

PERCHÉ VACCINARSI ?

Analisi dell'effetto delle vaccinazioni sull'andamento del Covid-19

Andrea Marinoni 799690, Dario Della Mura 793751,
Cesare Colzani 821253

Data Management e Visualization, Settembre 2021



Indice

PERCHÉ VACCINARSI ?	1
1. Introduzione	3
1.1. Obiettivo	4
I vaccini salvano vite	4
2. La raccolta dei dati	5
2.1 Dataset Popolazione	5
2.2 Dataset Covid	5
2.3 Dataset Vaccini	5
2.4 Dataset Indice RT	6
2.5 Integrazione dataset	6
3. Data exploration	7
3.1 BoxPlot	7
3.2 Correlazione	8
3.3 Conclusioni Analisi Correlazione.	11
4. Realizzazione delle infografiche.	12
4.1 Prima Infografica	13
4.2 Seconda Infografica	13
4.3 Terza Infografica	14
4.4 Quarta Infografica	15
4.5 Quinta Infografica	16
4.6 Sesta Infografica	17
4.7 Settima Infografica	18
4.8 Ottava Infografica	19
5. Valutazione della qualità delle visualizzazioni	20
5.1 La valutazione euristica	20
5.2 Modifiche alla visualizzazione in seguito alla valutazione euristica	21
5.3 Test Utente	23
5.4 Questionario psicometrico	26
6. Conclusioni	32
6.1 Sviluppi Futuri	32
7 Fonti	33

1. Introduzione



Figura 1. immagine covid-19

I coronavirus sono una grande famiglia di virus, la maggior parte dei quali sono innocui per l'uomo. Quattro tipi sono noti per causare raffreddori, altri due possono causare gravi infezioni polmonari (SARS e MERS) simili al Covid-19. Il COVID-19 è un virus che colpisce il sistema respiratorio causando un'infezione acuta che può portare sintomi più o meno gravi a seconda dei soggetti e delle circostanze.

Le prove emerse finora sottolineano come l'80% delle persone infettate soffrono solo di una forma moderata o lieve della malattia. Il 15% delle persone colpite sviluppa malattie gravi che richiedono il ricovero in ospedale. Mentre il 5% delle persone si ammala gravemente e di conseguenza necessita di cure intensive. Nonostante le cure intensive necessarie ai casi più gravi siano ampiamente disponibili, il rischio è che la forte pressione sulle strutture sanitarie provocata dall'alto numero di casi, ne possa provocare una paralisi, con conseguente drammatica riduzione della capacità di presa in carico.[1]

1.1. Obiettivo

I vaccini salvano vite

I vaccini proteggono e tutelano da malattie infettive gravi e potenzialmente letali che un tempo uccidevano milioni di persone ogni anno. Si vuole destinare questo progetto a tutti, per far capire l'importanza dell'immunizzazione e il ruolo centrale che in questo periodo ha nella lotta al Covid-19. A tale scopo verrà usato un'insieme di variabili raccolte da vari dataset, aggiornati giornalmente. L'analisi sarà limitata alle variabili sotto riportate, perché ritenute di ruolo cruciale e rappresentative della situazione pandemica.

- Regione¹: andremo ad analizzare i dati dallo scoppio della pandemia per le regioni Lombardia, Veneto, Lazio.
- L'indice di contagio Rt: esprime la quantità media di persone che possono essere contagiate da un solo positivo in un determinato periodo di tempo.
- Numero di persone in terapia intensiva²: quante persone a causa di sintomi molto gravi hanno dovuto ricorrere alla terapia intensiva.
- Numero di decessi : persone decedute a causa Covid-19.
- Vaccinati con prima dose³: numero di persone che hanno ricevuto almeno la prima dose di vaccino.
- Vaccinati con entrambe le dosi⁴: numero di persone che sono vaccinate con entrambe le dosi di vaccino.

L'obiettivo delle infografiche del progetto sarà quello di valutare l'importanza dei vaccini per quanto riguarda l'indice RT, Il numero di persone in terapie intensive e i decessi. Si mostrerà come sono legate tra di loro, il loro andamento e su quale di queste variabili il vaccino ha avuto un maggior effetto. Detto questo bisogna considerare che il vaccino non è l'unico elemento di influenza del trend, sia prima che dopo la sua introduzione ci sono stati periodi di Lock Down e di decisioni dell'apparato statale/regionale che hanno condizionato *temporaneamente*⁵ le variabili scelte.

I quesiti ai quali si vorrebbe rispondere sono i seguenti:

- L'introduzione dei vaccini ha influenzato l'indice RT ?
- L'introduzione dei vaccini ha influenzato l'andamento delle terapie intensive ?
- L'introduzione dei vaccini ha influenzato i decessi ?

Mentre le ipotesi che si vogliono verificare sono:

- L'introduzione dei vaccini ha avuto un'influenza meno significativa sull'indice RT, poiché anche dopo essersi vaccinati è comunque possibile riprendere e trasmettere il Covid-19.
- L'introduzione dei vaccini ha provocato una riduzione dei casi più gravi del Covid-19 (terapie intensive e decessi).

¹ Non verranno rappresentate tutte le regioni causa spazio, si analizzeranno solo quelle dove sono state attuate misure più efficaci: più dosi vaccino fatte, più casi covid, più morti ecc .

² Si è deciso di analizzare questa variabile poiché è il caso più grave dell' numero di totale ospedalizzati(totale ospedalizzati= ricoverati con sintomi + terapia intensiva).

³ Si analizzerà vaccinazioni percentuale dose 1 ottenuta rapportando il numero di vaccini prima dose alla popolazione della regione specifica

⁴ Si analizzerà vaccinazioni percentuale complete ottenute rapportando il numero di vaccinati completi alla popolazione della regione specifica.

⁵ Data exploration foglio 7(codice) : analisi change points dove data la serie storica vengono trovati i punti di inversione di tendenza.

2. La raccolta dei dati

Il database utilizzato per questa rappresentazione è il frutto dell'unione e dell'arricchimento di diversi dati. I dataset sono molto variegati, c'è la presenza di variabili cumulate che si sono dovute differenziare e alcune variabili che era più significativo cumulare (es vaccinazioni). Tutti i dati che si analizzano sono giornalieri, in seguito, dopo averli integrati verranno aggregati per lavorare su dati settimanali.

I database che sono stati integrati tra loro sono rispettivamente:

2.1 Dataset Popolazione [2]

Database fornito dal governo italiano e reso disponibile su Github come open data. Servirà per andare a calcolare i valori in percentuale delle variabili target sulla popolazione in base alla regione. Il dataset in questione contiene i seguenti attributi:

- Nome Area : che corrisponde al nome della regione.
- Totale Popolazione: numero persone nella regione.

2.2 Dataset Covid [3]

Dataset fornito da ISS e reso disponibile su Github come open data, contenente molti attributi utili per l'analisi:

- Regione : nome della regione
- Casi Covid: casi covid per regione, in seguito saranno analizzati rapportati alla popolazione (valore ottenuto dal dataset Popolazione)
- Deceduti: variabile cumulata di decessi totali per ogni regione
- Terapia intensiva

2.3 Dataset Vaccini [4]

Database fornito dal governo italiano e reso disponibile su Github come open data, contenente attributi fondamentali allo scopo dell'analisi:

- Regione : nome della regione
- Numero Vaccinazioni prima dose: quante persone hanno ricevuto la prima vaccinazione (verrà utilizzata questa variabile cumulandola e rapportandola alla popolazione regionale)
- Numero Vaccinazioni seconda dose : persone che hanno ricevuto entrambe le dosi di vaccino(verrà utilizzata questa variabile cumulandola e rapportandola alla popolazione regionale).

2.4 Dataset Indice RT [5]

Dataset ottenuto tramite web-scraping dal sito dell'ISS

- Regione : nome della regione
- Indice RT: indice che indica la contagiosità del virus.

2.5 Integrazione dataset

L'integrazione tra questi 4 dataset è stata effettuata attraverso una libreria di python chiamata Pandas. Si è deciso di partire dal dataset con più righe (descritto al punto **2.2**), il quale ha informazioni riguardo a tutto ciò che è accaduto dal 24-02-2020 fino all'8-08-2021. A quest'ultimo è stato aggiunto il numero di popolazione presente in modo da poter ottenere successivamente valori in rapporto alla popolazione della regione, in seguito è stato aggregato con il dataset con le informazioni riguardanti i vaccini e l'indice RT (rispettivamente descritti al punto **2.3** e **2.4**)

Il dataset finale è stato ottenuto integrando per data e regione.

Per quanto riguarda le analisi, tutte le variabili utilizzate sono state aggregate settimanalmente.

3. Data exploration

3.1 BoxPlot

Dopo aver aggregato settimanalmente il dataset ci si è focalizzati sull'analisi delle regioni d'interesse, in particolare sul come l'indice RT è mutato dal 2020 al 2021.

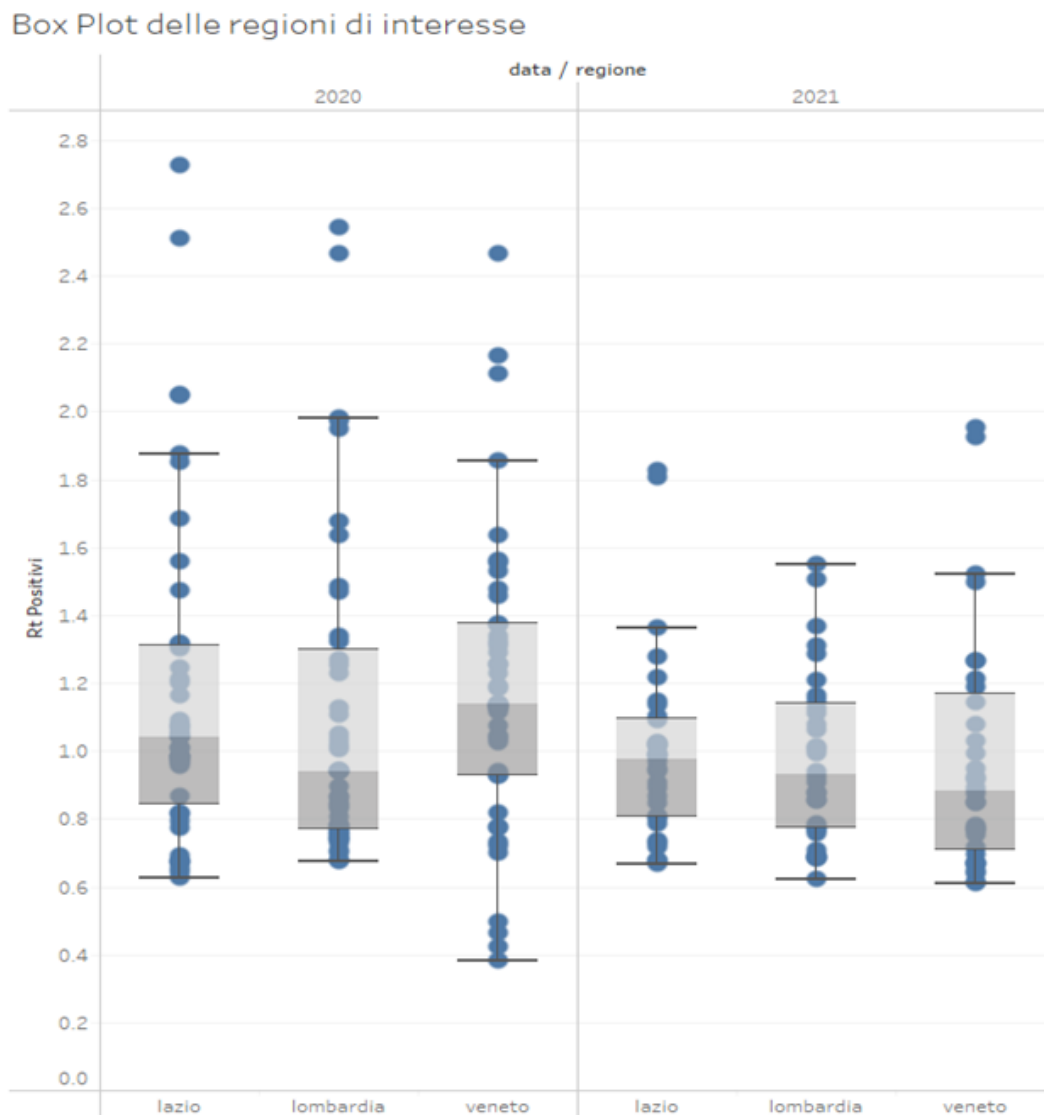


Figura 2. Box Plot regioni

Dai boxplot in questione si nota come l'indice RT sia cambiato dal 2020 al 2021. Questa riduzione visibile, ovviamente, non è data solo dall' introduzione del vaccino, ma anche alle varie manovre del governo e delle regioni .

Quello che salta subito all'occhio è come nell'anno 2020 l'indice RT per ciascuna regione analizzata presenti almeno 2 valori fuori limite (possibili outlier) e come questi valori corrispondono alle date in cui lo stato ha dichiarato il periodo di LockDown generale, a seguito di un indice RT molto elevato.

Spostando lo sguardo nell'anno 2021 si nota subito come le varie regioni sono riuscite a contenere meglio la diffusione del covid mediante il ricorso a nuovi Lockdown (sistema dei colori per regione), controlli più stringenti e, come verrà mostrato, all'introduzione dei vaccini. Questo cambiamento lo si nota nei vari box plot dove il range interquartile è più compatto e i baffi sono più corti.

L'ultima osservazione riguarda la Lombardia, è la regione che tra tutte è riuscita a contenere maggiormente la crescita dell'indice RT, poiché nell'anno 2021 non si notano valori fuori dal limite come le altre due regioni.

3.2 Correlazione⁶

In statistica una correlazione è una relazione tra due variabili. Essa rappresenta la capacità di una variabile di cambiare in funzione di un'altra, seguendo una certa regolarità.

In questo paragrafo si studia la correlazione⁷ per capire in che modo le variabili target (RT, Terapia Intensiva, Decessi) siano correlate alle dosi di vaccino (dose 1 e vaccinazioni complete). Il risultato auspicabile è quello di trovare un'alta correlazione negativa tra vaccini e variabili target, sintomo del fatto che al crescere delle persone immunizzate, le variabili target decrescono in maniera più o meno marcata.

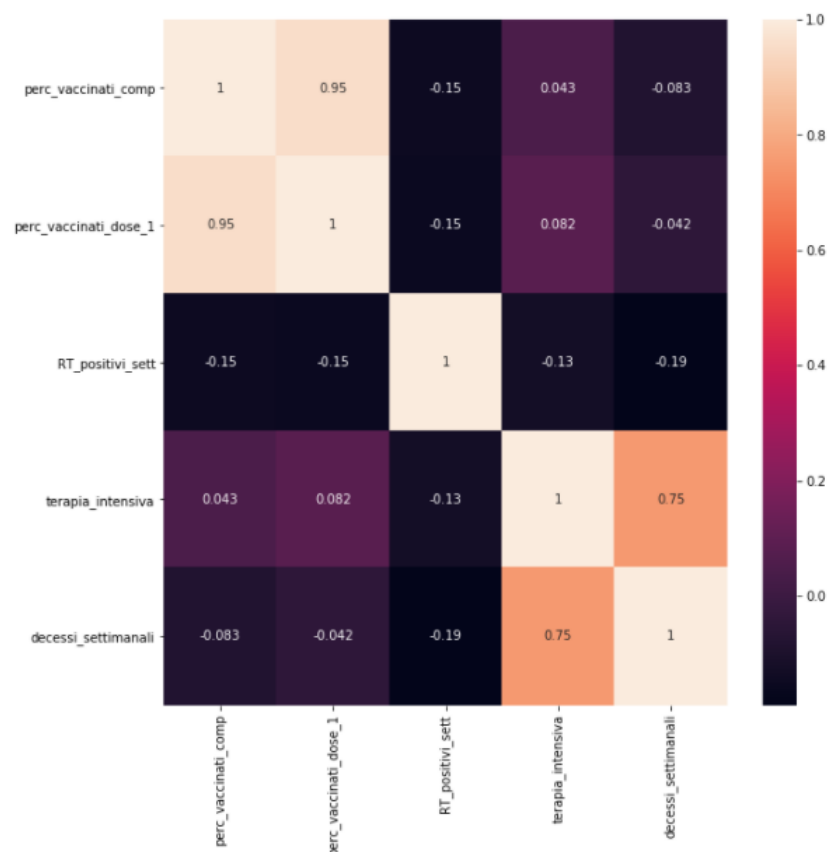


Figura 3. Correlogramma variabili(unico per le 3 regioni)

⁶ Scatterplot delle due dosi vaccino in relazione alle altre variabili(per quanto riguarda tutto il periodo di analisi) sono presenti nel foglio di lavoro (7 DATA EXPLORATION)

⁷ In questo caso non si può usare il classico indice di correlazione di Pearson, poiché non abbiamo una correlazione lineare, né la normalità. Per questo useremo la correlazione di Kendall, poiché non pone restrizioni né sulla normalità dei dati né sulla loro linearità.

Osservazioni:

Da questo primo grafico si può capire che solo l'indice RT è correlato in maniera negativa con la percentuale di vaccinati.

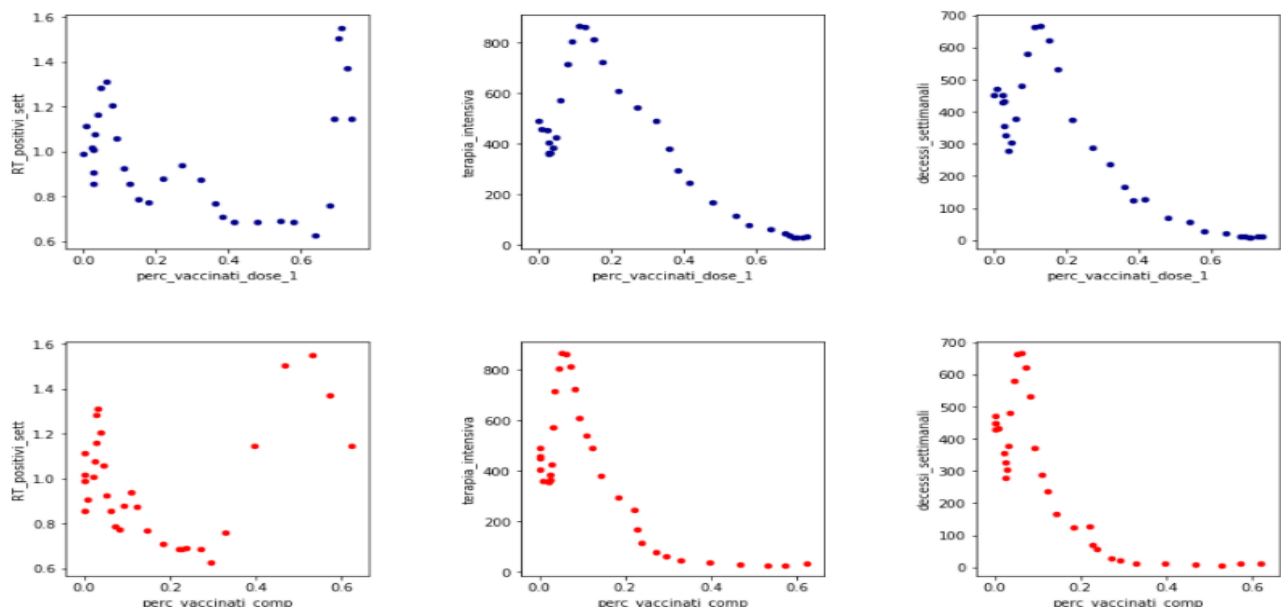
Per verificare la significatività della correlazione in base alle dosi di vaccino si userà un livello di significatività molto basso, ovvero verranno ritenute correlazioni significative tutti quei test che avranno un p-value (kendall tau test) ⁸ <0.001, valore di significatività usato in ambito medico.

RT: Possiamo notare che l'indice RT è l'unica variabile target che mostra una correlazione negativa (anche se molto bassa) non sia in realtà significativa.

Le variabili target vengono influenzate da tanti fattori non presenti nel dataset, variabili che non si conoscono e che non sono quantificabili, come ad esempio la responsabilità e il rispetto delle regole da parte di ogni individuo (rispetto del distanziamento sociale, uso di mascherine ecc) e soprattutto i periodi di LockDown. Si ritiene di poter aumentare la correlazione tra queste variabili. Il problema è che se si pensa alla variazione delle variabili target, prima dell'introduzione dei vaccini (periodo da marzo 2020 a gennaio 2021), si registra una diminuzione di tutte queste in corrispondenza dei periodi di LockDown quando i vaccini non erano stati ancora introdotti.

Per questo motivo, grazie alla presenza di variabili nascoste, il calo di morti, contagi e terapie non sarà direttamente attribuibile all'introduzione di vaccini (perchè zero), e così il test avrà bassa correlazione. In termini semplici significa che, questo brusco calo delle variabili target non risulta correlato all'introduzione e diffusione dei vaccini, ma ad un LockDown. Per questo si analizzerà (solo per questa parte) il periodo da quando vengono introdotti i vaccini andando così a rimuovere l'influenza del LockDown sulle variabili target poiché si vuole catturare l'effetto del vaccino, in modo da avere i dati il meno possibile influenzati dalle decisioni del Governo.

Scatter plot da introduzione vaccini Covid_19- Regione Lombardia⁹



⁸ 7 data exploration: test per ogni dose vaccino con variabili target mediante test: `"scipy.stats.kendalltau"`

⁹ Gli scatterplot nella fase di data exploration sono stati fatti per ogni regione analizzata, per ragioni di spazio non sono stati messi per tutte le regioni causa problemi spazio.

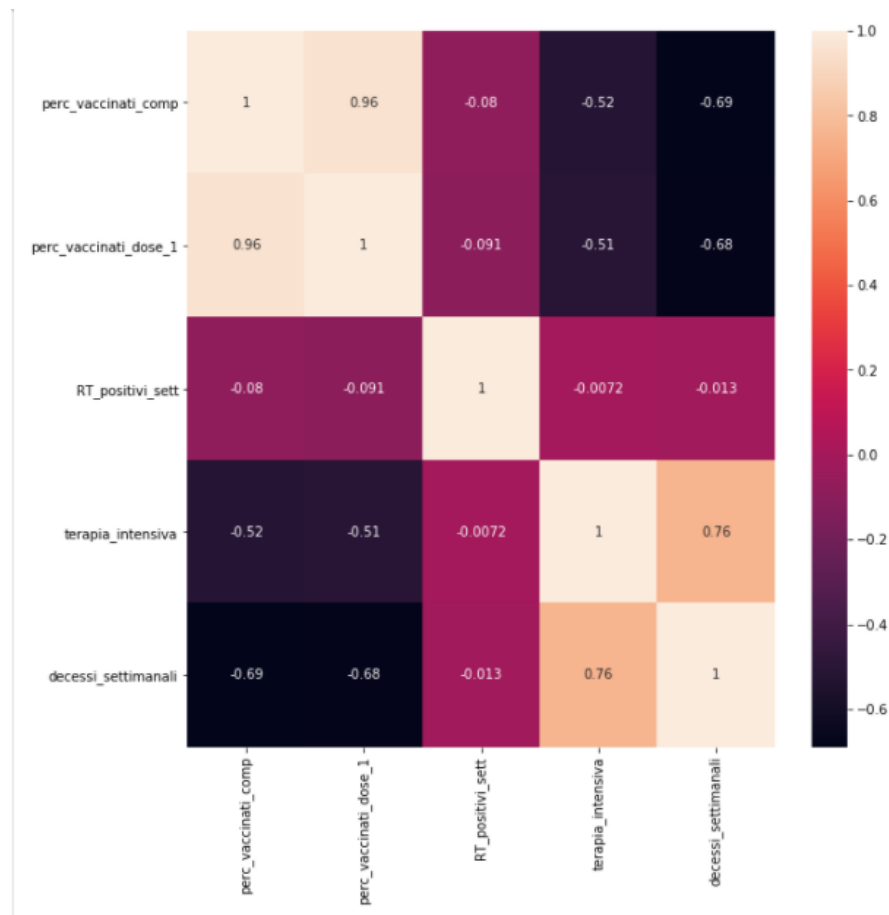


Figura 3. Correlogramma attributi (periodo vaccino, grafico unico per 3 regioni)

Si nota subito che, l'indice RT, non è influenzato nel modo desiderato dall'introduzione dei vaccini. Questo potrebbe significare che i vaccini NON hanno correlazione sull'indice RT. Quello che emerge dal grafico e dallo scatter plot è che l'introduzione dei vaccini ha una forte correlazione negativa su Terapia Intensiva (-0.52/-0.51) e Decessi (-0.69/-0.68), a dimostrazione del fatto che più le dosi di vaccino somministrate aumentano e più i casi gravi diminuiscono. Che è quello che si vuole dimostrare (l'introduzione dei vaccini ha effetto principalmente sui decessi e terapie intensive e meno sulla trasmissione del virus) e si sostiene la tesi che i vaccini tutelino maggiormente la salute dai casi più gravi.

Verificando di nuovo la significatività mediante test (kendall tau test)¹⁰ si può notare :

RT: Non più significativo, poiché non viene influenzato dall'introduzione del vaccino.

Terapie Intensive: Le terapie intensive in entrambe le dosi ora sono statisticamente significative. (si accetta fortemente ipotesi di correlazione significativa).

Decessi: E' statisticamente significativo per entrambe le dosi. (si accetta fortemente ipotesi di correlazione significativa)

¹⁰ 7 data exploration: test per ogni dose vaccino con variabili target mediante test: "scipy.stats.kendalltau"

3.3 Conclusioni Analisi Correlazione.

Grazie all'analisi si è individuata solo una debole correlazione significativa tra il vaccino (prima e seconda dose) sui Decessi. Questo significa che il vaccino non ha una relazione che si possa ritenere significativa sull'andamento dell'indice RT, e sulle terapie intensive. Successivamente si è rimossa l'influenza dell'LockDown.

Come da ipotesi, gli indici di correlazione si sono alzati più del doppio, ottenendo la significatività oltre che dei decessi anche delle terapie intensive con i vaccini. Possiamo concludere che l'indice RT non viene influenzato dall'introduzione del vaccino, mentre la riduzione delle terapie intensive è legata con coefficiente -0.52 /-51 all'introduzione del vaccino e i decessi con coefficiente -0.68/-69 all'introduzione del vaccino, e in maniera statisticamente molto significativa.

4. Realizzazione delle infografiche.

Lo sviluppo delle infografiche è stato pensato per poter dare delle risposte, quanto più accurate possibili, alle domande di ricerca chiarite nei paragrafi precedenti e, soprattutto, per poter verificare, laddove i dati raccolti lo dimostrassero, le ipotesi da noi definite.

Per questo motivo è stata studiata, anche per mezzo di indici statistici (quale la correlazione), ogni variabile di interesse sia singolarmente, sia attraverso una specifica infografica che meglio riuscisse a catturare la relazione con le variabili “vaccino dose 1” e “vaccino dose 2”.

È evidente come quando si parla di Covid19, si stia parlando di un argomento con importanza significativa a livello mediatico, a livello di interesse pubblico e personale; spesso, le molte tematiche che si affrontano non sono di facile comprensione, proprio per la difficoltà dell’argomento stesso e delle variabili che lo influenzano.

Uno degli obiettivi prefissati con lo sviluppo e la realizzazione delle infografiche è stato quello di rendere il più comprensibile possibile, soprattutto alle persone che non si occupano né di medicina né di analisi dati, come l’introduzione dei vaccini abbia influenzato positivamente l’andamento del Covid19 in Italia in termini di un minor numero di pazienti ricoverati in terapia intensiva e di un minor numero di decessi.

Per poterle rendere più accessibili e comprensibili, le infografiche sono state sviluppate assegnando ad ogni variabile analizzata un colore ben specifico in modo tale da facilitare l’utente ad individuare subito il contesto e la gravità analizzati. Per coerenza, e per non creare confusione verso l’utilizzatore e l’utente finale, il colore della singola variabile è stato mantenuto identico in ogni infografica dove viene analizzata. Nello specifico, si è deciso di assegnare i seguenti colori alla variabili d’interesse (la scelta del colore per ogni variabile, anche in questo caso, è dettata dal voler ridurre al minimo l’incomprensione e facilitare una corretta, esaustiva e più semplice visualizzazione):

- Tonalità di blu (da azzurro chiaro a blu scuro): Numero Positivi Covid19 e Indice Rt Covid19;
- Tonalità di viola (da rosa a viola scuro): Numero Terapie Intensive per Covid19;
- Tonalità di Nero (da grigio chiaro a nero) : Numero Decessi per Covid19;
- Giallo: rappresenta la percentuale di vaccinati con la sola dose 1 ;
- Arancione: rappresenta la percentuale di vaccinati con ciclo completo (ovvero chi ha ricevuto anche la dose 2);
- Rosso: Numero ospedalizzati (fanno parte degli ospedalizzati i pazienti ricoverati con sintomi e i pazienti ricoverati nelle terapie intensive).

In questo paragrafo, pertanto, si andranno a presentare singolarmente le infografiche realizzate, che dimostrano come l’introduzione del vaccino Covid19 abbia effettivamente portato enormi benefici soprattutto ai casi gravi della malattia (terapia intensive e decessi).

4.1 Prima Infografica

Percentuale annua di positivi Covid19 in rapporto alla popolazione in ogni regione

La prima infografica mostra come varia la percentuale di positivi a Covid19 dal primo anno in cui la malattia si è manifestata (2020) all'attuale anno in corso (2021) in ogni regione italiana. La tonalità di blu evidenzia la gravità della malattia (in termini di positivi) nelle varie regioni: tonalità chiare rappresentano una percentuale di positivi minore rispetto alle tonalità più scure.

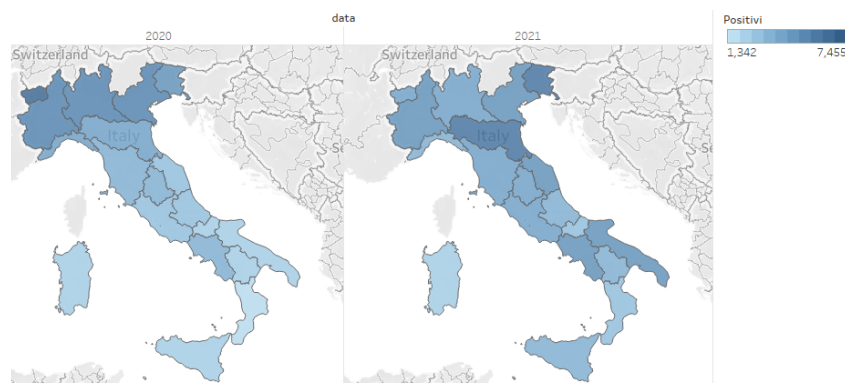


Figura 4. Infografica n.1

È facile notare come la tendenza sia cambiata da un anno all'altro: nel primo anno di pandemia le regioni più colpite sono state le regioni del Nord-Italia; nel secondo anno si vede un aumento dei positivi soprattutto nel centro-sud Italia con un leggero calo nelle regioni del Nord.

4.2 Seconda Infografica

Indice Rt positivi(annuo) Covid19 in ogni regione

Questa seconda infografica mostra la media relativa all'indice annuo R_t positivi in tutte le regioni italiane nei due anni di diffusione della malattia. Anche in questo caso si è scelta la tonalità di blu, che in egual modo evidenziano la gravità della malattia nelle varie regioni: tonalità chiare rappresentano un indice R_t annuo minore, tonalità scure rappresentano un indice R_t annuo maggiore.

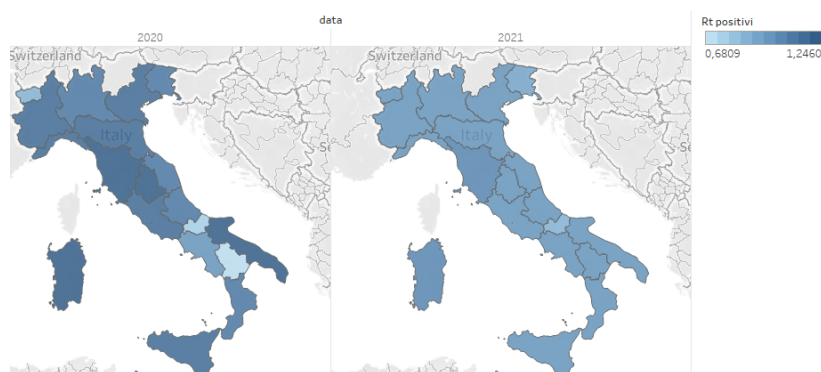


Figura 5. Infografica n.2

Si può notare subito come l'indice R_t annuo descriva una situazione differente rispetto all'analisi della percentuale di positivi in ogni regione: infatti, è evidente come nel primo anno di pandemia non solo le regioni del nord non fossero le sole a essere colpite dalla malattia, ma bensì l'intera penisola italiana è stata colpita duramente dalla diffusione del Covid19. Nell'anno successivo, il 2021, è evidente come ad oggi (8

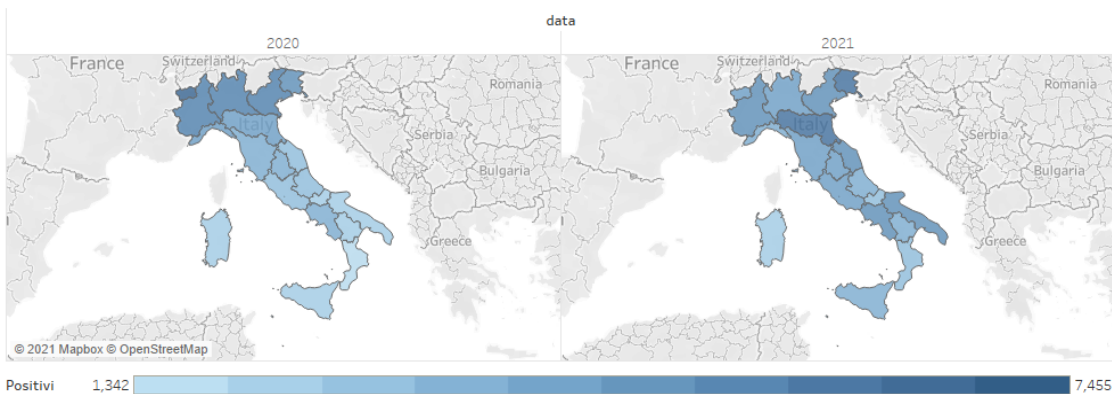
agosto 2021) le misure intraprese (introduzione vaccini, politiche di confinamento -lockdown-) hanno prodotto benefici per il Paese.

4.3 Terza Infografica

Confronto tra la percentuale di positivi Covid19(in rapporto alla popolazione) e Indice Rt annuo positivi Covid19 nelle regioni italiane

Si è deciso di fare un confronto tra le prime due infografiche per evidenziare come possa essere “pericoloso”, ed errato, parlare e spiegare il fenomeno Covid19 in Italia (e non solo) basandosi solo sul numero di positivi da Covid19 (in quanto è una variabile sì importante, ma fortemente influenzata).

Percentuale annua di positivi Covid19 in rapporto alla popolazione in ogni regione



Indice Rt positivi(annuo) Covid19 in ogni regione

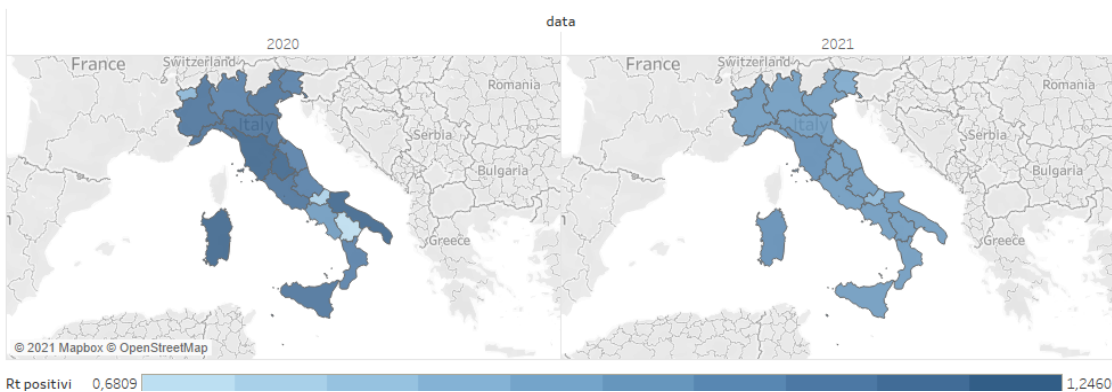


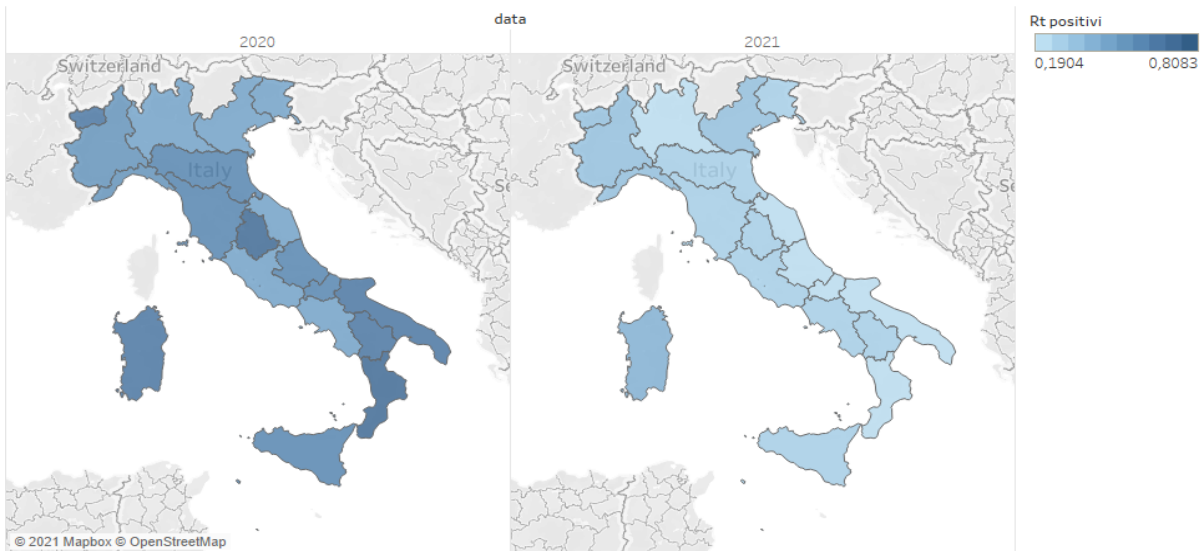
Figura 6. Infografica n.3

Questa infografica dimostra come con l'utilizzo di differenti variabili si possano raccontare storie differenti.

Per le analisi successive si è preferito non analizzare l'effetto dell'introduzione dei vaccini Covid19 sulla variabile “Percentuale Positivi”, ma piuttosto sulla variabile “RT positivi”, in quanto variabile per la quale le autorità prendevano le decisioni circa le misure di confinamento per ogni regione.

Avendo usato la media per quanto riguarda l'indice RT(fig.5), per rappresentare meglio i valori, è necessario disporre di informazioni ulteriori riguardanti la dispersione dei valori intorno alla misura di tendenza centrale. Per questo scopo verrà utilizzato la deviazione standard, esprimibile con la stessa unità di misura della variabile osservata¹¹.

Deviazione Standard Indice Rt positivi(annuo) Covid19 in ogni regione



Le regioni più scure nell'anno 2020 indicano una maggior dispersione¹² dei nostri dati, (dati sono molto sparsi intorno alla media) questo significa che per la rispettiva regione, ci saranno valori molto diversi tra di loro nell'intero anno. Questa situazione può essere interpretato come molto “altalenante” per quanto riguarda l'indice RT nel 2020 .

Al contrario quelli più chiari indicano una poca dispersione intorno alla media, quindi dei valori meno discordanti tra di loro. E' interessante notare come sia marcata la differenza cromatica tra i due anni, indice del fatto che la deviazione standard dei valori è più “vicino” alla media rispetto all'anno 2020. Questa variazione è intuibile (anche se relativa solo a 3 regioni) dal box plot (fig:2).

¹¹ Non presente nel questionario

¹² Variabilità è attribuibile a vari periodi di LockDown, e impennate dei contagi.

4.4 Quarta Infografica

Confronto tra l'indice Rt positivi (settimanale) e l'andamento delle vaccinazioni a Covid19

Come detto in precedenza, le analisi effettuate sull'effetto dell'introduzione del vaccino Covid19 si concentrano su tre variabili: indice Rt positivi (settimanale), Numero di pazienti in terapia intensiva (settimanale) e numero di pazienti deceduti (settimanalmente) causa Covid19.

Il focus di questa quarta infografica è incentrato a valutare la relazione tra l'indice Rt settimanale dei positivi (nelle tre regioni di interesse: Lombardia, Lazio e Veneto) e la percentuale di popolazione vaccinata con la prima dose (linea gialla) e con ciclo completo (linea arancione).

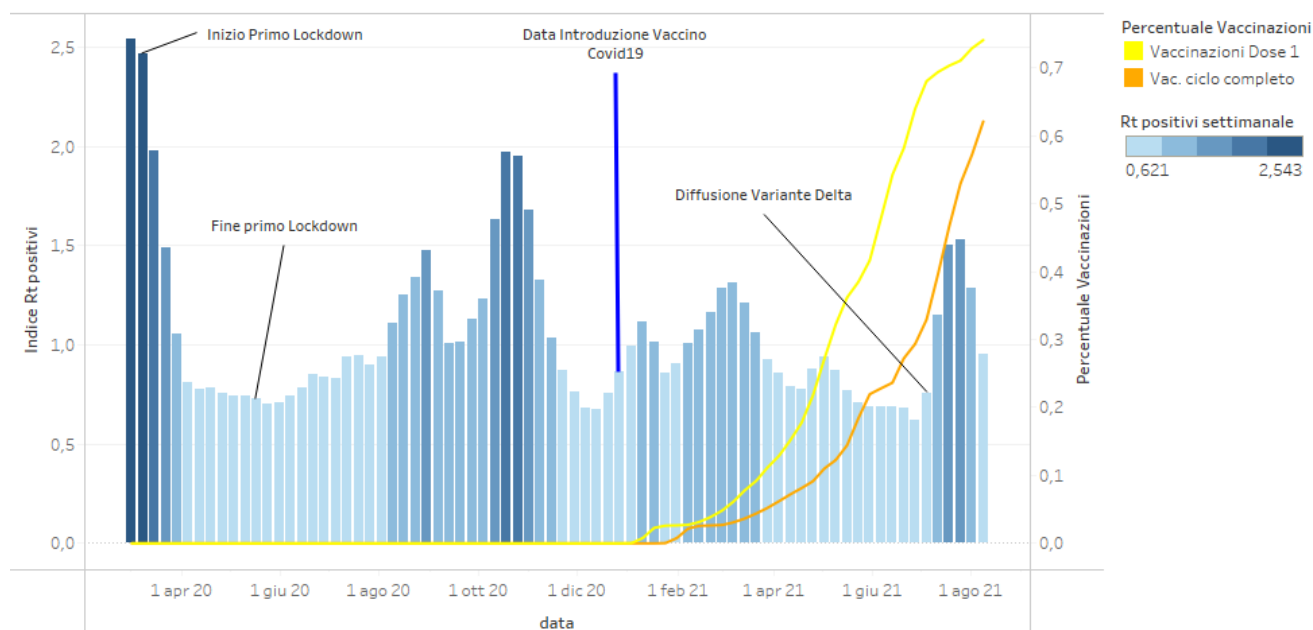


Figura 7. Infografica n.4 - Lombardia

A differenza delle prime tre infografiche, per facilitare l'utente finale nella comprensione di una nuova tipologia di visualizzazione, si è deciso di inserire dei visual cues che potessero fornire un contesto il più dettagliato possibile circa l'orizzonte temporale visualizzato. Infatti si può notare la presenza di puntatori nelle date in cui sono presenti eventi significativi della pandemia Covid19 dalla sua diffusione ad oggi: Inizio primo lockdown, fine primo lockdown, Data introduzione vaccino Covid19 e diffusione variante Delta.

È evidente come, prima dell'introduzione dei vaccini, le politiche di confinamento forzato abbiano aiutato molto a ridurre la diffusione del virus (primo lockdown); e come invece politiche meno restrittive abbiano causato, nei mesi successivi l'estate, un nuovo innalzamento dell'indice Rt.

Si può notare che, una volta introdotti i vaccini e vaccinata con almeno la prima dose una buona percentuale di popolazione (25%-30%), l'indice Rt sia diminuito anche se, non sorprende, un suo nuovo aumento causato dalla diffusione della variante Delta.

Tuttavia questo non deve scoraggiare o far dubitare sull'effettiva efficacia dei vaccini poichè il vaccino anti-Covid19 è pensato per poter proteggere, non dalla positività al virus, ma dalle aggravanti derivanti dalla malattia, come dimostrato nell'analisi di correlazione **3.2**

4.5 Quinta Infografica

Confronto andamento Rt positivi (settimanale) con vaccinazione dose 1¹³

L'intento di questa visualizzazione è quello di rendere semplice un concetto complesso come la correlazione tra due variabili. Il metodo utilizzato è quello di una serie storica interattiva che aiuti a comprendere, di settimana in settimana, come l'indice Rt varia al crescere della percentuale di vaccinazioni.

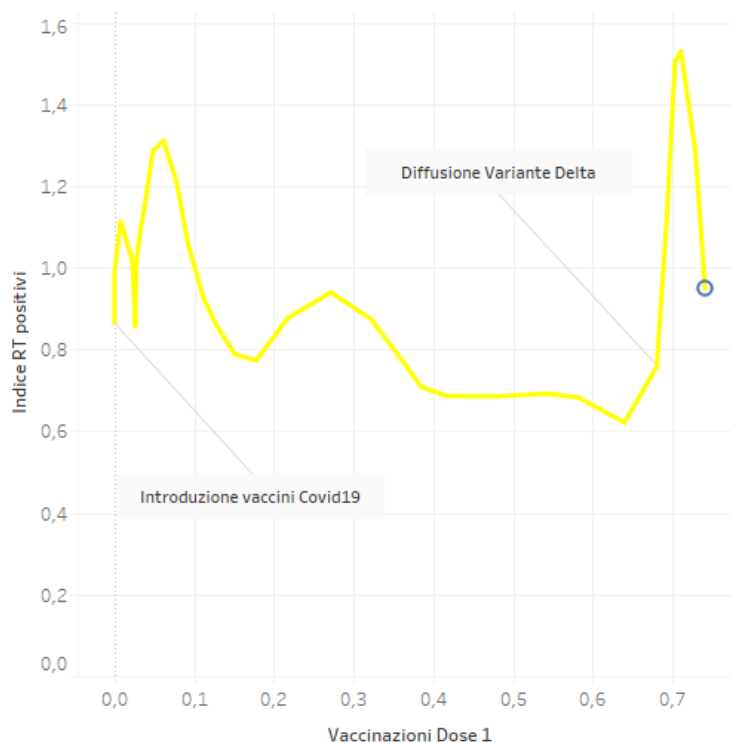


Figura 8. Infografica n.5 - Lombardia

Possiamo notare come dall'introduzione del vaccino, anche senza aver rappresentata la visione generale, si presenti un trend decrescente che porta l'indice Rt a 0.6 prima della crescita dovuta alla diffusione della variante Delta. Questo ci mostra come in questo periodo d'analisi, l'indice Rt sia inizialmente calato prima di risalire, ulteriore dimostrazione del fatto che gli effetti più evidenti del vaccino non si riscontrano sull'indice Rt ma su altre variabili.

Nonostante l'analisi di correlazione dimostri che non ci sia un legame significativo tra Rt e vaccini, qua si nota una "lieve" decrescita di trend prima di risalire causa variante Delta per quanto riguarda l'indice Rt.

Le due infografiche (quarta e quinta) raccontano la stessa storia attraverso visualizzazioni differenti, per questo motivo le considerazioni sono le medesime.

¹³ Rappresentiamo prima dose vaccino, ma i nostri dati subiscono effetto anche della seconda

4.6 Sesta Infografica

Confronto tra le terapie intensive settimanali causa Covid19 e l'andamento delle vaccinazioni a Covid19

Questa sesta infografica è stata pensata e sviluppata per mostrare, in confronto con la precedente, come la relazione tra la percentuale di popolazione vaccinata (almeno con la prima dose) e numero di terapia intensive settimanali causa Covid19 sia molto forte. In questo caso si è giocato con le tonalità del colore viola per poter facilitare la comprensione della stessa infografica: un colore tendente al rosa indica un minor numero di terapie intensive causa Covid19; tonalità sempre più tendenti al viola scuro indicano via via un crescente numero di terapie intensive settimanali causa Covid19. La percentuale di popolazione vaccinata è definita dalla linea gialla (per la sola dose 1) e dalla linea arancione (per chi ha concluso il ciclo completo).

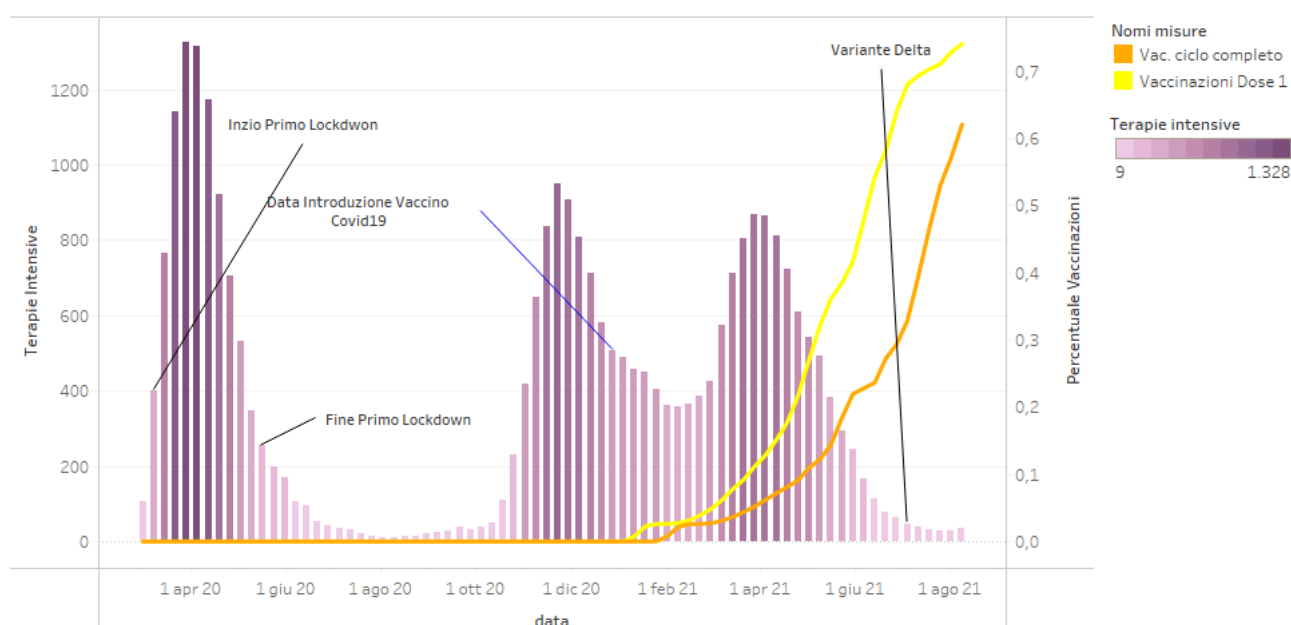


Figura 9. Infografica n.6 - Lombardia

Come per la precedente infografica, si è visto necessario introdurre delle visual cues che facilitassero la lettura dell'infografica, indicando le date degli avvenimenti più importanti dall'inizio della pandemia ad oggi.

Lo studio e l'analisi della variabile "terapia intensiva" è dovuto al fatto che un aumento dei ricoveri in terapia intensiva è sempre stato un forte segnale di pericolo per un eventuale collasso del sistema sanitario e determinava l'eventuale chiusura di una regione (zona rossa).

Risulta evidente come, grazie all'introduzione dei vaccini Covid19, una volta che una buona percentuale di popolazione abbia ricevuto la prima somministrazione del vaccino (circa il 25%-30%), il numero di terapie intensive vada a diminuire progressivamente (al crescere della popolazione vaccinata) senza la necessità di estreme misure di confinamento forzato.

La stessa diffusione della variante Delta, ad oggi (8 agosto 2021), non sembra comportare un aumento di casi gravi da covid19 che necessitano ricoveri nelle terapie intensive. Ciò evidenzia la bontà e gli effetti positivi che la vaccinazione al Covid19 comporta nel cercare di ridurre le aggravanti della malattia, salvaguardando il sistema ospedaliero regionale (e nazionale).

4.7 Settima Infografica

Confronto tra il numero di decessi settimanali causa Covid19 e l'andamento delle vaccinazioni a Covid19

Con questa settima infografica si è cercato di evidenziare l'importanza di come le vaccinazioni a Covid19 possano ridurre sensibilmente il numero di decessi provocati dalla malattia. Per un maggior impatto visivo, in questo caso, si è giocato con le tonalità del nero: tonalità tendenti al grigio chiaro evidenziano un minor numero di decessi da Covid19; le tonalità più scure, quelle che tendono al nero, evidenziano un numero di decessi da Covid19 sempre maggiore. Come per le precedenti infografiche, per coerenza e per una più semplice interpretabilità, il colore della percentuali di vaccinazioni con dose 1 (linea gialla) e ciclo completo (linea arancione) è rimasto invariato.

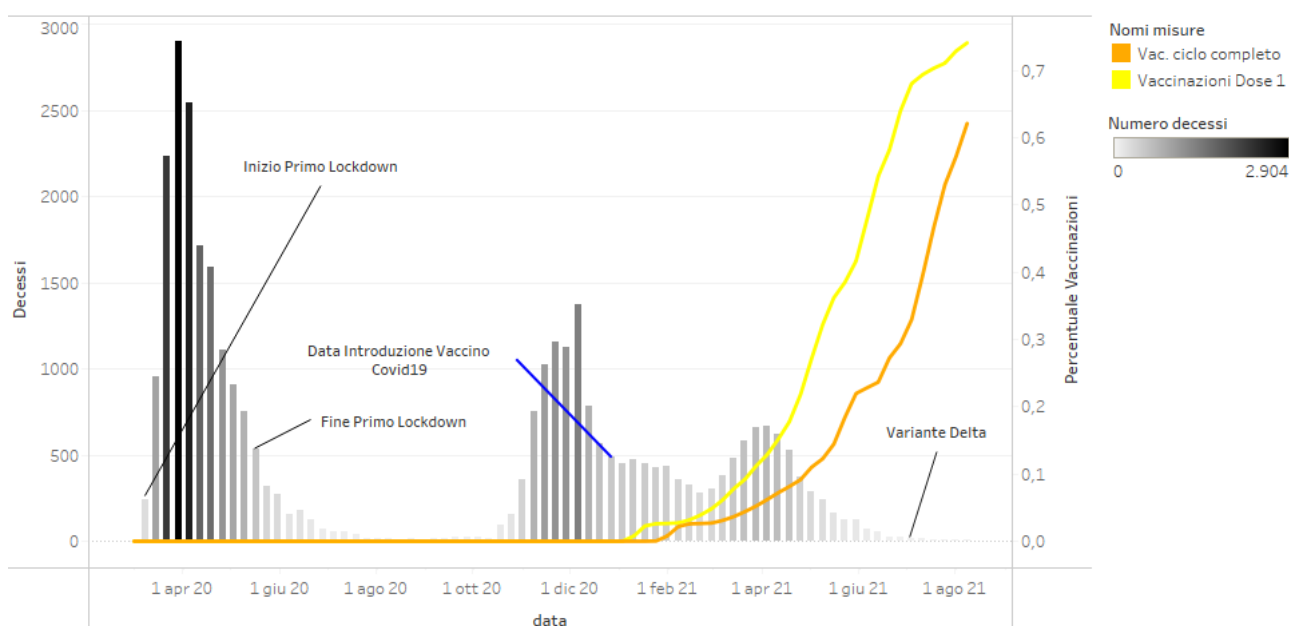


Figura 10. Infografica n.7 - Lombardia

È evidente come questa visualizzazione risulta essere una delle più importanti per l'intero progetto da noi sviluppato in quanto meglio rappresenta i benefici derivanti dall'introduzione e somministrazione del vaccino Covid19. Si nota subito come all'aumentare della popolazione vaccinata, il numero di decessi settimanali causati da Covid19 diminuisca in misura sempre più crescente fino ad arrivare agli stessi valori ottenuti dalle misure di confinamento forzato.

4.8 Ottava Infografica

Confronto tra andamento dei decessi, totali ospedalizzati e vaccini (dose 1)

L'ultima infografica fornisce un ulteriore punto di vista sul tema, migliorando la comprensione del problema affrontato dalle due precedenti infografiche, mostrando come l'introduzione del vaccino abbia avuto effetti positivi nella lotta contro il Covid19. Lo scopo principe è valutare la relazione tra andamento dei decessi (colore nero), totali ospedalizzati (somma terapie intensive e ricoverati con sintomi, colore rosso) e percentuale di somministrazioni del vaccino Covid19 (linea gialla¹⁴).

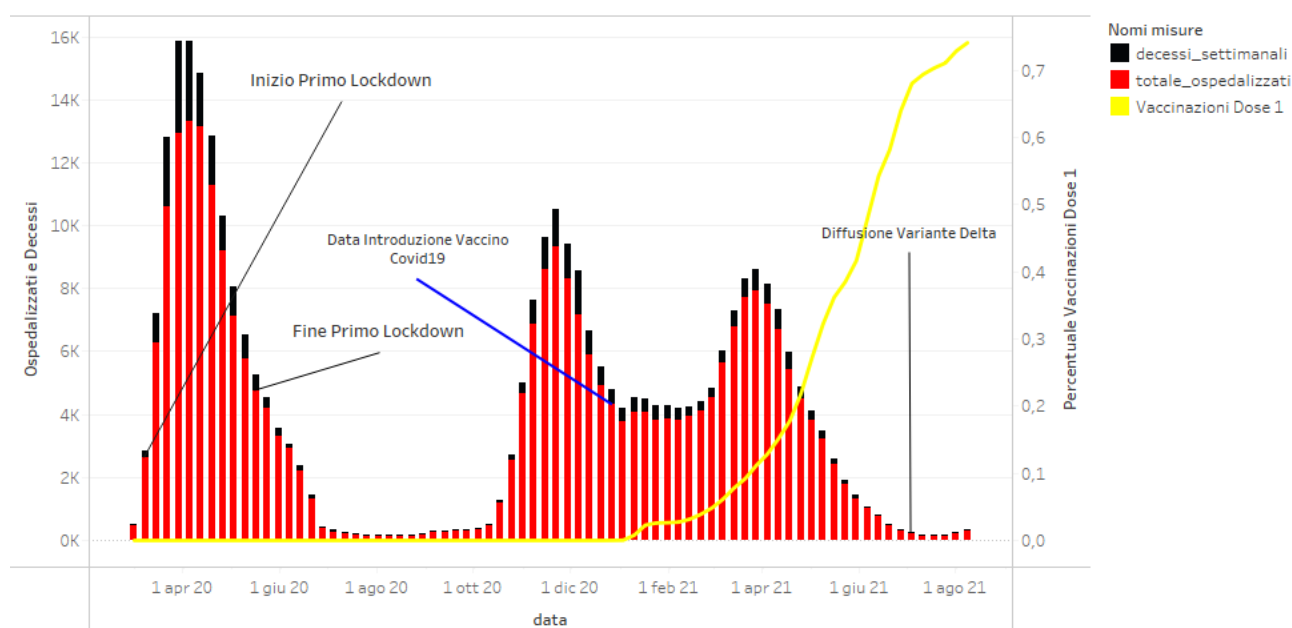


Figura 11. Infografica n.8 - Lombardia

Si è consapevoli che questa visualizzazione richieda un maggior sforzo interpretativo, soprattutto se presentata singolarmente, tuttavia, l'utente finale arriva con gli strumenti necessari per una sua corretta comprensione. Strumenti che ha acquisito gradualmente grazie alle visualizzazioni e interpretazione delle precedenti infografiche.

A inizio pandemia, al crescere dei pazienti ricoverati con sintomi o nelle terapie intensive cresceva di pari passo il numero di decessi dovuti a Covid19. Si è cercato di arginare il proliferare del virus e delle aggravanti della malattia a esso riconducibile, imponendo forti misure restrittive alla circolazione e alle attività sociali delle persone.

L'illusione creatasi con una maggior libertà, dovuta a misure meno stringenti, ha favorito una nuova diffusione del virus con relativo aumento dei decessi e dei casi gravi che necessitano di cure ospedaliere.

L'introduzione del vaccino anti Covid19, avvenuta in data 27-12-2020, mostra come, all'aumentare della popolazione vaccinata sia i casi che necessitano di cure ospedaliere e sia il numero di decessi decresce sensibilmente arrivando ad ottenere effetti sul trend riconducibili a misure molto invasive come il lockdown.

¹⁴ Rappresentiamo prima dose vaccini, ma i nostri dati subiscono effetto anche della seconda

5. Valutazione della qualità delle visualizzazioni

La valutazione delle qualità delle visualizzazioni è forse la parte più importante del progetto. Dopo aver studiato diversi metodi di valutazione delle infografiche si è deciso di utilizzare i seguenti:

- Valutazione euristica
- Questionario online
- Test utente

Ogni tipo di valutazione scelta ha sia punti positivi che mancanze, ma utilizzando questi tre metodi si ritiene di aver ottenuto un quadro generale adeguato per poter comprendere l'utilità e la correttezza delle infografiche.

5.1 La valutazione euristica

Questa metodologia di valutazione delle visualizzazioni è stata scelta per verificare il primo approccio di un utente con le infografiche e soprattutto con la loro interattività. Sono state selezionate pertanto 3 persone che non avevano mai avuto contatti con queste visualizzazioni e con questo progetto.

La prima cosa interessante da osservare è l'approccio iniziale dell'utente, se pare a disagio o se pare essere tranquillo ed abituato a certe tipologie di visualizzazioni. Da qui si è provato ad interagire con l'utente per verificare dubbi e perplessità, lo si è inoltre lasciato libero di esprimere la propria opinione riguardo a qualsiasi cosa percepita durante questo tipo di valutazione.

Ed ecco le cose più interessanti emerse:

Utente 1	Si intravede subito la poca abitudine alla visualizzazione di grafici, soprattutto sono evidenti le difficoltà nel loro utilizzo interattivo. Una volta superate le difficoltà iniziali, però la valutazione risulta essere positiva e la comprensione adeguata.
Utente 2	L'utente manifesta subito una difficoltà nel distinguere l'intensità del colore rosso, dove esso è presente. Durante la valutazione, l'utente rivelerà di essere leggermente affetto da daltonismo. Questa indicazione risulta essere molto importante per la modifica nella scelta dei colori.
Utente 3	Non presenta difficoltà nel comprendere le infografiche del progetto, ad eccezione del grafico interattivo che mette a confronto, l'indice RT e la seconda del vaccino.

5.2 Modifiche alla visualizzazione in seguito alla valutazione euristica

Dopo aver analizzato i risultati ed essersi confrontati riguardo i problemi emersi durante la valutazione euristica, si è deciso di apportare le seguenti modifiche cercando di risolvere i problemi evidenziati.

La modifica più importante è stata effettuata sulla visualizzazione riguardante la cartina italiana e la visualizzazione riguardante la serie temporale dell'indice RT; entrambe presentavano come caratteristiche in comune l'indicazione della gravità della situazione tramite una scala di colore che andava dal bianco al rosso.

Analizzando il problema riscontrato dall'utente numero 2, che evidenzia difficoltà nel comprendere la differenza di colore in queste due rappresentazioni, si è deciso di modificare la scala di colore dal bianco al rosso, in una che varia dal bianco al blu. Questa decisione è stata presa dopo essersi informati che il blu è un colore che causa molti meno problemi alle persone che soffrono di daltonismo (secondo IAPB¹⁵, le persone che soffrono di daltonismo in Italia si aggirano intorno i 2,5 milioni). Si è consci del fatto che il rosso fosse un colore più adatto, in quanto si voleva rappresentare la gravità della situazione, ma il rendere le visualizzazioni più fruibili e accessibili a tutti ha prevalso.

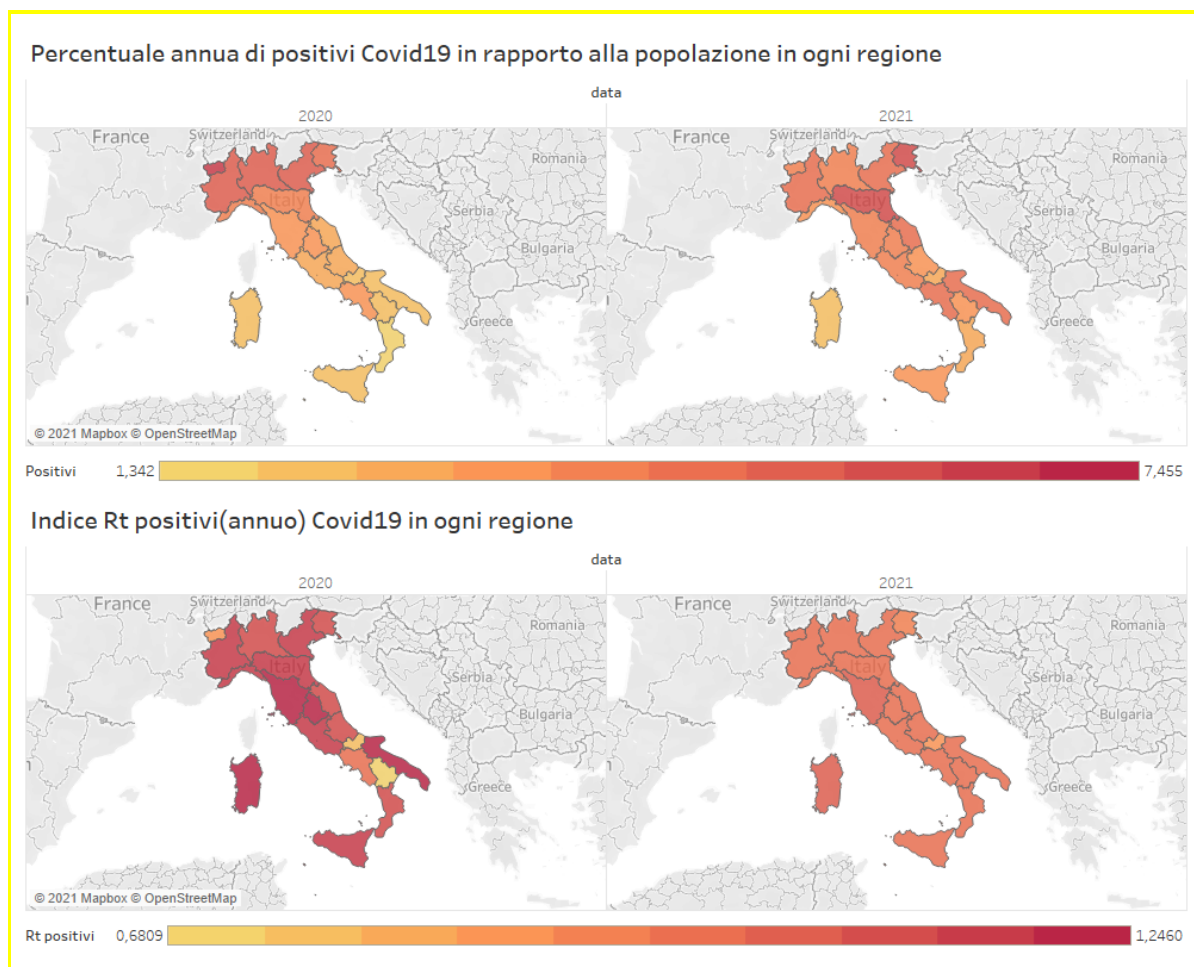


Figura 12. 1°Errore infografica

¹⁵ IAPB= Agenzia internazionale per la prevenzione della cecità

La seconda modifica deriva dalle indicazioni fornite dall'utente numero 3, il quale aveva mostrato difficoltà nel comprendere il grafico, che come obiettivo ha quello di rappresentare la correlazione, di settimana in settimana, tra la variabile RT positivi e la variabile percentuale vaccinazioni ciclo completo.

Il punto in cui l'utente ha difficoltà riguarda quasi tutto l'anno 2020, in quanto in quest'anno l'indice RT continua a muoversi e a cambiare di valore mentre le vaccinazioni rimangono sempre ferme (il vaccino è stato somministrato per la prima volta in Italia il 27 dicembre del 2020). Per ottenere una migliore visualizzazione allora si è deciso di eliminare tutte le date precedenti al 27 dicembre, in modo tale da poter rendere questa infografica molto più fruibile ed interessante.

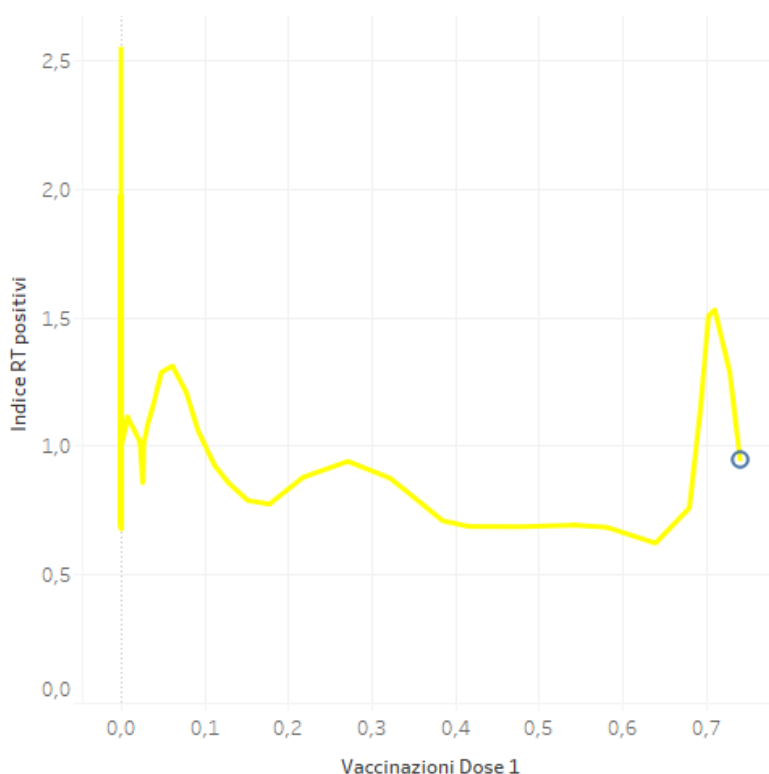


Figura 13. 2°Errore infografica - Lombardia

5.3 Test Utente

Si è proceduto ad una valutazione delle infografiche sotto forma di test somministrato a 4 utenti con un grado di difficoltà crescente nelle domande.

Gli utenti sottoposti al test sono stati catalogati per sesso, per età e per dimestichezza nella visione e nell'utilizzo dei grafici.

Utente	sesso	età	dimestichezza
1	M	20-30	buona
2	M	50-60	sufficiente
3	F	30-40	sufficiente
4	M	20-30	ottima

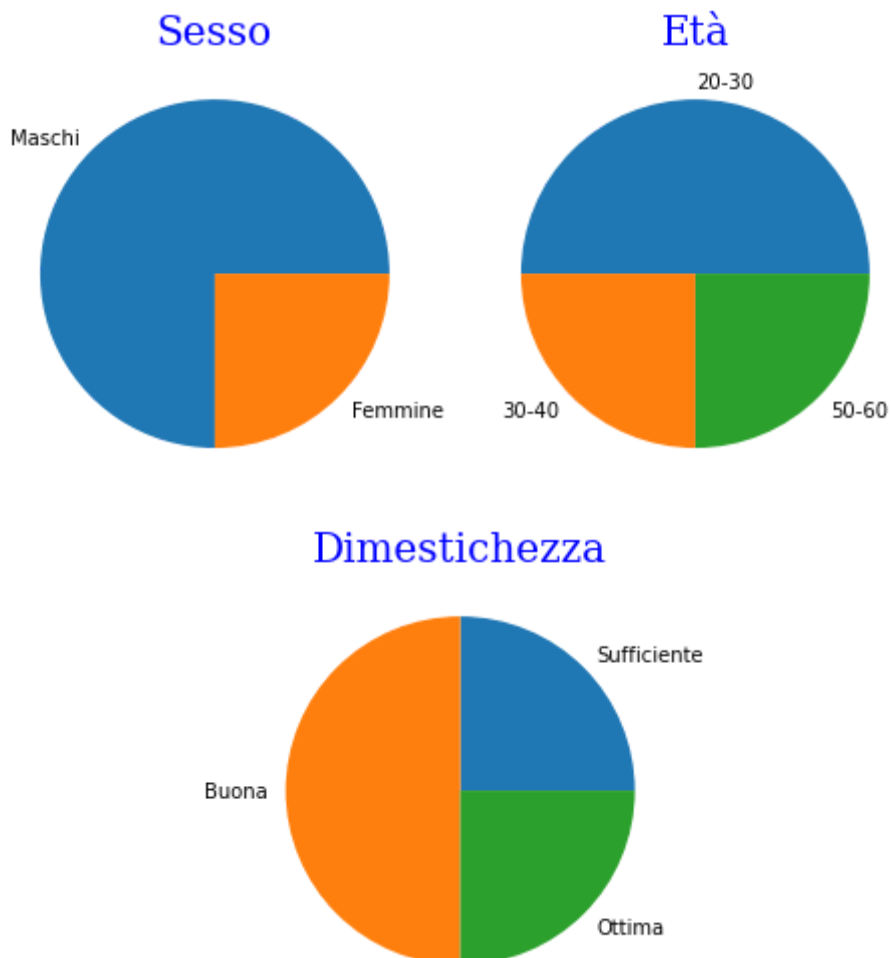


Figura 14. Pie chart test utente

Ecco i risultati delle risposte ottenute:

Test	Utente	Tempo (in secondi)	Correttezza
Domanda 1	1	33	C
	2	22	C
	3	17	S
	4	6	C
Domanda 2	1	60	C
	2	35	C
	3	45	C
	4	2	C
Domanda 3	1	44	S
	2	70	S
	3	30	C
	4	6	C

La correttezza delle risposte è stata valutata in termini di C = Corretto e S = Sbagliato.

Agli utenti non sono stati forniti aiuti di nessun genere, l'esecuzione del test è avvenuta in maniera individuale in modo che nessuno potesse venire influenzato. Le condizioni, l'ambiente e i mezzi tramite i quali gli utenti sono stati sottoposti erano uguali per tutti, così da non interferire in alcun modo nelle risposte.

Durante l'esecuzione del test sono state notate delle reazioni da parte degli utenti e sono state annotate in questo modo:

Test	Osservazioni
Domanda 1	Un utente ha mostrato difficoltà nel capire l'utilizzo del grafico multimediale inficiando così anche la sua possibile risposta.
Domanda 2	Il tempo richiesto è stato leggermente di più rispetto al previsto a causa del procedimento di apprendimento (riuscito) di questi grafici multimediali
Domanda 3	Alcuni utenti meno esperti sono stati tratti in inganno dalla visualizzazione andando ad osservare una parte sbagliata del grafico

Una cosa sicuramente positiva di questo test è stata la capacità che gli utenti acquisivano di domanda in domanda nel rapportarsi con i grafici e sfruttarne tutte le loro possibilità a livello interattivo

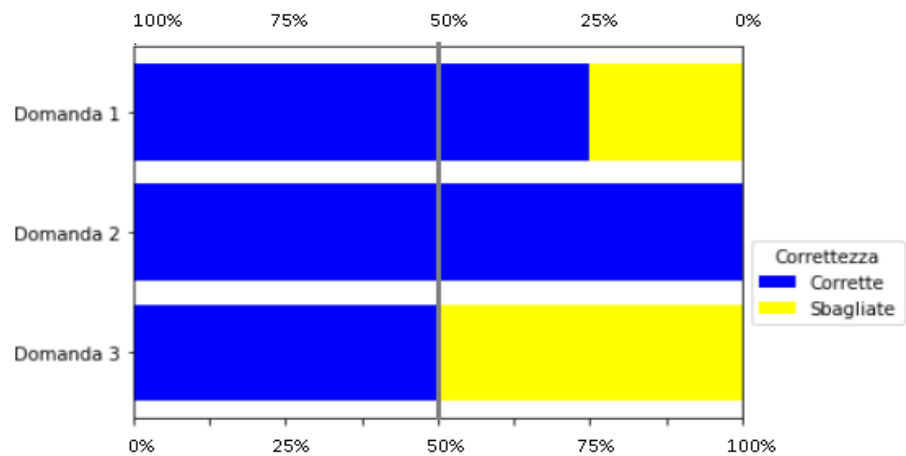
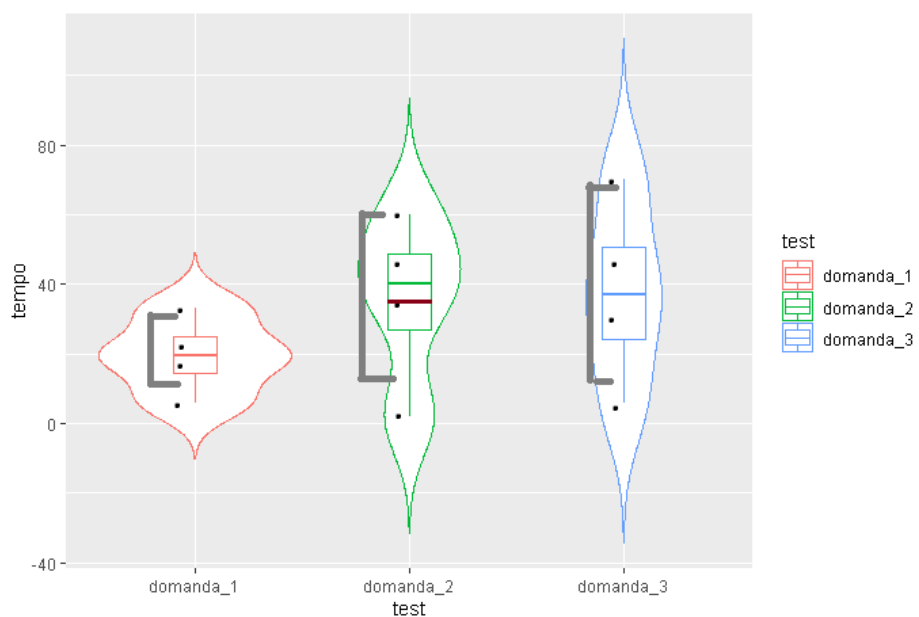


Figura 15. bar stacked chart test utente correttezza risposte

Dal grafico si nota che, sia la domanda 1 che la domanda 2 hanno una maggioranza significativa di risposte esatte, mentre la domanda 3 non presenta una maggioranza significativa di risposte esatte.

Si è deciso di effettuare una piccola visualizzazione sulla distribuzione dei tempi di risposta per tutti e tre i task richiesti agli utenti coinvolti. Il risultato è il seguente¹⁶:



test	mean	std	min	max
<chr>	<dbl>	<dbl>	<int>	<int>
domanda_1	19.5	11.2	6	33
domanda_2	35.5	24.6	2	60
domanda_3	37.5	26.8	6	70

¹⁶ Media linea marrone

Le barre grigie del grafico indicano una distanza 2sigma dalla media, tutti i valori fuori da queste “parentesi” indicano un ritardo o anticipo eccessivo nei tempi di esecuzione rispetto alla media. La barra marrone indica la media, l’unica domanda dove la media non coincide con la mediana è la domanda 2.

Il grafico è composto dal violin plot, dal boxplot e dai punti relativi alle osservazioni : in questo modo è possibile riassumere i risultati del test utente con completezza. Tutti e 3 i grafici mostrano come si distribuiscono i valori: il violin plot permette di vedere a colpo d’occhio la distribuzione dei dati; i punti rappresentano i tempi di esecuzione di ogni soggetto per un determinato task; il box plot permette di osservare dove si posizionano il massimo e il minimo valore (questo viene indicato dalla lunghezza dei baffi ed infatti i valori dello scatterplot più alti e più bassi si collocano in corrispondenza dei baffi).

Si può notare come per il task 2 e il task 3 i tempi di esecuzione del compito siano più vari rispetto al task 1 dove i tempi di esecuzione si collocano in un range più ristretto

5.4 Questionario psicometrico

Ultimo tipo di valutazione utilizzato è il questionario psicometrico.

Si è deciso di adottare la scala Cabitza-Locoro, che consiste nel valutare la qualità delle infografiche in una scala il cui range è compreso tra 1 e 6 .

Più la valutazione è positiva e più i voti saranno vicini al 6, più la valutazione è negativa e più i voti tenderanno all'1.

I campi della valutazione sono i seguenti:

- Chiarezza
- Utilità
- Bellezza
- Intuitività
- Informatività

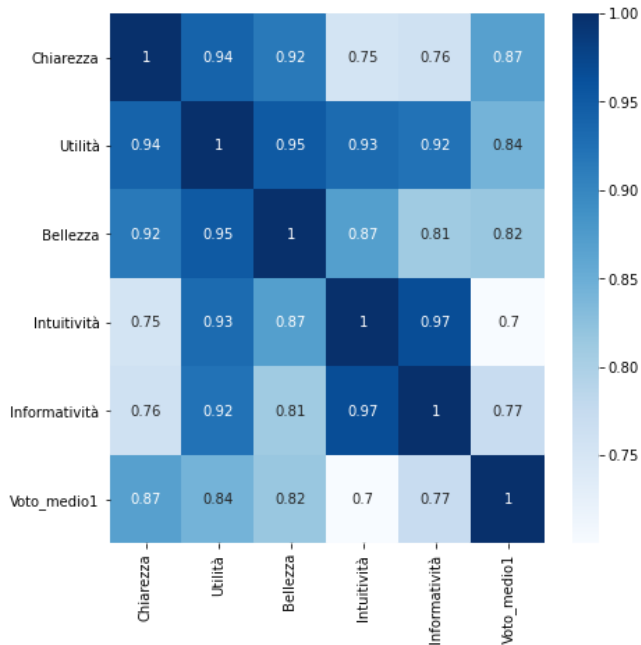
Con l’aggiunta di una domanda che chiede di esprimere una valutazione complessiva della visualizzazione (sempre con scala da 1 a 6).

Infine sono state inserite anche una domanda che punta a valutare la comprensione generale del progetto e una domanda sul grado di istruzione per avere anche un minimo di comprensione della possibile esperienza degli utenti con le visualizzazioni.

Ecco i risultati:

- **Domanda 1: Valutazione Terza Infografica**

Si è indagata l'eventuale correlazione tra gli attributi, per vedere come i voti di uno influenzassero quelli di un altro.



Emerge una correlazione molto positiva tra:

- Chiarezza e Utilità/Bellezza ($r = 0.94$; 0.92)
- Utilità e Bellezza/Intuitività/Informatività ($r = 0.95$; 0.93 ; 0.92)
- Intuitività e Informatività ($r = 0.97$)

Le altre correlazioni sono comunque rilevanti, ma meno significative

Figura 17. correlogramma valutazione domanda 1

Come si può notare dal grafico seguente, per la prima domanda si sono ottenuti voti molto positivi per tutti e sei gli attributi di analisi.

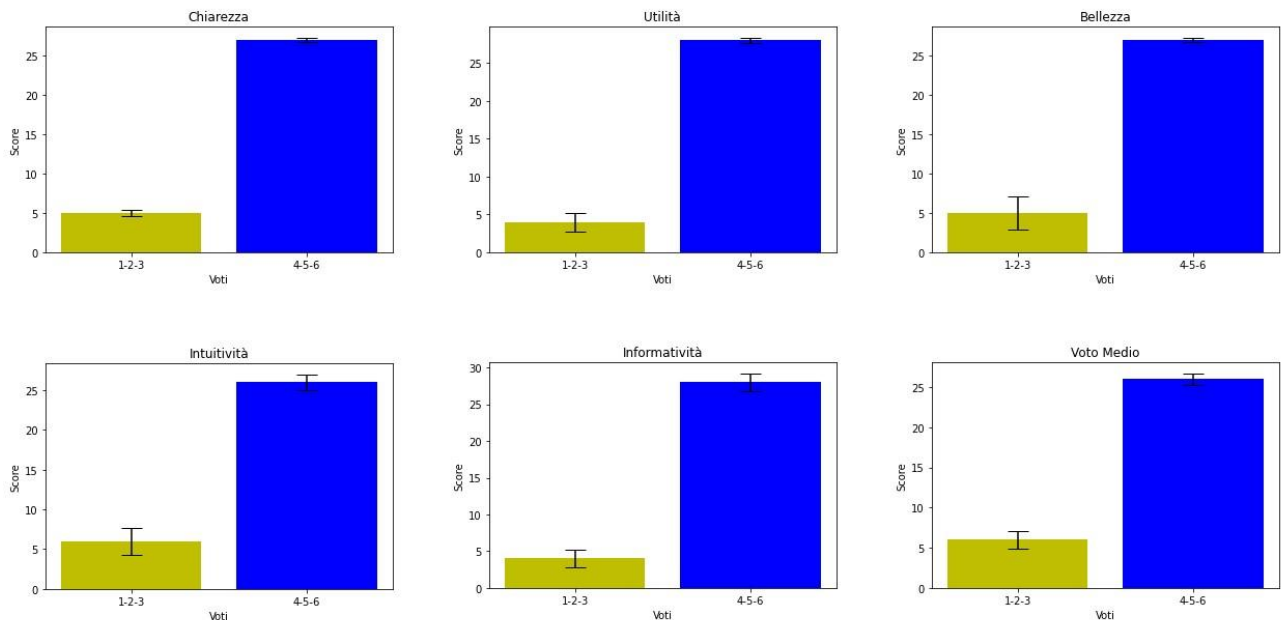
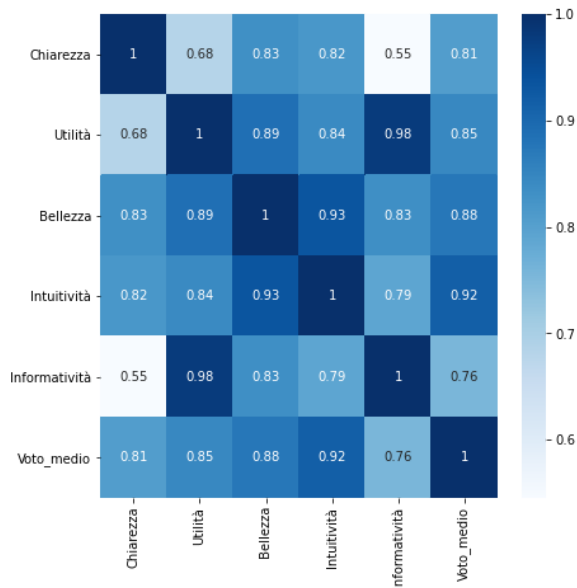


Figura 18. bar chart valutazione attributi domanda 1

● Domanda 2: Valutazione quarta infografica

Si è indagata l'eventuale correlazione tra gli attributi, per vedere come i voti di uno influenzassero quelli di un altro.



Si evidenzia una correlazione molto positiva tra:

- Utilità e Informatività ($r = 0.98$)

- Bellezza e Intuitività ($r = 0.93$)

- Voto_medio e Intuitività ($r = 0.92$)

Le altre correlazioni sono comunque rilevanti, ma meno significative

Figura 19. correlogramma valutazione domanda 2

Come si può notare dal grafico seguente, per la seconda domanda si sono ottenuti voti molto positivi per tutti e sei gli attributi di analisi.

L'utilità e l'informativa ottengono votazioni migliori rispetto alla domanda 1, questo era un risultato auspicabile alla vigilia, in quanto la visualizzazione proposta è più complessa, ma più esaustiva. La bellezza viene valutata in maniera "negativa", sempre in confronto alla domanda precedente.

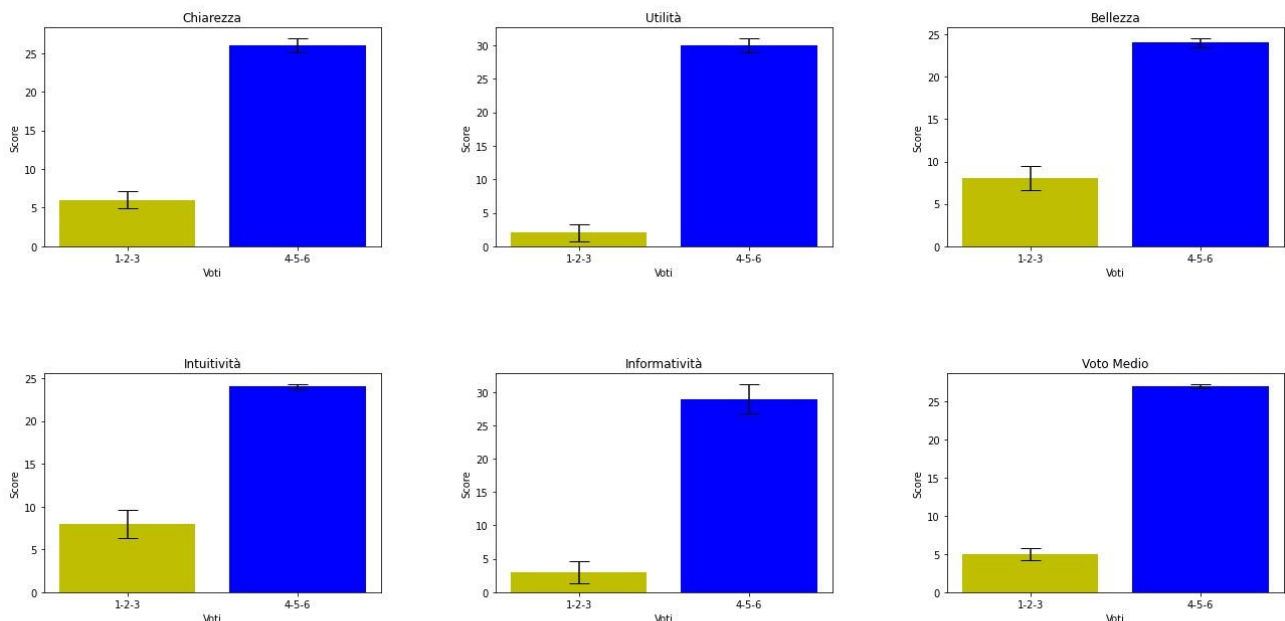
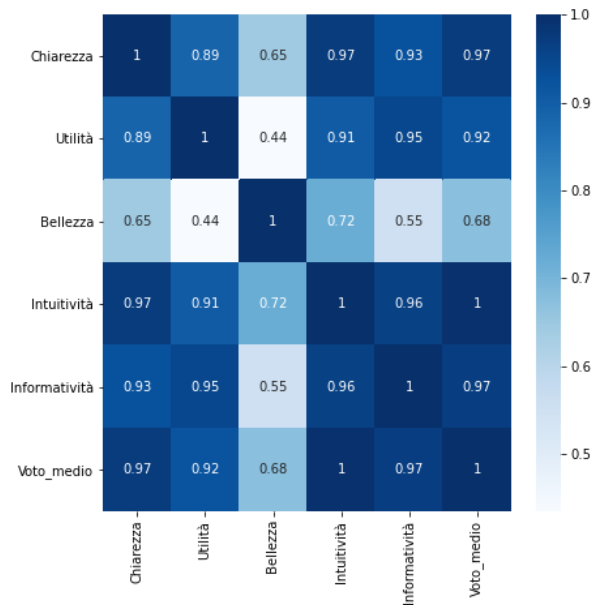


Figura 20. bar chart valutazione attributi domanda 2

● Domanda 3: Valutazione sesta Infografica

Si è indagata l'eventuale correlazione tra gli attributi, per vedere come i voti di uno influenzassero quelli di un altro.



E' evidente una correlazione molto positiva in quasi tutti gli attributi ad eccezione di Bellezza che presenta comunque correlazione positiva, ma nettamente inferiore.

Figura 21. correlogramma valutazione domanda 3

Essendo una visualizzazione simile alla precedente si sono ottenuti voti molto positivi per tutti e sei gli attributi di analisi.

Interessante il miglioramento dell'attributo Bellezza, probabilmente la scelta del colore è stata apprezzata.

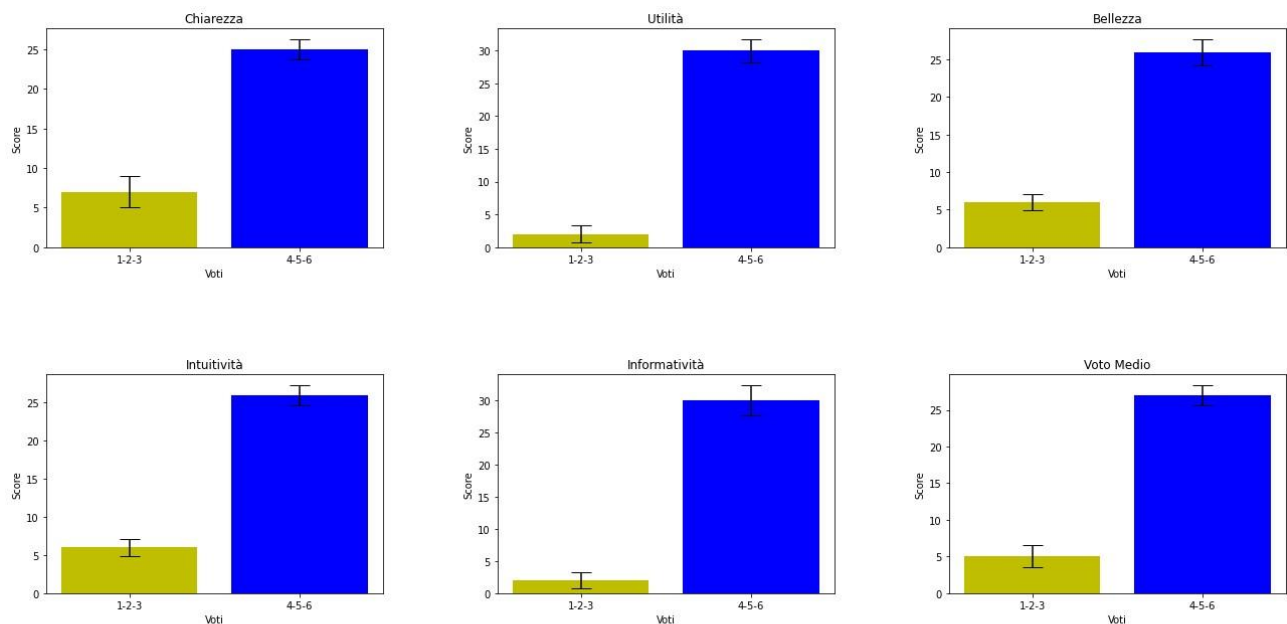
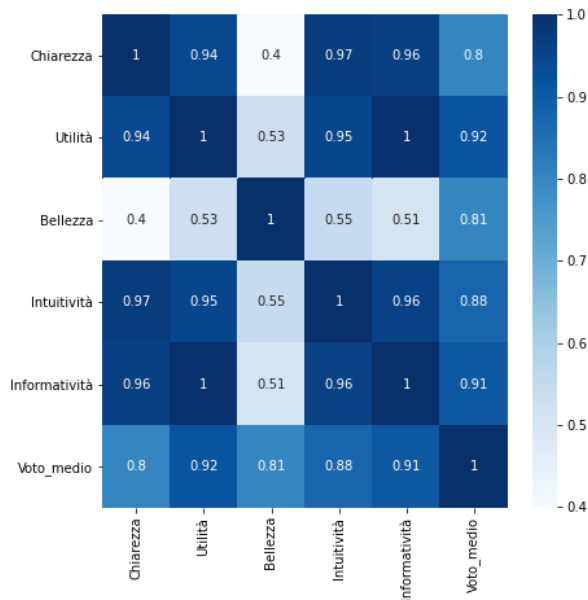


Figura 22. bar chart valutazione attributi domanda 3

- **Domanda 4: Valutazione Settima Infografica**

Si è indagata l'eventuale correlazione tra gli attributi, per vedere come i voti di uno influenzassero quelli di un altro.



Emerge una correlazione molto positiva in quasi tutti gli attributi ad eccezione di Bellezza e Voto_medio, che presentano comunque correlazione positiva, ma nettamente inferiore.

Figura 23. correlogramma valutazione domanda 4

Anche in questo caso si tratta di una visualizzazione simile alla precedente.

L'attributo bellezza risulta nettamente il peggiore, si paga probabilmente la scelta della scala di colore nero, che da un punto di vista estetico risulta essere infelice, ma da quello del messaggio sembra essere molto efficace, come si vede dal miglioramento generale degli altri attributi.

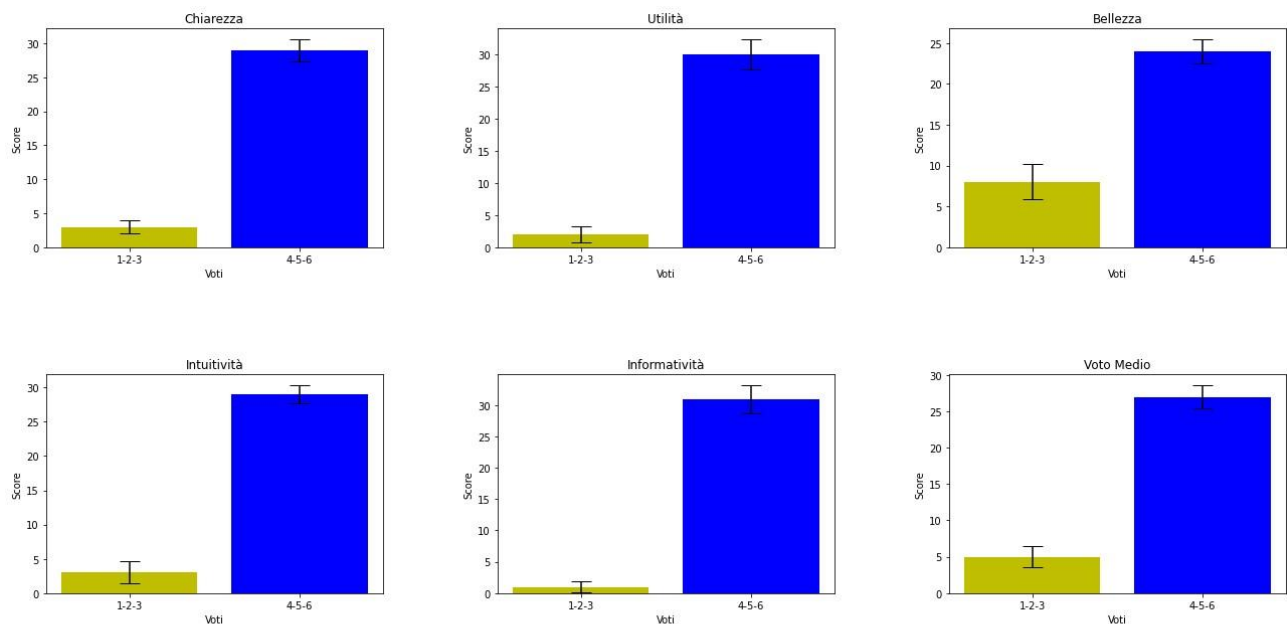
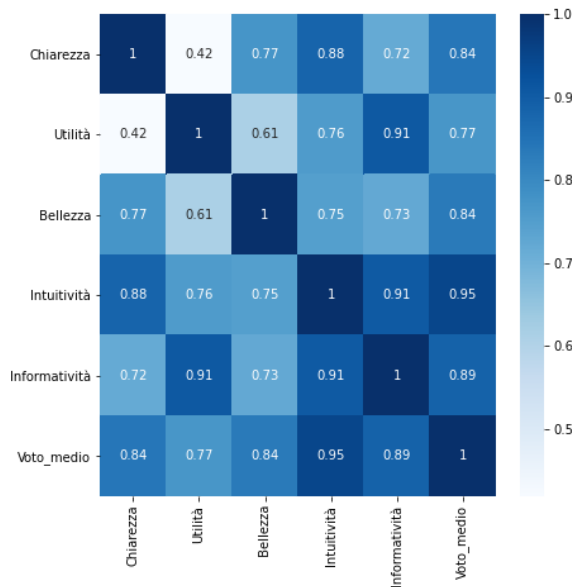


Figura 24. bar chart valutazione attributi domanda 4

● **Domanda 5: Valutazione ottava infografica**

Si è indagata l'eventuale correlazione tra gli attributi, per vedere come i voti di uno influenzassero quelli di un altro.



Si osserva una correlazione molto positiva tra:

- Utilità e Informatività ($r = 0.91$)

- Intuitività e informatività/voto_medio ($r = 0.91$; 0.95)

Le altre correlazioni sono comunque rilevanti, ma meno significative, spicca la correlazione più bassa tra chiarezza e utilità ($r=0.42$)

Figura 25. correlogramma valutazione domanda 5

Essendo una visualizzazione simile alla precedente si sono ottenuti voti molto positivi per tutti e sei gli attributi di analisi.

Tuttavia, data la maggior complessità dell'infografica, si può notare come per gli attributi "Chiarezza", "Intuitività" e "Informatività" ci sia stato un leggero aumento di votazioni più basse.

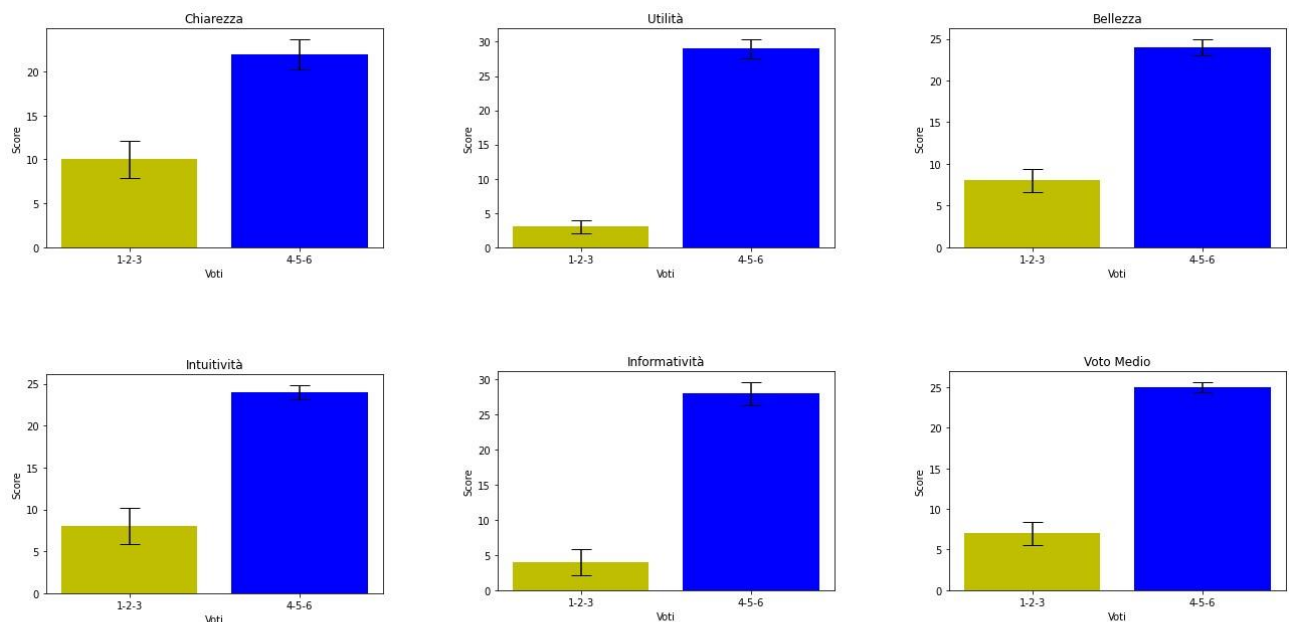


Figura 26. bar chart valutazione attributi domanda 5

6. Conclusioni

L'obiettivo del progetto era quello di mostrare come gli effetti del vaccino giochino un ruolo decisivo nella sconfitta del Covid-19. Per questo si sono analizzate le 2 variabili relative alla percentuale di popolazione vaccinata (dose 1, completa) e rispettivamente quelle relative a Indice RT, terapie intensive e decessi.

Dall'analisi si è andato a dimostrare la veridicità delle ipotesi¹⁷:

- Non esiste una correlazione tra percentuale popolazione vaccinata e indice RT, a dimostrazione del fatto che il vaccino non diminuisce la possibilità di prendere il Covid-19.
- Esiste una buona correlazione tra percentuale popolazione vaccinata e ricoveri terapia intensiva, visibile sia dal grafico che dallo studio relativo alla correlazione¹⁸.
- Esiste una buona correlazione tra percentuale popolazione vaccinata e numero di decessi, visibile sia dal grafico che dallo studio relativo alla correlazione¹⁹.

6.1 Sviluppi Futuri:

- A. Come si è visto, in questi ultimi mesi a preoccupare il Governo e le strutture ospedaliere in primis è una possibile terza ondata derivante dalla diffusione della variante Delta. Lo studio condotto in questo progetto, fatica a catturare al meglio l'effetto che la variante Delta potrà comportare per tutte le variabili target (RT positivi - terapie intensive - decessi), si intravede come i casi di positività siano aumentati sul finire del mese di luglio 2021. Per questo motivo, possibili sviluppi futuri potrebbero incentrarsi nel condurre uno studio più approfondito andando ad analizzare le variabili target facendo una distinzione per fasce d'età. Questo perché, la variante Delta, sembrerebbe colpire i più giovani (tra i 12 e i 25 anni), ovvero quell' range di età dove la campagna vaccinale è stata più tardiva. Ciò rafforza l'ipotesi iniziali su come i vaccini sono un potente strumento per contrastare:
- in primis i casi più gravi della malattia: ovvero persone che necessitano il ricovero in terapia intensiva e pazienti che decedono a causa del virus;
 - in secundis la diffusione del covid: è vero che un individuo vaccinato può contrarre il virus, ma ha poche probabilità di farlo mutare e diffonderlo a suo volta.
- B. Uno sviluppo futuro, potrebbe essere quello di andare a confrontare i dati tra le varie regioni, valutando come percentuale di popolazione vaccinata influenza i decessi e i casi di terapie intensive. Si suggerisce di calcolare le variabili decessi su popolazione e terapie intensive su popolazione. Inoltre per calcolare e confrontare le varie regioni si suggerisce di ottenere anche i dati relativi all'efficienza e alla qualità delle strutture ospedaliere nel territorio, poiché si ritiene che le variabili siano influenzate anche da questi aspetti, ritenuti di fondamentale importanza per questa analisi.
- C. Inoltre, per concludere, si potrebbe effettuare un'analisi di causalità. Questo perché lo studio della correlazione non implica un nesso di causalità tra le variabili studiate (vaccinazioni contro decessi/terapie intensive).

¹⁷ Cap 1.1(Obiettivi)

¹⁸ Cap 3.2(Correlazione)

¹⁹ Cap 3.2(Correlazione)

7 Fonti

[1] :

<https://www.medicisenzafrontiere.it/news-e-storie/news/coronavirus-sintomi-trasmissione/>

[2] Dataset Popolazione:

<https://github.com/italia/covid19-opendata-vaccini/blob/master/dati/platea.csv>

[3] Dataset Covid:

<https://github.com/pcm-dpc/COVID-19/blob/master/dati-regioni/dpc-covid19-ita-regioni.csv>

[4] Dataset Vaccini:

<https://github.com/italia/covid19-opendata-vaccini/blob/master/dati/somministrazioni-vaccini-latest.json>

[5] Dataset Indice RT :

<https://covid19.infn.it/iss/>