

# Università Politecnica delle Marche

Facoltà di Ingegneria

Dipartimento di Ingegneria dell'Informazione

Corso di Laurea in Ingegneria Informatica e dell'Automazione

---



**Tesi di Laurea**

**Progettazione e implementazione di una campagna di Data  
Science per valutare l'andamento della pandemia da  
COVID-19 nella Regione Marche**

**Design and implementation of a Data Science campaign to  
evaluate the trend of COVID-19 pandemics in the Marche  
Region**

Relatore

Prof. Domenico Ursino

Candidata

Silvia Ciuffreda

---

**Anno Accademico 2020-2021**



*Lentamente muore chi non capovolge il tavolo,  
chi è infelice sul lavoro,  
chi non rischia la certezza per l'incertezza,  
chi rinuncia ad inseguire un sogno,  
chi non si permette almeno una volta  
di fuggire ai consigli sensati.*

*M.Medeiros*





---

# Indice

<b>Introduzione</b> .....	3
<b>1 Pandemia da COVID-19</b> .....	7
1.1 Il virus e la pandemia .....	7
1.1.1 Origine del virus .....	8
1.1.2 La diffusione della pandemia .....	9
1.2 La seconda ondata di contagi da COVID-19 .....	9
1.2.1 La situazione in Italia .....	11
1.2.2 Prima e seconda ondata a confronto .....	12
1.3 I vaccini anti-COVID-19 .....	13
1.3.1 Sperimentazione, valutazione, approvazione .....	14
1.3.2 Quadro della campagna vaccinale in Italia .....	14
1.4 La Data Science a supporto della gestione della pandemia .....	15
<b>2 Strumenti software utilizzati</b> .....	17
2.1 Power BI .....	17
2.1.1 Architettura del software .....	18
2.1.2 Il flusso di azioni .....	22
2.2 Connessione al set di dati .....	22
2.2.1 Editor di query .....	23
2.3 Realizzazione di report .....	24
2.3.1 Il foglio di lavoro .....	25
2.3.2 Gli oggetti visivi supportati .....	27
2.3.3 Il pannello Filtri .....	33
2.3.4 Formule DAX .....	35
2.4 Pubblicazione e condivisione di dashboard .....	38
<b>3 Descrizione e costruzione del dataset</b> .....	39
3.1 Sorgenti dei dati .....	39
3.1.1 Dati forniti dalla Regione Marche .....	40
3.1.2 Struttura dei dati .....	44
3.2 Costruzione del dataset .....	51
3.2.1 Clonazione del repository remoto .....	52

## VI **Indice**

3.2.2	Realizzazione dei file JSON .....	52
3.2.3	Script Python per l'aggregazione dei file giornalieri .....	53
3.2.4	Upload dei file creati sul repository .....	53
<b>4</b>	<b>Progettazione e Implementazione dei Task di Data Science .....</b>	<b>55</b>
4.1	Attività di ETL svolte in Power BI .....	55
4.1.1	Importazione dei dati .....	56
4.1.2	Modifica dei dati .....	56
4.1.3	Creazioni di nuove query .....	57
4.1.4	Accodamento di query .....	59
4.1.5	Caricamento dei dati puliti .....	60
4.2	Dashboard panoramico sull'andamento dell'epidemia .....	60
4.2.1	Andamento giornaliero .....	63
4.3	Dashboard introduttivo su tamponi e positivi .....	65
4.3.1	Media settimanale .....	67
4.4	Cruscotto relativi ai malati attivi .....	68
4.4.1	Analisi ricoveri per provincia .....	69
4.4.2	Analisi quarantena domiciliare per provincia .....	72
4.5	Dashboard sui decessi .....	73
4.5.1	Distribuzione per comune .....	75
4.6	Confronto prima e seconda ondata di contagi .....	76
<b>5</b>	<b>Discussione in merito al lavoro svolto .....</b>	<b>81</b>
5.1	Premessa .....	81
5.2	Comunicare con i dati .....	82
5.3	Il software Power BI .....	82
5.4	I formati JSON e CSV per lo scambio di dati .....	83
5.5	L'analisi dei dati da cui estrarre conoscenza .....	84
<b>6</b>	<b>Conclusioni .....</b>	<b>87</b>
	<b>Riferimenti bibliografici .....</b>	<b>91</b>

---

## Elenco delle figure

1.1	Modello 3D del Covid-19	8
1.2	Mutazione del coronavirus SARS-COV-2	9
1.3	Vaccini Pfizer-BioNTech e Moderna	10
1.4	Disposizioni per differente livello di allerta	11
1.5	Infermiera Claudia Alivernini, prima vaccinata in Italia	12
2.1	Logo di Power BI	17
2.2	Il Magic Quadrant di Gartner relativo ai tool di analisi dei dati	18
2.3	L'ecosistema di Power BI	19
2.4	L'architettura di Power BI	20
2.5	Schermata di avvio di Power BI Desktop	21
2.6	Funzionalità fornite da Power BI Desktop e Power BI Service	21
2.7	Finestra di recupero dei dati Power in BI Desktop	23
2.8	Finestra dell'Editor di query	24
2.9	Esempio di una pagina di un report in Power BI	25
2.10	Foglio di lavoro in Power BI Desktop	26
2.11	Icone delle viste di Power BI	26
2.12	Riquadri laterali in visualizzazione <i>Report</i>	27
2.13	Visualizzazione <i>Dati</i>	28
2.14	Visualizzazione <i>Relazioni</i>	28
2.15	Grafici prefiniti in Power BI	29
2.16	Esempio di oggetto visivo "Scheda"	29
2.17	Esempio di istogramma	30
2.18	Esempio di tabella	31
2.19	Esempio di grafico a linee	31
2.20	Esempio di filtro dei dati	31
2.21	Esempio di grafico combinato	32
2.22	Esempio di grafico ad anello	33
2.23	Esempio di mappa a bolle	34
2.24	Esempio di indicatore KPI	34
2.25	Il riquadro "Filtri" nella visualizzazione <i>Report</i>	35
2.26	Il riquadro "Campi" dell'area <i>Report</i> con misure calcolate	36
2.27	Esempio di misura calcolata con DAX	37

## VIII Elenco delle figure

2.28	Esempio di colonna calcolata con DAX . . . . .	37
2.29	Esempio di dashboard pubblicata in Power BI Service . . . . .	38
3.1	Repository di GitHub utilizzato per l'importazione dei dati . . . . .	40
3.2	La pagina del sito della Regione Marche dedicata al Coronavirus . . . . .	41
3.3	Il report <i>Blu</i> pubblicato giornalmente dalla Regione Marche . . . . .	42
3.4	Il report <i>Giallo</i> pubblicato giornalmente dalla Regione Marche . . . . .	43
3.5	Il report <i>Arancio</i> pubblicato giornalmente dalla Regione Marche . . . . .	44
3.6	La directory Covid-19 con i file clonati dal repository di GitHub . . . . .	52
3.7	Gli Script Python utilizzati per l'aggregazione dei file giornalieri . . . . .	54
3.8	Istruzioni da riga di comando per la sincronizzazione tra repository locale e repository remoto . . . . .	54
4.1	Pre-view della cartella di lavoro selezionata per l'importazione dei dati	56
4.2	Tabella relativa ai dettagli dei decessi in Power Query . . . . .	58
4.3	Tabella relativa alle diverse province della regione Marche . . . . .	59
4.4	Dashboard panoramico sulla situazione COVID-19 . . . . .	61
4.5	Formattazione condizionale del valore del campo "Variazione Malati"	61
4.6	Formula DAX della colonna calcolata relativa alla "Variazione Malati" . . . . .	61
4.7	Il riquadro Filtro con i filtri applicati all'oggetto visivo tabella . . . . .	62
4.8	Il dashboard relativo alle variazioni giornaliere dei casi positivi . . . . .	65
4.9	Esempio dashboard con filtro dei dati in base alla provincia di Ancona	66
4.10	Il dashboard relativo al dato dei tamponi e dei casi positivi . . . . .	66
4.11	Formula DAX per il calcolo del rapporto tra casi positivi e casi diagnosticati . . . . .	67
4.12	Formula DAX per il calcolo della media settimanale . . . . .	67
4.13	Il dashboard relativo ai Malati Attivi . . . . .	68
4.14	Formula DAX per il calcolo della colonna relativa alla variazione dei malati . . . . .	69
4.15	Il dashboard relativo al dato dei ricoveri suddivisi per provincia . . . . .	70
4.16	Formula DAX per il calcolo della colonna calcolata relativa alla variazione dei ricoveri in terapia intensiva . . . . .	71
4.17	Il dashboard relativo al dato della quarantena domiciliare suddiviso per provincia . . . . .	72
4.18	Misura calcolata "ConPatologie" . . . . .	74
4.19	Misura calcolata "SenzaPatologie" . . . . .	74
4.20	Il dashboard relativo al dato dei decessi . . . . .	74
4.21	Il dashboard relativo al dato dei decessi comune per comune . . . . .	75
4.22	Il dashboard relativo all'andamento dei casi positivi suddivisi per ondata di contagi . . . . .	77
4.23	Il dashboard relativo all'andamento dei ricoverati suddivisi per ondata di contagi . . . . .	79
4.24	Il dashboard relativo all'andamento dei decessi suddivisi per ondata di contagio . . . . .	80

---

## Elenco dei listati

3.1	Struttura del file json che raccoglie i tre report giornalieri della Regione Marche .....	44
-----	---	----



---

## Introduzione

In tutto il mondo, ormai, appare evidente come la digitalizzazione dei sistemi informativi stia cambiando il modo in cui guardiamo diversi aspetti della nostra vita, da quello lavorativo a quello culturale. Abitudini e consumi si stanno evolvendo attraverso il nuovo mercato degli acquisti online e altrettanto stanno facendo le aziende, obbligate nel mantenersi costantemente al passo con i tempi.

In questo scenario si richiede, dunque, la capacità di gestire e manipolare enormi quantità di dati, al fine di estrapolare informazioni importanti e trarne conclusioni, a seconda del settore di interesse.

La Data Analytics assume, perciò, un ruolo fondamentale in tutti gli aspetti della vita quotidiana. Grazie alla crescente consapevolezza dell'importanza dei dati, la Data Analytics sta acquisendo sempre più una posizione determinante in ambito finanziario ed economico, ma non solo. Essa, infatti, sfruttando metodologie e algoritmi di analisi sempre più efficienti, consente di procurare, preparare e analizzare enormi moli di dati grezzi e non strutturati per estrarre informazioni significative che possano servire come input per identificare pattern, arricchire i dati aziendali già esistenti ed effettuare delle ricerche su larga scala. In particolare, attraverso le diverse fasi di analisi descrittiva, diagnostica, predittiva e prescrittiva, la Data Science permette di attribuire un vero e proprio significato ai dati, con lo scopo di comprendere e valorizzare la grande ricchezza che deriva da essi. In quest'ottica, il dato viene trasformato in un valore prezioso per l'azienda, supportandola in tutte quelle attività di decision-making, affinché le decisioni possano essere basate sui dati effettivi e non semplicemente sull'esperienza passata o sul solo intuito.

L'avvento dei Big Data, inoltre, ha acquisito importanza via via crescente nell'organizzazione delle attività di produzione e di scambio, al punto che i dati sono considerati una risorsa economica a tutti gli effetti, se non la ricchezza di gran lunga più importante in molti settori. Infatti, grazie agli avanzamenti nell'ambito dell'Information e Communication Technology (ICT), le organizzazioni tendono a raccogliere ed elaborare in tempo reale una vasta quantità di dati di vario tipo, per migliorare i propri processi decisionali e memorizzarli in maniera permanente, al fine di poterli riutilizzare in futuro.

In questo contesto, entra in gioco la cosiddetta Business Intelligence, la quale offre un insieme di strumenti software in grado di trasformare grandi quantità di

dati non correlati tra loro in un insieme di informazioni coerenti, mediante veri e propri modelli di visualizzazione che ne semplificano la lettura e la comprensione.

La versatilità nell'utilizzo della Data Analytics risulta, dunque, fondamentale in molti aspetti del nostro mondo. Questa potenzialità ha permesso, nel nostro specifico caso di studio, di applicare i processi della Data Analytics per analizzare la pandemia da COVID-19.

Negli ultimi anni la gestione dei dati è stato il tema centrale dei dibattiti sanitari, sociali ed economici. L'emergenza COVID-19, in particolare, ha rilevato l'importanza dei dati come risorsa fondamentale, utile alla comprensione dei fenomeni. L'interesse per i dati, inoltre, ha messo in luce il valore delle informazioni in essi contenute, rendendo consapevoli sulle potenzialità della Business Intelligence.

Durante quest'ultimo periodo è emerso il contributo della tecnologia nella gestione della salute pubblica e si è constatato come i modelli di analisi siano uno strumento indispensabile per guidare le decisioni, anche in ambito sanitario. Si è così sperimentato il potere del dato nella sua dimensione più concreta, come elemento centrale di un processo digitale, capace di estrarre informazioni dai dati grezzi e raccogliendo intuizioni dalle analisi effettuate. È proprio l'integrazione tra scienza medica e scienza del dato, infatti, ad aver fatto la differenza nel contrastare il diffondersi della pandemia.

I dati isolati tra loro hanno poco valore, ma lo acquisiscono quando sono organizzati. Per tale ragione, un ruolo centrale nell'intera filiera dei Big Data è rivestito dalla fase di elaborazione, che comporta l'organizzazione dei dati grezzi non strutturati in informazioni adatte ad essere divulgate in un formato compatto e facilmente interpretabile.

Nella gestione dell'emergenza COVID-19, nello specifico, è stato fondamentale riuscire a governare la moltitudine di dati prodotti dalle aziende sanitarie e valorizzarli in informazioni direttamente traducibili in decisioni pratiche. In sintesi, il tracciamento delle persone risultate positive al virus ha permesso di ottenere una prospettiva completa sull'evoluzione della pandemia, di tenere sotto controllo i focolai noti e di rilevarne la presenza di nuovi, grazie a tecniche di gestione del dato basate sulla Business Intelligence e il clustering. Le informazioni che ne derivano risultano decisamente preziose per chi deve stabilire i piani operativi dell'emergenza e tutelare la salute dei cittadini.

La presente tesi si colloca proprio in tale contesto e si ripropone di analizzare i dati relativi alla seconda ondata di contagi da COVID-19 della Regione Marche, al fine di descrivere l'andamento della diffusione della pandemia all'interno del territorio regionale ed ottenere, infine, analogie e differenze rispetto agli sviluppi della prima ondata.

I dati riguardanti l'evolversi della pandemia nella Regione Marche passano per le competenze tecniche del GORES (Gruppo Operativo Regionale Emergenza Sanitaria), un tavolo tecnico che la regione ha istituito da tempo, in modo strutturale, per fornire risposte tempestive alle emergenze sanitarie.

Il GORES, fin da subito, si è impegnato nel fornire con grande chiarezza e trasparenza i dati circa la diffusione del virus all'interno delle province della Regione Marche. I dati divulgati hanno una cadenza giornaliera e riguardano, in particolare, i numeri relativi ai tamponi, ai ricoveri e ai decessi, pubblicati sul sito della regione stessa, mediante tre report (*blu, giallo, arancio*).



Questi dati rappresentano la base del nostro studio, che si prefigge come obiettivo quello di descrivere e analizzare le informazioni relative all'andamento del COVID-19 durante la seconda ondata, realizzando uno schema di propagazione del contagio.

Nella fase preliminare dell'attività di studio sono state acquisite informazioni in merito alla pandemia da COVID-19, con particolare riferimento alle varie tappe del suo sviluppo nel territorio italiano e nelle Marche, dai primi casi segnalati fino alla somministrazione dei vaccini, passando per la seconda ondata epidemica. Sempre in tale fase, è stato introdotto lo strumento software Microsoft Power BI, uno dei leader del mercato dei tool di Data Analytics e Business Intelligence.

Successivamente a questa fase introduttiva, è stata illustrata, in primo luogo, la sorgente di provenienza dei dati ed, in seconda istanza, la struttura dei dati stessi. Dopodichè, sono state delineate le varie operazioni eseguite per la creazione del dataset, dalla realizzazione dei file JSON giornalieri all'aggregazione di questi ultimi utilizzando Script Python.

In seguito alla creazione del dataset, è iniziata la vera e propria attività di Data Analytics. Il processo realizzato durante questa fase è stato quello dell'importazione dei dati di nostro interesse all'interno di Power BI Desktop e la successiva creazione del modello dei dati. Durante quest'ultima attività, si è cercato di ripulire i dati, al fine di renderli facilmente utilizzabili dalle visualizzazioni grafiche disponibili in Power BI.

Una volta ricavato il modello dei dati, è stata svolta un'intensa attività di analisi descrittiva, la quale si occupa, nello specifico, di illustrare ciò che accade e di presentare le informazioni nel modo opportuno, servendosi di sofisticati strumenti di Business Intelligence e di visualizzazione. In particolare, si è deciso di suddividere l'analisi descrittiva per tematiche, in modo da coprire tutti gli aspetti della pandemia: sono stati realizzati dashboard che analizzano i dati riguardanti i casi positivi al virus, i tamponi effettuati, i ricoveri nelle strutture ospedaliere, i malati dimessi, i guariti dal virus e i decessi causati da esso, il tutto con riferimento alla Regione Marche. Infine, sono stati realizzati ulteriori dashboard che mettono a confronto la prima e la seconda ondata di contagi da COVID-19, considerando lo stesso numero di giorni ed esaminando i dati dei parametri che meglio sembrano caratterizzare la pandemia, ossia i casi positivi, gli ospedalizzati e i decessi registrati.

La maggior parte dei dashboard realizzati, inoltre, è stata suddivisa per singola provincia delle Marche ed è stata resa fortemente interattiva; selezionando una provincia all'interno di un oggetto visivo, quest'ultimo filtra i dati mostrandone l'andamento ad essa particolarizzato. La scelta di organizzare i dashboard per singole province ha permesso di effettuare confronti interni alla Regione e seguire l'evolversi della pandemia.

Nell'ultima fase del nostro studio, sono state tratte le conclusioni su tutto il lavoro svolto, evidenziandone i possibili sviluppi futuri come base di partenza per ulteriori studi e analisi nell'ambito delle diffusione di pandemie.

La presente tesi è strutturata come di seguito specificato:

- Nel Capitolo 1 verranno fornite, in maniera sintetica, informazioni sul nuovo coronavirus SARS-COV-2 responsabile della pandemia tutt'ora in corso. In particolare, si mostrerà una panoramica di ciò che è successo in Italia, dai primi casi segnalati fino alla somministrazione di vaccini. Successivamente, si analizzerà l'importanza della Data Science a supporto della gestione della pandemia.

- Nel Capitolo 2 verranno introdotte le funzionalità e le potenzialità del software di analisi dei dati utilizzato nel nostro studio, ovvero Power BI, con particolare riferimento alla sua componente Power BI Desktop.
- Nel Capitolo 3 verrà descritta la sorgente da cui provengono i dati utilizzati e la struttura degli stessi. Successivamente, verranno illustrate le operazioni eseguite per la creazione del dataset in un formato pulito e adatto per i nostri studi.
- Nel Capitolo 4 verrà trattata l'intensa attività di analisi descrittiva. All'interno del capitolo verranno realizzati dashboard interattivi per mostrare una panoramica completa delle principali tematiche dei dati a nostra disposizione per quanto riguarda la seconda ondata di contagi da COVID-19 nelle Marche. Infine, verranno messi a confronto gli andamenti della prima e della seconda ondata, al fine di tracciarne analogie e differenze. A tale scopo si faranno uso delle differenti tipologie di visualizzazione che Power BI Desktop mette a disposizione.
- Nel Capitolo 5 verranno illustrate le lezioni apprese durante lo sviluppo della presente tesi.
- Nel Capitolo 6 verranno tratte delle conclusioni in merito al lavoro svolto e si analizzeranno alcuni possibili sviluppi futuri.

## Pandemia da COVID-19

*Questo capitolo si occupa di fornire in maniera sintetica informazioni sul nuovo coronavirus SARS-COV-2 responsabile dell'attuale epidemia. In secondo luogo si mostrerà una panoramica di ciò che è successo in Italia, dai primi casi segnalati alla seconda ondata epidemica fino alla somministrazione di vaccini.*

### 1.1 Il virus e la pandemia

"Se pensiamo a qualcosa che potrebbe uccidere milioni di persone, una pandemia è la minaccia più grande. In termini di bilancio delle vittime una pandemia farebbe concorrenza persino alle grandi guerre del passato. L'economia si fermerebbe, il costo della vita sarebbe incredibile e nessun Paese sarebbe immune a problemi che un evento simile creerebbe."

Purtroppo tali parole appaiono quanto mai più attuali al giorno d'oggi, in cui la lotta al coronavirus è ancora aperta così come le ferite da esso causate.

Il nuovo coronavirus SARS-COV-2 è soltanto l'ultimo arrivato di una famiglia di ben sette coronavirus in grado di infettare gli esseri umani. La denominazione COVID-19, invece, è il nome della malattia causata, data dall'abbreviazione inglese di COrona VIRus Disease 2019 (anno di scoperta).

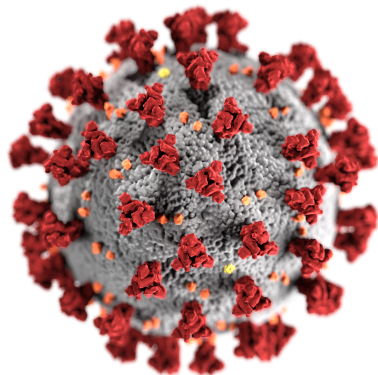
Il virus si diffonde tramite le goccioline di saliva prodotte starnutando, tossendo o parlando, le quali possono penetrare nell'organismo dagli occhi, dal naso o dalla bocca.

Una caratteristica peculiare del coronavirus SARS-COV-2 è la sua capacità di sopravvivenza sulle superfici, anche per diverse ore; di conseguenza l'infezione può avvenire anche semplicemente toccando queste ultime e, successivamente, portandosi le mani al viso, gesto che, in media, viene fatto venti volte in un'ora.

Una volta introdotte nell'organismo, le punte caratteristiche della struttura del coronavirus fungono da chiave incastrandosi tra le proteine che ricoprono molte cellule umane (Figura: 1.1)

Successivamente il virus dà l'istruzione a tali cellule infette di riprodurre nuove copie, invadendone potenzialmente in numero sempre crescente.

I pazienti contagiati dal virus solitamente accusano sintomi come febbre, tosse secca e affaticamento, ma va sottolineato che si può essere infetti o contagiosi anche senza alcun segnale manifesto. È questo uno dei motivi principali che rende il SARS-COV-2 così subdolo, oltre al fatto che la sintomatologia citata può essere facilmente scambiata come segnale di una tradizionale influenza stagionale. Ad ogni modo, nelle forme più gravi, i contagiati possono sviluppare un'infezione polmonare; perciò determinati gruppi della popolazione risultano particolarmente a rischio.



**Figura 1.1.** Modello 3D del Covid-19

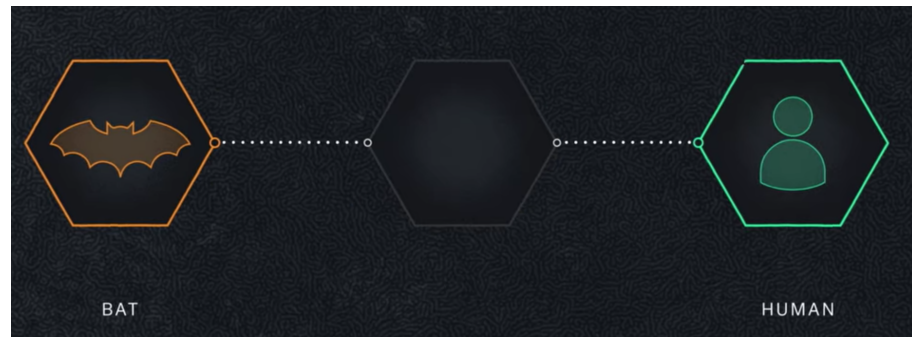
### 1.1.1 Origine del virus

La prima data ufficiale in cui ha avuto inizio la storia del nuovo coronavirus è il 31 Dicembre 2019 quando il governo cinese segnalò all'OMS (Organizzazione Mondiale della Sanità) l'emergenza di diversi casi di una misteriosa polmonite. In pochi giorni si contarono 41 casi, molti dei quali ebbero un legame epidemiologico con il mercato di Huanan Seafood, a Wuhan, nel sud della Cina, il che ha portato ad ipotizzare, fin dal principio, il coinvolgimento di animali vivi nella catena di contagio.

Già in precedenza, da quando fu scoperta la possibilità di trasmissione del virus dal mondo animale al mondo umano, un gruppo di studio chiamato EcoHealth Alliance ha condotto analisi su pipistrelli, considerati storicamente i reservoir naturali di molti virus ed alcuni anni fa scoprì un virus chiamato "virus del pipistrello RaTG13", considerato comunque a basso rischio.

Quando gli scienziati hanno sequenziato il genoma del virus responsabile del COVID-19 hanno scoperto che era per il 95% simile al virus del pipistrello; di conseguenza una teoria proposta ritiene che il virus del pipistrello sia stato mutato, direttamente o passando attraverso un'altra specie, nel nuovo coronavirus, in grado di infettare anche gli esseri umani (Figura: 1.2).

Tuttavia, ad oggi, l'esatta origine resta ancora un mistero e nessuna ipotesi è stata scartata a priori.



**Figura 1.2.** Mutazione del coronavirus SARS-COV-2

### 1.1.2 La diffusione della pandemia

A distanza di pochi giorni dall’annuncio del cluster di casi a Wuhan i ricercatori cinesi hanno condiviso la ”carta d’identità” del virus, ovvero la sequenza genomica. La conoscenza della sequenza è stato il primo passo nella lotta contro il coronavirus, permettendo la realizzazione di un test diagnostico che potesse essere messo a disposizione di molti altri paesi.

Dai primi casi segnalati in Cina, infatti, il virus si diffuse rapidamente in un territorio sempre più vasto raggiungendo, di volta in volta, quasi tutti i continenti. L’11 Marzo 2020, a fronte dell’aumento esponenziale del numero di contagi a livello globale, l’OMS ha definito la diffusione del COVID-19 non più un’epidemia confinata ad alcune zone geografiche, bensì una pandemia, estesa a tutto il pianeta.

A seguito della dichiarazione dell’OMS, ogni Stato ha messo in atto misure di contenimento del contagio per tenere sotto controllo l’espansione e i numeri del virus.

Nel giro di poco tempo, dunque, il virus è diventato parte della vita quotidiana cui doversi adattare; ad un tratto ci si è ritrovati a vivere una realtà nuova, a volte spaesati e impreparati ad affrontare mesi segnati dal susseguirsi di misure restrittive e nuove abitudini.

## 1.2 La seconda ondata di contagi da COVID-19

Dopo un’estate in cui si è registrato un rallentamento dell’epidemia, permettendo, quindi, una parziale riapertura, molti paesi europei hanno assistito ad una brusca risalita nel numero dei contagi, il che ha lasciato presagire la necessità di nuove misure di sicurezza.

In realtà, il nuovo incremento osservato era prevedibile, come preannunciato da molti esperti già nel corso della stagione estiva. Da sempre, infatti, esistono leggi biologiche che governano i rapporti tra organismi (uomini) e microrganismi (virus e batteri), le quali si adeguano ai ritmi delle stagioni e all’influenza dei diversi ambienti.

Nello specifico, è noto che le patologie respiratorie, come nel caso del COVID-19, crescono senza sosta fino alla primavera, sembrano sparire nel periodo estivo

per poi manifestarsi nuovamente con l'arrivo della stagione autunnale. A partire da Settembre 2020, periodo di inizio della seconda ondata, ci si è ritrovati, inoltre, in uno scenario nel quale l'età media del contagio si è abbassata, tant'è che si sono moltiplicati i casi anche tra gli individui al di sotto dei 30 anni, inizialmente ritenuti immuni o quasi.

Tali premesse hanno portato ben presto ad un nuovo collasso delle strutture sanitarie perciò molti paesi europei, come contromossa, hanno dichiarato mini lockdown nazionali con blocchi di aree territoriali e chiusure mirate o generalizzate di negozi e attività.

Novembre è stato il mese della speranza. Sono arrivati, a distanza di pochi giorni l'uno dall'altro, due annunci molto importanti sul fronte immunizzazione: il primo riguardante il vaccino anti-COVID messo a punto dalla azienda farmaceutica americana Pfizer in collaborazione con la tedesca BioNTech il quale è stato dimostrato essere efficace al 90%, il secondo relativo al vaccino sviluppato da Moderna il quale ha riportato un'efficacia che sfiora il 95% (Figura 1.3).



**Figura 1.3.** Vaccini Pfizer-BioNTech e Moderna

L'8 Dicembre 2020 è una data destinata a rimanere nella storia in quanto è stato il giorno della prima somministrazione del vaccino contro il COVID-19. La prima dose è stata iniettata a una cittadina 90enne dell'Irlanda del Nord, ospite di una casa di riposo. Ed è stata proprio l'Inghilterra, difatti, il primo paese al mondo a dare il via ad una campagna di vaccinazione adottando il vaccino prodotto da Pfizer-BioNTech.

Nel frattempo l'EMA (Agenzia Europea per i Medicinali) ha approvato la somministrazione del vaccino Pfizer-BioNTech, anche all'interno dei Paesi dell'Unione Europea, fissando il cosiddetto Vaccine-Day, simbolo della nuova fase di immunizzazione, al 27 Dicembre.

Se, da una parte, l'arrivo del 2021 ha riposto grande attesa e speranza nei vaccini, dall'altra l'esplosione di nuove varianti del virus ha iniziato a spaventare la comunità

scientifico. Degna di nota è la variante B117, denominata variante inglese, che ha messo in ginocchio in primo luogo la Gran Bretagna per poi trovare diffusione anche all'estero.

Nonostante l'allerta per la variante inglese, però, sono stati registrati anche i primi dati confortanti sul processo di immunizzazione in atto che ha visto un lodevole protagonista in Israele, Paese che, in poco meno di un mese, ha somministrato la prima dose del vaccino Pfizer a quasi un quarto della sua popolazione, provocando un drastico calo dei contagi già dal 14esimo giorno.

### 1.2.1 La situazione in Italia

La seconda ondata di contagi da COVID-19 in Italia è esplosa a fine ottobre 2020, lasciando, dunque, predire l'esigenza di nuove misure di contenimento.

Il 3 Novembre il Presidente del Consiglio, Giuseppe Conte, a seguito del rapporto del Comitato Tecnico-Scientifico, ha firmato un nuovo DPCM tramite il quale sono state previste nuove misure restrittive per bar e ristoranti, sport di contatto, nonché la didattica a distanza e lo smart working.

Successivamente, con il DPCM del 4 Novembre 2020, il premier Conte ha suddiviso l'Italia in tre zone a differente livello di allerta. Le disposizioni di tale DPCM sono entrate in vigore il 6 novembre.

Tale ripartizione è costituita da una fascia gialla, una arancione e una rossa, in riferimento alla gravità della situazione epidemiologica (Figura 1.4). Nella zona rossa, quella a più alto rischio, è previsto, di fatto, un lockdown, anche se con provvedimenti meno stringenti rispetto a quelli della trascorsa primavera.

AREA GIALLA	AREA ARANCIONE	AREA ROSSA
<p><b>Vieta circolare dalle ore 22 alle ore 5 del mattino, salvo comprovati motivi di lavoro, necessità e salute. Raccomandazione di non spostarsi se non per motivi di salute, lavoro, studio, situazioni di necessità.</b></p> <p><b>Chiusura dei centri commerciali nei giorni festivi e prefestivi ad eccezione delle farmacie, parafarmacie, punti vendita di generi alimentari, tabaccherie ed edicole al loro interno.</b></p> <p><b>Chiusura di musei e mostre.</b></p> <p><b>Didattica a distanza per le scuole superiori, fatta eccezione per gli studenti con disabilità e in caso di uso di laboratori; didattica in presenza per scuole dell'infanzia, scuole elementari e scuole medie. Chiuse le università, salvo alcune attività per le matricole e per i laboratori.</b></p> <p><b>Riduzione fino al 50% per il trasporto pubblico, ad eccezione dei mezzi di trasporto scolastico.</b></p> <p><b>Sospensione di attività di sale giochi, sale scommesse, bingo e slot machine anche nei bar e tabaccherie.</b></p> <p><b>Chiusura di bar e ristoranti alle ore 18. L'asporto è consentito fino alle ore 22. Per la consegna a domicilio non ci sono restrizioni.</b></p> <p><b>Restano chiuse piscine, palestra, teatri, cinema. Restano aperti i centri sportivi.</b></p>	<p><b>Vieta circolare dalle ore 22 alle ore 5 del mattino, salvo comprovati motivi di lavoro, necessità e salute.</b></p> <p><b>Vieta gli spostamenti in entrata e in uscita da una Regione all'altra e da un Comune all'altro, salvo comprovati motivi di lavoro, studio, salute, necessità. Raccomandazione di evitare spostamenti non necessari nel corso della giornata all'interno del proprio Comune.</b></p> <p><b>Chiusura di bar e ristoranti, 7 giorni su 7. L'asporto è consentito fino alle ore 22. Per la consegna a domicilio non ci sono restrizioni.</b></p> <p><b>Chiusura dei centri commerciali nei giorni festivi e prefestivi ad eccezione delle farmacie, parafarmacie, punti vendita di generi alimentari, tabaccherie ed edicole al loro interno.</b></p> <p><b>Chiusura di musei e mostre.</b></p> <p><b>Didattica a distanza per le scuole superiori, fatta eccezione per gli studenti con disabilità e in caso di uso di laboratori; didattica in presenza per scuole dell'infanzia, scuole elementari e scuole medie. Chiuse le università, salvo alcune attività per le matricole e per i laboratori.</b></p> <p><b>Riduzione fino al 50% per il trasporto pubblico, ad eccezione dei mezzi di trasporto scolastico.</b></p> <p><b>Sospensione di attività di sale giochi, sale scommesse, bingo e slot machine anche nei bar e tabaccherie.</b></p> <p><b>Restano chiuse piscine, palestra, teatri, cinema. Restano aperti i centri sportivi.</b></p>	<p><b>È vietato ogni spostamento, anche all'interno del proprio Comune, in qualsiasi orario, salvo che per motivi di lavoro, necessità e salute: vietati gli spostamenti da una Regione all'altra e da un Comune all'altro.</b></p> <p><b>Chiusura di bar e ristoranti, 7 giorni su 7. L'asporto è consentito fino alle ore 22. Per la consegna a domicilio non ci sono restrizioni.</b></p> <p><b>Chiusura dei negozi, fatta eccezione per supermercati, beni alimentari e di necessità.</b></p> <p><b>Restano aperte edicole, tabaccherie, farmacie e parafarmacie, lavanderia, parrucchieri e barbieri. Chiusi i centri estetici.</b></p> <p><b>Didattica a distanza per la scuola secondaria di secondo grado, per le classi di seconda e terza media. Restano aperte, quindi, solo le scuole dell'infanzia, le scuole elementari e la prima media. Chiuse le università, salvo specifiche eccezioni.</b></p> <p><b>Sono sospese tutte le competizioni sportive salvo quelle riconosciute di interesse nazionale dal CONI e CIP. Sospese le attività nei centri sportivi. Rimane consentito svolgere attività motoria nei pressi della propria abitazione e attività sportiva solo all'aperto in forma individuale.</b></p> <p><b>Sono chiusi musei e mostre; chiusi anche teatri, cinema, palestre, attività di sale giochi, sale scommesse, bingo, anche nei bar e nelle tabaccherie. Per i mezzi di trasporto pubblico è consentito il riempimento solo fino al 50%, fatta eccezione per i mezzi di trasporto scolastico.</b></p>

Figura 1.4. Disposizioni per differente livello di allerta

Al di là del numero totale di casi, l'Italia ha assistito ad un contagio diffuso su tutto il territorio nazionale, a differenza del quadro di Marzo, in cui la maggior parte dei casi e dei decessi si era registrata al Nord.



A Dicembre 2020, in seguito all'approvazione dei vaccini anti-COVID Pfizer e Moderna, anche l'Italia ha predisposto un opportuno piano di immunizzazione in concomitanza delle prime somministrazioni avvenute in molti Paesi europei.

La mattina del 27 dicembre sono state inoculate le prime dosi del vaccino Pfizer presso l'ospedale Spallanzani di Roma dove la prima vaccinata è stata l'infermiera Claudia Alivernini di 29 anni, inaugurando, di fatto, la campagna vaccinale nazionale (Figura 1.5).



**Figura 1.5.** Infermiera Claudia Alivernini, prima vaccinata in Italia

A Gennaio 2021, dopo la notizia della diffusione di una nuova e preoccupante variante del virus sviluppatasi in Gran Bretagna, l'Italia ha deciso di chiudere i voli da e verso il Regno Unito.

### 1.2.2 Prima e seconda ondata a confronto

La prima ondata dei contagi da COVID-19, come facilmente comprensibile, ha colto tutti di sorpresa ma, grazie allo sforzo eroico di medici ed infermieri, e i necessari sacrifici dei cittadini, è stato possibile uscire da un simile momento terribile.

Ciò che è accaduto nella seconda ondata, però, è stato, sotto certi aspetti, diverso in quanto i rischi erano già noti e nel corso dell'estate è stato commesso l'errore di sottovalutare l'incombente minaccia del riacutizzarsi dei picchi di contagi con il cambio di stagione.

La prima fase dell'epidemia è stata caratterizzata da un lungo periodo di lockdown, disposto in seguito ad una rapidissima diffusione del virus a cavallo tra Febbraio e Marzo 2020 e proseguita per circa quattro settimane.

Nel corso di questo intervallo temporale, l'andamento dell'epidemia ha mostrato un numero dei contagi che ha raggiunto un massimo a fine Marzo per poi iniziare una graduale discesa dovuta agli effetti del lockdown.



La riapertura di Giugno 2020 è avvenuta in un momento in cui in alcune regioni l'epidemia era di fatto scomparsa, mentre in altre era ancora in fase discendente, seppur non ancora del tutto debellata.

Questa situazione variegata a livello regionale, insieme alla ripresa di spostamenti interni e dei flussi di turismo in entrata ed uscita, ha fatto sì che, durante l'estate, il numero totale di infetti dapprima interrompesse la propria discesa, tra Luglio e Agosto, e poi addirittura riprendesse a salire, tra Agosto e Settembre.

Per quel che riguarda i dati, la seconda ondata di contagi da COVID-19 ha interessato un numero di italiani 8 volte superiore rispetto alla prima. Infatti, mentre nella prima ondata sono state infettate 236.134 persone, nella seconda fase il numero di contagiati è risultato pari a 1.822.841. Come emerso dal Focus dell'Instant report COVID-19, "La pandemia è una ma sembrano quasi due eventi diversi".

Un'ulteriore differenza riscontrata è che, da una parte, la prima ondata ha visto raggiungere il suo picco in poche settimane, dall'altra la seconda fase è stata caratterizzata da un'onda lunga che ha raggiunto più lentamente il plateau. Nello specifico, il picco giornaliero di contagi relativo alla prima ondata è stato di 6.557 persone e si è raggiunto il 21 marzo, a 27 giorni dall'inizio della crisi sanitaria; nella seconda ondata il picco giornaliero è stato di 40.906 contagi e si è raggiunto il 13 novembre, al 61esimo giorno.

Un'analogia riscontrata dai dati, invece, è una percentuale molto simile riferita ai contagiati che hanno dovuto ricorrere ad una terapia intensiva nella prima e seconda ondata, rispettivamente il 10,6% e il 9,3%.

Per quanto riguarda il tasso di positività adottato dal Ministero della Salute, calcolato come il rapporto tra nuovi positivi e tamponi effettuati, nella prima ondata si è toccato il massimo il 15esimo giorno, con un tasso pari al 46%; nella seconda ondata il massimo valore è stato raggiunto dopo 61 giorni ed il tasso è stato pari al 17,9%.

Durante la prima ondata sono decedute complessivamente 34.167 persone, mentre, nella seconda ondata, i decessi sono stati 38.549; nonostante ciò la prima ondata è stata più letale della seconda. Quest'ultimo riscontro si evince dalla letalità media apparente (settimanale), che, nella prima ondata, è stata del 14,9%, a differenza di quella della seconda ondata che è stata dell'1,9%, 7 volte inferiore.

Al di là dei numeri presentati in questo paragrafo, tale contesto verrà approfondito in un capitolo apposito nel quale saranno messe a confronto le analisi dei dati riguardanti le due ondate di contagi da COVID-19.

### 1.3 I vaccini anti-COVID-19

Da quando, l'11 Gennaio 2020, è stata pubblicata la sequenza genetica del virus SARS-CoV-2, scienziati e industrie di tutto il mondo hanno collaborato per sviluppare il prima possibile vaccini sicuri ed efficaci contro il COVID-19.

L'urgenza di creare un vaccino per il COVID-19 ha portato a programmi compressi, che hanno abbreviato la tempistica standard di sviluppo del vaccino, in alcuni casi combinando le fasi della sperimentazione clinica nell'arco di mesi, processo che tipicamente è sempre stato condotto sequenzialmente nel corso degli anni.

A luglio 2021, i vaccini anti-COVID-19 approvati da FDA (Food and Drug Administration) ed EMA (Agenzia Europea per i Medicinali) e, conseguentemente, messi in distribuzione, sono quattro, mentre sono diversi quelli ancora al vaglio della sperimentazione clinica o in attesa di approvazione.

Per quanto riguarda la struttura, alcuni vaccini sono stati realizzati per mezzo di tecnologie più tradizionali, secondo gli approcci utilizzati già nello sviluppo di vaccini contro SARS e Ebola; altri sono stati concepiti attraverso nuovi approcci, come nel caso del vaccino ad mRNA.

L'obiettivo, ad ogni modo, è stato quello di produrre una risposta immunitaria al fine di neutralizzare il virus e impedire l'infezione delle cellule.

Dei quattro vaccini anti-COVID-19 al momento somministrati, due sono a mRNA (Pfizer-BioNTech e Moderna) e due a vettore virale (AstraZeneca e Janssen di Johnson & Johnson).

I primi sono basati su RNA messaggero (mRNA) di SARS-CoV-2 in cui sono conservate le istruzioni per la codifica di una specifica proteina del virus, nota come spike o proteina S. I vaccini a vettore virale, invece, sono tipicamente basati su un virus esistente, generalmente un adenovirus incompetente per la replicazione, il quale trasporta la sequenza del codice genetico che codifica la proteina spike.

Per quanto concerne i vaccini anti-COVID-19 ancora in fase di sperimentazione o in attesa di approvazione, la lista include, oltre ad ulteriori preparati a mRNA e a vettore virale, vaccini inattivati, vaccini attenuati e vaccini a subunità.

### 1.3.1 Sperimentazione, valutazione, approvazione

Generalmente lo sviluppo di un vaccino è un processo lungo che necessita dai sette ai dieci anni, durante i quali le ricerche vengono condotte a tappe successive che includono i test di qualità, la sperimentazione preclinica e le fasi della sperimentazione clinica nell'uomo.

Il processo di produzione consiste in tre fasi: studi di fase 1, condotti su volontari sani per l'identificazione della dose ottimale e la valutazione della sicurezza nell'uomo; studi di fase 2 a carattere esplorativo e condotti su piccoli gruppi di persone, in genere meno di 100; studi di fase 3, disegnati allo scopo confermativo e condotti su migliaia o decine di migliaia di persone.

In Europa, in particolare, quando un'azienda farmaceutica ritiene di essere in grado di dimostrare qualità, sicurezza ed efficacia del prodotto, sottopone all'EMA una richiesta di autorizzazione per l'immissione sul mercato. Solo dopo l'approvazione dell'EMA e l'autorizzazione da parte della Commissione Europea, l'azienda inizia il processo di produzione su ampia scala.

### 1.3.2 Quadro della campagna vaccinale in Italia

Il 27 Dicembre 2020, il cosiddetto "Vaccine-Day", è stata la data che ha sancito il via ufficiale alla campagna di vaccinazione contro il COVID-19 in tutta Europa.

In Italia la distribuzione vera e propria del vaccino è iniziata il 31 Dicembre; i primi a riceverlo sono stati gli individui riconosciuti come appartenenti alle categorie da vaccinare prima di tutte le altre, ovvero gli operatori sanitari e sociosanitari, gli ospiti e i dipendenti delle Residenze Sanitarie Assistenziali (RSA).

Vaccinare subito medici e infermieri ha avuto un duplice valore; da un lato ha permesso di proteggere gli operatori maggiormente esposti a rischio a causa della natura del loro lavoro, dall'altro ha consentito di garantire l'assistenza sanitaria. Nel corso della prima ondata, infatti, sono state pesanti le conseguenze dovute a reparti ospedalieri decimati dalle assenze del personale sanitario infetto o costretto alla quarantena.

Con il decreto del 12 Marzo 2021, l'Italia ha adottato il nuovo piano strategico per l'esecuzione della campagna di vaccinazione nazionale, elaborato dal Ministero della Salute in collaborazione con i principali enti del settore tecnico-scientifico; attraverso il decreto in questione sono state definite le categorie di popolazione da vaccinare e le rispettive priorità da applicare su tutto il territorio nazionale. A tale proposito, dagli studi scientifici effettuati, è emerso che l'età e la presenza di patologie pregresse rappresentano le variabili principali di correlazione con la mortalità da COVID-19. Alla luce di ciò, sono state identificate cinque categorie entro cui suddividere la popolazione, ovvero

- *Categoria 1*: persone a elevata fragilità, ovvero estremamente vulnerabili in quanto affette da patologie oppure portatori di disabilità gravi.
- *Categoria 2*: persone di età compresa tra 70 e 79 anni.
- *Categoria 3*: persone di età compresa tra i 60 e 69 anni.
- *Categoria 4*: persone di età inferiore ai 60 anni con patologie e situazioni di compromissione immunologica che possono aumentare il rischio di ammalarsi e che non siano nella condizione di gravità delle persone estremamente vulnerabili.
- *Categoria 5*: resto della popolazione di età inferiore ai 60 anni.

Inoltre, a prescindere dall'età, sono stati considerati prioritari nella somministrazione del vaccino anche il personale docente e non docente, scolastico e universitario, le forze armate, di polizia e del soccorso pubblico.

L'obiettivo prefissato con il decreto 12 Marzo è stato quello di immunizzare, entro Ottobre, il 70% della popolazione, soglia minima di copertura affinché il virus circoli meno al fine di ottenere la cosiddetta immunità di gregge.

Al giorno 5 Luglio 2021, la popolazione italiana ad aver completato il ciclo vaccinale è poco più di 20 milioni mentre le dosi somministrate ogni giorno, in media, sono 544.207.

Procedendo con tale ritmo sono stati stimati ulteriori 2 mesi e 19 giorni per coprire fino all'80% della popolazione vaccinabile dunque il traguardo è stato posto al 16 Settembre 2021, in linea con le previsioni del governo per fine Settembre.

## 1.4 La Data Science a supporto della gestione della pandemia

Fin dal principio dell'epidemia, le varie nazioni e organizzazioni mondiali si sono impegnate nel divulgare i dati riguardanti la diffusione e la situazione del virus nei vari territori, attività di cui va riconosciuto sicuramente il merito.

La pandemia da COVID-19, ad ogni modo, ha dimostrato la necessità di un sistema informativo all'avanguardia per la gestione della diffusione delle malattie infettive che possa contare sulla disponibilità dei dati in tempo reale e sulla possibilità di incrociare fonti informative diverse.

Per combattere una pandemia, infatti, è di fondamentale importanza conoscere le dimensioni del fenomeno attraverso l'analisi dei dati allo scopo di avere a disposizione elementi oggettivi per prendere decisioni tempestive in ambito politico, economico, ma soprattutto, sanitario.

In Italia, con l'arrivo dei primi casi di contagi nel lodigiano e nel padovano, la Protezione Civile si è subito occupata di rilasciare dei riepiloghi dei dati giornalieri sull'evoluzione della pandemia ma anche di rendere pubblici i repository di tali dati in modo tale da consentirne il riutilizzo per monitorare l'emergenza o, semplicemente, per corretta informazione.

D'altra parte, però, a distanza di mesi, ci si è resi conto che i dati messi a disposizione dalla Protezione Civile erano sottostimati di almeno dieci volte rispetto alla situazione reale, in particolar modo nelle fasi iniziali dell'emergenza. Tale divergenza è imputabile non solo all'aspetto gestionale operativo, ma anche alla capacità di raccolta delle informazioni.

Gli esperti di Data Science e Statistica hanno chiesto fin da subito alle istituzioni di fornire i dati in maniera completa e trasparente per poter formulare previsioni affidabili utili ai primari degli ospedali ed, in particolare, ai reparti di terapia intensiva affinché questi potessero organizzarsi tempestivamente. Nonostante l'impegno da ambo le parti, però, spesso i dati non sono risultati del tutto accessibili in maniera aperta con la conseguenza che la politica, per quanto riguarda la tempestività delle decisioni, ha "tergiversato", in particolar modo durante la seconda ondata.

Già agli inizi di Ottobre 2020, infatti, le autorità politiche avrebbero potuto prevedere ed anticipare l'imminente ritorno, poi effettivamente avvenuto, di una crescita esponenziale dei contagi, come nella prima ondata, tuttavia ciò non è stato possibile, probabilmente perché non è stato fatto sufficiente affidamento su persone esperte in merito alla dinamica evolutiva del fenomeno.

In conclusione, avere a disposizione i dati serve non solo a capire cosa è accaduto, ma è fondamentale soprattutto per disegnare i possibili sviluppi futuri e, conseguentemente, stabilire quali misure intraprendere nella maniera più tempestiva possibile perché la rapidità di decisione dei provvedimenti è cruciale per limitare la diffusione di un virus che si propaga con velocità di diffusione potenzialmente esponenziale.

Il dato, soprattutto in ambito sanitario, è un segmento determinante e, al tempo stesso, delicato che, negli anni, non ha ricevuto un adeguato investimento né quell'attenzione particolare riguardo la sua prevenzione e l'utilizzo nei dataset che avrebbe meritato.

Le pandemie sono, sicuramente, una diversa dall'altra; perciò uno sviluppo eterogeneo in materia di dati può fornire elementi importanti al fine di creare un Data Warehouse "globale", che possa servire per l'individuazione precoce di informazioni, oggi per lo più raccolte "a canne d'organo".

L'ideale, dunque, sarebbe quello di un'integrazione con nuovi sistemi di gestione dei dati, che possano ricavare delle informazioni di trend e di andamento a partire da un caos meno visibile, ma non per questo inesistente.

---

## Strumenti software utilizzati

*In questo capitolo verrà introdotto il tool Power BI, una piattaforma di analisi business prodotta da Microsoft. Si tratta di una raccolta di strumenti software che estrapolano i dati sotto forma di parole e numeri per restituire grafici chiari e accattivanti sullo stato del business in questione. Il suo obiettivo è, infatti, la creazione di report e dashboard che possano essere condivisi e consultati da qualunque dispositivo e in qualsiasi momento.*

### 2.1 Power BI

Power BI (Figura 2.1) è un insieme di servizi software, app e connettori che interagiscono per trasformare le sorgenti di dati non correlate in un insieme di informazioni coerenti, visivamente accattivanti e interattive. I dati potrebbero essere un foglio di calcolo Excel o una raccolta di data warehouse basati su cloud o ibridi locali.

Power BI, in pratica, consente di connettersi facilmente alle sorgenti dati, visualizzare e scoprire le informazioni importanti ed, infine, condividerle con gli altri utenti.



**Figura 2.1.** Logo di Power BI

La soluzione adottata da Microsoft è considerata la leader sul mercato dei tool di Business Intelligence (BI) e Data Analytics, e ciò è confermato dal Magic Quadrant di Gartner, lo strumento per eccellenza per comparare i sistemi di BI attualmente

in commercio. Grazie ai grossi investimenti di Microsoft Power BI ha, in breve tempo, superato i già presenti leader del settore, quali QlikSense e Tableau (Figura 2.2).



**Figura 2.2.** Il Magic Quadrant di Gartner relativo ai tool di analisi dei dati

Tra le caratteristiche offerte dall'applicazione risaltano l'intuitività e la flessibilità con cui l'editor dei dati consente di elaborare e modellare le informazioni in input, in modo da restituire grafici, statistiche e tutta una serie di analisi dipendenti dalle richieste dell'utente. Infatti, Power BI permette di unificare dati provenienti dalle fonti più disparate; tuttavia, per poter operare su una mole di dati, tenendo anche conto del fatto che questi potrebbero essere in formati differenti, è necessario che le librerie di un sistema di Business Intelligence possiedano tutte le funzioni necessarie per effettuare opportune operazioni di Extraction, Transformation and Loading (ETL) sui dati.

### 2.1.1 Architettura del software

Power BI è considerato un vero e proprio ecosistema, in quanto l'applicazione è disponibile in svariate versioni (Figura 2.3).

Le componenti principali di tale ecosistema sono:

- *Power BI Desktop*: l'applicazione desktop che permette la modellazione e l'analisi dei dati e successivamente, la creazione e la pubblicazione di report e dashboard.

- *Power BI Service*: il SaaS basato sul servizio online che consente la condivisione e la visualizzazione dei dashboard in tempo reale.
- *Power BI Mobile*: le applicazioni per dispositivi Android e iOS che consentono la possibilità di portare dashboard creati dalla componente Desktop su qualsiasi dispositivo mobile affinché si possa controllare l'andamento dell'analisi dei dati in qualsiasi luogo.
- *Power BI Gateway*: componente utilizzata per sincronizzare i dati in entrata e in uscita affinché possa essere svolta l'attività di Business Intelligence.
- *Power BI Visuals Marketplace*: un punto dove la community può scaricare gratuitamente grafici, ma anche algoritmi di Intelligenza Artificiale e di predizione non presenti tra i tool predefiniti.

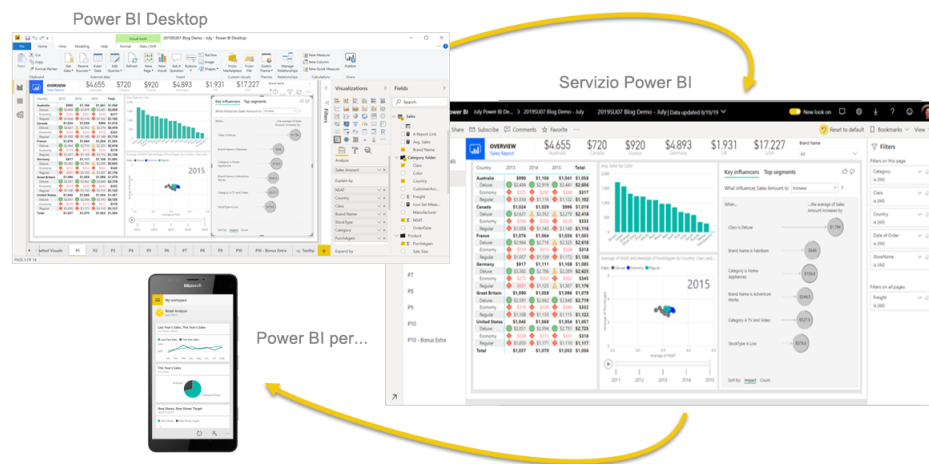
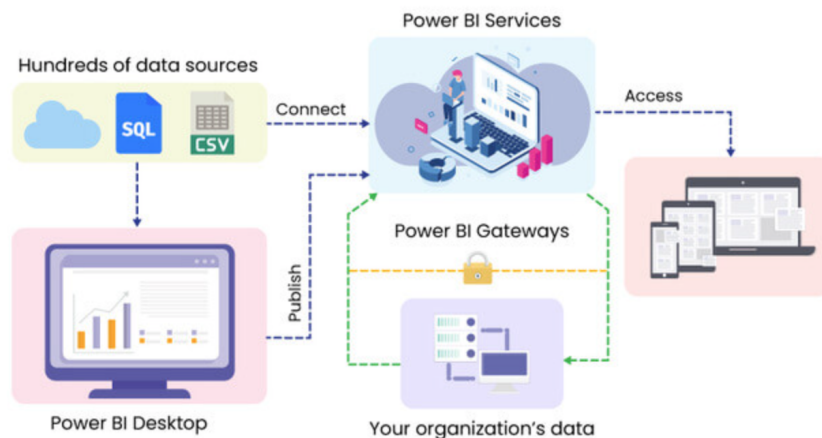


Figura 2.3. L'ecosistema di Power BI

Riassumendo, l'architettura del tool Power BI può essere schematizzata nel seguente modo (Figura 2.4): all'interno dell'applicazione Power BI Desktop vengono caricate tutte le sorgenti dati supportate in diversi formati, come cloud, SQL e CSV. Tali set di dati, però, possono essere pubblicati anche su Power BI Services, componente utilizzata principalmente per la condivisione di report e dashboard. Ad ogni modo, attraverso i gateway, vengono recuperati i dati che, magari, sono posizionati su un server locale dell'azienda. Infine, viene data la possibilità di visualizzare le dashboard da qualsiasi dispositivo mobile per un monitoraggio dell'andamento in tempo reale.

L'utilizzo di Power BI può essere basato sulle funzionalità o sul servizio e rappresenta lo strumento ottimale per ogni situazione. In particolare, l'utilizzo delle varie componenti di cui è costituito Power BI può variare in base al ruolo ricoperto all'interno di un team; ad esempio, può capitare che un utente usi principalmente Power BI Service, per visualizzare report e dashboard, e che i suoi colleghi addetti a calcoli e report aziendali usino, invece, Power BI Desktop, per creare i report e, successivamente, pubblicarli nel servizio Power BI, dove l'utente iniziale può vi-



**Figura 2.4.** L'architettura di Power BI

sualizzarli. Un altro collega che si occupa di vendite potrebbe sfruttare in misura maggiore l'app per smartphone di Power BI, per monitorare l'avanzamento delle quote di vendita e accedere ai dettagli sui nuovi clienti potenziali.

In sostanza, tutte le parti di Power BI vengono messe a disposizione degli utenti ed è questo il motivo per cui tale strumento è così flessibile ed efficace.

In questa tesi verrà utilizzato principalmente Power BI Desktop (Figura 2.5), grazie al quale è possibile connettersi a sorgenti dati diverse e combinarle in un modello di dati che consente di creare raccolte di oggetti visivi condivisibili come report con gli altri utenti all'interno dell'organizzazione.

### Power BI Desktop e Power BI Service a confronto

Le funzionalità di Power BI Desktop e Power BI Service sono piuttosto simili e potrebbero essere facilmente confuse. Per questo motivo, nel seguito verranno descritte singolarmente al fine di mettere in risalto differenze e analogie specifiche dell'applicazione e del servizio. Più specificatamente:

- Power BI Desktop è uno strumento completo per l'analisi dei dati e la creazione di report, installato come applicazione gratuita nel computer locale. Include l'Editor di query, tramite cui è possibile connettersi a molte sorgenti di dati diverse e combinare i dati (attività spesso denominata modellazione) in un modello opportuno. È quindi, possibile progettare un report basato sul modello di dati.
- Power BI Service è un servizio basato sul cloud. Esso consente di apportare piccole modifiche ai report e favorisce la collaborazione di team e aziende. Anche Power BI Service consente di connettersi alle sorgenti di dati, ma la modellazione è limitata.



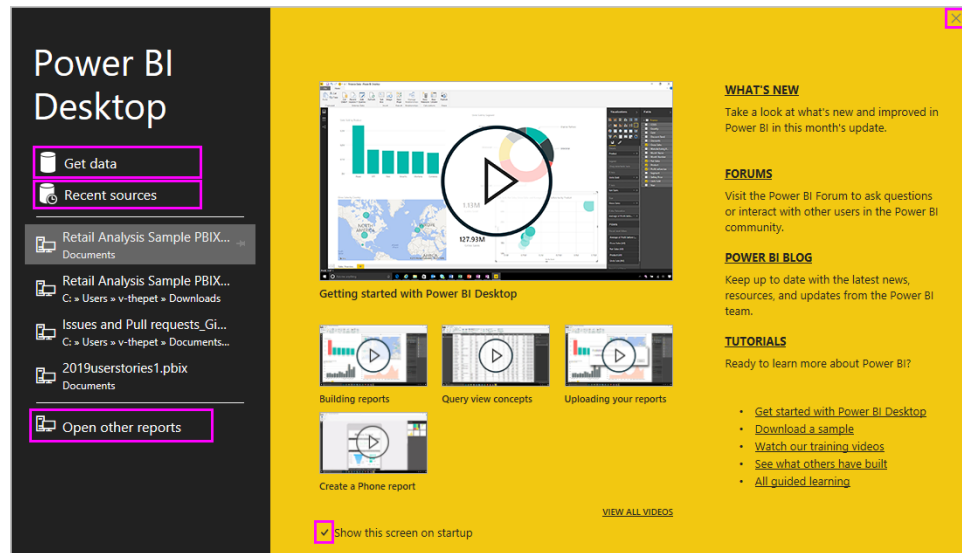


Figura 2.5. Schermata di avvio di Power BI Desktop

La maggior parte dei responsabili della progettazione di report usa Power BI Desktop per la realizzazione di report, e Power BI Service per la distribuzione e la collaborazione con gli altri utenti all'interno dell'organizzazione.

Le differenze e analogie delle due componenti, infine, vengono descritte perfettamente in un diagramma di Venn (Figura 2.6) in cui, in particolare, l'area al centro indica le funzionalità disponibili in entrambe le soluzioni e i due lati, invece, illustrano le funzionalità specifiche dell'applicazione e del servizio.

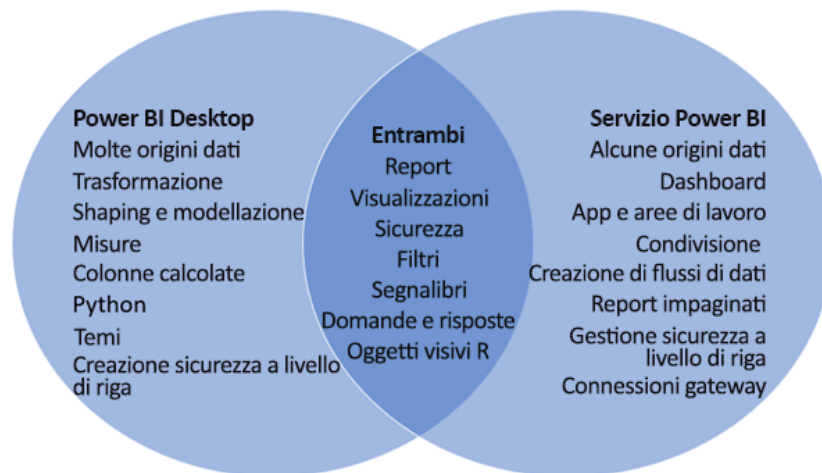


Figura 2.6. Funzionalità fornite da Power BI Desktop e Power BI Service

### 2.1.2 Il flusso di azioni

L'obiettivo comune in Power BI è quello di connettersi alle sorgenti dati, costruire un modello dei dati che sia adeguato per svolgere delle analisi e ricavare da essi informazioni importanti. Una volta terminata la fase di estrapolazione della conoscenza, il lavoro svolto viene organizzato in report e dashboard costituiti da grafici dinamici e accattivanti in Power BI Desktop allo scopo di essere pubblicati e condivisi nel servizio Power BI Service con gli utenti finali.

Di seguito vengono riportate in dettaglio le diverse fasi del flusso di azioni:

- connettersi ai dati;
- trasformare e pulire i dati per creare un modello degli stessi;
- creare oggetti visivi, come grafici o grafi, che forniscono rappresentazioni visive dei dati;
- creare report che raccolgono gli oggetti visivi in una o più pagine di report;
- condividere i report con altri utenti tramite il servizio Power BI.

Con Power BI è possibile creare report complessi e visivamente dettagliati usando dati provenienti da più sorgenti, in un unico report che è possibile condividere con altri utenti nell'organizzazione.

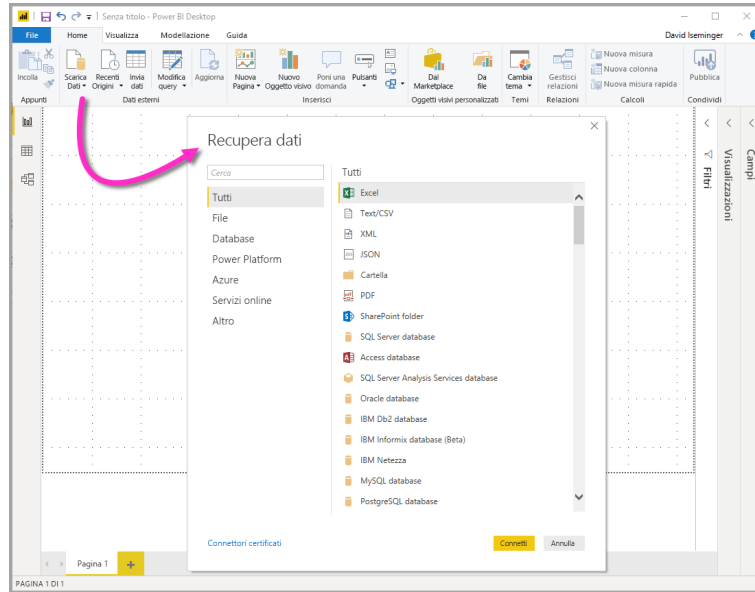
Nelle prossime sezioni verranno descritte nel dettaglio le fasi del flusso di azioni mostrate nell'elenco.

## 2.2 Connessione al set di dati

Come illustrato nella sezione precedente, il primo passaggio per poter lavorare con Power BI consiste nel connettersi ai dati. Per fare ciò, è necessario selezionare semplicemente la barra multifunzione “Home” e, dunque la voce “Get Data”. A seguito di ciò, viene visualizzata la finestra “Recupera dati” (Figura 2.7) con diversi connettori tramite i quali è possibile scegliere tra una serie di sorgenti differenti a cui Power BI può interfacciarsi. Tra queste sorgenti emergono

- File system, quali Excel, CSV, XML, JSON etc;
- Sistemi relazionali, quali MySQL, Oracle, SQL Server, etc;
- Sistemi cloud, quali Microsoft Azure, Google Analytics, Salesforce report, etc.

Una volta cliccato sulla sorgente dati di interesse, viene mostrata una pre-view della cartella di lavoro all'interno della quale Power BI Desktop cerca di interpretarne il contenuto. Questa sorta di anteprima dei dati è utile per capire se il software è riuscito a comprendere appieno la forma dei dati, valutando le intestazioni, la mancanza di alcune righe oppure l'errata interpretazione di un numero come una stringa. In caso di corretta esposizione, il foglio di lavoro può essere caricato direttamente nel modello dei dati e pronto per la creazione dei grafici; altrimenti è necessario la modifica e la pulizia dei dati.



**Figura 2.7.** Finestra di recupero dei dati Power in BI Desktop

### 2.2.1 Editor di query

L'applicazione Power BI Desktop include, anche, la componente Editor di query, una raccolta di strumenti estremamente utili nella fase di pulizia e manipolazione dei dati, dove vengono riorganizzate, secondo la logica e l'analisi che si vuole adottare, tutte le tabelle importate prima di poter iniziare a creare report.

L'Editor di query, dunque, viene aperto in una finestra separata e consente di trasformare dati, creare query, ed infine, caricare il modello dei dati ottimizzato in Power BI Desktop.

L'Editor di query permette, quindi, di realizzare operazioni di ETL sui dati; mediante tali operazioni si possono, ad esempio, rimuovere colonne o combinare i dati da più sorgenti; è possibile, inoltre, partire da un blocco di dati di grandi dimensioni e rimuovere o aggiungere porzioni in base alle necessità, finché le tabelle non ottengono esattamente la forma desiderata, adatta, all'uso nei report.

La Figura 2.8 mostra la schermata iniziale dell'Editor delle query costituita principalmente da quattro sezioni:

- Nella barra multifunzione ci sono diversi pulsanti per interagire con i dati nella query, quali "Home", "Trasforma", "Aggiungi colonna" e "Visualizza".
- Nel riquadro a sinistra, le query vengono elencate e rese disponibili per la selezione, visualizzazione e il data shaping. Quando si seleziona una query da sinistra, i relativi dati vengono visualizzati nella sezione centrale, dove è possibile, poi, applicare tutte le trasformazioni necessarie.
- Nel riquadro centrale, vengono visualizzati i dati della query selezionata, disponibili per il data shaping.

- La finestra “Impostazioni query” elenca tutte le proprietà della query e i passaggi applicati. Ogni passaggio effettuato durante la trasformazioni dei dati (ad esempio, rinominare una tabella, trasformare un tipo di dati o eliminare delle colonne) viene registrato dall’Editor di Power query ed eseguito ogni volta che la query creata si connette all’origine dati, in modo che quest’ultimi abbiano sempre la forma specificata.

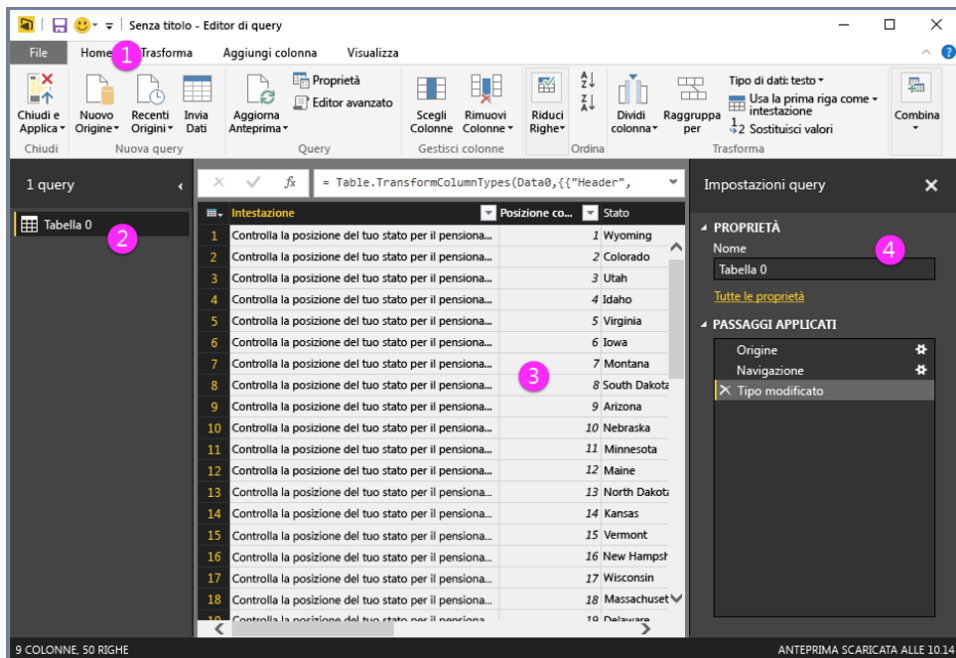


Figura 2.8. Finestra dell’Editor di query

## 2.3 Realizzazione di report

Dopo aver creato un modello di dati e, successivamente, importato in Power tramite l’Editor di Power query, è necessario creare una raccolta di oggetti visivi che illustrano i vari aspetti dei dati usati per creare il modello in Power BI Desktop.

Una raccolta di oggetti visivi in un file di Power BI viene chiamato report; questo a sua volta, può contenere una o più pagine.

Con Power BI Desktop, dunque, è possibile creare report altamente interattivi e personalizzabili, che si aggiornano al variare dei dati sottostanti; tali report, infine, una volta pronti, possono essere condivisi con gli altri utenti. Nella Figura 2.9 viene visualizzata la schermata di un esempio di report denominato “Panoramica” come si evince sulla scheda presente nella parte inferiore.

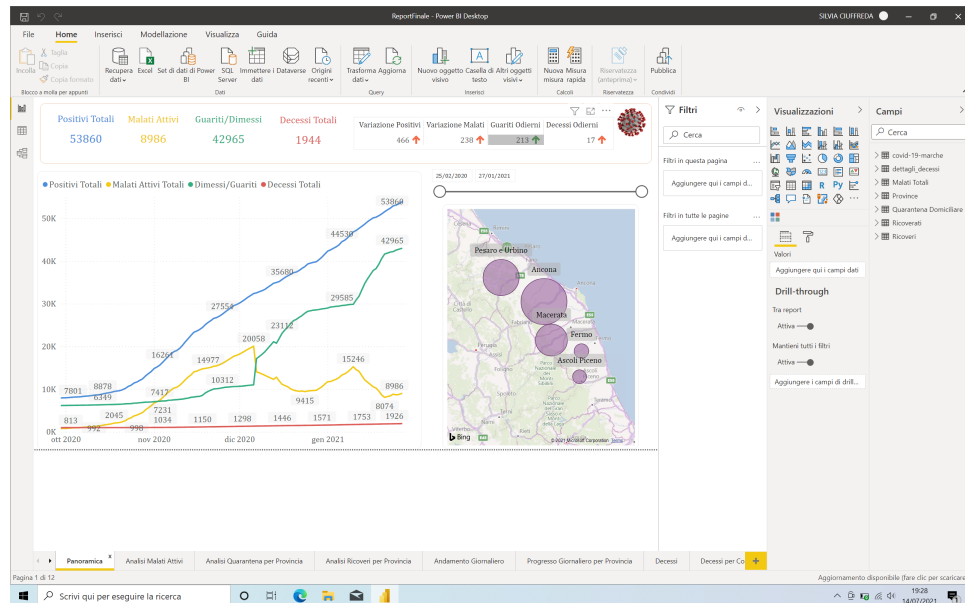


Figura 2.9. Esempio di una pagina di un report in Power BI

### 2.3.1 Il foglio di lavoro

Prima di esaminare il processo di analisi svolto per la realizzazione di report e dashboard in questa tesi, è opportuno descrivere la struttura di un tipico foglio di lavoro dell'applicazione Power BI Desktop (Figura 2.10). Una volta avviata l'applicazione, viene visualizzato un foglio con tre sezioni di lavoro ben definite, le cui icone sono illustrate nella Figura 2.11:

- *Report*: consente di usare le query create dall'utente per generare visualizzazioni interessanti, organizzate nella sequenza preferita e con più pagine, che è possibile condividere con altri utenti.
- *Dati*: consente di visualizzare i dati caricati nel report in un formato di modello di dati, in cui è possibile aggiungere misure, creare nuove colonne e gestire le relazioni.
- *Relazioni*: consente di ottenere una rappresentazione grafica delle relazioni che sono state stabilite nel modello dei dati, e di gestirle o modificarle in base alle esigenze.

Nella parte superiore vi è la barra multifunzione con una serie di strumenti utili per la creazione di grafici. Nella parte centrale c'è una pagina vuota dove si vanno a creare i diversi grafici e i relativi report. Sulla destra, invece, viene mostrata la finestra "Visualizzazioni", con cui è possibile scegliere quale grafico utilizzare per visualizzare i dati, e il menù "Filtri", che consente di impostare diversi filtri agli oggetti visivi e alle pagine create.

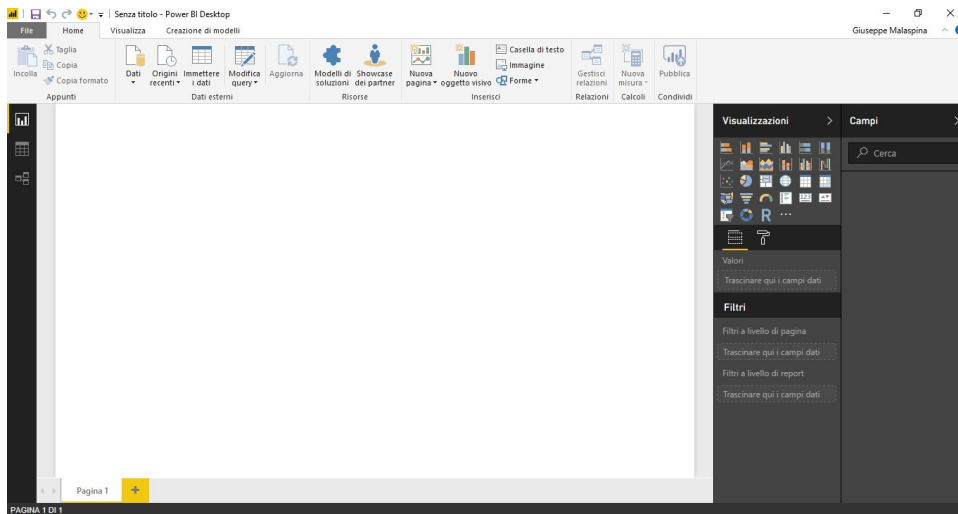


Figura 2.10. Foglio di lavoro in Power BI Desktop



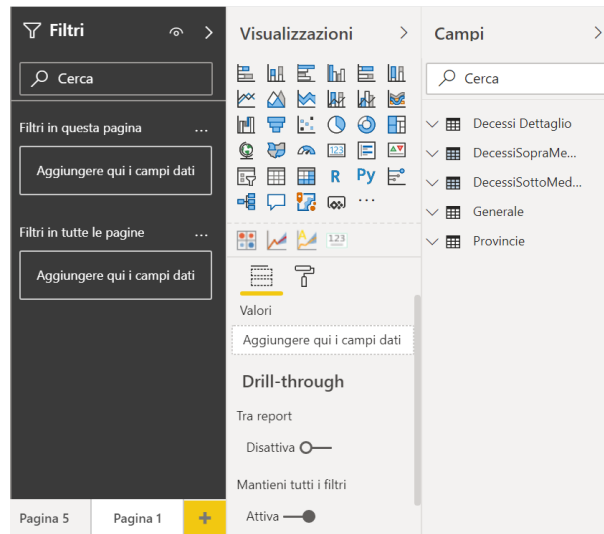
Figura 2.11. Icone delle viste di Power BI

### Visualizzazione Report

Nella schermata iniziale dell'applicazione Power BI ciò che domina gran parte l'occupazione è, sicuramente, la visualizzazione *Report*. Quando vengono caricati i dati in Power BI Desktop per la prima volta, ovviamente, la visualizzazione *Report* contiene un'area di disegno vuota. Da questo punto è possibile iniziare la creazione vera e propria di grafici e posizionarli all'interno del foglio di lavoro vuoto.

In questa pagina iniziale, in particolare, sono visibili tre riquadri: "Filtri", "Visualizzazioni" e "Campi" (Figura 2.12). I primi due riquadri a sinistra, "Filtri" e "Visualizzazioni", controllano l'aspetto delle visualizzazioni: tipo di colori, filtraggio e formattazione. L'ultimo riquadro a destra, "Campi", gestisce i dati sottostanti utilizzati nelle visualizzazioni. Il contenuto visualizzato nell'editor di report, infine, varia a seconda delle selezioni effettuate nell'area di visualizzazione di report.

Durante la modellazione dei dati, talvolta, si vuole vedere il contenuto effettivo di una tabella o una colonna senza creare un elemento visivo nell'area di disegno del report, spesso proprio a livello di riga.



**Figura 2.12.** Riquadri laterali in visualizzazione *Report*

### Visualizzazione *Dati*

La visualizzazione dei dati consente di esaminare, esplorare ed interpretare i dati nel modello di Power BI (Figura 2.13). Essa è diversa dalla visualizzazione di tabelle, colonne e dati nell'Editor di query, perché, in questo caso, i dati vengono visualizzati dopo essere stati caricati nel modello. Questo risulta particolarmente utile per la creazione di misure e colonne calcolate, o quando è necessario identificare un tipo di dati o una categoria di dati.

### Visualizzazione *Relazioni*

La visualizzazione relazioni mostra tutte le tabelle, le colonne e le relazioni presenti nel modello dati (Figura 2.14). Ciò può essere particolarmente utile quando un modello contiene relazioni complesse tra molte tabelle.

Le relazioni tra le tabelle sono necessarie per calcolare con precisione i risultati e visualizzare le informazioni corrette nei report.

Power BI, inoltre, semplifica la creazione di queste relazioni; infatti nella maggior parte dei casi, tutte le operazioni verranno eseguite automaticamente dalla funzionalità "Rilevamento automatico". Tuttavia, in alcuni casi, potrebbe essere necessario creare manualmente le relazioni, oppure apportare alcune modifiche a una relazione rilevata automaticamente dall'applicazione. È, quindi, importante comprendere le relazioni in Power BI e capire come crearle e modificarle.

### 2.3.2 Gli oggetti visivi supportati

Generalmente, in un set di dati, si nascondono informazioni dettagliate importanti che possono aiutare a sviluppare attività di Business Intelligence. Tuttavia, non

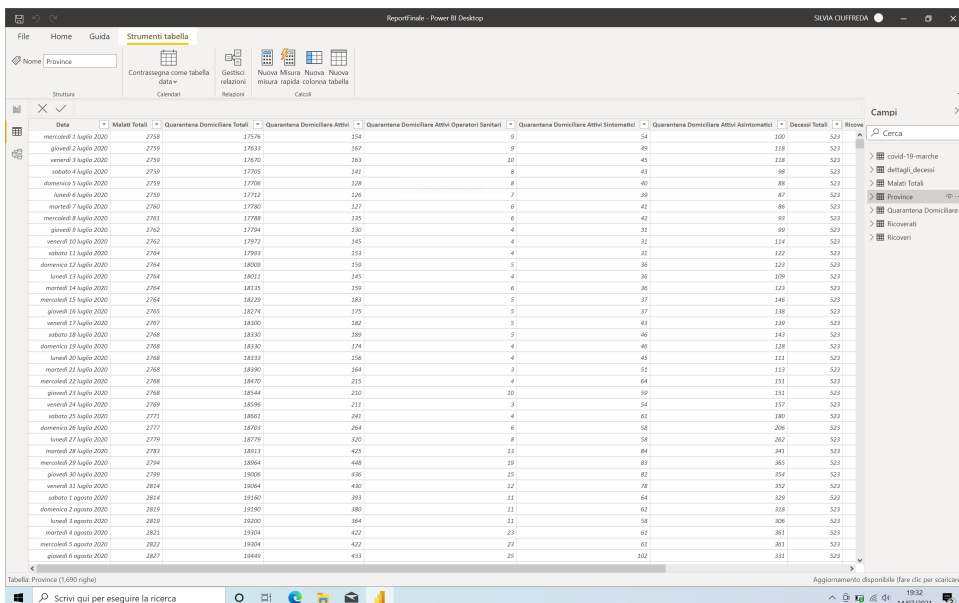


Figura 2.13. Visualizzazione *Dati*

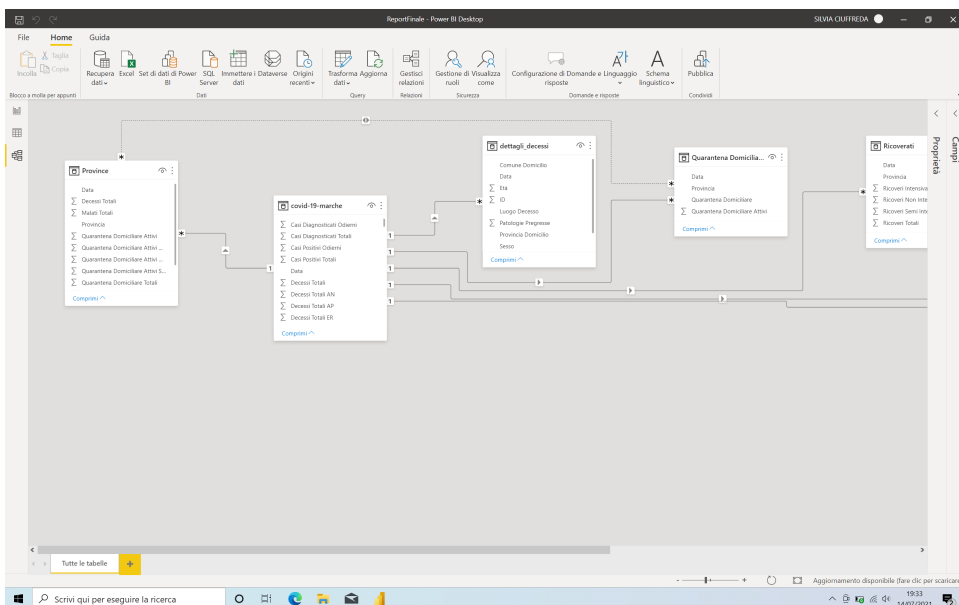


Figura 2.14. Visualizzazione *Relazioni*

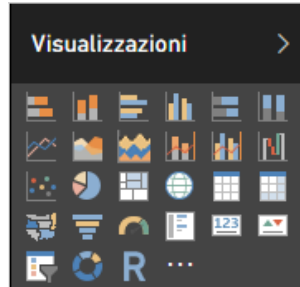
è sempre facile trovare i collegamenti esaminando semplicemente numeri non elaborati. La visualizzazione grafica dei dati, invece, risulta molto utile per la comprensione immediata di dati, in quanto, in un' enorme quantità di dati in formato visivo, è possibile notare connessione ed estrapolazione di conoscenza che, altrimenti,



rimarrebbero fuori.

Con la visualizzazione, dunque, i numeri prendono vita consentendo di accedere alle informazioni nascoste all'interno del dataset. Grazie a dashboard in tempo reale, report interattivi, grafici, diagrammi e altre rappresentazioni visive, la visualizzazione dei dati aiuta a sbloccare preziose informazioni chiave per il Business Intelligence in modo rapido ed efficace.

Power BI offre 24 visualizzazioni già predefinite all'interno dell'applicazione (Figura 2.15), ma molte altre sono reperibili dal Marketplace. Nel seguito verranno descritte in dettaglio gli oggetti visivi più utilizzati e funzionali per questa tesi.



**Figura 2.15.** Grafici prefiniti in Power BI

### L'oggetto visivo "Scheda"

A volte l'unico elemento che si vuole visualizzare in una dashboard o in un report di Power BI è, semplicemente, un numero, ad esempio il numero di persone attualmente positive o il numero di persone decedute a causa del Coronavirus. Questo tipo di visualizzazione è denominato "scheda". Come quasi tutte le visualizzazioni native di Power BI, è possibile creare le "Schede" tramite l'editor di report (Figura 2.16).

Positivi Totali	Malati Attivi	Guariti/Dimessi	Decessi Totali
53860	8986	42965	1944

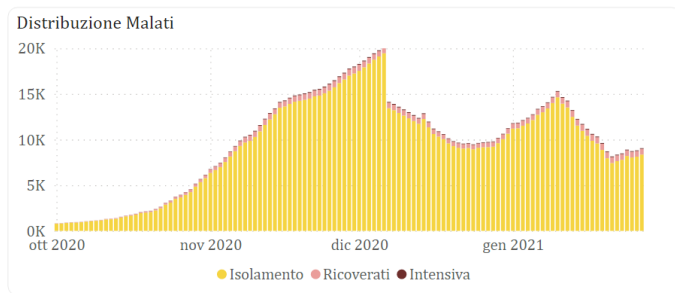
**Figura 2.16.** Esempio di oggetto visivo "Scheda"

### Gli istogrammi

Un istogramma è un diagramma che fornisce una rappresentazione di un insieme di dati statistici mediante un grafico a barre. Gli istogrammi possono essere rappresentati tramite barre orizzontali o verticali e consentono di aggregare determinati

numeri in gruppi di intervalli uguali, detti raccoglitori, nonchè di visualizzare la frequenza di valori all'interno di ciascun raccoglitore.

Un istogramma (Figura 2.17) è costituito da una successione di rettangoli adiacenti, ciascuno dei quali ha per base un certo intervallo della variabile e un'altezza, tale che la sua area rappresenti, nella scala prefissata, il relativo valore globale della funzione.



**Figura 2.17.** Esempio di istogramma

### Le tabelle

Una tabella è una griglia contenente dati correlati in una serie logica di righe e colonne. Può anche contenere intestazioni e un riga per i totali. Le tabelle funzionano bene con confronti quantitativi in cui si analizzano molti valori per una singola categoria. Ad esempio, la Figura 2.18 mostra il numero dei deceduti per ciascuna provincia della regione Marche.

Le tabelle, inoltre, rappresentano un'ottima scelta nelle seguenti situazioni:

- visualizzare e confrontare dati dettagliati e valori esatti (anziché rappresentazioni visive);
- visualizzare i dati in un formato tabulare;
- visualizzare i dati numerici per categorie.

### I grafici a linee

I grafici a linee (Figura 2.19) evidenziano la forma complessiva di un'intera serie di valori, in genere nel periodo temporale. Essi permettono, anche, di rappresentare l'andamento temporale di più serie di valori; si hanno, cioè, più linee in un'unica rappresentazione.

### Il filtro dei dati

Un filtro dei dati (Figura 2.20) è un grafico autonomo che può essere usato per filtrare gli altri oggetti visivi nella pagina. I filtri dei dati sono disponibili in molti formati diversi (categoria, intervallo, data, e così via) e possono essere formattati

Provincia	Malati Totali	Variazione Giornaliera
Ancona	15455	177 ↑
Ascoli Piceno	6133	42 ↑
Extra Regione	1970	16 ↑
Fermo	6337	54 ↑
Macerata	11385	79 ↑
Pesaro e Urbino	12580	98 ↑
<b>Totale</b>	<b>53860</b>	<b>466</b>

Figura 2.18. Esempio di tabella

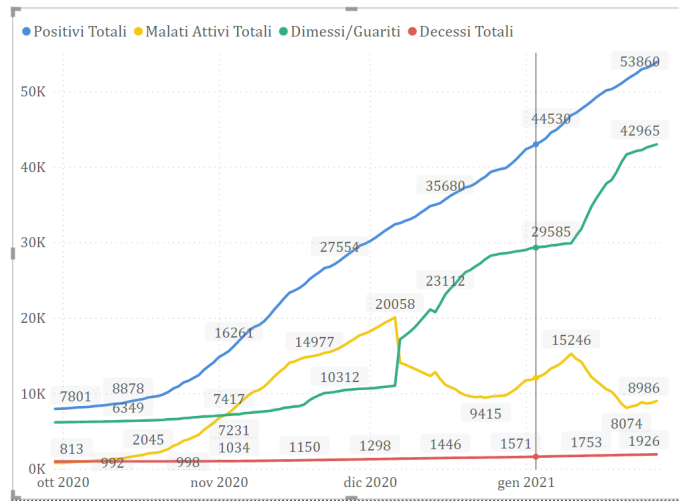


Figura 2.19. Esempio di grafico a linee

per consentire la selezione di uno solo, di molti o di tutti i valori disponibili. I filtri dei dati rappresentano un’ottima scelta per:

- visualizzare filtri importanti o di uso comune nell’area di disegno del report in modo da facilitare l’accesso;
- facilitare la visualizzazione dello stato filtrato corrente senza dover aprire un elenco a discesa;
- filtrare per colonne non necessarie e nascoste nelle tabelle dei dati;
- creare report più mirati inserendo i filtri dei dati in corrispondenza di oggetti visivi importanti.



Figura 2.20. Esempio di filtro dei dati

## I grafici combinati

Quando si vogliono visualizzare due misure che hanno scale diverse, ad esempio i casi positivi odierni e il totale casi positivi, è molto utile un grafico combinato, che include un istogramma e un grafico a linee con scale degli assi diverse.

Un grafico combinato (Figura 2.21) è costituito da un campo “Asse condiviso” (asse  $x$ ) e quindi i valori per i due campi, in questo caso una colonna e una linea. Le due legende dell’asse  $Y$  appaiono ai due lati della visualizzazione. È anche possibile, inoltre, suddividere ogni colonna per categoria, trascinando una categoria nel campo “Serie di colonne” nel riquadro “Visualizzazioni”. Così facendo, ogni barra viene colorata in modo proporzionale in base ai valori all’interno di ogni categoria.

I grafici combinati rappresentano un’ottima scelta nelle seguenti situazioni:

- quando si ha un grafico a linee e un istogramma con lo stesso asse  $X$ ;
- per confrontare più misure con intervalli di valori diversi;
- Per illustrare la correlazione con intervalli di valori diversi;
- per verificare se una misura raggiunge la destinazione definita da un’altra misura;
- per risparmiare spazio nell’area di disegno.

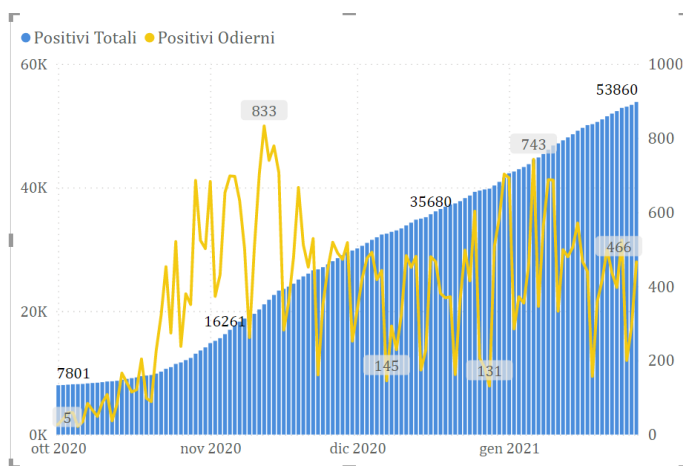
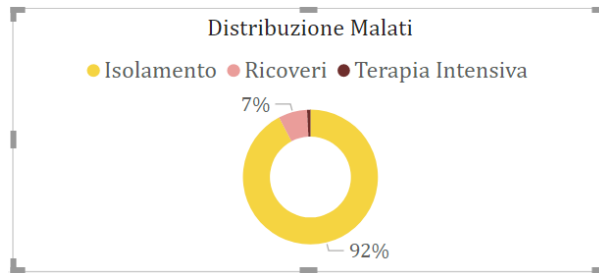


Figura 2.21. Esempio di grafico combinato

## I grafici ad anello

Un grafico ad anello (Figura ??) mostra la relazione delle parti rispetto a un intero. In questo grafico i dati vengono rappresentati sotto forma di anello a spicchi, con le relative percentuali. A differenza del grafico a torta, molto simile per certi aspetti, il centro è vuoto e consente di inserire un’etichetta o un’icona all’interno.



**Figura 2.22.** Esempio di grafico ad anello

### I grafici ad aree

Il grafico ad area si basa sul grafico a linee, in cui l'area compresa tra l'asse e la linea viene riempita. I grafici ad aree enfatizzano l'entità del cambiamento nel tempo e possono essere usati per attirare l'attenzione sul valore in una tendenza.

### Le mappe

Le mappe hanno la particolarità di associare informazioni categoriche e quantitative a posizioni nello spazio. Power BI dispone, in realtà, di diverse tipologie di visualizzazione delle mappe: la mappa a bolle, che sostituisce i dati con bolle su un punto geografico (Figura 2.23), e la mappa di forme, che confronta le aree di una mappa usando i colori. Nella presente tesi è stato utilizzato principalmente il grafico a bolle, il quale esegue il rendering dei dati di posizione come calcoli ridimensionati sulla mappa. In particolare, Power BI si integra con Bing Maps per fornire coordinate della mappa predefinite (un processo chiamato decodifica) in modo da poter creare mappe. Insieme usano algoritmi per identificare la posizione corretta; a volte, però, si tratta soltanto di un'ipotesi migliore. Infatti, se Power BI tenta di creare la visualizzazione della mappa in autonomia ma non ci riesce, richiede l'aiuto di Bing Maps.

### Indicatori KPI

Un indicatore di prestazioni chiave (Key Performance Indicator - KPI) è un segnale visivo che comunica lo stato di avanzamento verso un obiettivo misurabile (Figura 2.24).

Gli indicatori KPI rappresentano un'ottima scelta nelle seguenti situazioni:

- misura dello stato di avanzamento, in modo da stabilire se si è avanti o indietro;
- misura della distanza da un obiettivo, per stabilire quanto manca alla fine.

### 2.3.3 Il pannello Filtri

Una peculiarità di Power BI è la possibilità di filtraggio di dati al fine di estrapolare nuove informazioni dettagliate in essi. Nella visualizzazione Report, nel riquadro "Filtri" è possibile applicare diversi filtri agli oggetti visivi e alle pagine. Come si evince dalla Figura 2.25, Power BI mette a disposizione tre tipologie di filtri:

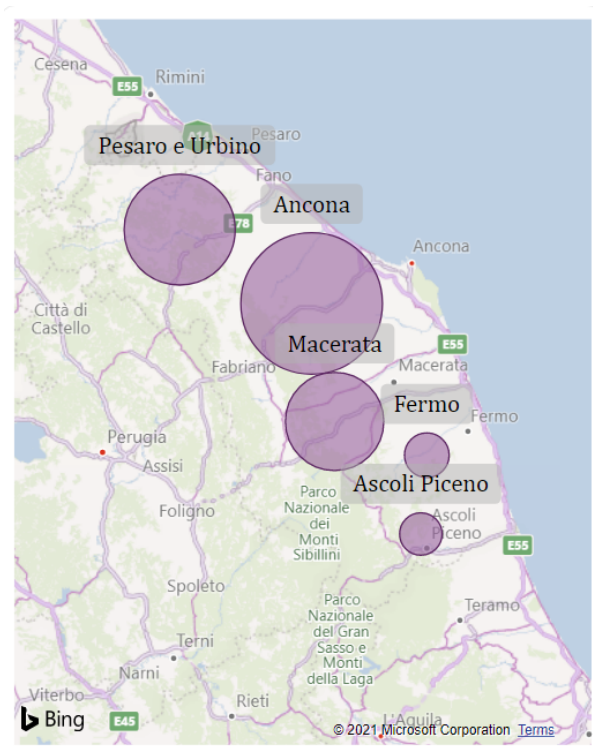


Figura 2.23. Esempio di mappa a bolle

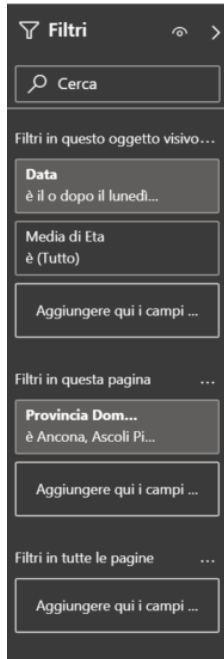


Figura 2.24. Esempio di indicatore KPI

- *Filtro della pagina*: si applica a tutti gli oggetti visivi nella pagina del report;
- *Filtro visivo*: si applica a un singolo oggetto visivo in una pagina del report;
- *Filtro di report*: si applica a tutte le pagine del report.

Il riquadro “Filtri” permette, inoltre, di filtrare un campo che non è già presente in una delle sezioni del report. Ogni singolo filtro ha dei sottofiltri, che servono per selezionare ulteriormente il dato in analisi. In particolare, si hanno i seguenti sottofiltri:

- *Filtro avanzato*: consiste nel filtrare i dati imponendo delle condizioni ben precise (ad esempio, grandezza, maggiore o minore di, uguale a, etc).
- *Filtro di base*: utile quando si devono selezionare dei valori ben precisi (ad esempio, anno, mese e giorno).
- *Filtro primi N*: utile nel caso in cui bisogna effettuare delle analisi su alcuni valori specifici.



**Figura 2.25.** Il riquadro “Filtri” nella visualizzazione *Report*

### 2.3.4 Formule DAX

A volte il modello di dati elaborato in Power query, e successivamente importato in Power BI Desktop, non contiene tutte le informazioni necessarie per l’attività di analisi della Business Intelligence. In questi casi, Power BI consente di creare nuove misure. In altre parole, tramite Power BI è possibile creare personalmente calcoli più avanzati usando una formula di tipo DAX.

DAX è l’acronimo di Data Analysis Expressions e indica il linguaggio delle formule usato in Power BI che permette di definire delle espressioni mettendo a disposizione una serie di funzioni.

DAX include una libreria con oltre 200 funzioni, operatori e costrutti per eseguire query e calcoli avanzati sui dati presenti in tabelle e colonne correlate nei modelli di dati tabulati. Tale libreria, infatti, offre un’enorme flessibilità per la creazione di misure di calcolo dei risultati per quasi tutte le esigenze di analisi dei dati.

Usare DAX all’interno di Power BI, dunque, è assolutamente indispensabile per sfruttarne pienamente la potenza e ottenere risultati altrimenti non realizzabili.

Quando si creano delle formule DAX è importante considerare il tipo dei dati per evitare errori o risultati non coerenti. In particolare quando si definisce un’espressione per una misura nella barra della formula, tramite una descrizione comanda, viene visualizzata un’anteprima sui possibili risultati per il totale nel contesto corrente. Generalmente i risultati non vengono restituiti immediatamente in nessuna posizione. Il motivo per cui non è possibile visualizzare i risultati (filtrati) del calcolo immediatamente è dovuto al fatto che il risultato di una misura non può essere determinato senza contesto.



**Figura 2.26.** Il riquadro “Campi” dell’area *Report* con misure calcolate

Nelle formule DAX, inoltre, vanno sempre specificate tabelle e colonne (se il nome della tabella contiene spazi, bisogna scriverlo tra apici). Le colonne e le misure sono racchiuse tra parentesi quadre.

È importante sottolineare che i risultati delle misure DAX calcolate vengono aggiornati costantemente in base alle operazioni dell’utente nei report e consentono un’esplorazione dei dati rapida e dinamica.

Una volta realizzata una misura, questa viene visualizzata in visualizzazione *Report* o in vista *Dati* e i rispettivi valori vengono mostrati nell’elenco di campi con un’icona a forma di calcolatrice (Figura 2.26).

La Figura 2.27 mostra la formula DAX utilizzata per la creazione di una misura



calcolata; essa è costituita dal nome della misura seguita dall'uguale e dalla funzione che si vuole utilizzare con i rispettivi argomenti da cui andare a prendere i dati.

In particolare, il linguaggio DAX consente, anche, di creare:

- colonne calcolate;
- tabelle calcolate.

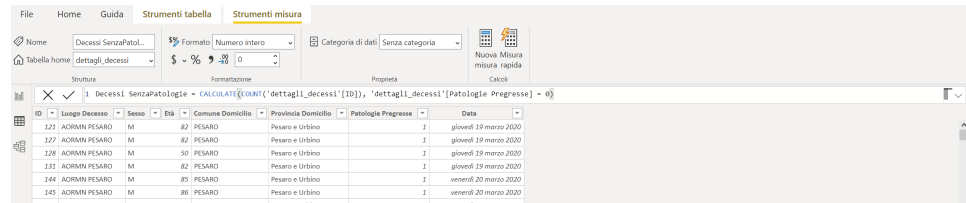


Figura 2.27. Esempio di misura calcolata con DAX

### Colonne calcolate

Con le colonne calcolate è possibile aggiungere nuovi dati a una tabella già presente nel modello. Tuttavia, invece di eseguire query e caricare i valori nella nuova colonna da un'origine dati, viene creata una formula DAX che definisce i valori della tabella. In Power BI Desktop, le colonne calcolate vengono create usando la funzionalità “Nuova colonna” nella visualizzazione Report.

Le colonne calcolate create vengono visualizzate nell'elenco “Campi” come qualsiasi altro campo, ma hanno un'icona speciale per indicare che i valori sono il risultato di una formula. È possibile, inoltre, assegnare qualsiasi nome alle colonne e aggiungerle a una visualizzazione del report con le normali procedure usate per gli altri campi.

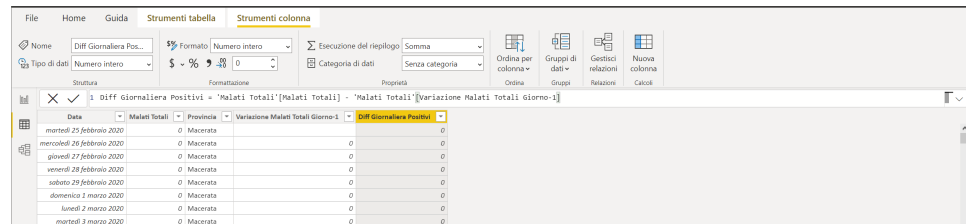


Figura 2.28. Esempio di colonna calcolata con DAX

### Tabelle calcolate

Una tabella calcolata è un oggetto calcolato, basato su un'espressione di formula o derivato da tutte le altre tabelle nello stesso modello, oppure da alcune di esse. Invece di eseguire query e caricare i valori nelle colonne della nuova tabella da un'origine

dati, una formula DAX definisce i valori della tabella. Creando una tabella calcolata è possibile ottenere una tabella autonoma disponibile per le query, perfettamente utilizzabile come qualsiasi altra tabella. Le tabelle calcolate sono utili anche per la configurazione di un set di righe filtrato o di un subset di colonne di altre tabelle esistenti. Ciò consente di mantenere intatta la tabella originale durante la creazione di varianti della stessa per supportare scenari specifici.

## 2.4 Pubblicazione e condivisione di dashboard

Una volta ultimata la fase di analisi dei dati all'interno delle aree di lavoro, il progetto è pronto per essere condiviso con utenti interni o esterni all'organizzazione.

La condivisione, infatti, è il modo più semplice per concedere agli utenti l'accesso ai report e alle dashboard nel servizio Power BI (Figura 2.29)

Una delle peculiarità più interessanti che Power BI mette a disposizione è la possibilità di creare aree di lavoro di gruppo; queste rappresentano un luogo comune per collaborare con i colleghi nelle dashboard, nei report e nei set di dati.

Dopo aver creato un file di Power BI Desktop, se questo viene pubblicato nell'area di lavoro del gruppo di Power BI, ogni membro del gruppo può collaborare alla sua elaborazione.

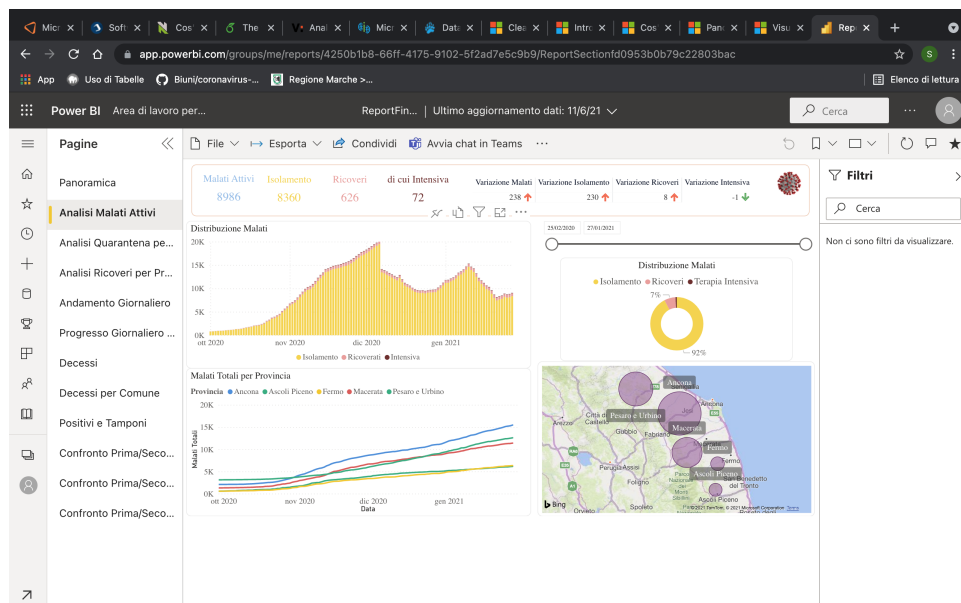


Figura 2.29. Esempio di dashboard pubblicata in Power BI Service

## Descrizione e costruzione del dataset

*In questo capitolo verranno illustrati i dati utilizzati nella presente tesi. Nella prima parte, in particolare, sarà descritta la sorgente da cui i dati provengono e la struttura dei dati stessi. Nella seconda parte, invece, si illustreranno le operazioni eseguite per la creazione del dataset, dalla realizzazione dei file JSON giornalieri all'aggregazione di quest'ultimi utilizzando Script Python.*

### 3.1 Sorgenti dei dati

Lo scopo della tesi è quello di analizzare i dati relativi alla seconda ondata di contagi da COVID-19 nella Regione Marche, al fine di descrivere l'andamento della diffusione della pandemia all'interno della Regione ed ottenere, infine, analogie e differenze rispetto gli sviluppi della prima ondata nel territorio marchigiano.

La sorgente dei dati utilizzata per realizzare l'analisi in Power BI Desktop è un repository pubblico (Figura 3.1), consultabile al sito <https://github.com/Biuni/coronavirus-marche>. Tale repository non è né collegato né sponsorizzato dalla Regione Marche, ed il suo autore è *Gianluca Bonifazi* (<https://biuni.it>).

I dati riguardanti la diffusione dell'epidemia all'interno della Regione Marche passano per le competenze tecniche del GORES (Gruppo Operativo Regionale per le Emergenze Sanitarie). Si tratta di un tavolo tecnico che la regione ha istituito da tempo, in modo strutturale, per fornire risposte tempestive alle emergenze sanitarie. Il GORES è attivo sul COVID-19 sin dal 27 Gennaio 2020, prima dell'arrivo del contagio in Italia, per strutturare il lavoro in base alle principali esigenze poste dalla specifica situazione. Ad esempio, esso traccia e trasmette le indicazioni operative definendo le modalità di presa in carico di un potenziale caso sospetto di COVID-19, le relative modalità di isolamento, diagnosi e trattamento ed, infine, l'attuazione delle più idonee misure di sanità pubblica per contenere l'infezione ed evitare eventuali ulteriori casi.

Il GORES, dunque, si è fin da subito impegnato nel fornire, con grande chiarezza e trasparenza, i dati circa la diffusione del virus all'interno delle province della Regione Marche. I dati divulgati, inoltre, hanno una cadenza giornaliera e riguardano, in particolare, gli aspetti relativi ai tamponi, ai ricoveri e ai decessi. Il repository

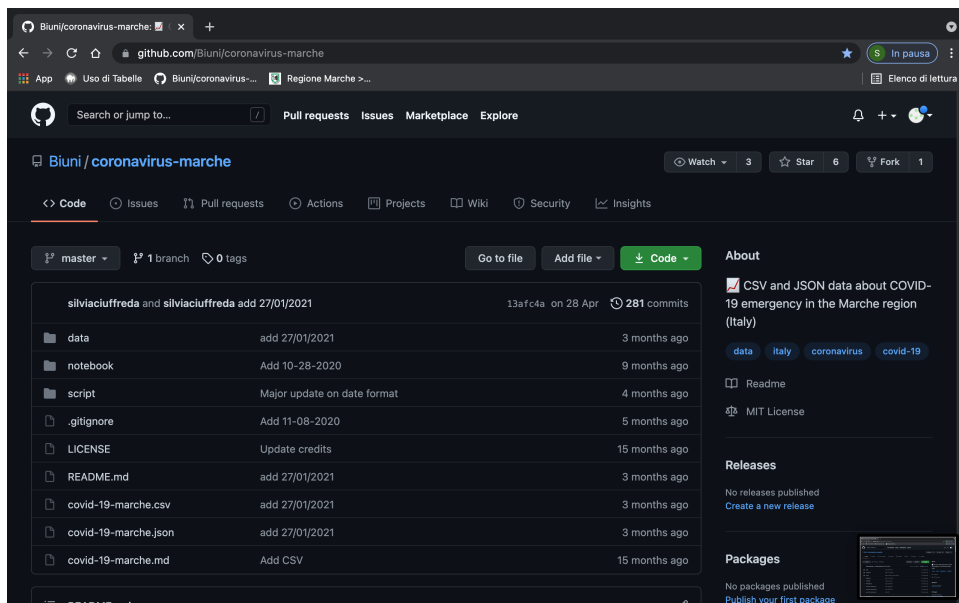


Figure 3.1. Repository di GitHub utilizzato per l’importazione dei dati

di riferimento aggrega proprio questi dati messi a disposizione dal GORES (diffusi soltanto in formato PDF) in un formato utile ai fini dell’analisi e delle statistiche.

### 3.1.1 Dati forniti dalla Regione Marche

Nel seguito si analizzeranno, nel dettaglio, i dati messi a disposizione dalla Regione Marche e, successivamente, raccolti all’interno del repository GitHub di interesse.

I dati, raccolti ed elaborati dal GORES, vengono rilasciati quotidianamente in tre diversi report (*Blu*, *Giallo* e *Arancio*). Questi ultimi vengono resi pubblici in un formato PDF (poco adatto alle analisi) e pubblicati direttamente sul sito della Regione stessa, in una sezione completamente dedicata al COVID-19 (Figura 3.2).

I tre report propagati quotidianamente contengono, in particolare, le seguenti informazioni:

- *Report Blu*: documento nel quale vengono rilasciati il numero di tamponi effettuati nelle ultime 24 ore ed il numero di nuovi positivi nelle ultime 24 ore.
- *Report Giallo*: documento nel quale viene aggiornata la panoramica sulla situazione dei casi risultati positivi al coronavirus (ricoverati, isolamento domiciliare, dimessi e guariti, etc.).
- *Report Arancio*: documento nel quale vengono rilasciati i numeri e i dettagli sui decessi avvenuti nelle ultime 24 ore.

Si esaminano, ora, più da vicino i tre report sopra esposti.

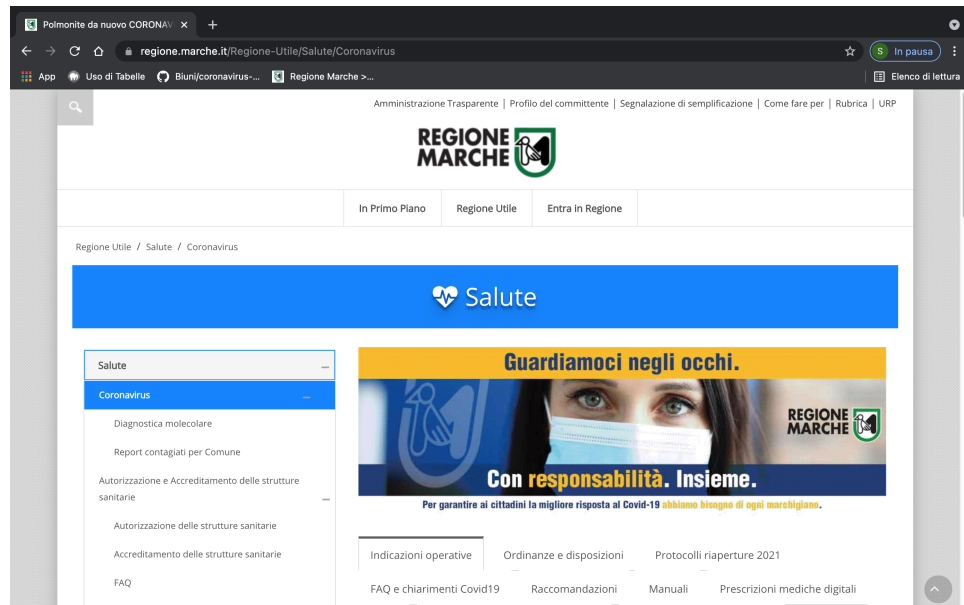


Figura 3.2. La pagina del sito della Regione Marche dedicata al Coronavirus

### Report Blu

Il *report Blu* è suddiviso in due sezioni:

- Nella prima viene riportata una tabella che illustra, dal giorno di inizio della raccolta dei dati (24 Febbraio 2020), e per ogni giorno successivo, i *Campioni Positivi giornalieri per SARS-Cov-2* e i *Campioni complessivamente testati/giorno*.
- Nella seconda, invece, viene riportata un'ulteriore tabella che, sempre con riferimento ad ogni singolo giorno, riporta i *Campioni Positivi totali: progressione* e i *Campioni testati totali: progressione*.

In questo report vengono, inoltre, rilasciati due grafici relativi ai *Campioni Positivi giornalieri per SARS-Cov-2* e ai *Campioni Positivi totali: progressione* (Figura 3.3).

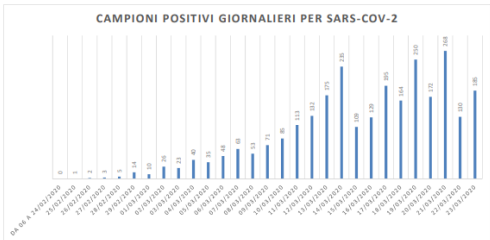
### Report Giallo

Il *report Giallo* (Figura 3.4) raccoglie i dati relativi sullo stato dei contagiati al COVID-19 ed è organizzato in 11 tabelle:

- La prima riassume le informazioni relative al totale dei test effettuati e diagnosticati e al totale dei test positivi.
- La seconda riassume le informazioni relative ai test effettuati e diagnosticati giornalieri, nonchè i casi positivi giornalieri.



giorno	Campioni Positivi giornalieri per SARS-Cov-2	Campioni complessivamente testati/giorno
da 06 a 24/02/2020	0	14
25/02/2020	1	5
26/02/2020	3	6
27/02/2020	3	19
28/02/2020	5	19
29/02/2020	14	33
01/03/2020	10	36
02/03/2020	26	64
03/03/2020	23	89
04/03/2020	40	126
05/03/2020	35	171
06/03/2020	48	230
07/03/2020	63	218
08/03/2020	53	216
09/03/2020	71	297
10/03/2020	85	229
11/03/2020	113	251
12/03/2020	132	311
13/03/2020	175	333
14/03/2020	281	386
15/03/2020	329	300
16/03/2020	129	335
17/03/2020	195	537
18/03/2020	164	403
19/03/2020	250	658
20/03/2020	172	570
21/03/2020	268	651
22/03/2020	180	373
<b>23/03/2020</b>	<b>185</b>	<b>465</b>



giorno	Campioni Positivi totali: progressione	Campioni testati totali: progressione
da 06 a 24/02/2020	0	14
25/02/2020	1	19
26/02/2020	3	27
27/02/2020	6	46
28/02/2020	11	65
29/02/2020	25	98
01/03/2020	35	134
02/03/2020	61	198
03/03/2020	86	287
04/03/2020	124	413
05/03/2020	159	586
06/03/2020	207	816
07/03/2020	279	1014
08/03/2020	323	1250
09/03/2020	394	1437
10/03/2020	479	1656
11/03/2020	597	1907
12/03/2020	724	2118
13/03/2020	898	2551
14/03/2020	1133	2937
15/03/2020	1442	3237
16/03/2020	1371	3570
17/03/2020	1567	4109
18/03/2020	1731	4512
19/03/2020	1981	5170
20/03/2020	2153	5760
21/03/2020	2421	6391
22/03/2020	2551	6764
<b>23/03/2020</b>	<b>2736</b>	<b>7229</b>

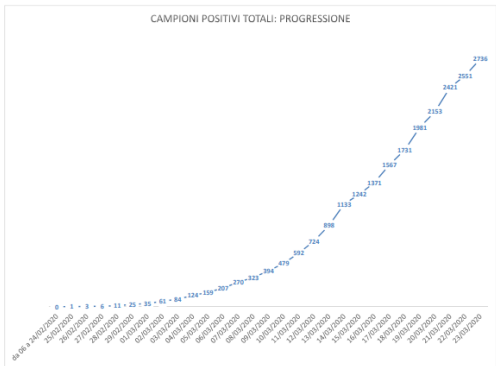


Figura 3.3. Il report Blu pubblicato giornalmente dalla Regione Marche

- La terza suddivide i malati in base alla criticità della loro situazione. In particolare, vengono riportati i casi positivi distribuiti per tipo di ricovero, i dimessi e i guariti, i malati in isolamento domiciliare e il totale dei deceduti.
- La quarta illustra i casi positivi per singola provincia marchigiana e comprende, anche, i cosiddetti casi positivi *EXTRA REGIONE*.
- La quinta illustra i ricoveri in terapia intensiva, suddivisi per struttura ospedaliera.
- La sesta illustra i ricoveri in terapia semi intensiva, suddivisi per struttura ospedaliera.
- La settima illustra i ricoveri in terapia non intensiva, suddivisi per struttura ospedaliera.
- L’ottava illustra i ricoveri post critica, suddivisi per struttura territoriale.
- La nona illustra i ricoveri in pronto soccorso, suddivisi per struttura ospedaliera.
- La decima riporta le variazioni dei casi positivi, dei ricoveri, dei dimessi e dei guariti nonché i test effettuati rispetto al giorno precedente.
- L’undicesima si concentra soltanto sui *casi e contatti in isolamento domiciliare*, distinguendo tra sintomatici, asintomatici e operatori sanitari, e riportando i dati per ogni provincia marchigiana.

**REGIONE MARCHE** SERVIZIO SANITÀ - Osservatorio Epidemiologico Regionale - ARS **aggiornamento: 27/01/2021 ore 12.00**

TEST PER SARS-CoV-2			
TOTALE casi diagnosticati da inizio pandemia	casi positivi	casi negativi	TOTALE test (tamponi) effettuati*
403.378	53.860	349.518	669.151 * compresi test ripetuti sulla stessa persona e percorso guariti

Casi giornalieri diagnosticati	casi positivi	casi negativi	tamponi effettuati*
3.301	466	2.835	5.452 * compresi test ripetuti sulla stessa persona e percorso guariti

\*\* di cui 1.287 Screening Antigenici

POSITIVI AL SARS-CoV-2								
Ricoverati in Terapia intensiva	Ricoverati (NON in terapia intensiva)**	TOTALE RICOVERATI**	Ospiti in strutture territoriali	DIMESSI /GUARITI	ISOLAMENTO DOMICILIARE	Totale Positivi alla data di oggi (ricoverati + isolamenti)	DECEDUTI	TOTALI CASI POSITIVI da inizio pandemia
72	554	626	230	42.965	8.360	8.986	1.909	53.860

\*\* I pazienti in Pronto Soccorso sono stati scorparati non essendo ricoverati

MARCHE casi positivi per domicilio/residenza	
PROVINCIA	n.casi
ANCONA	15.455
PESARO URBINO	12.580
MACERATA	11.385
FERMO	6.337
ASCOLI PICENO	6.133
extra regione/al	1.970
<b>TOTALE</b>	<b>53.860</b>

RICOVERI TERAPIA INTENSIVA	
STRUTTURE RICOVERO	n.
AORMN	20
AOOR Torrette	16
AOOR Pediatria	1
JESI	10
CIVITANOVA M. - Covid Hospital	12
FERMO	5
SAN BENEDETTO DEL TR.	8
<b>TOTALE</b>	<b>72</b>

Ospiti in Strutture territoriali	
STRUTTURE RICOVERO	n.
Campofilone - RSA	44
Chiaravalle - RSA	30
Gallatara - RSA	54
Macerata Feltria (Riab)	19
Fossombrone	26
Ripatransone	21
Residenza Dorica INRCA	36
<b>TOTALE</b>	<b>230</b>

Confronto indicatori sul giorno precedente	
	n. su ieri
Casi positivi diagnosticati	+466 su ieri
Ricoveri Terapia Intensiva	-1 su ieri
Ricoveri in Semi-Intensiva	+4 su ieri
Ricoveri in NON Intensiva	+5 su ieri
Totale Ricoveri	+8 su ieri
Dimessi	+34 su ieri
Tamponi Totali	+5.452 su ieri
Decessi*	+15 su ieri

# Dato riferito al giorno precedente

RICOVERI IN AREE DI SEMI INTENSIVA	
STRUTTURE RICOVERO	n.
AORMN	50
AOOR Torrette	19
SENIGALLIA - Covid	4
JESI	11
MACERATA	8
CIVITANOVA M. - Covid Hospital	28
FERMO	12
SAN BENEDETTO DEL TR.	11
ASCOLI PICENO	5
<b>TOTALE</b>	<b>148</b>

PAZIENTI IN PRONTO SOCCORSO	
STRUTTURE OSPEDALIERE	n.
AORMN	3
AOOR Torrette	6
URBINO	1
JESI	3
FABRIANO	4
CIVITANOVA MARCHE	6
FERMO	1
<b>TOTALE</b>	<b>24</b>

RICOVERI IN REPARTI NON INTENSIVI	
STRUTTURE RICOVERO	n.
AORMN Covid	30
AORMN Ostetricia	2
AOOR Torrette - Malattie inf.	36
AOOR Torrette - Cov4	19
AOOR Torrette - Ostetricia	1
INRCA Ancona	36
INRCA FERMO	15
PESARO - SPDC	1
JESI - Covid	52
SENIGALLIA - Covid	27
MACERATA - Malattie inf.	42
CIVITANOVA M. - Covid Hospital	34
CAMERINO	15
FERMO - Malattie inf.	49
SAN BENEDETTO del T. - Covid	37
ASCOLI PICENO - Malattie Inf.	12
CASA DI CURA VILLA SERENA	18
<b>TOTALE</b>	<b>406</b>

**TOTALE RICOVERI\*\*** 626

\*\* I pazienti in Pronto Soccorso sono stati scorparati non essendo ricoverati

MARCHE casi e contatti in isolamento domiciliare				dall'inizio dell'epidemia ad oggi N.tot.
PROVINCIA	situazione attuale			di cui Operatori sanitari
	n.asintomatici	n.sintomatici	n. Tot	
PESARO URBINO	1.789	1.697	3.486	51.329
ANCONA	3.413	1.778	5.191	54.101
MACERATA	3.120	274	3.394	33.181
FERMO	1.306	322	1.628	17.251
ASCOLI PICENO	1.199	339	1.538	18.054
<b>TOTALE</b>	10.827	4.410	15.237	173.916

Figura 3.4. Il report *Giallo* pubblicato giornalmente dalla Regione Marche

### Report Arancio

Il report Arancio si concentra solamente sui deceduti della Regione Marche e comprende, anche, i cosiddetti deceduti EXTRA REGIONE, ovvero persone non residenti nelle Marche (Figura 3.5), ma deceduti nel territorio marchigiano.

Questo report riporta il totale parziale dei deceduti, il totale dei deceduti giornalieri, la suddivisione per sesso e, anche, quella per provincia di domicilio. In particolare, viene anche riportata una tabella che fornisce il dettaglio di tutte le persone decedute nelle ultime 24 ore.

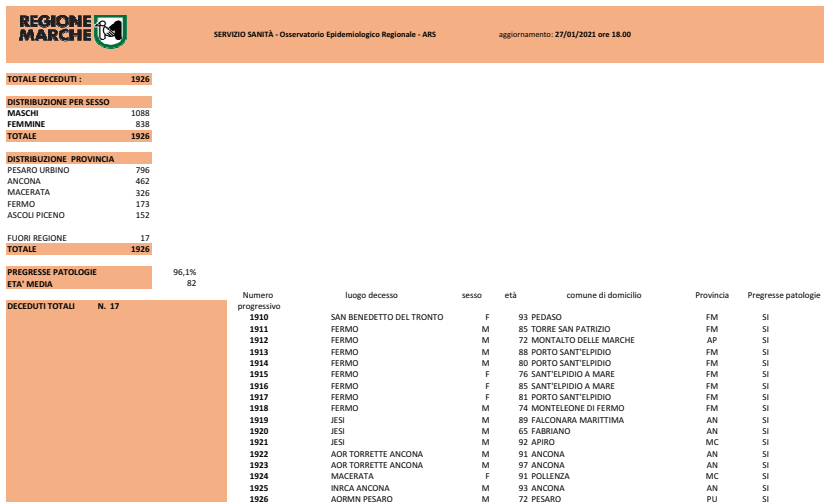


Figura 3.5. Il report Arancio pubblicato giornalmente dalla Regione Marche

### 3.1.2 Struttura dei dati

Il repository utilizzato come sorgente dei dati del caso di studio raccoglie le informazioni presenti nei tre report pubblicati quotidianamente dalla Regione Marche all'interno di un unico "file"; in questo modo, vengono forniti i PDF dal GORES in un formato utilizzabile per l'analisi di Business Intelligence.

In particolare, i tre report giornalieri vengono inizialmente aggregati in un unico file json, che riunisce le informazioni dei singoli report nella struttura riportata nel Listato 3.1.

```
{
  "data": "2021-01-27",
  "file_pdf": {
```



```

"locale": {
  "blu": "./data/PDF/GORES_Blu/27-01-2021.pdf",
  "giallo": "./data/PDF/GORES_Giallo/27-01-2021.pdf",
  "arancio": "./data/PDF/GORES_Arancio/27-01-2021.pdf"
},
"online": {
  "blu": "https://www.regione.marche.it/portals/0/Salute/CORONAVIRUS/DatiGORES/BLU.pdf",
  "giallo": "https://www.regione.marche.it/portals/0/Salute/CORONAVIRUS/DatiGORES/Gialla.pdf",
  "arancio": "https://www.regione.marche.it/portals/0/Salute/CORONAVIRUS/DatiGORES/Arancio.pdf"
}
},
"tamponi": {
  "totali": {
    "test_effettuati": 669151,
    "casi_diagnosticati": 403378,
    "casi_positivi": 53860
  },
  "odierni": {
    "test_effettuati": 5452,
    "casi_diagnosticati": 3301,
    "casi_positivi": 466
  }
},
"malati": {
  "totali": {
    "persone": 53860,
    "provincia": {
      "pesaro_urbino": 12580,
      "ancona": 15455,
      "macerata": 11385,
      "fermo": 6337,
      "ascoli_piceno": 6133,
      "extra_regione": 1970
    }
  },
  "dimessi_guariti": 42965,
  "attivi": {
    "persone": 8986,
    "isolamento_domiciliare": 8360,
    "ricoverati": {
      "totali": 626,
      "terapia_intensiva": {
        "totali": 72,
        "strutture": {
          "AO_Marche_Nord": 20,
          "AO_Ancona_Torrette": 17,
          "Senigallia": 0,
          "Fermo": 5,
          "Civitanova": 12,
          "Jesi": 10,
          "Urbino": 0,
          "Camerino": 0,
          "San_Benedetto_dT": 8,
          "Macerata": 0,
          "INRCA_Ancona": 0
        }
      },
      "semi_intensiva": {
        "totali": 148,
        "strutture": {
          "Civitanova": 28,
          "Camerino": 0,
          "San_Benedetto_dT": 16,
          "Jesi": 11,
          "Macerata": 8,
          "Urbino": 0,
          "Senigallia": 4,
          "Fermo": 12,
          "AO_Marche_Nord": 50,
          "AO_Ancona_Torrette": 19,
          "Villa_Pini": 0
        }
      }
    }
  },
  "post_critica": {
    "totali": 230,
    "strutture": {
      "Chiaravalle": 30,
      "Galantara": 73,
      "Fossombrone": 26,
      "Villa_Pini": 0,
      "RSA_Campofilone": 65,
      "AO_Ancona_Torrette": 36,
      "INRCA_Fermo": 0,
      "OspedaleDaCampo_Jesi": 0
    }
  }
},
"non_intensiva": {
  "totali": 430,
  "strutture": {
    "Fermo": 65,
    "Jesi": 73,
    "Urbino": 2,
    "Macerata": 42,
    "Senigallia": 27,

```

```

        "Fabriano": 4,
        "Camerino": 15,
        "Civitanova": 20,
        "San_Benedetto_dT": 49,
        "INRCA_Ancona": 36,
        "AO_Marche_Nord": 35,
        "AO_Ancona_Torrette": 62,
        "Villa_Pini": 0
    }
}
},
"quarantena_domiciliare": {
    "totali": 173916,
    "attivi": {
        "totali": 15237,
        "operatori_sanitari": 722,
        "sintomatici": 4410,
        "asintomatici": 10827
    },
    "provincia": {
        "pesaro_urbino": {
            "totali": 51329,
            "attivi": {
                "totali": 3486,
                "operatori_sanitari": 182,
                "sintomatici": 1697,
                "asintomatici": 1789
            }
        },
        "ancona": {
            "totali": 54101,
            "attivi": {
                "totali": 5191,
                "operatori_sanitari": 287,
                "sintomatici": 1778,
                "asintomatici": 3413
            }
        },
        "macerata": {
            "totali": 33181,
            "attivi": {
                "totali": 3394,
                "operatori_sanitari": 122,
                "sintomatici": 274,
                "asintomatici": 3120
            }
        },
        "fermo": {
            "totali": 17251,
            "attivi": {
                "totali": 1628,
                "operatori_sanitari": 82,
                "sintomatici": 322,
                "asintomatici": 1306
            }
        },
        "ascoli_piceno": {
            "totali": 18054,
            "attivi": {
                "totali": 1538,
                "operatori_sanitari": 49,
                "sintomatici": 339,
                "asintomatici": 1199
            }
        }
    }
}
},
"decessi": {
    "totali": {
        "decessi": 1926,
        "sesso": {
            "maschio": 1088,
            "femmina": 838
        }
    },
    "provincia": {
        "pesaro_urbino": 796,
        "ancona": 462,
        "macerata": 326,
        "fermo": 173,
        "ascoli_piceno": 152,
        "extra_regione": 17
    }
},
"odierni": {
    "decessi": 17,
    "dettaglio": [
        {
            "id": 1928,
            "luogo_decesso": "SAN_BENEDETTO_DEL_TRONTO",
            "sesso": "F",
            "eta": 93,
            "comune_domicilio": "PEDASO",

```

```

        "provincia_domicilio": "FM",
        "pregresse_patologie": 1
    },
    {
        "id": 1929,
        "luogo_decesso": "FERMO",
        "sesso": "M",
        "eta": 85,
        "comune_domicilio": "TORRE_SAN_PATRIZIO",
        "provincia_domicilio": "FM",
        "pregresse_patologie": 1
    },
    {
        "id": 1930,
        "luogo_decesso": "FERMO",
        "sesso": "M",
        "eta": 72,
        "comune_domicilio": "MONTALTO DELLE MARCHE",
        "provincia_domicilio": "AP",
        "pregresse_patologie": 1
    },
    {
        "id": 1931,
        "luogo_decesso": "FERMO",
        "sesso": "M",
        "eta": 88,
        "comune_domicilio": "PORTO_SANT'ELPIDIO",
        "provincia_domicilio": "FM",
        "pregresse_patologie": 1
    },
    {
        "id": 1932,
        "luogo_decesso": "FERMO",
        "sesso": "M",
        "eta": 80,
        "comune_domicilio": "PORTO_SANT'ELPIDIO",
        "provincia_domicilio": "FM",
        "pregresse_patologie": 1
    },
    {
        "id": 1933,
        "luogo_decesso": "FERMO",
        "sesso": "F",
        "eta": 76,
        "comune_domicilio": "SANT'ELPIDIO_A_MARE",
        "provincia_domicilio": "FM",
        "pregresse_patologie": 1
    },
    {
        "id": 1934,
        "luogo_decesso": "FERMO",
        "sesso": "F",
        "eta": 85,
        "comune_domicilio": "SANT'ELPIDIO_A_MARE",
        "provincia_domicilio": "FM",
        "pregresse_patologie": 1
    },
    {
        "id": 1935,
        "luogo_decesso": "FERMO",
        "sesso": "F",
        "eta": 81,
        "comune_domicilio": "PORTO_SANT'ELPIDIO",
        "provincia_domicilio": "FM",
        "pregresse_patologie": 1
    },
    {
        "id": 1936,
        "luogo_decesso": "FERMO",
        "sesso": "M",
        "eta": 74,
        "comune_domicilio": "MONTELEONE_DI_FERMO",
        "provincia_domicilio": "FM",
        "pregresse_patologie": 1
    },
    {
        "id": 1937,
        "luogo_decesso": "JESI",
        "sesso": "M",
        "eta": 89,
        "comune_domicilio": "FALCONARA_MARITTIMA",
        "provincia_domicilio": "AN",
        "pregresse_patologie": 1
    },
    {
        "id": 1938,
        "luogo_decesso": "JESI",
        "sesso": "M",
        "eta": 65,
        "comune_domicilio": "FABRIANO",
        "provincia_domicilio": "AN",
        "pregresse_patologie": 1
    },
    {

```

```

    "id": 1939,
    "luogo_decesso": "JESI",
    "sesso": "M",
    "eta": 92,
    "comune_domicilio": "APIRO",
    "provincia_domicilio": "MC",
    "pregresse_patologie": 1
  },
  {
    "id": 1940,
    "luogo_decesso": "AOR_TORRETTE_ANCONA",
    "sesso": "M",
    "eta": 91,
    "comune_domicilio": "ANCONA",
    "provincia_domicilio": "AN",
    "pregresse_patologie": 1
  },
  {
    "id": 1941,
    "luogo_decesso": "AOR_TORRETTE_ANCONA",
    "sesso": "M",
    "eta": 97,
    "comune_domicilio": "ANCONA",
    "provincia_domicilio": "AN",
    "pregresse_patologie": 1
  },
  {
    "id": 1942,
    "luogo_decesso": "MACERATA",
    "sesso": "F",
    "eta": 91,
    "comune_domicilio": "POLLENZA",
    "provincia_domicilio": "MC",
    "pregresse_patologie": 1
  },
  {
    "id": 1943,
    "luogo_decesso": "INRCA_ANCONA",
    "sesso": "M",
    "eta": 93,
    "comune_domicilio": "ANCONA",
    "provincia_domicilio": "AN",
    "pregresse_patologie": 1
  },
  {
    "id": 1944,
    "luogo_decesso": "AORMN_PESARO",
    "sesso": "M",
    "eta": 72,
    "comune_domicilio": "PESARO",
    "provincia_domicilio": "PU",
    "pregresse_patologie": 1
  }
]
}
}

```

**Listato 3.1.** Struttura del file `json` che raccoglie i tre report giornalieri della Regione Marche

### Dettaglio dei campi che compongono il file `json`

Si analizzano, dunque, in dettaglio, la struttura del file `json` che contiene i dati di interesse utilizzati per l'attività di analisi. È importante sottolineare come questo sia uno dei passaggi fondamentali della sopraccitata analisi, in quanto i campi di questo file sono, poi, gli stessi che verranno usati per creare i report e i grafici in Power BI Desktop.

Il file `json` è organizzato nel seguente modo e con le seguenti nidificazioni:

- *data*: giorno nel quale i tre report (*Blu*, *Giallo* e *Arancio*) sono stati pubblicati dalla Regione Marche. È importante sottolineare, però, che i dati si riferiscono alle 24 ore precedenti.
- *file\_pdf*
  - *locale*

- *blu*: percorso del file PDF blu (URL alla particolare pagina del sito della regione con il report di interesse);
- *giallo*: percorso del file PDF giallo;
- *arancio*: percorso del file PDF arancio.
- *online*
  - *blu*: URL del file PDF blu nel sito della regione;
  - *giallo*: URL del file PDF giallo nel sito della regione;
  - *arancio*: URL del file PDF arancio nel sito della regione.
- *tamponi*
  - *totali*
    - *test\_effettuati*: numero totale dei tamponi eseguiti, compresi anche i test ripetuti più volte sulla stessa persona, dall’inizio dell’epidemia;
    - *casi\_diagnosticati*: numero totale di persone a cui è stato eseguito un tampone dall’inizio dell’epidemia;
    - *casi\_positivi*: numero totale di persone risultate positive al tampone dall’inizio dell’epidemia.
  - *odierni*
    - *test\_effettuati*: numero dei tamponi eseguiti, compresi anche i test ripetuti più volte sulla stessa persona, nelle ultime 24 ore;
    - *casi\_diagnosticati*: numero di persone a cui è stato eseguito un tampone nelle ultime 24 ore;
    - *casi\_positivi*: numero di persone risultate positive al tampone nelle ultime 24 ore.
- *malati*
  - *totali*
    - *persone*: numero totale di persone risultate positive al tampone dall’inizio dell’epidemia.
    - *provincia*
      - *pesaro\_urbino*: numero totale di persone risultate positive al tampone, nella provincia di Pesaro Urbino, dall’inizio dell’epidemia;
      - *ancona*: numero totale di persone risultate positive al tampone, nella provincia di Ancona, dall’inizio dell’epidemia;
      - *macerata*: numero totale di persone risultate positive al tampone, nella provincia di Macerata, dall’inizio dell’epidemia;
      - *fermo*: numero totale di persone risultate positive al tampone, nella provincia di Fermo, dall’inizio dell’epidemia;
      - *ascoli\_piceno*: numero totale di persone risultate positive al tampone, nella provincia di Ascoli Piceno, dall’inizio dell’epidemia;
      - *extra\_regione*: numero totale di persone risultate positive al tampone, non residenti nella regione Marche, dall’inizio dell’epidemia.
  - *dimessi\_guariti*: numero totale di persone dimesse e guarite. In questo valore sono considerate sia le persone risultate negative al doppio tampone di controllo, sia le persone ancora positive ma che sono state dimesse dalle strutture ospedaliere.
  - *attivi*
    - *persone*: numero totale dei pazienti ricoverati o in isolamento domiciliare, attualmente positivi;

- *isolamento\_domiciliare*: numero totale di pazienti attualmente positivi in isolamento domiciliare.
- *ricoverati*
  - *totali*: numero totale di pazienti ricoverati in strutture ospedaliere;
  - *terapia\_intensiva*: *totali* (numero totale di pazienti ricoverati in terapia intensiva) e *strutture* (numero di ricoveri per ogni struttura);
  - *semi\_intensiva*: *totali* (numero totale di pazienti ricoverati in aree di terapia semintensiva.) e *strutture* (numero di ricoveri per ogni struttura);
  - *post\_critica*: *totali* (numero totale di pazienti ricoverati in degenze postcritiche) e *strutture* (numero di ricoveri per ogni struttura);
  - *non\_intensiva*: *totali* (numero totale di pazienti ricoverati in reparti non intensivi) e *strutture* (numero di ricoveri per ogni struttura).
- *quarantena\_domiciliare*
  - *totali*: numero totale, dall’inizio dell’epidemia, di pazienti che sono risultati positivi al tampone, o di persone che hanno avuto contatti con casi positivi, che sono attualmente, o sono stati, in quarantena preventiva.
  - *attivi*
    - *totali*: numero totale di persone positive al tampone, o che hanno avuto contatti con casi positivi, attualmente in isolamento domiciliare (questo valore è calcolato sommando asintomatici e sintomatici, nel cui conteggio sono già considerati gli operatori sanitari).
    - *operatori\_sanitari*: numero di operatori sanitari che sono positivi al tampone, o che hanno avuto contatti con casi positivi, attualmente in isolamento domiciliare (questo valore è già compreso nel conteggio di sintomatici e asintomatici).
    - *sintomatici*: numero di persone, con sintomi riconducibili al coronavirus, che sono positive al tampone, o che hanno avuto contatti con casi positivi, attualmente in isolamento domiciliare.
    - *asintomatici*: numero di persone, senza alcun sintomo, che sono positive al tampone, o che hanno avuto contatti con casi positivi, attualmente in isolamento domiciliare.
  - *provincia*
    - *pesaro\_urbino*: *totali* e *attivi* (*totali*, *operatori\_sanitari*, *sintomatici* e *asintomatici*);
    - *ancona*: *totali* e *attivi* (*totali*, *operatori\_sanitari*, *sintomatici* e *asintomatici*);
    - *macerata*: *totali* e *attivi* (*totali*, *operatori\_sanitari*, *sintomatici* e *asintomatici*);
    - *fermo*: *totali* e *attivi* (*totali*, *operatori\_sanitari*, *sintomatici* e *asintomatici*);
    - *ascoli\_piceno*: *totali* e *attivi* (*totali*, *operatori\_sanitari*, *sintomatici* e *asintomatici*).
- *decessi*
  - *totali*
    - *decessi*: numero totale di decessi di persone risultate positive al coronavirus, dall’inizio dell’epidemia.

- *Sesso*
  - *maschio*: numero totale maschi deceduti;
  - *femmina*: numero totale di femmine decedute.
- *provincia*
  - *pesaro\_urbino*: numero totale di decessi di persone risultate positive al coronavirus, dall’inizio dell’epidemia, nella provincia di Pesaro Urbino;
  - *ancona*: numero totale di decessi di persone risultate positive al coronavirus, dall’inizio dell’epidemia, nella provincia di Ancona;
  - *macerata*: numero totale di decessi di persone risultate positive al coronavirus, dall’inizio dell’epidemia, nella provincia di Macerata;
  - *fermo*: numero totale di decessi di persone risultate positive al coronavirus, dall’inizio dell’epidemia, nella provincia di Fermo;
  - *ascoli\_piceno*: numero totale di decessi di persone risultate positive al coronavirus, dall’inizio dell’epidemia, nella provincia di Ascoli Piceno.
- *odierni*
  - *decessi*: numero di decessi di persone risultate positive al coronavirus nelle ultime 24 ore.
  - *dettaglio*
    - *id*: numero identificativo incrementale di ogni deceduto;
    - *luogo\_decesso*: struttura ospedaliera o comune dove è avvenuto il decesso (se non disponibile, il valore sarà “ND”);
    - *sesso*: sesso del deceduto (se non disponibile, il valore sarà “ND”);
    - *eta*: età del deceduto (se non disponibile, il valore sarà “0”);
    - *comune\_domicilio*: comune di domicilio del deceduto (se non disponibile, il valore sarà “ND”);
    - *provincia\_domicilio*: provincia di domicilio del deceduto (se non disponibile, il valore “ND”);
    - *pregresse\_patologie*: se il valore è 1 il defunto era affetto da patologie pregresse, se il valore è 0 il defunto non era affetto da alcuna patologia pregressa (se non disponibile, il valore sarà “-1”).

## 3.2 Costruzione del dataset

Il repository di GitHub utilizzato come sorgente dati e, in particolare, il file CSV contenente i report strutturati secondo i campi di elementi precedentemente illustrati, rendono disponibili i dati relativi alla situazione del COVID-19 nelle Marche in una forma pulita. Questo significa che i dati della sorgente vengono già sottoposti a operazioni di ETL. Infatti, essendo pubblico il repository di interesse, e avendo come unico obiettivo quello di rendere i dati disponibili all’utente in un formato idoneo per le analisi, esso prevede una struttura dei dati stessi predefinita e ben precisa.

In questa sezione si pone l’attenzione sulle operazioni svolte per la creazione del dataset riguardante i dati della seconda ondata di contagi da COVID-19 nella Regione Marche. Una volta realizzato, il dataset è pronto per essere importato su Power BI Desktop ed effettuare l’attività di analisi.

Nelle sottosezioni seguenti verranno descritte singolarmente le varie operazioni svolte per la realizzazione del dataset, e come esse sono state applicate al caso di studio.

### 3.2.1 Clonazione del repository remoto

La prima fase del processo di creazione del dataset consiste nel prelievo dei dati giornalieri dal sistema sorgente e nella loro aggregazione con i dati già esistenti. In altre parole, è stato inizialmente clonato il repository di GitHub menzionato nella sezione precedente, e il file ottenuto da tale operazione è stato salvato all'interno della macchina locale in una directory, creando una copia locale di tutti i file Covid-19 (Figura 3.6). All'interno della cartella clonata, ovviamente, sono presenti i file `covid-19-marche.csv` e `dettagli_decessi.csv`, che, successivamente, verranno importati in Power BI Desktop.

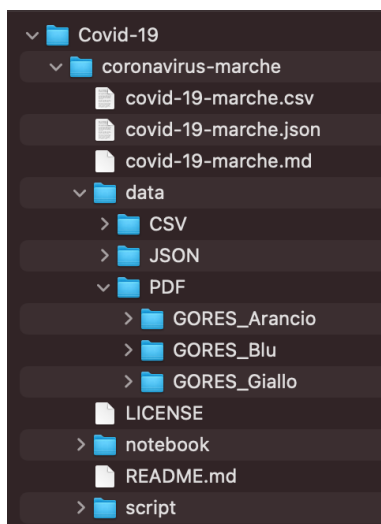


Figura 3.6. La directory Covid-19 con i file clonati dal repository di GitHub

### 3.2.2 Realizzazione dei file JSON

Una volta clonato il repository remoto e una volta salvato quest'ultimo in una directory della macchina locale, si può procedere con la modifica della struttura dati stessa, dalla creazione di nuovi file alla variazione dei file già esistenti.

Nel caso di studio della presente tesi, il repository è stato aggiornato con i dati giornalieri relativi alla seconda ondata di contagi da COVID-19, in quanto nella struttura dati, erano presenti soltanto le informazioni riguardanti la prima ondata.

Per realizzare i file JSON giornalieri, innanzitutto, occorre duplicare il file JSON dell'ultimo giorno inserito posizionato nella sottocartella `JSON`, ed aprire la copia con un editor di testo.



Una volta mostrata la finestra del file duplicato, si procede modificando soltanto i valori dei campi del rispettivo giorno, lasciando inalterata la struttura del formato di testo. I valori da inserire, inoltre, vengono copiati dalle tabelle dei tre report pubblicati quotidianamente, come specificato nella sezione precedente.

Terminata la fase di modifica, il file ottenuto viene salvato con il nome della data corrispondente, in modo tale che venga aggiunto nell'elenco seguendo l'ordine temporale.

In altre parole, i tre report pubblicati quotidianamente dal GORES vengono inizialmente aggregati in un unico file `json` che riunisce le informazioni dei singoli report nella struttura riportata nel Listato 3.1

### 3.2.3 Script Python per l'aggregazione dei file giornalieri

I file `json` dei singoli giorni così creati vengono, successivamente, raccolti in un unico file, sempre di tipo `json`, che include i precedenti come elementi di un sorta di "array di report"; quest'unico file `json`, cioè, non contiene altro che un array, dove ciascuno elemento è il report di un determinato giorno. In questo modo, per ogni giorno, si ha un unico report (che racchiude i tre report giornalieri), e i report dei singoli giorni formano l'array che rappresenta l'unica struttura del file `json` finale.

Oltre alla creazione del file `json` finale, inoltre, viene generato un unico file `CSV` che, al suo interno, contiene i dati relativi ai tre report dei singoli giorni separati da virgole, organizzati in modo da simulare una tabella. La struttura di un file `CSV`, però, è abbastanza complessa per essere valutata da un comune editor di testo. Infatti, aprendo un file con formato `CSV`, i dati all'interno rischiano di essere privi di senso e poco comprensibili, soprattutto se sono particolarmente complessi; invece, le informazioni acquisiscono un senso se mostrate con un programma in grado di interpretarle tramite fogli di calcolo e tabelle.

Quest'ultimo file `CSV` è proprio quello che è stato importato in Power BI Desktop per eseguire lo studio della presente tesi.

Per la realizzazione dei file `json` e `CSV` unici sono stati utilizzati Script Python, ovvero codici sorgenti realizzati con un semplice blocco note e con un'estensione ".py". In particolare, nel caso dell'attività di analisi di interesse sono stati usati gli script mostrati in figura 3.7.

Lo script `check_error.py` effettua un test statico che controlla tutti i file `JSON` dei singoli giorni verificando se ci sono incongruenze.

Lo script `generate_json.py` permette di generare il file `JSON` unico a partire dai file dei singoli giorni.

Lo script `generate_csv.py`, invece, genera il file `CSV` unico contenente le informazioni sui positivi, ospedalizzati e contagiati in quarantena domiciliare, e il file `CSV`, più specifico, riguardante i decessi.

Per eseguire uno script basta aprire il prompt dei comandi, inserire il percorso del file ed, infine, avviare lo script specificandone il nome.

### 3.2.4 Upload dei file creati sul repository

Dopo aver effettuato le modifiche necessarie, aggiungendo i dati relativi alla seconda ondata di contagi da COVID-19, è possibile sincronizzare il repository locale con

```
$ python3 check_error.py  
$ python3 generate_json.py  
$ python3 generate_csv.py
```

**Figura 3.7.** Gli Script Python utilizzati per l'aggregazione dei file giornalieri

quello remoto, in modo che le variazioni eseguite in locale risultino anche nel file remoto.

Per avviare la sincronizzazione tra i due repository è necessario aprire il prompt dei comandi e recarsi nella folder principale all'interno del quale è presente il repository locale. In seguito è necessario eseguire i comandi mostrati nella Figura 3.8

```
git add .  
git commi -m "add ultimo giorno inserito"  
git push origin master
```

**Figura 3.8.** Istruzioni da riga di comando per la sincronizzazione tra repository locale e repository remoto

L'istruzione `git add .` seleziona le modifiche effettuate in locale rispetto al repository remoto.

L'istruzione `git commit "add ultimo giorno inserito"` carica il progetto nell'HEAD di git, ma non ancora nel repository remoto. Tra le virgolette, inoltre, si digita l'ultimo giorno inserito nell'elenco anche se, in realtà, si tratta soltanto di un messaggio per il commit.

## Progettazione e Implementazione dei Task di Data Science

*In questo capitolo si entra nel dettaglio delle attività svolte in Power BI Desktop. In particolare, nella prima parte, si illustreranno le operazioni eseguite per l'importazione della sorgente dati e la creazione del modello di dati. Nella seconda parte, invece, si descriveranno e analizzeranno le informazioni relative all'andamento del COVID-19 durante la seconda ondata, con lo scopo di realizzare uno schema di propagazione del contagio. Infine, verranno messi a confronto gli andamenti della prima e della seconda ondata di contagi, al fine di tracciare analogie e differenze. A tale scopo si faranno uso delle differenti tipologie di visualizzazioni che Power BI mette a disposizione.*

### 4.1 Attività di ETL svolte in Power BI

Il repository di GitHub utilizzato come sorgente dati e, in particolare, il file CSV contenente le informazioni relative all'andamento del COVID-19, rendono disponibili i numeri in una struttura pulita. Questo significa che i dati della sorgente vengono sottoposti preliminarmente a operazioni di ETL. Infatti, trattandosi di un repository pubblico, e avendo come unico obiettivo quello di rendere disponibili i dati in un formato semplice per le analisi, esso prevede una struttura dati predefinita e ben precisa.

Ad ogni modo, Power BI permette di connettersi a diverse tipologie di sorgenti dati, tra cui anche i file CSV, file di testo semplici che contengono elenchi di dati. Questo formato, per la semplicità della struttura interna, consente di esportare facilmente i dati e importarli in altri programmi, per poterli visualizzare e riutilizzare in base all'utilità o alla necessità.

La scelta di utilizzare file CSV, piuttosto che file JSON, nell'importazione dei dati in Power BI risiede proprio nell'eterogeneità della struttura. Mentre un file CSV contiene, al proprio interno, una serie di dati organizzati in modo da strutturare una tabella, un file JSON, a causa della sua struttura complicata, necessita di alcune operazioni di manipolazione.

### 4.1.1 Importazione dei dati

Terminata la fase di costruzione del dataset precedentemente illustrata, e verificata l'omogeneità dei dati ivi rappresentati, il passaggio successivo è l'importazione dei dati all'interno di Power BI Desktop.

Per connettersi ai dati è possibile usare la barra multifunzione "Home" e, dunque, la voce "Get Data". A seguito di ciò, viene visualizzata la finestra "Recupera dati" la quale permette di scegliere, tra le tante tipologie di sorgenti dati disponibili, anche il tipo CSV.

Una volta aver cliccato sulla sorgente dati di interesse, viene mostrata una preview della cartella di lavoro all'interno della quale Power BI Desktop cerca di interpretarne il contenuto (Figura 4.1). Questa sorta di anteprima dei dati è utile per capire se il software è riuscito a comprendere appieno la forma dei dati, valutando le intestazioni, la mancanza di alcune righe oppure l'errata interpretazione di un numero o di una stringa. In caso di corretta decodifica, il foglio di lavoro può essere caricato direttamente nel modello dei dati ed utilizzabile per la creazione dei grafici; in caso contrario sono necessarie la modifica e la pulizia dei dati.

data	tempi_totai_test_effettuati	tempi_totai_casi_diagnostici	tempi_totai_casi_positivi	tempi_ademi_test_effettuati
25/02/2020	0	24	0	0
26/02/2020	0	29	1	0
27/02/2020	0	27	3	0
28/02/2020	0	46	6	0
29/02/2020	0	65	11	0
01/03/2020	0	98	25	0
02/03/2020	0	134	32	0
03/03/2020	0	198	41	0
04/03/2020	0	287	84	0
05/03/2020	0	413	124	0
06/03/2020	0	586	159	0
07/03/2020	0	816	207	0
08/03/2020	0	1034	270	0
09/03/2020	0	1291	327	0
10/03/2020	0	1417	394	0
11/03/2020	0	1656	479	0
12/03/2020	0	1907	592	0
13/03/2020	0	2216	724	0
14/03/2020	0	2511	889	0
15/03/2020	0	2917	1134	0

Figura 4.1. Pre-view della cartella di lavoro selezionata per l'importazione dei dati

### 4.1.2 Modifica dei dati

Come si evince dalla Figura 4.1, che mostra la pre-view del dataset importato, Power BI Desktop sembra aver interpretato correttamente il contenuto, dalle intestazioni ai tipi di dato delle colonne.

L'unica modifica apportata è stata la sostituzione dei nomi delle feature delle tabelle con dei nuovi nomi più intuitivi e di più facile lettura, al fine di facilitarne la fruizione.

### 4.1.3 Creazioni di nuove query

Se, da una parte, la modifica dei dati ha reso sicuramente il dataset più pulito rispetto a quello inizialmente importato, dall'altra le operazioni possibili nell'Editor di Power Query non si limitano a ciò.

A questo punto, infatti, si è resa necessaria un'ulteriore semplificazione nella struttura della tabella importata tramite il file CSV, creando nuove tabelle contenenti soltanto i dati relativi ad un particolare aspetto di interesse. Nel seguito verranno illustrate nel dettaglio queste operazioni.

#### Tabella relativa ai decessi giornalieri

Le informazioni relative ai decessi avvenuti nella Regione Marche a causa del COVID-19 sono già state isolate e separate in un'apposita tabella, diversa dalla principale, in fase di costruzione del dataset.

Tale scelta ha consentito, infatti, di suddividere in due relazioni distinte le informazioni messe a disposizione per il caso di studio della presente tesi. Nello specifico, le relazioni individuate sono le seguenti:

- **covid-19-marche**: in questa tabella sono riportati i dati generali sull'andamento dell'epidemia nella Regione Marche. Sono presenti, dunque, tutte quelle informazioni riguardanti tamponi, contagi, persone sottoposte alla quarantena domiciliare, ricoveri e decessi, illustrati giorno per giorno.
- **dettagli-decessi**: in questa seconda tabella, invece, sono riportate esclusivamente le informazioni relative ai decessi nella Regione Marche. È importante sottolineare come il campo "Data" non rappresenti un valore di tipo "unique", in quanto ci possono essere più tuple riportanti lo stesso valore.

Nella seconda tabella, inoltre, sono state applicate operazioni di modifica fondamentali per l'attività di analisi. In particolare, sono stati sostituiti i valori dei campi relativi alla provincia e al comune di domicilio delle persone decedute. La visualizzazione basata su mappe in Power BI, infatti, necessita delle notazioni ufficiali delle province e dei comuni per poterli geolocalizzare correttamente. Alla luce di ciò, i nomi delle province e dei comuni sono stati sostituiti direttamente nell'Editor di Power Query con i nomi ufficiali, riconosciuti dall'applicazione stessa (Figura 4.2).

#### Tabella relativa alle province

La tabella generale **covid-19-marche** contiene, tra le altre cose, alcune informazioni che sono specifiche per ogni singola provincia della Regione Marche. Ad esempio, è possibile trovare i dati riguardanti la quarantena domiciliare organizzati in tante colonne quante sono le province: una colonna per i contagiati in quarantena domiciliare della provincia di Macerata, una per i contagiati in quarantena domiciliare della provincia di Ancona, etc.

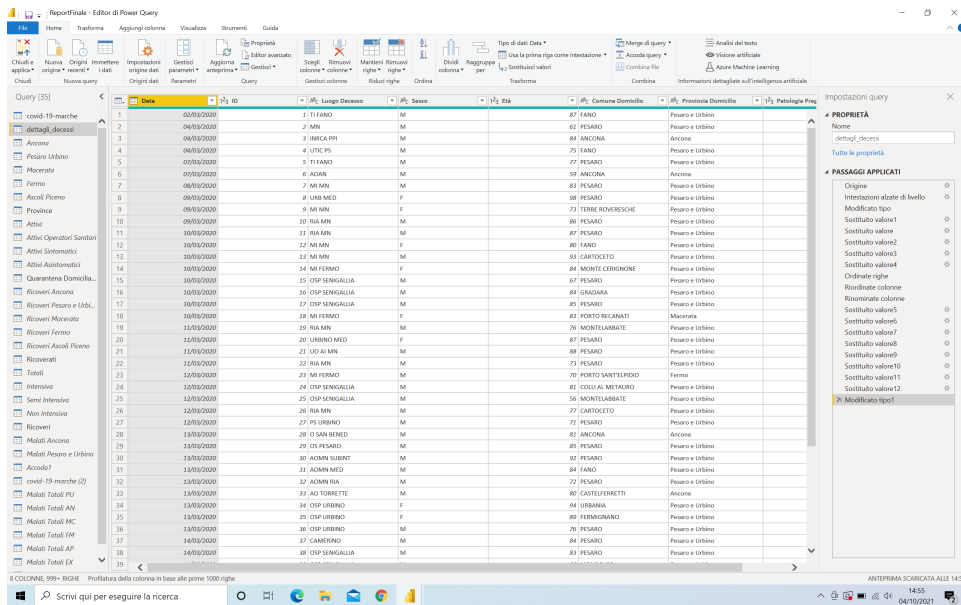


Figura 4.2. Tabella relativa ai dettagli dei decessi in Power Query

Per rendere più chiara ulteriormente tale tabella, ma anche per facilitare le attività di analisi che fanno riferimento ai dati delle singole province, si è deciso di creare una nuova query per ciascuna provincia, in modo da interessare esclusivamente i dati ad essa relativi.

Successivamente, le varie query create sono state accodate in una nuova query che definisce la relazione “Province”. Essa, dunque, riunisce le relazioni sulle singole province in un’unica tabella, di fondamentale importanza per realizzare analisi e confronti sui dati provincia per provincia (Figura 4.3).

Per realizzare tale operazione è stata utilizzata la funzione “Accoda query” presente nella barra “Home” dell’Editor di Power Query.

### Tabella relativa ai ricoveri

Ai fini dell’attività di analisi svolta e dei dashboard realizzati, si è deciso di creare una nuova query contenente soltanto le informazioni relative ai malati ricoverati all’interno delle strutture ospedaliere nei vari livelli di gravità di terapia per ognuna delle province della Regione Marche.

Analogamente a quanto fatto nelle precedenti situazioni, le varie query così create sono state accodate in una nuova query che definisce la relazione “Ricoveri”. Tale relazione, dunque, accorpa le relazioni sulle singole province e i livelli di gravità di terapia in un’unica tabella.

In particolare, come si vedrà nel seguito in dettaglio, sono stati esaminati e messi a confronto gli andamenti nel tempo dei ricoverati nei vari livelli di gravità per ciascuna provincia marchigiana, al fine di estrarre particolari informazioni successivamente illustrate.

The screenshot shows the Power BI Editor interface with a Power Query table. The table has 39 rows and 6 columns. The columns are: 'Data', 'Molai Totali', 'Quarantena Domiciliare Totali', 'Quarantena Domiciliare Attivi', 'Quarantena Domiciliare Attivi Operanti Sanitari', and 'Quarantena Domiciliare A...'. The rows list various provinces and municipalities in the Marche region, such as Ancona, Pesaro Libbio, Macerata, Fermo, Anelli Piceno, and others. The data values are mostly zeros, indicating no recorded activity for most locations.

Figura 4.3. Tabella relativa alle diverse province della regione Marche

Allo stesso modo, e per le stesse ragioni, è stata creata la relazione “Quarantena Domiciliare”, relativa al numero delle persone sottoposte, appunto, alla quarantena domiciliare, particolarizzato alle singole province della regione.

#### 4.1.4 Accodamento di query

Si analizza, brevemente, la funzionalità di accodamento messa a disposizione dall’Editor di Power Query.

L’operazione di accodamento crea una nuova query che contiene tutte le righe di una prima query seguite da tutte le righe di una seconda query; l’operazione richiede, infatti, almeno due query.

In particolare, si possono eseguire due tipi di accodamento:

- “*Accodamento intermedio*”: permette di creare una nuova query per ogni operazione di “accodamento”
- “*Accodamento inline*”: consiste nell’aggiunta dei dati alla query esistente fino a raggiungere un risultato finale. Quest’ultimo è un nuovo passaggio alla fine della query corrente.

Nel caso di studio è stata utilizzata la funzionalità di “Accodamento intermedio”. Per farlo è necessario recarsi nella barra multifunzione dell’Editor di Power Query e cliccare sulla voce “Accoda query”. Questa operazione determina l’apertura della finestra di dialogo “Accoda”, dalla quale è possibile selezionare la prima e la seconda query da accodare.

### 4.1.5 Caricamento dei dati puliti

Terminate le operazioni nell'Editor di Power Query, non rimane altro che salvare e applicare le modifiche apportate alle tabelle sul modello di dati in Power BI Desktop. A tal proposito, è sufficiente cliccare sulla funzione “Chiudi e applica”, posizionata nella barra multifunzione dell'Editor di Power Query, e l'operazione di modifica dei dati è conclusa.

I dati delle tabelle caricate assumono, così, la forma più adatta per ottenere tutti i grafici desiderati per la fase di studio.

## 4.2 Dashboard panoramico sull'andamento dell'epidemia

Prima di iniziare ad analizzare nel dettaglio lo sviluppo della pandemia da COVID-19 all'interno della Regione Marche, è stato creato un dashboard panoramico che riassume le principali statistiche del Coronavirus e le informazioni presenti nel modello di dati di Power BI Desktop relative alle province della Regione. In questo modo si ha una visione globale, seppur semplificata, del quadro epidemico generale.

È importante chiarire, però, la distinzione fra i due dati *Casi Positivi* e *Malati Attivi* che, ad un primo impatto, potrebbero sembrare identici e, come tali, essere confusi.

Il GORES, impegnato nella divulgazione dei dati utilizzati nel caso di studio, si esprime in maniera abbastanza chiara riguardo questi due concetti. Mentre il termine *Casi Positivi* si riferisce al numero totale di persone risultate positive al tampone a partire dall'inizio della pandemia, il dato relativo ai *Malati Attivi*, invece, rappresenta il numero dei contagiati attualmente positivi, ricoverati nelle strutture ospedaliere o in isolamento domiciliare.

Dopo tale premessa, è possibile illustrare il dashboard panoramico precedentemente introdotto, mostrato nella Figura 4.4.

Un aspetto fondamentale da sottolineare riguarda i valori numerici disposti nella fascia superiore del cruscotto. Da una parte, le schede riportano cifre che si riferiscono ad un valore cumulativo (“Positivi Totali”, “Malati Attivi”, “Guariti/Dimessi”, “Decessi Totali”); dall'altra, la tabella illustra le informazioni relative alle variazioni giornaliere (“Variazione Positivi”, “Variazione Malati”, “Guariti Odierni”, “Decessi Odierni”). Al fine di garantire una lettura più semplice e intuitiva, accanto al valore assoluto di ciascuna variazione viene visualizzata anche un'icona a forma di freccia di colore e verso differenti a seconda del valore di quel campo, sulla base della formattazione condizionale impostata.

Come esempio, nella Figura 4.5 viene mostrata la formattazione condizionale del valore del campo “Variazione Malati”.

È importante ricordare, però, che i dati forniti dal GORES non riportano esplicitamente le informazioni relative alle variazioni giornaliere, che, invece, risultano particolarmente significative nell'attività di analisi del caso di studio della presente tesi. Per tale motivo, è stata realizzata in Power BI Desktop una misura DAX che permette di ottenere una colonna calcolata direttamente nel modello dei dati. Nella



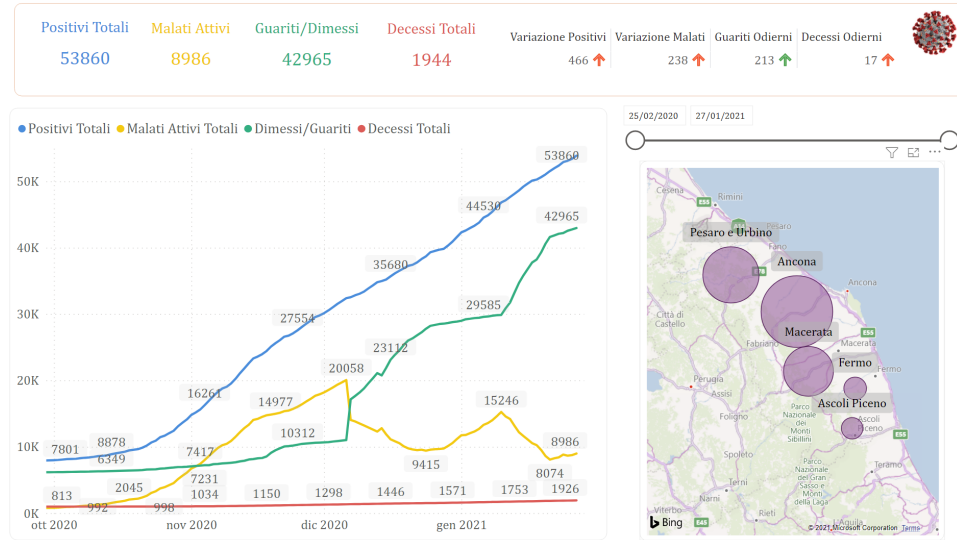


Figura 4.4. Dashboard panoramico sulla situazione COVID-19

Icone - *Variazione Malati*

Formatta per: Regole

Applica a: Solo valori

In base al campo: Somma di Variazione Malati

Esecuzione del riepilogo: Somma

Layout icona: A destra dei dati

Allineamento dell'icona: In mezzo

Stile: Personalizzato

Regole

- Se il valore è maggiore o uguale a 0 e è minore di 10000 allora ↑
- Se il valore è 0 allora →
- Se il valore è maggiore o uguale a -10000 e è minore o uguale a 0 allora ↓

Figura 4.5. Formattazione condizionale del valore del campo “Variazione Malati”

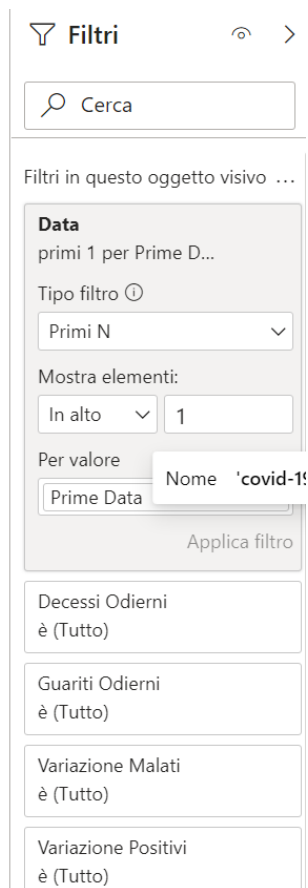
Figura 4.6 è riportata, a titolo di esempio, la misura DAX per ottenere la colonna calcolata relativa alla variazione dei malati.

```
1 Variazione Malati = (var selezionaData = CALCULATE(LASTDATE('covid-19-marche'[Data]), FILTER('covid-19-marche', 'covid-19-marche'[Data] < EARLIER('covid-19-marche'[Data]
))) return 'covid-19-marche'[Malati Attivi Totali] - CALCULATE(SUM('covid-19-marche'[Malati Attivi Totali]), FILTER('covid-19-marche', 'covid-19-marche'[Data] =
selezionaData)))
```

Figura 4.6. Formula DAX della colonna calcolata relativa alla “Variazione Malati”

Un oggetto su cui porre l’attenzione, e che tornerà utile anche nei dashboard successivi, è l’utilizzo di un filtro per visualizzare i dati relativi ad un determinato

giorno. Per realizzare tale scheda, è stata utilizzata la misura calcolata “Variazione Malati Attivi”, e all’oggetto visivo generato è stato applicato un filtro per quel che concerne il campo “Data” proveniente dalla tabella (Figura 4.7). Più in dettaglio, si tratta di un filtro avanzato che seleziona soltanto il valore più recente del campo “Data”, permettendo, così, di visualizzare esclusivamente i malati attivi relativi all’ultimo giorno disponibile, in questo caso il 27 Gennaio 2021.



**Figura 4.7.** Il riquadro Filtro con i filtri applicati all’oggetto visivo tabella

Il cruscotto, inoltre, presenta un grafico abbastanza interessante che permette di avere un’ampia panoramica degli andamenti complessivi dei “Positivi Totali”, “Malati Attivi”, “Dimessi/Guariti”, e “Decessi Totali”. Si tratta del grafico a linee mostrato nella parte sinistra del cruscotto il quale rappresenta l’evoluzione, in termini numerici, delle tematiche sopra elencate, dal giorno di inizio della seconda ondata di contagi da COVID-19 fino all’ultimo giorno di raccolta dei dati di interesse per il caso di studio. Da tale grafico si evince, in particolare, come, dopo una crescita per lo più costante riscontrata agli inizi del mese di Ottobre 2020, la curva dei “Positivi Totali” abbia raggiunto un andamento pressochè esponenziale,

in concomitanza del diffondersi della seconda ondata.

Ad ogni modo, in generale, il grafico rivela uno sviluppo che può considerarsi positivo; infatti, il numero dei dimessi/guariti, dopo una fase di lenta crescita, subisce un forte rialzo verso la metà del mese di Dicembre 2020, arrivando a oltrepassare il numero dei malati attivi (questi ultimi costituiti sia dai ricoverati che dai contagiati in isolamento domiciliare).

Per quanto riguarda il trend dei dimessi/guariti, inoltre, un aspetto che sicuramente necessita di chiarimento riguarda il passaggio dal dato di 11.008 relativo al 6 Dicembre 2020 a quello di 17.153 del 7 Dicembre 2020, con una differenza di 6.010, mai stata così alta dall'inizio della pandemia. Tale peculiarità è legata alla Circolare del Ministero della Salute del 12 Ottobre 2020, la quale aggiorna le indicazioni riguardo la durata e il termine dell'isolamento e della quarantena sulla base dell'evoluzione della situazione epidemiologica e del parere formulato dal Comitato Tecnico Scientifico. La Circolare definisce le seguenti possibilità:

- *Casi positivi asintomatici*: le persone asintomatiche risultate positive al COVID-19 possono rientrare in comunità dopo un periodo di isolamento di almeno 10 giorni dalla comparsa della positività, al termine del quale risulti eseguito un test molecolare con risultato negativo.
- *Casi positivi sintomatici*: le persone sintomatiche risultate positive al COVID-19 possono rientrare in comunità dopo un periodo di isolamento di almeno 10 giorni dalla comparsa dei sintomi accompagnato da un test molecolare con riscontro negativo eseguito dopo almeno 3 giorni senza sintomi.
- *Casi positivi a lungo termine*: le persone che, pur non presentando più sintomi, continuano a risultare positive al test molecolare, in caso di assenza di sintomatologia da almeno una settimana, potranno interrompere l'isolamento dopo 21 giorni dalla comparsa dei sintomi.

Le medesime informazioni mostrate nell'ultimo grafico descritto sono riportate, a livello provinciale, nella visualizzazione a mappe. Per realizzare tale oggetto visivo è stata utilizzata la relazione "Province", la cui struttura è stata analizzata nella sezione precedente.

Dalla mappa emerge chiaramente come la provincia di Ancona sia stata la più colpita per quanto riguarda i casi positivi nella seconda ondata del virus, seguita, poi, dalle province di Pesaro Urbino e Macerata.

### 4.2.1 Andamento giornaliero

Analizzare i dati sul COVID-19 non è sempre così semplice: ci sono, infatti, aspetti sensibilmente influenzati dalle diverse politiche regionali, e talvolta anche provinciali, principalmente in tema di tamponi.

A tale proposito, è necessario fare chiarezza sul perchè, per poter parlare di un rallentamento del virus, servono più giorni di decrescita del contagio. La motivazione sta nel fatto che si "sterilizzano" eventuali oscillazioni straordinarie dovute, magari, a esiti diagnostici che arrivano in massa in una sola giornata.

Per rendere più comprensibile il concetto è necessario l'utilizzo dei dati numerici.

Facendo riferimento al grafico della Figura 4.4, è possibile notare che il totale delle persone colpite dal COVID-19 dall'inizio della pandemia nelle Marche, aggiornato al 27 Gennaio 2021, è di 53.860, con una variazione positiva di 466 rispetto al giorno precedente. Ciò vuol dire che l'incremento c'è stato, anche se dell'1%.

Quel che però bisogna evitare, è lasciarsi ingannare dalla curva dei “Malati Attivi”, cioè quelle 8.986 persone con l'infezione attiva al 27 Gennaio 2021. Questo dato, infatti, può essere significativo per valutare la pressione del coronavirus sul sistema sanitario nazionale, ma non è indicativo, di per sé, per l'andamento generale dei contagi. Più utili, invece, sarebbero delle curve riguardanti i “nuovi casi”, gli “ospedalizzati” o le “terapie intensive”, che rifletterebbero meglio l'andamento dell'epidemia.

Basando l'analisi soltanto sulla curva dei “Malati Attivi”, si può affermare che, dopo un picco avvenuto il 10 Gennaio 2021 con 15.246 persone con l'infezione attiva, la curva ha iniziato a decrescere drasticamente raggiungendo, in meno di 10 giorni, la metà dei contagiati attivi. Dopodiché la curva ha iniziato a risalire, seppur in maniera lieve, con valori pressochè costanti.

Dunque, è vero che l'incremento appare meno consistente nelle ultime due settimane (in riferimento al 27 Gennaio 2021 come ultimo giorno del dataset utilizzato nel corso di studio) ma il dato è influenzato dall'andamento dei decessi e dei guariti. I “Nuovi Malati Attivi” possono essere calcolati dai nuovi contagi totali, cui vengono sottratti i deceduti e i guariti. Più morti si rilevano o più persone guariscono, inoltre, minore è il numero degli “Attualmente Positivi”. Ragionando per assurdo, se il sistema sanitario franasse e i medici non fossero più in grado di intubare i nuovi pazienti che arrivano in ospedale, la curva degli “Attualmente Positivi” crollerebbe. Ma non ci sarebbe nulla di cui andare fieri.

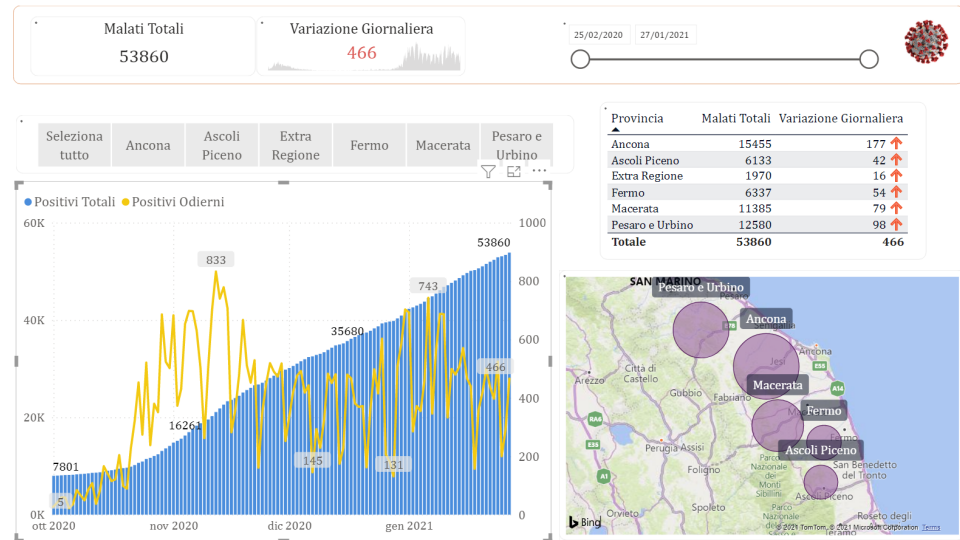
Per capire come procede l'epidemia nelle Marche bisogna dunque affidarsi all'andamento dei “Nuovi Casi”, cioè quante persone in più sono risultate infette rispetto al giorno precedente.

Date tali premesse, si passa ad analizzare il dashboard riportato nella Figura 4.8, relativo alle variazioni giornaliere dei casi positivi, il quale è stato suddiviso per ciascuna provincia delle Marche.

Il cruscotto in esame presenta, nella parte superiore, i dati numerici sul totale delle persone che hanno contratto l'infezione dall'inizio dell'epidemia e sulla variazione giornaliera di tale numero. In particolare, l'ultima informazione è riportata sotto l'oggetto visivo relativo all'indicatore di prestazione chiave (Key Performance Indicator - KPI), il quale comunica lo stato di avanzamento del dato considerato.

Al di sotto del KPI viene raffigurato l'andamento dei “nuovi casi” della Regione Marche. Quello che si può notare da tale visualizzazione è l'appiattimento della curva nei mesi caldi e l'aumento radicale dei contagi con l'avvento della stagione autunnale. Ciò conferma, ancora una volta, l'effetto di mitigazione estivo e la nota stagionalità della risposta immunitaria la quale, con la bella stagione, è più efficace e meno infiammatoria. Inoltre, si può osservare l'eccessivo contrasto tra la prima e la seconda ondata nel numero dei contagi, il che lascia presagire di quanto la seconda ondata sia stata peggiore della prima. Ad ogni modo, questo aspetto sarà trattato in una sezione dedicata.

Particolarmente interessante, in questo cruscotto, è il grafico misto riportato nella sezione inferiore sinistra, il quale è specializzato per ogni provincia delle Mar-



**Figura 4.8.** Il dashboard relativo alle variazioni giornaliere dei casi positivi

che. In tale oggetto visivo l'istogramma illustra il numero dei casi positivi totali, mentre la curva sovrapposta rappresenta il dato giornaliero dei casi risultati positivi al virus.

Osservando la curva relativa all'intera regione, si rileva dapprima un'accelerazione della pandemia dalla seconda metà del mese di Ottobre 2020 e per tutto il mese di Novembre, per poi registrare una tendenza con impatto minore nel mese di Dicembre. Questa discesa dimostra come le misure adottate dal Comitato Tecnico Scientifico e i DPCM (25 Ottobre 2020 e 4 Novembre 2020) riguardanti la chiusura delle scuole e la suddivisione dell'Italia in aree colorate, siano risultate efficaci per il rallentamento e la mitigazione del virus.

Il dashboard, inoltre, risulta fortemente interattivo infatti se si seleziona una provincia nella parte immediatamente superiore al grafico, quest'ultimo filtra i dati mostrandone, dunque, l'andamento particolarizzato per essa. In aggiunta, anche gli altri grafici presenti nel dashboard riporteranno il dato per la sola provincia selezionata. Nella Figura 4.9 è mostrato un esempio, ottenuto dopo aver selezionato la provincia di Ancona.

### 4.3 Dashboard introduttivo su tamponi e positivi

In questa sezione si analizza un dashboard che riassume le informazioni presenti nel modello di Power BI Desktop relative al dato dei tamponi e dei casi positivi rilevati all'interno della Regione Marche. Tale dashboard è riportato nella Figura 4.10

Nella fascia superiore del cruscotto sono presenti due righe dell'oggetto "Schede" fornito da Power BI. La prima riporta informazioni che fanno riferimento a un valore cumulativo; la seconda, invece, riporta informazioni riguardo le variazioni giornaliere.

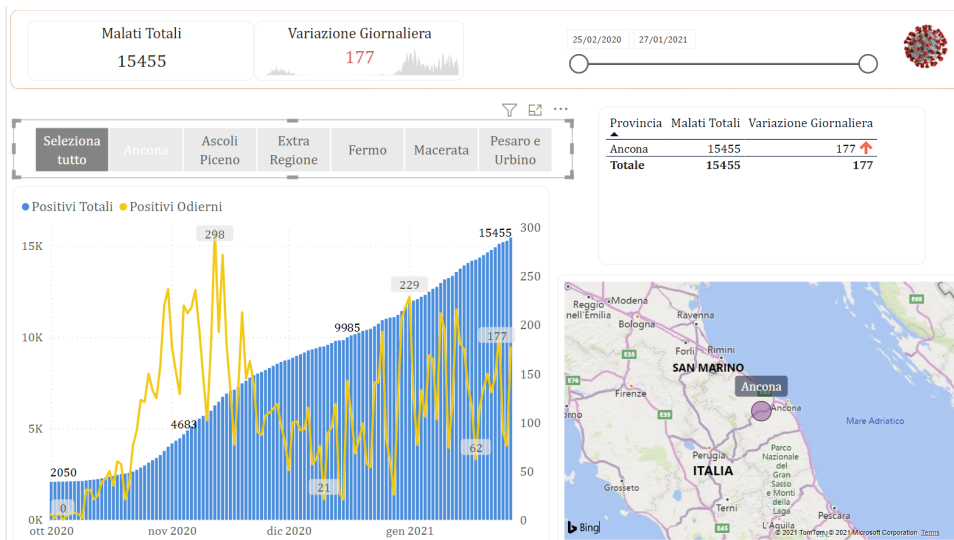


Figura 4.9. Esempio dashboard con filtro dei dati in base alla provincia di Ancona

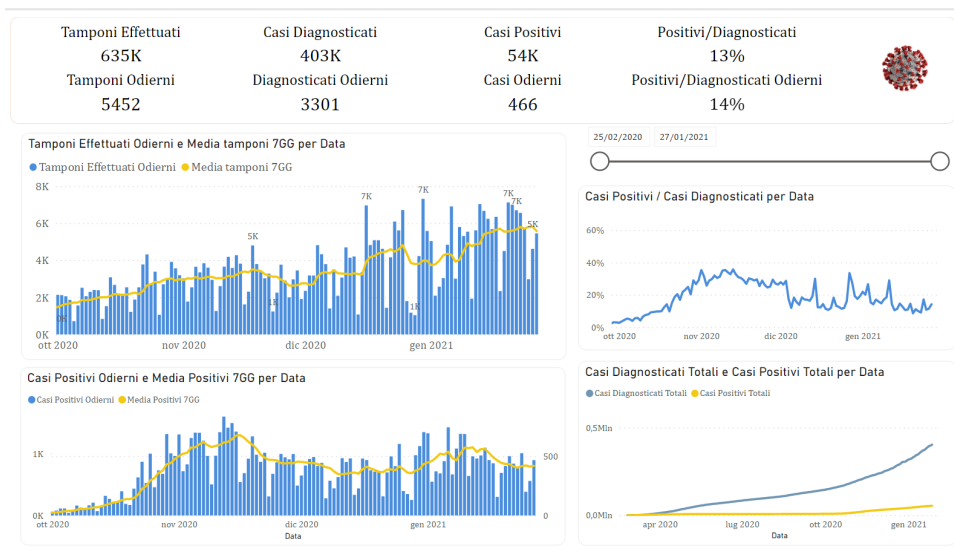


Figura 4.10. Il dashboard relativo al dato dei tamponi e dei casi positivi

Nelle prime due schede sulla destra è illustrato un dato particolarmente interessante, ovvero il rapporto percentuale tra casi diagnosticati e casi positivi. Tale relazione è stata ricavata dalle informazioni già presenti nel modello di dati, applicando semplicemente una funzione DAX di divisione, ed esprimendo il risultato in percentuale (Figura 4.11).

Tale misura rappresenta un dato rilevante, il quale permette di comprendere effettivamente ciò che accade nella dinamica epidemica e valutare il fattore di crescita

```

Positivi Odierni/Diagnosticati Odierni =
DIVIDE(
    SUM('covid-19-marche'[Casi Positivi Odierni]),
    SUM('covid-19-marche'[Casi Diagnosticati Odierni])
)

```

**Figura 4.11.** Formula DAX per il calcolo del rapporto tra casi positivi e casi diagnosticati

della pandemia.

A questo proposito, è di ulteriore informazione il grafico a linee mostrato nella fascia centrale sulla destra, il quale riporta il trend del rapporto percentuale precedentemente illustrato.

In tale visualizzazione si riscontra una sostanziale e progressiva crescita per tutta la durata del mese di Novembre 2020, seguita da una lenta decrescita nelle prime due settimane del mese di Dicembre. Nella prima settimana di Gennaio 2021, però, l'incidenza dei casi diagnosticati sui positivi è aumentata molto di più, e tale aumento è ancora più importante nella seconda settimana di Gennaio, nonostante i casi indagati siano stati più o meno costanti.

Tutto ciò può significare che non vi è una relazione stretta tra il numero di positivi individuati e il numero di casi diagnosticati; è chiaro che se non si indaga un numero sufficiente di soggetti non sarà possibile avere un numero di positivi che permetta di capire effettivamente cosa stia succedendo nella dinamica epidemica. Non è soltanto il numero dei test effettuati a determinare la percentuale di positività, ma anche, e soprattutto, la scelta dei casi da indagare.

Difatti, la positività dei casi indagati aumenta tra la fine di Ottobre e la fine di Novembre, e questo significa sia che è aumentata l'incidenza ma anche che la scelta dei casi da sottoporre a tampone si concentra maggiormente su quelli più a rischio; ciò conferma che l'attività di contact tracing ha funzionato, portando a concentrare i test sui contatti avvenuti con i casi diagnosticati, i quali sono, appunto, a maggior rischio di positività.

### 4.3.1 Media settimanale

Una funzionalità molto interessante utilizzata in Power BI Desktop è la media mobile, la quale permette di effettuare smoothing su un dato e valutarlo, nel complesso, come media piuttosto che nel suo valore assoluto. Lo scopo, infatti, è quello di “ammorbidire” i dati primitivi, che danno un'idea piuttosto relativa dell'informazione.

Nel caso di studio della tesi, in particolare, vi è fatto uso della media settimanale, creata attraverso una funzione DAX che porta alla misura della Figura 4.12.

```

1 Media tamponi 7GG = CALCULATE(AVERAGEX(DATESINPERIOD('covid-19-marche'[Data], LASTDATE('covid-19-marche'[Data]), -7, DAY), [Media di Tamponi Effettuati Odierni per Settimana]))

```

**Figura 4.12.** Formula DAX per il calcolo della media settimanale

Gli ultimi due grafici da analizzare, mostrati nella fascia sinistra del dashboard, si basano proprio su tale misura. Il grafico che illustra il numero dei tamponi giornalieri mostra un dato interessante. Con l'arrivo della seconda ondata di contagi, il numero dei tamponi giornalieri elaborati è aumentato a partire da Ottobre, ed è cresciuto

in maniera più accentuata per tutto il mese di Novembre e Dicembre, arrivando a una situazione di stallo nel mese di Gennaio. Tale costanza, probabilmente, dichiara qual è il livello del sistema sanitario, ovvero il valore massimo di tamponi in grado di processare, con lo scopo di identificare immediatamente ogni singolo caso sospetto.

Come si può notare dalla visione simultanea di questi ultimi due grafici, inoltre, la diminuzione del numero dei casi positivi giornalieri non è correlata a un calo di tamponi effettuati, i quali, invece, sono addirittura triplicati.

## 4.4 Cruscotto relativi ai malati attivi

In questa sezione si prosegue l'analisi concentrandosi esclusivamente sul dato dei positivi al COVID-19. In particolare, si analizzeranno la suddivisione dei casi positivi per gravità e la crescita nel numero di persone infette nell'intera Regione Marche e nelle singole province.

Prima di esaminare i dashboard creati, però, è necessario chiarire che si utilizzeranno, indistintamente, i termini “Casi Positivi” e “Malati” per riferirsi alle persone risultate positive al virus.

Nella Figura 4.13 è riportato il dashboard introduttivo sulla situazione attuale dei malati e sulla suddivisione per gravità all'interno della Regione Marche aggiornata al 27 Gennaio 2021.

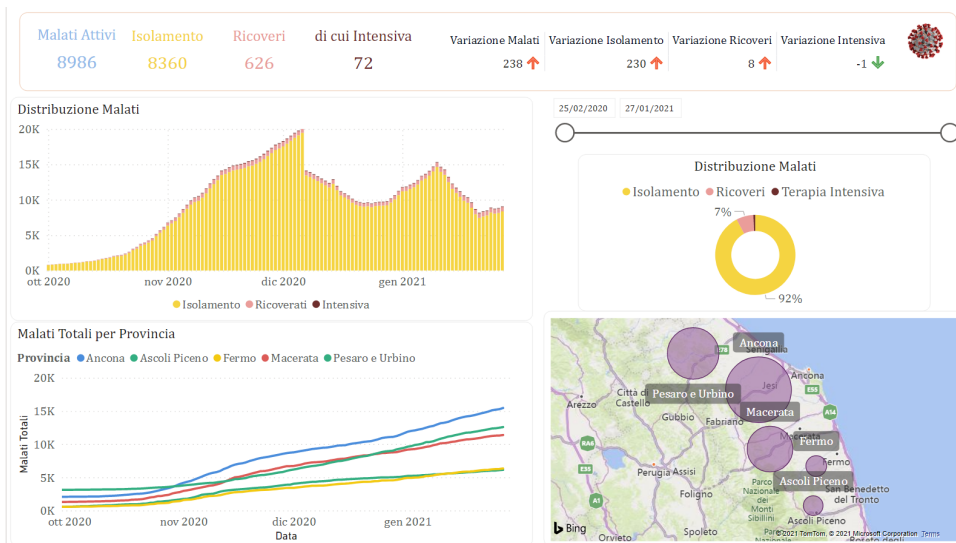


Figura 4.13. Il dashboard relativo ai Malati Attivi

Questo cruscotto, così come quelli illustrati nelle sezioni precedenti, presenta una parte superiore caratterizzata dalla presenza di soli oggetti visivi di tipo “Scheda”. Tali schede riportano, in particolare, il valore assoluto di “Malati Attivi”, “Isolamento”, “Ricoveri” e “di cui Intensiva”. Di questi dati, inoltre, sono riportate le



variazioni giornaliere all'interno di un oggetto visivo di tipo “Tabella”, posizionato sulla parte destra della fascia superiore.

Si ricordi che per, ottenere le variazioni rispetto al giorno precedente, sono state realizzate delle apposite misure DAX che permettono di ottenere colonne calcolate direttamente nel modello dei dati (Figura 4.12).

```
1 Variazione Malati = (var selezionaData = CALCULATE(LASTDATE('covid-19-marche'[Data]), FILTER('covid-19-marche', 'covid-19-marche'[Data] < EARLIER('covid-19-marche'[Data]
))) return 'covid-19-marche'[Malati Attivi Totali] - CALCULATE(SUM('covid-19-marche'[Malati Attivi Totali]), FILTER('covid-19-marche', 'covid-19-marche'[Data] =
selezionaData)))
```

**Figura 4.14.** Formula DAX per il calcolo della colonna relativa alla variazione dei malati

Ritornando all’analisi del cruscotto di Figura 4.13, quest’ultimo presenta un grafico a torta che mostra la distribuzione dei malati attivi tra coloro che sono in isolamento domiciliare, quelli che sono ricoverati nelle strutture ospedaliere e malati che sono in terapia intensiva. La percentuale dei ricoverati, e ancor di più dei ricoverati in terapia intensiva, è nettamente inferiore rispetto a quella dei malati in isolamento domiciliare. Tale risultato permette di comprendere come il virus, nella maggior parte dei casi, si manifesti con sintomi lievi, senza aggravare le condizioni di salute delle persone contagiate.

Considerazioni analoghe valgono anche per l’“istogramma a colonne in pila” presente nella parte sinistra del cruscotto, il quale riporta, giorno per giorno, il valore relativo dei malati in isolamento, dei malati ricoverati e dei malati ricoverati in terapia intensiva.

I due grafici mostrati in basso, infine, illustrano la situazione dei malati nelle singole province marchigiane. Si noti ancora una volta come, per realizzare i grafici relativi alle situazioni provinciali, sia stata impiegata l’apposita relazione “Province”.

In particolare, il grafico a linee rappresenta l’andamento del numero totale di malati per ciascuna provincia della Regione Marche. Dal confronto dei vari andamenti si può osservare come nel mese di Ottobre 2020, fase iniziale della seconda ondata di contagi da COVID-19, la provincia più colpita sia stata Pesaro e Urbino. A partire da Novembre 2020, oltre al sostanziale aumento dei contagiati in tutta la Regione, la provincia di Ancona diventa quella più colpita.

In ultima istanza, è presente una visualizzazione a mappa che permette di illustrare immediatamente la gravità della situazione nelle diverse province, riportando, di fatto, le medesime condizioni dei grafici precedenti.

#### 4.4.1 Analisi ricoveri per provincia

In questa sezione si pone l’attenzione sul dato dei contagiati ricoverati nelle strutture ospedaliere della Regione Marche, aspetto negativo della pandemia.

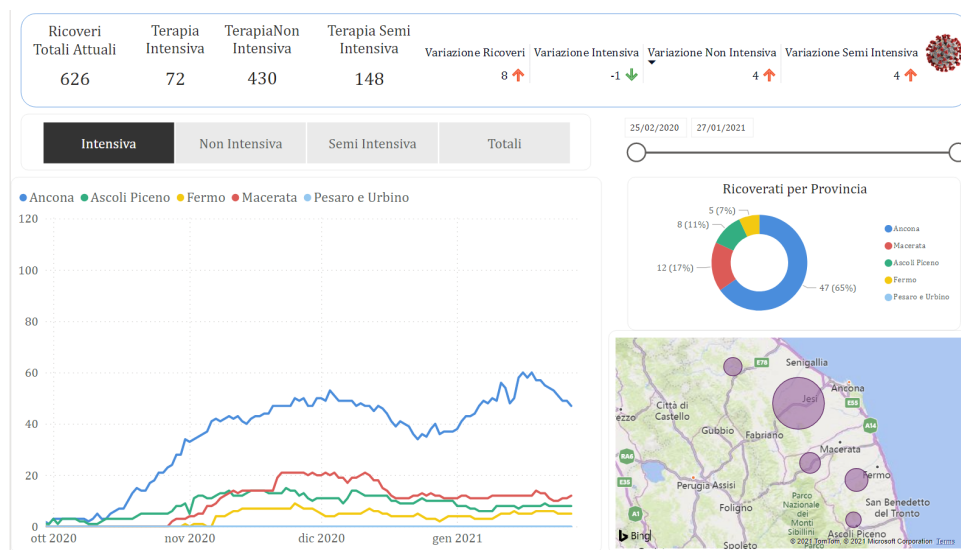
Come si è potuto constatare durante l’evoluzione del virus nel territorio italiano ed, in generale, in tutto il mondo, il COVID-19, in alcuni casi, comporta un aggravamento delle condizioni di salute tale da indurre il malato a sottoporsi al ricovero e, nei casi più gravi, alla terapia intensiva.

Fin dagli inizi della pandemia, l'Italia si è trovata, inaspettatamente, di fronte ad un incremento esponenziale di ricoveri che ha dovuto fronteggiare con i pochi mezzi a disposizione, facendo i conti con la carenza di personale sanitario.

Per tale motivo, fin da subito, lo stato italiano si è disposto al potenziamento dei reparti ospedalieri, soprattutto nelle regioni più colpite, diramando linee di indirizzo assistenziali per i pazienti affetti da COVID-19.

In particolare, la Regione Marche, per far fronte all'imminente crescita di richieste di posti letto nelle varie strutture sanitarie, ha stanziato fondi per la costruzione di nuovi centri da adibire a reparti COVID-19 aumentando, dunque, il numero dei posti letto a disposizione. Tuttavia, la criticità della carenza di personale sanitario è rimasta a causa di un ingente carico di pazienti.

Nella Figura 4.15 è mostrato il cruscotto sulla situazione dei ricoveri della Regione.



**Figura 4.15.** Il dashboard relativo al dato dei ricoveri suddivisi per provincia

Tale cruscotto fornisce informazioni precise sul numero dei malati ricoverati nelle varie strutture ospedaliere delle Marche, suddivisi sulla base della gravità epidemiologica del paziente. Così come negli altri dashboard finora realizzati, nella parte superiore sono posizionati gli oggetti visivi di tipo “Scheda” che riportano, nel caso specifico, il valore assoluto dei malati ricoverati nei reparti di terapia intensiva, non intensiva e semi-intensiva, nonché il numero esatto dei ricoverati totali, il tutto con la possibilità di selezionare interattivamente il giorno di interesse. Inoltre, le variazioni giornaliere di tali valori sono riportate in un oggetto visivo di tipo “tabella”. Si ricorda, ancora una volta, che i dati nella presente tesi sono aggiornati al giorno 27 Gennaio 2021.

Per ottenere i dati relativi alle variazioni giornaliere sono state create apposite colonne utilizzando le formule DAX, secondo le stesse procedure dei dashboard

analizzati nelle sezioni precedenti.

Si riporta, a titolo di esempio, la formula utilizzata per il calcolo della colonna “Variazione Intensiva” che riporta, riga per riga, la variazione giornaliera del numero dei ricoverati in strutture di terapia intensiva (Figura 4.16)

```
1 Variazione Terapia Intensiva = (var selezionaData = CALCULATE(LASTDATE('covid-19-marche'[Data]), FILTER('covid-19-marche', 'covid-19-marche'[Data] < EARLIER('covid-19-marche'[Data]))) return 'covid-19-marche'[Terapia Intensiva Totali] - CALCULATE(SUM('covid-19-marche'[Terapia Intensiva Totali]), FILTER('covid-19-marche', 'covid-19-marche'[Data] = selezionaData)))
```

**Figura 4.16.** Formula DAX per il calcolo della colonna calcolata relativa alla variazione dei ricoveri in terapia intensiva

Particolarmente interessante, all’interno del cruscotto in analisi, è sicuramente il grafico a linee riportante l’andamento dei ricoveri nei vari reparti ospedalieri, suddivisi per le singole province della Regione.

Per avere un quadro temporale sull’andamento dei ricoveri, invece, si può focalizzare l’interesse sulle curve relative ai ricoveri totali giornalieri delle singole province.

La provincia di Ancona è stata, indubbiamente, la più colpita in termini di ospedalizzati nel corso della seconda ondata di contagi del virus. La curva in questione, infatti, mostra un andamento assimilabile a due campane con un primo picco di 445 ricoverati il 2 Dicembre 2020 e un secondo picco di 476 ricoverati il 16 Gennaio 2021. La fase di discesa relativa alla prima campana, nel periodo che intercorre tra i primi di Dicembre e la fine dello stesso mese, è l’effetto delle rigide misure di contenimento messe in atto dallo stato italiano nelle settimane precedenti. Tali restrizioni, entrate in vigore a Novembre e successivamente rallentate nel periodo delle festività natalizie, infatti, hanno portato ad un aumento di malati costretti all’ospedalizzazione.

Tale dato non è confortante perchè, ancora una volta, è sintomo del fatto che ci si è lasciati trascinare da un segnale di decrescita dei ricoveri per attenuare i parametri delle restrizioni, dimenticandosi, d’altro canto, della continua circolazione del virus.

Nel capitolo della presente tesi relativo alla diffusione della pandemia da COVID-19 è stata accennata la scelta del premier Conte di suddividere l’Italia in tre zone a differente livello di allerta. La valutazione per le fasce di colore tiene in considerazione l’andamento regionale e, a tal proposito, spesso si guarda uno degli indicatori più d’impatto ma, allo stesso tempo, meno precoci del peggioramento della situazione: l’andamento del numero di persone in terapia intensiva. Pur essendo un indicatore che arriva “in ritardo”, si tratta, però, anche di un indice che consente di descrivere con maggior precisione la gravità dell’infezione, dal momento che, mentre il numero di nuovi casi dipende anche dalla capacità di testing regionale, quello dei ricoveri (e in particolare dei ricoveri in condizioni più gravi) è pressochè indipendente dalla capacità di testing.

In seguito a tali chiarimenti, sembra abbastanza intuitivo collocare la Regione Marche nella zona più ad alto rischio, in corrispondenza dei due picchi delle campane dell’ultimo grafico preso in considerazione.

Nel dashboard trattato in questa sezione, inoltre, è presente un filtro dei dati con la possibilità di selezionare le terapie per gravità ed effettuare un confronto tra le

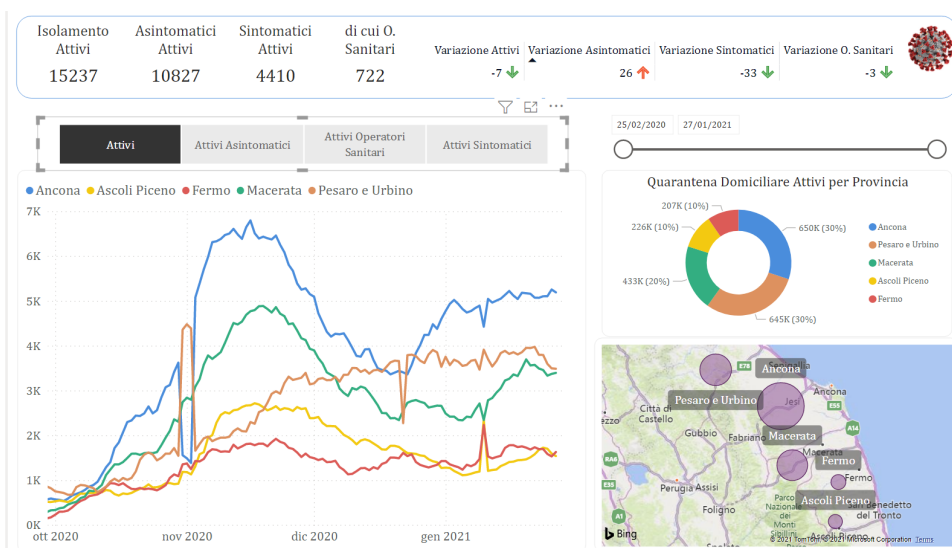
diverse province. In altre parole, cliccando su una determinata terapia, tutti i grafici presenti all'interno del cruscotto vengono filtrati sulla base del dato selezionato.

#### 4.4.2 Analisi quarantena domiciliare per provincia

Con il ritorno dei contagi e la difficoltà delle regioni nel tracciare la diffusione della pandemia con la seconda ondata, uno degli aspetti cruciali per il controllo è quello della corretta gestione di isolamenti e quarantene per le persone infette o potenzialmente contagiate.

Mentre l'isolamento è riferito a persone risultate effettivamente positive, la quarantena interessa le persone sane che sono venute a contatto con un positivo e potrebbero essere potenzialmente esposte al rischio infettivo.

Alla luce di ciò, si è deciso di realizzare un dashboard riportante i dati sulla quarantena domiciliare, particolarizzato alle singole province della regione. Il cruscotto è riportato nella Figura 4.17 e mostra una struttura analoga a quella dei cruscotti analizzati finora.



**Figura 4.17.** Il dashboard relativo al dato della quarantena domiciliare suddiviso per provincia

Nella parte superiore del dashboard sono illustrate, sotto l'oggetto visivo “Scheda”, le seguenti informazioni:

- *Isolamento Attivi*: numero totale di persone attualmente in isolamento domiciliare.
- *Isolamento Sintomatici*: numero totale di persone attualmente in isolamento domiciliare con sintomi riconducibili al COVID-19.
- *Isolamento Asintomatici*: numero totale di persone attualmente in isolamento domiciliare senza alcun sintomo.

- *di cui Operatori Sanitari*: personale sanitario in isolamento domiciliare perchè risultato positivo o ha avuto contatti con malati.

Sono presenti, inoltre, le variazioni giornaliere di tali numeri attraverso un oggetto visivo di tipo “tabella”, ottenuto mediante colonne calcolate con formule DAX secondo le procedure viste nei dashboard precedentemente analizzati.

Un oggetto visivo interessante, all’interno del dashboard, è il grafico a linee che riporta gli andamenti delle persone attualmente in isolamento domiciliare, distinti tra asintomatici e sintomatici, ed, inoltre, particolarizzato per le singole province. Ciò permette di avere un quadro generale sull’andamento dell’epidemia e, allo stesso tempo, effettuare confronti per provincia.

Il dato che emerge da tale grafico è che il numero degli attivi asintomatici è nettamente superiore a quello dei sintomatici. L’informazione che si può estrarre è che, come nella maggior parte dei casi da COVID-19, la malattia si manifesta in forma lieve con la possibilità di rimanere a casa a riposo finchè non ci si riprende.

Infine, si può notare, ancora una volta, come la provincia di Ancona sia la più colpita anche da questo punto di vista.

## 4.5 Dashboard sui decessi

La pandemia da COVID-19 ha, purtroppo, colpito tutta la regione Marche provocando decessi in tutte le sue province.

È bene ricordare come i cruscotti presentati in questa sezione siano stati realizzati utilizzando la tabella `dettaglio-decessi`, la cui struttura è mostrata nella Figura 4.2

Si tratta di una tabella contenente i dettagli dei decessi giornalieri, in particolare le caratteristiche di ogni persona deceduta a causa del COVID-19. Come si può notare dalla figura, non sono presenti soltanto le colonne della tabella stessa, bensì anche alcune misure ottenute utilizzando le formule DAX, piuttosto facili da riconoscere, in quanto rappresentate mediante il simbolo della calcolatrice.

Le misure sono state usate per arricchire il modello dei dati e per rispondere a tutti quei quesiti importanti per i quali, con i soli dati originali, non si riuscirebbe a trovare una risposta.

In particolare, sono state create due misure legate alla relazione `dettaglio-decessi`; tali misure sono:

- *ConPatologie*: questa misura conta il numero di persone decedute nelle quali, al momento del decesso, sono state riscontrate una o più patologie. Effettua, quindi, un semplice conteggio delle righe (Figura 4.18).
- *SenzaPatologie*: questa misura, invece, esegue il calcolo complementare a quello precedente, facendo, cioè, riferimento alle persone decedute e senza patologie pregresse (Figura 4.19).

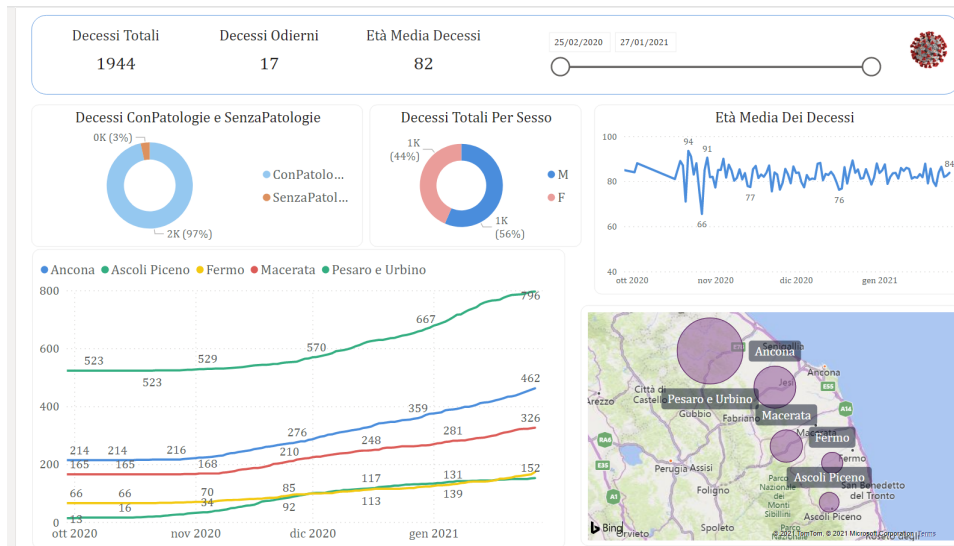
Dunque, avendo a disposizione la relazione relativa al dettaglio dei decessi e le misure calcolate appena illustrate, è stato possibile costruire un dashboard relativo alla situazione dei decessi nelle Marche. Tale dashboard riporta gli oggetti visivi che

```
1 Decessi ConPatologie = CALCULATE(COUNT('dettagli_decessi'[ID]), 'dettagli_decessi'[Patologie Progressive] = 1)
```

**Figura 4.18.** Misura calcolata “ConPatologie”

```
1 Decessi SenzaPatologie = CALCULATE(COUNT('dettagli_decessi'[ID]), 'dettagli_decessi'[Patologie Progressive] = 0)
```

**Figura 4.19.** Misura calcolata “SenzaPatologie”



**Figura 4.20.** Il dashboard relativo al dato dei decessi

riassumono i campi della relazione per dare una visione panoramica sulla situazione dei decessi (Figura 4.20).

Nel cruscotto sono presenti, nella parte superiore, delle semplici “Schede”, che permettono di mostrare direttamente i numeri più importanti legati ai decessi.

Tali schede riportano, infatti, il numero totale dei decessi, il numero dei decessi giornalieri e l’età media dei decessi, aggiornati al 27 Gennaio 2021.

Al di sotto delle schede sono presenti, invece, due grafici a torta che illustrano l’uno la ripartizione dei decessi per sesso, e l’altro la ripartizione delle persone decedute tra quelle con almeno una patologia pregressa e quelle senza patologie pregresse.

In particolare, dal grafico a torta relativo alla suddivisione dei decessi tra femmine e maschi, si nota facilmente come il virus sia più letale nel sesso maschile. Dei 1944 decessi nelle Marche, infatti, 1098 (56,48%) sono maschi e 846 (43,52%) sono femmine.

Un’informazione significativa si può trarre dal grafico relativo alle patologie pregresse tra i deceduti a causa del COVID-19 nella Regione Marche. A tal proposito, è importante sottolineare come i dati messi a disposizione per il caso di studio sui dettagli delle persone decedute non forniscono il numero esatto di patologie pregresse. Il report fornito dalla Regione Marche e dal GORES riporta, infatti, semplicemente, il numero intero 1 per indicare i deceduti con una o più patologie pregresse e

il numero 0 per indicare i deceduti senza patologie pregresse. Avendo, comunque, solo tale dato a disposizione, si è in grado di evidenziare come il 97% dei deceduti fossero affetti da patologie, e soltanto il restante 3% degli stessi fossero deceduti senza alcuna patologia pregressa.

Un aspetto interessante di questo dashboard è il grafico riportante l'età media giornaliera dei decessi; si tratta di un grafico a linee che, nonostante abbia una struttura a “dente di sega”, permette di verificare chiaramente un'età media dei decessi molto alta, circa 82 anni. Tale dato conferma, dunque, come il virus sia letale soprattutto tra la popolazione più anziana.

Osservando i due grafici in basso, infine, emerge il grande impatto della provincia di Pesaro Urbino nel numero di deceduti dell'intera regione, con quasi 800 decessi su 1944 totali, dunque, oltre il 40%. Questo valore così discordante è da ricercare su ciò che è accaduto nella prima ondata di contagi. All'inizio dell'evolversi dell'epidemia, infatti, Pesaro Urbino fu sorpresa da un consistente focolaio, con strutture ospedaliere al collasso e gli operatori sanitari stremati. Non sono mancate, in aggiunta, polemiche sul perchè i focolai si sono sviluppati soltanto in alcune città e se il numero dei decessi rispondeva alla realtà delle cose.

#### 4.5.1 Distribuzione per comune

Si analizza, ora, la situazione dei decessi avvenuti nelle Marche comune per comune. La Figura 4.21 mostra un dashboard che illustra, mediante una visualizzazione a mappa, la distribuzione geografica dei decessi da COVID-19 nella regione.

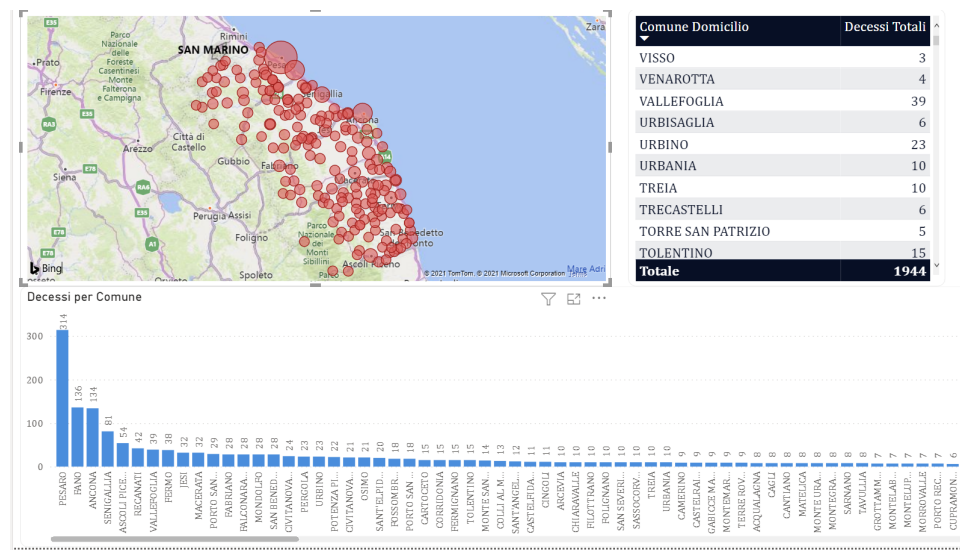


Figura 4.21. Il dashboard relativo al dato dei decessi comune per comune

Il cruscotto presenta, inoltre, una tabella riportante il numero dei decessi per ciascun comune. Tale tabella, in realtà, funge anche da filtro dei dati; cioè, se si



seleziona un comune, tutti gli altri grafici presenti nel dashboard riportano soltanto i dati del comune selezionato.

Un comportamento analogo si ottiene anche selezionando un comune in un qualsiasi altro grafico presente all'interno del cruscotto.

Ancora una volta si può notare come la provincia di Pesaro Urbino sia stata la più colpita per quanto riguarda il numero dei decessi.

## 4.6 Confronto prima e seconda ondata di contagi

La prima ondata pandemica da COVID-19, come facilmente comprensibile, ha colto tutti di sorpresa ma, grazie allo sforzo eroico di medici ed infermieri, e i necessari sacrifici dei cittadini, è stato possibile uscire da un simile momento terribile.

Ciò che è accaduto nella seconda ondata, però, è stato diverso, sotto certi aspetti, in quanto i rischi erano già noti e nel corso dell'estate è stato commesso l'errore di sottovalutare l'incombente minaccia del riacutizzarsi dei picchi di contagi con il cambio di stagione.

Tra la fine della prima ondata e l'inizio della seconda, infatti, c'è stata una finestra in cui le persone, probabilmente, hanno avuto la sensazione che tutto fosse tornato alla normalità, per poi cadere in una situazione più critica rispetto alla prima ondata, nonostante ci fosse una maggiore consapevolezza, conoscenza e capacità di diagnosticare il COVID-19.

A tal proposito, per quanto riguarda i tamponi, mentre nella primavera 2020 si aspettava che il paziente arrivasse al pronto soccorso con sintomi del virus, nell'autunno 2020 ci si è preferito, invece, attuare un'indagine di contact tracing. Quest'ultimo termine che, letteralmente, esprime il tracciamento dei contatti, è l'attività di ricerca e gestione dei contatti di un caso confermato di COVID-19. Identificare e gestire la rete di contatti dei casi confermati di COVID-19, di fatto, permette di individuare e isolare rapidamente gli eventuali casi secondari e interrompere così la catena di trasmissione.

Se, da una parte, il contact tracing ha permesso di ricostruire le catene di infezioni ed impedire nuovi contagi, dall'altra ha generato ansia e preoccupazione nella popolazione in quanto i numeri di casi positivi sono stati importanti. Progressivamente, infatti, si è verificato il nuovo affollamento ospedaliero e un ripristino dei reparti di COVID-19 che, durante il periodo estivo, erano stati adibiti a ricoveri di altra natura.

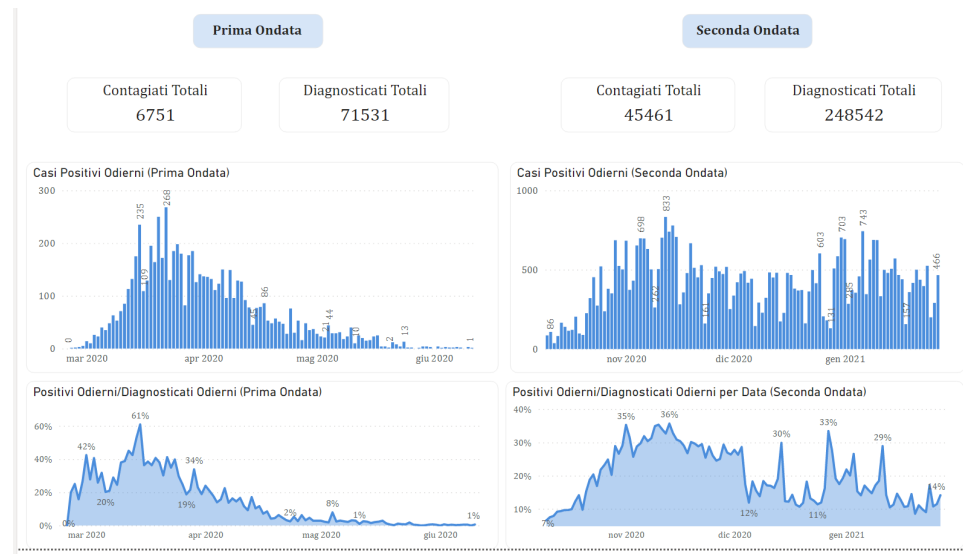
Nell'attività di analisi della presente tesi sono state messe a confronto la prima e la seconda ondata di contagi da COVID-19 considerando lo stesso numero di giorni (109), nel caso specifico della Regione Marche. In particolare, la prima ondata è rappresentata dal periodo che va dal 25 Febbraio 2020 al 10 Giugno 2020 mentre la seconda ondata, invece, va dal 10 Ottobre 2020 al 27 Gennaio 2021.

Si è pensato, dunque, di paragonare i dati dei parametri che meglio sembrano caratterizzare la pandemia: i casi positivi, gli ospedalizzati ed, infine, i decessi registrati. Per ciascuno dei parametri è stato creato un apposito report, utilizzato per le successive discussioni.

La Figura 4.22 mostra i dati relativi all'andamento dei casi risultati positivi al COVID-19, suddivisi per ondata di contagi. Il dashboard, infatti, è diviso a metà:



nella parte sinistra ci sono gli oggetti visivi che rappresentano la prima ondata, a destra, invece, quelli che raffigurano la seconda ondata.



**Figura 4.22.** Il dashboard relativo all'andamento dei casi positivi suddivisi per ondata di contagi

Come è facilmente comprensibile dalle schede che mostrano i valori assoluti dei casi positivi totali, la seconda ondata ha interessato un numero di contagiati 7 volte superiore rispetto alla prima. Infatti, mentre nella prima ondata sono state infettate 6.751 persone, nella seconda fase il numero dei contagiati è risultato pari a 45.461.

Interessanti risultano gli istogrammi che seguono l'andamento dei casi positivi giornalieri, mostrati nella parte intermedia del cruscotto.

L'informazione che si può trarre da tali istogrammi è che, mentre la prima fase della pandemia è stata caratterizzata da un numero di contagi che ha raggiunto un massimo a fine Marzo 2020 per poi iniziare una graduale discesa, la seconda ondata è stata caratterizzata da un'onda lunga che ha raggiunto più lentamente il plateau.

La contagiosità del virus, dunque, è enormemente aumentata: a differenza di quanto accadeva nei mesi del primo lockdown, dalla fine di Agosto in poi il virus ha fatto riscontrare una maggiore facilità di propagazione e contagio, mantenendo al tempo stesso un'aggressività invariata rispetto a marzo, a causa del calo dell'attenzione generale da parte di gran parte della popolazione e per l'allentamento delle norme preventive.

Un'altra sostanziale differenza riscontrata è che, da una parte, la prima ondata ha visto raggiungere il suo picco in poche settimane, dall'altra la seconda è stata caratterizzata da un flusso lungo, che ha visto più lentamente l'apice. Nello specifico, il picco giornaliero di contagi relativo alla prima ondata è stato di 268 persone e si è raggiunto il 22 Marzo 2020, a 28 giorni dall'inizio della pandemia; nella seconda

ondata il picco giornaliero è stato di 833 contagi e si è raggiunto il 12 Novembre 2020, al 60esimo giorno.

Un altro aspetto di notevole importanza in questo dashboard è il rapporto percentuale tra casi positivi e casi diagnosticati. Tale dato è stato ricavato dalle informazioni già presenti nel modello di dati in Power BI, applicando una semplice funzione DAX di divisione, ed esprimendo il risultato in notazione percentuale.

Questa misura rappresenta un indice abbastanza rilevante per valutare il fattore di crescita della pandemia, in quanto non indica un valore assoluto, ma permette di mettere in relazione il numero di nuovi casi positivi di ciascun giorno con il numero dei casi diagnosticati dello stesso giorno. A tal proposito, particolarmente significativi sono i confronti dei grafici ad aree riportati nella parte inferiore del cruscotto, i quali mostrano giornalmente questo rapporto percentuale.

La prima ondata di contagi vede una prima fase di crescita della misura in questione per quasi tutta la durata del mese di Marzo. Il 15 Marzo 2020 la percentuale arrivò addirittura al 61% dei tamponi effettuati. Dall'altro lato, però, si deve tenere conto, anche, che nei dati del mese di Marzo non si faceva distinzione tra tamponi e nuovi casi indagati, per cui è probabile che, spesso, per ciascun individuo fossero stati eseguiti due tamponi, e quindi la percentuale di casi testati risultati positivi doveva essere, in realtà, molto maggiore, probabilmente attorno alla metà dei soggetti indagati.

A partire dalla fine di Marzo 2020 ci fu una sostanziale e progressiva decrescita.

La positività dei casi indagati è aumentata dalla metà di Ottobre 2020 fino agli inizi di Dicembre 2020, e questo significa sia che è aumentata l'incidenza ma anche che la scelta dei casi da sottoporre a tampone si fosse concentrata maggiormente sui casi più a rischio. Ciò ha fatto sperare in un buon funzionamento dell'attività di contact tracing allora in atto.

È, forse, questa una conclusione paradossale, ma riscontrare una maggior positività nei test eseguiti, anche se comporta rilevare una maggior incidenza, è probabilmente anche un buon segno del funzionamento dell'attività preventiva, e non, invece, del fatto che aumenta la percentuale di casi asintomatici spontanei, tra i sottoposti al tampone.

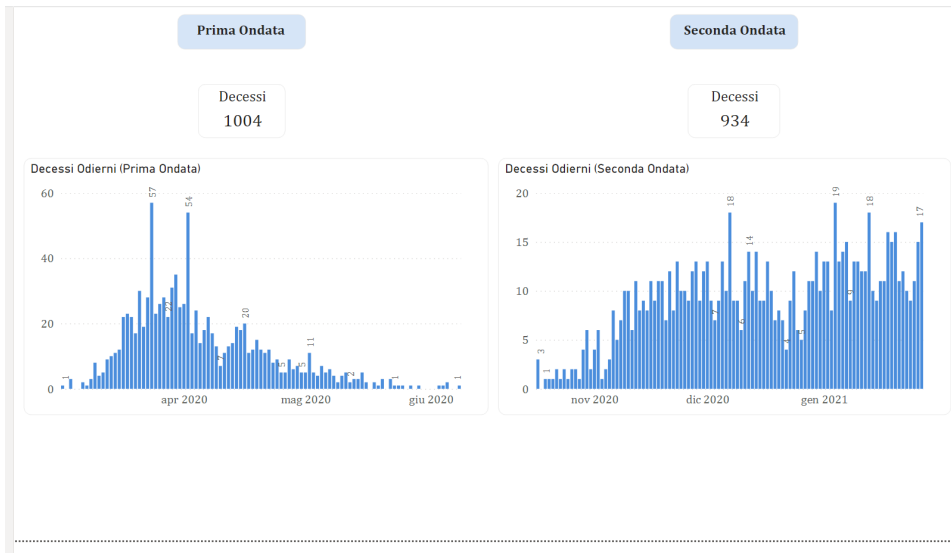
Come è evidente dalle precedenti considerazioni, non è facile confrontare i due periodi. Nel caso della prima ondata non si sa, effettivamente, quando la pandemia abbia avuto inizio, poiché quando sono iniziati i primi tamponi, probabilmente, il contagio era già molto esteso.

Quando, a Settembre 2020, ci si è resi conto che la pandemia era ricominciata a manifestarsi, si è verificato un incremento dei ricoverati e delle terapie intensive, come riportato nella Figura 4.23.

In tale dashboard, nella parte inferiore, sono raffigurati gli istogrammi che mostrano l'andamento delle ospedalizzazioni rispetto al numero di infetti attivi. È molto evidente che il numero di infetti attivi nella seconda fase è dominato dalle quarantene, quindi da malati con sintomi lievi o assenti.

Inoltre, la crescita degli ospedalizzati è molto più lenta rispetto all'inizio dell'epidemia. Durante la prima ondata la frazione di positivi ospedalizzati ha superato il 15%, mentre nella seconda ondata si è assestata intorno allo 0,3%.





**Figura 4.24.** Il dashboard relativo all'andamento dei decessi suddivisi per ondata di contagio

ondata di contagi è stato possibile ricavare le seguenti informazioni:

- Le statistiche sugli infetti sono state dominate dagli asintomatici, fatto che ha creato moltissimo disagio sociale, considerati i numeri, ma, da un punto di vista diagnostico, è stata un'ottima notizia. Infatti si è trattato di persone che, nel corso della prima ondata, non venivano identificate. È stata, dunque, una sorta di lockdown mirato che è risultato molto efficace.
- La crescita dell'epidemia è stata molto più lenta che nel passato: anche questa è stata una buona notizia.
- L'impatto sulla società è stato, comunque, notevole, raggiungendo la stessa intensità di terapie intensive della prima ondata, seppur con andamenti diversi. Ciò ha portato, inevitabilmente, a situazioni di emergenza con la concretizzazione di un secondo lockdown nella stagione autunnale 2020.

## Discussione in merito al lavoro svolto

*In questo capitolo si illustreranno le principali lezioni apprese dalla presente tesi, al fine di poterle utilizzare come base di conoscenza per i progetti futuri.*

### 5.1 Premessa

Nel corso dell'attività di tale tesi è stato analizzato l'andamento della pandemia da COVID-19 durante la seconda ondata di contagi all'interno della Regione Marche. Tale pandemia ha avuto origine in Cina il 31 Dicembre 2019, ma ben presto si è diffusa con rapidità e in un territorio sempre più vasto, raggiungendo, di volta in volta, quasi tutti i continenti.

In Italia, la seconda ondata di contagi iniziò a manifestarsi dopo la stagione estiva, assistendo ad una brusca risalita nel numero dei contagi.

La Regione Marche, in particolare, si è fin da subito impegnata nel fornire, con grande chiarezza e trasparenza, i dati circa la diffusione del virus all'interno delle province. Lo studio di analisi effettuato si è basato proprio su tali dati, messi a disposizione direttamente sul sito della Regione Marche.

Lo strumento software utilizzato per l'attività di analisi è stato Power BI che ha permesso di affrontare le fasi principali della Data Analysis.

In particolare, nel corso delle attività connesse alla presente tesi, si è avuto modo di trarre delle lezioni utili sulle seguenti tematiche:

- l'importanza della comunicazione mediante i dati;
- il formato CSV per lo scambio di dati;
- il software Power BI;
- l'analisi dei dati da cui estrarre conoscenza.

Tali lezioni saranno illustrate in dettaglio nelle prossime sezioni.

## 5.2 Comunicare con i dati

L'era digitale in cui ci si trova è caratterizzata dall'enorme quantità di dati che, sotto diversi aspetti, rappresenta una rivoluzione che influenza fortemente la vita quotidiana dell'individuo.

La trasformazione digitale, infatti, ha introdotto alcune novità strutturali nel modo di accedere alle informazioni. Innanzitutto, i contenuti sono oggi sganciati da qualsiasi supporto fisico, il che ne rende l'accesso molto più libero e condiviso, risolvendo le problematiche di archiviazione e veicolazione di enormi quantità di dati in maniera semplice e veloce.

Il fluire delle informazioni ha permesso, inoltre, una forte riduzione degli intermediari che, fino a qualche tempo fa, svolgevano il compito di mediare la conoscenza, a vantaggio del grande pubblico; oggi l'individuo ha potere di accesso ai dati senza vincoli, al punto che ne è lui stesso creatore, mediante i dispositivi e le interazioni in real time.

Tale flusso viene registrato dalle organizzazioni per una conoscenza più approfondita del proprio pubblico, non più solo in termini di fattori anagrafici ma, soprattutto, nel campo psicografico, in tutto quel che riguarda abitudini, attitudini, percezioni e preferenze.

Durante lo studio di attività del suddetto elaborato, ci si è resi conto che, nonostante i diversi strumenti digitali a disposizione, comunicare i dati della pandemia, in modo semplice e rigoroso, non è affatto facile, soprattutto se ci si rivolge a un pubblico ampio e di non esperti.

Da quando il COVID-19 ha iniziato a diffondersi in Italia, improvvisamente ci si è trovati immersi in una comunicazione monotematica, fatta di numeri e dati, trasmessi, giorno dopo giorno, per descrivere l'andamento epidemiologico del Paese. Si tratta dell'ormai famoso bollettino del Ministero della Salute che dal 24 febbraio 2020 arriva, ogni giorno, tra le 17 e le 18.

Ma nel trasmettere ogni giorno il dato nudo e crudo, il rischio è quello di generare l'effetto "montagne russe": un giorno i contagi salgono e l'epidemia peggiora, un altro giorno scendono e allora le cose migliorano. Purtroppo le cose sono ben più complicate, e spesso i media, nel comunicare i dati sul coronavirus, a causa di una eccessiva semplificazione, non riescono a trasmettere tale complessità.

Nel corso dell'attività di tale tesi, ci si è immersi nell'analisi dei dati della pandemia e nella loro divulgazione, cercando di cogliere tutta la complessità che portano con sé e provando a comunicarli nella maniera più naturale possibile, senza ricorrere, però, a semplificazioni. Per farlo sono stati utilizzati diversi grafici, consapevoli, però, che essi possano rivelarsi un'arma a doppio taglio se non spiegati in modo semplice e accurato, anche a chi non possiede delle buone basi scientifiche.

## 5.3 Il software Power BI

L'attività di studio ha permesso di entrare in contatto con il mondo dell'analisi dei dati, che riveste un ruolo fondamentale al giorno d'oggi. In particolare, ha consentito di affrontare, nel dettaglio, le varie fasi che caratterizzano una qualunque attività di Business Intelligence.

In tale tesi, le diverse operazioni tipiche di Data Analytics sono state eseguite mediante l'utilizzo del software Power BI Desktop, cercando di sfruttare al meglio le potenzialità messe a disposizione dallo strumento.

Gli strumenti di Business Intelligence usano, in genere, il metodo ETL (Extract, Transform, and Load) per aggregare dati strutturati e non strutturati da più origini. Tali dati vengono, quindi, trasformati e rimodellati prima di essere archiviati in un'ubicazione centrale ed, infine, sottoposti all'attività di analisi.

Un aspetto fondamentale di Power BI che è stato affrontato è la realizzazione di report e dashboard mediante l'utilizzo di oggetti visivi messi a disposizione dallo stesso strumento.

Infine, nell'attività di analisi, è stato approfondito il linguaggio delle formule DAX, di fondamentale importanza nel corso di studio, che ha permesso di completare i dati a disposizione mediante la realizzazione di misure, colonne e tabelle calcolate, senza dover modificarne direttamente il modello.

## 5.4 I formati JSON e CSV per lo scambio di dati

La progettazione dei task della presente tesi è iniziata con la costruzione del dataset riguardante i dati della seconda ondata di contagi da COVID-19 nella Regione Marche. Una volta realizzato, il dataset è stato importato su Power BI Desktop ed è stata effettuata l'attività di analisi.

In particolare, in fase di realizzazione del dataset, sono stati generati, mediante l'utilizzo di script Python, file con formati JSON e CSV, anche se, nell'importazione della sorgente di dati di interesse su Power BI Desktop, è stato utilizzato soltanto il file di tipo CSV.

Nel seguito verranno descritte brevemente le caratteristiche di entrambi i formati, soffermandosi sulla questione che ci ha spinti all'utilizzo del formato CSV piuttosto che di quello JSON in fase di importazione su Power BI Desktop per la creazione del modello di dati.

JSON (JavaScript Object Notation) è un formato leggero per lo scambio di dati, facile da leggere e scrivere per gli esseri umani e facile da generare e analizzare per le macchine. Si basa su un sottoinsieme del linguaggio di programmazione JavaScript; in particolare, rappresenta una notazione che, all'interno di tale linguaggio, consente di rappresentare, in maniera semplice, un oggetto per mezzo di una stringa che racchiude tra parentesi graffe le coppie *nome:valore* delle proprietà dell'oggetto, separate da virgola.

JSON è basato su due strutture:

- *Un insieme di coppie nome/valore*: in diversi linguaggi, questo è realizzato come un oggetto, un record, una struct, un dizionario, una tabella hash, un elenco di chiavi o un array associativo.
- *Un elenco ordinato di valori*: nella maggior parte dei linguaggi questo si realizza con un array, un vettore, un elenco o una sequenza.

I formati CSV (Comma Separated Values), invece, sono file di testo semplici che contengono elenchi di dati. Tale formato consente di esportare facilmente i dati, ed

importarli in altri programmi, per poterli visualizzare e riutilizzare in base all'utilità o alla necessità.

Nel formato CSV i valori di ciascuna cella sono separati da virgole (o punti e virgola); ogni riga di testo corrisponde a una riga della tabella, mentre ogni virgola corrisponde a una separazione tra le colonne.

Il formato CSV rappresenta uno dei modi più semplici per rappresentare dati in forma tabellare all'interno di un semplice file di testo. Non a caso, questo formato di file è uno dei primi ad essersi diffuso per l'interscambio di dati. E, nonostante la presenza di altri formati altrettanto conosciuti ed utilizzati, è ancora oggi il più utilizzato in molte applicazioni.

Per quanto riguarda le differenze che contraddistinguono i due formati, i file CSV sono più comunemente usati per esportare e importare dati, o per elaborarli per l'analisi e l'apprendimento automatico. I file JSON, invece, hanno gli stessi vantaggi, ma sono più comuni nelle soluzioni di scambio di dati a caldo.

Infine, la scelta di utilizzare file CSV, piuttosto che file JSON, nell'importazione in Power BI risiede proprio nell'eterogeneità della loro struttura. Mentre un file CSV contiene, al proprio interno, una serie di dati organizzati in modo da strutturare una tabella, un file JSON, a causa della sua struttura complicata, necessita di alcune operazioni di manipolazione.

Dunque, saper lavorare con i dati richiede la conoscenza dei diversi formati che occorre analizzare, al fine di organizzarne la gestione e l'utilizzo nel modo più appropriato ed efficiente e per poter permettere lo scambio di qualunque tipo di informazioni.

## 5.5 L'analisi dei dati da cui estrarre conoscenza

Il percorso di studio effettuato in tale tesi ha consentito di inoltrarsi nel mondo della Data Science, in particolare nell'analisi dei dati finalizzata all'estrazione di conoscenza dagli stessi.

Nel corso dell'attività svolta, infatti, sono state affrontate le varie fasi che caratterizzano un qualunque impiego di Data Analytics, a partire dalla costruzione del dataset, per proseguire con l'operazione di ETL, fino ad arrivare all'attività di analisi descrittiva e diagnostica e a qualche cenno su quella predittiva.

La pandemia da COVID-19, dunque, ha evidenziato l'importanza della raccolta tempestiva dei dati, anche in situazioni di emergenza, e della predisposizione di sistemi che consentano di assestare in corsa le procedure per la loro estrazione, organizzazione ed elaborazione.

Inoltre, è emerso come raccogliere e analizzare dati possa essere utile per comprendere ciò che accade nel presente per provare, successivamente, ad effettuare previsioni su ciò che potrebbe accadere in futuro.

In generale, la grande rivoluzione dei Big Data sta rendendo più agili ed efficienti migliaia di imprese. I Big Data trovano impiego in ogni settore, come, ad esempio, il Marketing, per intercettare le esigenze dei consumatori, conoscere le tendenze e pianificare strategie efficaci. Un settore che sta investendo molto sulla Data Analytics è la banca, la quale, lavorando con i Big Data, utilizza le informazioni transazionali



di un cliente per monitorare continuamente il proprio comportamento in tempo reale, fornendo il tipo esatto di risorsa necessaria in ogni momento.

Le conoscenze acquisite durante l'intero percorso di studio consentono, dunque, di esaminare più in dettaglio i processi che caratterizzano una campagna di analisi di dati che, come rilevato poc'anzi, possono essere propagati ad una vasta classe di applicazioni.

Inoltre, si ritiene che tali conoscenze siano di fondamentale importanza nell'attuale era digitale, che ha visto una crescita esponenziale di dati in ogni aspetto della società. Tale patrimonio di dati, infine, necessita di operazioni di analisi per poterne estrarre informazioni utili, al fine di trasformare i dati stessi in una vera fonte di conoscenza e rendere i processi decisionali più efficaci.



## Conclusioni

In questa tesi è stata studiata la seconda ondata di contagi da COVID-19, al fine di descrivere l'andamento della diffusione della pandemia ed ottenere un quadro generale con analogie e differenze rispetto agli sviluppi della prima ondata all'interno della Regione Marche. In particolare, l'attività di analisi effettuata è stata suddivisa per tematiche, in modo da coprire tutti gli aspetti legati alla pandemia, comprendere ciò che è accaduto e consolidare le basi per una ripresa post-pandemia. A tal proposito, sono stati utilizzati i dati forniti direttamente dalla Regione Marche mediante il "Gruppo Operativo Regionale per le Emergenze Sanitarie", noto anche come GORES. Tali dati, infatti, fanno riferimento ai vari temi legati allo sviluppo della pandemia, e rappresentano la sorgente su cui sono stati applicati gli strumenti di Business Intelligence e Data Analytics.

Per garantire una maggiore chiarezza del caso di studio, il primo passo affrontato è stato quello di introdurre, seppur in maniera sintetica, informazioni sul coronavirus SARS-COV-2, responsabile dell'attuale pandemia. È stata presentata, dunque, una panoramica con le diverse fasi di sviluppo, dai primi casi segnalati fino alla somministrazione di vaccini.

Terminata la fase introduttiva, è stato studiato il tool Power BI, strumento software che trasforma le sorgenti di dati non correlate in un insieme di informazioni coerenti, visivamente accattivanti e interattive. Tra i vari tool di analisi disponibili sul mercato, Power BI è considerato il leader tra i tool di Business Intelligence e Data Analytics. Grazie alla vasta gamma di funzionalità che possiede al proprio interno, è stato possibile effettuare l'intera attività di analisi descrittiva.

Nella fase successiva della tesi, si è posta l'attenzione sulle operazioni svolte per la creazione del dataset riguardante i dati della seconda ondata di contagi da COVID-19 nella Regione Marche. Agli atti pratici, si è trattato del prelievo dei dati giornalieri dai report pubblicati quotidianamente dal GORES e della loro aggregazione con i dati già esistenti nel repository GitHub di interesse. Tale repository, infatti, è stato aggiornato con i dati giornalieri relativi alla seconda ondata, in quanto erano presenti soltanto le informazioni riguardanti la prima. Successivamente, i dati aggiunti sono stati raggruppati utilizzando opportuni Script Python e sono stati creati dei file JSON giornalieri, al fine di generare un unico file in formato CSV. Quest'ultimo, infatti, è quello che è stato importato in Power BI Desktop per eseguire l'intera attività di analisi della tesi.

Per quanto riguarda la progettazione di cruscotti, abbiamo deciso di suddividere le analisi dei dati per tematiche cercando, in qualche modo, di coprire tutti gli aspetti legati alla pandemia. Sono stati realizzati, infatti, dashboard panoramici che riassumono le principali statistiche del virus e le informazioni presenti nel modello dei dati di Power BI Desktop relative alle singole province della Regione Marche. La maggior parte dei dashboard realizzati, inoltre, è stata resa fortemente interattiva; infatti se si seleziona una provincia di un oggetto visivo, quest'ultimo filtra i dati mostrando l'andamento ad essa relativo. La scelta di organizzare i dashboard per singole province ha permesso di effettuare confronti interni alla Regione e seguire l'evolversi della pandemia.

L'attività di analisi effettuata nel corso di questa tesi ci ha permesso di dimostrare come sia necessaria la progettazione di un sistema informativo all'avanguardia per la gestione della diffusione delle malattie infettive, che possa contare sulla disponibilità dei dati in tempo reale e sulla possibilità di incrociare fonti informative diverse. Per combattere una pandemia, infatti, è di fondamentale importanza conoscere le dimensioni del fenomeno attraverso l'analisi dei dati, allo scopo di avere a disposizione elementi oggettivi per prendere decisioni tempestive in ambito politico, economico, ma, soprattutto, sanitario.

Gli esperti di Data Science e Statistica hanno chiesto fin da subito alle istituzioni di fornire i dati in maniera completa e trasparente per poter formulare previsioni affidabili utili ai primari degli ospedali ed, in particolare, ai reparti di terapia intensiva, affinché questi potessero organizzarsi tempestivamente. Nonostante l'impegno da ambo le parti, però, spesso i dati non sono risultati del tutto accessibili in maniera aperta, con la conseguenza che la politica, in quanto a tempestività delle decisioni, ha "tergiversato", in particolar modo durante la seconda ondata. Già agli inizi di Ottobre 2020, infatti, le autorità politiche avrebbero potuto prevedere ed anticipare l'imminente ritorno, poi effettivamente avvenuto, di una crescita esponenziale dei contagi, come nella prima ondata. Tuttavia ciò non è avvenuto, probabilmente perchè non si è fatto sufficiente affidamento su persone esperte in merito alla dinamica evolutiva del fenomeno.

In conclusione, avere a disposizione i dati serve non solo a capire cosa è accaduto, ma è fondamentale soprattutto per disegnare i possibili sviluppi futuri e, conseguentemente, stabilire quali misure intraprendere nella maniera più risoluta possibile perchè la rapidità di decisione dei provvedimenti è cruciale per limitare la diffusione di un virus che si propaga con velocità di diffusione potenzialmente esponenziale. Il dato, soprattutto in ambito sanitario, è un segmento determinante e, al tempo stesso, delicato che, negli anni, non ha ricevuto un adeguato investimento e la necessaria attenzione. Le pandemie sono, sicuramente, una diversa dall'altra perciò uno sviluppo eterogeneo in materia di dati può fornire elementi importanti al fine di creare un Data Warehouse "globale", che possa servire per l'individuazione precoce di informazioni, oggi per lo più raccolte "a canne d'organo". L'ideale, dunque, è un'integrazione con nuovi sistemi di gestione dei dati, in modo da ricavare delle informazioni di trend a partire da un caos meno visibile, ma non per questo inesistente.

La pandemia tutt'ora in corso, quindi, rappresenta una delle dimostrazioni più evidenti del ruolo della tecnologia e dei dati nell'era digitale in cui ci troviamo, diventando la sfida per eccellenza di chiunque vorrà competere nella nuova normalità,

o meglio, post-normalità.

La forza dei dati risiede, però, non solo nella grandezza, ma ancor prima nel loro carattere “diffuso”: dagli strumenti tecnologici più semplici, come gli smartphone, i tablet e i computer, alle soluzioni più complesse, come quelle legate al paradigma dell’IoT o all’Industria 4.0, ogni tassello della rivoluzione tecnologica è una fonte di informazioni preziose. Tuttavia, la raccolta di questi dati non è di per sé sufficiente in quanto la vera sfida è metterli in connessione, individuare delle relazioni e analizzarli in tempo reale.

Solo così, infatti, le informazioni assumono un valore e possono essere trasformate in veri e propri suggerimenti e input in grado di aprire nuovi scenari di supporto alle decisioni strategiche e di predizione a medio e lungo termine.

La grande lezione di questa pandemia lascia spazio a grandi interpretazioni: le aziende, se vogliono essere competitive e innovative, devono essere sempre più “data-driven”, ossia guidate dai dati. I dati, infatti, giocano sempre più un ruolo chiave nella crescita di ogni business, indipendentemente dal settore e dalle dimensioni. La possibilità di estrapolarne informazioni permette di coglierne i benefici in ogni settore produttivo. Solo le aziende in grado di trasformare questa nuova abbondanza di dati in conoscenza, attraverso gli analytics, saranno in grado di creare valore reale.



---

## Riferimenti bibliografici

1. Covid-19: intelligenza artificiale e big data per combattere la pandemia. <https://contaminationuniversity.it/covid-19-intelligenza-artificiale-e-big-data-per-combattere-la-pandemia/>, 2020.
2. Gartner positions Microsoft as a leader in BI and Analytics Platforms for thirteen consecutive years. <https://info.microsoft.com/ww-landing-2020-gartner-magic-quadrant-for-analytics-and-business-intelligence.html?LCID=EN-US>, 2020.
3. Introduzione a Power BI Desktop. <https://docs.microsoft.com/it-it/power-bi/fundamentals/desktop-getting-started>, 2020.
4. La battaglia contro il virus si combatte anche con i dati. <https://www.ilfoglio.it/scienza/2020/03/23/news/la-battaglia-contro-il-coronavirus-si-combatte-anche-con-i-dati-306991/>, 2020.
5. Gestire la pandemia nell'era dei big data. <https://webmagazine.unitn.it/innovazione/93521/gestire-la-pandemia-nell-era-dei-big-data>, 2021.
6. L. Abenavoli, P. Cinaglia, F. Luzza, I. Gentile, and L. Boccuto. Epidemiology of coronavirus disease outbreak: the Italian trends. *Rev Recent Clin Trials*, 15, 2020.
7. A. Aspin. *High Impact Data Visualization with Power View, Power Map, and Power BI*. Apress, 2014.
8. A. Aspin. *Pro Power BI Desktop: Self-service Analytics and Data Visualization for the Power User*. Apress, 2020.
9. H. Chen, R.H.L. Chiang, and V.C. Storey. Business Intelligence and Analytics: From big data to big impact. *MIS quarterly*, pages 1165–1188, 2012.
10. F. Conte. *La comunicazione digitale nell'era dei Big Data. Un'indagine empirica sulla data-driven communication nel contesto italiano*. Franco Angeli, 2020.
11. M. Delmastro and A. Nicita. *Big data. Come stanno cambiando il nostro mondo*. Il Mulino.
12. A. Ferrari and M. Russo. *Introducing Microsoft Power BI*. Microsoft Press, 2016.
13. A. Ferrari and M. Russo. *Analyzing Data with Power BI and Power Pivot for Excel*. Microsoft Press, 2017.
14. A. Ferrari and M. Russo. *The Definitive Guide to DAX: Business Intelligence with Microsoft Excel, SQL Server Analysis Services, and Power BI*. Microsoft Press, 2019.
15. S. Finlay. *Predictive analytics, data mining and big data: Myths, misconceptions and methods*. Springer, 2014.
16. A. Gandomi and M. Haider. Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2):137–144, 2015.
17. A. Haleem, M. Javaid, H. Khan, Ibrahim, and R. Vaishya. Significant applications of big data in covid-19 pandemic. 2020.

18. F.A. Hamzah, C.H. Lau, H. Nazri, D.V. Ligot, G. Lee, C.L. Tan, M.K. Shaib, U.H. Zaidon, A.B. Abdullah, M.H. Chung, C.H. Ong, P.Y. Chew, and R.E. Salunga. CoronaTracker: worldwide COVID-19 outbreak data analysis and prediction. *Bull World Health Organ*, 1:32, 2020.
19. T. Lachev and E. Price. *Applied Microsoft Power BI: Bring Your Data to Life!* Prologika, 2021.
20. E. Livingston and K. Bucher. Coronavirus disease 2019 (COVID-19) in Italy. *Jama*, 323(14):1335–1335, 2020.
21. S. Machiraju and S. Gaurav. *Power BI Data Analysis and Visualization*. De Gruyter, 2018.
22. P.N. Mahalle, N.P. Sable, N.P. Mahalle, and G.R. Shinde. Data analytics: COVID-19 prediction using multimodal data. *Preprints*, 2020.
23. F. Marinuzzi and M. Liciani. *Basi di dati e Big Data: come estrarre valore dai propri dati*. Youcanprint, 2016.
24. B. Marr. *Data Strategy: How to Profit from a World of Big Data, Analytics and the Internet of Things*. Kogan Page, 2017.
25. World Health Organization. Coronavirus disease 2019 (COVID-19): situation report, 72. 2020.
26. B. Powell. *Microsoft Power BI Cookbook: Creating Business Intelligence Solutions of Analytical Data Models, Reports, and Dashboards*. Packt Publishing Ltd, 2017.
27. H.J. Qiu, L.X. Yuan, X.K. Huang, Y.Q. Zhou, Q.W. Wu, R. Zheng, and Q.T. Yang. Using the big data of internet to understand coronavirus disease 2019’s symptom characteristics: a big data study. *Chinese journal of otorhinolaryngology head and neck surgery*, 55:E004, 2020.
28. A. Remuzzi and G. Remuzzi. COVID-19 and Italy: what next? *The Lancet*, 2020.
29. A. Rezzani. *Big Data Analytics. Il manuale del data scientist*. Apogeo Education, 2017.
30. P. Russom. Big data analytics. *TDWI best practices report, fourth quarter*, 19(4):1–34, 2011.
31. T. Singhal. A review of coronavirus disease-2019 (COVID-19). *The Indian Journal of Pediatrics*, pages 1–6, 2020.
32. D. Tosi and A. Campi. How data analytics and big data can help scientists in managing covid-19 diffusion: modeling study to predict the covid-19 diffusion in italy and the lombardy region. 2020.
33. C. Webb. *Power query for power BI and Excel*. Apress, 2014.