**ESERCITAZIONE DATI DEL 7-2-23**

**Fonti Utilizzate:**

* World Health Organization (WHO)
* Centro Europeo per la Prevenzione e il Controllo delle Malattie (ECDC)
* Ministero della Salute (Italia)
* Lab24 Il Sole 24ore
* Worldometer

**World Health Organization (WHO)**

<https://covid19.who.int/table?tableChartType=heat>

La fonte mostra i dati dei casi Covid-19 nel mondo. In particolare, si vanno ad analizzare quelli riferiti alla situazione presente in Italia da inizio 2023.

I campi descritti dalla fonte riguardano:

-Il totale dei casi da inizio pandemia;

-Il numero di casi degli ultimi 7 giorni (ma con la presenza di un grafico si può ricavare anche il dato riferito al mese di gennaio 2023);

-il numero di morti totali; il numero di morti negli ultimi 7 giorni (ma con la presenza di un grafico si può ricavare anche il dato riferito al mese di gennaio 2023);

-non è presente il numero di persone ricoverate ma ci sono dati riferiti al numero di persone vaccinate

La fonte utilizzata è il sito ufficiale dell’Organizzazione Mondiale della Sanità.

I dati sono scaricabili in formato .CSV

Classificazione 2

**Centro Europeo per la Prevenzione e il Controllo delle Malattie (ECDC)**

<https://www.ecdc.europa.eu/en/publications-data/data-national-14-day-notification-rate-covid-19>

<https://www.ecdc.europa.eu/en/publications-data/download-data-hospital-and-icu-admission-rates-and-current-occupancy-covid-19>

La fonte da la possibilità all’utente di poter scaricare vari formati di file di dati relativi ai casi e alle morti di Covid-19 nel panorama europeo. Formati di file disponibili: XLSX, CSV, JSON, XML.

Dai file scaricati si possono facilmente filtrare solo quelli relativi all’Italia. I link alla fonte sono due perché le relative entità sono diverse: una riguarda il numero di positivi/morti in uno specifico lasso di tempo mentre l’altro indica il numero di ricoveri per ogni 100k abitanti in un determinato lasso di tempo.

La fonte utilizzata è il sito ufficiale del Centro Europeo per la Prevenzione delle Malattie.

Data la ottima fruibilità e descrizione dei dati, e la possibilità di scaricarli e modificarli in diversi formati la classificazione della fonte è 3

**Ministero della Salute**

<https://www.salute.gov.it/portale/nuovocoronavirus/dettaglioContenutiNuovoCoronavirus.jsp?lingua=italiano&id=6001&area=nuovoCoronavirus&menu=vuoto>

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

La fonte ci permette di analizzare nel dettaglio i dati relativi alla situazione del Covid-19 in Italia.

Le entità sono disponibili sulla base di un report settimanale e per singola regione, quindi è più semplice andare a ricostruire nel dettaglio la situazione dei dati.

I dati sono molto dettagliati e specifici e ci indicano:

-il numero dei casi sintomatici notificati per mese in cui è indicata la data di inizio dei sintomi

- il numero dei casi notificati per mese con storia di ricovero in ospedale

-Il numero di casi notificati per mese con storia di trasferimento/ricovero in reparto di terapia intensiva

Non si riescono a ricavare dei dati sul numero dei morti, probabilmente perché la bassa incidenza del contagio e la copertura vaccinale hanno quasi eliminato quel dato come indicatore.

L’archivio dati non è molto open dato che questi report sono scaricabili, ma solo in formato PDF.

In ogni caso molto precisi e dettagliati.

Da segnalare anche la presenza di una sezione Open Data che rimanda ad una repository su GitHub con pubblicazioni sempre in aggiornamento e in formati .CSV

Classificazione 1

**Lab 24 Il sole 24ore**

<https://lab24.ilsole24ore.com/coronavirus/?refresh_ce=1>

Il sito ufficiale della testata giornalistica del Sole 24ore ci mostra dei dati molto dettagliati sulla situazione Covid-19 in Italia.

In particolare, ci mostrano i dati relativi al numero di casi, di morti e di ricoveri sul territorio italiano in periodi di tempo e divisioni territoriali ben definiti.

I dati sono mostrati sul sito con l’aiuto di grafici bene realizzati ma sono disponibili ad essere scaricati anche in formato .CSV

Nei file .CSV scaricabili non sono presenti dati riguardanti l’aspetto territoriale (regione o provincia).

Classificazione 2

**Worldometer**

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

<https://github.com/pcm-dpc/COVID-19/blob/master/dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale.csv#L1>

Una fonte da un sito internazionale di informazioni e statistica.

I dati che andremo a estrapolare sono relativi alla situazione del Covid-19 in Italia.

Come per il sito del Sole 24 ore anche quello di Worldometer ci dà informazioni, aiutandosi con dei grafici chiari e ben strutturati, riguardo:

-il numero dei ricoveri

-il numero degli ospedalizzati

-il numero delle persone in terapia intensive

-il numero totale dei positivi

-in numero totale dei decessi

Tuttavia, è possibile scaricare questi dati in formato .CSV.

Classificazione 2

**Esempi di tabelle CSV**

La tabella di esempio ci mostra l’andamento dei casi di Covid-19 in Italia su base giornaliera dal 24/02/2020 al 01/02/2023, quindi di tutto il periodo della pandemia.

La tabella è molto precisa e ci mostra giorno per giorno i vari dati.

Anche i campi delle varie colonne sono molto dettagliati e qui li vediamo in maniera sintetica:

-Data

-Stato

-Ricoverati\_con\_sintomi

-Terapia\_intensiva

-Totale\_ospedalizzati

-Isolamento\_domiciliare

-Totale\_positivi

-Variazione\_totale\_positivi

-Nuovi\_positivi

-Dimessi\_guariti

-Deceduti

**COVID Case\_Italia\_trend\_2022\_2023**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| data | stato | ricoverati\_con\_sintomi | terapia\_intensiva | totale\_ospedalizzati | isolamento\_domiciliare | totale\_positivi | variazione\_totale\_positivi | nuovi\_positivi | dimessi\_guariti | deceduti |
| 2020-02-24T18:00:00 | ITA | 101 | 26 | 127 | 94 | 221 | 0 | 221 | 1 | 7 |
| 2020-02-25T18:00:00 | ITA | 114 | 35 | 150 | 162 | 311 | 90 | 93 | 1 | 10 |
| 2020-02-26T18:00:00 | ITA | 128 | 36 | 164 | 221 | 385 | 74 | 78 | 3 | 12 |
| 2020-02-27T18:00:00 | ITA | 248 | 56 | 304 | 284 | 588 | 203 | 250 | 45 | 17 |
| 2020-02-28T18:00:00 | ITA | 345 | 64 | 409 | 412 | 821 | 233 | 238 | 46 | 21 |
| 2020-02-29T18:00:00 | ITA | 401 | 105 | 506 | 543 | 1049 | 228 | 240 | 50 | 29 |
| 2020-03-01T18:00:00 | ITA | 639 | 140 | 779 | 798 | 1577 | 528 | 566 | 83 | 34 |
| 2020-03-02T18:00:00 | ITA | 742 | 166 | 908 | 927 | 1835 | 258 | 342 | 149 | 52 |
| 2020-03-03T18:00:00 | ITA | 1034 | 229 | 1263 | 1000 | 2263 | 428 | 466 | 160 | 79 |
| 2020-03-04T18:00:00 | ITA | 1346 | 295 | 1641 | 1065 | 2706 | 443 | 587 | 276 | 107 |
| 2020-03-05T18:00:00 | ITA | 1790 | 351 | 2141 | 1155 | 3296 | 590 | 769 | 414 | 148 |
| 2020-03-06T18:00:00 | ITA | 2394 | 462 | 2856 | 1060 | 3916 | 620 | 778 | 523 | 197 |
| 2020-03-07T18:00:00 | ITA | 2651 | 567 | 3218 | 1843 | 5061 | 1145 | 1247 | 589 | 233 |
| 2020-03-08T18:00:00 | ITA | 3557 | 650 | 4207 | 2180 | 6387 | 1326 | 1492 | 622 | 366 |
| 2020-03-09T18:00:00 | ITA | 4316 | 733 | 5049 | 2936 | 7985 | 1598 | 1797 | 724 | 463 |
| 2020-03-10T18:00:00 | ITA | 5038 | 877 | 5915 | 2599 | 8514 | 529 | 977 | 1004 | 631 |
| 2020-03-11T17:00:00 | ITA | 5838 | 1028 | 6866 | 3724 | 10590 | 2076 | 2313 | 1045 | 827 |
| 2020-03-12T17:00:00 | ITA | 6650 | 1153 | 7803 | 5036 | 12839 | 2249 | 2651 | 1258 | 1016 |
| 2020-03-13T17:00:00 | ITA | 7426 | 1328 | 8754 | 6201 | 14955 | 2116 | 2547 | 1439 | 1266 |
| 2020-03-14T17:00:00 | ITA | 8372 | 1518 | 9890 | 7860 | 17750 | 2795 | 3497 | 1966 | 1441 |
| 2020-03-15T17:00:00 | ITA | 9663 | 1672 | 11335 | 9268 | 20603 | 2853 | 3590 | 2335 | 1809 |
| 2020-03-16T17:00:00 | ITA | 11025 | 1851 | 12876 | 10197 | 23073 | 2470 | 3233 | 2749 | 2158 |
| 2020-03-17T17:00:00 | ITA | 12894 | 2060 | 14954 | 11108 | 26062 | 2989 | 3526 | 2941 | 2503 |
| 2020-03-18T17:00:00 | ITA | 14363 | 2257 | 16620 | 12090 | 28710 | 2648 | 4207 | 4025 | 2978 |
| 2020-03-19T17:00:00 | ITA | 15757 | 2498 | 18255 | 14935 | 33190 | 4480 | 5322 | 4440 | 3405 |
| 2020-03-20T17:00:00 | ITA | 16020 | 2655 | 18675 | 19185 | 37860 | 4670 | 5986 | 5129 | 4032 |
| 2020-03-21T17:00:00 | ITA | 17708 | 2857 | 20565 | 22116 | 42681 | 4821 | 6557 | 6072 | 4825 |
| 2020-03-22T17:00:00 | ITA | 19846 | 3009 | 22855 | 23783 | 46638 | 3957 | 5560 | 7024 | 5476 |
| 2020-03-23T17:00:00 | ITA | 20692 | 3204 | 23896 | 26522 | 50418 | 3780 | 4789 | 7432 | 6077 |
| 2020-03-24T17:00:00 | ITA | 21937 | 3396 | 25333 | 28697 | 54030 | 3612 | 5249 | 8326 | 6820 |
| 2020-03-25T17:00:00 | ITA | 23112 | 3489 | 26601 | 30920 | 57521 | 3491 | 5210 | 9362 | 7503 |
| 2020-03-26T17:00:00 | ITA | 24753 | 3612 | 28365 | 33648 | 62013 | 4492 | 6153 | 10361 | 8165 |
| 2020-03-27T17:00:00 | ITA | 26029 | 3732 | 29761 | 36653 | 66414 | 4401 | 5959 | 10950 | 9134 |
| 2020-03-28T17:00:00 | ITA | 26676 | 3856 | 30532 | 39533 | 70065 | 3651 | 5974 | 12384 | 10023 |
| 2020-03-29T17:00:00 | ITA | 27386 | 3906 | 31292 | 42588 | 73880 | 3815 | 5217 | 13030 | 10779 |
| 2020-03-30T17:00:00 | ITA | 27795 | 3981 | 31776 | 43752 | 75528 | 1648 | 4050 | 14620 | 11591 |
| 2020-03-31T17:00:00 | ITA | 28192 | 4023 | 32215 | 45420 | 77635 | 2107 | 4053 | 15729 | 12428 |
| 2020-04-01T17:00:00 | ITA | 28403 | 4035 | 32438 | 48134 | 80572 | 2937 | 4782 | 16847 | 13155 |
| 2020-04-02T17:00:00 | ITA | 28540 | 4053 | 32593 | 50456 | 83049 | 2477 | 4668 | 18278 | 13915 |
| 2020-04-03T17:00:00 | ITA | 28741 | 4068 | 32809 | 52579 | 85388 | 2339 | 4585 | 19758 | 14681 |
| 2020-04-04T17:00:00 | ITA | 29010 | 3994 | 33004 | 55270 | 88274 | 2886 | 4805 | 20996 | 15362 |
| 2020-04-05T17:00:00 | ITA | 28949 | 3977 | 32926 | 58320 | 91246 | 2972 | 4316 | 21815 | 15887 |
| 2020-04-06T17:00:00 | ITA | 28976 | 3898 | 32874 | 60313 | 93187 | 1941 | 3599 | 22837 | 16523 |
| 2020-04-07T17:00:00 | ITA | 28718 | 3792 | 32510 | 61557 | 94067 | 880 | 3039 | 24392 | 17127 |
| 2020-04-08T17:00:00 | ITA | 28485 | 3693 | 32178 | 63084 | 95262 | 1195 | 3836 | 26491 | 17669 |
| 2020-04-09T17:00:00 | ITA | 28399 | 3605 | 32004 | 64873 | 96877 | 1615 | 4204 | 28470 | 18279 |
| 2020-04-10T17:00:00 | ITA | 28242 | 3497 | 31739 | 66534 | 98273 | 1396 | 3951 | 30455 | 18849 |
| 2020-04-11T17:00:00 | ITA | 28144 | 3381 | 31525 | 68744 | 100269 | 1996 | 4694 | 32534 | 19468 |
| 2020-04-12T17:00:00 | ITA | 27847 | 3343 | 31190 | 71063 | 102253 | 1984 | 4092 | 34211 | 19899 |

A seguire la tabella d’esempio della nostra classe.



Una **chiave primaria** che potrebbe definire la relazione tra le due tabelle può essere la data, in quanto ci aiuterebbe a capire in quale periodo alcuni elementi della nostra classe hanno avuto il Covid e confrontarli con la tabella generale per vedere se i periodi coincidono con la presenza di picchi a livello nazionale.

Della fase di raccolta dei dati, il **Data Engineer** potrebbe essere coinvolto nell'acquisire e integrare i dati da diverse fonti.

Il **Data Scientist** potrebbe utilizzare i dati per analizzare la situazione del COVID-19.

Il **Data Analyst** potrebbe essere coinvolto nella creazione della tabella CSV e nella visualizzazione dei dati.

I metadati che possono essere identificati in questa esercitazione potrebbero includere:

* data di raccolta dei dati
* fonte dei dati
* definizione dei campi
* modalità di raccolta dei dati
* aggiornamento frequente o meno frequente dei dati