BW3 PERSONALE - LIOSSIS SOFIA

Per la presentazione finale con il gruppo abbiamo analizzato il dataset, <https://github.com/owid/co2-data/>, riguardante le emissioni di Co2.

Inizialmente abbiamo pulito e analizzato entrambi i dataset Co2 e Energy ma alla fine la scelta è ricaduta nell’utilizzo soltanto del primo, poiché le tematiche trattate le abbiamo ritenute più interessanti.

Per effettuare un’analisi più approfondita abbiamo ritenuto necessario cercare altre fonti.

Io sono riuscita a trovare diversi dataset riguardanti tematiche affini come:

* Annual Surface Temperature Change (https://climatedata.imf.org/datasets/4063314923d74187be9596f10d034914/explore)
* Change in Mean Sea Levels (https://climatedata.imf.org/datasets/b84a7e25159b4c65ba62d3f82c605855/explore)
* Climate-related Disasters Frequency(https://climatedata.imf.org/datasets/b13b69ee0dde43a99c811f592af4e821)

Ho scelto questi 3 dataset poichè li ho considerati validi e affidabili essendo stati prodotti dal Dipartimento di Statistica del “International Monetary Fund (IMF)” in collaborazione con diversi enti internazionali come: the European Statistical Office (Eurostat), the Food and Agriculture Organization (FAO), the National Oceanic and Atmospheric Administration (NOAA) and National Aeronautics and Space Administration Goddard Institute for Space Studies (NASA-GISS).

### **Dataset Co2**

Ho utilizzato lo strumento **Power Query** in PowerBI per:

* rimozione colonne vuote e duplicati
* sostituzione dei valori nulli
* assegnazione del corretto formato alle colonne
* creazione di una colonna personalizzata per estrarre l’anno nel formato corretto

### **Dataset Annual Surface Temperature Change**

Per questo dataset ho effettuato le seguenti operazioni:

* trasformazione con Python
* assegnazione del corretto formato alle colonne
* sostituzione valori nulli

All’interno sono presenti colonne per ogni anno che riportano il valore misurato.

Per poter effettuare un’analisi temporale ho aggiunto con un **notebook Python** una colonna univoca per la data e una per il valore (vedi allegato *Transform\_temperatura.ipynb*).

Poiché la colonna “Year” non veniva riconosciuta come “Date” su **Power Query** ho aggiunto una nuova colonna : (#date( [Year], 1,1) e poi ho cambiato il formato colonna in “Date” e cancellato la vecchia colonna.

### **Dataset Change in Mean Sea Levels**

Per questo dataset ho effettuato le seguenti operazioni:

* rimozione colonne vuote e duplicati
* sostituzione dei valori nulli
* assegnazione del corretto formato alle colonne
* raggruppamento per le colonne “Anno” e “Mare” facendo la Media di “Value”

Poiché la colonna “Year” non veniva riconosciuta per poter fare un’analisi temporale, ho trasformato la colonna “Date” nel formato “Data inglese - Stati Uniti”. Successivamente ho duplicato la colonna “Date” ed estratto le ultime 4 cifre andando a selezionare solo l’anno. Per poter fare delle analisi su base annua, ho proceduto a raggruppare le colonne in base a “Anno” e “Mare”, facendo la media di “Value”. In questo modo ho ottenuto una tabella con solo le 3 colonne di nostro interesse (Anno,Mare e Variazione).

### **Dataset Climate-related Disasters Frequency**

Per questo dataset ho effettuato le seguenti operazioni:

* trasformazione e manipolazione con Python
* assegnazione del corretto formato alle colonne
* sostituzione valori nulli

Come il dataset riguardante l’aumento di temperatura, anche all’interno di questo sono presenti colonne per ogni anno che riportano il numero di disastri ambientali.

Per poter effettuare un’analisi temporale, ho aggiunto con un **notebook Python** una colonna univoca per la data e una per il valore, andando anche a sostituire alcuni valori al suo interno (vedi allegato *Transform\_disastri.ipynb*).

Infine ho replicato la stessa trasformazione, fatta precedentemente per le temperature, per la colonna “Year”.

### **PowerBI**

Dopo aver importato i vari dataset ed effettuato le operazioni di pulizia, ho definito le relazioni tra i vari modelli per il valore “Year” così da poter inserire un filtro temporale condiviso.

Infine ho contribuito alla scelta dello stile, delle immagini e dei grafici per la dashboard.