**1 – ISTITUTO SUPERIORE DI SANITÀ -** ★★

File: **covid\_19-iss fonte 1.xlsx**

È un file excel, ottenuto dal sito dell’Istituto superiore di sanità, contenente i dati sull’andamento dei casi, decessi, ricoveri e altro dall’inizio pandemia ad oggi.

La chiave primaria varia a seconda della tabella:

* casi\_prelievo\_diagnosi: DATA\_PRELIEVO\_DIAGNOSI
* casi\_inizio\_sintomi: DATA\_INIZIO\_SINTOMI
* casi\_inizio\_sintomi\_sint: DATA\_INIZIO\_SINTOMI
* casi\_regione: NOME\_NUTS2
* casi\_provincie: COD\_PROV
* ricoveri: DATARICOVERO1
* decessi: DATA\_DECESSO
* sesso\_eta: per questa tabella sarà opportuno fare una suddivisione tra i sessi, creando 3 diverse tabelle, e utilizzare come chiave primaria AGE\_GROUP per ognuna di esse. Oppure si potrebbe creare un identificativo numerico (es. serie di numeri)
* stato\_clinico: qui bisognerà procedere come per la tabella “sesso\_età” e poi creare delle ulteriori tabelle distintive per AGE\_GROUP. Così la chiave primaria sarà STATO\_CLINICO. Oppure con id numerico

In questo caso dietro alla raccolta dati ci sarà il data engineer che svilupperà un metodo di inserimento/prelievo dati attraverso gli strumenti della sanità pubblica, essendo una fonte dati del governo. Successivamente un data analyst dovrà ispezionare i dati e creare le tabelle citate.

**2 - REPOSITORY DATI ANDAMENTO NAZIONALE GOVERNO -** ★★★

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

<https://github.com/pcm-dpc/COVID-19/blob/cae4442f4dd4e157fe18677a372d7be1c915e083/dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale-20230202.csv> (02/02/2023)

Questa è una fonte dati trovata all’interno del sito del Ministero della Salute messa a disposizione liberamente per il download in formato CSV. Questa è, probabilmente, la fonte dati utilizzata per l’elaborazione del file della fonte 1. Il primo link rimanda alla repository Github contenente i file suddivisi per data dell’andamento giornaliero e da inizio pandemia.

La chiave primaria è la data.

È il lavoro svolto dal data engineer della fonte 1.

Licenza CC-BY-4.0

**3 – OPEN WORLD IN DATA -** ★★★★★

https://ourworldindata.org/covid-vaccinations?country=OWID\_WRL

Il link rimanda al sito completo di mappe, grafici, tabelle ed è disponibile il download in file CSV. Contiene l’andamento covid suddiviso per paese da inizio pandemia.

La chiave primaria è, dopo aver preso come riferimento solo i dati riguardanti l’Italia, la data.

In questo caso il lavoro sarà affidato ad un data engineer che si occuperà della ricerca dei dati e il controllo della loro correttezza. Successivamente un data analyst dovrà ispezionare i dati e creare i modelli visivi (grafici, mappe).

Licenza CC-BY

**4 – CORRIERE DELLA SERA -** ★

File**: fonte 4.png**

È un’immagine contenente una tabella che mostra l’aggiornamento giornaliero dei casi covid e il totale dei casi da inizio pandemia.

La chiave primaria è la data.

In questo caso, essendo una testata giornalistica l’autrice di questo file, sarà un data journalist a svolgere il lavoro di ottenimento e visualizzazione dati. Quindi farà sia la parte del data engineer che del data analyst.

**5 – HEALTH DATA -** ★★★★★

<https://covid19.healthdata.org/italy?view=cumulative-deaths&tab=trend>

<https://www.healthdata.org/covid/data-downloads> (Reference scenario, per i dati; Data dictionary, per la legenda)

Questa fonte contiene grafici dell’andamento covid suddiviso per paese da inizio pandemia, con la possibilità di scaricare in formato CSV i dati.

I campi sono specificati nel file allegato, nominato “data\_dictionary”, che contiene la legenda.

La chiave primaria è, dopo aver filtrato per location o location id, la data.

Le figure coinvolte saranno le stesse della fonte dati nr. 3, ma con la differenza che all’interno di questo file ci sono anche dei dati di stima che verranno eseguiti da un data scientyst.

**6 - FILE CLASSE CASI COVID DA INIZIO PANDEMIA -** ★★

In questa fonte vengono esposti i dati riguardanti i casi covid all’interno della classe da inizio pandemia, con in aggiunta la specifica dei giorni di degenza.

La chiave primaria è l’ID.

Questo file non è un’open data poiché non è disponibile a tutti. Però supponendo che sia di libero accesso sarebbe considerato un ★★.

**METADATI**

I metadati che possiamo individuare per le fonti descritte sono:

* autore
* nome del file
* tipo di file
* posizione del file
* dimensione del file
* data di creazione
* parole chiave contenute all’interno dei file, per esempio:
  + casi regione
  + casi province
  + totale casi
  + ricoveri
  + decessi
  + sesso
  + età
  + covid
  + andamento
  + ecc..

**CONFRONTO**

Da un veloce confronto tra i dati ottenuti dai vari open data e i dati sui casi della classe, possiamo notare che i positivi nella classe sono stati in % di più rispetto all’andamento a livello nazionale. Infatti in Italia la % media di casi di positività è stata del 43% (rispetto al totale della popolazione), mentre all’interno della classe il 60% degli studenti ha contratto il covid (rispetto al totale studenti esaminati). Fortunatamente non ci sono stati casi di ricovero tra gli studenti presi in esame. Questo valore invece è in linea con i dati nazionali pari a meno del 1%.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 02/02/2023 | TOTALE POSITIVI | TOTALE RICOVERI | RICOVERI 02/02/2023 | TOTALE MORTI |
| FONTE 1 | 25.700.453 | 402.617 | <5 | 185.882 |
| FONTE 2 | 25.488.166 |  | 179 | 187.272 |
| FONTE 3 | 25.453.789 | 578.785 | 179 | 186.833 |
| FONTE 4 | 25.488.166 |  | 179 | 187.272 |
| FONTE 5 |  |  | 447 | 185.231 |
| FONTE 6 | 18 | 0 | 0 | 0 |
|  |  |  |  |  |
| SU TOT. POPOLAZ. | 59.110.000 | SU TOTALE STUDENTI | 30 |  |
|  | TOTALE POSITIVI | TOTALE RICOVERI | TOTALE MORTI |  |
| FONTE 1 | 43,48% | 0,68% | 0,31% |  |
| FONTE 2 | 43,12% |  | 0,32% |  |
| FONTE 3 | 43,06% | 0,98% | 0,32% |  |
| FONTE 4 | 43,12% |  | 0,32% |  |
| FONTE 5 |  |  | 0,31% |  |
| FONTE 6 | 60,00% | 0,00% | 0,00% |  |

Commento tabella:

I dati ottenuti per il totale ricoveri per la fonte 1 e 3 sono stati ottenuti elaborando i dati dei file principali.

Il dato dei ricoveri del 02/02/2023 per la fonte 5 è una stima, mentre per la fonte 1 è specificato nel file che i dati più recenti possono non essere del tutto aggiornati e quindi grezzi.

Le celle lasciate vuote sono dati non riusciti a reperire.