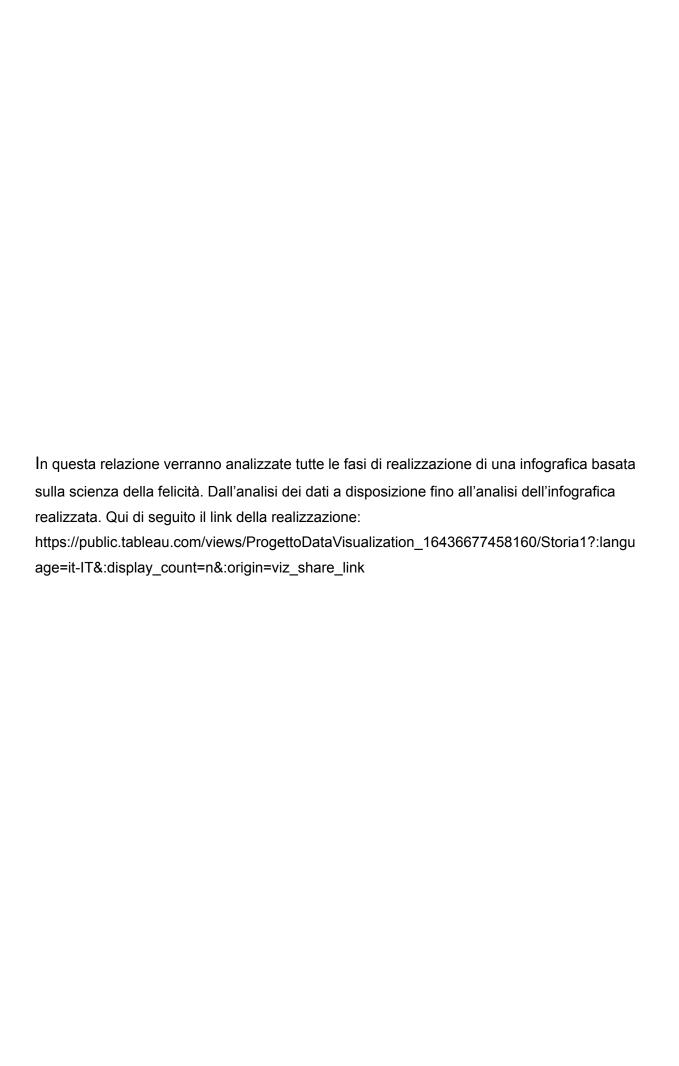
## Università degli Studi di Milano-Bicocca Corso di Laurea in Data Science



# Alla ricerca della felicità

Simone Farallo 889719 Edoardo Rana 828625 Michele Salvaterra 891109 Mirko Tritella 887196



## Indice

## Introduzione

- 1. La progettazione della data visualization
  - 1.1. Gli obiettivi
  - 1.2. La raccolta dei Dati
  - 1.3. Target di riferimento
- 2. Realizzazione
  - 2.1. Accessibilità
  - 2.2. Realizzazione finale
- 3. Valutazione
  - 3.1. Valutazione euristica
    - 3.1.1. Modifiche alla Data Visualization
  - 3.2. Test utente
    - 3.2.1. Risultati
  - 3.3. Questionario psicometrico
    - 3.3.1. Risultati
- 4. Conclusioni e implementazioni future
- 5. Fonti

## **Abstract**

Il seguente report riguarda la realizzazione di una data visualization, lo scopo del progetto è quello di studiare le differenze socio-economiche nei vari paesi e di come esse sono variate nel tempo.

I dati provengono dal World Happiness Report [1], sostanzialmente un sondaggio sullo stato della felicità globale, le visualizzazioni sono state create mediante il software Tableau e la fase di esplorazione e analisi dei dati è stata svolta con Python.

Analizzando i dati raccolti si evince che il punteggio di felicità può essere correlato a diverse variabili e appare chiara la differenza che c'è tra i paesi Nord Europei e i paesi Africani, è stata inoltre analizzata la variazione del punteggio della felicità nel tempo, facendo emergere alcune ipotesi sulle possibili cause della variazione.

Questo percorso ha portato alla realizzazione di due grafici interattivi dove l'utente può analizzare le differenze socio-economiche attraverso i punteggi delle variabili, con particolare attenzione al punteggio riguardante la felicità e come esso è cambiato nel tempo.

#### Introduzione

Il World Happiness Report è un sondaggio di riferimento sullo stato della felicità globale, che classifica 155 paesi per i loro livelli di felicità; ed è stato pubblicato dalle Nazioni Unite durante un evento che celebra la Giornata Internazionale della Felicità, il 20 marzo. Il rapporto continua ad ottenere un riconoscimento globale in quanto i governi, le organizzazioni e le società civili utilizzano sempre più gli indicatori di felicità per supportare le loro decisioni politiche. Esperti di spicco in diversi campi affermano che le misurazioni del benessere possano essere utilizzate efficacemente per valutare il progresso delle nazioni. I punteggi e le classifiche sulla felicità utilizzano i dati del Gallup World Poll [2], infatti questi punteggi si basano sulle risposte date dagli intervistati al sondaggio. Per assegnare il punteggio viene utilizzata la scala di Cantril, che opera nel seguente modo: chiedere agli intervistati di pensare a una scala in cui la migliore vita possibile sia identificata dal numero dieci e la peggiore dallo zero, e di valutare la propria vita attuale su questa scala. Le scale di Cantril sono state sviluppate a partire dagli anni '50 e richiedono un'attività di valutazione su una serie di oggetti. Essa è operata scegliendo un punteggio entro un arco di posizioni semanticamente ancorate ad un valore minimo e a un valore massimo. Tra i due ancoraggi le posizioni degli individui vengono identificati con numeri naturali o con punti grafici (a cui vengono assegnati valori).

Queste scale riducono al minimo l'autonomia semantica delle categorie, mentre i punteggi provengono da campioni rappresentativi a livello nazionale e usano i pesi Gallup per rendere le stime rappresentative.

In questo report, quindi, viene presentata la progettazione di una data visualization con l'obiettivo di dare una panoramica globale sulla felicità del mondo.

La progettazione è stata creata in modo iterativo, in una prima fase sono stati raccolti dati e informazioni sull'argomento, successivamente è stata progettata una data visualization che potesse comunicare al meglio l'obiettivo.

La fase successiva è stata la valutazione euristica, ovvero sono stati usati dei principi di buona progettazione con cui si è valutato il prodotto al fine di capire se qualche principio è stato violato o non seguito, in questo modo si rilevano i problemi di usabilità. Inoltre è stato creato un test utente che permette di valutare efficacia ed efficienza ed un questionario volto a raccogliere pareri soggettivi da parte di utenti esterni al processo di progettazione, per valutare la soddisfazione. Infine, abbiamo apportato delle modifiche alla data visualization in base ai risultati ottenuti durante la fase di valutazione.

## 1. La progettazione della data visualization

#### Gli obiettivi

L'idea alla base di questo lavoro è stata quella di compiere un percorso simbolico tra alcune caratteristiche riguardanti la scienza della felicità, misurando ciascuno dei seguenti fattori:

- Punteggio di felicità: è un indicatore ottenuto tramite il sondaggio World Happiness
   Report, il cui intervallo è compreso tra 0 e 10;
- *PIL pro-capite*: è un indicatore statistico ottenuto dal rapporto tra il valore del PIL del Paese nel periodo considerato e il suo numero di abitanti;
- Sostegno sociale: questo indicatore misura gli aiuti, i contributi,e le agevolazioni per il sostegno economico e sociale di famiglie o persone in difficoltà;
- Aspettativa di vita: rappresenta l'aspettativa di vita alla nascita;
- Libertà di espressione: è un indicatore che misura la libertà di opinione, di ricevere o
  di comunicare informazioni o idee senza che vi possa essere ingerenza da parte
  delle autorità pubbliche e senza limiti di frontiera;
- Percezione della corruzione: è un indicatore riguardante la corruzione nel Paese, più
   è basso il punteggio, più c'è fiducia nelle istituzioni;
- Generosità: rappresenta la nobiltà d'animo che si manifesta soprattutto come altruismo, prontezza al sacrificio e al perdono;
- Popolazione: La quantità delle persone che vivono in un determinato territorio.

L'obiettivo della nostra data visualization è quello di offrire agli utenti una panoramica dei punteggi ottenuti dal World Happiness Report per ogni Paese e far cogliere le differenze sociali, culturali ed economiche.

#### Le **domande di ricerca** sono le seguenti:

- Esiste una correlazione tra la ricchezza di uno Stato e la felicità dei suoi cittadini e come si comportano i punteggi delle altre variabili?
- Quali sono i Paesi con una maggiore variazione del punteggio di felicità nel tempo e perchè?

## Le **ipotesi** che si vogliono verificare sono le seguenti:

- Maggiore è il PIL di una nazione, maggiore è il punteggio di felicità.
- Determinati eventi storici influiscono sul punteggio di felicità, nel corso del tempo.

#### La raccolta dei dati

Dopo aver fissato gli obiettivi, la fase successiva è stata la raccolta dati.

I dataset utilizzati sono stati presi dal World Happiness Report e sono stati raccolti dalla Gallup Inc., una società americana di analisi e consulenza con sede a Washington, D.C., fondata da George Gallup nel 1935. La società è nota per i suoi sondaggi d'opinione condotti in tutto il mondo.

I dati nel dataset comprendono i punteggi di numerose variabili, ma sono state prese in considerazione solo le seguenti: *felicità*, *PIL pro-capite*, *sostegno sociale*, *aspettativa di vita*, *libertà di espressione*, *percezione della corruzione*, *generosità*.

A questi punteggi è stato affiancato un dataset riguardante la popolazione mondiale per avere un quadro più completo, questi ultimi sono stati raccolti da <u>Our World In Data</u> [3], un sito di pubblicazione scientifica appartenente alla categoria di editoria digitale, che presenta ricerca empirica e dati che mostrano come stanno cambiando le condizioni di vita nel mondo.

#### Target di riferimento

Il target di riferimento della data visualization riguarda coloro che sono interessati ad avere una panoramica dei dati riguardanti la qualità di vita nei vari Paesi e come essa è cambiata nel tempo. Nel dettaglio riguarda le organizzazioni e le società civili che utilizzano sempre più frequentemente gli indicatori di felicità per supportare le loro decisioni politiche.

#### 2. La realizzazione delle Data Visualization

Dopo una fase esplorativa dei dati è stato svolto un percorso di analisi per poter scegliere le visualizzazioni più adatte per la nostra domanda di ricerca. Di seguito ci sono alcuni esempi usati come punti di partenza.

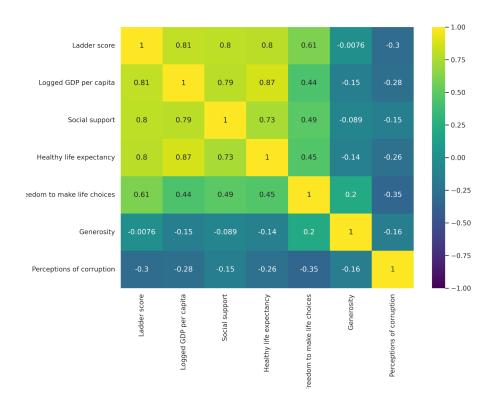


Fig 1 - Matrice di correlazione dei punteggi.

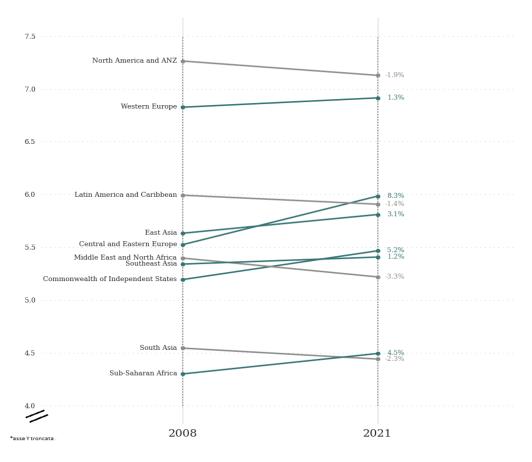


Fig 2 - Variazione percentuale del punteggio di felicità dal 2008 al 2021 per ogni area geografica nel mondo.

I risultati hanno evidenziato quali variabili fossero più correlate con il punteggio di felicità, rispettivamente: *PIL*, *aspettativa di vita* e *supporto sociale*. Anche la variabile *freedom to make a life choices*, risulta correlata ma in maniera più parsimoniosa. La variazione del punteggio di felicità è stata più alta in *Central and Eastern Europe* con un aumento del 8.3%, seguito da *Commonwelt of Indipendent States* che ha avuto un aumento del 5.2%, al contrario il *North America and ANZ* ha avuto un calo del 4.9%, seguito da *Middle East and North Africa* con un calo del 3.3%. Sulla base di questi risultati abbiamo deciso di mostrare la correlazione tra il punteggio di felicità e il PIL pro capite.

Il primo grafico è uno scatter plot, nonché un diagramma a dispersione, il quale permette di visualizzare la correlazione tra due variabili e al tempo stesso la distribuzione dei dati. All'interno della visualizzazione ogni punto rappresenta un Paese, sull'asse delle ascisse viene rappresentato il PIL pro-capite, mentre sull'asse delle ordinate il punteggio della felicità. La grandezza di ogni punto, rappresenta il numero di abitanti del Paese considerato. E' presente una divisione dei paesi in tre gruppi: i più felici, altri, i più tristi; ad ogni gruppo è assegnato un colore. Allo scatter plot è affiancato un bar-chart, esso viene utilizzato per rappresentare i dati visivamente con barre di diversa altezza o lunghezza.

Quest'ultimo è usato per supportare lo scatterplot e dare ulteriori informazioni sui punteggi ottenuti da ogni paese e quindi ottenere una panoramica completa.

Le variabili sono state normalizzate, ottenendo così valori compresi tra 0 e 1.

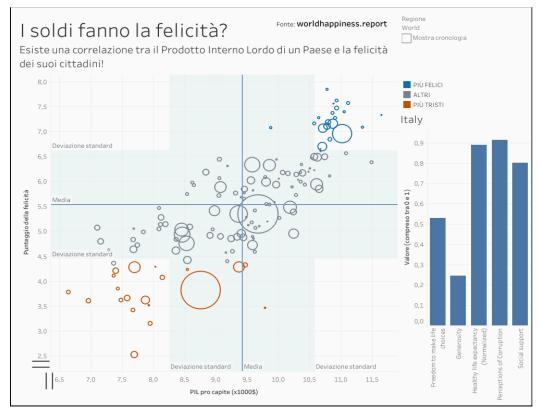


Fig. 3 - Bozza della prima dashboard contenente scatter-plot e il bar-chart

Il secondo grafico mostra la differenza di punteggio tra il 2011 e il 2021 per ogni paese ed è rappresentata da una mappa coropletica.

I Paesi sono contraddistinti da un colore divergente che va dal blu all'arancione, che rappresentano rispettivamente un aumento o una diminuzione del punteggio di felicità. Alla mappa è stata affiancata una time series, che mostra l'andamento del punteggio di felicità nei vari anni di ogni paese. In alcuni grafici, all'interno delle serie storiche di un determinato paese, sono stati riportati gli eventi che in quell'anno lo hanno influenzato. Per provare a dare una motivazione dell'eventuale miglioramento o peggioramento abbiamo fatto ricerche sull'enciclopedia di Wikipedia [4].

Zimbabwe: nel 2013, in Zimbabwe, la crisi dell'acqua e benzina provoca forti proteste nel Paese.

Afghanistan: nel 2017, i talebani attaccano il quartier generale dell'esercito afgano uccidendo oltre 200 soldati.

Jordan: nel 2011, in Giordania, il Re scioglie il governo e dà inizio alla crisi politica.

Venezuela: nel 2012, in Venezuela, ha inizio una grave crisi economica, politica e sociale.

*India*: nel 2018, una grave crisi finanziaria causata da una nuova legge sulla bancarotta e nuovi sistemi finanziari, ha messo in difficoltà l'economia del paese.

Brasile: tra il 2014 e il 2017 il Brasile ha subito una grave crisi economica.

*Egitto*: nel 2013, il malcontento verso il Presidente Morsi scatena delle proteste nelle piazze, portando, addirittura, ad un colpo di stato.

*Giappone*: nel 2011, in Giappone è avvenuto un terremoto e, successivamente, un maremoto che ha causato circa 16000 vittime e ha avuto gravi conseguenze economiche.

*Bulgaria*: nel 2014, la Commissione Europea annuncia la possibilità per i cittadini Bulgari di lavorare nei 28 Paesi membri dell'UE.

*Turchia:* nel 2016, le forze armate turche falliscono un golpe militare contro il regime del presidente Erdogan.

Ucraina: nel 2014, ricomincia il conflitto russo-ucraino.

*Niger:* il 2015 inizia con il Niger Raid: una serie di attentati firmati da Boko Haram in due città importanti del Paese.

South Africa: a fine 2012, avviene il cosiddetto Massacro di Marikama: 34 lavoratori furono uccisi in uno scontro con le forze di polizia.

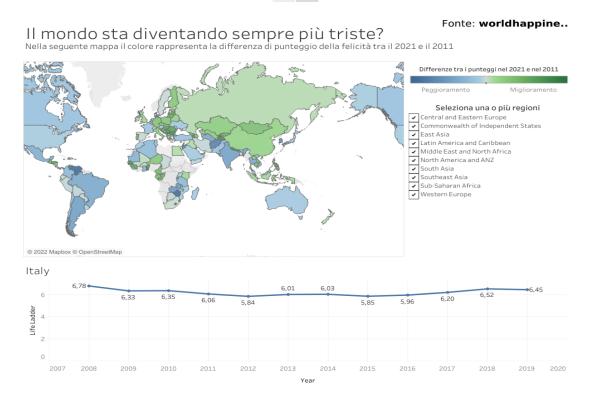


Fig. 4 - Bozza della seconda dashboard contenente mappa coropletica e serie storica

## L'accessibilità

Per dare una visione a colpo d'occhio delle nazioni più felici e infelici si è dato un colore diverso in base al punteggio di felicità. Tableau permette di visualizzare le informazioni di ogni stato presente nel grafico semplicemente posizionando il cursore sul pallino corrispondente, di conseguenza le persone affette da daltonismo possono comunque fruire delle informazioni relative agli Stati. Nonostante ciò, è stato deciso di cambiare i colori, cercando di utilizzare una palette colore che aiutasse i daltonici. Il colore nelle varie data visualization e nell'infografica non deve essere trattato in modo isolato ma deve essere considerato come un aspetto fondamentale nel design incentrato sull'utente.

## 3. Valutazione

Nella valutazione di qualità della data visualization sono state adottate le seguenti metodologie:

La valutazione euristica che ha permesso di individuare le problematiche di usabilità relative alla nostra data visualization;

Il questionario psicometrico;

Il test utente che ha permesso di valutare *l'efficacia*, misurando il tempo necessario per eseguire un determinato compito e *l'efficienza*, considerando i compiti eseguiti con successo, della data visualization.

#### La valutazione euristica

La valutazione euristica ha coinvolto 6 utenti ai quali abbiamo chiesto di interagire liberamente per qualche minuto con la data visualization commentando a voce alta quello che stavano facendo. Interpretando i comportamenti e commenti degli utenti, abbiamo individuato alcune problematiche relative alla data visualization:

Utente 1	A primo impatto l'utente non coglie che l'infografica ha tre colori, dopo pochi secondi c'è stupore nel fatto che i paesi più felici siano di colore blu e i paesi più infelici siano di colore arancione.
Utente 2	L'utente riesce a comprendere il grafico facilmente ma accenna al fatto che la deviazione standard va specificata, non riesce a cogliere l'intervallo che la comprende poichè l'azzurro è troppo chiaro ed è poco distinguibile dal bianco.
Utente 3	L'utente alla visione delle variabili non comprende se il valore è correlato positivamente o negativamente, infatti pone l'attenzione sulla percezione della corruzione poiché è l'unica inversamente proporzionale rispetto alla felicità.
Utente 4	L'utente dopo un'attenta analisi dei grafici, non ritiene appropriata la scelta di progettazione riguardante il nome della legenda, la posizione e il contesto.
Utente 5	L'utente denota la mancanza della barra di ricerca.
Utente 6	L'utente interagisce senza particolari problemi con il primo grafico, nel secondo grafico riscontra un problema di usabilità nel cercare gli stati che vuole perchè non è specificato che la ricerca ha come lingua l'inglese.

#### Modifiche alla data visualization:

In seguito a quanto emerso dalla valutazione euristica sono state apportate diverse modifiche ad entrambe le data visualization.

1. Al sottotitolo è stata data l'importanza di dare un quadro generale della visualizzazione, in più sono state aggiunte le istruzioni per l'uso e un sommario. Si è deciso di cambiare una parte del design, la barra di ricerca è stata valorizzata, il bar chart è stato reso orizzontale, i colori grigi della categoria altri sono stati opacizzati per rendere più chiara la visione dei paesi con i punteggi migliori e peggiori.

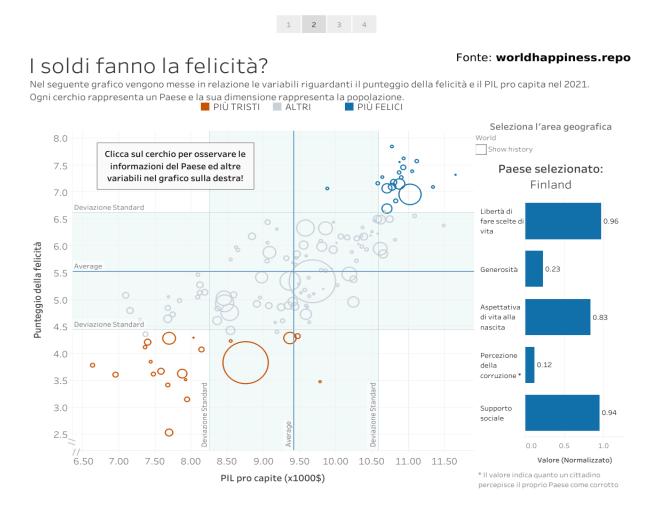


Fig. 5 - Prima dashboard modificata dopo le valutazione euristica

2. Il design ha subito un cambiamento, il contesto è stato integrato con una descrizione della visualizzazione. Sono state aggiunte le istruzioni d'uso, il colore della mappa è stato modificato, diverge tra un colore arancione se la variazione del punteggio di felicità peggiora negli anni e un colore blu se il punteggio migliora.

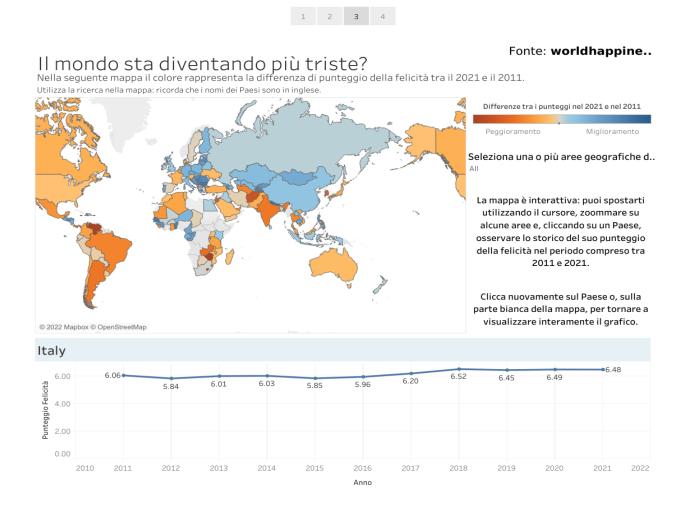


Fig 6 - Seconda dashboard modificata dopo la valutazione euristica

#### Il test utente

Dopo aver risolto le problematiche riscontrate nella valutazione euristica, abbiamo sottoposto 12 soggetti al test utente. Durante il test abbiamo somministrato 3 compiti (task) agli utenti; misurando il tempo di esecuzione del task abbiamo potuto valutare l'efficienza, considerando il numero di task portati a termine abbiamo potuto valutare l'efficacia.

Sono stati formulati dei task di facile comprensione per utenti con poca o nessuna esperienza nell'utilizzo di Tableau.

Nella tabella a seguito vengono descritti i compiti richiesti agli utenti:

	obiettivo	compito
Task 1	Interazione con la legenda. Comprensione del grafico e interpretazione delle variabili.	Osservando il primo grafico, indica qual è il paese più felice, elencando i punteggi delle variabili.
Task 2	Filtrare una regione del mondo utilizzando l'apposita barra.	Osservando il primo grafico seleziona una regione specifica, considera il paese più triste ed indica la variabile con il punteggio più basso.
Task 3	Ricercare uno Stato specifico nella barra di ricerca. Comprensione del motivo per cui il punteggio di felicità negli anni è cambiato.	Osservando il secondo grafico, seleziona uno dei paesi che ha avuto un grande peggioramento dal 2011 al 2021 e spiega quale evento storico si è verificato.

## I risultati

Per valutare *l'efficienza* abbiamo cronometrato i tempi di esecuzione di ogni task effettuato dall'utente. Nella tabella vengono mostrati i tempi delle task espressi in secondi:

Utenti	Task 1	Task 2	Task 3
U1	25"	31"	36"
U2	16"	40"	30"
U3	29"	46"	46"
U4	63"	88"	58"
U5	41"	58"	42"
U6	33"	42"	49"
U7	27"	66"	31"
U8	24"	51"	45"
U9	38"	35"	56"

Utenti	Task 1	Task 2	Task 3
U1	25"	31"	36"
U10	21"	47"	42"
U11	26"	39"	32"
U12	31"	52"	55"

Nel seguente grafico sono riportati i tempi di esecuzione dei tre compiti:

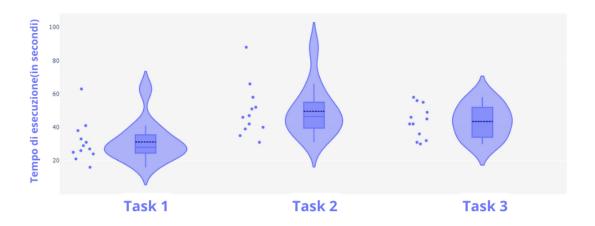


Fig 7 - Violin Plot e Box Plot; ogni pallino corrisponde ad un utente.

Il grafico è composto da: i violin plot che permettono di vedere a colpo d'occhio la distribuzione dei dati; dai boxplot che permettono di osservare dove si posiziona il valore massimo, il valore minimo, la media, la mediana e l'ampiezza della deviazione standard delle distribuzioni dei task; a questi abbiamo affiancato dei puntini che mostrano il tempo di esecuzione che ciascun utente ha impiegato per completare i task.

Dai risultati osservati, emerge che la seconda task ha avuto un tempo di esecuzione medio leggermente maggiore rispetto alle altre, la distribuzione risulta distorta a causa di un utente che ha impiegato ben 88 secondi per svolgere il compito; a detta dell'utente c'è stata confusione nell'interpretazione del grafico.

Inoltre si può notare come per la seconda task i tempi di esecuzione del compito siano più vari rispetto alle altre, dove i tempi di esecuzione si collocano in un range più ristretto. L'efficacia è massima essendo che tutti gli utenti hanno portato a termine i loro compiti.

## Questionario psicometrico

Successivamente alla valutazione euristica e al test utente, si è scelto di somministrare un questionario per valutare la qualità della data visualization utilizzando la Cabitza-Locoro Scale. Questa è composta da due sezioni. Nella prima sezione l'utente deve valutare la qualità dell'infografica attribuendo un valore su un continuum che va da 1(pochissimo) a 6 (moltissimo) a ciascuno dei seguenti aggettivi:

- Utile
- Chiara
- Informativa
- Bella
- Intuitiva

Nella seconda sezione viene invece richiesto all'utente di valutare, su un continuum che va da 1 (bassissimo) a 6 (altissimo), il valore complessivo percepito della data visualization. Il target al quale è stato sottoposto il questionario è composto da studenti universitari del corso di Data science e non solo. Al questionario hanno partecipato 27 soggetti.

I risultati
Per la prima infografica i risultati sono i seguenti:

	Utile	Chiara	Informativa	Bella	Intuitiva	Valore complessivo
1	0	0	0	1	1	0
2	2	3	2	2	7	1
3	3	10	4	3	5	2
4	9	4	7	7	6	9
5	11	7	9	7	4	11
6	2	3	5	7	4	4

Per la seconda infografica i risultati sono i seguenti:

	Utile	Chiara	Informativa	Bella	Intuitiva	Valore complessivo
1	0	0	0	0	0	0
2	2	0	1	2	0	0

3	6	6	4	4	7	1
4	9	7	10	10	4	8
5	6	7	7	4	9	13
6	4	7	5	7	7	5

Di seguito viene mostrata l'analisi dei risultati del questionario, entrambi i grafici vengono rappresentati con dei divergent stacked bar charts((noto anche con il nome di back-to-back stacked bar Chart). Vengono raggruppati i valori, in modo da considerare i valori centrali (3 e 4) come incerti.

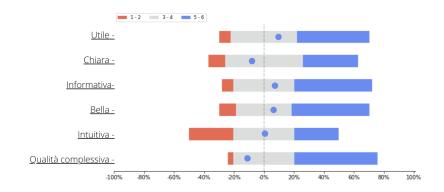


Fig 8 - Dati raccolti per mezzo del Questionario psicometrico - Divergent Stacked Bar Chart, prima visualizzazione

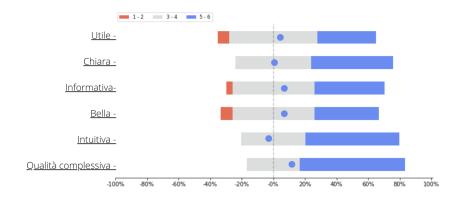


Fig 9 - Dati raccolti per mezzo del Questionario psicometrico - Divergent Stacked Bar Chart , seconda visualizzazione

Il primo grafico mostra risultati soddisfacenti per quanto riguarda la qualità complessiva, d'altra parte risulta poco intuitiva e i restanti parametri tendono a rimanere su una scala positiva. Il secondo grafico ha il valore più alto che è la qualità complessiva della

visualizzazione, l'utilità e la bellezza hanno gli stessi valori. Il campione di soggetti tende a rientrare maggiormente nei valori centrali, tre e quattro anche se complessivamente possiamo notare che i risultati tendono verso valori positivi, maggiori risposte corrispondenti al valore 4 rispetto al valore 3.

La fase di valutazione si conclude con lo studio delle correlazioni tra i vari punteggi, dove viene utilizzato un correlogramma, in cui è possibile osservare i risultati ottenuti.

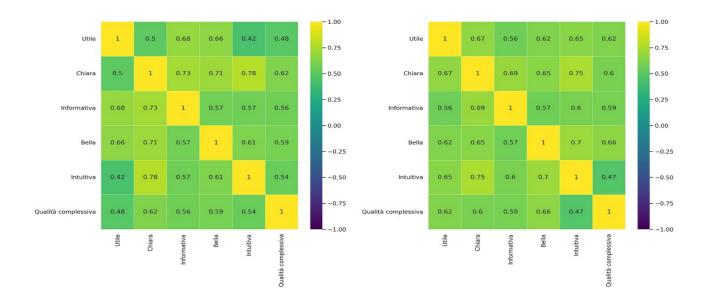


Fig 10 - Correlogramma questionario psicometrico

## Conclusione e implementazioni future

In conclusione la realizzazione dell'infografica ha evidenziato le differenze socio-economiche nei vari paesi e di come esse sono variate nel tempo. La prima domanda di ricerca riguardava la correlazione tra la ricchezza di uno Stato e la felicità dei suoi cittadini, la seconda i paesi con una maggiore variazione del punteggio di felicità nel tempo e la motivazione. Le ipotesi prefissate sono state due, la prima è che se è maggiore il PIL di una nazione allora maggiore è il punteggio di felicità, la seconda è che determinati eventi storici influiscono sul punteggio di felicità nel corso del tempo.

Il grafico mostra come la maggior parte dei paesi ricchi ha un punteggio alto di felicità elevato e viceversa per quelli più poveri.

Ciò che emerge dai dati è che per il quarto anno consecutivo la Finlandia è lo stato con il punteggio di felicità più alto, 9 Stati su 10 con il punteggio maggiore sono in Europa e 7 di quelli peggiori appartengono all'Africa.

Inoltre è interessante vedere come gli eventi storici che hanno coinvolto certi paesi hanno avuto conseguenze positive o negative sul punteggio della felicità.

La valutazione euristica ha permesso di risolvere i problemi di usabilità che violavano i principi di Design di buona progettazione.

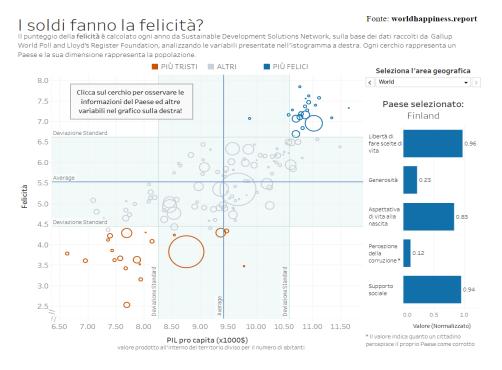
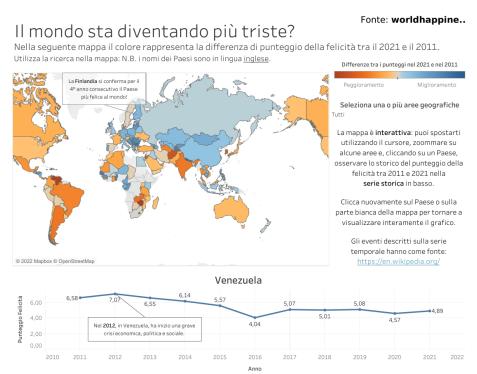


Fig 10 - Realizzazione finale della prima dashboard



1 2 3

Fig 11 - Realizzazione finale della seconda dashboard

# Fonti

- [1] World Happiness Report https://worldhappiness.report/
- [2] Gallup Https://www.gallup.com/analytics/349487/gallup-global-happiness-center.aspx
- [3 Our World in Data https://ourworldindata.org/
- [4] Timeline of events https://it.wikipedia.org/wiki