

UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Scuola di Scienze

Corso di Laurea Magistrale in Data Science

## Data Management and Visualization

**#EARTHDAY**

**Gruppo di lavoro:**

*Alberto Carlone matr. 726894*

*Davide Miori matr. 813692*

*Alice Ondeì matr. 826399*

**Anno Accademico 2020-2021**

## Introduzione

Il cambiamento climatico da anni è al centro di dibattiti internazionali a causa dell'impatto che sta avendo sulle nostre vite. Per dare una svolta alla situazione ambientale che sempre più preoccupa, molti stati hanno aderito al protocollo di Kyoto, trattato internazionale (pubblicato l'11 dicembre 1997 e ratificato nel 2004) che ha come oggetto principale il surriscaldamento globale.

Tale protocollo è posto alla base del UNFCCC (*United Nations Framework Convention on Climate Change*, noto anche come "Accordi di Rio") a cui hanno aderito 197 stati; tra questi 43 fanno parte del cosiddetto Allegato I (Annex I), ossia un insieme di stati che si sono impegnati nel tempo nell'adozione di politiche nazionali atte alla diminuzione delle emissioni e, in generale, nel mitigare gli effetti negativi del cambiamento climatico. Di questi 43 stati ve ne sono 24 facenti parte dell'Allegato II (Annex II) che si impegnano ulteriormente, sia in termini economici sia in termini di politiche nazionali, nella lotta al cambiamento climatico. L'Unione Europea, intesa nel suo insieme, è considerata separatamente oltre agli altri 23 membri dell'Allegato II.

A questi stati, spesso virtuosi in termini relativi alle politiche climatiche adottate, si contrappone l'insieme delle cinque maggiori economie in via di sviluppo, chiamato BRICS (acronimo degli stati che ne fanno parte: Brasile, Russia, India, Cina e Sudafrica). Queste nazioni sono frequentemente oggetto di critiche riguardo l'impatto ambientale delle loro politiche legate all'industrializzazione.

Il lavoro alle spalle di questo elaborato mira a studiare la situazione socioeconomica di 36 stati, confrontandola a una serie di indicatori che raccontano come il singolo paese si approcci al cambiamento climatico. Per tale analisi si prendono in considerazione i 23 membri dell'Allegato II, 3 membri del BRICS - fatto salvo per la Cina e la Russia, a causa della mancanza di dati provenienti dal social network Twitter - e altre 10 nazioni particolarmente attive sull'argomento.

In particolare, sono stati utilizzati indici come la qualità dell'aria o la quantità di emissioni inquinanti per indagare se esista una relazione con la situazione economica e sociale di ogni paese. Verrà inoltre studiato come la popolazione degli stati scelti risponda a queste problematiche mediante un'analisi del sentimento dei tweet pubblicati in eventi attinenti al tema in questione. Ciò verrà fatto confrontando l'indice di polarità medio relativo al singolo paese e il numero di utenti Twitter che si sono preoccupati di lasciare una propria opinione sul social.

## Velocità

Tra i social più utilizzati al giorno d'oggi si trova Twitter, piattaforma in cui è possibile condividere liberamente (o quasi) post relativi a ciò che accade nel mondo. Per questa ragione si è deciso di recuperare tweet con specifici hashtag relativi al cambiamento climatico, pubblicati durante la Giornata della Terra (avvenuta il 22 Aprile) con l'intento di salvarne il contenuto e la location associata all'autore.

Per fare ciò ci si è avvalsi della libreria **Tweepy** - che permette di effettuare streaming in tempo reale - e di **Kafka**, il quale ci ha permesso di salvare i dati per recuperarli in un secondo momento. Gli hashtag utilizzati al fine di fare una selezione dei tweet sono diversi; tra questi è bene segnalare quelli principalmente presi in considerazione: *#climatechange*, *#earthday* e *#plasticfree*, i quali sono stati tradotti in diverse lingue per facilitarne la ricerca in paesi differenti.

Prima di effettuare l'analisi testuale dei tweet, tramite la funzione **Geocode** della libreria **Geopy**, sono state associate le città di provenienza dei corrispondenti account Twitter al loro paese di provenienza, per poi essere tradotti in lingua inglese. Poiché la location del profilo Twitter è totalmente personalizzabile dall'utente, è stata necessaria una pulizia manuale di luoghi geografici non geolocalizzabili (esempio: World, Everywhere, Somewhere, nomi di interi continenti, etc). Successivamente il

nome dei paesi è stato normalizzato mediante un dizionario ad hoc, al fine di avere nomi comparabili con gli altri dati.

## Sentiment Analysis e Text Mining

La fase di **Sentiment Analysis** inizia mediante la traduzione in lingua inglese del testo di ogni tweet. A causa della presenza di parole inserite come hashtag, quindi scritte senza la presenza dello spazio, è necessaria una successiva verifica e pulizia manuale affinché anche tali parole, molto frequenti negli ambienti social, vengano processate correttamente.

Si procede quindi alla costruzione del dataset su cui verrà poi svolta la fase di sentiment, decidendo di mantenere come attributi solamente il testo originale, quello tradotto, la lingua con cui era scritto il tweet originale e il paese.

Al fine di rendere più veloce il processo di traduzione, è stata scritta una funzione con cui si analizza preliminarmente la lingua del tweet e procedendo alla traduzione soltanto se il tweet non è scritto in lingua inglese.

I testi così tradotti vengono poi ripuliti da numeri, cancelletto (#) e link. Tutti i segni di punteggiatura vengono mantenuti poiché la funzione **TextBlob**, utilizzata per svolgere l'analisi, è in grado di riconoscere e tenere in considerazione le emoticon nel computo delle metriche tipiche della Sentiment Analysis. Sono stati quindi calcolati i valori di polarity, subjectivity e di score (positivo, negativo e neutro). Si decide poi di considerare come parametro chiave per assegnare il sentiment al tweet il valore assunto dalla polarity: tale metrica assume valori all'interno dell'intervallo  $[-1, 1]$ , verranno quindi considerati negativi i tweet con polarity negativa, viceversa nel caso positivo.

Successivamente si è deciso di procedere applicando tecniche di *Text Mining* per individuare le parole (o gruppi di parole) maggiormente utilizzate. A tal fine il testo di ogni tweet è stato ripulito da punteggiatura, caratteri speciali e *stopwords* per poi applicare due funzioni che procedessero alla *tokenizzazione* del testo e allo *stemming* delle parole.

È stata poi definita una funzione ad hoc che restituisce come output un dataframe contenente l'elenco delle  $n$  parole più utilizzate (con  $n$  parametro di input) associate alla frequenza corrispondente. Tale funzione richiede due aggiuntivi parametri di ingresso: il testo e il numero di parole da considerare all'interno di una singola occorrenza (una singola parola, bigrammi, trigrammi, ...).

Si decide di individuare e salvare, per ognuno dei paesi considerati, le 300 parole singole più utilizzate e i 50 bigrammi e trigrammi maggiormente frequenti.

Tra le numerose informazioni ottenute mediante queste tecniche si è deciso di conservare la polarità media per ogni stato con la relativa deviazione standard come indicatore principale. I dataframe costruiti con le parole più frequenti vengono conservati per una successiva rappresentazione grafica.

## Varietà

Per completare il dato ottenuto dal social network sono state recuperate informazioni circa la natura socioeconomica del paese dal sito della *WorldBank*<sup>1</sup>. Tra i numerosi indicatori forniti dalla fonte ne sono stati selezionati una decina in grado di definire un quadro generale ma accurato della situazione economica e sociale dello stato. I valori recuperati risalgono all'ultima rilevazione effettuata negli ultimi dieci anni, perciò è possibile che alcune informazioni siano mancanti. L'output ottenuto a seguito del processo è un dataset che associa ad ogni stato un set di indicatori con il valore relativo all'ultima rilevazione. Questo verrà successivamente integrato all'output della sentiment analysis.

I dati relativi alla situazione climatica ed ambientale degli stati sono stati recuperati dal sito ufficiale

---

<sup>1</sup><https://www.worldbank.org>

dell'Environmental Performance Index <sup>2</sup> creato e gestito dall'Università di Yale che riporta, oltre all'indice principale EPI, numerosi indicatori che descrivono la situazione ambientale in ogni stato del mondo.

In particolare l'indice principale è formato dalla media ponderata di due indicatori Ecosystem Vitality ed Environmental Health rispettivamente ECO ed HLT. Analogamente, questi due sono il risultato della media ponderata di altri indici che vengono riportati nel terzo anello dell'immagine 1.

L'ultimo settore non è stato preso in considerazione. Ad ogni stato viene accostato ogni indicatore



Figura 1: Costruzione indice EPI

con il valore più recente (2020) e la sua variazione assoluta dal 2010.

Infine sono state recuperate informazioni circa il numero di utenti social nello specifico paese e la percentuale di utenti Twitter per ogni stato, utilizzando i dati riportati in due diversi portali <sup>3</sup>.

Ottenuti tali dati è stato possibile creare due nuovi indicatori, uno ottenuto rapportando il numero di tweet pubblicati per stato al numero di utenti twitter nello stesso, l'altro ottenuto nel medesimo modo ma considerando il numero di user che ha effettuato una pubblicazione nel periodo di tempo scelto. Entrambi i valori sono per milione di abitanti.

I tre dataset puliti sono stati quindi uniti al risultato della sentiment e all'elenco dei tweet per poi essere caricati su **MongoDB**.

## Esplorazione dei risultati

Ottenuto il dato finale, è stata svolta un'analisi esplorativa volta a comprendere relazioni o pattern interni. Prima è stato analizzato l'output della sentiment analysis, il quale ci ha fornito, come indicato nei paragrafi precedenti, un valore di polarità per ogni paese. Tale valore, insieme alla deviazione standard, è stato necessario per confrontare il sentimento medio tra i singoli stati. Osservando la Figura 2, la costruzione dell'intervallo di confidenza relativo alla media al 95% del valore della polarità mostra come effettivamente non ci sia una differenza statistica significativa tra i sentiment che risultano neutrali poiché comprendenti lo zero. È stato necessario dunque l'utilizzo di un nuovo valore per cogliere una possibile differenza circa la preoccupazione nei confronti della situazione climatica. Tale indice è il numero di utenti Twitter che hanno pubblicato nel periodo della streaming rapportato al

<sup>2</sup><https://www.epi.yale.edu>

<sup>3</sup>[datareportal.com](http://datareportal.com) e [gs.statcounter.com](http://gs.statcounter.com)

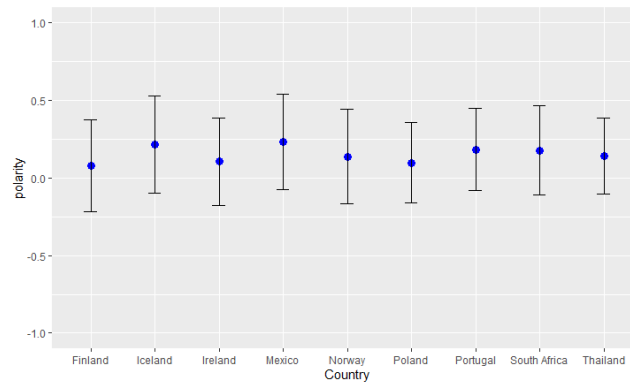


Figura 2: IC polarity

numero di utenti Twitter nel singolo stato. Questo ci ha permesso di ottenere un valore in grado di riassumere approssimativamente quanto gli abitanti del paese ci tenessero a far sentire la propria voce in merito.

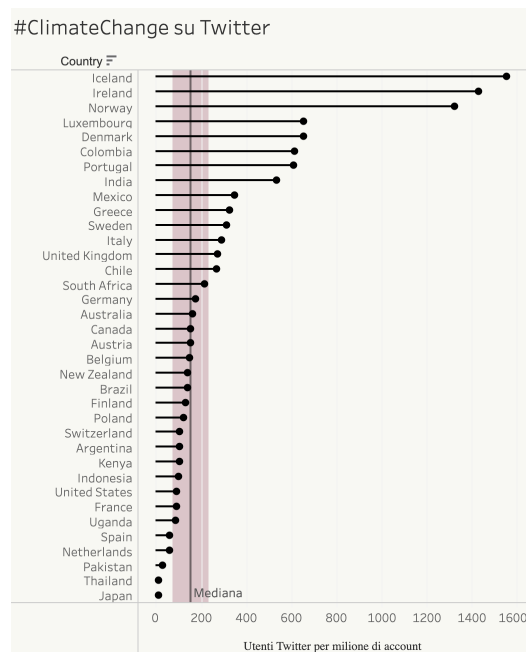


Figura 3: Climate change su twitter

L'Islanda, la Norvegia e l'Irlanda, come mostra la Figura 3, ottengono un valore nettamente sopra la media confermando come la popolazione di questi paesi abbia a cuore la drammatica situazione in cui ci troviamo. Inoltre, in questi tre stati, come nella maggior parte di quelli selezionati, osservando la Figura 4, è possibile notare come la curva abbia una maggiore altezza nei valori positivi, quindi maggiori di 0. La maggior presenza di tweet giudicati positivi rispetto ai negativi può essere sintomo di un approccio fiducioso piuttosto che aggressivo ai metodi utilizzati dal paese per risolvere la questione.

Successivamente sono stati studiati gli indici socio economici ed ambientali, prima singolarmente poi in relazione, mediante l'uso di visualizzazioni reperibili sulla piattaforma **Tableau Public**<sup>4</sup>. Gli stati europei, nord americani, l'Oceania e Giappone possiedono un valore di EPI superiore alla media mentre i restanti stati asiatici e gli stati africani hanno valori molto bassi causati principalmente dalla

<sup>4</sup><https://public.tableau.com/app/profile/alberto.carlone/viz/Earthday-ProgettoDMDVCarloneMioriOndei/Storia1>

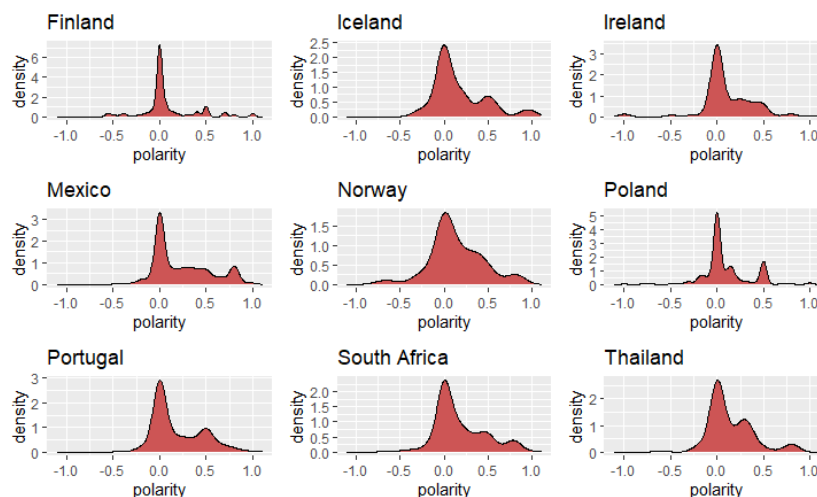


Figura 4: Densità

pessima qualità dell'aria per i primi e dalla scarsa qualità dell'acqua per i secondi.

Una suddivisione simile è si evince anche dallo studio del GDP per capita. Gli stati europei, con il Lussemburgo come leader, insieme al Nord America e Oceania possiedono un GDP per capita superiore alla media. Utilizzando la mediana invece, solo cinque stati europei e gli USA superano la soglia di significatività. Gli stati sudamericani, il Messico, quelli asiatici (sempre senza il Giappone) e quelli africani cadono sotto sia l'intervallo della media sia della mediana.

L'indice di Gini, indice che misura la distribuzione del reddito e che in un mondo spezzato in due assume sempre più importanza, mostra come i paesi più poveri abbiano un valore molto elevato, sintomo di ingegualità. Specialmente il Sud Africa racconta storie di povertà e miseria fuori dai centri cittadini moderni e tecnologici sempre più simili a città europee (basti pensare a Cape Town).

Evidenziata una correlazione positiva tra GDP ed EPI, ci si è chiesti se effettivamente respirare aria più pulita potesse portare ad una vita più longeva. Non avendo gli strumenti per studiarne la causalità, ci si è limitati ad osservarne la correlazione che risulta essere positiva (pari allo 0.89). A dominare la scena sono i paesi che hanno aderito ad Annex II mentre i paesi del BRICS e quelli senza una vera e propria appartenenza mostrano valori sotto la mediana per entrambi gli indici. L'unico stato con valore superiore all'intervallo della mediana sia dell'aspettativa di vita sia della qualità dell'aria è l'Australia la quale, a partire dagli anni 70 del XX secolo, ha inaugurato una vasta gamma di cambiamenti sociali e ambientali, come migliorie della condizioni di lavoro e di alloggio, e della salute pubblica facendo affidamento quando possibile a campagne vaccinali.

Infine, un tema molto delicato e centrale nel protocollo di Kyoto sono le emissioni inquinanti. Si è studiato dunque la variazione assoluta dal 2010 al 2020 al fine di capire quali stati si siano maggiormente impegnati a combattere questa problematica. Prima su tutti la Nuova Zelanda, che con un aumento di quasi 40 punti, possiede un valore dell'indice APE (Pollution Emissions) pari ad 87. Tale aumento è sicuramente dovuto all'adozione dell'Emission Trading Scheme nel 2012 che ha permesso di ridurre notevolmente le emissioni inquinanti sull'isola, grazie al tetto massimo imposto su di esse agli impianti partecipanti. Contrariamente, stati come il Cile e la Colombia, a causa dell'utilizzo di legna da ardere per il riscaldamento nonché dei numerosi trasporti via terra, hanno ridotto notevolmente l'APE dal 2010 rispettivamente di 54 e 44 punti aumentando di fatto le emissioni inquinanti.

## Data Visualization - Valutazioni

Si analizzano ora i risultati ottenuti a seguito delle valutazioni effettuate in merito alla dashboard pubblicata su Tableau Public <sup>5</sup>. Nello specifico si effettua una valutazione euristica, un test utente e si diffonde un questionario psicometrico in riferimento alla seconda pagina della storia riportata al link di cui sopra.

### Valutazione Euristica

La valutazione euristica ha coinvolto 8 soggetti, di età compresa tra i 20 e i 60 anni, 4 di genere femminile e 4 di genere maschile. Agli utenti è stato chiesto di interagire con la dashboard commentando liberamente ciò che stavano osservando. Da questa prima verifica se ne evince

- Mancanza di familiarità con concetti quali: intervallo di confidenza, mediana;
- Difficoltà di interpretazione del significato dei termini in lingua inglese;
- Necessità di una maggior spiegazione teorica degli indicatori;
- Valori della sentiment non compresi (da alcuni non considerati, erano inseriti nella tendina a comparsa a seguito della selezione del paese ma il caricamento risultava troppo lento ed alcuni chiudevano la tendina prima che venisse visualizzata la sentiment) <sup>6</sup>;
- Mancata comprensione della tipologia di energia "Altro";
- Mancato corretto utilizzo del glossario (non capito che nei quadratini era presente la descrizione degli indici). Conseguente mancata comprensione degli indici;
- Segnalata eccessiva presenza del colore blu;
- Tabella con indici socio - economici valutata poco bella / elegante. Suggerimento: inserire tali valori nella tendina che si apre selezionando il paese sulla mappa.

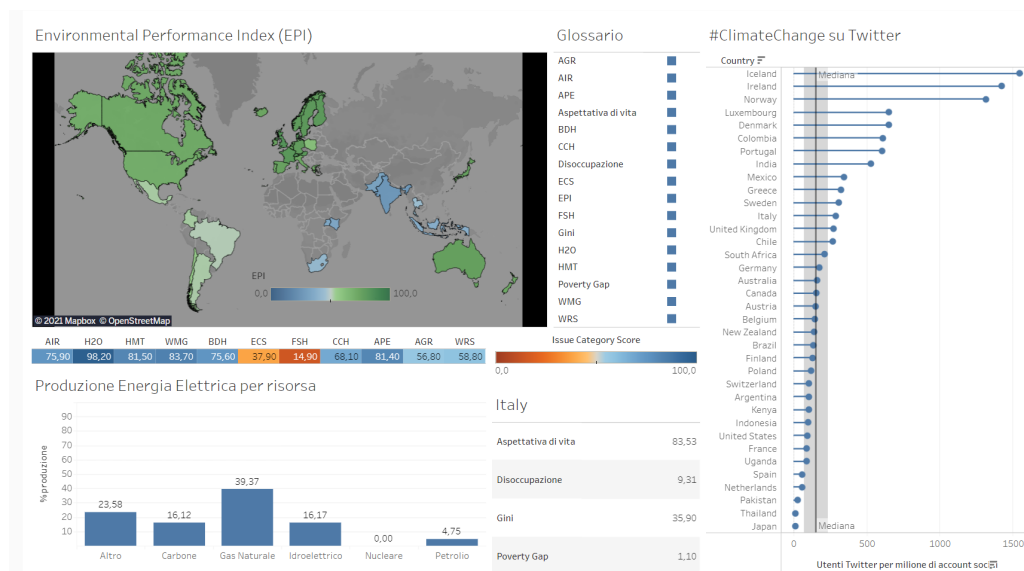


Figura 5: Dashboard originale

<sup>5</sup><https://public.tableau.com/app/profile/alberto.carlone/viz/Earthday-ProgettoDMDVCarloneMioriOndei/Storia1>

<sup>6</sup>Circa  $\frac{1}{3}$  degli intervistati

## User Test

Dopo aver modificato la dashboard in base alle problematiche riscontrate nella valutazione euristica, sono stati effettuati dei test per verificare che gli utenti comprendessero realmente le informazioni contenute nella visualizzazione. Sono stati quindi somministrati 4 task, di difficoltà crescente, ai 12 utenti coinvolti. Nello specifico le domande poste sono:

1. Relativamente al numero di utenti Twitter per milione di account , quanti ne possiede la nazione che si trova al terzo posto?
2. Qual è il valore dell'indice Climate Change (CCH) relativo alla Francia?
3. Considerando il Sud America, qual è la parola più frequente associata allo stato con maggior indice di disoccupazione?
4. Quali sono i valori medi di energia green relativi al continente nord americano? (Hint: non è necessario selezionare gli stati singolarmente).

Si nota, sia dalla distribuzione dei tempi sia dalla distribuzione delle risposte che i task più complessi sono risultati il numero 3 e il numero 4, come secondo aspettative. In particolare i risultati del quarto task mostrano una differenza statisticamente significativa nei confronti dei primi due.

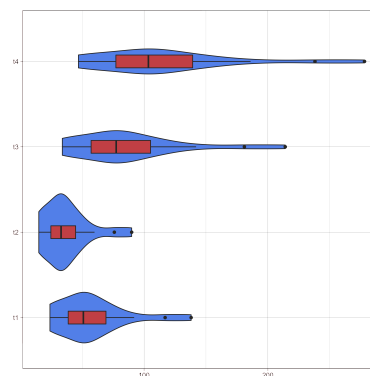


Figura 6: Distribuzione dei tempi, differenziati per task

Per il task numero 4 si specifica inoltre che, viste le risposte date e il conseguente intervallo di confidenza, non è possibile considerare tale compito come superato con un livello di confidenza del 95%.

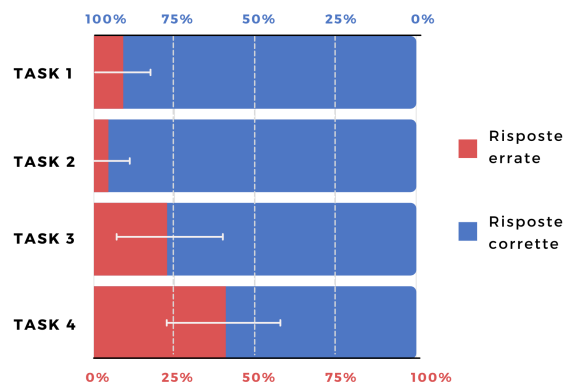


Figura 7: Confronto risposte corrette - errate, con relativo intervallo di confidenza



Alcuni utenti hanno segnalato una mancanza di dimestichezza con Tableau, anche da ciò sembra derivare la difficoltà nel portare a termine alcuni task e nel comprendere a pieno il potenziale della visualizzazione proposta. Nonostante il suggerimento specificato nella domanda, alcuni utenti non sono riusciti a selezionare più paesi contemporaneamente e quindi rispondere correttamente alla domanda.

## Questionario Psicometrico

Al fine di realizzare un questionario psicometrico si decide di far riferimento alla scala Cabitza - Locoro <sup>7</sup>. Viene quindi chiesto agli intervistati di valutare la qualità del prodotto assegnando un valore compreso tra 1 e 6 a ciascuno dei seguenti campi: utilità, chiarezza, informatività, bellezza, intuitività per poi assegnare un punteggio anche al valore complessivo.

I risultati ottenuti vengono analizzati tramite l'utilizzo di un correlogramma e uno *Stacked Bar Chart*.

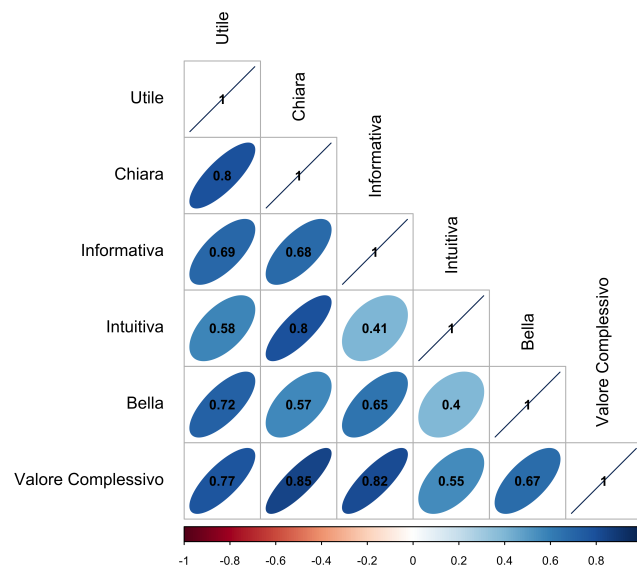


Figura 8: Correlogramma

Il correlogramma indaga la correlazione tra i diversi item del questionario psicometrico. Come osservabile anche dal grafico, non vi è la presenza di correlazioni negative. In generale tutti gli item risultano molto correlati, soltanto Intuitività - Bellezza e Intuitività - Informatività hanno una correlazione inferiore a 0,55.

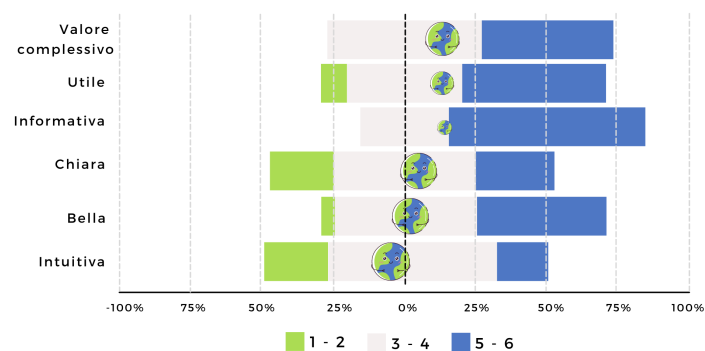


Figura 9: Stacked Bar Chart risposte

<sup>7</sup>Actis-Grosso R. & Batini C. Locoro A. Cabitza F. "Static and interactive infographics in daily tasks: A value-in-use and quality of interaction user study." In: Computers in Human Behavior (2017).

Nella Figura 9 si possono vedere le votazioni attribuite agli item del questionario psicometrico. Si nota una generale preponderanza di risposte positive, per alcuni però l'infografica risulta essere poco intuitiva. Attraverso l'utilizzo di un Violin Plot si può notare con maggior precisione la distribuzione delle risposte date.

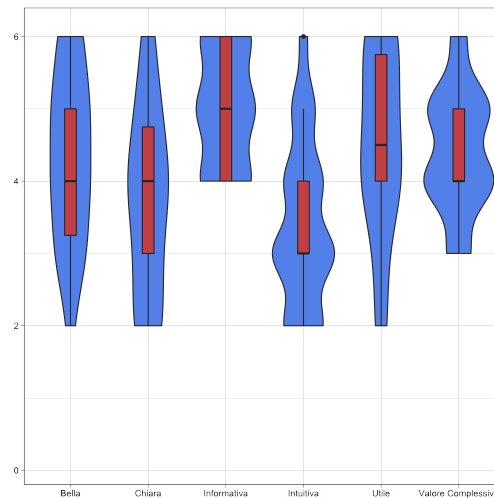


Figura 10: Violin Plot - Questionario Psicometrico

Il questionario dava infine la possibilità di lasciare un commento generale libero, su base volontaria. La maggior parte dei commenti evidenziano una mancanza di comprensione di parte della dashboard a causa della carenza di nozioni di carattere tecnico. In particolare il grafico raffigurante la sentiment analysis è stato il meno compreso ha creato confusione. La difficoltà nella lettura di quest'ultimo era prevedibile in quanto si tratta di un concetto poco diffuso e poco noto al di fuori della comunità statistica. Alcuni hanno sottolineato come il nucleare, secondo il loro parere, dovesse rientrare tra le energie green. Benché rappresenti in gran parte una forma di energia pulita dal punto di vista delle emissioni, non rientra per definizione nelle fonti di energia rinnovabile in quanto non vengono reintegrate naturalmente in una scala temporale umana.

La Figura 11 rappresenta infine la dashboard definitiva, modificata sulla base di tutti i suggerimenti ricevuti.

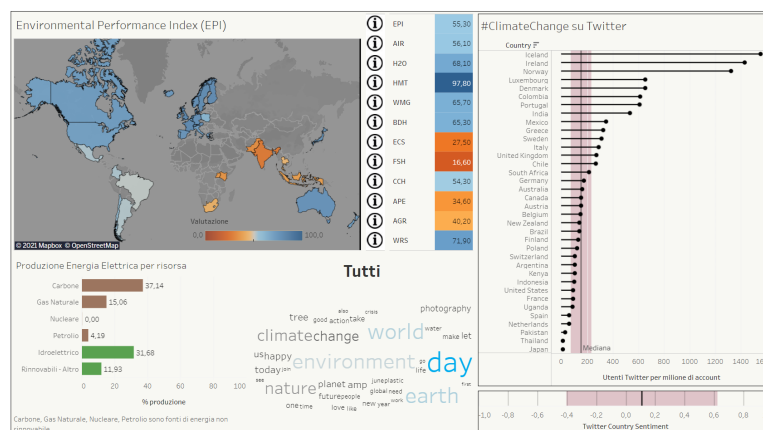


Figura 11: Dashboard dopo le modifiche effettuate a seguito delle valutazioni ricevute

## Conclusione

Lo studio svolto ci ha permesso sia di cogliere relazioni tra gli indicatori socio economici e quelli ambientali sia di comprendere quale fosse l'approccio delle varie popolazioni degli stati selezionati nei confronti del cambiamento climatico. In particolare, facendo riferimento alla seconda questione, nessuno degli stati ha mostrato un indice di polarità statisticamente differente dagli altri. Contrariamente, studiando il numero di utenti Twitter attivi sul cambiamento climatico si è potuto evincere che, principalmente, i paesi del Nord Europa abbiano a cuore la questione. Inoltre, la maggior parte degli stati presenta curve di densità con picchi per valori maggiori di zero mentre per valori negativi la curva resta piatta dimostrando come apparentemente l'atteggiamento degli utenti di quel paese sia fiducioso.

Da un punto di vista sociale ed economico, gli stati con un maggiore reddito pro capite come l'Europa, il Nord America, l'Oceania e il Giappone si confermano in corsa per un futuro "*green*" infatti i valore dell'EPI sono molto alti. Tuttavia solo alcuni stati europei (principalmente del nord) producono ed utilizzano per la maggior parte, se non interamente, energia rinnovabile. Gli stati poveri, invece, o a causa di governi instabili o per i metodi arretrati utilizzati in vari settori, presenta indici decisamente non in linea con gli obiettivi posti a Kyoto.

Il tema in questione è molto delicato, ogni giorno che passa la situazione si aggrava poiché la serie di feedback negativi causati dal riscaldamento globale sono sempre più frequenti e dannosi. Risulta, dunque, molto difficile per molti stati preoccuparsi di tale problema viste le difficoltà giornaliere che affrontano. D'altra parte, quei paesi considerati avanzati, i quali possiedono valori dell'environmental performance index indubbiamente buoni, sono molto spesso legati all'energia non rinnovabile la cui produzione o consumo danneggia il nostro ecosistema e la nostra atmosfera costantemente.

# Appendice

## Indicatori

Nota: per ogni singolo indicatore relativo all' *Enviromental Performance Index* è presente una variabile *new* (esempio EPI.new) che rappresenta il valore al 2020, ed una variabile *change* (esempio EPI.change) che rappresenta la variazione del medesimo indicatore rispetto al 2010.

- AIR.new: Air Quality, misura gli impatti diretti della qualità dell'aria sulla salute dell'uomo per ogni nazione.
- AGR.new: Agriculture, misura gli sforzi nel supportare una popolazione in salute e al contempo di minimizzare le minacce dell'agricoltura all'ambiente.
- APE.new: Pollution Emission, misura i progressi nella gestione delle emissioni dei due principali agenti inquinanti dell'aria (Biossido di Zolfo e Ossido di Azoto).
- BDH.new: Biodiversity and habitat, valuta il comportamento delle nazioni rispetto al mantenimento dell'ecosistema naturale e alla protezione delle biodiversità all'interno dei propri confini.
- CCH.new: Climate change, misura il progresso nella lotta al cambiamento climatico globale, che aggrava tutti gli altre minacce ambientali e mette in pericolo la salute e la sicurezza dell'uomo.
- ECO.new: Ecosystem Vitality, misura quanto bene i paesi preservano, proteggono e migliorano gli ecosistemi e i servizi che forniscono. Rappresenta il 60% dell'indicatore EPI ed è composto da BDH, ECS, FSH, CCH, APE, AGR, WRS.
- ECS.new: Ecosystem Services, riconosce gli importanti servizi che gli ecosistemi forniscono al benessere umano e ambientale, tra cui il sequestro e lo stoccaggio del carbonio, la biodiversità, il ciclo dei nutrienti e la protezione delle coste.
- EDU\_bachelor: Educational attainment, at least Bachelor's or equivalent, population 25+, total (%) (cumulative), la percentuale di popolazione di età pari o superiore a 25 anni che ha conseguito o completato un diploma di laurea o equivalente.
- EDU\_secondary: Educational attainment, at least completed post-secondary, population 25+, total (%) (cumulative), la percentuale di popolazione di età pari o superiore a 25 anni che ha conseguito o completato l'istruzione post-secondaria non terziaria.
- ELE\_coal: Electricity production from coal sources (% of total), l'energia elettrica generata dal consumo di fonti a base di carbone.
- ELE\_gas: Electricity production from natural gas sources (% of total), l'energia elettrica generata da fonti a base di gas naturale.
- ELE\_hydr: Electricity production from hydroelectric sources (% of total), l'energia elettrica generata da centrali idroelettriche.
- ELE\_nuc: Electricity production from nuclear sources (% of total), l'energia elettrica generata da centrali nucleari.
- ELE\_oil: Electricity production from oil sources (% of total), l'energia elettrica generata da fonti a base petrolifera.
- ELE\_renew: Renewable electricity output (% of total electricity output), è la quota di elettricità generata da impianti di energia rinnovabile rispetto al totale dell'elettricità generata da tutti i tipi di impianti.

- EMP\_prim: Employment in agriculture (% of total employment) (modeled ILO estimate), l'occupazione all'interno del settore agricolo. È costituito da agricoltura, caccia, silvicoltura e pesca.
- EMP\_seco: Employment in industry (% of total employment) (modeled ILO estimate), l'occupazione all'interno del settore industriale. È costituito da miniere e cave, produzione, edilizia e servizi pubblici (elettricità, gas e acqua).
- EMP\_terz: Employment in services (% of total employment) (modeled ILO estimate), l'occupazione all'interno dei servizi. È costituito da commercio all'ingrosso e al dettaglio, ristoranti e alberghi; trasporto, stoccaggio e comunicazioni; servizi finanziari, assicurativi, immobiliari e aziendali; servizi comunitari, sociali e personali.
- EPI.new: Environmental Performance Index, L'EPI è l'indicatore globale che fornisce una base quantitativa per comparare, analizzare e capire le performance ambientali di 180 nazioni. Il suo punteggio è composto da una somma pesata delle categorie HLT ed ECO.
- FSH.new: Fisheries, misura la salute e la sostenibilità mondiale della pesca.
- GDP: il prodotto interno lordo pro capite.
- GDP\_growth: GDP per capita growth (annual %), l'aumento annuale pro capite del prodotto interno lordo.
- Gini: Gini index (World Bank estimate), misura quanto la distribuzione della ricchezza nella popolazione sia più o meno vicina ad una distribuzione perfettamente equa. Un indice di Gini pari a zero rappresenta una perfetta equità mentre un valore di 100 rappresenta una totale iniquità.
- H2O.new: Sanitation and drinking water, misura quanto bene le nazioni proteggono la salute umana dai rischi ambientali.
- HLT.new: Environmental Health, misura quanto bene i paesi stanno proteggendo le loro popolazioni dai rischi per la salute ambientale. Rappresenta il 40% dell'indicatore EPI ed è composto da AIR, H2O, HMT e WMG.
- HMT.new: Heavy Metals, misura quanto bene vengono limitati gli impatti diretti dovuti all'inquinamento da esposizione dei metalli pesanti sulla salute umana in ciascuna nazione.
- Iso: il codice ISO Alpha-3 della nazione considerata
- Labor\_force: Labor force participation rate, total (% of total population ages 15-64) (modeled ILO estimate), il tasso di partecipazione alla forza lavoro è la proporzione della popolazione di età 15-64 che è economicamente attiva: tutte le persone che forniscono lavoro per la produzione di beni e servizi durante un determinato periodo.
- Life\_exp: Life expectancy at birth, total (years), indica il numero di anni che un neonato vivrebbe se tutti i fattori di mortalità al momento della sua nascita rimanessero i medesimi per tutta la sua vita.
- NUC\_energy: Alternative and nuclear energy (% of total energy use), è l'energia senza idrocarburi che non produce anidride carbonica quando viene generata. Include energia idroelettrica e nucleare, geotermica e solare, tra le altre.
- Pop: Population (total), si basa sulla definizione de facto di popolazione, che conta tutti i residenti indipendentemente dallo status giuridico o dalla cittadinanza.

- Poverty\_gap: Poverty gap at \$1.90 a day (2011 PPP) (%), rappresenta quanto, in media, la popolazione di una nazione è distante dalla soglia di povertà valutata ad \$1.90 al giorno. Essa quindi rappresenta sia quanto è radicata la povertà sia la sua incidenza.
- Region: l'area geografica d'appartenenza
- Unemployment: Unemployment, total (% of total labor force), fa riferimento alla porzione di forza lavoro che è senza lavoro disponibile pur essendone alla ricerca.
- WMG.new: Waste Management, misura gli sforzi nella gestione dei rifiuti per mitigare la minaccia degli stessi sulla salute umana.
- WRS.new: Water resources, misura quanto l'uomo stia cercando di mitigare le minacce agli ecosistemi acquatici.