Esercitazione 7/02/2023

SITUAZIONE COVID IN ITALIA

1. **FONTE:** <https://lab24.ilsole24ore.com/coronavirus/>

Ne **Il Sole 24 Ore** in cui vengono riportati i nuovi casi, gli attualmente positivi, i guariti e i morti dal primo rilevamento a oggi, cioè dal 24/02/2020 al 02/02/2023.

**CLASSIFICAZIONE**: Il dataset scaricabile è un .csv; classificazione **3 stelle**.

**ATTRIBUTI:** Gli attributi contenuti in tabella sono i seguenti

Data, ricoverati, terapia\_intensiva, ospedalizzati, isolamento\_domiciliare, totale\_positivi, variazione\_totale\_positivi, nuovi\_positivi, guariti, deceduti, totale\_casi, tamponi, casi\_testati, ingressi\_terapia\_intensiva, tamponi\_test\_molecolare, tamponi\_test\_antigenico\_rapido,

diff\_ricoverati, diff\_terapia\_intensiva, diff\_variazione\_totale\_positivi, diff\_nuovi\_positivi, diff\_guariti, 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

**ENTITA’:** andamento complessivo della pandemia.

**CHIAVE PRIMARIA:** campo data.

**METADATI:** Non vi è presenza di metadati.

1. **FONTE:** <https://www.rainews.it/ran24/speciali/2020/covid19/>

Il sito sopra riportato è **Rai News,** la cui fonte è la Protezione Civile, in cui vengono riportati i dati sugli attualmente positivi, casi totali, nuovi casi, l’incremento dei decessi, i decessi totali, % di positività sui test eseguiti, ricoveri, ricoveri in terapia intensiva, test effettuati al giorno e i guariti.

**CLASSIFICAZIONE:** I dataset scaricabili sono dei .csv caricati sulla piattaforma github; classificazione **4 stelle**.

**ATTRIBUTI:** I dati non sono cumulativi e vi è una tabella per ogni giorno da inizio pandemia con il riferimento di ogni regione.

Gli attributi contenuti nelle tabelle sono i seguenti:

data, stato, ricoverati\_con\_sintomi, terapia\_intensiva, totale\_ospedalizzati, isolamento\_domiciliare, totale\_positivi, variazione\_totale\_positivi, nuovi\_positivi, dimessi\_guariti, deceduti, casi\_da\_sospetto\_diagnostico, casi\_da\_screening, totale\_casi, tamponi, casi\_testati, Nota ingressi\_terapia\_intensiva, nota\_test, note\_casi

**ENTITA’:** aggiornamento giornaliero della pandemia.

**CHIAVE PRIMARIA**: campo data.

**METADATI:** Nei file, essendo questi scaricabili da github, sono presenti metadati che forniscono informazioni sugli autori, sulla data di creazione e sulla data di modifica.

1. **FONTE**:<https://opendatamds.maps.arcgis.com/apps/dashboards/0f1c9a02467b45a7b4ca12d8ba296596>

Il sito sopra riportato è la dashboard del **Ministero della Salute.**

**CLASSIFICAZIONE:** I dati sono scaricabili attraverso un collegamento ipertestuale all’interno del sito attraverso dei file pdf, i quali contengono però immagini di tabelle non modificabili.

In questo file pdf esiste un collegamento ipertestuale che rimanda a github dove posso trovare i dati salvati come .csv.

La classificazione dei dati open è dunque **5 stelle.**

I dati non sono cumulativi, ma settimanali.

**ATTRIBUTI:** gli attributi contenuti nella tabella sono

data, stato, ricoverati\_con\_sintomi, terapia\_intensiva, totale\_ospedalizzati, isolamento\_domiciliare, totale\_positivi, variazione\_totale\_positivi, nuovi\_positivi, dimessi\_guariti, deceduti, casi\_da\_sospetto\_diagnostico, casi\_da\_screening, totale\_casi, tamponi, casi\_testati, Nota ingressi\_terapia\_intensiva, nota\_test, note\_casi

**ENTITA’**: andamento settimanale della pandemia.

**CHIAVE PRIMARIA:** è il campo settimana per i vari report delle regioni.

**METADATI:** Nei file, essendo questi scaricabili da github, sono presenti i metadati che forniscono informazioni sugli autori, sulla data di creazione e sulla data di modifica.

1. **FONTE**: <https://www.epicentro.iss.it/coronavirus/open-data/covid_19-iss.xlsx>

Il sito sopra riportato è quello dell’Istituto Superiore di Sanità.

**CLASSIFICAZIONE:** permette di scaricare direttamente un file excel;

classificazione del dato open **2 stelle**.

**ENTITA’:**

-nota metodologica (solo testo)

-contenuto (in cui vengono chiariti i significati degli attributi)

-casi\_prelievo\_diagnosi (Numero casi per data prelievo/diagnosi)

-casi\_inizio\_sintomi (Numero casi per data inizio sintomi)

-casi\_inizio\_sintomi\_sint (Numero casi per data inizio sintomi, ovvero i casi di infezione confermata da virus SARS-CoV-2 per cui è indicata una data inizio sintomi, tranne i casi dichiarati asintomatici)

-casi\_regione

-casi\_province

-ricoveri

-decessi

-sesso\_età

-stato\_clinico

**ATTRIBUTI E CHIAVI PRIMARIE:**

Prendo in considerazione solo 3 entità

* L’entità ricoveri ha come attributi:

-data ricovero

-numero ricoveri

-data registrazione dato

La chiave primaria è il campo data ricovero.

* L’entità decessi ha come attributi:

-data decesso

-numero decessi

-data registrazione dato

La chiave primaria è il campo data decesso.

* L’entità casi\_regioni ha come attributi:

-nome regione

-casi cumulativi

-data registrazione dato

La chiave primaria è il campo nome regione.

**METADATI:** data di creazione, modifica e autore.

Altri metadati all’interno del file

1. **FONTE:** <https://covid19.infn.it/iss/>

I dati sono stati forniti dall'[**Istituto Superiore di Sanità**](https://www.iss.it/)**(ISS)** con il quale l'**INFN** ha stabilito un accordo di collaborazione scientifica. L'accordo non prevede la possibilità, da parte dell'INFN, di rendere pubblici i dati ricevuti dall'ISS. Tuttavia, è possibile scaricare i dati elaborati, relativi ai grafici pubblicati.

**CLASSIFICAZIONE:** I dati sono scaricabili in formato .csv; classificazione **3 stelle**.

**ATTRIBUTI:** data, positivi, ricoveri, deceduti.

**ENTITA’:** andamento complessivo della pandemia.

**CHIAVE PRIMARIA:** campo data.

**METADATI:** Non vi è presenza di metadati.

**N.B.:** Ho trovato 5 fonti inerenti alla ricerca richiesta, ma solo tre di questi (1,4,5) soddisfano il requisito dei dati cumulativi.

**FIGURE PROFESSIONALI**

La figura del Data Engineer interviene per quanto riguarda la reperibilità dei dati e l’attendibilità delle fonti.

La figura del Data Analyst interviene sull’efficace visualizzazione grafica di tali dati e sulla loro corretta interpretazione.

La figura del Data Scientist interviene nel momento in cui si ottengono modelli predittivi sull’andamento della pandemia.

Il Data Journalist interviene sull’esposizione informativa dei dati sulle testate giornalistiche o sui siti di informazione mediatica.

Nell’eseguire questa ricerca ho ricoperto i ruoli di Data Engineer e Data Analyst.

**CONCLUSIONI**

I casi avuti in classe sono 18/30, ovvero il 60%.

Secondo la prima fonte de “Il Sole 24 Ore” i positivi ammontano a 25’186'059.

Considerando che la popolazione italiana ammonta a circa 59'000'000 abitanti, raggiungiamo una percentuale di incidenza del 42,7%.

Secondo la quarta fonte “Istituto Superiore di Sanità” i positivi ammontano a 25’700'453, dunque raggiungiamo una percentuale di incidenza del 43,6%.

Possiamo concludere quindi che la percentuale di incidenza della classe, rispetto all’incidenza nazionale, si alza del 17% circa.