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

**SVOLGIMENTO - PARTE 1**

**ESERCIZIO 1 - 2**

1. **Our World in Data**

Scaricato il 07/06/2023, ore 18:00. Classificazione: ⭐⭐⭐⭐⭐

Link: <https://ourworldindata.org/explorers/coronavirus-data-explorer>.

Il dataset contiene numerosissime statistiche riguardanti l’andamento dell’epidemia Covid 19 a livello globale negli ultimi 3 anni. L’interrogazione dei dati è resa rapida ad efficace grazia alla possibilità di filtraggio per paese, campo (esempio numero totale deceduti, vaccinazioni, casi, ospedalizzazioni e così via) e data e all’utilizzo di numerose infografiche interattive. Si tratta dunque di dati strutturati, liberamente accessibili e scaricabili in più formati (.csv, .xslx, .json). Vengono inoltre riportate le fonti dei dati allegando i link specifici.

1. **World Health Organization**

Scaricato il 07/06/2023, ore 18:00. Classificazione: ⭐⭐⭐⭐⭐

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

Anche in questo caso ci si trova a che fare con una grande quantità di dati strutturati, liberamente e facilmente navigabili e scarabilli. Sono citate le fonti e le modalità di raccolta. Come prima i dati sono raccolti a livello globale.

1. **DATI COVID ITALIA**

Scaricato il 07/06/2023, ore 18:00. Classificazione: ⭐⭐⭐⭐

Link: <https://github.com/pcm-dpc/COVID-19/tree/master/dati-andamento-nazionale>

Si tratta di una repository GitHub che raccoglie numerosi dati a livello nazionale raccolti in vari sub cartelle. I dataset sono aggiornati in modo frequente. Per l’esercizio si prende in considerazione il seguente database: <https://github.com/pcm-dpc/COVID-19/blob/master/dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale-20230601.csv>, liberamente scaricabile in formato .csv. Inoltre, questa repository viene frequentemente utilizzata da terzi per elaborazioni proprio, come nel caso del Sole 24 Ore (<https://lab24.ilsole24ore.com/coronavirus/>)

1. **GOOGLE**

Scaricato il 07/06/2023, ore 18:00. Classificazione: ⭐⭐⭐⭐⭐

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

Si tratta di un database piuttosto dettagliata che può essere navigato comodamente. Si può filtrare la ricerca per nazione e regioni del mondo. Numerosi link di collegamento al altri database. Pecca importante è che il database non è oggetto di aggiornamento a partire da novembre 2022.

1. **European Centre for Disease Prevention and Control ECDPC**

Scaricato il 07/06/2023, ore 18:00. Classificazione: ⭐⭐⭐⭐⭐

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

Non si tratta di un singolo database ma piuttosto di una raccolta di più database specifici. Non sono presenti infografiche in sito, ma i dati sono ampiamente fruibili e liberamente scaricabili in diversi formati, incluso .csv.

**ESERCIZIO 3-4, TABELLA: File Excel in GitHub**

**PARTE 2**

**ESERCIZIO 3:** il Data Engineer avrà sicuramente provveduto alla raccolta e all’immagazzinamento della grande mole e varietà di dati, provvedendo alla corretta formattazione al fine di produrre dati ben strutturati e facilmente interrogabili. Il data analyst avrà probabilmente lavorato su questi dati per aiutare nella modulazione della produzione e della messa in commercio di prodotti richiesti durante la pandemia (per esempio, un data analyst di una casa farmaceutica o di un produttore di DPI quali mascherine ecc.). Il data scientist avrà certamente elaborato dei modelli statistici predittivi per cercare di anticipare l’andamento del virus. Infine, il data journalist avrà elaborato dei documenti divulgativi ricchi di dati al fine di informare puntualmente gli utenti sui punti di maggior interesse.

**ESERCIZIO 4:** In tutti i dataset sono presenti numerosissimi metadati descrittivi, gestionali e strutturali. Essi informano sulla natura dei dati, sul loro significato, sul modo in cui sono stati raccolti ed organizzati. La chiarezza di tali metadati è fondamentale in dataset di questo tipo, i quali oltre ad essere densi di dati, presentano evidenti difficoltà interpretative.