ESERCIZIO D3 = ANALISI DATI COVID IN ITALIA.

**Fonte n°1**

<https://github.com/pcm-dpc/COVID-19/blob/master/dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale-20221119.csv>

All’interno troviamo un report ‘’Dati andamento nazionale‘’ che fa riferimento ai dati covid in Italia alla data 19.11.2022, tra gli attributi troviamo: deceduti, tot. Positivi, ecc...

Il file che prendiamo in riferimento all’interno della fonte è un csv e viene classificato come dato a tre stelle in quanto fanno riferimento a: dati strutturati, dati codificati in un formato che può essere aperto da diversi programmi.

All’interno del csv troviamo: Tot. Ricoverati (vedi attributo Tot. ospedalizzati) 7350, Tot. Positivi: 470933; Tot. Decessi: 180679.

METADATI INDIVIDUATI:

* Parola Chiave: covid, covid Italia
* Data di pubblicazione: 25.11.2022
* Nome file: dpc-covid19-ita-andamento-nazionale-20221119
* Formato file: csv
* Autore: Presidenza del Consiglio dei Ministri - Dipartimento della Protezione Civile

**Fonte n°2**

<https://github.com/pcm-dpc/COVID-19/blob/master/schede-riepilogative/regioni/dpc-covid19-ita-scheda-regioni-20221126.pdf>

All’interno troviamo il report ‘’Aggiornamento casi covid-19’’ che fa riferimento a dati covid per ogni regione datati al 26.11.2022, tra gli attributi troviamo: decessi, positivi, ricoverati, ecc…

Il file che prendiamo in riferimento all’interno della fonte è un pdf che può può essere classificato come dato a una stella in quanto: è un dato libero, ha un tipo di licenza aperta, è facilmente leggibile, ricondivisibile, stampabile, ma difficile da modificare, possono essere aperti solo da determinati programmi e non può essere convertito in dato aperto.

L’ultima striga riporta il totale, quindi la situazione che fa riferimento all’intero Stato alla data 26.11.2022, ovvero, totali decessi : 181.271; totali positivi: 24.327.664; totali ricoverati: 6.359.

METADATI INDIVIDUATI:

* Parola Chiave: covid, covid Italia
* Data di pubblicazione: 02.12.2022
* Nome file: dpc-covid19-ita-scheda-regioni-20221126
* Formato file: pdf
* Autore: Presidenza del Consiglio dei Ministri - Dipartimento della Protezione Civile

**Fonte n°3**

<https://github.com/pcm-dpc/COVID-19/blob/cae4442f4dd4e157fe18677a372d7be1c915e083/dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale-20230127.csv>

All’interno troviamo un report ‘’Andamento report nazionale covid’’ che fa riferimento ai dati covid in Italia alla data 27.01.2023, tra gli attributi troviamo: deceduti, tot. Positivi, ecc...

Il file che prendiamo in riferimento all’interno della fonte è un csv e viene classificato come dato a tre stelle in quanto fanno riferimento a: dati strutturati, dati codificati in un formato che può essere aperto da diversi programmi.

All’interno del csv troviamo: Tot. Ricoverati (vedi attributo Tot. ospedalizzati) 4.203, Tot. Positivi (vedi attributo tot. Casi): 25.458.763; Tot. Decessi: 186871.

METADATI INDIVIDUATI:

* Parola Chiave: covid, covid Italia
* Data di pubblicazione: 03.02.2023
* Nome file: dpc-covid19-ita-andamento-nazionale-20230127
* Formato file: csv
* Autore: Presidenza del Consiglio dei Ministri - Dipartimento della Protezione Civile

Ho effettuato il confronto riguardante la situazione covid di questi anni, tra i dati che siamo riusciti ad estrapolare all’interno della classe stessa, organizzati in un file csv e i dati che sono riuscita a trovare sul web, in particolar modo per quest’ultimi, ho fatto riferimento alla fonte n°2 riportata sopra.

Il confronto è stato fatto in base agli indicatori comuni, ovvero: casi positivi, ricoveri, decessi, ecc…

Su 30 campioni presi in considerazione per effettuare un’analisi in base agli indicatori (positivi, giorni di degenza, ricoveri), si può evincere che quasi il 60 % di noi è risultato essere positivo, mentre nessuno di noi è stato ricoverato e tra quelli che sono stati positivi e quindi in isolamento, lo sono stati per una media di 10 giorni.

Il numero dei positivi della classe ed il numero dei positivi riportati nella fonte presa in considerazione, potrebbero essere in linea, mentre per quanto riguarda l’attributo ricoveri, in Italia c’è stata un’elevata percentuale di ricoveri che non trova corrispondenza rispetto e dati della classe da noi estrapolati.

La chiave primaria nel file della classe è rappresentata dall’attributo ‘’Id’’ delle persone che ne fanno parte, mentre la chiave primaria della tabella riferita alla situazione covid in Italia (fonte n°2) è rappresentata dall’attributo ‘’Regione’’.

Tra le figure che hanno dato un contributo all’elaborazione del progetto abbiamo:

La figura del data engineer la troviamo nella ricerca dei diversi dati inerenti alla situazione covid in Italia basandosi su diverse fonti, nell’organizzazione dei dati selezionati che poi andranno ad essere analizzati, sulla veridicità dei dati selezionati e sulla qualità dei formati in cui ha trovato tali dati.

La figura del data analyst la ritroviamo quando vengono analizzati i diversi dati messi a disposizione dal data engineer, grazie a queste analisi si è potuti arrivare ad un quadro definitivo della situazione covid in Italia, rendendosi così conto del rischio che si stava attraversando avendo dati reali in mano.

La figura del data scientist ha dato il suo contributo nel fare analisi predittive di quella che magari poteva essere l’avanzamento della curva epidemiologica sia nel breve che nel lungo periodo.

La figura del data journalist ha messo insieme tutti i dati e tutte le analisi in suo possesso per rendere partecipe l’intero mondo dell’andamento di questa epidemia, in modo tale che tutti fossero al corrente dei dati effettivi di questo fenomeno.