# TRACCIA, OBIETTIVI E STRUMENTI UTILIZZATI:

* Presentazione: Inquinamento atmosferico per emissioni di CO2 con le maggiori conseguenze portate da esso, cioè, tasso di mortalità per inquinamento dell'aria, innalzamento delle temperature, scioglimento dei ghiacciai con innalzamento dei livelli del mare e produzione ed utilizzo delle energie rinnovabili;
* Obiettivi: evidenziare le possibili correlazioni tra causa ed effetto (emissioni CO2 e le conseguenze sopra elencate), constatare l'urgenza di un maggior utilizzo delle rinnovabili;
* Strumenti utilizzati: Power BI e Python.

# FASE 1 ETL:

1. Ho caricato i file CSV ed Excel su Power BI e ho iniziato ad esplorarli;
2. Confrontando i file sulle Emissioni di CO2, ho rinominato la prima colonna dei file da “Entity” in “Country” per avere allinearli;
3. Nei file ho notato nelle colonne dei valori “(spazio vuoto)” e dei valori null. Entrambi i tipi di valori dovrebbero rappresentare mancanza di dati, per quei paesi in quelle annate, quindi per una validazione e coerenza, gli spazi vuoti li ho sostituiti con i null;
4. Lo stesso procedimento l’ho ripetuto per i file delle Energie rinnovabili, non essendoci ulteriori particolarità;
5. Nel file “Tasso di mortalità per inquinamento atmosferico” invece ho rinominato la tabella e essendo un file Excel, e ho alzato le righe rendendo la prima come intestazione;
6. Per i file sui ghiacciai, erano presenti disuguaglianze tra il file del ghiacciaio Artico e quello sulla massa delle calotte polari: sostanzialmente venivano chiamati diversamente, cosi ho deciso di unire i 2 nomi in un'unica soluzione per entrambi i file e quindi maggiore chiarezza, cioè “Arctic Ocean/Greenland” e nel codice di quest’ultimo, dove mancava ho inserito lo stesso presente altrove e cioè GRL;
7. Per i file delle temperature: in “Aumento temperature Globali (per mese)” la colonna “Entity” l’ho rinominata “Month” siccome contiene unicamente i 12 mesi e la colonna “Code” dello stesso file l’ho eliminata in quanto li è decontestualizzata.  
   invece nel file “Temperature medie globali”, nella colonna “Code”, essendo già presente il codice per l’entità “world”, ho deciso simbolicamente di dare dei codici anche al “Northern Hemisphere” ed al “Southern Hemisphere” rispettivamente “NH” e “SH”;
8. Soltanto 1 colonna in tutti i datasets aveva date per intero in formato testo che ho corretto.
9. Mi sono reso conto che Power Query, caricando i dati, aveva automaticamente modificato il formato delle colonne dei dati da decimale a intero, cosi ho cancellato il passaggio del “modificato tipo”, ho trovato tutte le colonne con decimali, ho sostituito il “.” con la “,” e poi ho cambiato il formato da “Testo” in “Decimale” in modo da non invalidare i dati;
10. Dalla tabella delle emissioni di Co2 pro capite ho eliminato tutte le colonne che non mi sarebbero servite alle analisi, confrontandole con le colonne presenti nella tabella delle emissioni di Co2 per fonte.

# FASE 2 RELAZIONI:

1. Mi sono reso conto che per creare delle relazioni tra le tabelle avessi bisogno di tabelle delle “Dimensioni”;
2. Tramite l’utilizzo di python e pandas, su jupyterlab ho caricato prima il file delle Emissioni pro capite che era quello con la colonna dei paesi più popolata, cosi mi sono estratto unicamente quella colonna eliminando tutti i duplicati ed il data frame me lo sono salvato in formato csv ed importato su power BI, ottenendo così la tabella delle dimensioni dei paesi: ”Country list”, una semplice lista dei paesi per relazionare tutte le tabelle che li possedessero;
3. Per i file dei ghiacciai invece ho creato un unico file della loro estensione cosi da ridurre la quantità di tabelle: sempre su python ho importato i 2 csv e ho rinominato le colonne delle “massima” e “minima” estensione dei ghiacciai, per avere tutte colonne uguali e le ho riordinate per metterle nella stessa posizione, dopodiché le ho unite con il concat di python e caricato il file su power bi eliminando le 2 tabelle precedenti;
4. Per relazionare i file dei ghiacciai ho creato una tabella delle dimensioni contenente semplicemente 2 righe e 2 colonne con entrambi i ghiacciai ed i loro codici allo stesso modo per la lista paesi;
5. Le relazioni, una volta create le tabelle delle dimensioni, si sono create automaticamente, ma controllandole, sono tutte giuste.