**Esercizio 3 – Modulo 1 @Epicode**

Task:

COVID – PARTE 1

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”, “CITTÀ”, “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 STUDIATE (Data Scientist, Data Engineer, Data Analyst e Data Journalist) SE PRESENTI

4. QUALI METADATI POTETE IDENTIFICARE IN QUESTA ESERCITAZIONE?

Prima fonte:

<https://ourworldindata.org/explorers/coronavirus-data-explorer?zoomToSelection=true&time=2025-08-02&facet=none&country=~ITA&pickerSort=asc&pickerMetric=location&Metric=Confirmed+cases&Interval=7-day+rolling+average&Relative+to+Population=true&Color+by+test+positivity=false>

Data di consultazione: 08/06/2023 ore 18.05

Status: attivo

Questo dataset presenta dei contenuti accessibili e fruibili in chiave interattiva; molto utili i filtri che consentono di selezionare la nazione di riferimento e visualizzare le metriche correlate, come le morti confermate, i casi confermati, i casi di ospedalizzazione, i casi di terapia intensiva, e molte altre) e si possono scegliere visualizzazioni ad istogramma, a mappa ecc. Dal dataset è possibile ricavare i dati tramite un file .csv scaricabile, e ci sono diversi metadati e link ad altri dataset/ai dati GitHub, per cui ritengo sia classificabile come cinque stelle.

Seconda fonte:

<https://www.ecdc.europa.eu/en/covid-19/data>

Data di consultazione: 08/06/2023 ore 18.10

Status: attivo

A questa pagina è possibile accedere ad un insieme di diversi dataset, che possono essere confrontati per raccogliere tipi di dati differenti e trarre conclusioni sull’andamento della pandemia. A differenza del dataset precedente, non vi sono dashboard interattive con cui visualizzare i dati, ma vi sono diversi dati accessibili corredati da metadati, in particolare descrittivi. Vista la presenza di link, ritengo si possa classificare anche questo dataset come cinque stelle.

Terza fonte:

<https://health.google.com/covid-19/open-data/explorer?loc=IT>

Data di consultazione: 08/06/2023 ore 18.12

Status: aggiornato al 15/09/2022, ma attivo

Anche in questo caso, il dataset presenta una modalità di visualizzazione interattiva arricchita con filtri di selezione; nel caso in esame sono inoltre presenti dati relativi alle singole regioni italiane, aggiornati però al 15/09/2022. Selezionando il periodo di riferimento mediante il filtro in alto a sinistra, è possibile filtrare per nazione e regione, e visualizzare il numero di casi e decessi; in base ai filtri applicati, sono visualizzabili dati differenti sui vari paesi, anche a seconda della disponibilità di dati. Il dataset riporta dei metadati e link alle data sources disponibili, che rimandano ai dati forniti dai vari paesi e dagli enti dedicati. Classificherei anche questo dataset come cinque stelle.

Quarta fonte:

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

Data di consultazione: 08/06/2023 ore 18.15

Status: attivo

Questo dataset risulta molto più schematico rispetto a quelli precedentemente analizzati; sono presenti dei link per il download dei dati e vengono anche riportate tabelle di dati che specificano le tipologie di dati riportati (in un certo senso, la tabella costituisce una sorta di chiave di lettura dei dati presenti, in quanto indica ad esempio “Field name = New\_cases”, “Type = Integer” e “Description = New confirmed cases. Calculated by subtracting previous cumulative case count from current cumulative cases count”). All’interno del dataset sono reperibili dati come nuovi casi, nuove morti, casi totali, morti totali, nazione di riferimento, ma non sono presenti dati relativi ai ricoveri e ai pazienti guariti. In ogni caso, assegnerei cinque stelle anche a questo dataset.

Quinta fonte:

<https://dati-covid.italia.it/>

Data di consultazione: 08/06/2023 ore 18.23

Status: attivo

Questo è il dataset migliore per un’analisi a livello nazionale poiché ci sono dati molto aggiornati e dettagliati a livello territoriale, aggiornati peraltro ogni ora, e i dati risultano sparsi in più file .csv che è necessario confrontare. Sono presenti dati a livello nazionale sui ricoverati su sintomi, sugli ospedalizzati, sui pazienti positivi, sui dimessi, sulla variazione del totale di positivi e molto altro, nonché dati relativi alle sole regioni o province. Inoltre, sono presenti anche dei dataset in formato JSON, che riportano le informazioni come mostrato di seguito:

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

Assegnerei cinque stelle anche in questo caso, considerando la presenza di link ai dataset su GitHub e alle Dashboard (che però mi risultano inaccessibili).

Tabella

|  |  |  |  |
| --- | --- | --- | --- |
| Paese | Tot. Nuovi Casi Positivi | Totale Ricoverati | Totale Nuovi Morti |
| Italia | 2858 | 1922 | 24 |
| Francia | 902 | n.d. | 34 |
| Germania | 964 | n.d. | 16 |
| Spagna | 0 | 2279 | 0 |
| Portogallo | 360 | n.d. | 5 |

Dati presi da un file .csv relativo al dataset accessibile a questo link (non inserisco il file su GitHub solo perché è davvero molto grande come dimensioni): <https://ourworldindata.org/explorers/coronavirus-data-explorer>

I totali sono riferiti al numero di nuovi casi positivi e di nuove morti, e al numero di pazienti ricoverati al 31/07/2023, stando a quanto risulta nel dataset e nel file csv. Purtroppo per alcuni paesi non sono disponibili dati in merito al numero di persone ricoverate, ma di per sé è stato molto complicato reperire dataset completi riportanti tutte le informazioni richieste per l’esercizio.

Parte 2:

Dataset di riferimento (dashboard Worldometers Italia): <https://www.worldometers.info/coronavirus/country/italy/>

Data Engineer: Questa figura progetta, sviluppa e gestisce le infrastrutture necessarie alla raccolta e memorizzazione dei dati, e si assicura che vengano organizzati nei modi e formati più adatti al caso.

Data Scientist: Questa figura sfrutta modelli predittivi e conoscenze di matematica, statistica, calcolo delle probabilità e machine learning per fare delle previsioni derivanti dai dati in suo possesso.

Data Analyst: Questa figura ispeziona e analizza i dati, e nel caso specifico ha curato anche l’implementazione grafica delle dashboard in modo che siano facilmente fruibili.

Nell’esercitazione si possono identificare diversi set di metadati nel corso dello svolgimento, e durante la consultazione dei diversi dataset / dashboard interattive reperite. Nello specifico, si possono identificare metadati prevalentemente descrittivi, che descrivono il documento o gli oggetti cui sono associati, ma in alcuni casi è possibile trovare anche metadati amministrativi/gestionali (che forniscono info su come è stato o deve essere trattato il documento, es. le licenze cui è sottoposto) ed anche metadati strutturali (che permettono di localizzare il documento all’interno di un sistema di conservazione, o un database, es. i link al dataset su github); un esempio parlante è questo dataset (esempio extra rispetto a quelli precedentemente riportati), nel quale sono presenti tutti e tre i tipi di metadati: <https://www.kaggle.com/datasets/sudalairajkumar/covid19-in-italy>

