**Esercitazione Dataset Covid19**

* **Fonte:** Presidenza del Consiglio dei ministri - Dipartimento Protezione Civile

<https://github.com/pcm-dpc/COVID-19/blob/master/dati-json/dpc-covid19-ita-andamento-nazionale.json>

* **Descrizione:** File JSON, quotidianamente aggiornato dall’inizio della pandemia, riporta tutti i dati rilevanti per il monitoraggio del virus (licenza CC-BY-4.0)
* **Campi:** Riferimento Giorno, Ospedalizzazioni (Sintomatiche, Isolamento, Terapia Intensiva), Positivi totali, Nuovi Positivi, Dimessi, Deceduti, Info tamponi\*
* **Classificazione: 🟊 🟊 🟊 🟊 🟊**
* **Fonte:** Presidenza del Consiglio dei ministri - Dipartimento Protezione Civile

<https://raw.githubusercontent.com/JacopoTrono/EPICODE/main/Esempio%20Elab%20orazione%201%20stella.jpg>

* **Descrizione:** Immagine contenente grafici cartesiani
* **Campi: ᴓ**
* **Classificazione: 🟊**
* **Fonte:** OurWorldInData

<https://github.com/owid/covid-19-data/tree/master/public/data>

* **Descrizione:** Dataset disponibile in formato JSON e CSV costituito dall’aggregazione dei dati giornalieri comunicati da ogni nazione del globo (licenza Creative Commons BY)
* **Campi:** Casi confermati, morti, ospedalizzazioni, test condotti, vaccinazioni\*
* **Classificazione: 🟊 🟊 🟊 🟊 🟊**
* **Fonte:** Worldometers + Johns Hopkins University

<https://github.com/imdevskp/covid_19_jhu_data_web_scrap_and_cleaning/blob/master/worldometer_data.csv>

* **Descrizione:** Database aggregato delle fonti citate
* **Campi:** Nazione,continente, popolazione, casi totali, nuovi casi, morti, guarigioni, casi critici, casi attivi, stime riportate su 1M di popolazione.
* **Classificazione: 🟊 🟊 🟊 🟊** (non viene citata la licenza)
* **Fonte:** Worldometers

<https://www.worldometers.info/coronavirus/>

* **Descrizione:** Dashboard di visualizzazione dati. Non è possibile scaricare i file dati che vengono riportati nelle tabelle.
* **Campi:** Nazione,casi totali, nuovi casi, decessi, guariti, casi attivi, casi gravi, popolazione, indici su 1M di popolazione
* **Classificazione: 🟊 🟊** (la tabella è copiabile, due stelle anche se non è un file?)

**Resoconto Covid19 2022 classe Epicode DA**

<https://github.com/JacopoTrono/EPICODE/blob/main/Esercizio_Covid_CSV_1.csv>

* Entità: studenti della classe Epicode DA
* Chiave primaria: ID (codice di riferimento univoco di ciascuno studente)
* Campi: informazioni cronologiche riferite all’invio dei dati, nome e cognome degli studenti, città di residenza e informazioni relative al Covid-19 articolate in 3 colonne:
* “Positività” fornisce l’informazione dell’avvenuto contagio nell’anno preso in esame.
* “Ricovero” dice se sia stato registrato un ricovero per ogni entità.
* “Giorni Degenza” definisce la lunghezza della malattia registrata per ogni positività.

**Analisi CSV di classe**

* 18 studenti su 30 sono risultati positivi al virus (60% della popolazione)
* Dividendo gli studenti in base al sesso è interessante notare come il 100% della popolazione femminile della classe sia risultata positiva nell’arco del 2022, a fronte di un 40% di positività della porzione maschile.
* Prendendo in esame i giorni di degenza di chi ha contratto il virus (della porzione femminile un solo caso riporta positività con 0 giorni di degenza), emerge una media di 14 giorni riscontrabile anche separando i campioni in base al sesso.
* Nessun caso di positività al virus ha portato a ricovero

**Analisi Comparata Classe / Dataset Nazionali**

* Nel dataset nazionale sono riportati 19067148 casi totali per l’anno 2022, che, su una popolazione di 59030133 individui (dato istat 1 gennaio 2022), rappresentano circa il 24%. La notevole differenza tra questa percentuale e quella registrata nella classe è imputabile alla mancanza di rappresentatività dell’ultima.

**Data Jobs e Case Study**

Il ***Data Engineer*** è responsabile della costruzione del sistema di raccolta dei dati, che potrebbe includere la raccolta dei dati epidemiologici sul COVID-19 da diverse fonti e l'archiviazione di questi dati in un database adatto.

ll ***Data Scientist*** utilizza i dati raccolti dal Data Engineer per eseguire analisi e modelli statistici che possono aiutare a comprendere la diffusione del virus e i fattori che contribuiscono alla sua propagazione. Potrebbe anche utilizzare tecniche di apprendimento automatico per prevedere l'evoluzione futura della pandemia.

Il ***Data Analyst*** sarebbe responsabile dell'interpretazione dei dati raccolti dal Data Engineer e analizzati dal Data Scientist. Potrebbe creare visualizzazioni dei dati che rendano più facile per i decisori comprendere la situazione e prendere decisioni informate sulle misure da adottare per controllare la diffusione del virus. Potrebbe inoltre eseguire analisi di tendenza e comparare i dati con le performance di altri paesi per identificare opportunità per migliorare la gestione della pandemia.

**Metadati Esercitazione**

* Rispetto ai dataset dati da cercare, i nomi delle fonti sono metadati.
* Nel CSV di classe i campi *Positività, Ricoveri, Giorni degenza* sono metadati (il file CSV non ha metadati).
* Data la consegna, abbiamo identificativo delle task (numero progressivo), dati che coincidono con le task stesse e ogni informazione contenuta nel loro testo è un metadato.