

### **D3 - ESERCIZIO 3**

#### **COVID**

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 TESTO INDICARE LA FONTE, BREVE DESCRIZIONE DEL CONTENUTO, DESCRIZIONE DEI CAMPI E LA CLASSIFICAZIONE 5 STELLE DELLE FONTI SCELTE
3. DEFINIRE UNA TABELLA 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

#### **COVID – PARTE 2**

1.
  1. DEFINIRE LA CHIAVE PRIMARIA PER OGNI TABELLA
  2. CONFRONTARE I DATI CLASSE/ITALIA E DESCRIVERE LA SITUAZIONE
  3. DESCRIVERE IN QUALE FASE INTERVIENE OGNUNA DELLE FIGURE STUDIAE (Data Scientist, Data Engineer, Data Analyst e Data Journalist) SE PRESENTI
  4. QUALI METADATI POTETE IDENTIFICARE IN QUESTA ESERCITAZIONE?

## SOLUZIONE COVID PARTE 1 & 2

1ST DATASET: World Health Organization COVID-19 Dashboard.

1. <https://covid19.who.int/data>

- a) Data consultazione: 07/06/2023 ore 22:24
- b) Stato servizio: attivo (dato aggiornato il mercoledì di ogni settimana)
- c) La WHO Covid-19 Dashboard presenta dati giornalieri ufficiali sul numero di casi, decessi e vaccini, organizzati per nazioni e aree geografiche.
- d) METADATA: vengono indicate informazioni riguardanti il dataset, la modalità di raccolta dei dati e le rispettive fonti riportate per categoria (casi, positivi, decessi, popolazione, etc.)
- e) 5-Stars classification: \*\*\*\*\* dati strutturati, in formato aperto e dotati di un URI. Hanno, inoltre, collegamenti ad altri dataset (LDO)
- f) Licenza: non trovata
- g) Campi:
  - i. Data
  - ii. Codice della nazione
  - iii. Nazione
  - iv. WHO region (AFRO, AMRO, EURO, etc.)
  - v. Nuovi casi
  - vi. Casi cumulativi
  - vii. Nuovi decessi
  - viii. Decessi cumulativi

Ai fini della mia analisi, tuttavia, ho utilizzato unicamente le variabili *Data*, *Nazione*, *Casi cumulativi*, *Decessi cumulativi*.

NAZIONE	NUOVI POSITIVI	NUOVI DECESSI	TASSO DI MORTALITA' PER CASI
Italy	678002	5599	0,83%

Dalla tabella possiamo vedere che da inizio 2023 al 07/06/2023 in Italia ci sono stati *678.002 nuovi casi di positività*, *5599 decessi*, con una percentuale dello 0,83 % di decessi per numero di casi.

## 2nd DATASET: Our World In Data

### 1. <https://ourworldindata.org/coronavirus>

- a) Data consultazione: 07/06/2023 ore 22:43
- b) Stato servizio: attivo – I dati vengono aggiornati giornalmente
- c) OWI Coronavirus Pandemic data è una dashboard che permette agli utenti di esplorare le statistiche sulla situazione pandemica da Covid-19 per ogni nazione nel mondo
- d) METADATA: vengono indicate informazioni, organizzate per categoria (Profilo della nazione, Vaccini, Casi, Decessi, Ospitalizzazioni, etc.), attraverso le quali l'utente può esplorare il dataset
- e) 5-Stars classification: \*\*\*\*\* dati strutturati, in formato aperto e dotati di un URI. Hanno, inoltre, collegamenti ad altri dataset (LDO)
- f) Licenza: CC-BY-4.0, la quale consente la condivisione e l'utilizzo dei dati in modo aperto e flessibile. I dati possono essere modificati, trasformati o adattati secondo le esigenze dell'utilizzatore, creando così opere derivate. Tuttavia, l'utilizzatore deve rispettare i termini della licenza e fornire l'attribuzione all'autore originale
- g) Campi:
  - i. Continente
  - ii. Nazione
  - iii. Data
  - iv. Casi totali
  - v. Nuovi casi
  - vi. Decessi totali
  - vii. Nuovi decessi
  - viii. Ospedalizzazioni
  - ix. Totale vaccinati
  - Etc.

Ai fini della mia analisi, tuttavia, ho utilizzato unicamente le variabili *Data*, *Nazione*, *Casi Totali*, *Morti totali*, *Ospedalizzazioni*.

NAZIONE	NUOVI POSITIVI	NUOVI DECESSI	Δ OSPEDALIZZAZIONI
Italy	664741	5474	-77,78%

Dalla tabella risulta che da inizio 2023 al 31/05/2023 in Italia ci sono stati 664.741,00 nuovi casi di positività, 5.474 nuovi decessi e una variazione percentuale di ospedalizzazioni del -78%, ad indicare una diminuzione del 78 % di ricoveri dall'inizio del 2023 a fine maggio.

### 3rd DATASET: IISole24Ore IL LAB – Corona Virus

1) <https://lab24.ilsole24ore.com/coronavirus/>

- a) Data consultazione: 07/06/2023 ore 23:08
- b) Stato servizio: attivo – Ultimo aggiornamento 09/06/2023
- c) Il dataset indica i dati giornalieri della pandemia da Covid-19 in Italia per numero di positivi, ricoverati, guariti, deceduti, etc.
- d) METADATA: nella definizione delle risultanze statistiche, vengono indicati: il trend giorno per giorno, la crescita dei casi su base settimanale, le cinque regioni con più contagi ed altri contenuti.  
Rilevante è l'identificazione delle Fonti utilizzate per definire il dataset con relativa indicazione dei diversi URI.
- e) 5-Stars classification: \*\*\*\*\* dati strutturati, in formato aperto e dotati di un URI. Hanno, inoltre, collegamenti ad altri dataset (LDO).
- f) Licenza: CC-BY-4.0
- g) Campi:
  - i. Data
  - ii. Ricoverati
  - iii. Terapia intensiva
  - iv. Ospedalizzazioni
  - v. Totale positivi
  - vi. Variazione totale positivi
  - vii. Nuovi positivi
  - viii. Tamponi
  - ix. Deceduti
  - Etc.

Ai fini della mia analisi, tuttavia, ho utilizzato unicamente le variabili *Data*, *Ricoverati*, *Totale positivi*, *Ospedalizzazioni*, *Deceduti*.

Δ POSITIVI	Δ RICOVERATI	Δ DECEDUTI
-71,99%	-77,75%	3,00%

La tabella mostra che da inizio 2023 al 01/06/2023 in Italia c'è stata una diminuzione del numero di positivi del 72 %, dei ricoverati del 77% e un incremento del numero dei morti del 3%.

#### 4th DATASET: KAGGLE Covid Worldwide Dataset

- 1) <https://www.kaggle.com/datasets/themrityunjaypathak/covid-cases-and-deaths-worldwide>
  - a) Data consultazione: 10/06/2023 ore 10:43
  - b) Stato servizio: attivo – Aggiornato annualmente
  - c) Il dataset riporta il numero di casi e decessi mondiali giornalieri causati dal Covid-19
  - d) METADATA: viene descritto il Corona Virus e le sue implicazioni in chiave generalizzata. Sono indicati gli autori e gli eventuali collaboratori dell'analisi.
  - e) 5-Stars classification: \*\*\*\*\* dati strutturati, in formato aperto e dotati di un URI. Hanno, inoltre, collegamenti ad altri dataset (LDO)
  - f) Licenza: CC0 Public Domain, la quale consente ai detentori dei diritti di un'opera di rinunciare a tutti i loro diritti in tutto il mondo, concedendo il massimo livello di libertà possibile. La licenza consente a chiunque di utilizzare, modificare, distribuire e sfruttare l'opera a fini personali, commerciali, educativi o altri, senza restrizioni. Non sono richiesti attribuzione o riconoscimenti all'autore originale
  - g) Campi:
    - i. Serial number (identificativo per le nazioni)
    - ii. Nazioni
    - iii. Casi totali
    - iv. Decessi totali
    - v. Totale ricoverati
    - vi. Testo totali
    - vii. Popolazione (per nazione)

Ai fini della mia analisi, tuttavia, ho utilizzato unicamente le variabili *Nazioni*, *Casi totali*, *Decessi totali*, *Totale ricoverati*. In questo caso non abbiamo la serie storica dei dati, bensì viene indicato il numero di ogni variabile considerata a febbraio del 2023.

Country	Total Cases	Total Deaths	Total Recovered
Italy	25.453.789,00	186.833,00	25.014.986,00

## 5th DATASET: GITHUB Covid-19, Italia, Andamento nazionale

1. <https://github.com/pcm-dpc/COVID-19/blob/master/dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale.csv>
  - a) Data consultazione: 10/06/2023 ore 10:48
  - b) Stato servizio: attivo – Aggiornato quotidianamente alle 18:30, ultimo aggiornamento in data 09/06/2023
  - c) Il dataset è contenuto nella repository github della Presidenza del Consiglio dei Ministri – Dipartimento della Protezione Civile. Riporta i dati complessivi (cumulandoli per tutte le regioni) dell'andamento nazionale della pandemia.
  - d) METADATA: viene indicato il fine del Dipartimento della Protezione Civile, ovvero mettere a disposizione dei cittadini i dati sull'andamento della pandemia raccolti e aggiornati quotidianamente. Viene rappresentata anche una road map del flusso del dato:  
Regioni>Ministero della Salute>Dipartimento della Protezione Civile>Pubblicazione su Github e Dashboard
  - e) 5-Stars classification: \*\*\*\*\* dati strutturati, in formato aperto e dotati di un URI. Hanno, inoltre, collegamenti ad altri dataset (LDO)
  - f) Licenza: CC-BY-4.0
  - g) Campi:
    - i. Data
    - ii. Stato
    - iii. Ricoverati con sintomi
    - iv. Terapia intensiva
    - v. Totale ospedalizzati
    - vi. Totale positivi
    - vii. Deceduti
    - viii. Totale casi
    - ix. Nuovi positivi
    - Etc.

Ai fini della mia analisi, tuttavia, ho utilizzato unicamente le variabili *Data*, *Totale casi*, *Totale ospedalizzati*, *Deceduti*.

Totale ospedalizzati	Totale positivi	Deceduti
-81,60%	-72,34%	3,06%

La tabella mostra che da inizio 2023 al 08/06/2023 in Italia c'è stata una diminuzione del numero di positivi del 72,34 %, degli ospedalizzati del 81,6% e un incremento del numero dei morti del 3,06 %.

Concludendo, le evidenze sui dati analizzati risultano coerenti tra i diversi dataset utilizzati.

In particolar modo, i dataset 1 e 2 (WHO e OWI), analizzati con modalità speculari, mostrano le stesse risultanze, nonostante il leggero scostamento numerico derivante dalla differente data di aggiornamento del dato.

I dataset 3 e 5 (ILSOLE24H e GITHUB) sono stati analizzati con lo stesso approccio, con l'intento di identificare le variazioni percentuali delle variabili nel periodo. Anche in questo caso i risultati ottenuti risultano coerenti tra i due dataset.

Il dataset 4 non aveva le serie storiche dei dati, bensì i valori assoluti a febbraio 2023; tuttavia, i valori riportati risultano coerenti con gli altri dataset.

