**COVID-19**

Sommario

[Raccolta dati 2](#_Toc202014091)

[Analisi 3](#_Toc202014092)

[Fonte 6](#_Toc202014093)

# Raccolta dati

Per effettuare l’analisi sulla situazione sanitaria COVID-19, abbiamo principalmente consultato varie fonti, tra cui, Ministero della Salute, Protezione Civile, Lab24. Ogni sito ci ha riportato ad una fonte primaria (GitHub) sulla quale sono stati caricati i risultati di monitoraggio di tale infezione nella cartella dati-regioni della directory pcm-dpc/COVID-19.

Nella cartella di GitHub risultavano varie informazioni, due cartelle in particolare, sono state suddivise per Provincia e Regione, da quest’ultima abbiamo esportato i dati in CSV da data 2020 al 2024.

Per quanto riguarda l’andamento delle vaccinazioni nelle regioni, risulta anch’esso un report dei dati raccolti giornalmente in un Open Data nella Directory di GitHub covid-19opendata-vaccini in cui sono stati caricati i seguenti documenti: “vaccini\_summary\_latest” in CSV, aggiornato al 20 novembre 2023, “somministrazioni-vaccini-latest” in CSV, aggiornato al 17 novembre 2023, e “somministrazioni-vaccini-summary-latest” in CSV, aggiornato al 20 novembre 2023.

Come Metadati che si trovano per gli open data sulla situazione nazionale sono:

I dati vengono in principio raccolti da ogni regione che dovranno passarli su una piattaforma dell’Istituto Superiore della Sanità (controllo dati-warning).

Successivamente controllati dal Ministero della Salute il quale verifica e invia i dati al DPC (Dipartimento della Protezione Civile), il quale controlla la qualità dei dati ed elabora i dataset per la pubblicazione finale su GitHub. L’ultimo aggiornamento è del 10 gennaio 2025.

Come Metadati che si trovano per gli open data sui vaccini:

Authors / Copyright - 2021 (c) Commissario straordinario per l'emergenza Covid-19 - Presidenza del Consiglio dei Ministri.

Il mantenimento/aggiornamento è riservato al -datateam-opendata- utenza di appartenenza alla Struttura Commissariale per l'Emergenza Covid-19.

A partire dal 18/01/2021 i dati sono in fase di consolidamento da parte delle regioni e potranno subire alcune variazioni.

Tutte e due le analisi, sia sulla situazione nazionale rispetto al COVID-19 e sia sul monitoraggio dei vaccini per regioni, sono protette con la licenza CC-BY-4.0

BY: impone l’indicazione della paternità, fornire un link alla licenza e indicare se sono state effettuate delle modifiche.

[Dalla licenza denota che: è possibile di:

1. Condividere — riprodurre, distribuire, comunicare al pubblico, esporre in pubblico, rappresentare, eseguire e recitare questo materiale con qualsiasi mezzo e formato per qualsiasi fine, anche commerciale.
2. Modificare — remixare, trasformare il materiale e basarti su di esso per le tue opere per qualsiasi fine, anche commerciale.
3. Il licenziante non può revocare questi diritti fintanto che tu rispetti i termini della licenza.

Alle seguenti condizioni:

1. Attribuzione — Devi riconoscere una menzione di paternità adeguata, fornire un link alla licenza e indicare se sono state effettuate delle modifiche. Puoi fare ciò in qualsiasi maniera ragionevole possibile, ma non con modalità tali da suggerire che il licenziante avalli te o il tuo utilizzo del materiale.
2. Divieto di restrizioni aggiuntive — Non puoi applicare termini legali o misure tecnologiche che impongano ad altri soggetti dei vincoli giuridici su quanto la licenza consente loro di fare.

Note:

Non sei tenuto a rispettare i termini della licenza per quelle componenti del materiale che siano in pubblico dominio o nei casi in cui il tuo utilizzo sia consentito da una eccezione o limitazione prevista dalla legge.

Non sono fornite garanzie. La licenza può non conferirti tutte le autorizzazioni necessarie per l'utilizzo che ti prefiggi. Ad esempio, diritti di terzi come i diritti all’immagine, alla riservatezza e i diritti morali potrebbero restringere gli usi che ti prefiggi sul materiale.]

https://creativecommons.org/licenses/by/4.0/deed.it

# Analisi

Per l’analisi sulla situazione nazionale del COVID-19 sono state eseguite alcune riflessioni prima di passare a pulire ed elaborare il DataBase per capire il significato di ogni colonna, inoltre la Fonte, ovvero il Ministero della Salute, hanno lasciato nel documento “dati-andamento-covid19-italia.md” una descrizione di ogni campo.

Inizialmente sono stati tenuti i seguenti dati in considerazione:

* *totale\_positivi* Totale attualmente positivi (ospedalizzati + isolamento domiciliare);
* *variazione\_totale\_positivi* Variazione del totale positivi (totale\_positivi giorno corrente - totale\_positivi del giorno precedente);
* *nuovi\_positivi* Nuovi attualmente positivi (totale\_casi giorno corrente - totale\_casi giorno precedente);
* *totali\_casi* Totale casi positivi.

Successivamente è stato deciso di lasciare solo i campi: totale\_casi rinominato in “CONTAGIATI”, dimessi\_guariti in “GUARITI” e deceduti lasciato in “DECEDUTI”; le altre colonne sono state eliminate in quanto non strettamente necessarie allo studio corrente.

Sui tre campi lasciati sono state eseguite le rispettive unioni degli anni e trimestri in PowerQuery con la funzione accoda, inoltre sono state unite le province autonome di Bolzano e Trento in una unica voce “Trentino”.

DataBase sulla Situazione nazionale “Report\_COVID\_trimestri\_anni”

Si è lasciato solo la colonna totali casi come Contagiati.

A seguire si analizza il significato del database originale.

**Totali Positivi** 🡪 totale delle 2 colonne precedenti (ospedalizzati (terapia intensiva e ricoverati), isolamento domiciliare).

**Variazione** 🡪 Totali\_positivi - (totali ospedalizzati e isolamento\_domiciliare) del gg prima.

**Totali** **casi** 🡪 Totali\_positivi + dimessi\_guariti + totali\_deceduti

Non sono stati presi in considerazione i nuovi positivi, in quanto:

Nuovi\_positivi 🡪 n° positivi di quel giorno

Variazione\_totale\_positivi 🡪 l’operazione che si fa: prendendo il valore di nuovi\_positivi del gg meno la somma complessiva di dimessi\_guariti e deceduti del giorno prima.

I dati dei dimessi\_guariti e deceduti 🡪 sono dati cumulativi

Esempio: 28 Febbraio 2020

Nuovi\_positivi: 128 persone

Se analizziamo anche il dato Variazione\_totale\_positivi ci rendiamo conto che non è altro l’operazione della differenza delle 128 persone positive di quel giorno, meno, la somma tra i guariti e deceduti di quel giorno, che risulta essere 3, perché risultano solo 3 deceduti.

Esempio: 27 Gennaio 2021

Nuovi\_positivi: 2293 persone

Variazione\_totale\_positivi ci rendiamo conto che non è altro l’operazione della differenza delle 2293 persone positive di quel giorno (nuovi\_positivi), meno, la differenza tra i guariti (2161) e deceduti (62) di quel giorno, che risulta essere 70. Calcolo: [2293 – (2161 - 62)].

I DataBase sulle Vaccinazioni sono stati raggruppati per: Regione, trimestre, e per anno.

Per unire i due DataBase, ovvero sulla situazione nazionale e sull’andamento delle vaccinazioni, si è usato lo strumento di PowerQuery merge di query, prendendo le cartelle excel “Report\_COVID\_trimestri\_anni” e “vaccini\_summary\_latest”, eseguendo il merge left outer prendendo come colonne in comune -trimestre e regione-.

“somministrazioni-vaccini-summary-latest” (con l’aggiornamento più recente e i dati fino a fine settembre 2023) dal quale sono stati presi solo due campi, le “d1” nonché chi ha eseguito la prima dose -e in seguito può aver eseguito più dosi- e i “dpi” ovvero chi ha eseguito il vaccino entro i 3-6 mesi dal contagio da COVID-19 e che concludono il ciclo vaccinale con quella unica dose.

Le successive dosi indicano un proseguire del ciclo vaccinale, dunque il campo relativo ai Vaccini comincia dal quarto trimestre del 2020 e si arresta al terzo trimestre del 2023.

In seguito, si è creata una colonna della somma dei due campi per avere il risultato dei totali VACCINATI, richiamando il foglio come “Record\_vaccinati\_per\_regioni”.

(Analizzando i Data a disposizione per comprendere quale fosse il DataBase migliore da prendere in esame, si sono comparati i due CSV “somministrazioni-vaccini-summary-latest” e “somministrazioni-vaccini-latest” dal quale ci si accorge che sono presenti alcuni dati leggermente discrepanti, potrebbe essere risultato di un riconteggio e ulteriori controlli come enunciato nel capitolo “Raccolta dati”.

Immagine che contiene testo, schermata, menu, numero

Il contenuto generato dall'IA potrebbe non essere corretto.

Infine, come detto inizialmente, si è preso il DataBase più recente.)

Avendo il Dataset comprensivo di: Contagiati, Guariti, Deceduti e Vaccinati. Si è proceduto a creare un primo foglio dal nome MASCHERA, nella quale è stata inserito il menu a tendina che dà la possibilità di selezionare la regione e di reindirizzare i rispettivi valori dei totali Contagiati, Deceduti e Guariti usando la funzione cerca.x e usando come periodo solo il quarto trimestre dell’ultimo anno preso in esame (2024).

Nello stesso foglio si è aggiunta una tabella nella quale si osserva in ordine crescente le regioni dalla meno contagiata, alla più contagiata. Aggiungendo due grafici nel quale si osserva più chiaramente la situazione finale dell’epidemia nel nostro Paese.

Si nota che: la regione maggiormente contagiata è stata la Lombardia con 4.391.455 contagiati alla fine del 2024, in successione troviamo le seguenti regioni: del Veneto, Campania, Lazio, Emilia-Romagna, Sicilia, Piemonte, Puglia e Toscana superiori al milione; le restanti regioni rimangono invece con un numero inferiore, per finire con la Valle d’Aosta con 52.784 contagiati.

Sono state create tre ulteriori tabelle Pivot con i relativi grafici per analizzare l’andamento dei Contagiati, Guariti e Deceduti.

dove si è posizionato le regioni nelle righe e gli anni nelle colonne, inoltre filtrato per il quarto trimestre per vedere l’andamento fino al quinto anno; in aggiunta si può vedere l’andamento per singola regione scegliendola dal filtro. La stessa cosa è stata fatta per i guariti e vaccinati.

Per il calcolo del rapporto tra vaccinati e deceduti

Su PowerQuery si è preso in esame il DataBase “Report\_COVID\_trimestri\_anni” iniziale, si è proseguito prendendo in esame solo il terzo trimestre del 2023 tramite i filtri, si è poi mergiato con il DataBase “Record\_vaccinati\_per\_regioni” per regione, e successivamente eliminato i campi superflui lasciando come campi: ID\_regione, Regione, Vaccinati, Deceduti.

Si è creata una nuova colonna “Tasso\_mortalità\_vaccinati” eseguendo il calcolo Vaccinati/Deceduti.

Si osserva che all’ultimo aggiornamento della fine del 2023, il tasso di mortalità rispetto ai vaccini rimane sotto lo 0%, le regioni con tasso minore sono Sardegna e Basilicata al 0,22% si osserva che il Friuli ha il tasso più alto con lo 0,62%.

# Fonte

* Ministero della Salute

<https://www.salute.gov.it/portale/nuovocoronavirus/menuContenutoNuovoCoronavirus.jsp?lingua=italiano&area=nuovoCoronavirus&menu=reportCampagna>

* Protezione civile - Aggiornato al 2022

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

* COVID Situazione Italia - In fase di aggiornamento

<https://opendatamds.maps.arcgis.com/apps/dashboards/0f1c9a02467b45a7b4ca12d8ba296596>

* Caratteristiche dei pazienti deceduti positivo all’infezione da SASRS-CoV-2 in Italia

<https://www.epicentro.iss.it/coronavirus/sars-cov-2-decessi-italia>

* Coronavirus in Italia, i dati e la mappa. - Aggiornato al 10 Gennaio 2025

<https://lab24.ilsole24ore.com/coronavirus/?refresh_ce>

* Ministero della Salute

<https://github.com/italia/covid19-opendata-vaccini>