

## Progetto

### Analisi e raccolta dei requisiti per un progetto di analisi dati

#### 1. Descrizione del caso di studio scelto

Si intende realizzare un sistema per l'analisi di dati per il Centro di Ricerca di una nota Università italiana, impegnato da più di 2 anni nella raccolta quotidiana di dati relativi alla pandemia da COVID-19. Coadiuvando il team di ricerca, lo scopo del progetto sarà quello di preparare i dati per l'analisi e procedere con la creazione di una dashboard interattiva sul sito web dell'Università. Il progetto mira a sensibilizzare la popolazione sulla tematica e promuovere comportamenti che possano contenere il contagio.

#### 2. Analisi e raccolta dei requisiti (generici e poi più specifici)

Si hanno a disposizione i dati relativi ai pazienti COVID-19 ricoverati in Italia, alle ospedalizzazioni e alle regioni d'Italia dal mese di gennaio 2020 alla data odierna.

Si allega al documento un estratto dei tre dataset in formato CSV, XML e JSON.

#### 3. Strutturazione dei requisiti in gruppi di frasi omogenee

Per i **pazienti ricoverati**, identificati attraverso un codice univoco, si hanno a disposizione nome, cognome, età, genere, numero di carta d'identità, nazionalità, altezza, peso, comorbidità, data di ricovero, data di dimissione, farmaci prescritti e note.

Per le **ospedalizzazioni**, identificate attraverso un codice univoco, si hanno a disposizione codice regione, ospedale, città, data, pazienti covid ricoverati.

Per le **regioni**, identificate attraverso un codice, si hanno a disposizione capoluogo, popolazione, densità, età media, data, casi totali, nuovi casi, morti totali, nuovi morti, persone vaccinate, nuove vaccinazioni.

#### Dizionario dei dati:

Colonna	Tipo	Descrizione	Formato Dataset
Codice	text	Codice univoco paziente ricoverato	CSV
Nome	text	Nome paziente ricoverato	CSV
Cognome	text	Cognome paziente ricoverato	CSV
Eta	numeric	Età paziente ricoverato	CSV

Sesso	boolean	Sesso del paziente ricoverato	CSV
Carta identita	text	Carta d'identità del paziente ricoverato	CSV
nazionalita	text	Nazionalità del paziente ricoverato	CSV
Altezza(cm)	numeric	Altezza espressa in centimetri del paziente ricoverato	CSV
Peso(kg)	numeric	Peso espresso in kilogrammi del paziente ricoverato	CSV
Comorbidita	text	Comorbidità al Covid-19 del paziente ricoverato	CSV
Data Ricovero	date	Data ricovero del paziente ricoverato	CSV
Data Dimissione	date	Data dimissione del paziente ricoverato	CSV
Farmaci	text	Farmaci prescritti per il trattamento del Covid-19	CSV
Note	text	Trascrizione delle note prodotte durante il ricovero	CSV
cod_ospedalizzazione	text	Codice ospedalizzazione	JSON
cod_regione	text	Codice regione	JSON
ospedale	text	Ospedale dove è avvenuto il ricovero	JSON
giorno	numeric	Giorno di ospedalizzazione	JSON

mese	numeric	Mese di ospedalizzazione	JSON
anno	numeric	Anno di ospedalizzazione	JSON
paz_covid	text	codice del/dei paziente/i ricoverato/i	JSON
capoluogo	text	Capoluogo di regione	XML
popolazione	numeric	Popolazione della regione	XML
densita	numeric	Densità della popolazione della regione	XML
eta_media	numeric	Età media della popolazione della regione	XML
giorno	numeric	Giorno della data entry	XML
mese	number	Mese della data entry	XML
anno	numeric	Anno della data entry	XML
casi_totali	numeric	Numero totale di casi COVID-19. Il conteggio può includere casi probabili, quando riportati	XML
nuovi_casi	numeric	Numero di nuovi casi COVID-19. Il conteggio può includere casi probabili, quando riportati	XML
morti_totali	numeric	Numero totale di morti attribuite a COVID-19. Il conteggio può includere morti probabili, quando riportate	XML
nuovi_morti	numeric	Numero di nuove morti attribuite a COVID-19. Il	XML

		conteggio può includere morti probabili, quando riportate	
persone_vacc	numeric	Numero totale di vaccinazioni di persone che hanno ricevuto almeno una dose di vaccino	XML
nuove_vacc	numeric	Numero di nuove vaccinazioni di persone che hanno ricevuto almeno una dose di vaccino	XML

La dashboard interattiva sarà accessibile al pubblico attraverso l'URL <https://coronavirus.uniroma1.it>

I requisiti sono accompagnati da una serie di immagini estrapolate dal mockup progettato dal team di design.

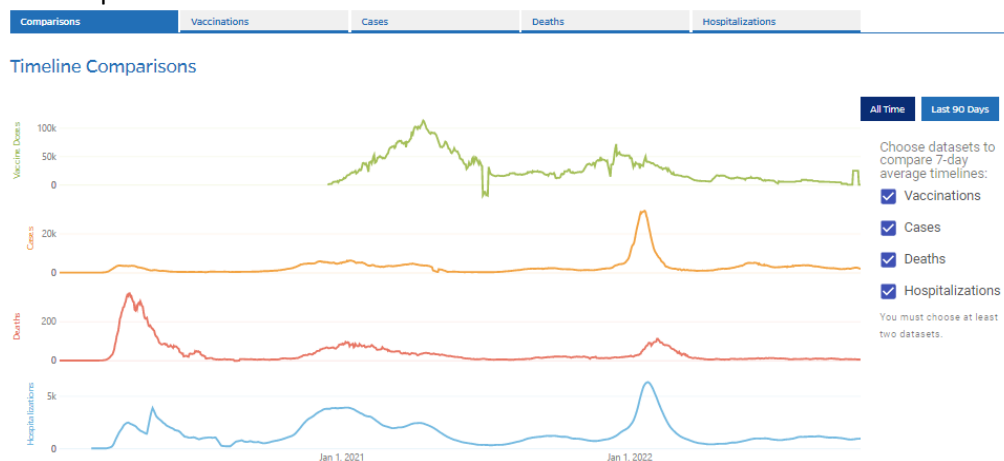
Effettuato l'accesso l'utente potrà:

Selezionare una regione italiana.

- 1.1.1 Visualizzare l'elenco delle regioni in ordine alfabetico.
- 1.1.2 Scrivere nella barra di ricerca la regione di suo interesse e navigare nella regione.
- 1.1.3 Cliccare sulla regione di suo interesse dall'elenco regioni e navigare nella regione.
- 1.1.4 In caso di ricerca o di navigazione, il risultato sullo schermo conterrà una sintesi dei dati cumulati relativi al COVID-19 per la regione (totale casi, totale morti, totale vaccinazione).
- 1.1.5 Un menù permetterà di visualizzare i dati per giorno/mese/anno.

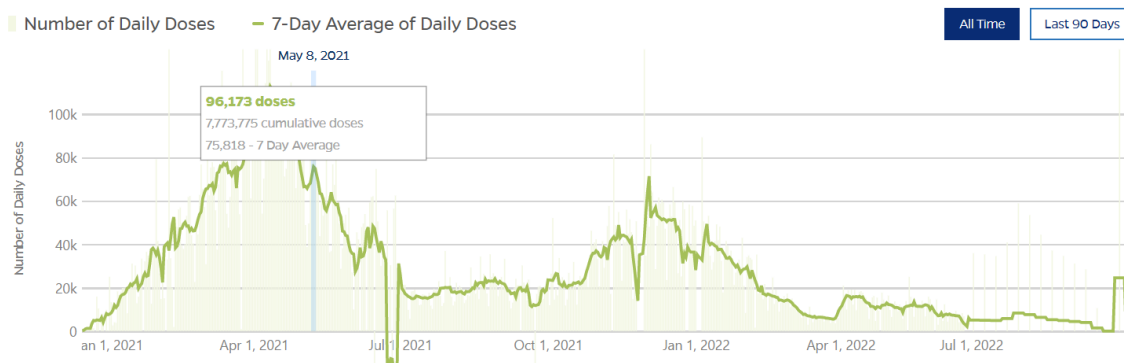
Accedere alla data timeline di una regione italiana

- 1.2.1 Scegliere il dataset da comparare dal menù laterale (vaccinazioni, casi, morti, ospedalizzazioni). Due bottoni permetteranno di selezionare la visualizzazione dei dati cumulati o trimestrali.



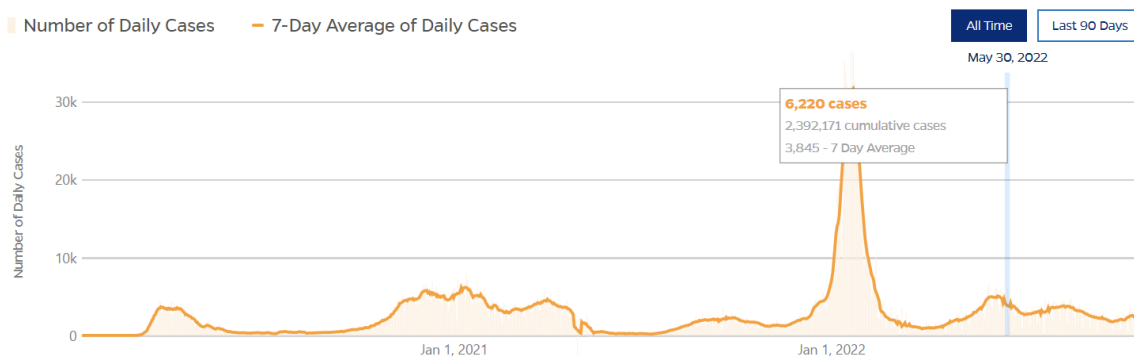
1.2.2 Selezionare i dati relativi alle sole vaccinazioni.

1.2.3 In caso di selezione dei dati relativi alle vaccinazioni, il risultato sullo schermo mostrerà una visualizzazione delle dosi di vaccino erogate giornalmente. Il dato riportato farà riferimento alla media settimanale. Due bottoni permetteranno di selezionare la visualizzazione dei dati cumulati o trimestrali.



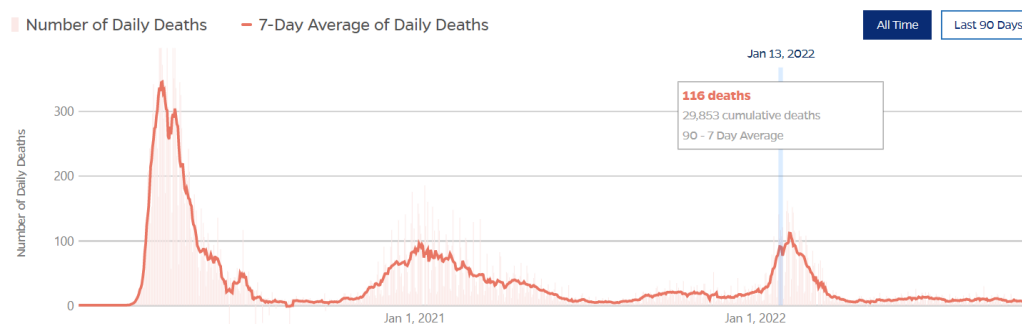
1.2.4 Selezionare i dati relativi ai soli casi di COVID-19.

1.2.5 In caso di selezione dei dati relativi ai casi, il risultato sullo schermo mostrerà una visualizzazione dei casi giornalieri di COVID-19. Il dato riportato farà riferimento alla media settimanale. Due bottoni permetteranno di selezionare la visualizzazione dei dati cumulati o trimestrali.



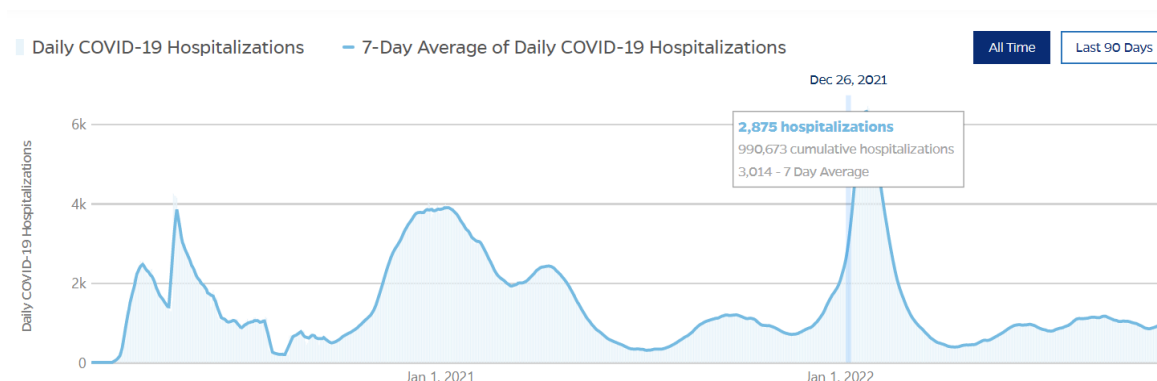
1.2.6 Selezionare i dati relativi alle sole morti attribuite a COVID-19.

1.2.7 In caso di selezione dei dati relativi alle morti, il risultato sullo schermo una visualizzazione delle morti giornaliere attribuite al COVID-19. Il dato riportato farà riferimento alla media settimanale. Due bottoni permetteranno di selezionare la visualizzazione dei dati cumulati o trimestrali.



1.2.8 Selezionare i dati relativi alle sole ospedalizzazioni dovute a positività al COVID-19.

- 1.2.9 In caso di selezione dei dati relativi alle ospedalizzazioni, il risultato sullo schermo mostrerà una visualizzazione delle morti giornaliere attribuite al COVID-19. Il dato riportato farà riferimento alla media settimanale. Due bottoni permetteranno di selezionare la visualizzazione dei dati cumulati o trimestrali.

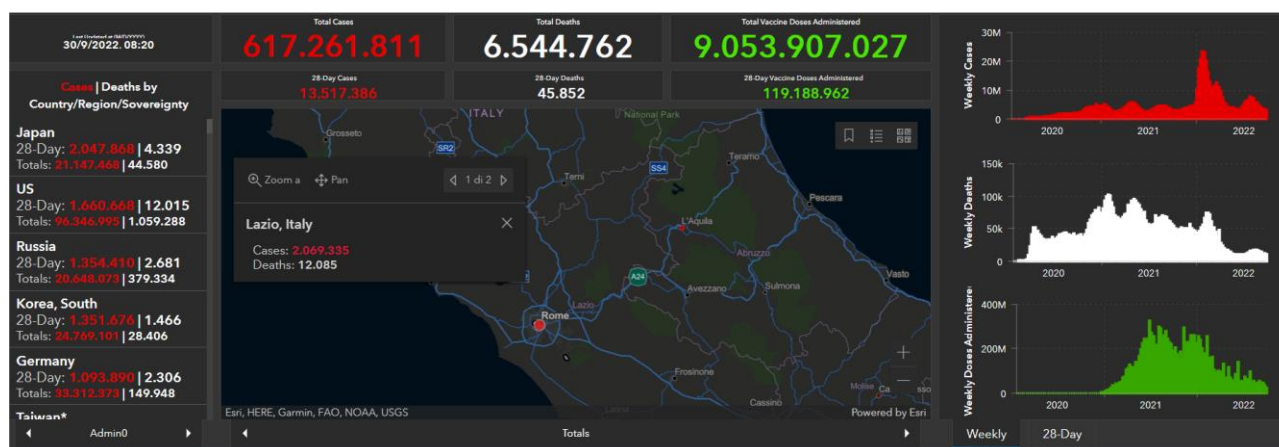


Accedere ai dataset

- 1.3.1 Visualizzare i dati grezzi in essi contenuti.
- 1.3.2 Scaricare liberamente i dataset.
- 1.3.3 Il sistema, i dataset ed i dati prodotti saranno distribuiti con licenza CC BY 4.0 (Creative Commons Attribuzione 4.0 Internazionale).

Accedere alla mappa nazionale

- 1.4.1 Muoversi liberamente sulla mappa.
- 1.4.2 Cliccare su un punto della mappa.
- 1.4.3 Nel caso in cui venga cliccato un punto della mappa, il risultato sullo schermo mostrerà i dati relativi ai casi e morti attribuiti a COVID-19 nella regione.



- 1.4.4 Cliccare sulla barra inferiore per visualizzare i dati dei casi mensili/totali, incidenza, rapporto tra casi e morti, vaccinazioni.

#### 4. Glossario dei termini

Termine	Descrizione	Sinonimi	Collegamenti
COVID-19	Malattia infettiva respiratoria causata da virus appartenente alla famiglia dei coronavirus	/	Paziente ricoverato
Italia	Stato situato in Europa Meridionale	Repubblica Italiana	Regione
Ospedale	Istituzione per l'assistenza sanitaria, il ricovero e la cura dei pazienti.	/	Regione, Paziente ricoverato
Ospedalizzazione	Ricovero in ospedale dovuto a COVID-19	/	Ospedale, Paziente ricoverato
Paziente ricoverato	Persona positiva al COVID-19	/	Ospedalizzazione
Regione	Divisione amministrativa della Repubblica Italiana. In Italia ce ne sono 20	/	Italia, Università, Ospedale
Università	Ente che ha il compito di promuovere la ricerca, il progresso delle scienze e l'istruzione di livello superiore	/	Regione
Utente	Persona che accede alla pagina web dell'Università e usufruisce dei servizi in esso presenti.	/	Università

#### 5. Outcome atteso dal sistema, come potrebbe essere di supporto nelle decisioni ed eventualmente quali 'storie' è possibile ricavare

Svolgendo operazioni logico-matematiche quotidiane sui diversi dati disponibili nel dataset è possibile estrapolare delle informazioni condivisibili alla popolazione attraverso una dashboard interattiva.

Seguono esempi:

- Casi totali / Morti totali: permette di esplorare la probabilità di contrarre il COVID-19 a livello nazionale o regionale
- Casi totali / Popolazione: permette di esplorare la percentuale di popolazione infettata a livello nazionale o regionale
- Popolazione / Totale vaccinazioni: permette di esplorare la percentuale di popolazione con almeno 1 dose di vaccino
- Regioni con il più alto tasso di infezioni da COVID-19 rispetto alla popolazione
- Regioni con il più alto tasso di morte attribuita a COVID-19 rispetto alla popolazione

Queste operazioni saranno affiancate da modelli predittivi, che permetteranno di stimare con un grado accettabile di incertezza il comportamento della pandemia, stabilendo quando e sotto quali condizioni il Paese italiano può aspettarsi picchi, riduzioni di nuovi casi (incidenza) e morte (mortalità). Con queste informazioni, sarà possibile calcolare la domanda per servizi medici in acuzie, così come predire bisogni futuri per successive ondate della pandemia da COVID-19.

Stimare la domanda di assistenza sanitaria permetterà di pianificare per tempo l'invio di strumentazioni sanitarie dove necessarie e garantire livelli adeguati di supply chain e distribuzione, come anche una corretta gestione delle risorse umane per un'appropriata e tempestiva risposta.

**Ruoli:**

- **Data Engineer:** affiancherà il team di ricerca e si assicurerà che i dati necessari per le analisi siano raccolti nel formato e nel modo corretto.
- **Data Analyst:** si occuperà di individuare trend e pattern esaminando i dataset universitari. Ispezionerà, analizzerà e trasformerà i dati oggetto del progetto.
- **Data Scientist:** si occuperà di elaborare modelli predittivi sull'andamento della pandemia
- **Data Journalist:** si occuperà di rendere fruibili e facilmente comprensibili alla popolazione le informazioni derivate da dati complessi grazie allo storytelling e alla dashboard interattiva. Il suo è un ruolo chiave nel sensibilizzare i cittadini sulla tematica e promuovere comportamenti che possano contenere il contagio attraverso la produzione di articoli sull'argomento.

Roma, 30/09/2022

Filippo Testa