

Università dell'Insubria
Dipartimento di Scienze Teoriche e Applicate - DiSTA
Corso di Laurea Triennale in Informatica



Realizzazione di una Dashboard di Business Intelligence per la pandemia da COVID-19

Tesi di:
Alessandro Riva
Matricola 734572

Relatore:
Prof. Davide Tosi

Anno Accademico 2019/2020

*Per questo traguardo devo ringraziare innanzitutto i miei genitori, che mi hanno
sostenuto e incoraggiato durante questi anni.
Senza mia madre e mio padre, non avrei avuto la possibilità di studiare e di scrivere
questo elaborato.*

Contenuti

1	Introduzione	1
1.1	Scopo del lavoro	2
1.2	La Business Intelligence	2
1.3	L'importanza dei dati	3
2	Contesto del lavoro	4
3	Obiettivi	5
4	La visualizzazione dati	6
4.1	Cos'è la visualizzazione dati	6
4.2	Breve storia	6
5	Design e sviluppo della soluzione	9
5.1	Origine dei dati	9
5.2	Tecnologie utilizzate	10
5.2.1	Python	10
5.2.2	Plotly	10
5.2.3	Pandas	11
5.2.4	Dash	11
5.2.5	Github	12
5.2.6	Docker	12
5.3	Lo sviluppo	13
5.3.1	Funzionamento	13
5.3.2	Calcolo dei dati	13
5.3.3	Diagramma di attività	14
5.3.4	Organizzazione file	14
5.4	Problemi riscontrati	14
5.4.1	Errore nei tamponi	16
6	Le dashboard	17
6.1	Dashboard Italia	17
6.1.1	Nuovi casi	18
6.1.2	Totale casi	19
6.1.3	Isolamento domiciliare	19

6.1.4	Terapia intensiva	20
6.1.5	Casi normalizzati	20
6.1.6	Terapia intensiva e ospedalizzati	21
6.1.7	Decessi e contagi giornalieri	21
6.1.8	Percentuale dei casi positivi sui casi testati	22
6.1.9	Nuovi casi e decessi	22
6.2	Dashboard Lombardia	23
6.2.1	Andamento dei contagi	23
6.2.2	Percentuale positivi sui casi testati	23
6.2.3	Casi normalizzati	23
6.2.4	Ospedalizzazioni	25
6.2.5	Decessi giornalieri	25
6.2.6	Terapia intensiva	26
6.3	Dashboard Regioni	26
7	Il portale Covid19-Italy	28
7.1	Rassegna stampa	31
7.2	Statistiche	32
8	Conclusioni	33
	Appendice	35
A	Manuale utente	37
A.1	Esecuzione in locale	37
A.1.1	Requisiti software	37
A.1.2	Clonazione del repository	38
A.1.3	Installazione dipendenze	38
A.1.4	Avvio del programma	38
A.2	Configurazione della VPS	39
A.2.1	Creazione account utente	39
A.2.2	Installazione di Docker	39
A.2.3	Installazione e avvio dei container	40
A.2.4	Portainer	40
A.2.5	Ngnix	40
A.2.6	Installazione certificato SSL	42
	Bibliografia	45

1

Introduzione

Viviamo in un mondo sempre più data-driven che tradotto significa “guidato dai dati”, un’espressione efficace per indicare che le decisioni vengono prese in base all’analisi dei dati e quindi su fatti oggettivi e non su interpretazioni personali. Raccogliere, analizzare e comprendere i dati è utile in ogni campo, dalla scienza, all’ingegneria, alla tecnologia alla matematica e così via, per poter attuare delle decisioni strategiche.

I dati sono fondamentali nei processi produttivi, nella logistica, nelle risorse umane, anche per la divisione amministrativa e nella gestione finanziaria.

La quantità di dati prodotti ogni giorno cresce in modo esponenziale, quindi è importante saper gestire questa mole di dati nel modo più efficiente ed efficace possibile, utilizzando le nuove tecnologie.

La visualizzazione dei dati aiuta a spiegare i contenuti, organizzandoli in un modo più comprensibile e mettendo in evidenza tendenze e valori anomali.

Una visualizzazione efficace consente di esporre i contenuti eliminando dai dati il superfluo e portando in primo piano le informazioni utili.

Sono moltissimi i metodi di visualizzazione disponibili per presentare i dati in modo efficace e interessante.

1.1 Scopo del lavoro

Giornalmente il dipartimento di Protezione Civile comunica i numeri dei contagi relativi alla pandemia di Covid 19. Per una migliore consultazione si è pensato di sviluppare un'applicazione in grado di rappresentare i dati in modalità grafica, per facilitare la lettura e la comprensione dell'andamento della pandemia. Lo scopo del seguente elaborato è quello di illustrare la struttura dell'applicazione creata per la realizzazione di questo servizio, accessibile online, attraverso l'elaborazione dei dati relativi all'emergenza di Coronavirus Covid-19 utilizzando dei grafici interattivi.

1.2 La Business Intelligence

La Business Intelligence o BI si può definire come un insieme di metodologie processi e strumenti che utilizzano le architetture e le tecnologie per implementare tutti i processi necessari per la raccolta dei dati, che attraverso elaborazioni, analisi e aggregazioni, ne permettono la trasformazione in informazioni, la loro conservazione, e la presentazione in una forma semplice, flessibile ed efficace, tale da costituire un supporto alle decisioni strategiche ed operative.

Il concetto è stato introdotto negli anni '60, quando si è iniziato a parlare di sistemi per aiutare i manager nei processi decisionali, ma è a partire dagli anni '80 che si è diffuso in modo massiccio.

La BI consiste nella trasformazione dei dati grezzi in informazioni utili e significative al fine dell'utilizzo in processi decisionali strategici, in modo che questi siano supportati da informazioni coerenti, aggiornate e altamente precise.

Un'applicazione di BI è un software che permette di trasformare grandi quantità di dati strutturati e anche de-strutturati in informazioni strategiche. Il compito della BI è infatti quello di fornire report, statistiche, indicatori, grafici aggiornati utili ai decisori per effettuare le scelte più corrette.

1.3 L'importanza dei dati

Nelle prime settimane della pandemia, il Covid-19 ha rapidamente conquistato le prime pagine di giornali, telegiornali e siti web con informazioni scientifiche spesso prive di fondamento o provenienti da fonti poco attendibili.

Per far fronte a emergenze sanitarie di questo tipo è molto importante poter disporre di informazioni tempestive, comprensibili e il più possibile accurate e condividerle tra gli operatori sanitari.

I medici e gli operatori sanitari durante la pandemia hanno dovuto, oltre al loro consueto lavoro, raccogliere e fornire alle istituzioni una enorme quantità di informazioni impiegando la tecnologia moderna.

Oggigiorno soltanto una minoranza degli ospedali italiani dispone di cartelle cliniche elettroniche, per poter raccogliere i dati in tempo reale.

Si è reso necessario costruire ad hoc dei sistemi informativi solidi per l'analisi e la condivisione dei dati su larga scala. Questi sistemi informativi di Business Intelligence, di cui ne soffriamo la mancanza, sono doverosi per migliorare la conoscenza e la qualità della cura.

Attualmente non è disponibile un database unificato a livello nazionale liberamente accessibile con i dati sul Covid-19 per ogni paziente.

2

Contesto del lavoro

Il 30 dicembre 2019, Li Wenliang, un oftalmologo dell'ospedale di Wuhan, ha condiviso online in una chat dei suoi amici la notizia di un nuovo virus sospetto simile alla SARS. Il 3 gennaio 2020, Li fu intimato dalla polizia di non diffondere false notizie online.

Li fu infettato dal nuovo virus e morì poco dopo, il 7 febbraio all'età di 33 anni. Il virus scoperto in Cina, si diffuse rapidamente, fino a diventare un'emergenza sanitaria a livello globale.

A causa dell'elevata contagiosità del virus, e alla necessità di cure in terapia intensiva per il numero elevato di ricoveri, gli ospedali andarono sotto pressione e i vari governi del mondo misero in atto diverse misure restrittive per contenere il propagarsi del virus, quando l'epidemia si trasformò in pandemia.

Il seguente lavoro di tesi, in collaborazione con il Professor Davide Tosi Ricercatore dell'Università degli Studi dell'Insubria di Varese, autore di diversi articoli scientifici[8][7][9], ha lo scopo di illustrare mediante dei grafici interattivi l'evoluzione in tempo reale dei dati dell'epidemia in Italia, e rappresenta il prosieguo del lavoro di tesi intrapreso dalla Dottorssa Alice Schiavone sul calcolo degli indici di contagio $R_0(t)$ [14].

3

Obiettivi

L'obiettivo di questo progetto è la realizzazione di una dashboard in grado di mostrare le informazioni sulla pandemia di Coronavirus Covid-19 in maniera ottimale e immediata.

Il termine *dashboard* comprende un insieme di oggetti grafici che, strutturati e raffigurati in un determinato modo, rendono visibile in tempo reale molte informazioni anche di diversa natura e complessità. La dashboard deve avere le seguenti caratteristiche:

- Rappresentare in modo semplice e chiaro i dati sulla pandemia;
- Aggiornarsi automaticamente con la pubblicazione dei nuovi dati;
- Interattività dei grafici.

4

La visualizzazione dati

4.1 Cos'è la visualizzazione dati

La rappresentazione grafica dei dati o in inglese *data visualization* è l'attività di comunicazione realizzata tramite la proiezione di dati in una forma grafica strutturata.

Mediante l'utilizzo di elementi visivi come grafici o mappe, si ha una lettura dei dati che evidenzia le informazioni rilevanti e consente di riconoscere eventuali valori anomali. Nel mondo dei big data, gli strumenti e le tecnologie di data visualization sono essenziali per analizzare enormi quantità di informazioni e prendere decisioni data oriented.

4.2 Breve storia

Con la tecnologia moderna, la visualizzazione di dati non è mai stata così semplice. Con pochi click si può trasformare un enorme tabella di dati grezzi in un diagram-

ma accattivante e semplice da interpretare. Rappresentare i dati sotto forma di grafici, fornisce una via molto rapida per rendere comprensibile e far capire il proprio punto di vista. Molte tecniche di visualizzazione dati che si usano oggi sono state inventate durante la prima rivoluzione industriale. Quello che può sembrare semplice e ovvio al giorno d'oggi, come un grafico a linea o a barre, sarebbe stato strano e incomprensibile duecento anni fa.

Nel 18esimo secolo uno tra i primi pionieri della visualizzazione dati fu William Playfair, uno statista scozzese, visse nel periodo dell'Illuminismo e che introdusse la rappresentazione grafica in statistica ed economia.

In pochi anni introdusse modi innovativi di rappresentare graficamente i dati, oggi passati alla storia. Nella sua opera del 1786 dal titolo *The Commercial and Political Atlas* incluse 44 grafici, tutti particolarmente *creativi* per l'epoca, come dimostra quello che illustra la differenza tra import ed export dell'economia inglese, nella figura 4.2.

Ai tempi di Playfair, i dati venivano presentati in tabelle asettiche di non facile lettura con poca attenzione per la loro interpretazione. Per comprenderli, non c'erano scorciatoie intuitive, ma un laborioso lavoro di esaminare attentamente i numeri, ricordare, copiare e confrontare i dati.

La figura 4.1 mostra il diagramma inventato da Joseph Priestly dove con una linea temporale, rende immediatamente visibile quali personaggi storici fossero contemporanei.

Playfair fu anche l'inventore del grafico a torta pubblicato per la prima volta nel 1801 nel *Statistical Breviary* per descrivere la distribuzione dell'Impero Ottomano tra i vari continenti europeo, africano e asiatico (nella figura 4.3).

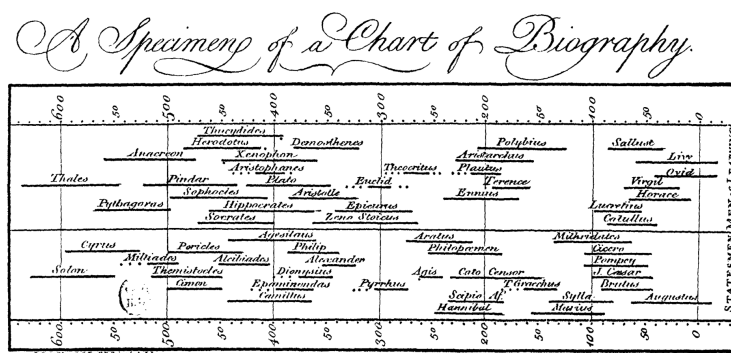


Fig. 4.1: Joseph Priestley — Diagramma temporale

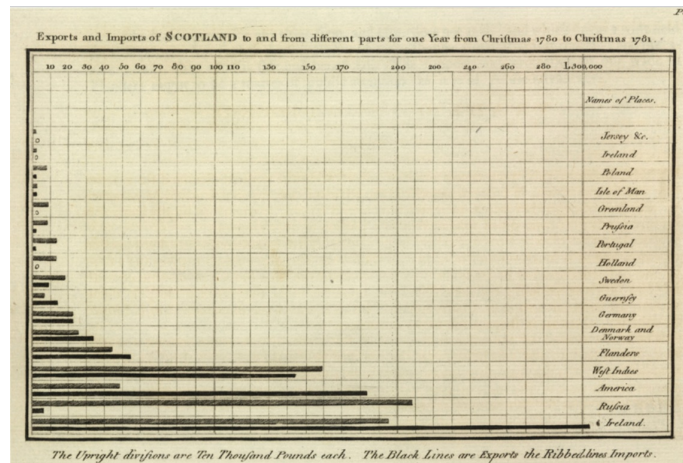


Fig. 4.2: William Playfair - Atlante Politico e Commerciale

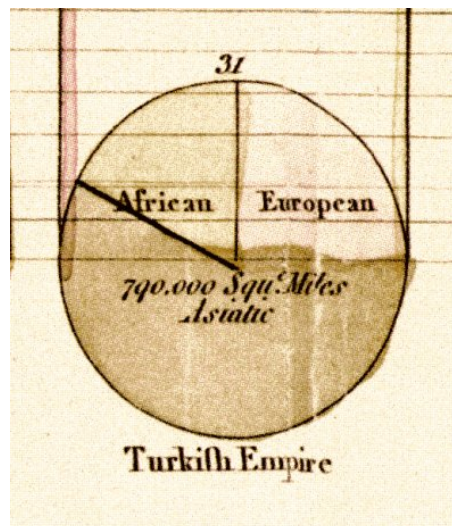


Fig. 4.3: William Playfair — Grafico a torta

5

Design e sviluppo della soluzione

5.1 Origine dei dati

La principale fonte di dati riguardanti la diffusione della pandemia da COVID-19 in Italia è il Dipartimento della Protezione Civile, che giornalmente a partire dal 24 febbraio pubblica i relativi dati sul repository Github[13].

Alcuni dei dati resi disponibili sul repository sono i seguenti:

- Ricoverati con sintomi: numero di soggetti ricoverati non in terapia intensiva.
- Terapia intensiva: Numero di soggetti ricoverati in terapia intensiva.
- Totale ospedalizzati: Totale dei soggetti ricoverati in ospedale.
- Isolamento domiciliare: Soggetti ricoverati presso la propria abitazione.
- Deceduti: Persone decedute positive al virus.
- Tamponi: numero di tamponi effettuati (antigenici + molecolari)



Fig. 5.1: Logo del Dipartimento della Protezione Civile

5.2 Tecnologie utilizzate

Le tecnologie software disponibili per l'analisi, la manipolazione e la visualizzazione dei dati sono molteplici, in questo progetto sono state utilizzate esclusivamente librerie gratuite ed open source in alternativa alle soluzioni commerciali. Di seguito una breve descrizione di ciascuna delle tecnologie utilizzate.

5.2.1 Python

Python è un linguaggio di programmazione dinamico orientato agli oggetti utilizzabile per molti tipi di sviluppo software, è completamente gratuito ed è utilizzabile senza restrizioni di copyright.

Python è un linguaggio portabile sviluppato in ANSI C, è caratterizzato da una sintassi essenziale, ma consente lo sviluppo di applicazioni molto complesse.

È multiplatforma e quindi sfruttabile su diversi sistemi operativi quali Windows, MacOS e GNU/Linux. Offre un forte supporto all'integrazione con altri linguaggi e programmi, dispone di una estesa libreria standard e funzioni matematiche integrate, che semplificano il calcolo dei problemi matematici e l'esecuzione dell'analisi dei dati. Per lo sviluppo del progetto è stato utilizzato interamente il linguaggio Python versione 3.8.



Fig. 5.2: Logo di Python

5.2.2 Plotly

Plotly è una libreria Python open source per la creazione di grafici interattivi, supporta oltre 40 tipologie diverse di grafici che consentono l'applicazione in vari campi scientifici, tra i quali la statistica, la finanza, e la geografia. Plotly basa

il suo funzionamento sulla libreria Javascript plotly.js, che permette agli utenti Python di realizzare stupende applicazioni di visualizzazione dati interattive, che si possono mostrare in un file HTML a sé stante, in un Notebook Jupyter¹, oppure all'interno di una applicazione web utilizzando Dash.



Fig. 5.3: Logo di Plotly

5.2.3 Pandas

Pandas è una libreria Python open source, adatta per l'analisi e la manipolazione di dati in formato sequenziale o tabellare. Pandas consente il Caricamento e il salvataggio di formati standard per dati tabellari, come CSV (Comma-separated Values). Una delle strutture dati utilizzate da Pandas prende il nome di DataFrame, una struttura dati bi-dimensionale, con dimensione variabile che può contenere dati eterogenei.



Fig. 5.4: Logo di Pandas

5.2.4 Dash

Dash è un framework Python ideato per la costruzione di applicazioni web di analytics, si appoggia a Flask², Plotly.js e React.js³. Dash è ideale per la costruzione di applicazioni di visualizzazione dati molto personalizzabili, usando soltanto il linguaggio Python. È particolarmente indicato per chiunque lavori con i dati in Python. Attraverso l'utilizzo di semplici pattern, Dash semplifica notevolmente la costruzione di applicazioni basate su web, come nel caso della dashboard, in

¹Jupyter Notebook è un'applicazione web open source che consente di creare e condividere documenti che contengono codice Python eseguibile in tempo reale, equazioni, visualizzazioni.

²Flask è un framework leggero sviluppato da Armin Ronacher per applicazioni web WSGI (Web Server Gateway Interface)

quanto non è necessario conoscere approfonditamente le tecnologie e i protocolli che ne determinano il funzionamento.



Fig. 5.5: Logo di Dash

5.2.5 Github

GitHub è un servizio web basato su cloud che aiuta gli sviluppatori ad archiviare e gestire il loro codice e a tracciare e controllare le modifiche.

Git è un sistema che consente il controllo delle versioni e permette a più sviluppatori di collaborare allo stesso progetto contemporaneamente. In sostanza Github, semplifica l'utilizzo del software a riga di comando Git per il controllo delle versioni e la collaborazione.



Fig. 5.6: Logo di Github

5.2.6 Docker

Docker è un progetto open source nato con lo scopo di automatizzare la distribuzione di applicazioni sotto forma di “contenitori” detti *container*. I container hanno al loro interno tutto il necessario per la corretta esecuzione dell'applicazione, incluse librerie, strumenti di sistema, codice e runtime.

Un container è basato su un'immagine che rappresenta tutto il suo contenuto e definisce quale applicativo sarà eseguito al suo interno. Per il progetto è stato scelto Docker in quanto consente di isolare una dashboard dalle altre e pertanto semplifica il processo di distribuzione e aggiornamento.

³React è una libreria JavaScript per la creazione di interfacce utente.

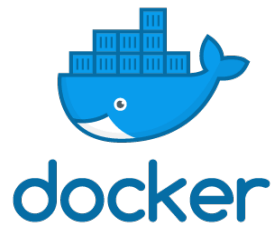


Fig. 5.7: Logo di Docker

5.3 Lo sviluppo

Lo sviluppo del progetto è iniziato partendo dai dati contenuti in un foglio di calcolo Excel fornito dal professore, dal quale sono state estrapolate le formule necessarie per l'elaborazione degli stessi atti alla costruzione della dashboard.

5.3.1 Funzionamento

La dashboard creata è in grado di prelevare i dati direttamente dal file in formato CSV presente nel repository Github fornito dalla protezione civile. La dashboard ha la caratteristica di essere stateless, cioè senza memoria, pertanto i dati non vengono memorizzati localmente, ma prelevati dal repository ad ogni caricamento della pagina.

5.3.2 Calcolo dei dati

I dati prelevati vengono memorizzati all'interno di un dataframe di Pandas, su cui avvengono le elaborazioni necessarie al fine della visualizzazione nei grafici, che consistono in:

- Calcolo decessi giornalieri;
- Calcolo tamponi giornalieri;
- Calcolo dei casi normalizzati;
- Calcolo del rapporto dei nuovi casi meno i casi testati;
- Calcolo delle medie mobili

In figura 5.8 sono mostrate le istruzioni relative ai calcoli.

```
# data calculation
df['nuovi_decessi'] = df.deceduti.diff().fillna(df.deceduti)

# normalized cases
df['delta_tamponi'] = df.tamponi.diff().fillna(df.tamponi)
df['tamp_norm'] = MIN_DELTA_TAMP / df['delta_tamponi'] * df['nuovi_positivi']
df['nuovi_casi_norm'] = df['nuovi_positivi'] * REF_TAMP / df['delta_tamponi']

# ratio cases - tests
df['delta_casi_testati'] = df.casi_testati.diff().fillna(df.casi_testati) # U
df['tamponi_meno_casi_testati'] = df['tamponi'] - df['casi_testati'] # S
df['delta_tamponi_casi'] = df.tamponi_meno_casi_testati.diff().fillna(df.tamponi_meno_casi_testati)
df['rapp_casi_test'] = (df['nuovi_positivi'] / df['delta_casi_testati']) * 100
df['perc_tamponi_meno_testati'] = (df['nuovi_positivi'] / df['delta_tamponi_casi']) * 100

# averages
df['terapia_intensiva_avg'] = df['terapia_intensiva'].rolling(7).mean()
df['nuovi_positivi_avg'] = df['nuovi_positivi'].rolling(7).mean()
df['nuovi_decessi_avg'] = df['nuovi_decessi'].rolling(7).mean()
df['totale_ospedalizzati_avg'] = df['totale_ospedalizzati'].rolling(7).mean()
df['nuovi_casi_norm_avg'] = df['nuovi_casi_norm'].rolling(7).mean()
df['rolling_tested'] = df['rapp_casi_test'].rolling(7).mean()
df['rolling_swabs_tested'] = df['perc_tamponi_meno_testati'].rolling(7).mean()
```

Fig. 5.8: Codice Python per il calcolo dei dati

5.3.3 Diagramma di attività

Un diagramma di attività UML è un diagramma di flusso che rappresenta il flusso del programma da un'attività all'altra, da un punto iniziale (punto nero pieno) a un punto finale (punto nero cerchiato). Nella figura 5.9 è raffigurato il diagramma delle attività per la dashboard.

5.3.4 Organizzazione file

Una applicazione Dash è composta essenzialmente da un singolo programma Python. La cartella *assets* contiene il file *bootstrap.min.css*, una versione minimale del framework Bootstrap,⁴ questo assicura dei tempi di caricamento della pagina decisamente più rapidi. Nella cartella sono contenuti anche i loghi delle regioni, della dashboard regioni. (nella figura 5.12)

5.4 Problemi riscontrati

I dati forniti dal Dipartimento della Protezione Civile non sono esenti da errori, ciò può essere dovuto a errori di trascrizione oppure a causa di comunicazioni parziali

⁴Bootstrap è un framework CSS utilizzato per la semplificare la gestione dello stile della dashboard

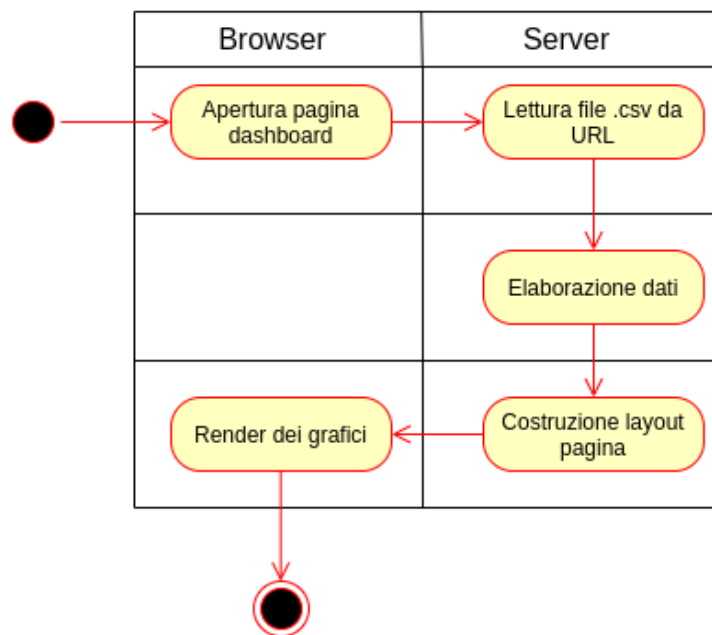


Fig. 5.9: Diagramma di attività

```

italy
├── assets
│   └── bootstrap.min.css
└── dash_italy.py
  
```

Fig. 5.10: Struttura file dashboard Italia

```

lombardia
├── assets
│   └── bootstrap.min.css
└── dash_
    └── lombardia.py
  
```

Fig. 5.11: Struttura file dashboard Lombardia

```

regioni
├── assets
│   ├── bootstrap.min.css
│   └── img
│       ├── Abruzzo.png
│       ├── Basilicata.png
│       ├── Calabria.png
│       └── ...
└── dash_regioni.py
  
```

Fig. 5.12: Struttura file dashboard regioni

o in ritardo dal parte delle regioni. Alcune serie di dati, come ad esempio quelli cumulativi come guariti, deceduti, totale casi e tamponi, che dovrebbero essere strettamente crescenti, alcune volte non lo sono.

5.4.1 Errore nei tamponi

Il 17 dicembre 2020 è stato registrato un valore errato nel numero cumulativo dei tamponi, ciò ha causato un'anomalia nel grafico dei casi normalizzati.

Per la correzione di questo problema è stato effettuato un intervento di pulizia del dato. Il dato "incriminato" è stato sostituito con la media tra il giorno precedente e quella del giorno successivo.

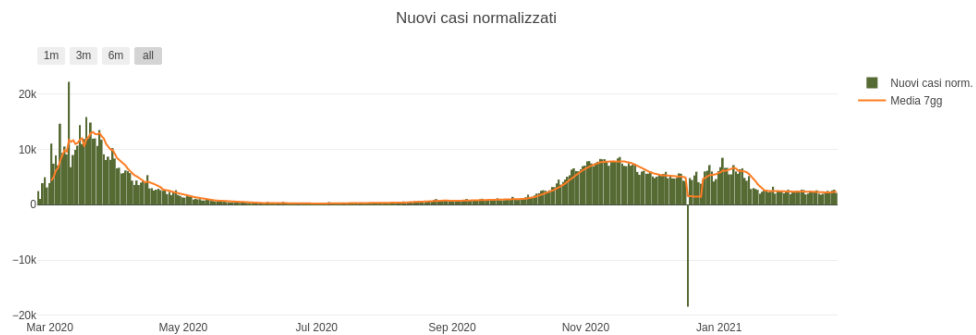


Fig. 5.13: Problema causato dal valore errato dei tamponi

6

Le dashboard

In questo capitolo, a titolo di esempio verranno presentate delle immagini relative alle schermate estratte dall'applicazione. È possibile visualizzare l'andamento dei contagi riferito a 1, 3, 6 mesi e per tutto il periodo di riferimento a partire da marzo 2020. Tutti i grafici contenuti nella dashboard, sono interattivi. Passando con il cursore del mouse sopra al grafico, vengono evidenziate le barre che indicano il numero dei contagi per quel determinato giorno. Sul grafico è possibile agire sull'intervallo di tempo selezionando uno dei pulsanti indicati nel riquadro 1 della figura, oppure cliccando sulla legenda del riquadro 2, cambiare la configurazione delle tracce.

6.1 Dashboard Italia

Nella dashboard Italia sono presenti nove diversi grafici che illustrano l'andamento della pandemia attraverso le curve di contagi, le ospedalizzazioni, le terapie intensive e i decessi.

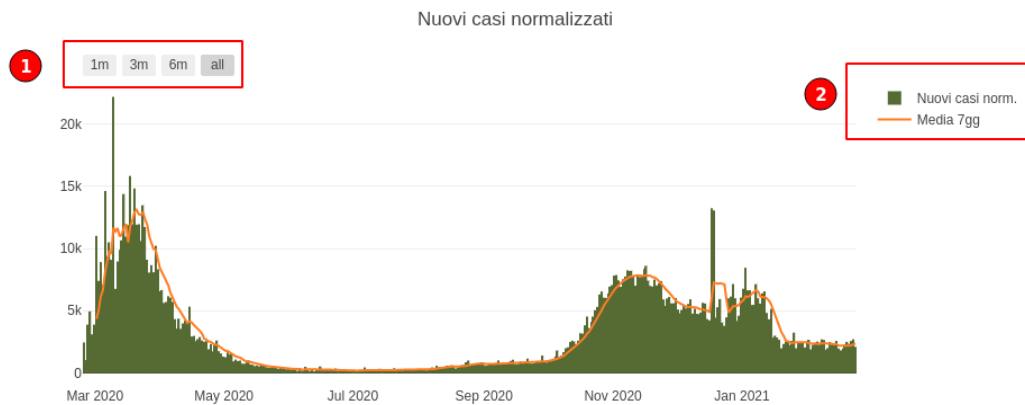


Fig. 6.1: Esempio grafico

6.1.1 Nuovi casi

La figura 6.2 mostra il numero di nuovi casi giornalieri in Italia a partire dall'inizio della pandemia. In ascissa sono riportati i giorni, e in ordinata il numero dei nuovi casi riportati in quel determinato giorno.

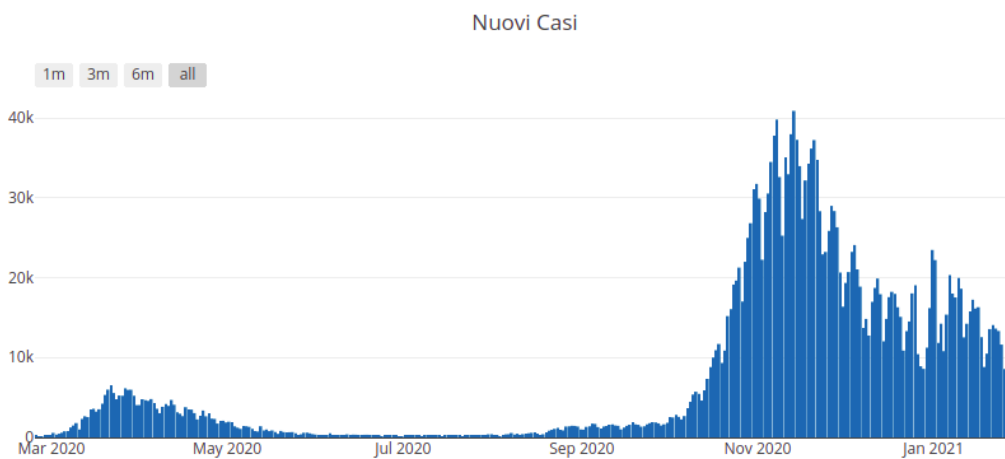


Fig. 6.2: Nuovi casi

6.1.2 Totale casi

Nella figura 6.3 sono rappresentati i casi totali delle persone che hanno contratto il virus, che comprende gli attuali positivi, i guariti e i decessi.

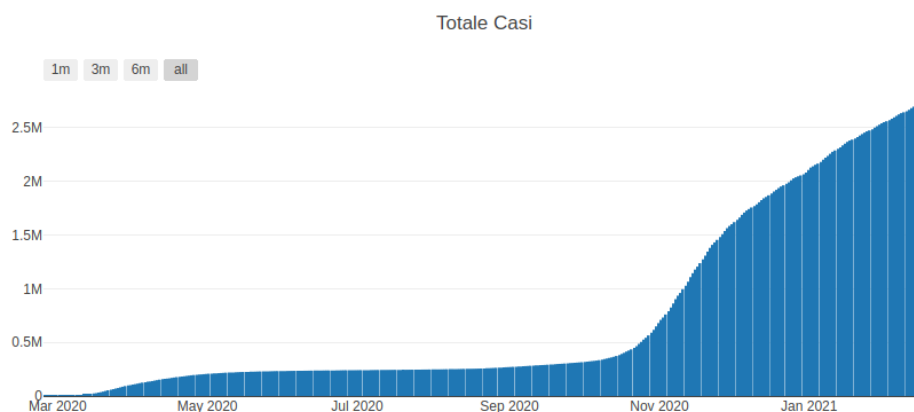


Fig. 6.3: Totale casi

6.1.3 Isolamento domiciliare

Nella figura 6.4 è raffigurato il grafico delle persone che si trovano in isolamento presso il loro domicilio.

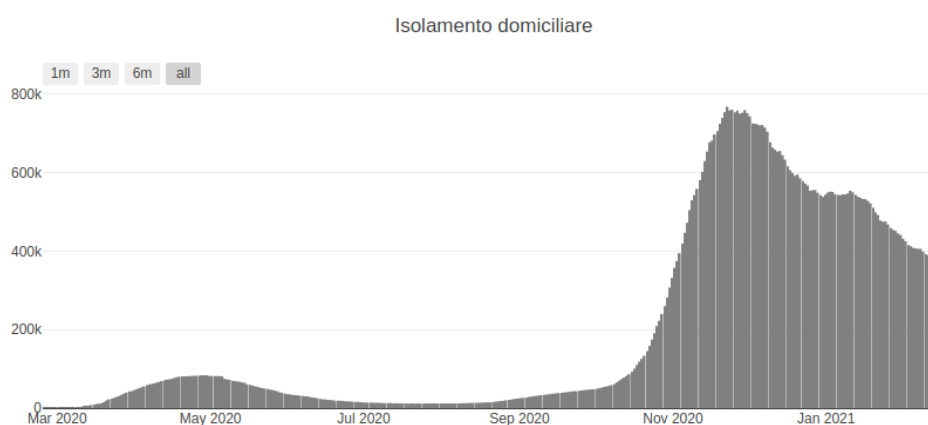


Fig. 6.4: Isolamento domiciliare

6.1.4 Terapia intensiva

La figura 6.5 mostra l'andamento dei pazienti ricoverati in ospedale in terapia intensiva dall'inizio della pandemia. La linea indica la media mobile a 7 giorni.

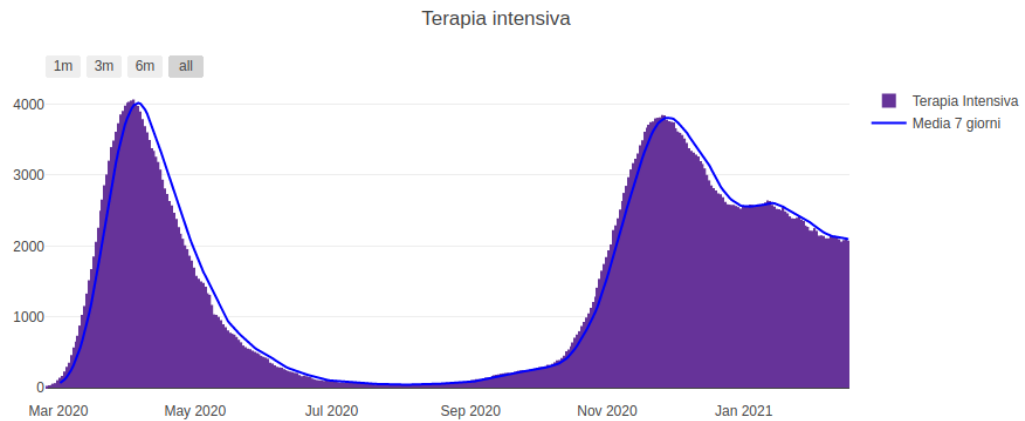


Fig. 6.5: Terapia intensiva

6.1.5 Casi normalizzati

Il grafico nella figura 6.6 mostra il numero di casi normalizzati, ossia resi indipendenti dai tamponi effettuati ogni giorno. In arancione è indicata anche la linea della media mobile a 7 giorni.

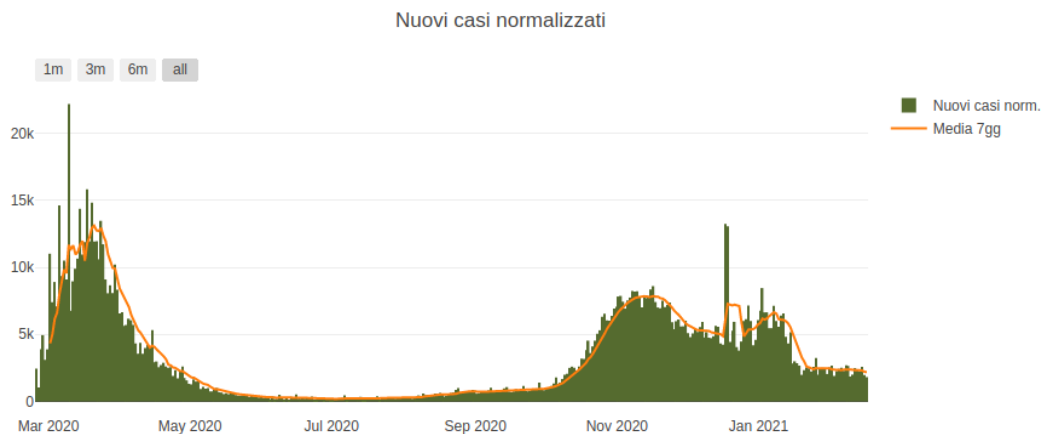


Fig. 6.6: Casi normalizzati

6.1.6 Terapia intensiva e ospedalizzati

Nel grafico della figura 6.7 è possibile osservare l'andamento delle persone ospedalizzate, cioè la somma dei ricoverati con sintomi e di quelli in terapia intensiva.

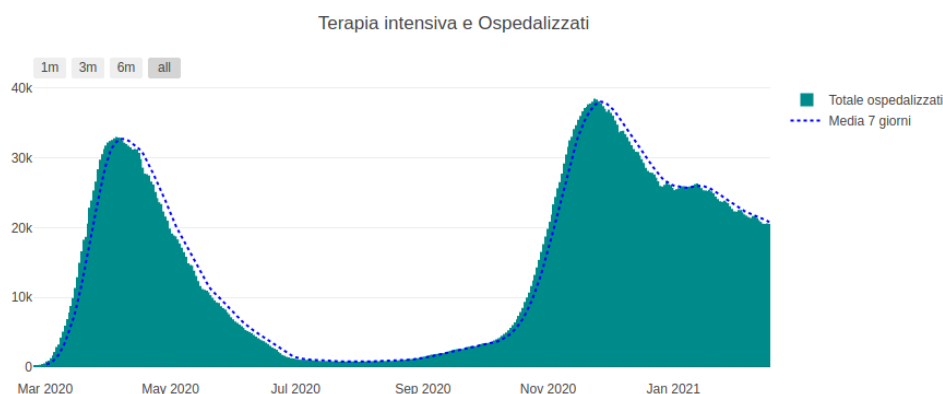


Fig. 6.7: Terapia intensiva e ospedalizzati

6.1.7 Decessi e contagi giornalieri

Nella figura 6.8 è rappresentato l'andamento dei decessi (linea blu punteggiata) e dei contagi giornalieri (linea arancione punteggiata). Le linee blu e arancione sono rispettivamente la media mobile a 7 giorni per decessi e contagi giornalieri.

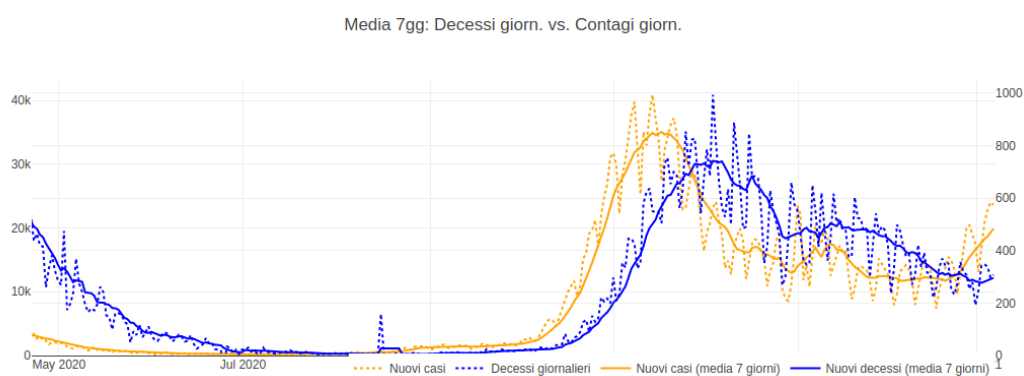


Fig. 6.8: Andamento di decessi e contagi

6.1.8 Percentuale dei casi positivi sui casi testati

Nella figura 6.9 si può osservare l'andamento della percentuale dei casi positivi sui casi testati con tamponi.

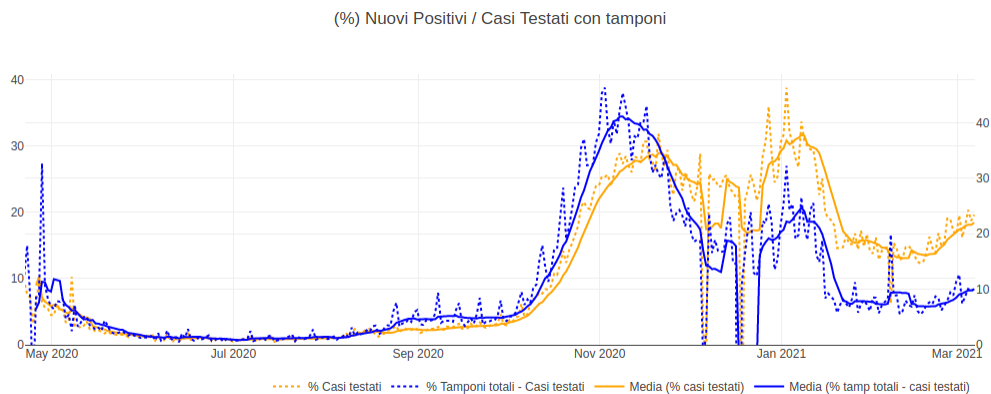


Fig. 6.9: Percentuale dei nuovi positivi sui casi testati

6.1.9 Nuovi casi e decessi

Nella figura 6.10 viene mostrato l'andamento dei positivi e dei decessi. In questo grafico è anche possibile selezionare l'intervallo temporale, agendo sulla barra in basso.

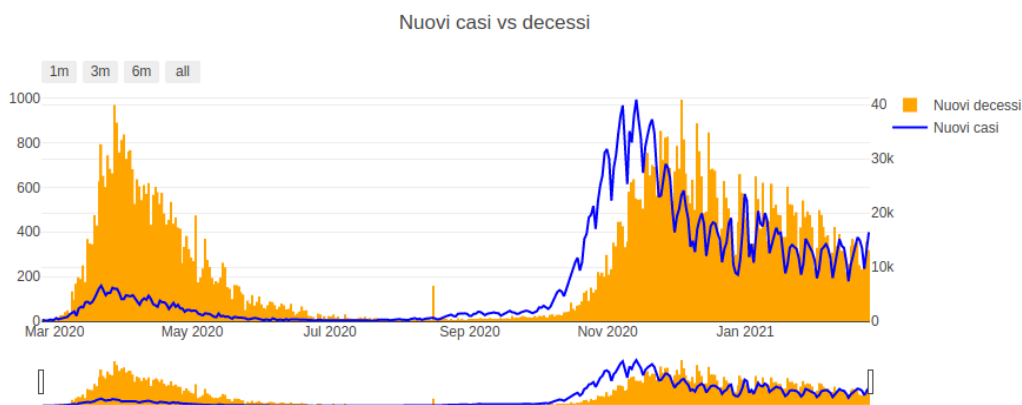


Fig. 6.10: Nuovi casi e nuovi decessi

6.2 Dashboard Lombardia

6.2.1 Andamento dei contagi

Nella figura 6.11 è raffigurato l'andamento dei contagi nella regione Lombardia. In giallo la linea della media mobile a 7 giorni.



Fig. 6.11: Andamento dei contagi in Lombardia

6.2.2 Percentuale positivi sui casi testati

La figura 6.12 rappresenta la percentuale del rapporto tra i nuovi casi rispetto al totale dei casi e al numero dei casi testati tramite tamponi. La linea gialla indica l'andamento dei *nuovi casi testati*, mentre quella blu indica il *totale dei casi testati*. Le linee "punteggiate" indicano le medie mobili a 3 giorni.

6.2.3 Casi normalizzati

Nella figura 6.13 è possibile osservare l'andamento dei casi normalizzati nella regione Lombardia.

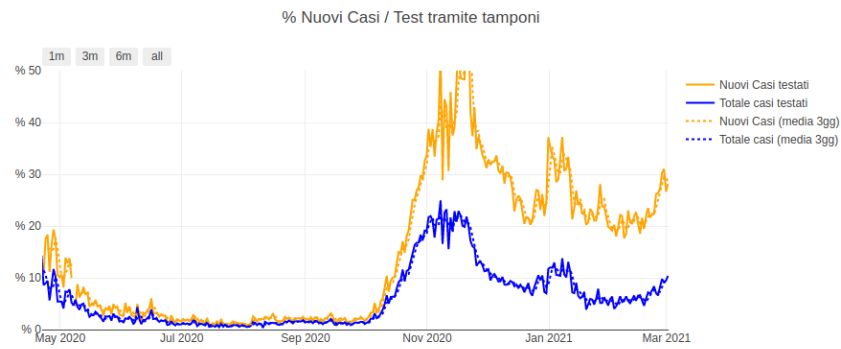


Fig. 6.12: Andamento dei contagi in Lombardia

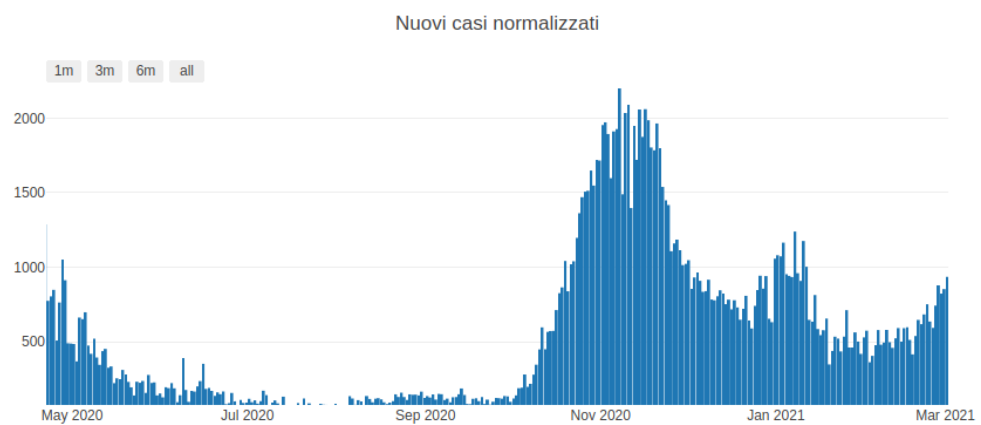


Fig. 6.13: Andamento dei casi normalizzati

6.2.4 Ospedalizzazioni

Nella figura 6.14 si può osservare l'andamento delle ospedalizzazioni nella regione Lombardia. In arancione la linea della media mobile a 7 giorni.

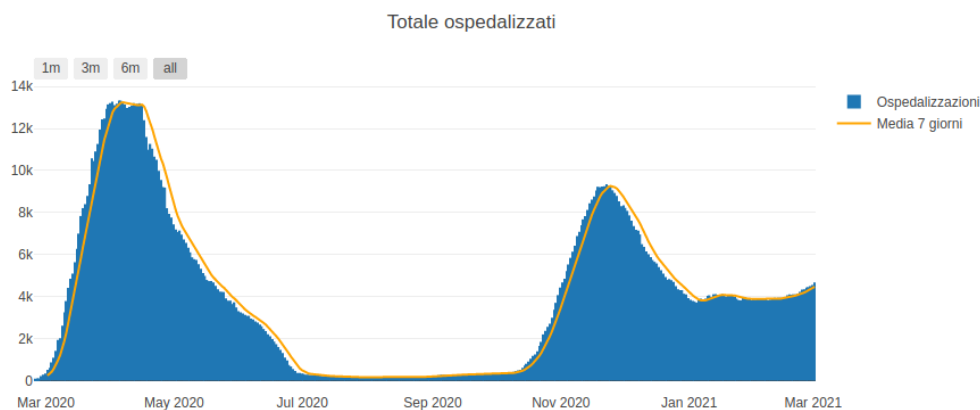


Fig. 6.14: Andamento delle ospedalizzazioni in Lombardia

6.2.5 Decessi giornalieri

La figura 6.15 mostra l'andamento dei decessi nella regione Lombardia. La linea blu indica la media mobile a 7 giorni.

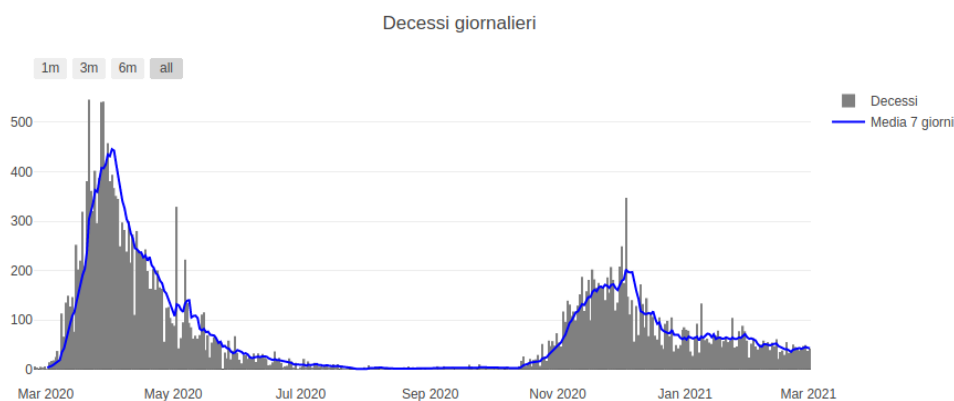


Fig. 6.15: Andamento dei deceduti

6.2.6 Terapia intensiva

La figura 6.16 mostra l'andamento dei pazienti ricoverati negli ospedali lombardi in terapia intensiva dall'inizio della pandemia.

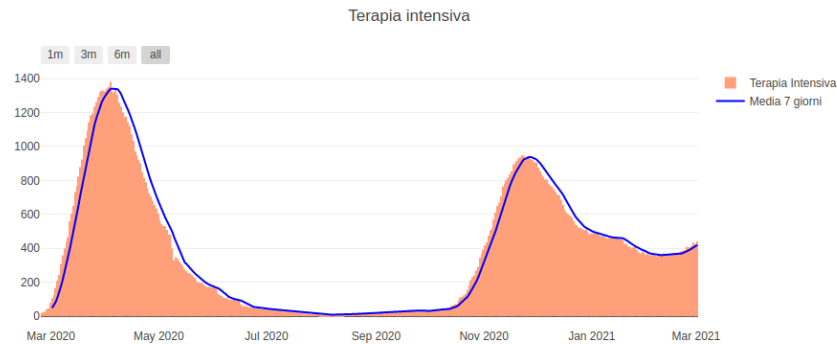


Fig. 6.16: Andamento della terapia intensiva

6.3 Dashboard Regioni

Nella dashboard regioni, selezionando dal menù a tendina il nome della regione d'interesse (nella figura 6.17), è possibile visualizzarne i dati. I grafici contenuti all'interno della dashboard regioni sono analoghi ai precedenti, e riguardano:

- Andamento dei contagi
- Percentuale nuovi casi sui tamponi effettuati
- Totale ospedalizzati
- Decessi giornalieri
- Terapia intensiva



Fig. 6.17: Selezione della regione

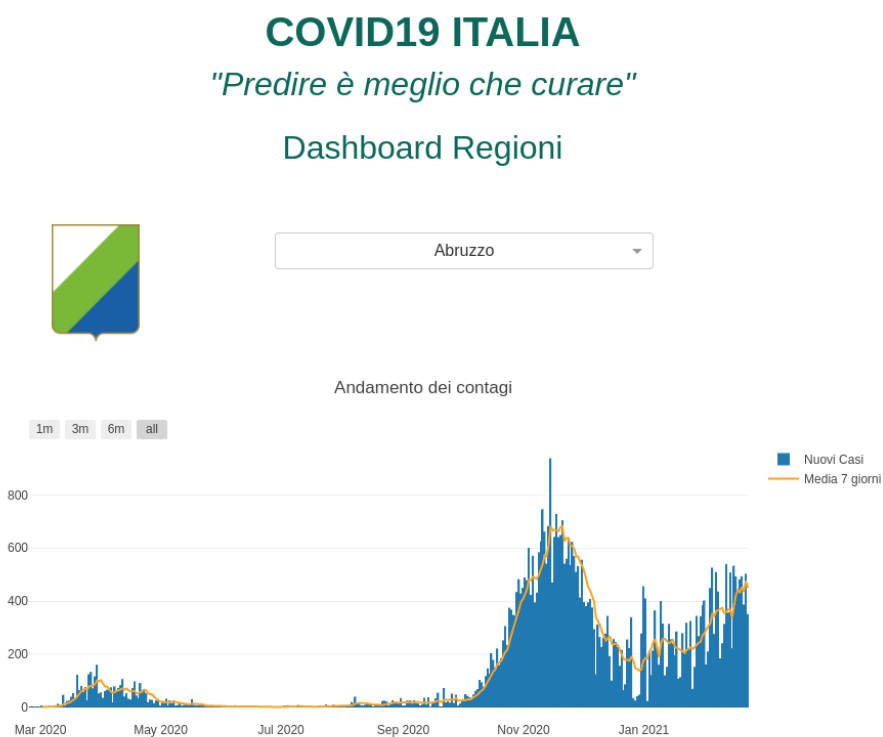


Fig. 6.18: Visualizzazione della regione Abruzzo

7

Il portale Covid19-Italy

Con il patrocinio dell'Università degli studi dell'Insubria è stato ideato il portale Covid 19-Italy.it¹, un sito che permette di visualizzare i dati aggregati e ufficiali sull'andamento della pandemia regione per regione e a livello nazionale, attraverso grafici e confronti. Sulla pagina principale del sito sono presenti 4 card, che permettono di accedere alle seguenti pagine:

- *Italy R0(t) Indices* permette di accedere alla pagina sulla stima degli indici di riproduzione $R_0(t)$ usando il metodo Wallinga and Teunis, a livello nazionale, regionale e provinciale. (in figura 7.2)
- *Italy dashboard* visualizza il monitoraggio del Covid-19 in Italia, attraverso 9 diversi grafici, sempre aggiornati con la pubblicazione di nuovi dati;
- *Lombardy dashboard* visualizza l'andamento nella pandemia in dettaglio nella regione Lombardia;
- *Region dashboard* mostra l'andamento della pandemia in una regione italiana a scelta.

¹<https://covid19-italy.it>

COVID19 ITALIA

"Predire è meglio che curare"



Italy R0(t) Indices

View the page with the R0(t) of the Italian regions and provinces.

[View now all R0\(t\)](#)



Italy Dashboard

Access the dashboard to monitor the Covid19 situation in Italy.

[Go to the dashboard](#)



Lombardy Dashboard

Access the dashboard to monitor the Covid19 situation in Lombardy.

[Go to the dashboard](#)



Regions Dashboard

Access and Select the dashboard of your Italian Region.

[Go to the dashboard](#)

Fonte dati: ISS / Dip. Protezione Civile.

Authors: Prof. Davide Tosi, Alessandro Riva, Alice Schiavone, Università degli Studi dell'Insubria.

Fig. 7.1: Pagina principale del sito

Reproduction Number Estimate $R_0(t)$ in Italian Regions and Provinces

Reproduction Number Estimate $R_0(t)$ using the Time-Dependent method (Wallinga and Teunis).
Data are taken from the Department of Italian Civil Protection for the 2020 SARS-Cov2 pandemic in Italy.
Indexes and Charts are automatically updated on a weekly basis.

Authors:

Davide Tosi (Prof. Tenure Track at Università degli Studi dell'Insubria)
Alice Schiavone (Thesis Student at Università degli Studi dell'Insubria)
Alessandro Riva (Thesis Student at Università degli Studi dell'Insubria)

Update: Mar. 07, 2021

Note: Starting from 15 Jan. 2021, the Civil Protection reports in the daily positive count also those who test positive in the antigen rapid test. This explains the drastic drop in the R_t indices.

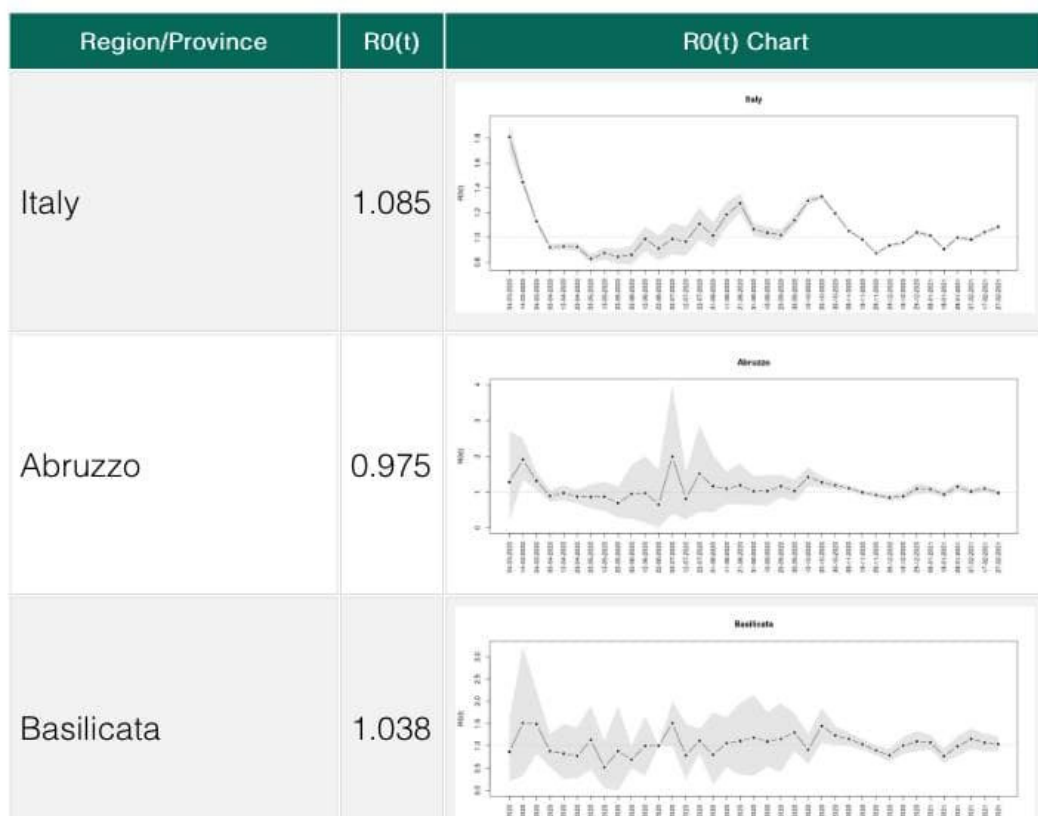


Fig. 7.2: Pagina degli indici $R_0(t)$

7.1 Rassegna stampa

Sono stati pubblicati numerosi articoli riguardanti il portale *Covid19-italy.it*. Alcuni sono riportati di seguito in ordine cronologico.

29 ottobre 2020:

- *Covid: online nuovo portale grafici indice $R0$ in Lombardia*[6]
- *La pandemia spiegata con numeri e grafici nel portale dell'Insubria*[12]

14 novembre 2020:

- *Covid: "La discesa dei contagi inizierà a dicembre"*[5]

19 novembre 2020:

- *Covid, Davide Tosi (Insubria) e Massimo Galli insieme per la divulgazione scientifica*[4]

14 febbraio 2021

- *"Ho raccontato il Covid con i numeri. Che non mentono mai"*[1]

7.2 Statistiche

Il portale *covid19-italy.it* ha riscosso sin da subito un discreto successo, sono stati registrati mediamente 200 utenti attivi al giorno per un totale di quasi 19.000 visitatori, nel periodo compreso tra la fine di settembre 2020 e l'inizio di marzo 2021.



Fig. 7.3: Statistiche da Google Analytics

8

Conclusioni

Lo scopo dell'elaborato di tesi era quello di illustrare la struttura dell'applicazione creata per la visualizzazione dei grafici sulla pandemia di Covid-19.

Gli obiettivi prefissati sono stati raggiunti con successo attraverso la realizzazione di una dashboard utilizzando il linguaggio open source Python 3.

A causa dell'emergenza sanitaria tutt'ora in corso, il tirocinio si è svolto a distanza, in modalità *smart working*.

Questa situazione non ne ha condizionato in modo negativo lo svolgimento, in quanto sono sempre rimasto in contatto con il mio tutor.

Lavorare su questo progetto mi ha permesso di acquisire nuove conoscenze e mettere in pratica le mie abilità. La pandemia di Covid19 ha stravolto le nostre vite, ha cambiato il modo di lavorare e comunicare.

Appendice



Manuale utente

A.1 Esecuzione in locale

A.1.1 Requisiti software

Il progetto è stato realizzato utilizzando l'*ambiente di sviluppo integrato* (IDE) Pycharm Community. Per l'esecuzione della dashboard sono richieste le seguenti dipendenze.

- Git
- Python 3.8.5
- Pandas
- Plotly
- Dash
- Dash Bootstrap Components

A.1.2 Clonazione del repository

È possibile scaricare il progetto direttamente dal mio repository Github, con il comando `git clone https://github.com/alex27riva/Covid-dashboard`

A.1.3 Installazione dipendenze

Per installare le librerie richieste è sufficiente lanciare i seguenti comandi dal terminale integrato in Pycharm.

- `$ pip install pandas`
- `$ pip install plotly==4.12.0`
- `$ pip install dash==1.17.0`
- `$ pip install dash-bootstrap-components`

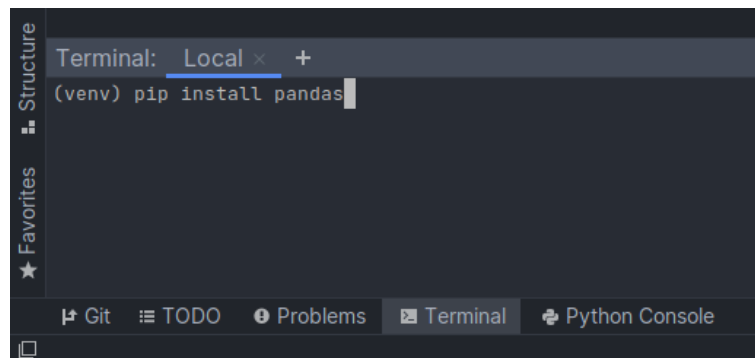


Fig. A.1: Terminale integrato in Pycharm

Ad installazione completata, sarà possibile avviare il progetto.

A.1.4 Avvio del programma

Per l'avvio del programma, cliccare sul pulsante con la freccina verde (nella figura A.2). Verrà avviato un server locale, il cui suo URL verrà mostrato nella console di debug a cui si potrà visualizzare la pagina della dashboard.

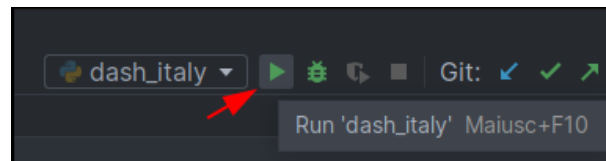


Fig. A.2: Avvio del programma

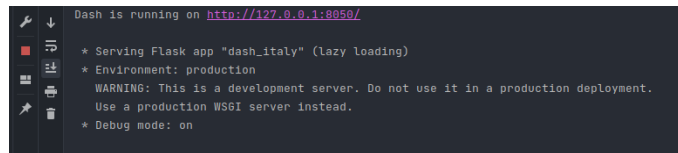


Fig. A.3: Programma Python in esecuzione

A.2 Configurazione della VPS

È necessario disporre di un server privato virtuale (VPS) con sistema operativo Ubuntu Server.

A.2.1 Creazione account utente

Per questioni di sicurezza è meglio non utilizzare l'account privilegiato *root*.

1. Collegarsi alla VPS tramite SSH `ssh root@<ip-vps>`
2. Creare l'account per l'utente admin `useradd -m admin`
3. Impostare una password per l'utente admin `passwd admin`
4. Chiudere la connessione con il comando `exit`

A.2.2 Installazione di Docker

1. Collegarsi alla VPS con l'account admin `ssh admin@<ip-vps>`
2. Scaricare lo script per l'installazione di Docker `curl -fsSL https://get.docker.com -o get-docker.sh`
3. Eseguire lo script con `sudo sh get-docker.sh`

4. Eseguire questo comando per interagire con Docker senza privilegi di root
`sudo usermod -aG docker admin`

A.2.3 Installazione e avvio dei container

I container sono presenti nel repository di Docker, eseguire i seguenti comandi per l'installazione e avvio delle dashboard.

- `docker run -d -p 8050:8050 --name=dash_italy alex27riva/dash_italy`
- `docker run -d -p 8051:8050 --name=dash_lomb alex27riva/dashboard_-lombardia`
- `docker run -d -p 8052:8050 --name=dash_regioni alex27riva/dashboard_-regioni`

A.2.4 Portainer

Per una gestione più semplice dei container Docker si installare *Portainer*, un'interfaccia web che facilita l'avvio, aggiornamento, rimozione dei container. Per installare Portainer, eseguire questi due comandi:

1. `docker volume create portainer_data`
2. `docker run -d -p 8000:8000 -p 9000:9000 --name=portainer --restart=always -v /var/run/docker.sock:/var/run/docker.sock -v portainer_data:/data portainer/portainer-ce`

Ora è possibile accedere all'interfaccia web di *Portainer* all'indirizzo `http://<ip-vps>:9000`

A.2.5 Ngnix

Ngnix è un web server, che è stato utilizzato come reverse proxy, ciò permette di raggiungere la dashboard attraverso un URL, senza specificare un numero di porta differente.

Installazione di Ngnix

1. `sudo apt update`
2. `sudo apt install nginx`
3. `unlink /etc/nginx/sites-enabled/default`
4. `sudo nano /etc/nginx/sites-available/reverse-proxy.conf`

```
server
{
    listen 80;
    listen 443;
    listen [::]:80;
    ssl on;
    ssl_certificate /etc/letsencrypt/live/dash.covid19-italy.it/fullchain.pem;
    ssl_certificate_key /etc/letsencrypt/live/dash.covid19-italy.it/privkey.pem;

    server_name dash.covid19-italy.it

    access_log /var/log/nginx/reverse-access.log;
    error_log /var/log/nginx/reverse-error.log;

    location /italy/ {
        proxy_pass http://localhost:8050;
    }

    location /lombardy/ {
        proxy_pass http://localhost:8051;
    }

    location /regions/ {
        proxy_pass http://localhost:8052;
    }
}
```

Creazione del link simbolico

```
$ ln -s /etc/nginx/sites-available/reverse-proxy.conf  
/etc/nginx/sites-enabled/reverse-proxy.conf
```

Test della configurazione

```
nginx -t
```

A.2.6 Installazione certificato SSL

1. Assicurarsi che snap sia aggiornato `sudo snap install core; sudo snap refresh core`
2. Installare Certbot `sudo snap install --classic certbot`
3. Creazione simlink per l'eseguibile `sudo ln -s /snap/bin/certbot /usr/bin/certbot`
4. Richiesta del certificato `sudo certbot --nginx`

Riavvio di Nginx

Riavviare il servizio di Nginx con il seguente comando:
`sudo systemctl restart nginx`

Colophon

Questa tesi è stata scritta in L^AT_EX utilizzando Overleaf,¹ un editor collaborativo basato su cloud utilizzato per scrivere, modificare e pubblicare documenti scientifici.

¹<https://overleaf.com>

Bibliografia

- [1] "Ho raccontato il Covid con i numeri. Che non mentono mai". URL: <https://www.ilgiorno.it/notizie/unannodicovid/covid-curve-contagio-1.6031412>.
- [2] *A Short History of Data Visualisation*. URL: <https://towardsdatascience.com/a-short-history-of-data-visualisation-de2f81ed0b23>.
- [3] Christopher V Cosgriff. «Data sharing in the era of COVID-19». In: *The Lancet* (). DOI: [https://doi.org/10.1016/S2589-7500\(20\)30082-0](https://doi.org/10.1016/S2589-7500(20)30082-0).
- [4] *Covid, Davide Tosi (Insubria) e Massimo Galli insieme per la divulgazione scientifica*. URL: <https://www.corrieredicomo.it/covid-davide-tosi-insubria-e-massimo-galli-insieme-per-la-divulgazione-scientifica>.
- [5] *Covid: "La discesa dei contagi inizierà a dicembre"*. URL: <https://www.ilgiorno.it/varese/cronaca/covid-curva-contagi-1.5715611>.
- [6] *Covid: online nuovo portale grafici indice R0 in Lombardia*. URL: https://www.ansa.it/lombardia/notizie/2020/10/29/covid-online-nuovo-portale-grafici-indice-r0-in-lombardia_3822b218-3b5d-48bf-a02d-4aa327ade61a.html.
- [7] A. Campi D. Tosi. «How Data Analytics and Big Data can Help Scientists in Managing COVID-19 Diffusion: A Model to Predict the COVID-19 Diffusion in Italy and Lombardy Region.» In: *Int. Journal of Medical Internet Research* (2020). DOI: 10.2196/21081.
- [8] M. Chiappa D. Tosi. «Understanding the geographical spread of COVID-19 in relation with Goods Regional Routes and Governmental Decrees: The Lombardy Region Case Study.» In: *Int. Journal of Medical Internet Research* (2020). DOI: 10.13140/RG.2.2.10634.18887.

-
- [9] M. Verde D. Tosi A. Verde. «Clarification of Misleading Perceptions of COVID-19 Fatality and Testing Rates in Italy: Data Analysis. Int. Journal of Medical Internet Research.» In: *Int. Journal of Medical Internet Research* (2020).
- [10] *Data sharing in the era of COVID-19*. 2020. URL: [https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(20\)30082-0/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30082-0/fulltext).
- [11] *Getting Started with Plotly in Python*. URL: <https://plotly.com/python/getting-started/>.
- [12] *La pandemia spiegata con numeri e grafici nel portale dell'Insubria*. URL: <https://www.varesenews.it/2020/10/la-pandemia-spiegata-numeri-grafici-nel-portale-dellinsubria/975336/>.
- [13] *pcm-dpc/COVID-19: COVID-19 Italia - Monitoraggio situazione*. URL: <https://github.com/pcm-dpc/COVID-19>.
- [14] Alice Schiavone. «Development of a tool to study COVID-19 contagiousness via $R_0(t)$ in Italy at national, regional and provincial level». URL: https://aliswh.github.io/docs/thesis_2020.pdf.
- [15] *Sito ufficiale Python*. URL: <https://www.python.org/>.