

Pratica W2D1

Tutte le fonti citate riguardano il monitoraggio della situazione Covid19 dall'inizio delle rilevazioni fino ad oggi, data la pressoché inesistenza di dataset riferiti al periodo indicato nella consegna, ma che ovviamente sono stati inclusi nei dati di monitoraggio fino ad oggi. Le fonti selezionate in una o più rappresentazioni grafiche indicano e seguono le indicazioni di ricerca ("POSITIVI COVID", "RICOVERATI", "DECEDUTI"), inoltre alcune sono direttamente archivi riferiti ai file csv dei monitoraggi (Mondiale, Italiano). Le fonti selezionate presentano derivati o copie del database da cui si assumono le informazioni, ovvero:

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

Licenza (comune tra le fonti) = CC-BY-4.0

- Fonti

- 1) https://lab24.ilsole24ore.com/coronavirus/?refresh_ce

Descrizione:

La fonte citata è un sito internet che presenta varie illustrazioni grafiche che presentano la possibilità di variare gli archi temporali di visualizzazione.

Classificazione 1 stella

- 2) <https://www.dati.gov.it/view-dataset/dataset?id=cb732617-c72e-4b38-b843-752422f4bdd6>

Descrizione:

In questa fonte ci si riferisce con un collegamento al dataset nazionale di monitoraggio covid, in sono contenuti tutti i monitoraggi giornalieri con report per regione e/o provincia in formato .csv.

| | | | | | | | | | | | | | |
|------------------------------------|----|---------------------|-------|----------------|-----------------------|-------------|--------------------|------------------------|-------------------|----------------------|------------------------|-----------------|----------------------------|
| dpc-covid19-ita-regioni-2020022... | 1 | data | stato | codice_regione | denominazione_regione | lat | long | ricoverati_con_sintomi | terapia_intensiva | totale_ospedalizzati | isolamento_domiciliare | totale_positivi | variazione_totale_positivi |
| dpc-covid19-ita-regioni-2020022... | 2 | 2020-03-05T17:00:00 | ITA | 13 | Abruzzo | 42.35122196 | 13.39843823 | 8 | 0 | 8 | 0 | 8 | 1 |
| dpc-covid19-ita-regioni-2020030... | 3 | 2020-03-05T17:00:00 | ITA | 17 | Basilicata | 40.63947052 | 15.80514834 | 0 | 0 | 0 | 1 | 1 | 0 |
| dpc-covid19-ita-regioni-2020030... | 4 | 2020-03-05T17:00:00 | ITA | 18 | Calabria | 38.90597598 | 16.59440194 | 1 | 0 | 1 | 1 | 2 | 1 |
| dpc-covid19-ita-regioni-2020030... | 5 | 2020-03-05T17:00:00 | ITA | 15 | Campania | 40.83956555 | 14.25084984 | 12 | 0 | 12 | 33 | 45 | 14 |
| dpc-covid19-ita-regioni-2020030... | 6 | 2020-03-05T17:00:00 | ITA | 08 | Emilia-Romagna | 44.49436681 | 11.341720800000001 | 327 | 32 | 359 | 299 | 658 | 142 |
| dpc-covid19-ita-regioni-2020030... | 7 | 2020-03-05T17:00:00 | ITA | 06 | Friuli Venezia Giulia | 45.6494354 | 12.76813649 | 4 | 0 | 4 | 17 | 21 | 3 |
| dpc-covid19-ita-regioni-2020030... | 8 | 2020-03-05T17:00:00 | ITA | 12 | Lazio | 41.89277044 | 12.48366722 | 20 | 7 | 27 | 14 | 41 | 14 |
| dpc-covid19-ita-regioni-2020030... | 9 | 2020-03-05T17:00:00 | ITA | 07 | Liguria | 44.41149315 | 8.9326992 | 11 | 3 | 14 | 7 | 21 | 0 |
| dpc-covid19-ita-regioni-2020030... | 10 | 2020-03-05T17:00:00 | ITA | 03 | Lombardia | 45.46679409 | 9.190347404 | 1169 | 244 | 1413 | 364 | 1777 | 280 |
| dpc-covid19-ita-regioni-2020031... | 11 | 2020-03-05T17:00:00 | ITA | 11 | Marche | 43.61675973 | 13.5188753 | 57 | 19 | 76 | 44 | 120 | 40 |
| dpc-covid19-ita-regioni-2020031... | 12 | 2020-03-05T17:00:00 | ITA | 14 | Molise | 41.55774734 | 14.65916051 | 4 | 0 | 4 | 3 | 7 | 4 |
| dpc-covid19-ita-regioni-2020031... | 13 | 2020-03-05T17:00:00 | ITA | 21 | P.A. Bolzano | 46.49933453 | 11.35662422 | 1 | 0 | 1 | 0 | 1 | 0 |
| dpc-covid19-ita-regioni-2020031... | 14 | 2020-03-05T17:00:00 | ITA | 22 | P.A. Trento | 46.06893511 | 11.12123097 | 2 | 0 | 2 | 5 | 7 | 2 |
| dpc-covid19-ita-regioni-2020031... | 15 | 2020-03-05T17:00:00 | ITA | 01 | Piemonte | 45.0732745 | 7.680687483 | 43 | 17 | 60 | 46 | 106 | 24 |
| dpc-covid19-ita-regioni-2020031... | 16 | 2020-03-05T17:00:00 | ITA | 16 | Puglia | 41.12559576 | 16.86736889 | 5 | 1 | 6 | 6 | 12 | 5 |
| dpc-covid19-ita-regioni-2020031... | 17 | 2020-03-05T17:00:00 | ITA | 20 | Sardegna | 39.21531192 | 9.110616306 | 2 | 0 | 2 | 0 | 2 | 0 |
| dpc-covid19-ita-regioni-2020031... | 18 | 2020-03-05T17:00:00 | ITA | 19 | Sicilia | 38.11569725 | 13.362356699999998 | 5 | 0 | 5 | 11 | 16 | 0 |
| dpc-covid19-ita-regioni-2020031... | 19 | 2020-03-05T17:00:00 | ITA | 09 | Toscana | 43.76923077 | 11.25588885 | 26 | 3 | 29 | 31 | 60 | 23 |

Classificazione 3 / 4 stelle

- 3) https://cerca.ministerosalute.it/search?q=covid&client=defaultPORT_front-end&proxy_stylesheet=defaultPORT_front-end&site=default_collection&output=xml_no_dtd&filter=p

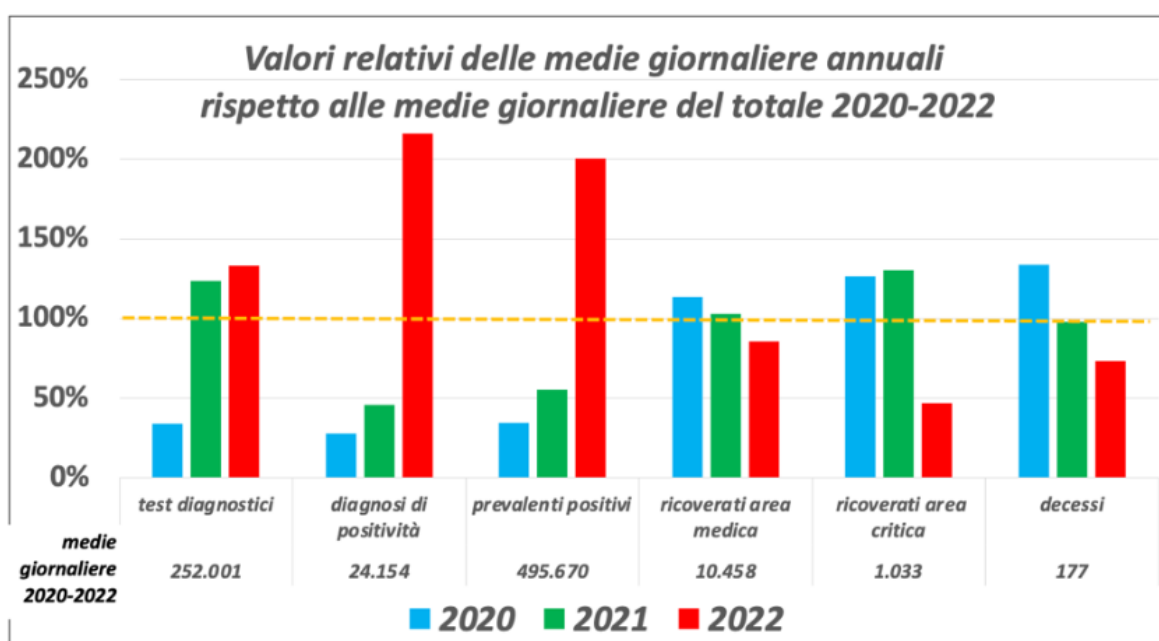
Descrizione:

Nella presente fonte sono presenti i numerosi report che partono dall'inizio della pandemia (primo monitoraggio) fino ad oggi; i report però sono fruibili unicamente nel formato .pdf, nonostante siano tabulati o espressi in rappresentazioni grafiche.

Classificazione 1 stella

- 4) <https://www.welforum.it/tre-anni-di-pandemia-covid/>

| | <i>test diagnostici</i> | <i>diagnosi di positività</i> | <i>prevalenti positivi</i> | <i>ricoverati area medica</i> | <i>ricoverati area critica</i> | <i>decessi</i> |
|---------------|-----------------------------|-----------------------------------|--------------------------------|-----------------------------------|------------------------------------|----------------|
| 2020 | 84.980 | 6.732 | 171.804 | 11.885 | 1.306 | 237 |
| 2021 | 311.066 | 11.010 | 275.140 | 10.760 | 1.349 | 173 |
| 2022 | 336.164 | 52.239 | 993.925 | 8.932 | 483 | 130 |
| totale | 252.001 | 24.154 | 495.670 | 10.458 | 1.033 | 177 |

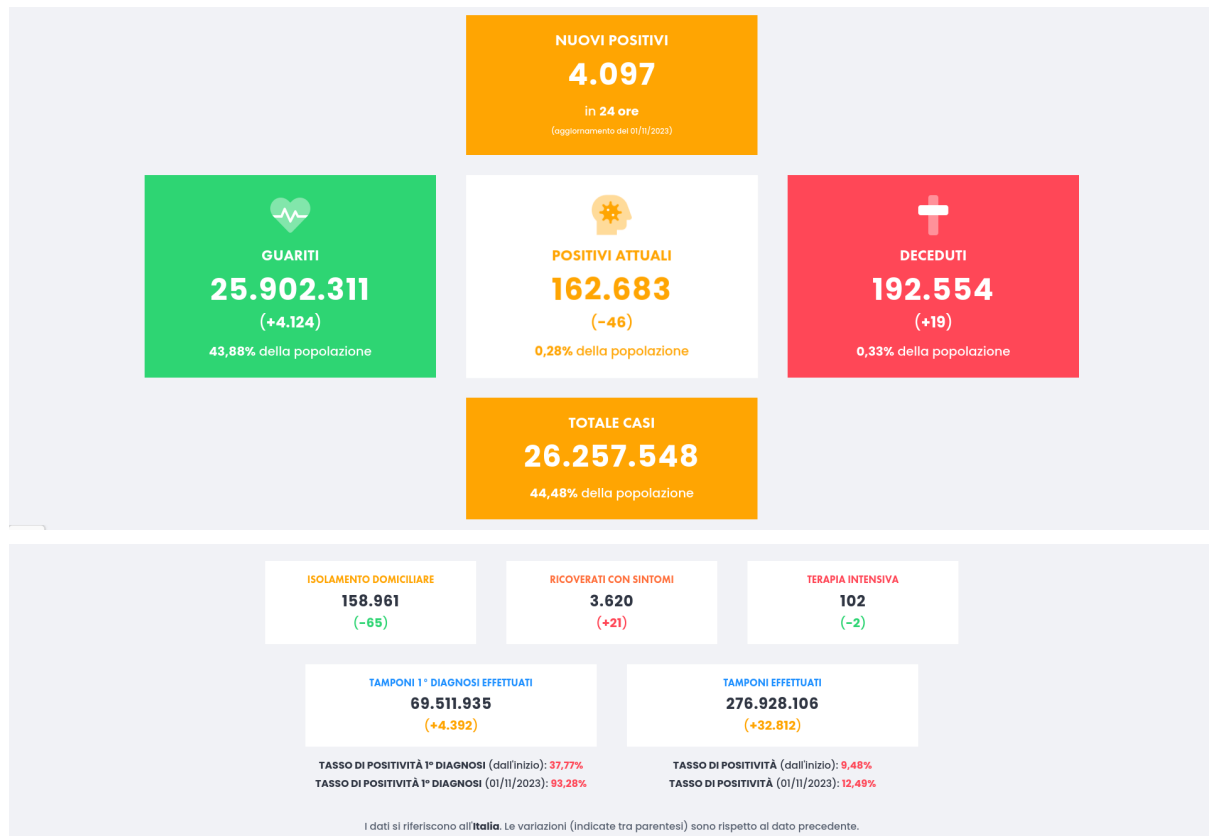


Descrizione:

La fonte selezionata è un sito internet in cui è incluso un report degli anni di pandemia.

Classificazione 1 stella

5) <https://www.curaitalia.it/>



Descrizione:

La fonte indicata e' un sito internet che riporta illustrazioni grafiche in continuo aggiornamento.

Classificazione 1 stella

- Altre fonti:

<https://covid19.who.int/region/euro/country/it>

<https://github.com/owid/covid-19-data/tree/master/public/data>

<https://www.epicentro.iss.it/coronavirus/sars-cov-2-sorveglianza-dati>

Csv per Tabella:

https://drive.google.com/file/d/1jdNciCi257ljeflSmRurMlnsFKiN_828/view?usp=sharing

| | C | D | E | F | G | H |
|----|--------------------------|-----------------|---------------------------|---------------|---------------------------|-----------------------------------|
| 1 | Nome e Cognome | Dove abiti? | Sei mai stato contagiato? | Quante volte? | Sei mai stato ricoverato? | Quanti giorni sei stato positivo? |
| 4 | Campana Fabiola | Trieste | Si | 2 | No | 37 |
| 5 | Claudio Gugliotta | Modica | Si | 2 | No | 20 |
| 6 | Daniele Manes | Carugate | Si | 1 | No | 7 |
| 7 | Fulvio Carraro | Roma | Si | 2 | No | 28 |
| 8 | Giovanni Mocchi | Bagheria | Si | 3 | No | 21 |
| 9 | Laura Scocco | Venezia | Si | 1 | No | 5 |
| 10 | Maddalena Maggiore | Venezia | Si | 1 | No | 10 |
| 11 | Marco Di Giacomo | Atri | No | 0 | No | 0 |
| 12 | Matteo Forti | Montappone | Si | 1 | No | 6 |
| 13 | Melissa Nora | Roma | Si | 1 | No | 3 |
| 14 | Nicola Tortul | Medea | Si | 1 | No | 10 |
| 15 | Noemi Trentinella | Roma | Si | 1 | No | 4 |
| 16 | Noemi von Mallasz | Gaggio Montano | Si | 1 | No | 7 |
| 17 | Raffaele Nappi | Somma Vesuviana | No | 0 | No | 0 |
| 18 | Roberta de Angelis | Milano | Si | 2 | No | 18 |
| 19 | Sara Amiri | Sicilia | No | 0 | No | 0 |
| 20 | Sara Amiri | Sicilia | No | 0 | No | 0 |
| 21 | Sara Moret | Pordenone | Si | 1 | No | 14 |
| 22 | Serena Delugas | Roma | No | 0 | No | 0 |
| 23 | Serena Pavani | Roma | No | 0 | No | 0 |
| 24 | Simone Petrini | Roma | Si | 1 | No | 15 |
| 25 | Virginia Cogoli Salducco | Estero | Si | 1 | No | 6 |
| 26 | Vitalii Bota | Varese | Si | 1 | No | 20 |
| 27 | Vito Longo | Bari | Si | 1 | No | 7 |
| 28 | Marco Franzese | Campania | Si | 1 | No | 14 |
| 29 | Luana Privitera | Torino | Si | 2 | No | 20 |
| 30 | Michael Naponiello | Campania | No | 0 | No | 0 |
| 31 | Alessandro Alvisi | Bologna | Si | 1 | No | 10 |
| 32 | Elsa Palermo | Milano | Si | 2 | No | 7 |
| 33 | Samuel David Magnin | Monza | Si | 2 | No | 5 |
| 34 | Moreno D'Urso | Roma | Si | 1 | No | 10 |
| 35 | Giovanni ripa | Torino | No | 0 | No | 0 |
| 36 | Giuseppe De Lorenzo | Saragozza | Si | 1 | No | 7 |
| 37 | Davide Zatterin | Foligno | Si | 1 | No | 5 |
| 38 | Rocco Cifone | Palagiano | Si | 1 | No | 21 |
| 39 | Francesca Maroni | Asti | No | 0 | No | 0 |

Relazione

Nel nostro report si puo' evincere che la maggior parte dei soggetti, indipendentemente dalla localita', ha contratto il covid19 (stima 80%) e per una stima del 40% si hanno avuto piu' contag, aumentando sommariamente la media dei giorni di contagio, il cui valore piu' basso e' 5 giorni; fortunatamente nessuno e' stato ricoverato.

Confronto

Purtroppo per l'impossibilita' manuale di elaborare l'ingente mole di dati del dataset nazionale, e' quasi impossibile confrontare i dati a nostra disposizione (aggiungendo anche la mancanza dati temporali del contagio), ma si puo' valutare che i nostri dati hanno seguito l'andamento nazionale.

La chiave primaria del nostro report tabulato per la situazione Covid19 e' evidenziata in giallo.

- Ruoli su dataset

Data Engineer:

La figura del Data Engineer interviene nella prima fase in quanto ricerca, identifica e crea la pipeline per la fruizione dei dati all'interno di un dataset piu' o meno "sgrezzato" da parte delle figure del Data Analyst e del Data Scientist.

In questo caso il Data Engineer si dovrebbe collocare nei dipartimenti di raccolta dei Dati del Ministero della Salute e Protezione Civile di ogni regione e degli istituti superiori di sanita'.

Data Analyst:

Il Data Analyst ottenuto/i i dataset dal Data Engineer in forma diretta o indiretta (internet), passa allo "sgrezzamento" dei dati, andando a rilevare i dati utili da quelli che non lo sono e non appena finisce procede alla fase di data visualization, in cui, tramite la formulazione di una rappresentazione grafica, riassume i dati in un grafico in cui sono espresse le "tendenze" del momento, ma, se richieste (nel caso covid, si), anche quelle passate per poter evidenziare un cambiamento o meno nel trend dei dati.

Data Journalist:

Il data journalist ha le stesse indicazioni del Data Analyst ma aggiunge un aspetto critico alla data visualization per i fini giornalistici dell'informazione da lui apportata.

Data Scientist:

Il Data Scientist, assunti i dati dai database forniti (come anche per il Data Analyst), seleziona possibilmente tutti i dati disponibili per poter effettuare delle previsioni sul breve e lungo periodo per fornire un'informazione che sia piu' possibilmente veritiera e probabile, grazie alle elaborazioni fatte da lui con modelli statistici predittivi oppure tramite Machine Learning ed IA.

- Metadati

I metadati piu' ricorrenti della ricerca sono stati:

Il numero di contagi, la sede dei contagi da una scala macroscopica (mondiale) ad una sempre piu' specifica (Italia, regioni, Città, non indicate nel documento), il numero di ricoverati, il numero di morti.