# Esercitazione dati 7/02/2023

Fonti dati:

1. Dati covid-19 Italia – protezione civile:

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

La repository della protezione civile su github contiene i dati dall’inizio della pandemia ad oggi.

La repository ha la licenza creative commons BY.

I file csv riportati nella cartella dati andamento nazionale riportano i seguenti campi:

|  |  |
| --- | --- |
| **Campo** | **Descrizione** |
| data |  |
| stato |  |
| ricoverati\_con\_sintomi | pazienti che risultano ricoverati alla data di riferimento |
| terapia\_intensiva | pazienti in terapia intensiva alla data di riferimento |
| totale\_ospedalizzati | somma dei due campi precedenti |
| isolamento\_domiciliare | pazienti in terapia intensiva alla data di riferimento |
| totale\_positivi | Somma dei due campi precedenti |
| variazione\_totale\_positivi | Differenza con il totale positivi del giorno precendente |
| nuovi\_positivi |  |
| dimessi\_guariti |  |
| deceduti | Dato cumulativo da inizio pandemia alla data di riferimento |
| casi\_da\_sospetto\_diagnostico |  |
| casi\_da\_screening |  |
| totale\_casi | Dato cumulativo dall’inizio della pandemia ad oggi |
| tamponi | Dato cumulativo dall’inizio della pandemia ad oggi |
| casi\_testati |  |
| note |  |
| ingressi\_terapia\_intensiva |  |
| note\_test |  |
| note\_casi |  |
| totale\_positivi\_test\_molecolare |  |
| totale\_positivi\_test\_antigenico\_rapido |  |
| tamponi\_test\_molecolare |  |
| tamponi\_test\_antigenico\_rapido |  |

Classificazione: 3 stelle

Fonte utilizzata per il confronto con la casse

1. World Health organization

<https://covid19.who.int/WHO-COVID-19-global-data.csv>

Il file in formato csv riporta i dati divisi per data e paese dall’inizio della pandemia.

Contiene anche i dati relativi all’Italia.

I campi riportati in questo dataset sono i seguenti:

| **Campo** | **Descrizione** |
| --- | --- |
| Date\_reported | Data |
| Country\_code | ISO Alpha-2 country code |
| Country | Stato |
| WHO\_region | L’organizzazione mondiale della salute divide gli stati membri in sei aree così codificate: Regional Office for Africa (AFRO), Regional Office for the Americas (AMRO), Regional Office for South-East Asia (SEARO), Regional Office for Europe (EURO), Regional Office for the Eastern Mediterranean (EMRO), and Regional Office for the Western Pacific (WPRO). |
| New\_cases | Nuovi casi confermati. |
| Cumulative\_cases | Casi cumulate dall’inizio del monitoraggio della pandemia alla data di riferimento |
| New\_deaths | Nuove morti confermate |
| Cumulative\_deaths | Morti confermate dall’inizio della pandemia alla data di riferimento |

Classificazione: 3 stelle

1. Istituto superiore di sanità:

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

File in formato excel, riporta i dati sull’andamento covid divisi in fogli.

Il foglio “contenuto” è un metadato perché descrive il contenuto del resto dei fogli.

I fogli relativi all’andamento covid riportano le seguenti informazioni:

|  |  |
| --- | --- |
| **Foglio** | **Contenuto** |
| casi\_prelievo\_diagnosi | Numero casi per data prelievo/diagnosi aggiornati alla data riportata in iss\_date |
| casi\_inizio\_sintomi | Numero casi per data inizio sintomi aggiornati alla data riportata in iss\_date |
| casi\_inizio\_sintomi\_sint | Numero casi per data inizio sintomi i casi di infezione confermata da virus SARS-CoV-2 per cui è indicata una data inizio sintomi tranne i casi dichiarati asintomatici aggiornati alla data riportata in iss\_date |
| casi\_regioni | Numero totale casi dall'inizio dell'epidemia per regione/PA di diagnosi aggiornati alla data riportata in iss\_date |
| casi\_province | Numero totale casi dall'inizio dell'epidemia per provincia di domicilio/residenza aggiornati alla data riportata in iss\_date |
| ricoveri | Numero ricoveri per data di ricovero aggiornati alla data riportata in iss\_date |
| decessi | Numero decessi per data di decesso aggiornati alla data riportata in iss\_date |
| sesso\_eta | Numero totale casi e decessi dall'inizio dell'epidemia per sesso e per fascia d'età aggiornati alla data riportata in iss\_date |
| stato\_clinico | Numero totale casi con malattia ancora in corso e con esito finale ancora non definito dall'inizio dell'epidemia, differenziati per sintomatologia più recente, per sesso e per fascia d'età aggiornati alla data riportata in iss\_date |

Classificazione: 2 stelle (i dati sono strutturati ma per aprire il file bisogna avere un software proprietario)

1. Our world in data

<https://covid.ourworldindata.org/data/owid-covid-data.csv>

File in formato CSV che riporta numerosi dati di diversi paesi tra cui l’Italia.

I campi sono 67 e oltre ai dati su nazione, data, casi, decessi e ricoveri (riportati di seguito) includono anche dati su vaccini, tamponi, fasce d’età dei pazienti.

|  |  |
| --- | --- |
| **Campi** | **descrizione** |
| iso\_code | Codice Iso della nazione |
| continent | Continente |
| location | Nazione |
| date | Data |
| total\_cases | Casi totali alla data di riferimento |
| new\_cases | Nuovi casi |
| total\_deaths | Morti totali alla data di riferimento |
| new\_deaths | Nuove morti |
| icu\_patients | Pazienti in terapia intensiva |
| hosp\_patients | Pazienti ospedalizzati |
| weekly\_icu\_admissions | Ingressi settimanali in terapia intensiva (data di riferimento più i 6 giorni precedenti) |
| weekly\_hosp\_admissions | Ingressi settimanali in ospedale (data di riferimento più i 6 giorni precedenti) |

Classificazione: 3 stelle

**Confronto con la classe:**

I dati sull’andamento covid della classe sono stati raccolti tramite un sondaggio a cui hanno risposto 30 partecipanti.

I risultati del sondaggio sono stati estratti in formato CSV.

I campi definiti nel CSV sono stati:

* ID (1,2,3,4 ecc.) – (chiave primaria)
* Informazioni cronologiche (data e ora di raccolta di ogni record)
* Nome
* Cognome
* Città
* Positivo (Si/no)
* Ricoverato (Si/no)
* Giorni di degenza (giorni di positività)

Sul totale classe 18 persone su 30 hanno avuto il covid, nessuna è stata ricoverata, in media i giorni di positività per chi è risultato positivo sono stati 12.

Confrontano la situazione della classe con la situazione nazionale possiamo dire che in percentuale il tasso di positività della classe (60%) è stato più alto rispetto al totale nazionale (43%).

La percentuale di positivi ricoverati in classe è dello 0%, la percentuale di positivi ricoverati a livello nazionale è del 1,25%.

Nello svolgimento dell’esercizio possiamo dire che alcuni di noi hanno ricoperto il ruolo di data engineer strutturando il sondaggio per la raccogliere i dati della classe e condividendo i dati raccolti in formato csv.

Ci siamo poi calati tutti nei panni del data analyst nel confronto dei dati nazionali con i dati della classe.