

Autore: Francesco Sapia

Data: 07/06/2023

Obiettivo parte 1:

- Recuperare 5 fonti dati da internet che descrivano la situazione Covid al 2023 (Contenete almeno i campi "Positivi al Covid", "Ricoverti", "Morti")
- Indicare la fonte, effettuarne una breve descrizione e classificarla secondo la classificazione a 5 stelle
- Definire una tabella in cui vengono confrontati i dati
- Dare un nome alle entità ed agli attributi in modo descrittivo

Obiettivo parte 2:

- Definire la chiave primaria per ogni tabella
- Descrivere in quale fase intervengono le figure studiate
- Identificare i metadati

Svolgimento parte 1:

- Fonte consultata in data 08/06/2023 ore 07:27, stato del servizio attivo;
Link: https://data.humdata.org/dataset/coronavirus-covid-19-cases-and-deaths?force_layout=desktop
La fonte consente di visualizzare graficamente i dati e di filtrarli per singolo paese. Consente di scaricare il dataset in formato csv.
È un'organizzazione privata che lavora in ambito sanitario.
Nel dataset vengono riportati i seguenti campi:
 - Data;
 - Codice paese;
 - Paese;
 - Continente;
 - Nuovi positivi;
 - Positivi totali
 - Nuovi morti;
 - Morti totali;

Utilizzando la classificazione "5 stelle" tale fonte può essere classificata come 4 Stelle.

- Fonte consultata in data 08/06/2023 ore 07:38, stato del servizio attivo;
Link: <https://www.epicentro.iss.it/coronavirus/sars-cov-2-dashboard>
La fonte consente di visualizzare graficamente i dati. Consente di scaricare il dataset in formato xlsx.
È un'organizzazione pubblica che lavora in ambito sanitario. Fa riferimento all'istituto superiore della sanità.
Nel dataset sono presenti diversi fogli di calcolo ciascuno caratterizzato da diversi campi, i fogli d'interesse per l'esercizio sono stati posti in grassetto e sottolineati per evidenziarli. Si hanno i seguenti fogli:
 - **Casi prelievo diagnosi;**
 - Casi inizio sintomi;
 - Casi per regione;
 - Casi per provincia;
 - **Ricoverti;**
 - **Decessi;**

- Sesso ed età;
- Stato clinico;

Per ogni foglio come campi sono riportati i seguenti campi:

- Data;
- Argomento del foglio (es. ricoveri);
- La data relativa al download del dataset;

Utilizzando la classificazione “5 stelle” tale fonte può essere classificata come 2 Stelle in quanto i dati sono resi disponibili in un formato proprietario, excel nello specifico.

- Fonte consultata in data 08/06/2023 ore 07:27, stato del servizio attivo;

Link: <https://www.curaitalia.it/>

La fonte consente di visualizzare graficamente i dati e di filtrarli per singolo paese. Non consente di scaricare il dataset.

È un’organizzazione privata senza fini di lucro allo scopo di monitorare l’andamento del covid.

Nel dataset vengono riportati i seguenti campi:

- Totale casi;
- Nuovi positivi;
- Positivi attuali;
- Guariti;
- Deceduti;
- Terapia intensiva;
- Ricoverati con sintomi;
- Isolamento domiciliare;
- Tamponi;
- Persone testate;

Utilizzando la classificazione “5 stelle” tale fonte può essere classificata come 1 Stelle in quanto i dati non possono essere elaborati ma sono solo consultabili.

- Fonte consultata in data 10/06/2023 ore 11:54, stato del servizio attivo;

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

La fonte consente di scaricare il dataset in vari formati.

È un’organizzazione pubblica che opera in ambito sanitario.

Nel dataset vengono riportati i seguenti campi:

- Paese;
- Codice paese;
- Continente;
- Popolazione;
- Indicatore;
- Anno/settimana;
- Ecc;

Utilizzando la classificazione “5 stelle” tale fonte può essere classificata come 4 Stelle.

- Fonte consultata in data 10/06/2023 ore 11:54, stato del servizio attivo;

Link: <https://data.world/ondata/covid-19-italia-dati-dipartimento-protezione-civile>

La fonte consente di visualizzare e filtrare i dati. Consente di scaricare il dataset in formato csv.

È un’organizzazione privata che opera in ambito raccolta dati.

Nel dataset vengono riportati i seguenti campi:

- Data;
- Stato;
- Ricoverati con sintomi;
- Terapia intensiva;
- Totale ospedalizzati;
- Isolamento domiciliare;
- Totale positivi;
- Nuovi positivi;
- Dimessi guariti;
- Decessi;
- Ecc.;

Utilizzando la classificazione “5 stelle” tale fonte può essere classificata come 4 Stelle.

Di seguito viene riportata una tabella in cui vengono confrontati i vari dati ottenuti dalle 5 fonti.

	Fonte	Numero positivi	Numero di ricoveri	Numero di decessi
1	Humadata	702395	-	5599
2	Istituto superiore di sanità	657056	34224	4395
3	Curaitalia	669913	527938	5707
4	ECDC	861249	-	6461
5	Data.world	685385	549757	5653

Note:

- Nelle fonti 1 e 4 non viene riportato il campo dei ricoveri;
- Nel caso della seconda fonte vengono riportati i nuovi ricoveri per singolo giorno.
- Nelle fonti 3 e 5 i dati relativi ai ricoveri tengono del numero di persone effettivamente ricoverate per una specifica data;

All'interno della tabella avremo i seguenti attributi:

- Fonte;
- Numero di positivi;
- Numero di ricoveri;
- Numero di decessi;

Svolgimento parte 2:

Come chiave primaria della tabella precedentemente riportata possiamo considerare la “fonte”.

- Data Engineer. Se consideriamo la seconda fonte si ha che i dati vengono comunicati dall'Istituto Superiore di Sanità. Il compito del data engineer è quello di collezionare, analizzare i vari dati siano memorizzati nel modo corretto.
- Data Analyst. Una volta ricevuti i dati provenienti dall'ISS il compito del data analyst è stato quello di andare analizzare e trasformare i dati in modo tale da renderli fruibili al consumatore.
- Data Scientist. Nel caso in esame al momento della consultazione del link si ha che non è stata individuata una fase in cui si evince la presenza di un data scientist. Comunque il data scientist

utilizzando come base i dati raccolti dal data engineer utilizzando modelli statistici potrebbe realizzare delle previsioni su quelle che potrebbero essere i futuri sviluppi del covid.

- All'interno dell'esercitazione possiamo individuare sia metadati descrittivi che metadati strutturali.