

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. 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 STUDIALE (Data Scientist, Data Engineer, Data Analyst e Data Journalist) SE PRESENTI
4. QUALI METADATI POTETE IDENTIFICARE IN QUESTA ESERCITAZIONE?

PRIMO DATASET:

https://ourworldindata.org/explorers/coronavirus-data-explorer?zoomToSelection=true&time=2020-03-01..latest&facet=none&country=~CAN&pickerSort=desc&pickerMetric=hospital_beds_per_thousand&Metric=Confirmed+cases&Interval=7-day+rolling+average&Relative+to+Population=true&Color+by+test+positivity=false

Questo dataset risulta di facile fruizione, i dati sono strutturati e separati in forma tabellare, è disponibile un download del CSV. I dati sono fruibili direttamente dal link oppure si può accedere alla loro repository su github, il dato è stato visualizzato anche su excel e risulta chiara e senza errori.

I metadati sono presenti ed è presente anche la possibilità di vedere l'excursus del covid attraverso un diagramma cartesiano interattivo.

Gli attributi sono chiari.

La licenza risulta essere Creative Common.

La fonte è **WHO COVID-19 Dashboard. Geneva: World Health Organization, 2020**

Darei 4 stelle a questo Dataset a causa di una sovrapposizione troppo accentuata sui menu di impostazione dei dataset. Non è stato facile l'accesso alle licenze disponibili.

SECONDO DATASET:

<https://www.kaggle.com/datasets/chitrakumari25/corona-virus-la-test-data-2023>

Questo dataset è chiaro, i metadati sono accurati e di facile fruizione. I dati sono strutturati, ma gli attributi hanno alcune digressioni testuali. A disposizione nel download è presente il CSV. La repository non risulta essere caricata su Github, ma permettono il download diretto dal loro link

La licenza è **ODC Public Domain Dedication and Licence (PDDL)**

La fonte è **WHO COVID-19 Dashboard. Geneva: World Health Organization, 2020**

Darei 3 stelle a questo dataset, per quanto sia molto comoda la possibilità di accesso a tutti i dati dalla pagina trovo che i dati forniti siano meno accurati e lavorati del precedente.

TERZO DATASET:

<https://www.kaggle.com/datasets/anandhuh/covid19-weekly-trends-in-europe-latest-data>

Questo dataset risulta parzialmente completo, la mancanza di metadati dettagliati aumentano la difficoltà di interpretazione. I dati sono lavorati. È presente un download in formato CSV.

La fonte è https://www.worldometers.info/coronavirus/weekly-trends/#weekly_table.

La licenza è un open data Commons.

Darei 2 stelle a questo dataset, la mancanza di metadati approfonditi lo rende decisamente meno efficiente dei precedenti, anche se in questo caso la fonte risulta essere differente.

QUARTO DATASET:

<https://www.kaggle.com/datasets/anandhuh/covid19-in-world-countrieslatest-data>

Questo dataset risulta chiaro e dalla semplice fruizione sulla piattaforma Kaggle, gli attributi sono chiari e messi in ordine, è possibile scaricare il CSV. I dati sono lavorati e le informazioni sono chiare a prima vista.

La fonte è <https://www.worldometers.info/coronavirus/#countries>

La licenza è Open data Common.

Darei 3 stelle a questo dataset, gli attributi sono chiari, ma i dataset risultano non approfonditi.

QUINTO DATASET: <https://ourworldindata.org/covid-deaths>

Come nel primo caso questo dataset risulta molto fruibile e approfondito, mette a disposizione molte informazioni ma la loro percezione non è semplice. Gli attributi sono dettagliati e anche i metadati risultano adeguatamente preparati.

Il dataset è disponibile in download sia in .CSV che in .xlsx che in .json.

la licenza è Open source.

Darei 4 stelle a questo dataset, per la sua semplicità di utilizzo.

<https://www.kaggle.com/datasets/chitrakumari25/corona-virus-la-test-data-2023> (dataset2)

Data Engineer:

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

Il lavoro di data engineering è stato sviluppato dal di WolrdOmeters che recuperano i dati riguardo la popolazione e i dettagli sul covid costantemente.

Data Analyst:

Il lavoro fatto dal data analyst è stato la trasformazione dei dati grezzi in dati lavorati, in uno schema che ha selezionato 9 attributi su cui basare le sue analisi e mettere in evidenza I casi totali, i casi attivi e metterli in relazione con la popolazione.

Data scientist:

Non mi sembra che in questo dataset sia stato effettuato alcun lavoro di previsione.

I metadati sono quei dati che vanno a darci informazioni riguardo i dati che stiamo interrogando.

