



Esercizio: progetto di analisi dati

Analisi Diffusione COVID-19

Il committente richiede di avere un report su casi e vaccinazioni in diverse aree del mondo; a tal fine, richiede di utilizzare il dataset, curato da Our World in Data, all'indirizzo <https://github.com/owid/covid-19-data/tree/master/public/data> alla voce "📦 Download our complete COVID-19 dataset" scaricare il dataset nel formato che si preferisce.

1. Si richiede di verificare le dimensioni del dataset e i relativi metadati
2. Si chiede di trovare, per ogni continente:
 - a. il numero di casi fin dall'inizio della pandemia
 - b. la percentuale rispetto al totale mondiale del numero di casi
3. Selezionare i dati relativi all'Italia nel 2022 e, poiché i nuovi casi vengono registrati settimanalmente, filtrare via i giorni che non hanno misurazioni; quindi mostrare con dei grafici adeguati:
 - a. l'evoluzione dei casi totali dall'inizio alla fine dell'anno
 - b. il numero di nuovi casi rispetto alla data
4. Riguardo le nazioni di Italia, Germania e Francia:
 - a. mostrare in un boxplot la differenza tra queste nazioni riguardo il numero di pazienti in terapia intensiva (Intensive Care Unit, ICU), considerare quindi la colonna `icu_patients` da maggio 2022 (incluso) ad aprile 2023 (incluso)
 - b. scrivere un breve commento (una o due righe) riguardo che conclusioni possiamo trarre osservando il grafico risultante
5. Riguardo le nazioni di Italia, Germania, Francia e Spagna in tutto il 2021:
 - a. mostrare, in maniera grafica oppure numerica, la somma dei pazienti ospitalizzati per ognuna (colonna `hosp_patients`)
 - b. se ci sono dati nulli, con un breve commento scrivere se può essere possibile gestirli tramite sostituzione o meno

Note

Prima di effettuare le analisi, effettuate un'EDA del dataset per meglio comprenderne la natura (e non dimenticate ulteriori metadati e descrizioni presenti sulla pagina Github da cui avete scaricato il dataset); in particolare:

- Fate attenzione alle colonne `continent` e `location`; contengono sia dati puntuali che dati aggregati
 - *Prima di iniziare l'analisi, assicuratevi di aver compreso cosa contengono*
- Comprendete bene il contenuto delle colonne `new_cases` e `total_cases`
 - *Hanno logiche differenti e ovviamente vanno trattate diversamente*

Consegna

- Per consegnare, effettuate direttamente l'upload del file `.ipynb` su Github, senza passare per file `.zip` o similari
 - *Caricati direttamente in questo modo, i Notebook possono essere direttamente consultati dal browser; possono quindi essere utilizzati, ad esempio, per costruire un portfolio*



GRAZIE
EPCODE