M1 – D3 - Esercitazione 3 Open Data

# Fonti andamento Covid19 all’anno 2023

## Fonte 1 Covid19 - ISS

### Descrizione e fonte

Al seguente sito <https://www.epicentro.iss.it/coronavirus/sars-cov-2-dashboard> l’Istituto Superiore di Sanità (= fonte) ha messo a disposizione una dashboard interattiva che descrive l’andamento del covid19 in Italia, includendo un possibile collegamento ai dati mondiali e le descrizioni delle modalità di analisi.

Nella fonte citata sono messi a disposizione i dati originali scaricabili e in formato proprietario (xlsx), <https://www.epicentro.iss.it/coronavirus/open-data/covid_19-iss.xlsx>.

La licenza non è specificata.

Il dataset xlsx è composto da 9 entità (Excel sheets), che permettono di descrivere l’andamento dei casi di positività, ospedalizzazioni e decessi da covid19 come totale, per ogni regione, provincia, per sesso, età e stato clinico, dall’inizio della pandemia (febbraio 2020) ad oggi (febbraio 2023):

* Casi prelievo diagnosi

= Numero di casi per data prelievo/diagnosi

Campi:

* iss\_date: data di aggiornamento dati ISS
* DATA\_PRELIEVO\_DIAGNOSI: data prelievo o, se mancante, di diagnosi Chiave primaria
* CASI: numero casi cumulativi
* Casi inizio sintomi

= Numero di casi per data dell’inizio sintomi

Campi:

* iss\_date: data di aggiornamento dati ISS
* DATA\_INIZIO\_SINTOMI: data di comparsa sintomi Chiave primaria
* CASI: numero casi cumulativi
* Casi inizio sintomi sint

= Numero di casi per data dell’inizio sintomi indicata per gli asintomatici

Campi:

* iss\_date: data di aggiornamento dati ISS
* DATA\_INIZIO\_SINTOMI: data di comparsa sintomi Chiave primaria
* CASI\_SINT: numero casi di infezione confermata da virus SARS-CoV-2 per cui è indicata una data inizio sintomi tranne i casi dichiarati asintomatici
* Casi regioni

Numero di casi cumulativi per regione

Campi:

* iss\_date: data di aggiornamento dati ISS
* NOME\_NUTS2: nome regione/PA Chiave primaria
* CASI\_CUMULATIVI: numero casi cumulativi
* Casi province

Numero di casi cumulativi per provincia

Campi:

* iss\_date: data di aggiornamento dati ISS
* COD\_PROV: codice provincia Chiave primaria
* PROV\_names: nome provincia Chiave primaria
* CASI\_CUMULATIVI: numero casi cumulativi
* Ricoveri

Numero di ricoveri cumulativi

Campi:

* iss\_date: data di aggiornamento dati ISS
* DATARICOVERO1: data di primo ricovero Chiave primaria
* RICOVERI: numero casi ricoverati cumulativi
* Decessi

Numero casi deceduti

* iss\_date: data di aggiornamento dati ISS
* DATA\_DECESSO: data di decesso Chiave primaria
* DECESSI: numero casi deceduti
* Sesso\_età

= Numero totale casi e decessi dall'inizio dell'epidemia per sesso e per fascia d'età

Campi:

iss\_date: data di aggiornamento dati ISS

* SESSO: sesso casi
* AGE\_GROUP: fascia d'età casi
* DECEDUTI: numero casi deceduti
* CASI\_CUMULATIVI: numero casi cumulativi

Chiave primaria = combinazione di SESSO + AGE\_GROUP

* Stato\_clinico

= Numero totale casi con malattia ancora in corso e con esito finale ancora non definito dall'inizio dell'epidemia, differenziati per sintomatologia più recente, per sesso e per fascia d'età

Campi:

* iss\_date: data di aggiornamento dati ISS
* SESSO: sesso casi
* AGE\_GROUP: fascia d'età casi
* STATO\_CLINICO: diviso nelle seguenti categorie

-ASINTOMATICO

-LIEVE

-PAUCI-SINTOMATICO

-SEVERO

-CRITICO

* CASI: numero casi

Chiave primaria = combinazione di SESSO + AGE\_GROUP + STATO\_CLINICO

### Metadati

All’interno dell’entità sono anche presenti i seguenti metadati descrittivi (organizzati in due differenti schede):

* Nota metodologica

Include la descrizione della metodologia e vari link che rimandano ad un Open Data della Protezione civile classificazione 1 stella

* Contenuto

Include le descrizioni delle entità e dei campi di ognuna

I metadati del file presenti sono:

* Size = 195KB
* Data e ora di creazione = 11/24/2020 10:52 AM
* Data di ultima modifica = oggi (10/02/2023) 9:54 AM
* Autore = Matteo
* Ultima modifica da = Elisabetta De Vito-Francesco
* Open File Location = file path

### Valori dei campi richiesti (estrapolati)

|  |  |
| --- | --- |
| Nr positivi  (Dati cumulativi al 2023 2 febbraio) | 25700453 |
| Nr ricoverati  (Dati cumulativi al 2023 31 gennaio) | 740095 |
| Nr ricoverati  (Giornalieri 31 gennaio 2023) | 170 |
| Nr decessi  (Dati cumulativi al 2023 3 febbraio) | 185882 |

### Classificazione del dataset xlsx “covid\_19-iss”

**Stelle 2**

* dati aperti scaricabili (ri)condivisibili + modificabili + formato proprietario (xlsx) + ma eventualmente convertibili

## Fonte 2 Covid19 – Ministero della Salute - salute.gov.it

### Descrizione e fonte

Il Ministero della Salute ha messo a disposizione sul sito salute.gov.it, in particolare su <https://www.salute.gov.it/portale/nuovocoronavirus/dettaglioNotizieNuovoCoronavirus.jsp?lingua=italiano&menu=notizie&p=dalministero&id=6150>, i monitoraggi e bollettini settimanali (correnti e archiviati) sull’andamento del covid19. Il Ministero della Salute mette inoltre e disposizione una dashboard interattiva, che permette di consultare i dati dell’andamento in modo interattivo, ad esempio è possibile selezionare l’arco di tempo e/o la regione voluti <https://opendatamds.maps.arcgis.com/apps/dashboards/0f1c9a02467b45a7b4ca12d8ba296596>.

Nella stessa fonte, il Ministero della Salute permette inoltre di consultare e scaricare i dati usati per le analisi di monitoraggio del covid19, nella seguente repository GitHub pubblica <https://github.com/pcm-dpc/COVID-19> (creata dalla protezione civile). All’interno del progetto GitHub pubblico si trovano cartelle contenenti dati consultabili, scaricabili e in formato aperto, CSV e JSON. Alcuni esempi di cartelle presenti sono: andamento nazionale, regionale, provinciale, dati statistici di riferimento.

La licenza dell’intera repository è CC-BY-4.0.

Come esempio, è stato preso in considerazione l’entità che descrive l’andamento regionale covid19 al giorno 19 novembre 2022 (ultimo file csv disponibile), <https://github.com/pcm-dpc/COVID-19/blob/master/dati-regioni/dpc-covid19-ita-regioni-20221119.csv> denominato “dpc-covid19-ita-regioni-20221119.csv”. In questa tabella, i campi/attributi sono 30:

* Data
* Stato
* Codice\_regione (= Chiave primaria)
* Denominazione\_regione (= Chiave primaria)
* Lat (=latitudine)
* Long (=longitudine)
* Ricoverati\_con\_sintomi
* Terapia\_intensiva
* Totale\_ospedalizzati (= somma di ricoverati\_con\_sintomi + terapia\_intensiva)
* Isolamento\_domiciliare
* Totale\_positivi
* Variazione\_totale\_positivi
* Nuovi\_positivi
* Dimessi\_guariti
* Deceduti
* Casi\_da\_sospetto\_diagnostico
* Casi\_da\_screening
* Totale\_casi
* Tamponi
* Casi\_testati
* Note
* Ingressi\_terapia\_intensiva
* Note\_test
* Note\_casi
* Totale\_positivi\_test\_molecolare
* Totale\_positivi\_test\_antigenico\_rapido
* Tamponi\_test\_molecolare
* Tamponi\_test\_antigenico\_rapido
* Codice\_nuts\_1 (=area geografica)
* Codice\_nuts\_2 (=area geografica)

### Metadati

I metadati del file presenti sono:

* Size = 5.31 KB
* Data e ora di creazione = 10/02/2023 4:07 PM
* Data di ultima modifica = 10/02/2023 4:07 PM
* Autore = Elisabetta De Vito-Francesco
* Ultima modifica da = Elisabetta De Vito-Francesco
* Open File Location = file path

I metadati della repository della protezione civile sono presenti all’interno di una cartella che contiene informazioni in formato xml riguardo:

* Covdi19 aree comuni
* Covdi19 aree nuove IT
* Covdi19 aree
* Covid19 ita contratti DPC pagamenti
* Covid19 ita andamento nazionale
* Covid19 monitoraggio

Il file README presente nella repository contiene metadati riguardo la descrizione e struttura della repository stessa, e riguardo la gestione, analisi e il formato dei dati.

### Valori dei campi richiesti (estrapolati)

|  |  |
| --- | --- |
| Nr positivi  (Dati cumulativi al 2022, 19 novembre) | 24099206 |
| Nr ospedalizzati  (Dati cumulativi al 2022, 19 novembre) | 7350 |
| Nr decessi  (Dati cumulativi al 2023, 3 febbraio) | 180679 |

### Classificazione “dpc-covid19-ita-regioni-20221119.csv”

**Stelle 4**

* dati aperti consultabili + scaricabili + modificabili + formato aperto (CSV) + URL

## Fonte 3 – IlSole24h Lab24

### Descrizione e fonte

Il Sole 24 Ore ha creato una dashboard che descrive l’andamento covid19 e l’andamento delle vaccinazioni in Italia (con possibile visualizzazione di dati mondiali), al seguente URL <https://lab24.ilsole24ore.com/coronavirus/?refresh_ce=1>. Il sito rimanda alle fonti usate: protezione civile, ISS, Ministero della salute, Areu, John Hopkins University, Our World in Data, Gisaid.

In nota ai grafici visibili, il sito rende disponibili e scaricabili dati utilizzati in formato aperto CSV o JSON.

L’entità “trend\_giornaliero.csv” descrive l’andamento dei casi di positività (includendo numero di tamponi effettuati), ospedalizzazioni e decessi da covid19 come totale, per ogni regione, provincia, per sesso, età e stato clinico, dall’inizio della pandemia (febbraio 2020) ad oggi (febbraio 2023). L’entità è composta da 34 campi:

* data (Chiave primaria)
* ricoverati
* terapia\_intensiva
* ospedalizzati (= ricoverati + terapia\_intensiva, totale giornaliero)
* isolamento\_domiciliare
* totale\_positivi (= numero totale di casi positivi giornaliero)
* variazione\_totale\_positivi
* nuovi\_positivi (= numero di nuovi positivi giornaliero)
* guariti
* deceduti
* totale\_casi (= numero casi positivi cumulativi)
* tamponi
* casi\_testati
* ingressi\_terapia\_intensiva
* totale\_positivi\_test\_molecolare
* totale\_positivi\_test\_antigenico\_rapido
* tamponi\_test\_molecolare
* tamponi\_test\_antigenico\_rapido
* diff\_ricoverati
* diff\_terapia\_intensiva
* diff\_ospedalizzati
* diff\_isolamento\_domiciliare
* diff\_totale\_positivi
* diff\_variazione\_totale\_positivi
* diff\_nuovi\_positivi
* diff\_guariti
* diff\_deceduti
* diff\_totale\_casi
* diff\_tamponi
* diff\_totale\_positivi\_test\_molecolare
* diff\_totale\_positivi\_test\_antigenico\_rapido
* diff\_tamponi\_test\_molecolare
* diff\_tamponi\_test\_antigenico\_rapido
* diff\_casi\_testati

### Metadati

I metadati dell’entità “trend\_giornaliero.csv” sono:

* Size = 208 KB
* Data e ora di creazione = 8/02/2023 8:54 PM
* Data di ultima modifica = 10/02/2023 3:47 PM

I metadati della dashboard sono:

* Fonti
* Coordinamento
* Design Director
* Design
* Data Visualization
* Sviluppo

### Campi richiesti

|  |  |
| --- | --- |
| Nr positivi  (Dati cumulativi al 2023, 2 febbraio) | 25488166 |
| Nr ospedalizzati  (Dati cumulativi al 2023, 2 febbraio) | 12179408 |
| Nr ospedalizzati  (Dati totali giornalieri al 2023, 2 febbraio, includendo dimissioni) | 3891 |
| Nr ospedalizzati  (Dati totali giornalieri al 2023, 2 febbraio, escludendo dimissioni, contando ospedalizzazioni) | 121448 |
| Nr decessi  (Dati cumulativi al 2023, 2 febbraio) | 187272 |

### Classificazione entità “Trend giornaliero”,

**Stelle 4**

* Dati aperti consultabili + scaricabili + modificabili + formato aperto (CSV, JSON) + URL

## Fonte 4 covid19 – Our World in data

### Descrizione e fonte

La fonte Our World in Data mette a disposizione delle dashboards e i dati utilizzati che descrivono l’andamento del covid19 e delle vaccinazioni dall’inizio della pandemia, al seguente url <https://ourworldindata.org/coronavirus>.

Licenza dei dati:

“Reuse this work freely

All visualizations, data, and code produced by Our World in Data are completely open access under the Creative Commons BY license. You have the permission to use, distribute, and reproduce these in any medium, provided the source and authors are credited.

The data produced by third parties and made available by Our World in Data is subject to the license terms from the original third-party authors. We will always indicate the original source of the data in our documentation, so you should always check the license of any such third-party data before use and redistribution.

All of our charts can be embedded in any site.”

L’entità “owid-covid-data.csv”, consultabile e scaricabile in formato aperto <https://covid.ourworldindata.org/data/owid-covid-data.csv>, descrive l’andamento dei casi di positività (includendo numero di tamponi effettuati), ospedalizzazioni e decessi da covid19 come totale, per ogni regione, provincia, per sesso, età, stato clinico, caratteristiche sociali, numero di vaccinazioni, dall’inizio della pandemia (febbraio 2020) ad oggi (febbraio 2023), di tutti i paesi del mondo (con disponibilità di dati). L’entità è composta da 67 campi:

* iso\_code
* continent
* location
* date
* total\_cases
* new\_cases
* new\_cases\_smoothed
* total\_deaths
* new\_deaths
* new\_deaths\_smoothed
* total\_cases\_per\_million
* new\_cases\_per\_million
* new\_cases\_smoothed\_per\_million
* total\_deaths\_per\_million
* new\_deaths\_per\_million
* new\_deaths\_smoothed\_per\_million
* reproduction\_rate
* icu\_patients
* icu\_patients\_per\_million
* hosp\_patients
* hosp\_patients\_per\_million
* weekly\_icu\_admissions
* weekly\_icu\_admissions\_per\_million
* weekly\_hosp\_admissions
* weekly\_hosp\_admissions\_per\_million
* total\_tests
* new\_tests
* total\_tests\_per\_thousand
* new\_tests\_per\_thousand
* new\_tests\_smoothed
* new\_tests\_smoothed\_per\_thousand
* positive\_rate
* tests\_per\_case
* tests\_units
* total\_vaccinations
* people\_vaccinated
* people\_fully\_vaccinated
* total\_boosters
* new\_vaccinations
* new\_vaccinations\_smoothed
* total\_vaccinations\_per\_hundred
* people\_vaccinated\_per\_hundred
* people\_fully\_vaccinated\_per\_hundred
* total\_boosters\_per\_hundred
* new\_vaccinations\_smoothed\_per\_million
* new\_people\_vaccinated\_smoothed
* new\_people\_vaccinated\_smoothed\_per\_hundred
* stringency\_index
* population\_density
* median\_age
* aged\_65\_older
* aged\_70\_older
* gdp\_per\_capita
* extreme\_poverty
* cardiovasc\_death\_rate
* diabetes\_prevalence
* female\_smokers
* male\_smokers
* handwashing\_facilities
* hospital\_beds\_per\_thousand
* life\_expectancy
* human\_development\_index
* population
* excess\_mortality\_cumulative\_absolute
* excess\_mortality\_cumulative
* excess\_mortality
* excess\_mortality\_cumulative\_per\_million

### Valori dei campi richiesti

|  |  |
| --- | --- |
| Nr positivi  (Dati cumulativi al 2023, 9 febbraio) | 25488166 |
| Nr ospedalizzati  (Dati totali giornalieri al 2023, 2 febbraio, includendo dimissioni) | 3891 |
| Nr decessi  (Dati cumulativi al 2023, 9 febbraio) | 187272 |

### Classificazione entità “owid-covid-data.csv”,

**Stelle 4**

* Dati aperti consultabili + scaricabili + modificabili + formato aperto (CSV, JSON) + URL

## Fonte 5 Covid19 – Google News

### Descrizione e fonte

Google News mette a disposizione una dashboard interattiva che descrive l’andamento covid19 in Italia (possibile collegamento al mondo), al seguente URL <https://news.google.com/covid19/map?hl=it&mid=%2Fm%2F03rjj&gl=IT&ceid=IT%3Ait&state=1> .

Dati visibili della dashboard: casi totali, nuovi casi, decessi, ricoveri ospedalieri, dosi tot somministrate + nuove dosi, persone con vaccinazione completa.

### Metadati

Sono presenti metadata descrittivi e gestionali. La dashboard rimanda alle Fonti John Hopkins University e Our World in Data.

### Campi richiesti (estrapolati)

|  |  |
| --- | --- |
| Nr positivi  (Dati cumulativi al 2023 3 febbraio) | 25488166 |
| Nr ospedalizzati  (Dati cumulativi al 2023 3 febbraio) | NA |
| Nr ospedalizzati  (Giornaliero 2 febbraio) | 3891 |
| Nr decessi  (Dati cumulativi al 2023 3 febbraio) | 187272 |

### Classificazione

**Stelle 1**

* dati aperti consultabili e condivisibili + non modificabili + non riconvertibili + non scaricabili

## Fonte 6 – RaiNews

### Descrizione e fonte

RaiNews ha creato una dashboard che descrive l’andamento covid19 l’andamento delle vaccinazioni in Italia (possibile collegamento al mondo), al seguente URL <https://www.rainews.it/ran24/speciali/2020/covid19/> .

### Campi richiesti (estrapolati)

|  |  |
| --- | --- |
| Nr positivi  (Dati cumulativi al 2023, 3 febbraio) | 25488166 |
| Nr ospedalizzati  (Dati cumulativi al 2023, 3 febbraio) | NA |
| Nr ospedalizzati  (Giornaliero 2 febbraio) | 3891 |
| Nr decessi  (Dati cumulativi al 2023, 3 febbraio) | 187272 |

### Metadati

I metadati della dashboard sono:

* Autore
* Fonte
* data di aggiornamento

### Classificazione

**Stelle 1**

* dati aperti consultabili e condivisibili + non modificabili + non riconvertibili + non scaricabili

## Fonte 7 Covid19 - John Hopkins Hospital

### Descrizione e fonte

La cartella pubblica “CSSEGISandData/COVID-19” in GitHub creata dalla John Hopkins University include i link per la dashboard e i vari collegamenti alle fonti dei dati che descrivono l’andamento dei casi, e relative analisi, della pandemia dovuta al Covid19.

La cartella è consultabile al seguente URL <https://github.com/CSSEGISandData/COVID-19>.

Questa cartella rimanda sia al progetto GitHub della protezione civile <https://github.com/pcm-dpc/COVID-19/tree/master/> che al sito del governo italiano <https://www.salute.gov.it/portale/nuovocoronavirus/dettaglioNotizieNuovoCoronavirus.jsp?lingua=italiano&menu=notizie&p=dalministero&id=6150> .

In questo progetto tutti i dati time-series e aggiornamenti quotidiani sono in formato aperto CSV.

### Campi richiesti (estrapolati)

Disponibili previa analisi degli open data, non disponibili come informazione.

### Classificazione “CSSEGISandData/COVID-19”

**Stelle 5**

* formato aperto (CSV) + URL + collegamento ad altri database

# Andamento Covid19 della classe DAPT0123 Epicode

### Descrizione e fonte

Il dataset in formato aperto (CSV) “Esercizio\_covid\_CSV.csv” descrive la situazione di contagio da Covid19 all’interno di un campione di 30 persone.

I campi/attributi che descrivono l’entità sono:

* ID (=numero identificativo)
* Informazioni cronologiche (= data di immissione dati)
* Nome
* Cognome
* Città (= di residenza)
* Positivo (= se la persona è stata contagiata e risultata positiva)
* Ricoverato (= se la persona risultata positiva abbia avuto bisogno di un ricovero ospedaliero)
* Giorni Degenza (= numero di giorni di positività al Covdi19)

Nel caso di questa entità la chiave primaria è il campo ID. Anche la combinazione dei campi Nome + Cognome può essere considerata come chiave primaria.

### Metadati

I metadati di questa entità sono le informazioni cronologiche (data di immissione dati).

### Valori dei campi richiesti (estrapolati)

|  |  |
| --- | --- |
| Nr positivi  (Dati cumulativi al 2023) | 18 |
| Nr ospedalizzati  (Dati cumulativi al 2023) | 0 |
| Nr decessi  (Dati cumulativi al 2023) | 0 |

# Confronto andamento Covid19 Italia – DAPT0123

Per confrontare i dati dell’andamento del covid19 del campione classe DAPT0123 e della popolazione Italia sono stati calcolati i valori relativi di percentuali di positivi e percentuale di ricoveri, considerando il totale di partecipanti al sondaggio di classe (30) e la popolazione dell’Italia ad oggi (59.11 milioni). La Tabella 1 mostra i valori ottenuti dopo il calcolo della percentuale.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Positività % | Ospedalizzazioni % | Ospedalizzazioni cumulativi % | Ospedalizzazioni senza dimissioni % | Decessi % |
| DAPT0123 | 60 | 0 | 0 | 0 | 0 |
| IT | 43 | 0.007 | 20.6 | 0.2 | 0.32 |

Tabella 1. Dati relative di positività e ricovero da Covid19 della classe DAPT0123 e dell’Italia, ad oggi (2023).

Dalla Tabella 1 si osserva una maggiore positività all’interno del campione DAPT0123 (60% di positività) rispetto che nell’andamento Italiano nazionale (~43%).

Per quanto riguarda la percentuale di ricoveri e decessi, il campione DAPT0123 è pari a zero, inferiore rispetto all’andamento nazionale, seppur bassa.

NOTA: i dati dell’andamento nazionale sono considerati dalla fonte de Il Sole 24 Ore.

# Ruoli data engineer, data analyst, data scientist

Nell’ambito dell’analisi e paragone dell’andamento del covid19, le figure di data engineer/analyst/scientist/journalist sono interevnuti nelle seguenti fasi:

* Data engineer -> Fase di raccolta e gestione dati (nazionali e della classe)
* Data analyst -> Fase di analisi dei dati
* Data scientist -> NA
* Data journalist -> NA