

ESERCITAZIONE DATI DEL 7-2-2023

1. RECUPERARE 5 FONTI DATI DA INTERNET CHE DESCRIVANO LA SITUAZIONE COVID AL 2023 (CI DEVONO ESSERE ALMENO I CAMPI "POSITIVI COVID", "RICOVERATI", "MORTI")
2. IN UN DOCUMENTO WORD INDICARE LA FONTE, BREVE DESCRIZIONE DEL CONTENUTO, DESCRIZIONE DEI CAMPI E LA CLASSIFICAZIONE DELLE FONTI SCELTE
3. DEFINIRE UNA TABELLA (formato CSV) PER CONFRONTARE I DATI NUMERO POSITIVI COVID, PERSONE RICOVERATE, "CITTA'", "GIORNI POSITIVI" DI OGNI PARTECIPANTE
4. DARE UN NOME ALLE ENTITA' E AGLI ATTRIBUTI (CAMPI) IN MODO DESCRITTIVO
5. DEFINIRE LA CHIAVE PRIMARIA PER OGNI TABELLA
6. CONFRONTARE I DATI CLASSE/ITALIA E DESCRIVERE LA SITUAZIONE
7. DESCRIVERE IN QUALE FASE INTERVIENE OGNUNA DELLE FIGURE STUDIALE (Data Scientist, Data Engineer, Data Analyst e Data Journalist) SE PRESENTI
8. QUALI METADATI POTRETE IDENTIFICARE IN QUESTA ESERCITAZIONE?

1. FONTI:

Fonte 1: GIMBE

<https://coronavirus.gimbe.org/>

la Fondazione GIMBE pubblica settimanalmente i dati confermati dal Ministero della Salute.

Fonte 2: ISS

https://www.epicentro.iss.it/coronavirus/open-data/covid_19-iss.xlsx

I **dati** riportati nel file rappresentano aggregazioni dei casi positivi al SARS-CoV-2 desunti dalla Sorveglianza Integrata **Covid-19** coordinata dall'Istituto Superiore di Sanità (Ordinanza n. 640 del 27 febbraio 2020). Nel file sono presenti diversi fogli ognuno dei quali rappresenta una diversa distribuzione dei dati a partire dall'inizio dell'epidemia.

Fonte 3:

<https://www.salute.gov.it/portale/nuovocoronavirus/>

Monitoraggio settimanale dei casi covid da parte del Ministero della Salute.

Fonte 4:

<https://github.com/pcm-dpc/COVID-19>

Lo stato Italiano per informare i cittadini a messo a disposizione i dati raccolti dal Dipartimento della Protezione Civile, che vengono aggiornati quotidianamente.

FONTE: Stato italiano - Dipartimento della Protezione Civile

DESCRIZIONE: File CSV, Lo stato Italiano per informare i cittadini a messo a disposizione i dati raccolti dal Dipartimento della Protezione Civile, con licenza CC-BY-4.0, che vengono aggiornati quotidianamente.

CAMPI: ID, data, stato, ricoverati_con_sintomi, terapia_intensiva, totale_ospedalizzati, isolamento_domiciliare, totale_positivi, variazione_totale_positivi, nuovi_positivi, dimessi_guariti, deceduti

CLASSIFICAZIONE: Tre Stelle (★★★) per i file CSV importati in Excel passano a Due Stelle (★★).

CONSIDERAZIONI: Delle fonti dati trovate tutte portano alla fonte principale che è quella della Protezione Civile (Fonte 4). Essendo un sito con licenza CC-BY-4.0, questi dati possono essere liberamente riutilizzabili sempre con la fonte indicata.

Per l'inserimento dei dati della classe abbiamo utilizzato un forum di Google con gli attributi : ID, Data inserimento, Nome, Cognome, Città, Positivo, Ricoverato e Giorni Degenza.

Il confronto si può fare con i dati nazionali della Protezione Civile (fonte 4).

Notiamo un maggiore tasso di positività da parte del dato classe rispetto al dato nazionale dove i soggetti sono di numero inferiore ed di un' età che va dai 20 ai 40 anni, quindi con vita sociale più attiva.

Le figure Professionali che intervengono durante il processo del nostro progetto, sono:

DATA ANALYST

Raccolta dei i dati da fonti interne ed esterne.

Analisi dei dati nei formati arrivati Identificare gli obiettivi dell'analisi dei dati

DATA SCIENTIST

Creazione da parte del collega del forum Google per la raccolta dati della classe.

DATA JOURNALIST

Utilizzo de i dati per creare visualizzazioni per report, che semplificano alla portata della popolazione informazioni, raccontando storie per le trasmissioni, il web e i social media.

METADATI :

dati tabella exexcel

https://www.epicentro.iss.it/coronavirus/open-data/covid_19-iss.xlsx