**ANALISI SITUAZIONE COVID**

La prima fonte che ho individuato si trova nel sito dati.gov.it accedendo al seguente link:

[COVID-19 Andamento nazionale (Open Data) | dati.gov.it](https://www.dati.gov.it/view-dataset/dataset?id=79f4c2db-c3d4-4a7c-ba1e-dc20f68f7196)

Si tratta di un database riepilogativo che contiene la serie temporale dei totali degli indicatori di monitoraggio della situazione relativa all’emergenza collegata alla diffusione del Coronavirus in Italia. Da questo documento si può evincere che il formato dell’open data dal quale si possono estrapolare le informazioni è definito CSV e può quindi essere classificato come un open data di 3 stelle. Di seguito il link per accedere al file:

<https://github.com/pcm-dpc/COVID-19/tree/master/dati-andamento-nazionale>

Altre caratteristiche che vengono definite come metadato relative al file dal quale ho scaricato il CSV sono:

TITOLARE: PCM dipartimento della protezione civile,

EDITORE: <http://dati.protezionecivile.it/PCM#redazione>

AUTORE: Soggetto attuatore MinSALUTE DCDPC 414/7-2-2020 “

KEYWORD: covid 19

DATA ULTIMA MODIFICA: 31/03/2021

LICENZA: Creative Commons Attribuzione 2.0 Italia (CC BY 2.0 IT)

Un'altra fonte può essere identificata nel sito dati.gov.it al link:

[COVID-19 Monitoraggio situazione Italia (RNDT - Serie) - Versione 2.0 | dati.gov.it](https://www.dati.gov.it/view-dataset/dataset?id=327cbe0a-4737-4e2c-9567-c7d90f69570b)

In questo open data si può scoprire il riepilogo nazionale quotidiano del monitoraggio sanitario che descrive il quadro riassuntivo della situazione epidemiologica relativa al Covid 19. Il formato del documento scaricabile può essere classificato 3 stelle, in quanto si tratta di un CSV.

Ecco il link del download:

<https://github.com/pcm-dpc/COVID-19/tree/master/dati-regioni>

Le caratteristiche (metadati) individuabili dalla fonte sono:

KEYWORDS: covid 19, human health, nazionale, open data, salute umana e sicurezza

DATA ULTIMA MODIFICA: 10/12/2021

LICENZA: Creative Commons Attribuzione 4.0 Internazionale (CC BY 4.0)

La terza fonte è stata individuata al seguente url:

[Coronavirus in Italia, i dati e la mappa (ilsole24ore.com)](https://lab24.ilsole24ore.com/coronavirus/)

In questa dashboard si ha una panoramica completa dei dati e dei grafici relativi al coronavirus in Italia e nel mondo. I metadati presenti in questo sito sono molteplici e facilmente visibili nel codice d’origine della pagina. Ovvero per esempio i titoli e la descrizione dei titoli (“coronavirus, i dati e le mappe” “Tutti i numeri del coronavirus in Italia: contagiati, morti, guariti e numero di tamponi. Informazioni per regioni e province fornite dal ministero della Salute”). O ancora il nome del sito stesso(www.sole21ore.com), il creatore(sole24ore). Dalla pagina web si possono scaricare dei file csv a 3 stelle, come il seguente, avente come chiave primaria il campo data:



Ulteriore fonte si può trovare a questo link:

[Pandemia di coronavirus (COVID-19) - Il nostro mondo nei dati (ourworldindata.org)](https://ourworldindata.org/coronavirus)

In questo open linked data (con classificazione 5 stelle) ci viene concessa la possibilità di indagare istantaneamente sulla propagazione della pandemia in tutto il mondo, attraverso grafici molto chiari e precisi ed eseguendo il download di file csv raccolto su GitHub :

[covid-19-data/owid-covid-data.csv at master · owid/covid-19-data (github.com)](https://github.com/owid/covid-19-data/blob/master/public/data/owid-covid-data.csv)

Dalla prima fonte è possibile ricavare il seguente link:

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

In questo repository si trova una cartella zip contenenti due documenti significativi per l’analisi della situazione pandemica dall’inizio dei primi sintomi al 02 febbraio 2023, data dell’ultimo aggiornamento. Il primo open data è un pdf, quindi con classificazione pari a 1 stella, che descrive per ogni regione i casi covid 19 confermati e anche il numero di tamponi effettuati. La chiave primaria di questa tabella rappresentata nel pdf è il campo regione.



Il secondo documento invece è un CSV, avente classificazione 3 stelle, e mostra una panoramica progressiva dei casi covid registrati in tutta Italia. La chiave primaria è il campo data.



È stato creato poi un report su 30 studenti inscritti alla scuola telematica EPICODE-EPIC EDUCATION al corso data Analyst per stabilire quali sono stati i casi positivi al covid 19 dall’inizio della pandemia e confrontare i dati raccolti con quelli relativi a tutta Italia. Questa nuova entità è denominata “casi covid data Analyst” e spiega appunto quanti ragazzi di quel corso sono stati contagiati, per quanto tempo sono rimasti degenti e se è stato necessario per loro anche il ricovero. Il campo informazioni cronologiche descrive a che ora ogni studente ha risposto al sondaggio relativo al coronavirus, dal quale poi è stato possibile ricavare le informazioni individuali. In questa tabella la chiave primaria è individuata nel campo identificativo ID e i metadata possono essere trovati nelle informazioni del file, sono descrittivi e riguardano la data di creazione (07/02/2023 alle 21.07), l’ultima data di modifica(10/02/2023 alle 21.18), il creatore (Krauly) e l’autore dell’ ultima modifica (Mattia Brugnaro).

Di seguito vengono riportati i file CSV ed EXCEL della tabella “casi covid data Analyst” e il file EXCEL dpc\_covid19\_andamento-nazionale:







Mettendo a confronto i dati raccolti nella tabella dpc-covid19-ita-andamento-nazionale.xlsx e quelli raccolti nel report casi covid data Analyst.xlsx si denota innanzitutto che la percentuale di positività della tabella casi covid data Analyst (18 casi su 30 individui : 60%) è superiore rispetto a quella della tabella andamento nazionale.xls (25.186.059 casi su 59 milioni di abitanti : 42,69%).Questo perché gli individui che compongono la classe data Analyst hanno una vita sociale e lavorativa movimentata e sono stati esposti più facilmente al virus. Un'altra considerazione importante riguarda il campo RICOVERATO nella tabella casi covid data Analyst. Nessun studente ha infatti manifestato sintomi gravi a tal punto da dover essere ospedalizzato, a differenza invece del numero degli ospedalizzati in tutta Italia (12.179.409 su 59 milioni: 20%). Questa osservazione potrebbe dipendere dal fattore media età, che è superiore in Italia rispetto alla classe data Analyst. Tenendo conto, infatti, che il covid ha colpito duramente e maggiormente individui con una fascia d’età alta, si può dedurre che coloro sotto un determinato range abbiano manifestato quindi sintomi più lievi. Basandosi invece sulla tabella in pdf che rappresenta l’andamento pandemico regionale e raffrontandola con quella in Excel della classe data Analyst, si può invece notare che nonostante il Veneto sia stata una delle regioni maggiormente infettate, nessun ragazzo veneto ha contratto la malattia. Sorge un ultimo spunto di riflessione dalla relazione tra questi due prospetti che però rispetto all’osservazione precedente sottolinea una correlazione. Si può riscontrare, infatti, che lo studente che ha dovuto affrontare più giorni di degenza proviene dal Piemonte, una delle regioni più colpite dalla malattia infettiva.

Un grande ruolo nell’elaborazione di queste informazioni e nella battaglia contro l’epidemia l’hanno avuto le seguenti figure del team data:

1. Il data engineer si è occupato della prima elaborazione dati, verificando l’autenticità delle informazioni. Ha definito i campi di ricerca, eliminando quindi fonti non attendibili e in base a dei criteri ben definiti ha scelto quali documenti potevano essere accettati e soggetti a nuove elaborazioni. Nello specifico si è occupato di raccogliere i dati pandemici nella loro interezza e integrità da numerose open source, garantendo la qualità dei dati e migliorando la pipeline.
2. Il data Analyst tramite potenti grafici ha sensibilizzato le informazioni pulite e autenticate dal data engineer e ha contribuito in modo importante a combattere l’avanzare dell’epidemia. Raccogliendo i dati sulla mobilità umana, per esempio, e trasformandoli è riuscito a informare il governo se le misure funzionavano o meno. Tramite l’analisi spaziale è stato poi in grado di determinare quali aree potevano sovrappopolarsi, e dare quindi dei consigli sulle linee guida da adottare per far si che questo non si verificasse.
3. Il data scientist attraverso l’utilizzo della IA e della machine learning ha elaborato modelli predittivi con il fine di aiutare le strutture ospedaliere a preparare nuovi posti letto per i ricoverati in terapia intensiva, e ha permesso di gestire la distribuzione di mascherine ipotizzando la curva crescente dei casi asintomatici
4. Il data journalist infine ha raccolto e pulito dati da numerose fonti diverse, ne ha riconosciuto il potenziale e ha raccontato la storia della pandemia, informando il mondo costantemente ma non solo in modo narrativo e distaccato, ma anche con animo e forza. Ha reso le informazioni accessibili a tutti tramite infografici e ha usato gli strumenti narrativi con uno scopo giornalistico per sensibilizzare i cittadini e responsabilizzarli.