Dapt2023: Andrea D’Amore

Analisi e ricerca su dataset covid 19 prendendo in considerazione come pivot di ricerca “ casi, deceduti, e ricoveri”

Fonti Ispezionate:

1. <https://www.worldometers.info/coronavirus/#main_table><https://github.com/pcm-dpc/COVID-19/blob/master/dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale.csv>  
     
   A. Data consultazione 08/06/2023

B. Stato del servizio: attivo

C. Descrizione della fonte: interfaccia API, dati scaricabili, possibilità di filtrare. Ente privato, settore sanitario

D. Campi

i.Casi: numero intero che conta gli esiti positivi ai tamponi

ii. Ricoverati

iii. Deceduti  
  
 E. Tabella estratta da fonte sopraindicata: [dpc-covid19-ita-andamento-nazionale](https://docs.google.com/spreadsheets/d/1ydo_kvsGi_MDs50mQC9mjGGW2_0uG1aImR_wUNA0zng/edit#gid=944068528), filtrata al 2022

1. <https://www.ecdc.europa.eu/en/publications-data/data-national-14-day-notification-rate-covid-19>  
     
   A. Data consultazione 08/06/2023

B. Stato del servizio: attivo

C. Descrizione della fonte: dati scaricabili. Ente pubblico, settore sanitario

D. Campi

i.Casi: numero intero che conta gli esiti positivi ai tamponi

ii. Ricoverati(null)

iii. Deceduti  
  
 E. Tabella estratta da fonte sopraindicata: [Dataset Europa.eu](https://docs.google.com/spreadsheets/d/1IhAPlZRHI52-_Rsk7cdm7VCKGi2Fr6_rTaoIvSe9srA/edit#gid=690816822),filtrata al 2022.

1. <https://ourworldindata.org/>   
     
   A. Data consultazione 08/06/2023

B. Stato del servizio: attivo

C. Descrizione della fonte: Dati scaricabili. Ente pubblico, settore sanitario

D. Campi

i.Casi: numero intero che conta gli esiti positivi ai tamponi

ii. Ricoverati

iii. Deceduti  
  
 E. Tabella estratta da fonte sopraindicata: [Casi 2022](https://docs.google.com/spreadsheets/d/1IyyUkZfhyjWAWbgQnvwrhykfm2AjPMtdCRoyaKiICtc/edit#gid=2007548001) & [biweekly\_deaths](https://docs.google.com/spreadsheets/d/1Kfv1HMWuORmSnwtsNhtEGyNetzrmRF4noeXogvdZ5sk/edit#gid=680892184) & [Ricoveri ITA](https://docs.google.com/spreadsheets/d/1u8fLbTCkwopcqgvqhR6z_DmrWz-6ecHXWbxGnfEShdk/edit#gid=2063504222)

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

A. Data consultazione 09/06/2023

B. Stato del servizio: attivo

C. Descrizione della fonte: Dati scaricabili. Ente pubblico, settore sanitario

D. Campi

i.Casi: numero intero che conta gli esiti positivi ai tamponi

ii. Ricoverati

iii. Deceduti  
  
 E. Tabella estratta da fonte sopraindicata: [WHO-COVID-19-global-data](https://docs.google.com/spreadsheets/d/1gaCWTaE-6noMPh08srAzI4qwDFgIDULd6TZ5QfdIQmw/edit#gid=881188217)

1. <https://www.kaggle.com/datasets/sudalairajkumar/covid19-in-italy?resource=download&select=covid19_italy_region.csv>

A. Data consultazione 09/06/2023

B. Stato del servizio: inattivo

C. Descrizione della fonte: Dati scaricabili. Ente pubblico, settore sanitario

D. Campi

i.Casi: numero intero che conta gli esiti positivi ai tamponi

ii. Ricoverati

iii. Deceduti  
  
 E. Tabella estratta da fonte sopraindicata: [kaggle dataset](https://docs.google.com/spreadsheets/d/1N-ZeO20XsGH6nJT9L3xavxDKH-m9gygzCkONfBDkYIE/edit#gid=1063659607)

| Fonti | casi | decessi | ricoveri |
| --- | --- | --- | --- |
| worldometers | 18925796 | 47405 | 3436491 |
| edc.europa.eu | 25441023 | 192166 | null |
| ourworldindata | AVG daily 2022: **746794** | 668571 | 1747922 |
| WHO.int | 25168438 | 184792 | Null |
| kaggle.com | 1729061 | 8866632 | 3405170 |

Parte 2  
  
Prendendo come esempio questa fonte:<https://covid19.who.int/data>, cercherò di rispondere alla domanda : “ DESCRIVERE IN QUALE FASE INTERVIENE OGNUNA DELLE FIGURE STUDIATE (Data Scientist, Data Engineer, Data Analyst e Data Journalist) SE PRESENTI”  
  
  
Il data engineer si occuperà della fase di raccolta dati iniziale, può essere una figura che si interfacci con le varie istituzioni che si sono occupate dell’emergenza sanitaria in prima linea, tra cui la protezione civile e il sistema sanitario.   
  
Successivamente il data engineer, dopo aver raccolto tutti questi input, aggrega i dati a sua disposizione in un unico database, rendendo così disponibile una base di dati sgrezzata e pulita per il data analyst che si impegnerà a studiarla e interpretarla, rendendola capibile per persone terze o per un ramo aziendale che ha richiesto un determinato studio sull’argomento.   
  
Il data scientist può intervenire su questo dataset magari utilizzandolo per scopi predittivi e di sviluppo della pandemia, creando delle previsioni e anticipando l’andamento del covid 19, ipotizzo possa anche verificare l’incidenza di vaccini/non vaccini a lungo termine.

QUALI METADATI POTETE IDENTIFICARE IN QUESTA ESERCITAZIONE?

I Metadati incontrati durante questa ricerca sono stati molteplici, tra cui le descrizioni trovate sui vari siti che “mi hanno guidato” alla lettura dei vari dataset strutturati,