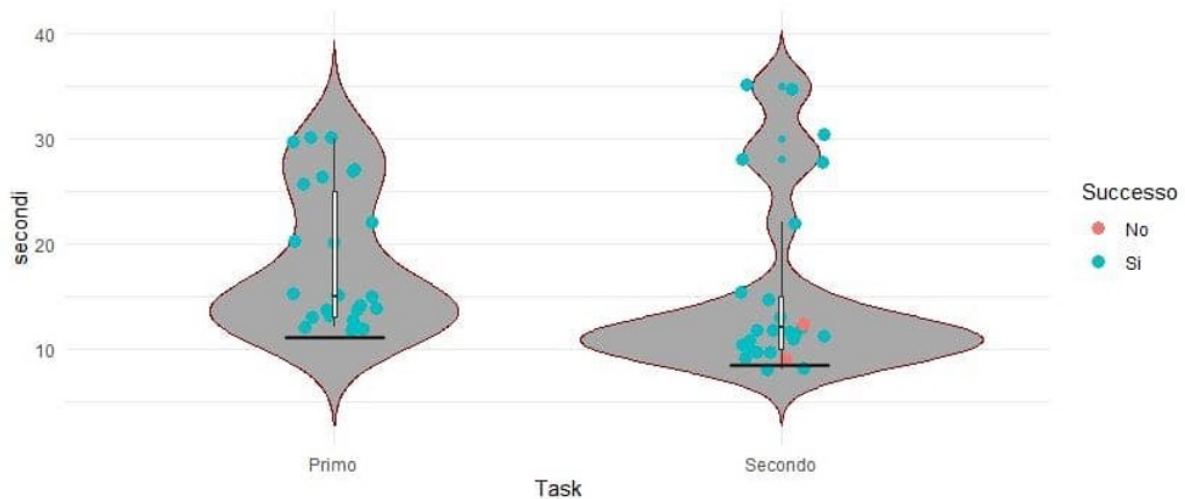


Figura 16

La grande maggioranza degli utenti impiega meno di 20 secondi per rispondere. Probabilmente la pienezza di questa infografica non permette una immediata comprensione a discapito del tempo necessario. Si nota tuttavia una differenza tra i due task, per il secondo i tempi di risposta sono mediamente più rapidi, mentre per il primo task il massimo è 30 secondi, probabilmente perché l'hashtag corrispondente suscita ad alcuni utenti difficoltà nella ricerca. Di compenso gli errori registrati sono bassi. Dunque, la visualizzazione si conferma chiara.

## 5.4 QUARTA E QUINTA INFOGRAFICA

- Quando ritieni che gli utenti positivi siano stati molti, indica due date?
- Quando ritieni che il numero degli utenti negativi siano stati molti, indica la data con il valore massimo?

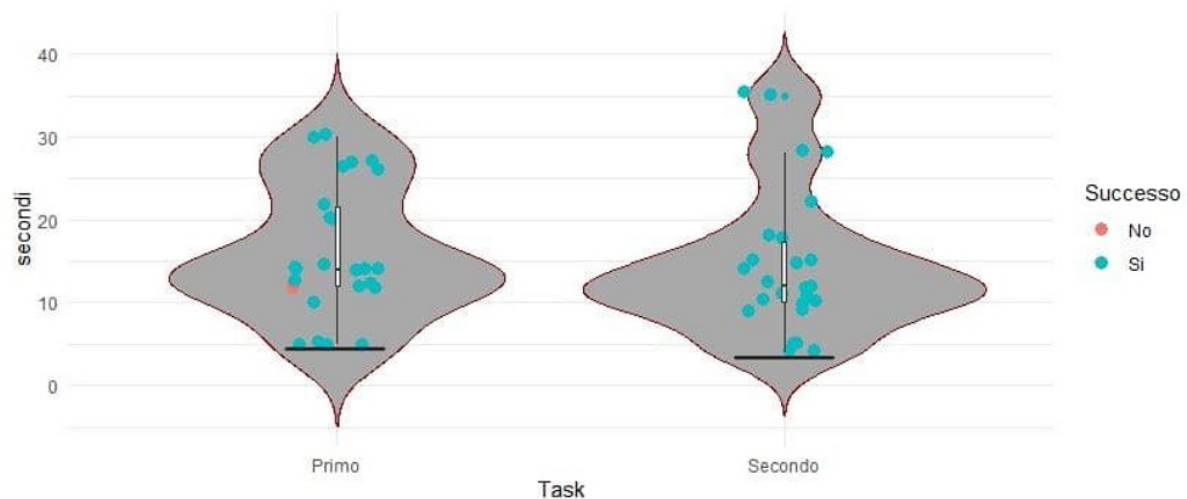


Figura 17

La maggioranza degli utenti impiega meno di 10 secondi per rispondere. Tuttavia, l'aver compreso la prima visualizzazione comporta automaticamente la comprensione anche della seconda. L'errore si attesta a 1 solo utente, il quale ha scambiato il riferimento temporale di due bin posti pressoché alla medesima altezza.

## Analisi successo e insuccesso del completamento dei task



Figura 18

Infine, una valutazione complessiva dimostra come l'errore si sia attestato attorno all'4 %, avendo fissato una soglia di tolleranza del 10%, dato il dominio piuttosto complesso e specifico, si può ritenere una generale buona riuscita dell'analisi. Si può considerare che un errore così basso sia dovuto anche alla presenza di molto context e al fatto che l'utenza ha avuto modo, prima dell'avvio del test, di familiarizzare con le infografiche.

## 6. Link a Tableau Public

Per quanto le nostre visualizzazioni siano statiche, si riporta un link con l'insieme completo delle infografiche dalle quali è possibile trarre ulteriori informazioni selezionando oggetti o aree con il cursore.

### **LINK ALLE PRIME 4 VISUALIZZAZIONI<sup>7</sup>**

[https://public.tableau.com/profile/emanuele7773#!/vizhome/BitCoinCharts\\_def/Storia1?publish=yes](https://public.tableau.com/profile/emanuele7773#!/vizhome/BitCoinCharts_def/Storia1?publish=yes)

### **LINK ALLA VISUALIZZAZIONE FINALE PER L'HASHTAG**

[https://public.tableau.com/profile/emanuele7773#!/vizhome/Tweecoin\\_hashtag/hashtag?publish=yes](https://public.tableau.com/profile/emanuele7773#!/vizhome/Tweecoin_hashtag/hashtag?publish=yes)

---

<sup>7</sup> Il troncamento dell'asse è stato effettuato con un'altra modalità in quanto si è adoperato un software esterno a Tableau per tale scopo.