Università degli Studi di Milano-Bicocca

Progetto Finale di *Data Visualization*

"Just Repeat-Hit"

Misurare la ripetitività della musica
tramite la compressione testuale

Valutazione della qualità delle infografiche

Borroni Alessandro / mat. **800069** / a.borroni2@campus.unimib.it
Giugliano Mirko / mat. **800226** / m.giugliano@campus.unimib.it
Prade Angela / mat. **838540** / a.prade@campus.unimib.it

Introduzione

Il progetto nasce dall'idea di verificare se la *ripetitività dei testi delle canzoni sia in aumento* con il passare del tempo. Dopo aver calcolato per ogni testo un indice di ripetitività, si è cercato di verificare dapprima se ci fosse un trend di ripetitività in crescita negli ultimi anni e successivamente si è cercato di verificare quale sia il genere o i generi che trainano questo trend. Una volta completato il lavoro di raccolta dati, di pre-processing ed elaborazione e analisi dei dati, sono state create delle infografiche e, successivamente, ne è stata valutata la qualità sottoponendole al parere degli utenti. Il software utilizzato per implementare le visualizzazioni è stato *Tableau*, mentre per i grafici riguardanti la valutazione delle infografiche è stato utilizzato *R*.

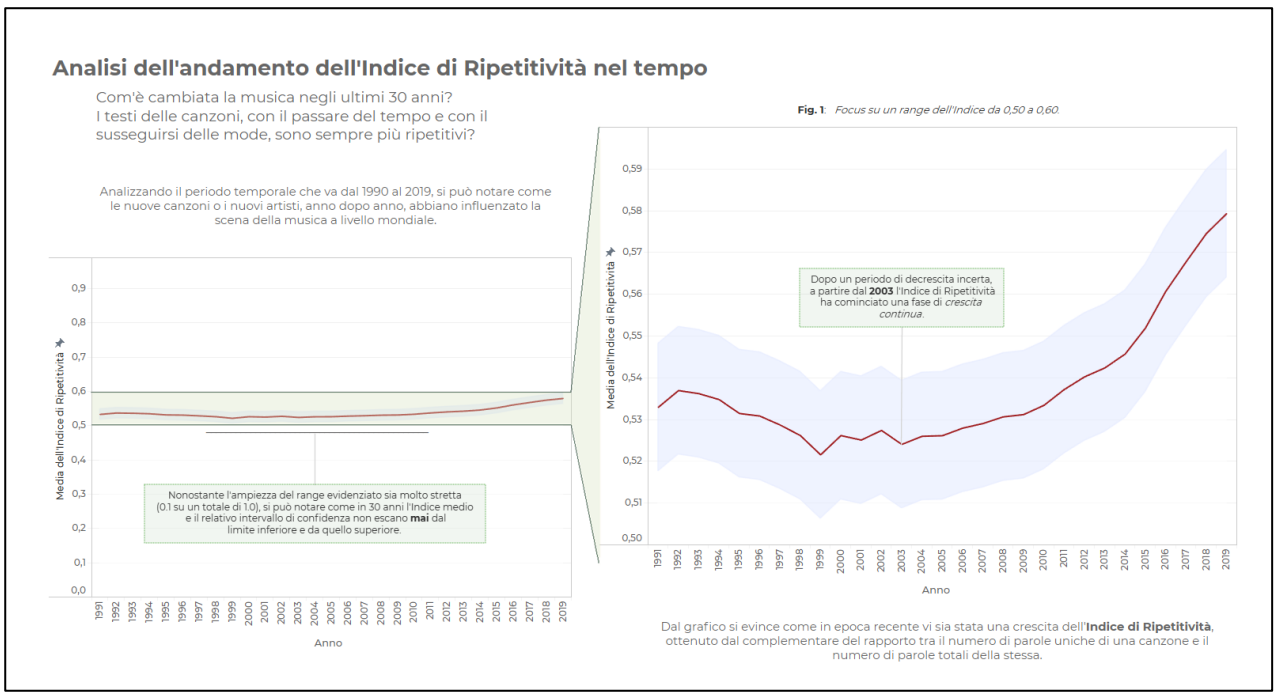
Sono state sviluppate, in totale, 5 infografiche:

- **Infografica 0** Introduzione e Descrizione dell'Indice di Ripetitività
- **Infografica 1** Analisi dell'andamento dell'Indice di Ripetitività nel tempo
- **Infografica 2** Indice di Ripetitività per Genere
- **Infografica 3** Distribuzione dell'Indice di Ripetitività
- **Infografica 4** Analisi della ripetitività delle singole canzoni per artista

Think aloud

Prima di procedere con i questionari psicometrici e con i task da sottoporre agli utenti, è stato chiesto a *tre* persone di esprimere a voce alta il loro pensiero mentre interagivano con l'infografica in modo da poter comprendere gli aspetti della User Experience che potessero creare delle difficoltà.

- **Infografica 1**



Il problema che è emerso mostrando questa visualizzazione è il fatto che agli utenti non risultasse immediatamente chiaro che il grafico a destra rappresentasse le stesse informazioni del grafico a destra ma su una scala diversa. Per ovviare a questo problema, si è deciso di specificare in modo più visibile cosa rappresentasse il grafico a destra e di evidenziare tramite un segmento colorato la porzione di asse y che assume i valori da 0.5 a 0.6 del line chart a sinistra e l'asse del line chart a destra. Inoltre, è stata aggiunta nel line chart di sinistra una linea indicante l'Indice medio di Ripetitività misurato su 280 articoli di giornale e nel line chart l'indicazione dei valori di massimo e di minimo assunti dall'indice stesso. Il risultato ottenuto dopo le modifiche è mostrato nella figura successiva.

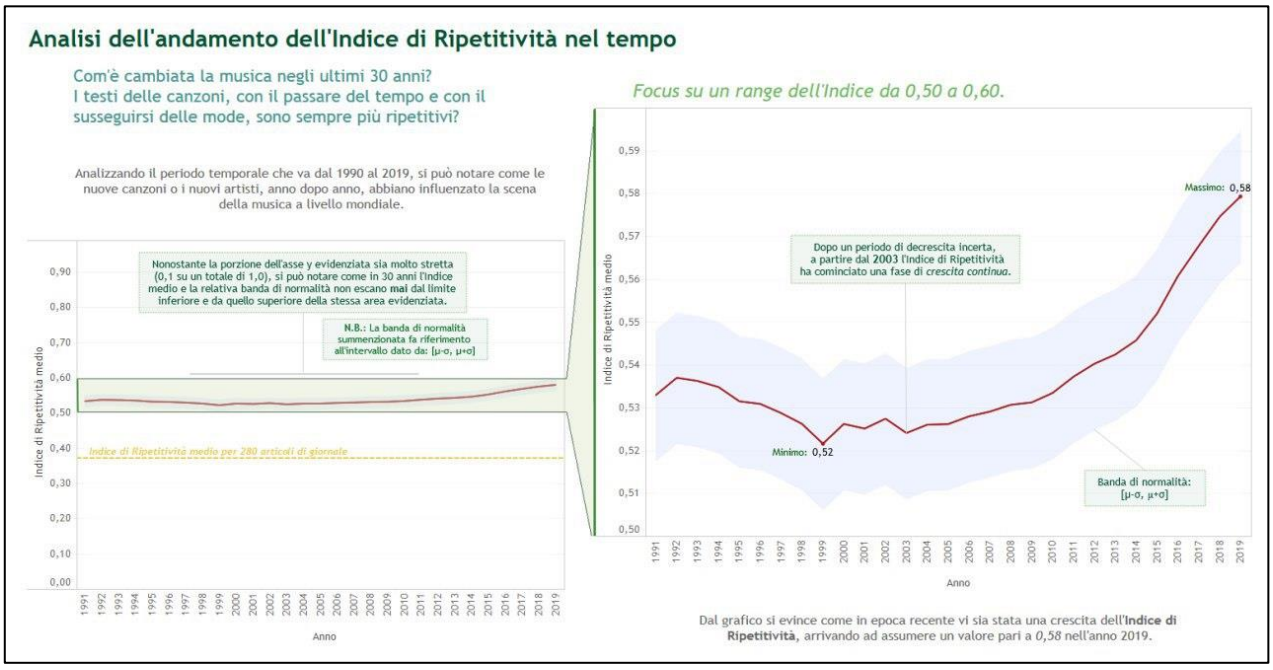
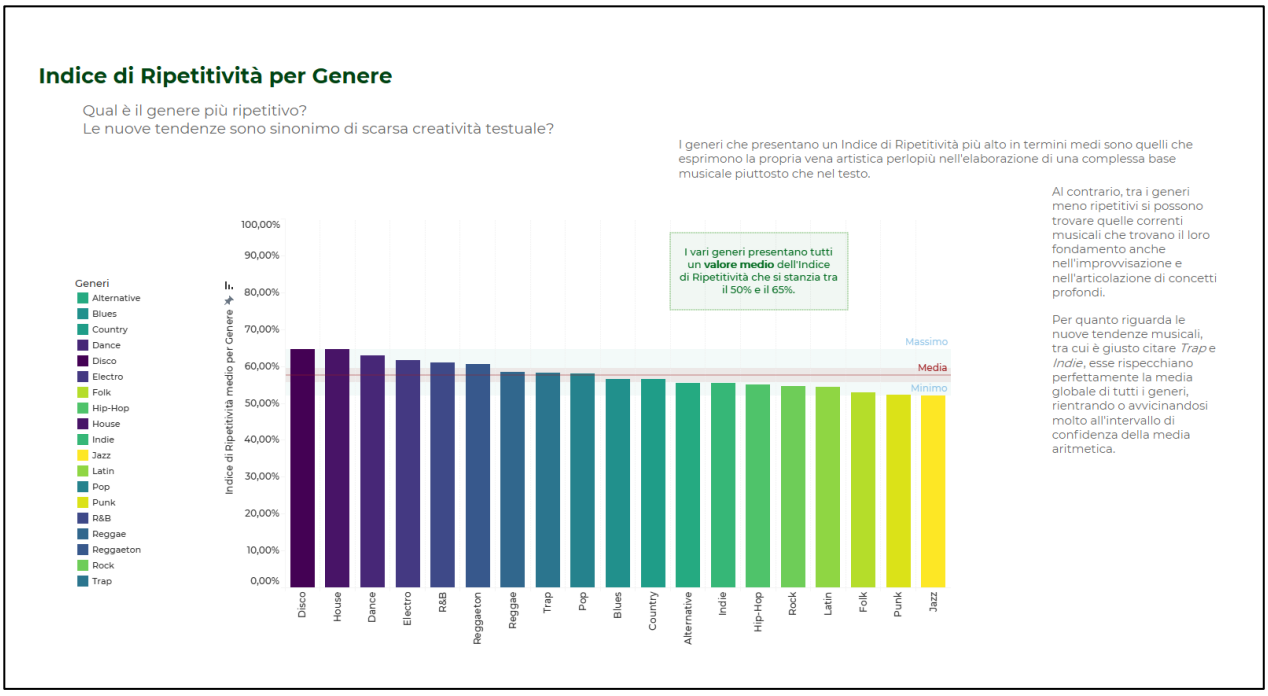
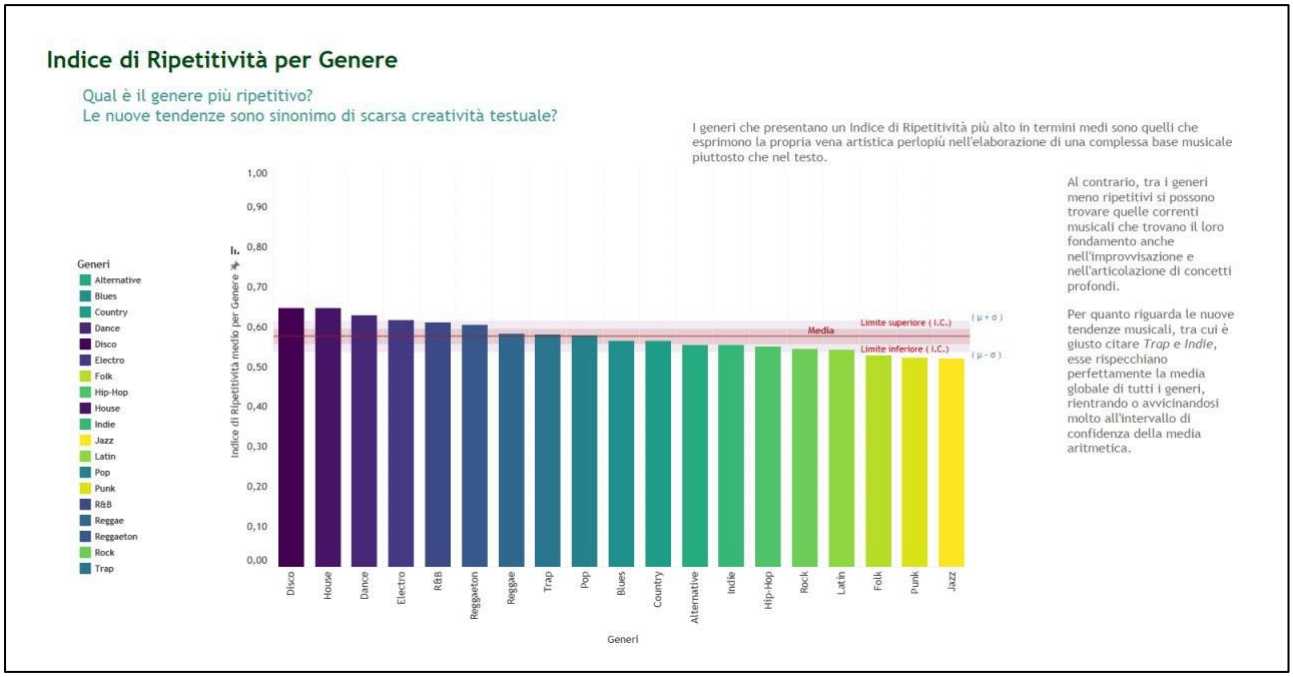


Figura 1

- **Infografica 2**



L'approccio degli utenti con questa Infografica ha evidenziato come l'indicazione di "Massimo" e "Minimo" sia stata ritenuta non rilevante. Inoltre, la banda di colore rosa intorno alla media ha creato confusione. Si è preferito quindi eliminare l'indicazione di "Massimo" e "Minimo", indicare l'utilità della banda rosa e aggiungere la banda di normalità data da $[\mu - \sigma, \mu + \sigma]$.



Successivamente sono stati aggiunti gli intervalli di confidenza per ogni singolo genere e, al fine di renderli più evidenti agli utenti, si è deciso di troncare l'asse y evidenziando solo i valori appartenenti al range [0.30, 0.80].

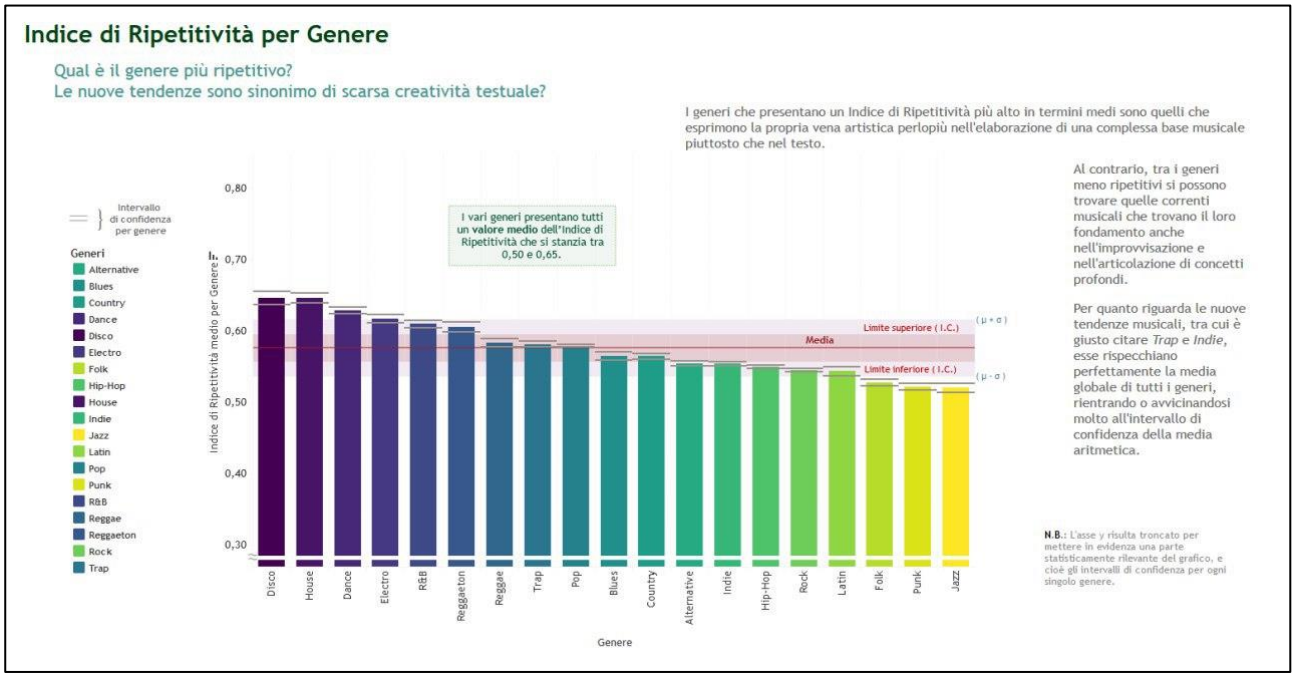
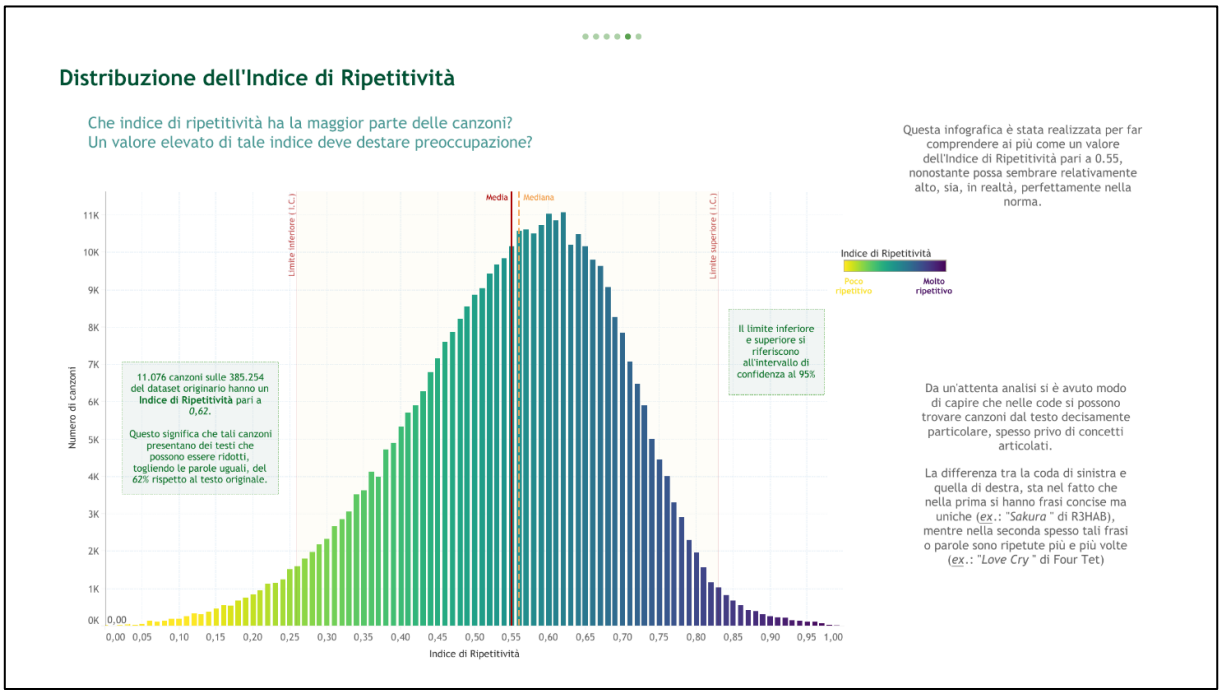


Figura 2

- **Infografica 3**



Inizialmente si è deciso di calcolare l'intervallo di confidenza al 95% intorno alla media e di riportarlo sul grafico. Utenti più esperti che hanno interagito con l'infografica ci hanno fatto notare la stranezza del valore di tale intervallo portandoci a individuare un errore nel calcolo.

Visto che l'intervallo corretto sarebbe stato eccessivamente stretto e non avrebbe portato ad alcuna informazione aggiuntiva, si è scelto di indicare la banda di normalità $[\mu - \sigma, \mu + \sigma]$. Il risultato è quello riportato nella figura successiva.

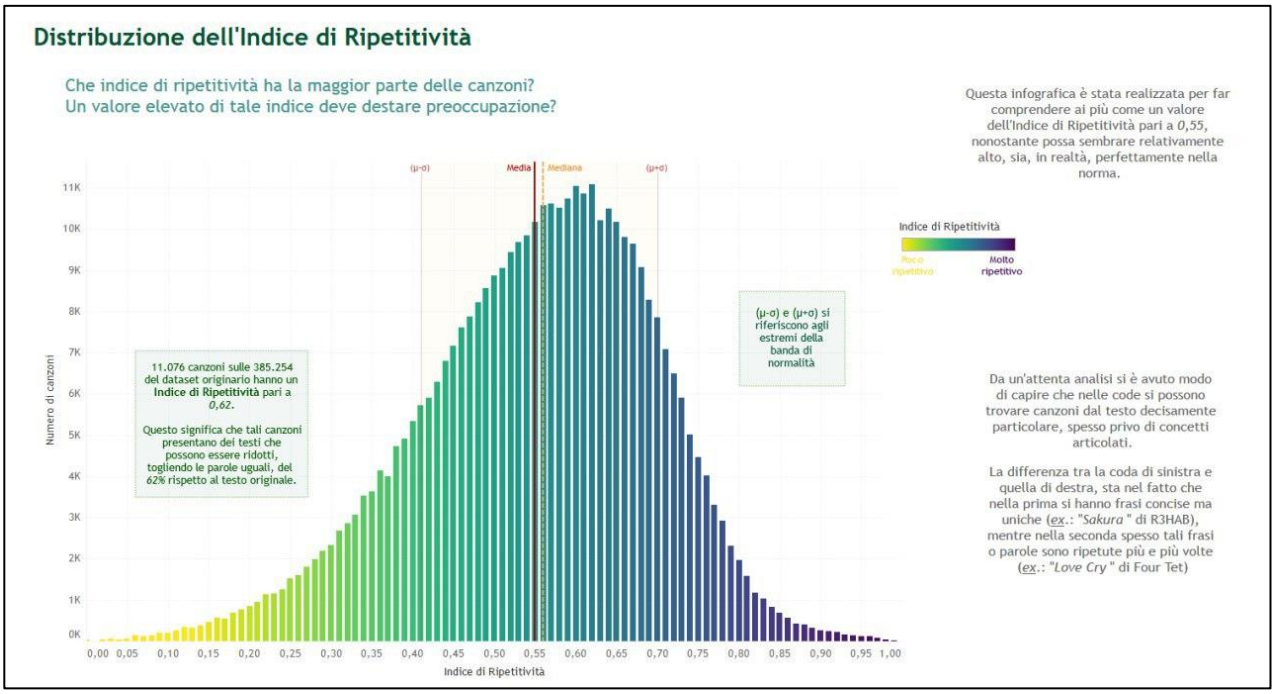


Figura 3

- **Infografica 4**

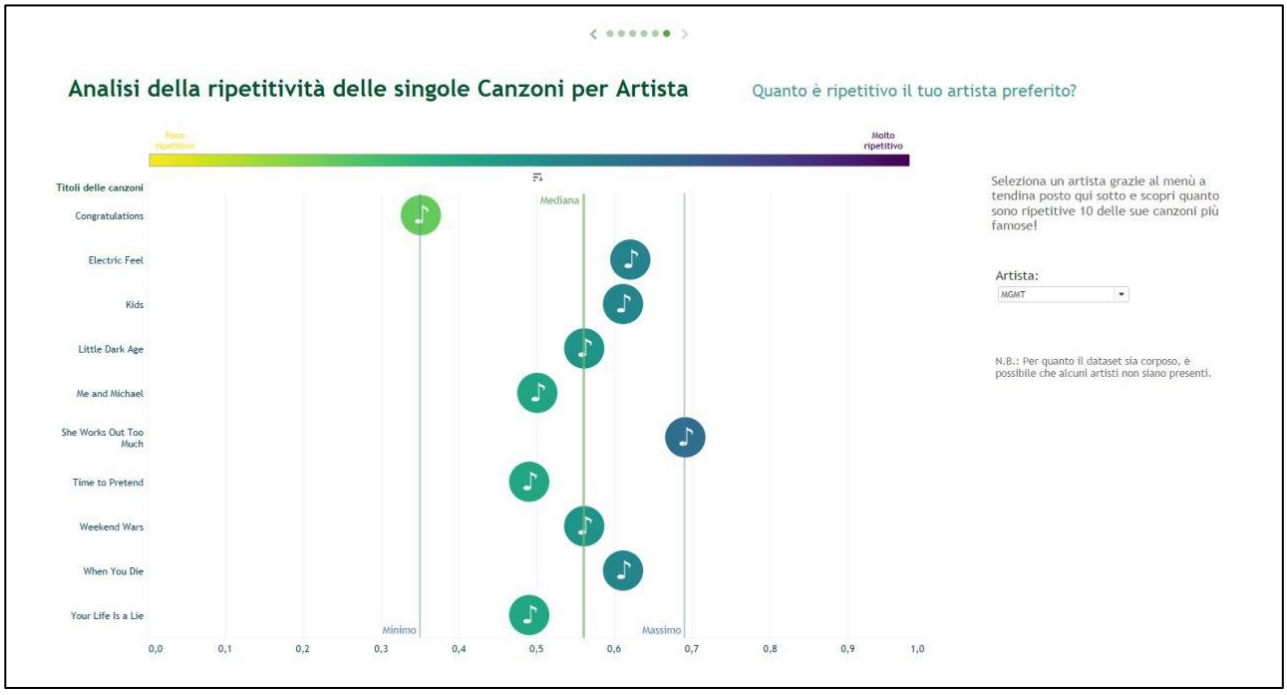
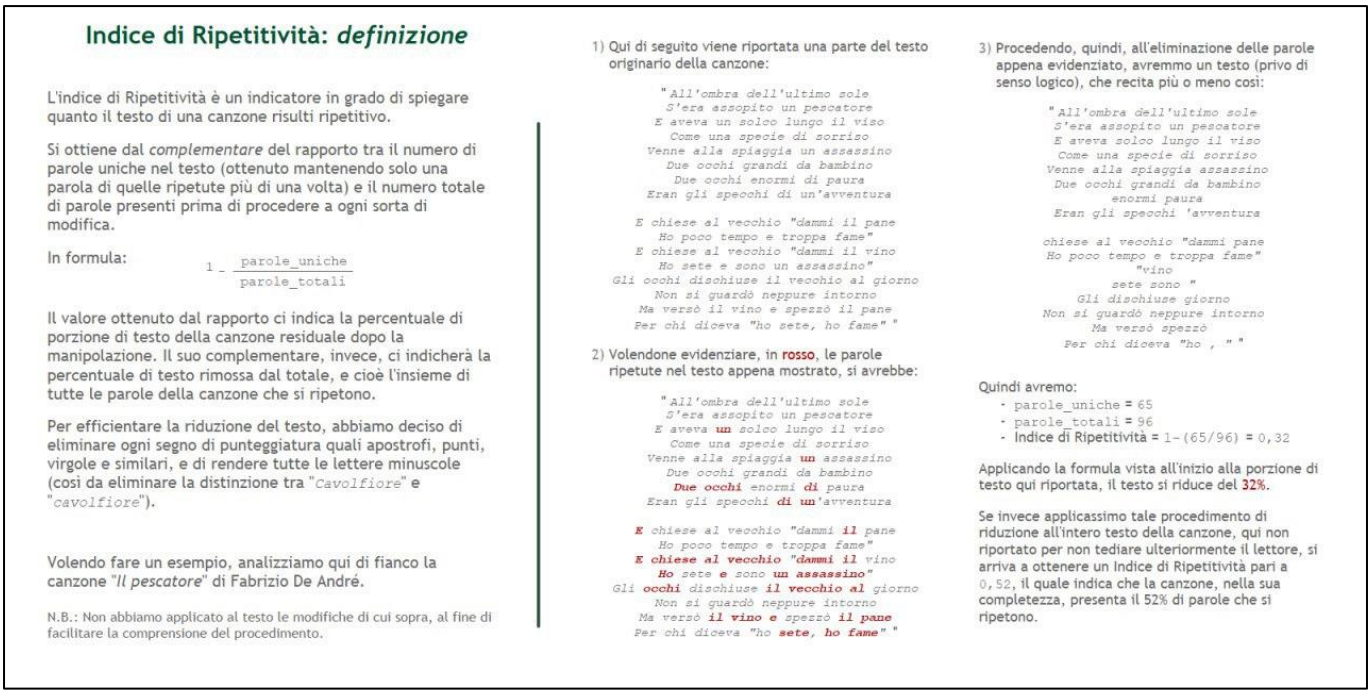


Figura 4

L'infografica 4 non ha presentato particolari difficoltà di comprensione da parte degli utenti.

Inoltre, vista la curiosità dei soggetti intervistati nel comprendere come sia stato calcolato l'indice di ripetitività, si è scelto di inserire un'infografica iniziale nella quale venisse approfondito questo aspetto.



N.B.: Le Figure 1, 2, 3 e 4, e cioè le versioni finali delle infografiche, sono state successivamente mostrate agli utenti per completare la sezione "*Task*".

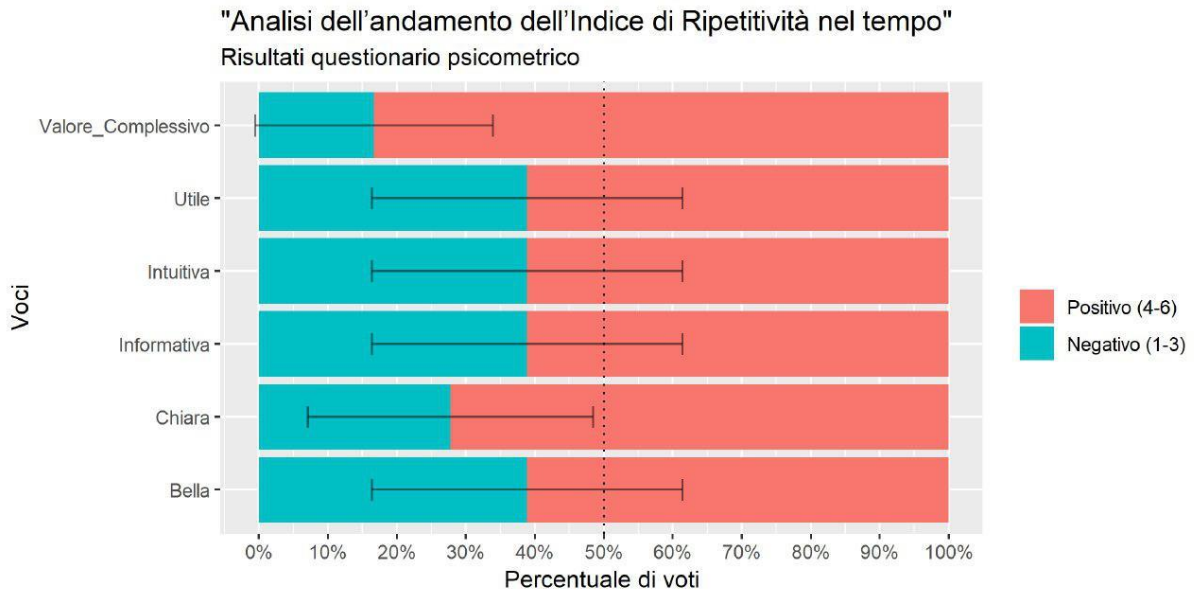
Questionario psicometrico

Sono inoltre stati sottoposti ad ulteriori *18* utenti i questionari psicometrici. Una volta visualizzata l'infografica, agli utenti venivano sottoposti 5 quesiti:

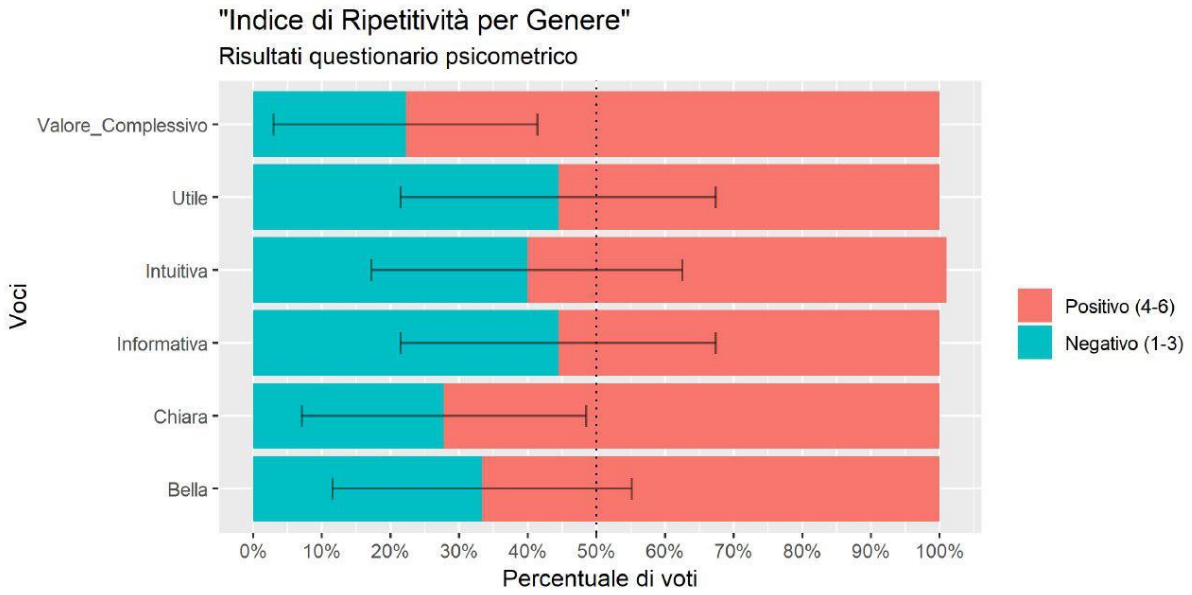
- Quanto ritieni l'infografica Utile?
- Quanto ritieni l'infografica Intuitiva?
- Quanto ritieni l'infografica Chiara?
- Quanto ritieni l'infografica Informativa?
- Quanto ritieni l'infografica Bella?
- Qual è il valore complessivo?

Di seguito vengono riportati i risultati ottenuti attraverso gli stacked barchart.

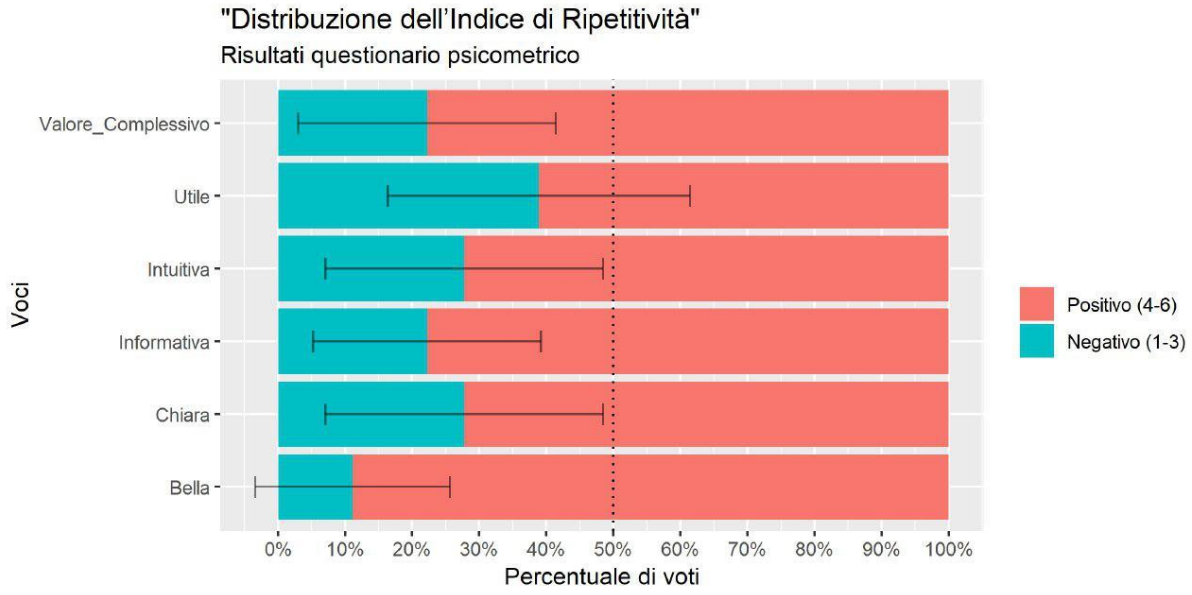
◦ **Infografica 1**



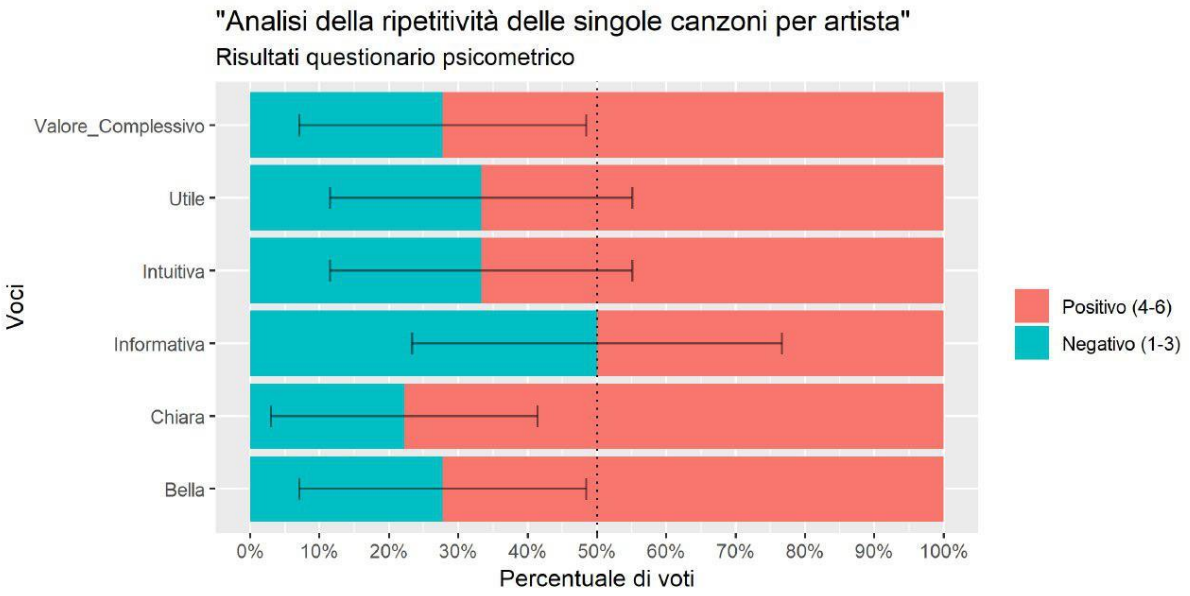
- **Infografica 2**



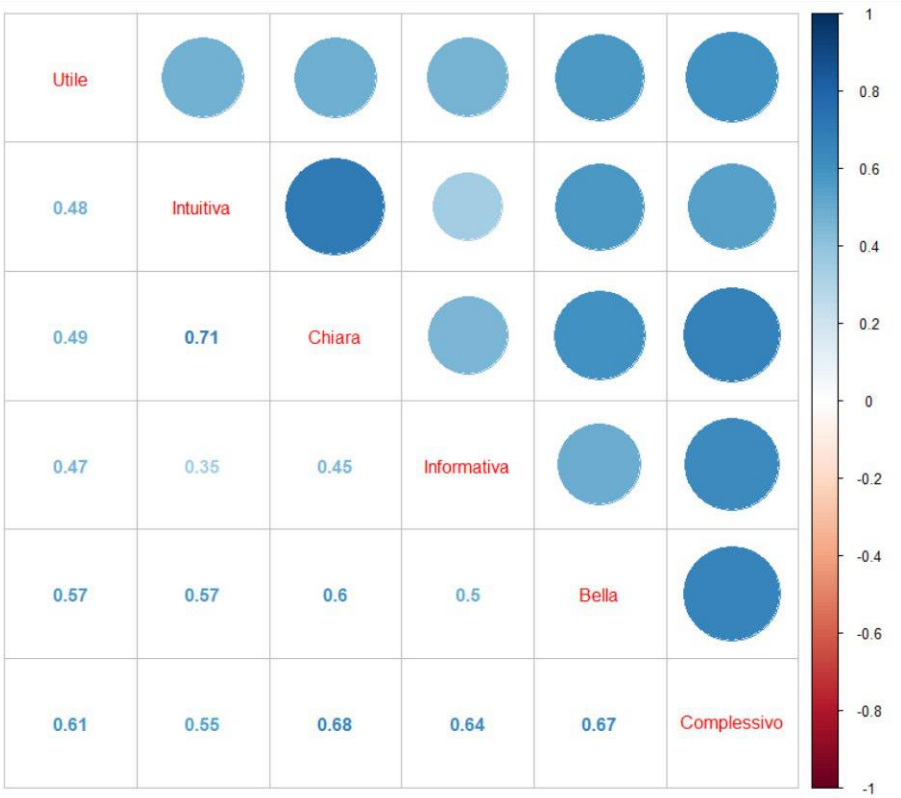
- **Infografica 3**



- **Infografica 4**



Si è inoltre verificato se esistesse correlazione tra le risposte date ai questionari.



Gli indici di correlazione più alti si presentano nelle coppie *complessivo - bella*, *complessivo - utile*, e *complessivo - informativa*, a testimoniare l'apprezzamento sia estetico che di contenuto.

Task

Successivamente a 12 utenti sono stati sottoposti dei task, ovvero delle domande le cui risposte potevano essere dedotte osservando l'infografica. Questo è stato fatto al fine di verificare quanto le informazioni che si volevano trasmettere attraverso l'infografica fossero effettivamente percepite.

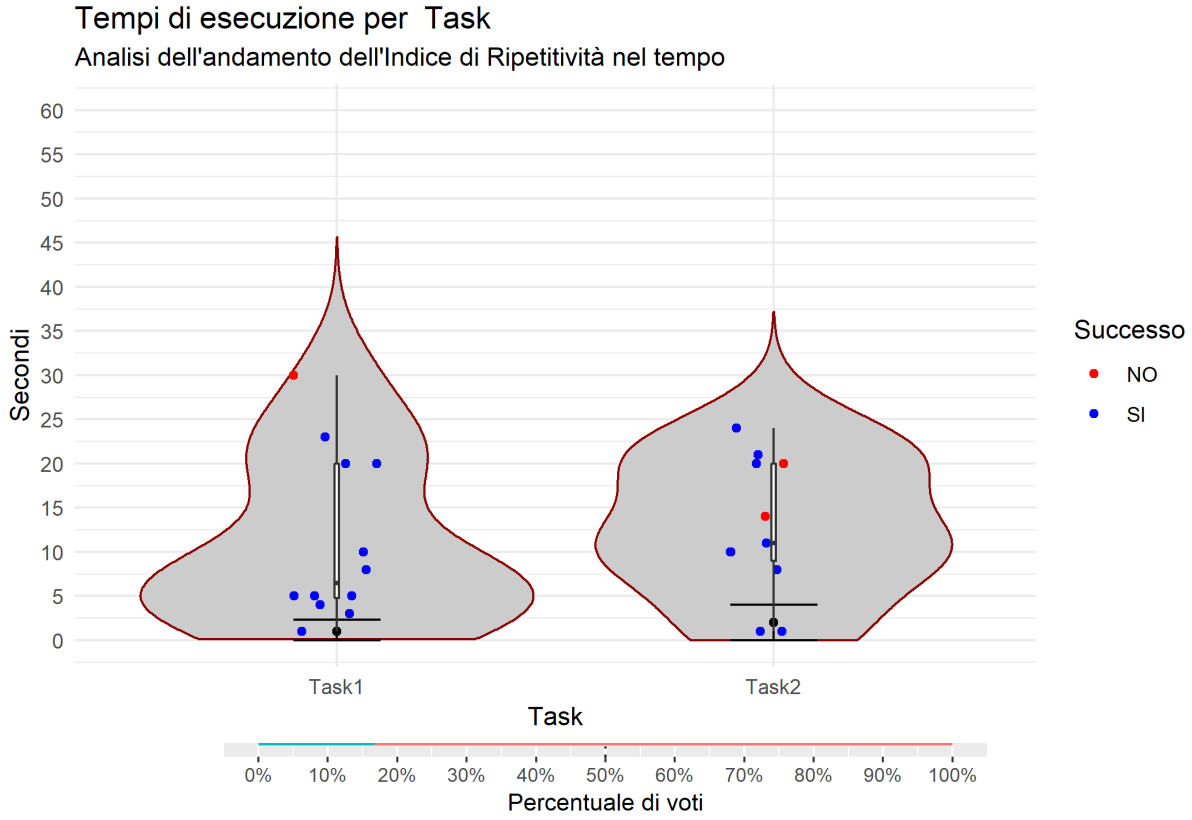
Successivamente a questi utenti sono stati sottoposti i questionari psicometrici.

Si riportano di seguito dapprima i violin plot, i quali riportano i risultati dei task, e subito dopo vengono mostrati gli stacked barchart indicanti le risposte ai questionari psicometrici.

- **Infografica 1** - *Analisi dell'andamento dell'Indice di Ripetitività nel tempo*

Per la prima visualizzazione i task richiesti agli utenti sono stati:

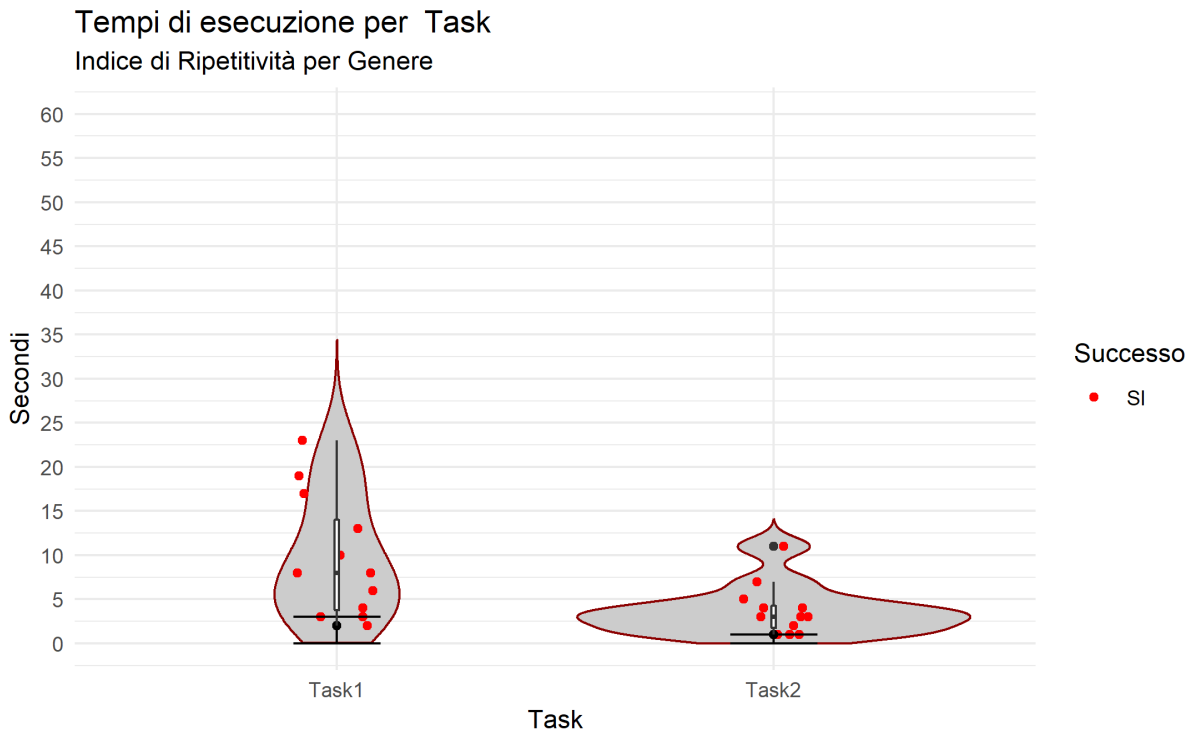
1. Indicare se le tendenze degli ultimi anni presentano qualche trend visibile;
2. Indicare se grafico a destra e quello a sinistra rappresentano l'andamento delle stesse variabili.

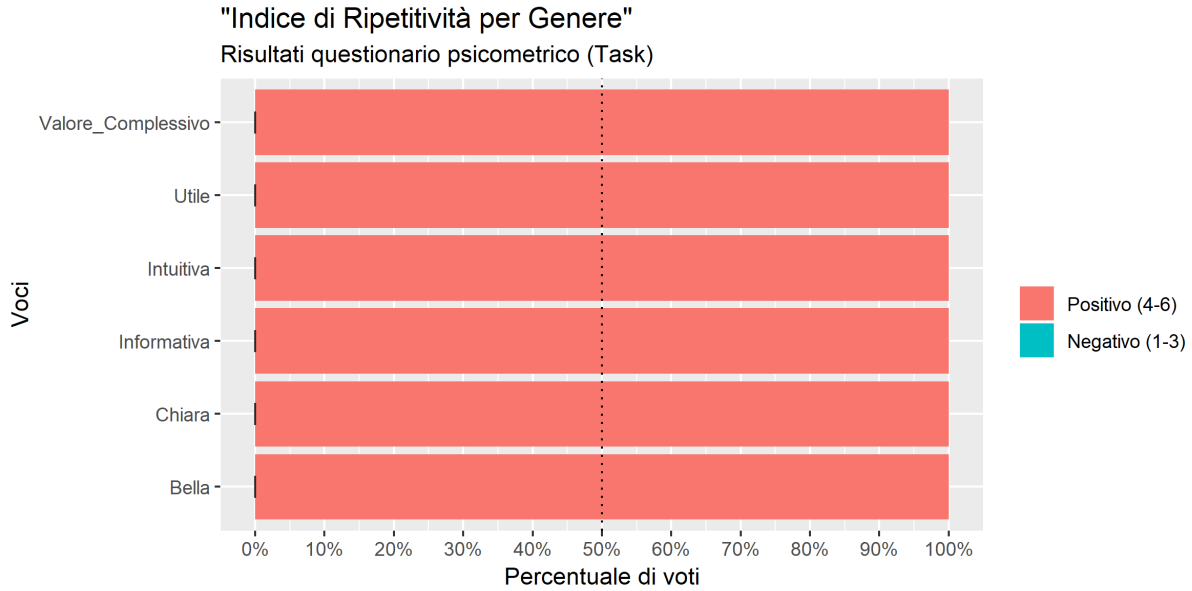


◦ **Infografica 2** - *Indice di Ripetitività per Genere*

Per la seconda visualizzazione i task effettuati sono stati:

1. Indicare la colonna “Trap” e dire se è significativamente più ripetitiva rispetto alla media;
2. Indicare il genere più ripetitivo.

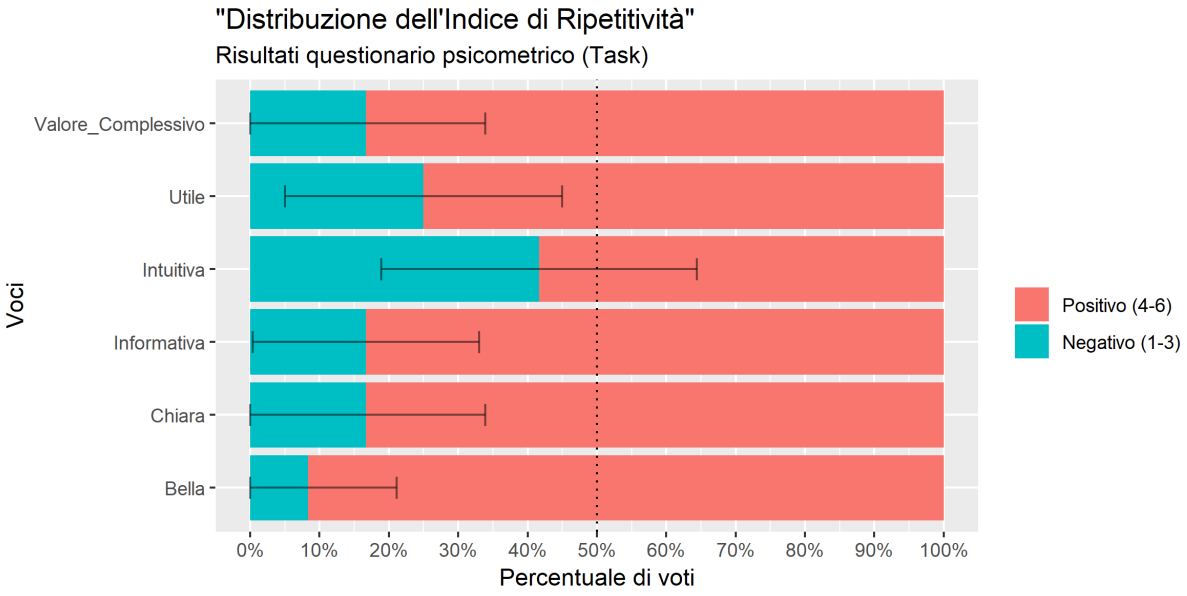


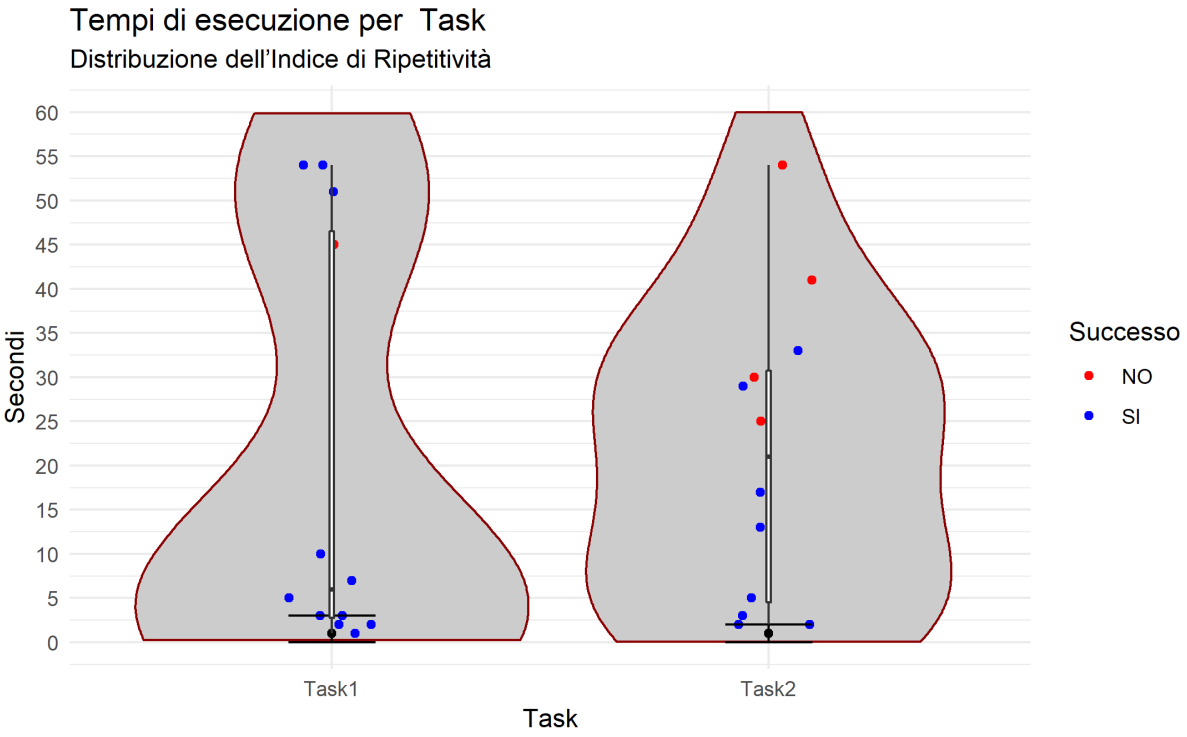


◦ **Infografica 3** - *Distribuzione dell'Indice di Ripetitività*

Per la terza visualizzazione i task effettuati sono stati:

1. Indicare qual è il valore mediano dell'indice di ripetitività;
2. Indicare i limiti della banda di normalità.

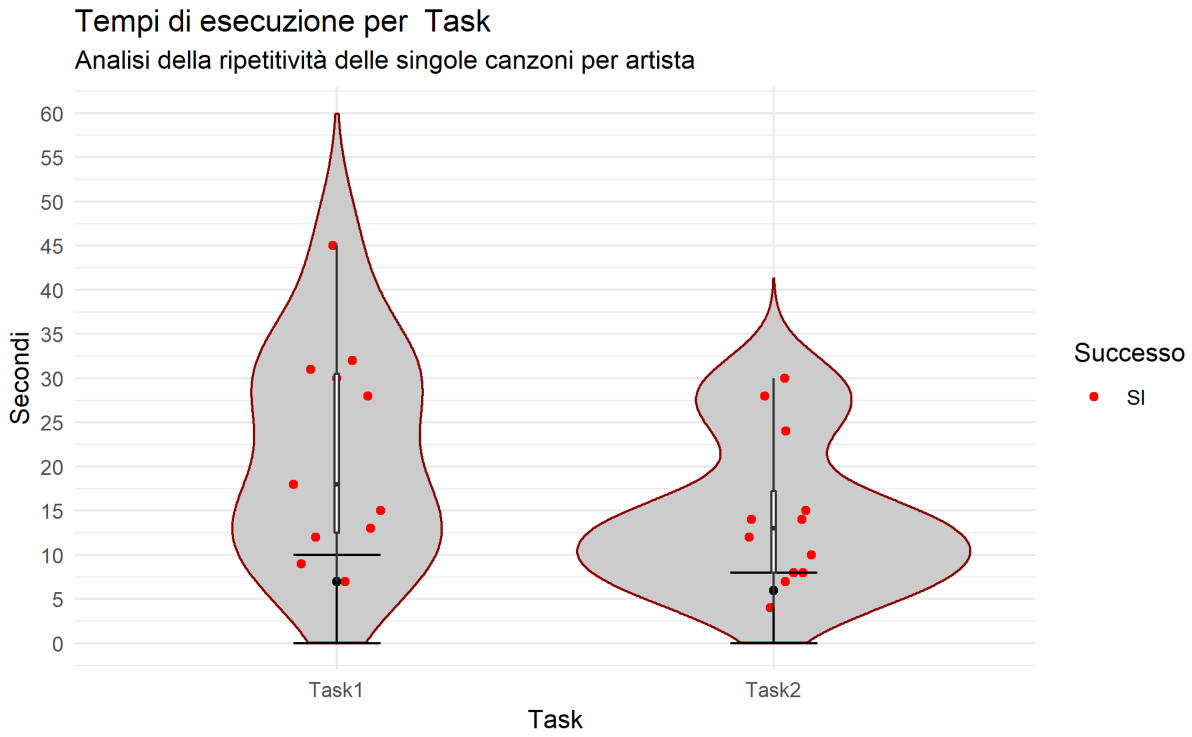


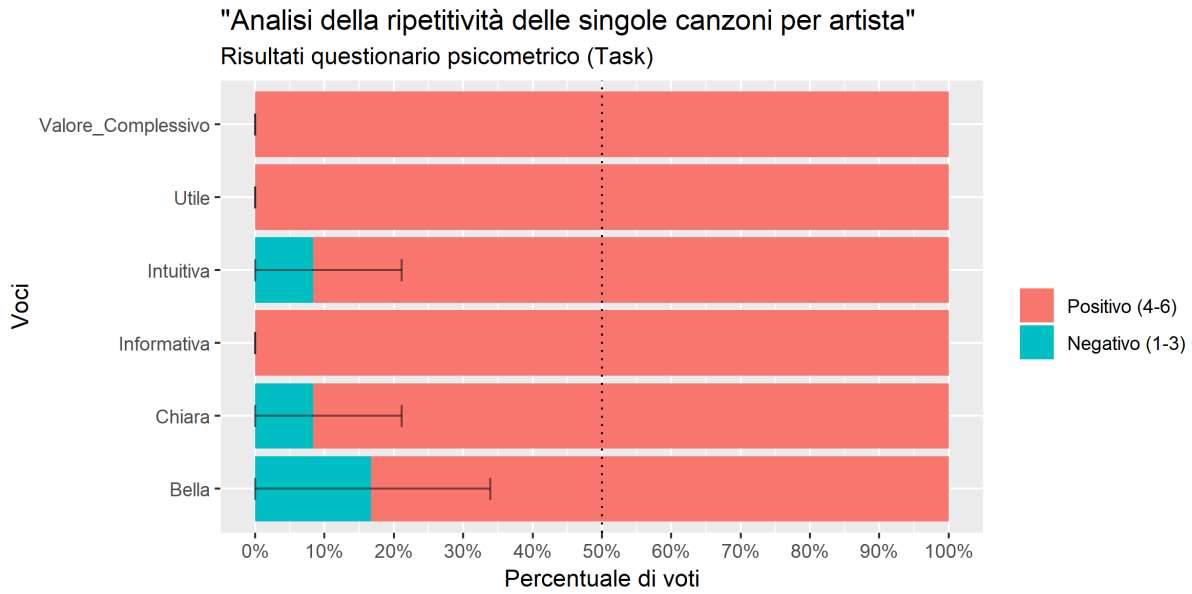


◦ **Infografica 4** - *Analisi della ripetitività delle singole canzoni per artista*

Per la quarta visualizzazione i task effettuati sono stati:

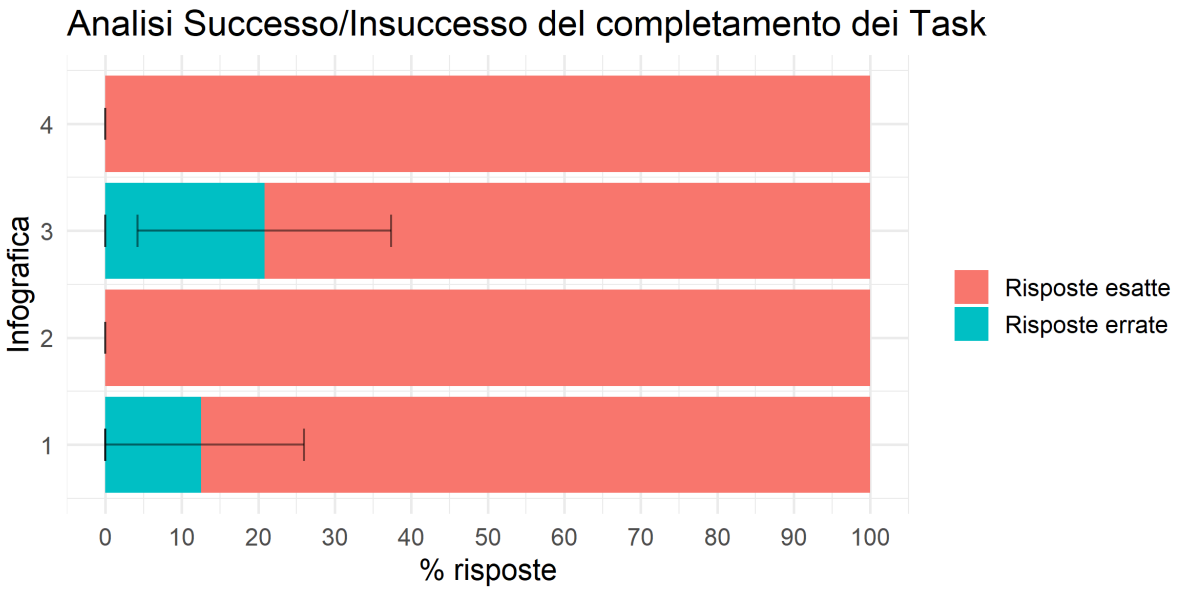
1. Indicare qual è il valore dell'indice di ripetitività della canzone "A casa de Sandro?" di Achille Lauro
2. Qual è la canzone meno ripetitiva di Capo Plaza





È interessante notare come gli utenti a cui è stato richiesto di svolgere i task, abbiano apprezzato maggiormente le infografiche.

Più precisamente, le visualizzazioni più apprezzate sono state la seconda (*Figura 2*) e la quarta (*Figura 4*).



Si può notare come i task relativi alle Infografiche 2 e 4 siano stati tutti completati con successo, mentre per quanto riguarda le Infografiche 1 e 3 vi sono state delle risposte errate. Inoltre, anche la densità di distribuzione dei tempi di risposta emersa dai violin plot si può differenziare tra il gruppo di infografiche 1 e 3 e il gruppo di infografiche 2 e 4. Questa differenza si suppone possa essere dovuta al diverso background degli intervistati, il quale non è stato tuttavia registrato. Si è però notato che utenti con background umanistico facessero fatica a comprendere concetti prettamente statistici/ matematici (quali intervallo di confidenza, variabili e trend), nonostante questi fossero riportati in modo chiaro e ben esposti all'interno delle visualizzazioni stesse, e nonostante la risposta ai task non richiedesse una vera e propria comprensione del concetto ma una semplice comprensione della "legenda" (e cioè saper leggere le indicazioni poste nelle infografiche).

Link alle Infografiche

Qui seguito si riporta il link alla pagina di Tableau in cui è possibile visualizzare una versione interattiva delle Infografiche qui analizzate:

- https://public.tableau.com/views/ProgettodiDataVisualization-I/DataVisualization?:embed=y&:display_count=yes&:toolbar=no&:origin=viz_share_link