**ESERCIZIO 3** – 5 GIUGNO

**Autore:** Sara G.

**Data:** 7th June 2023

**Obiettivo:** Confronto di 5 fonti da dati internet che descrivano la situazione covid al 2023

Immagine che contiene testo, schermata, Carattere, documento

Descrizione generata automaticamente

1. ***Source:*** **WHO World Health Organization**

<https://app.powerbi.com/view?r=eyJrIjoiYWRiZWVkNWUtNmM0Ni00MDAwLTljYWMtN2EwNTM3YjQzYmRmIiwidCI6ImY2MTBjMGI3LWJkMjQtNGIzOS04MTBiLTNkYzI4MGFmYjU5MCIsImMiOjh9>

* ***Data di consultazione:***

*7 Giugno*

* ***Stato servizio:***

*ATTIVO*

* ***Descrizione della fonte:***

La fonte è il corrispettivo dell’OMS (Organizzazione Mondiale della Sanità) parte specializzata dell’ONU per la sanità, quindi ***ente pubblico*** mondiale.   
I dati forniti sono ***consultabili online*** molto facilmente da esperti e non vista la ***comprensibile interfaccia grafica***. In una sezione dedicata è possibile modificare i vari attributi e anche scaricare i relativi ***dataset filtrati***.

* ***Campi:***
* *Country Name*
* *Country Code*
* *ISO Year*
* *ISO Week*
* *ISO Start Date*
* *Sex*
* *Age group*
* *Daily Cases*
* Detailed Cases deaths
* *Daily Cases deaths*
* Detailed Cases
* Hospitalized
* *⋆⋆⋆⋆⋆* Dati aperti. Possibile modificare senza necessita del software proprietario. Dati strutturati e codificati in *formato aperto* come CSV, *collegati ad altri dataset* e raggiungibili anche da URL.

1. ***Source:*** **The European Surveillance System (TESSy) from Member States**

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

* ***Data di consultazione:***

7 Giugno

* ***Stato servizio:***

ATTIVO

* ***Descrizione della fonte:***

Ente pubblico a supporto della sanità europea per casi specifici. I dati forniti sono ***consultabili online*** ma solo infografiche **non interattive**. In una sezione dedicata è possibile **scaricare i dataset** completi divisi in categorie e successivamente da filtrare manualmente.

* *⋆⋆⋆* Dati aperti. Possibile modificare senza necessita del software proprietario. Dati strutturati e codificati in *formato aperto* come CSV, *collegati ad altri dataset* e raggiungibili anche da URL.
* ***Campi:***

Immagine che contiene testo, Carattere, numero, schermata

Descrizione generata automaticamente

Immagine che contiene testo, ricevuta, schermata, Carattere

Descrizione generata automaticamente

1. ***Source:*** **Protezione Civile data in Worldometres**

<https://www.worldometers.info/coronavirus/country/italy/>

<https://github.com/pcm-dpc/COVID-19>

* ***Data di consultazione:***

7 Giugno

* ***Stato servizio:***

ATTIVO

* ***Descrizione della fonte:***

La protezione civile è un ente pubblico a supporto della sanità italiana per casi specifici. I dati forniti sono ***consultabili online*** ma solo infografiche **non interattive**. In una sezione dedicata è possibile **scaricare i dataset** completi divisi in categorie e successivamente da filtrare manualmente.

**Licenza Creative Commons Attribution 4.0 International**

* *⋆⋆⋆⋆⋆* Dati aperti. Possibile modificare senza necessita del software proprietario. Dati strutturati e codificati in *formato aperto* come CSV, *collegati ad altri dataset* e raggiungibili anche da URL.
* ***Campi:***
* **Casi totali** = conteggio cumulativo totale riferito di rilevato e di laboratorio (e talvolta, a seconda del paese che li segnala e dei criteri adottati al momento, anche clinicamente) confermato positivo e talvolta - a seconda dei paesi standard di segnalazione - anche casi presuntivi, sospetti o probabili di infezione rilevata. Poiché rappresenta un conteggio cumulativo (piuttosto che un'istantanea del numero di casi correnti in un dato momento), questo numero non può diminuire. L'entità del divario tra i casi rilevati (confermati, sospetti o probabili) e i casi segnalati rispetto ai casi reali dipenderà dal numero di test eseguiti e dalla trasparenza del paese nella segnalazione. La maggior parte delle stime hanno messo il numero di casi non rilevati a diversi multipli di casi rilevati.
* **Casi attivi** = Questa cifra rappresenta il numero attuale di persone rilevate e confermate infettate dal virus. Questa cifra può aumentare o diminuire, e rappresenta una metrica importante per la salute pubblica e le autorità di risposta alle emergenze quando si valutano le esigenze di ospedalizzazione rispetto alla capacità.
* **Recuperi = questa statistica è altamente imperfetta, perché la segnalazione può essere mancante, incompleta, errata, basata su definizioni diverse, o datata (o una combinazione di tutti questi) per molti governi, sia a livello locale che nazionale, a volte con differenze tra stati all'interno dello stesso paese o contee all'interno dello stesso stato. L'OMS raccomanda di seguire i criteri di [i sintomi si risolvono + 2 test negativi entro 24 ore] o [i sintomi si risolvono + altri 14 giorni], ma questa è solo una raccomandazione. In alcuni paesi, quando un paziente viene dimesso dall'ospedale viene contato come "recuperato" anche se non viene eseguito alcun test. Alcuni funzionari sanitari ora considerano chiunque sia stato diagnosticato con COVID-19 tre o più settimane fa e non è morto per essere recuperato dalla malattia. In considerazione di ciò, "Casi Attivi" e "Casi Chiusi Risultato" che dipendono entrambi dal numero di recuperi (oltre ad un conteggio accurato dei decessi e un tasso soddisfacente di rilevamento dei casi, entrambi carenti nella stragrande maggioranza dei paesi) può essere influenzato da questo difetto intrinseco per molti paesi e per il conteggio totale mondiale**.
* **Grave e Critico** = analogamente ai recuperi, è molto imperfetto, per molte delle stesse ragioni. Quando il 99% dei casi era in Cina, la cifra corrispondeva praticamente al numero di casi "gravi" segnalati dalla NHC cinese. Oggi, rappresenta per la maggior parte il numero di pazienti attualmente trattati nell'unità di terapia intensiva (ICU), se e quando questa cifra è riportata.
* **Morti totali = numero cumulativo di morti tra i casi rilevati.**

1. ***Source:*** **Our World in Data**

https://ourworldindata.org/coronavirus/country/italy

* ***Data di consultazione:***

7 Giugno

* ***Stato servizio:***

ATTIVO

* ***Descrizione della fonte:***

Progetto open source che colleziona dati da varie fonti. I dati sono forniti in **forma iconografica** e consultabili online da qualsiasi utente; sono inoltre **scaricabili** in **vari formati** ed è anche possibile partecipare al progetto di collezione se lo si desidera. La piattaforma interattiva e il database fornisce una vasta gamma di dati differenti.

* *⋆⋆⋆⋆⋆* Dati aperti. Possibile modificare senza necessita del software proprietario. Dati strutturati e codificati in *formato aperto* come CSV, *collegati ad altri dataset* e raggiungibili anche da URL.
* ***Campi:***

|  |  |  |  |
| --- | --- | --- | --- |
| **column** | **source** | **category** | **description** |
| **iso\_code** | International Organization for Standardization | Others | ISO 3166-1 alpha-3 – three-letter country codes. Note that OWID-defined regions (e.g. continents like 'Europe') contain prefix 'OWID\_'. |
| **continent** | Our World in Data | Others | Continent of the geographical location |
| **location** | Our World in Data | Others | Geographical location |
| **date** | Our World in Data | Others | Date of observation |
| **total\_cases** | COVID-19 Dashboard by the WHO | Confirmed cases | Total confirmed cases of COVID-19. Counts can include probable cases, where reported. |
| **new\_cases** | COVID-19 Dashboard by the WHO | Confirmed cases | New confirmed cases of COVID-19. Counts can include probable cases, where reported. In rare cases where our source reports a negative daily change due to a data correction, we set this metric to NA. |
| **new\_cases\_smoothed** | COVID-19 Dashboard by the WHO | Confirmed cases | New confirmed cases of COVID-19 (7-day smoothed). Counts can include probable cases, where reported. |
| **total\_deaths** | COVID-19 Dashboard by the WHO | Confirmed deaths | Total deaths attributed to COVID-19. Counts can include probable deaths, where reported. |
| **new\_deaths** | COVID-19 Dashboard by the WHO | Confirmed deaths | New deaths attributed to COVID-19. Counts can include probable deaths, where reported. In rare cases where our source reports a negative daily change due to a data correction, we set this metric to NA. |
| **new\_deaths\_smoothed** | COVID-19 Dashboard by the WHO | Confirmed deaths | New deaths attributed to COVID-19 (7-day smoothed). Counts can include probable deaths, where reported. |
| **total\_cases\_per\_million** | COVID-19 Dashboard by the WHO | Confirmed cases | Total confirmed cases of COVID-19 per 1,000,000 people. Counts can include probable cases, where reported. |
| **new\_cases\_per\_million** | COVID-19 Dashboard by the WHO | Confirmed cases | New confirmed cases of COVID-19 per 1,000,000 people. Counts can include probable cases, where reported. |
| **new\_cases\_smoothed\_per\_million** | COVID-19 Dashboard by the WHO | Confirmed cases | New confirmed cases of COVID-19 (7-day smoothed) per 1,000,000 people. Counts can include probable cases, where reported. |
| **total\_deaths\_per\_million** | COVID-19 Dashboard by the WHO | Confirmed deaths | Total deaths attributed to COVID-19 per 1,000,000 people. Counts can include probable deaths, where reported. |
| **new\_deaths\_per\_million** | COVID-19 Dashboard by the WHO | Confirmed deaths | New deaths attributed to COVID-19 per 1,000,000 people. Counts can include probable deaths, where reported. |
| **new\_deaths\_smoothed\_per\_million** | COVID-19 Dashboard by the WHO | Confirmed deaths | New deaths attributed to COVID-19 (7-day smoothed) per 1,000,000 people. Counts can include probable deaths, where reported. |
| **reproduction\_rate** | Arroyo Marioli et al. (2020). <https://doi.org/10.2139/ssrn.3581633> | Reproduction rate | Real-time estimate of the effective reproduction rate (R) of COVID-19. See <https://github.com/crondonm/TrackingR/tree/main/Estimates-Database> |
| **icu\_patients** | National government reports and European CDC | Hospital & ICU | Number of COVID-19 patients in intensive care units (ICUs) on a given day |
| **icu\_patients\_per\_million** | National government reports and European CDC | Hospital & ICU | Number of COVID-19 patients in intensive care units (ICUs) on a given day per 1,000,000 people |
| **hosp\_patients** | National government reports and European CDC | Hospital & ICU | Number of COVID-19 patients in hospital on a given day |
| **hosp\_patients\_per\_million** | National government reports and European CDC | Hospital & ICU | Number of COVID-19 patients in hospital on a given day per 1,000,000 people |
| **weekly\_icu\_admissions** | National government reports and European CDC | Hospital & ICU | Number of COVID-19 patients newly admitted to intensive care units (ICUs) in a given week (reporting date and the preceeding 6 days) |
| **weekly\_icu\_admissions\_per\_million** | National government reports and European CDC | Hospital & ICU | Number of COVID-19 patients newly admitted to intensive care units (ICUs) in a given week per 1,000,000 people (reporting date and the preceeding 6 days) |
| **weekly\_hosp\_admissions** | National government reports and European CDC | Hospital & ICU | Number of COVID-19 patients newly admitted to hospitals in a given week (reporting date and the preceeding 6 days) |
| **weekly\_hosp\_admissions\_per\_million** | National government reports and European CDC | Hospital & ICU | Number of COVID-19 patients newly admitted to hospitals in a given week per 1,000,000 people (reporting date and the preceeding 6 days) |
| **total\_tests** | National government reports | Tests & positivity | Total tests for COVID-19 |
| **new\_tests** | National government reports | Tests & positivity | New tests for COVID-19 (only calculated for consecutive days) |
| **total\_tests\_per\_thousand** | National government reports | Tests & positivity | Total tests for COVID-19 per 1,000 people |
| **new\_tests\_per\_thousand** | National government reports | Tests & positivity | New tests for COVID-19 per 1,000 people |
| **new\_tests\_smoothed** | National government reports | Tests & positivity | New tests for COVID-19 (7-day smoothed). For countries that don't report testing data on a daily basis, we assume that testing changed equally on a daily basis over any periods in which no data was reported. This produces a complete series of daily figures, which is then averaged over a rolling 7-day window |
| **new\_tests\_smoothed\_per\_thousand** | National government reports | Tests & positivity | New tests for COVID-19 (7-day smoothed) per 1,000 people |
| **positive\_rate** | National government reports | Tests & positivity | The share of COVID-19 tests that are positive, given as a rolling 7-day average (this is the inverse of tests\_per\_case) |
| **tests\_per\_case** | National government reports | Tests & positivity | Tests conducted per new confirmed case of COVID-19, given as a rolling 7-day average (this is the inverse of positive\_rate) |
| **tests\_units** | National government reports | Tests & positivity | Units used by the location to report its testing data. A country file can't contain mixed units. All metrics concerning testing data use the specified test unit. Valid units are 'people tested' (number of people tested), 'tests performed' (number of tests performed. a single person can be tested more than once in a given day) and 'samples tested' (number of samples tested. In some cases, more than one sample may be required to perform a given test.) |
| **total\_vaccinations** | National government reports | Vaccinations | Total number of COVID-19 vaccination doses administered |
| **people\_vaccinated** | National government reports | Vaccinations | Total number of people who received at least one vaccine dose |
| **people\_fully\_vaccinated** | National government reports | Vaccinations | Total number of people who received all doses prescribed by the initial vaccination protocol |
| **total\_boosters** | National government reports | Vaccinations | Total number of COVID-19 vaccination booster doses administered (doses administered beyond the number prescribed by the vaccination protocol) |
| **new\_vaccinations** | National government reports | Vaccinations | New COVID-19 vaccination doses administered (only calculated for consecutive days) |
| **new\_vaccinations\_smoothed** | National government reports | Vaccinations | New COVID-19 vaccination doses administered (7-day smoothed). For countries that don't report vaccination data on a daily basis, we assume that vaccination changed equally on a daily basis over any periods in which no data was reported. This produces a complete series of daily figures, which is then averaged over a rolling 7-day window |
| **total\_vaccinations\_per\_hundred** | National government reports | Vaccinations | Total number of COVID-19 vaccination doses administered per 100 people in the total population |
| **people\_vaccinated\_per\_hundred** | National government reports | Vaccinations | Total number of people who received at least one vaccine dose per 100 people in the total population |
| **people\_fully\_vaccinated\_per\_hundred** | National government reports | Vaccinations | Total number of people who received all doses prescribed by the initial vaccination protocol per 100 people in the total population |
| **total\_boosters\_per\_hundred** | National government reports | Vaccinations | Total number of COVID-19 vaccination booster doses administered per 100 people in the total population |
| **new\_vaccinations\_smoothed\_per\_million** | National government reports | Vaccinations | New COVID-19 vaccination doses administered (7-day smoothed) per 1,000,000 people in the total population |
| **new\_people\_vaccinated\_smoothed** | National government reports | Vaccinations | Daily number of people receiving their first vaccine dose (7-day smoothed) |
| **new\_people\_vaccinated\_smoothed\_per\_hundred** | National government reports | Vaccinations | Daily number of people receiving their first vaccine dose (7-day smoothed) per 100 people in the total population |
| **stringency\_index** | Oxford COVID-19 Government Response Tracker, Blavatnik School of Government | Policy responses | Government Response Stringency Index: composite measure based on 9 response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest response) |
| **population** | United Nations, Department of Economic and Social Affairs, Population Division, World Population Prospects 2019 Revision | Others | Population (latest available values). See <https://github.com/owid/covid-19-data/blob/master/scripts/input/un/population_latest.csv> for full list of sources |
| **population\_density** | World Bank World Development Indicators, sourced from Food and Agriculture Organization and World Bank estimates | Others | Number of people divided by land area, measured in square kilometers, most recent year available |
| **median\_age** | UN Population Division, World Population Prospects, 2017 Revision | Others | Median age of the population, UN projection for 2020 |
| **aged\_65\_older** | World Bank World Development Indicators based on age/sex distributions of United Nations World Population Prospects 2017 Revision | Others | Share of the population that is 65 years and older, most recent year available |
| **aged\_70\_older** | United Nations, Department of Economic and Social Affairs, Population Division (2017), World Population Prospects 2017 Revision | Others | Share of the population that is 70 years and older in 2015 |
| **gdp\_per\_capita** | World Bank World Development Indicators, source from World Bank, International Comparison Program database | Others | Gross domestic product at purchasing power parity (constant 2011 international dollars), most recent year available |
| **extreme\_poverty** | World Bank World Development Indicators, sourced from World Bank Development Research Group | Others | Share of the population living in extreme poverty, most recent year available since 2010 |
| **cardiovasc\_death\_rate** | Global Burden of Disease Collaborative Network, Global Burden of Disease Study 2017 Results | Others | Death rate from cardiovascular disease in 2017 (annual number of deaths per 100,000 people) |
| **diabetes\_prevalence** | World Bank World Development Indicators, sourced from International Diabetes Federation, Diabetes Atlas | Others | Diabetes prevalence (% of population aged 20 to 79) in 2017 |
| **female\_smokers** | World Bank World Development Indicators, sourced from World Health Organization, Global Health Observatory Data Repository | Others | Share of women who smoke, most recent year available |
| **male\_smokers** | World Bank World Development Indicators, sourced from World Health Organization, Global Health Observatory Data Repository | Others | Share of men who smoke, most recent year available |
| **handwashing\_facilities** | United Nations Statistics Division | Others | Share of the population with basic handwashing facilities on premises, most recent year available |
| **hospital\_beds\_per\_thousand** | OECD, Eurostat, World Bank, national government records and other sources | Others | Hospital beds per 1,000 people, most recent year available since 2010 |
| **life\_expectancy** | James C. Riley, Clio Infra, United Nations Population Division | Others | Life expectancy at birth in 2019 |
| **human\_development\_index** | United Nations Development Programme (UNDP) | Others | A composite index measuring average achievement in three basic dimensions of human development—a long and healthy life, knowledge and a decent standard of living. Values for 2019, imported from <http://hdr.undp.org/en/indicators/137506> |
| **excess\_mortality** | Human Mortality Database (2021), World Mortality Dataset (2021) | Excess mortality | Percentage difference between the reported number of weekly or monthly deaths in 2020–2021 and the projected number of deaths for the same period based on previous years. For more information, see <https://github.com/owid/covid-19-data/tree/master/public/data/excess_mortality> |
| **excess\_mortality\_cumulative** | Human Mortality Database (2021), World Mortality Dataset (2021) | Excess mortality | Percentage difference between the cumulative number of deaths since 1 January 2020 and the cumulative projected deaths for the same period based on previous years. For more information, see <https://github.com/owid/covid-19-data/tree/master/public/data/excess_mortality> |
| **excess\_mortality\_cumulative\_absolute** | Human Mortality Database (2021), World Mortality Dataset (2021) | Excess mortality | Cumulative difference between the reported number of deaths since 1 January 2020 and the projected number of deaths for the same period based on previous years. For more information, see <https://github.com/owid/covid-19-data/tree/master/public/data/excess_mortality> |
| **excess\_mortality\_cumulative\_per\_million** | Human Mortality Database (2021), World Mortality Dataset (2021) | Excess mortality | Cumulative difference between the reported number of deaths since 1 January 2020 and the projected number of deaths for the same period based on previous years, per million people. For more information, see <https://github.com/owid/covid-19-data/tree/master/public/data/excess_mortality> |

1. ***Source:*** **Statista da IlSole24Ore**

https://www.statista.com/statistics/1102808/coronavirus-cases-by-status-italy/

* ***Data di consultazione:***

*7 Giugno 13.00 PST*

* ***Stato servizio:***

*ULTIMO DATO 25 MAGGIO 2023*

* ***Descrizione della fonte:***

Sito privato che fornisce dati dopo **registrazione** al sito e preclude lo scaricamento di alcuni formati solo previo **abbonamento**. I dati sono però totalmente consultabili online tramite infografiche filtrabili. Con licenza abbastanza limitante i dati sono infatti ripubblicabili solo da utenti “PRO”.

Dati aggiornati mensilmente fino ad oggi.

* *⋆⋆* **Dati strutturati codificati in proprietario** per i quali serve un programma specifico o una chiave di lettura per essere usati. (ex. file Excel). Inoltre, necessaria la registrazione al sito per l’accesso ai dati.
* ***Campi:***
* Number of COVID-19 patients hospitalized in Italy as of May 2023

|  |
| --- |
| * Deceased |
| * Recovered/discharged from hospital |
| * Active cases |
| * Total cases |

**TABELLA DI CONFRONTO**

**Date:** May 2023 (mostly 5th of May but it may vary)

|  |  |  |  |
| --- | --- | --- | --- |
| **FONTI - Chiave primaria** | **PERSONE POSITIVE** | **PERSONE OSPEDALIZZATE** | **DECEDUTI** |
| WHO | 3580 | 1542 | 31 |
| WORLDOMETERS | 3416 | 2852 | 32 |
| TESSy | Imprecise | imprecise | imprecise |
| OWD | 2974 | 1301 | 23.71 |
| STATISTA | 3417 | 2387 | 32 |

**PARTE 2**

1. La chiave primaria è tendenzialmente il codice paese/regione. Nel particolare caso della nostra tabella di confronto la chiave diventa la fonte di origine.
2. I dati risultano essere per lo più regolari sulle varie fonti anche se in alcuni casi non totalmente; potrebbe essere dovuto al fatto che non tutte le fonti prendono i dati a cadenza giornaliera bensì alcuni a cadenza settimanale e spesso non nelle stesse giornate. Alcuni addirittura a livello di report mensile.
3. **Data engineer 🡪** Progetta e sviluppa la pipeline per riuscire a procurarsi e poi stoccare i vari dati dai relativi report.

**Data Analyst 🡪** Prende visione dei dati collezionati e ne sviluppa i relativi grafici. Trova eventuali stranezze presenti magari tra le diverse fonti. Interpreta i dati.

**Data Scientist 🡪** Interviene per riuscire a sviluppare ipotesi basate su modelli scientifico matematici e per fare in mondo che le informazioni fornite siano accurate. Con modelli predittivi anche per la situazione futura.

**Data journalist** 🡪 In questa situazione diciamo il ruolo del data journalist diventa fondamentale in quanto parte della popolazione potrebbe non essere completamente formata nel campo e quindi essenziale spiegare i vari dati trovati, magari comuni con maggior numero di contagi etc.

4) I metadati di questa esercitazione sono infiniti vista anche la notevole quantità di dati presenti e la collezione da varie fonti.

Questo un esempio di una delle fonti presenti:

https://github.com/pcm-dpc/COVID-19/tree/master/metadata