

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 cruscotti per l'analisi
dell'andamento della pandemia da Covid-19 nella Regione
Marche**

**Design and implementation of scoreboards for analyzing the
trend of the Covid-19 pandemic in the Marche Region**

Relatore

Prof. Domenico Ursino

Candidato

Denaldo Lapi

Anno accademico 2019-2020

Indice

Introduzione	3
1 Pandemia da Covid-19	7
1.1 Il virus e la pandemia	7
1.1.1 Origine del virus	8
1.1.2 La pandemia.....	9
1.2 Tappe della diffusione	10
1.2.1 La situazione italiana	13
1.2.2 Il caso delle Marche	15
1.3 L'importanza dei dati nell'analisi della diffusione del virus	16
1.3.1 La corretta interpretazione di dati e statistiche	18
2 Power BI	21
2.1 Architettura e descrizione dell'applicazione	21
2.1.1 Il flusso di lavoro	24
2.2 Connessione e importazione dei dati	26
2.2.1 Editor di query	26
2.3 Creazione di report	28
2.3.1 Il foglio di lavoro	28
2.3.2 Tipi di visualizzazione supportati.....	32
2.3.3 Oggetti visivi con script in linguaggio R	36
2.3.4 Filtri	38
2.3.5 Formule DAX.....	39
2.4 Pubblicazione e condivisione di dashboard	42
3 Acquisizione ed elaborazione dei dati	43
3.1 Origine dei dati	43
3.1.1 Dati forniti dalla Regione Marche	43
3.1.2 Struttura dei dati	46
3.2 Attività di ETL	53
3.2.1 Estrazione	54
3.2.2 Trasformazione.....	54
3.2.3 Caricamento	55

IV **Indice**

3.3	Attività di ETL svolte in Power BI	55
3.3.1	Importazione dei dati	55
3.3.2	Modifica dei dati	58
3.3.3	Creazione di nuove query	59
3.3.4	Caricamento dei dati puliti	62
4	Realizzazione di cruscotti relativi ai decessi	65
4.1	Dashboard introduttiva sui decessi	65
4.2	Distribuzione geografica delle morti	69
4.2.1	Distribuzione per provincia	69
4.2.2	Distribuzione per comune	70
4.3	Analisi sulla crescita dei decessi	72
4.3.1	Confronto temporale	72
4.3.2	Crescita dei decessi per provincia	73
4.4	Analisi su età, sesso e patologie dei decessi	75
5	Realizzazione di cruscotti relativi a casi positivi, ricoveri e guariti	79
5.1	Dashboard introduttiva su tamponi e positivi	79
5.1.1	Confronti temporali	82
5.2	Analisi dei casi positivi	83
5.2.1	Analisi per Provincia	86
5.3	Analisi dei ricoveri	90
5.3.1	Ricoveri per strutture	92
5.3.2	Focus sulle terapie intensive	94
5.4	Cruscotti relativi alla quarantena domiciliare	97
5.4.1	Situazione per provincia	98
5.4.2	Analisi sugli operatori sanitari in quarantena domiciliare	98
6	Realizzazione di cruscotti di tipo predittivo	103
6.1	L'analisi predittiva	103
6.1.1	Il forecasting	103
6.2	Forecasting in Power BI mediante l'utilizzo di oggetti visivi R	104
6.2.1	Le serie temporali	105
6.2.2	Il modello ARIMA	107
6.3	Attività di forecasting svolte in Power BI	107
6.3.1	Decessi	108
6.3.2	Malati	110
6.3.3	Ricoveri	111
7	Lezioni apprese	113
7.1	Premessa	113
7.2	L'analisi dei dati	113
7.3	La Business Intelligence	114
7.4	Il tool Power BI	115
7.5	Il formato JSON per lo scambio di informazioni	116
8	Conclusioni	119

Riferimenti bibliografici.....	123
Ringraziamenti	125

Elenco delle figure

1.1	Modello 3D del Covid-19	8
1.2	Il mercato “Huanan Seafood Wholesale Market” di Wuhan	9
1.3	Logo dell’Organizzazione Mondiale della Sanità (OMS/WHO)	9
1.4	Epidemia VS Pandemia	10
1.5	“Infondemia” da Coronavirus	12
1.6	Decessi per Covid-19 al 27 Maggio 2020	13
1.7	Situazione generale in Italia al 27 Maggio 2020	14
1.8	Situazione generale nella Regione Marche al 27 Maggio 2020	16
1.9	Dashboard Johns Hopkins University	17
1.10	Dashboard Protezione Civile	17
2.1	Logo di Power BI	21
2.2	Logo del linguaggio R	22
2.3	Il Magic Quadrant di Gartner relativo ai tool di analisi dei dati	22
2.4	L’ecosistema di Power BI	23
2.5	Schermata di avvio di Power BI Desktop	24
2.6	Power BI Desktop e Power BI Service	25
2.7	Componenti principali del flusso di lavoro in Power BI	25
2.8	Finestra di recupero dei dati Power in BI Desktop	26
2.9	Finestra dell’Editor di query	27
2.10	Esempio di pagina di un report	28
2.11	Foglio di lavoro in Power BI Desktop	29
2.12	Icone delle viste di Power BI	29
2.13	Riquadri laterali in visualizzazione <i>Report</i>	30
2.14	Visualizzazione <i>Dati</i>	31
2.15	Visualizzazione <i>Relazioni</i>	31
2.16	Grafici prefiniti in Power BI	32
2.17	Esempio di oggetto visivo “Scheda”	32
2.18	Esempio di istogramma	33
2.19	Esempio di grafico a linee	33
2.20	Esempio di grafico combinato	34
2.21	Esempio di grafico ad area	35
2.22	Esempio di grafico a torta	35

VIII Elenco delle figure

2.23	Esempio di mappa a bolle	35
2.24	Esempio di grafico a dispersione a bolle	36
2.25	Finestra “Oggetti visivi di Power BI”	37
2.26	Esempio di grafico basato sul modello Arima	38
2.27	Il riquadro “Filtri” nella visualizzazione <i>Report</i>	38
2.28	Il riquadro “Campi” dell’area <i>Report</i> con misure calcolate	40
2.29	Esempio di misura calcolata con DAX	40
2.30	Esempio di colonna calcolata con DAX	41
2.31	Esempio di tabella calcolata con DAX	41
2.32	Esempio di dashboard pubblicata in Power BI Service	42
3.1	Repository di GitHub utilizzato per l’importazione dei dati	44
3.2	La pagina del sito della Regione Marche dedicata al Coronavirus	44
3.3	Il <i>report Blu</i> pubblicato giornalmente dalla Regione Marche	46
3.4	Il <i>report Giallo</i> pubblicato giornalmente dalla Regione Marche	47
3.5	Il <i>report Arancio</i> pubblicato giornalmente dalla Regione Marche	47
3.6	Il processo ETL e le tre fasi di <i>Data Preparation</i>	53
3.7	La directory <code>coronavirus-marche</code> con i file clonati dal repository di GitHub	56
3.8	Caricamento del file <code>json</code> in Power BI Desktop	56
3.9	Funzionalità “In tabella” nell’Editor di Power Query	57
3.10	Funzionalità “Espandi in nuove righe” dell’Editor di Power Query	57
3.11	Esempio di espansione delle righe nell’Editor di Power Query	57
3.12	Tabella risultante dall’espansione delle righe	58
3.13	Tabella con voci di tipo “null” e celle vuote	59
3.14	Tabella relativa ai decessi da Covid-19 nella Regione Marche	60
3.15	Esempio di sostituzione di valori nell’Editor di Power Query	61
3.16	Tabella relativa alla provincia di Ascoli Piceno	62
3.17	Finestra di dialogo “Accoda” nell’Editor di Power Query	62
4.1	Tabella relativa al dettaglio dei decessi da Covid-19	66
4.2	Misura calcolata “Decessi Totali”	66
4.3	Misura calcolata “Parziale Decessi”	66
4.4	Misura calcolata “ConPatologie”	66
4.5	Misura calcolata “SenzaPatologie”	67
4.6	Dashboard introduttiva sui decessi	67
4.7	Il riquadro Filtri con i filtri applicati all’oggetto visivo scheda	68
4.8	Cruscotto relativo al numero di decessi nelle province marchigiane	70
4.9	Formula DAX della colonna calcolata relativa all’impatto dei decessi di ciascuna provincia sul totale della regione	70
4.10	Cruscotto relativo al numero di decessi nei comuni marchigiani	71
4.11	Esempio di cruscotto con filtro dei dati in base al comune selezionato	71
4.12	Cruscotto relativo al confronto temporale del numero dei decessi	72
4.13	Esempio di applicazione della funzione DAX DATEADD	73
4.14	Misure calcolate relative agli abitanti delle province marchigiane	73
4.15	Misure calcolate sui decessi ogni 100.000 abitanti nelle province marchigiane	74

4.16	Calcolo del tasso cumulativo grezzo di mortalità per la provincia di Ancona	74
4.17	Cruscotto relativo al numero di decessi ogni 100.000 abitanti nelle province marchigiane	74
4.18	Tabella con dati sui decessi delle regioni italiane più colpite dal Covid-19	75
4.19	Cruscotto relativo all'analisi di età, sesso e patologie dei decessi	76
4.20	Tabella calcolata relativa ai decessi di persone di età inferiore alla mediana.....	77
4.21	Istogramma relativo ai decessi in Italia sulla base delle età	78
5.1	Dashboard introduttiva su tamponi e casi positivi	80
5.2	Misura DAX per il calcolo del rapporto percentuale tra casi positivi e casi diagnosticati	81
5.3	Cruscotto relativo al confronto temporale dei casi positivi	82
5.4	Cruscotto introduttivo sulla situazione dei malati.....	83
5.5	Colonna calcolata relativa alla variazione del numero degli attivi	84
5.6	Grafico della Protezione Civile relativo alla variazione giornaliera degli attivi in Italia.....	85
5.7	Grafico della Protezione Civile relativo all'andamento di casi positivi, decessi e guariti in Italia.....	86
5.8	Cruscotto relativo all'andamento dei casi positivi nelle province marchigiane	87
5.9	Formula DAX della colonna calcolata relativa alla percentuale dei casi positivi provinciali sul totale	88
5.10	Cruscotto relativo al numero di positivi ogni 100.000 abitanti nelle province marchigiane	88
5.11	Mappa dei casi positivi per 100.000 abitanti in Italia	89
5.12	Cruscotto introduttivo sui ricoveri da Covid-19 nella regione Marche	90
5.13	Colonna calcolata relativa alla variazione giornaliera dei ricoverati in terapia intensiva	91
5.14	Grafico relativo all'andamento dei ricoveri da Covid-19 in Italia	92
5.15	Cruscotto relativo all'andamento dei ricoveri in terapia intensiva per struttura	93
5.16	Cruscotto relativo all'andamento dei ricoveri in terapia semi-intensiva per struttura.....	94
5.17	Cruscotto relativo all'andamento dei ricoverati in reparti post-critici, suddivisi per struttura	94
5.18	Cruscotto relativo all'andamento dei ricoveri in reparti non intensivi, suddivisi per struttura	95
5.19	Cruscotto relativo all'analisi dei ricoveri in terapia intensiva nella regione Marche.....	95
5.20	Misura calcolata relativa all'occupazione di posti letto di terapia intensiva	96
5.21	Cruscotto introduttivo sulle persone in quarantena domiciliare nella regione Marche.....	97

X **Elenco delle figure**

5.22	Cruscotto relativo all'andamento del numero delle persone in quarantena domiciliare casi nelle province marchigiane	99
5.23	Cruscotto relativo agli operatori sanitari in quarantena domiciliare ..	100
5.24	Misura calcolata relativa al rapporto tra gli operatori sanitari in quarantena domiciliare e il totale di casi in isolamento	100
5.25	Grafico relativo a operatori sanitari in isolamento domiciliare, pazienti ricoverati e decessi totali	101
6.1	Le caratteristiche dell'analisi predittiva	104
6.2	Importazione dell'oggetto visivo relativo al forecasting mediante ARIMA	105
6.3	Esempio di serie temporale	106
6.4	Il menù "Previsione" in Power BI	108
6.5	Esempi di previsioni in Power BI	108
6.6	Cruscotto di tipo predittivo relativo ai decessi	109
6.7	Impostazioni relativa al forecasting con ARIMA	110
6.8	Cruscotto di tipo predittivo relativo ai malati totali	110
6.9	Cruscotto di tipo predittivo relativo ai malati attivi	111
6.10	Cruscotto di tipo predittivo relativo ai malati ricoverati in terapia intensiva	112
7.1	La Business Intelligence e le sue fasi	115
7.2	La struttura di un file in formato JSON	117

Elenco dei listati

3.1	Struttura del file json che raccoglie i tre report giornalieri della Regione Marche	47
-----	---	----

Introduzione

Il mondo digitalizzato in cui viviamo richiede la capacità di gestire e manipolare enormi moli di dati, al fine di ricavare conclusioni e informazioni preziose, a seconda dell'ambito di interesse. La Data Analytics assume, perciò, ai giorni d'oggi, un ruolo fondamentale in tutti gli aspetti della vita quotidiana. Grazie, soprattutto, alla crescente consapevolezza dell'importanza dei dati e delle informazioni, la Data Analytics sta diventando sempre più un aspetto cruciale in ambito finanziario ed economico, ma non solo. Essa consente, infatti, attraverso l'utilizzo di tecnologie e strumenti sempre più automatizzati ed efficienti, di eseguire un processo di ispezione, pulizia, trasformazione e modellazione dei dati, al fine di trarre informazioni significative che siano in grado di supportare le decisioni aziendali. In particolare, attraverso processi di analisi descrittiva, diagnostica e predittiva, la Data Analytics permette di dare un vero e proprio significato ai dati, trasformando una possibile fonte di conoscenza inutilizzata in una base di partenza per acquisire importanti informazioni strategiche o, anche, per supportare le attività di decision-making, fondamentali all'interno di aziende e organizzazioni.

L'avvento dei Big Data, inoltre, ha portato ad una vera e propria rivoluzione nel mondo odierno, causando una crescente necessità di collegare fra di loro una vasta quantità di informazioni per fornire un approccio visuale ai dati, mediante veri e propri modelli di visualizzazione che ne semplificano la lettura e la comprensione. A tal proposito entra in gioco, anche, la cosiddetta Business Intelligence, la quale sempre in ambito economico e finanziario, si propone di raccogliere dati aziendali di diversa natura, per analizzarli e trarne decisioni strategiche attraverso un insieme di processi e strumenti appositi. Sappiamo, infatti, come ogni giorno le aziende e, in generale, le organizzazioni, producono una quantità davvero incredibile di dati che spesso, però, sono inutili per la società, se non per lo scopo originario per cui sono stati creati. Tuttavia, con il progresso degli strumenti di Business Intelligence, le organizzazioni riescono ad utilizzare i dati raccolti per prendere decisioni strategiche in maniera sempre più veloce ed efficace. Gli strumenti di Business Intelligence, in particolare, si fondono con i processi della Data Analytics, e permettono di visualizzare gli aspetti più rilevanti dei dati di interesse, mediante la realizzazione di dashboard e report interattivi.

Il mondo dei Big Data e della Data Analytics si estende non solo ai mercati di business e agli ambiti aziendali, ma in qualunque settore e ambiente in cui sono

presenti grandi quantità di dati e si ha la necessità di convertirli in vera e propria informazione.

La versatilità nell'utilizzo della Data Analytics risulta, dunque, fondamentale in molti aspetti del nostro mondo. Questa potenzialità ha permesso, nel nostro caso di studio, di applicare i processi della Data Analytics per analizzare la pandemia da Covid-19. Tale pandemia, sviluppatasi a partire dalla fine del 2019, è arrivata, nel giro di pochi mesi, a colpire il mondo intero, provocando importanti danni sociali, economici e, soprattutto, in termini di vite umane. Lo sviluppo della pandemia ha letteralmente bloccato e mandato in confusione i sistemi sanitari in molti paesi del mondo, tra cui anche l'Italia. Fin dal suo nascere, le varie organizzazioni mondiali, ma anche le stesse nazioni, si sono impegnate nel diffondere i dati relativi alla diffusione del virus.

Come precedentemente illustrato, tali dati, se non raccolti ed organizzati a dovere, rappresentano più un problema che una risorsa. In un contesto come quello di una pandemia, inoltre, raccogliere ed analizzare dati diventa un aspetto fondamentale per capire l'andamento della diffusione del virus e per provare, anche, a fare previsioni su ciò che potrebbe accadere in futuro.

Nel corso dei mesi dello sviluppo del virus, infatti, si è compresa sempre di più l'importanza del distanziamento sociale e dell'individuazione dei focolai, con quest'ultimi che possono essere considerati come un vero e proprio nemico nella lotta al virus. Ai fini dell'individuazione dei focolai, risulta fondamentale un'intensa attività di analisi dei tamponi e degli spostamenti delle persone. Tutto ciò contribuisce, naturalmente, alla produzione di una grande quantità di dati, la cui elaborazione, risulta, perciò, di grande importanza per la lotta al virus stesso. Inoltre, la raccolta dei dati risulta altresì fondamentale agli enti governativi e sanitari, per tenere sotto controllo la pandemia, nonché per stabilire se è necessario o meno attuare misure di contenimento in caso di una eventuale espansione del virus a velocità molto alte.

Questa tesi si colloca proprio in tale contesto e si propone di realizzare un'analisi dei dati relativi alla diffusione del Covid-19 all'interno della Regione Marche. Durante il periodo di maggior diffusione del virus in Italia, sono state realizzati molti grafici e dashboard relativi al suo andamento in tutto il territorio italiano. In realtà, anche le singole regioni si sono impegnate nella diffusione dei dati, ma pochi grafici e poche analisi si sono realizzate a livello regionale. In particolare, ad eccezione della Regione Lombardia, la più studiata, nonché la più colpita dalla pandemia, pochi progetti sono scesi a livello regionale e poche analisi sono state effettuate sui dati regionali.

La Regione Marche, a differenza di altre regioni, ha cercato, fin dai primi sviluppi del virus, di mantenere massima chiarezza e trasparenza sui dati riguardanti i tamponi effettuati all'interno della regione, i ricoveri nelle strutture ospedaliere regionali ed i decessi avvenuti nel territorio regionale (anche da parte di residenti fuori regione).

In particolare, attraverso il GORES (Gruppo Operativo Regionale Emergenza Sanitaria), tutti i dati relativi agli aspetti del virus sopra riportati sono stati pubblicati sul sito della regione stessa, mediante tre report giornalieri (*blu*, *giallo*, *arancio*).

Questi dati rappresentano la base del nostro studio, che si propone come obiettivo quello di fornire una chiara ed ampia panoramica riguardante gli aspetti più

importanti della diffusione del virus all'interno della Regione Marche, nonché di effettuare una attività di forecasting per capire le possibili evoluzioni future della pandemia all'interno del territorio regionale.

Nella fase preliminare dello studio è stata acquisita conoscenza in merito alla pandemia da Covid-19, con particolare riferimento alle varie tappe del suo sviluppo nel territorio italiano e nelle Marche. Sempre in questa fase, è stato analizzato il tool Microsoft Power BI, uno dei leader del mercato delle piattaforme di Data Analytics e Business Intelligence.

Successivamente a questa fase preliminare, 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. A tale proposito, per arrivare alla creazione di un vero e proprio modello dei dati all'interno del tool in questione, sono state svolte delle attività di ETL sui dati utilizzando le funzionalità di Power BI. Durante questa attività, si è cercato di ripulire i dati, al fine di renderli facilmente utilizzabili dalle visualizzazioni grafiche disponibili in Power BI.

Una volta ottenuto il modello dei dati, è stata svolta un'intensa attività di analisi descrittiva, la quale si occupa, proprio, 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, durante questa fase, è stata realizzata una serie di cruscotti, all'interno di Power BI, al fine di fornire una rappresentazione grafica, semplice ed intuitiva, dei dati a nostra disposizione. Sono stati effettuati confronti temporali e analisi a livello provinciale andando, anche, nel dettaglio dei dati relativi alle singole province delle Marche. I cruscotti realizzati in questa fase sono relativi agli aspetti fondamentali dello sviluppo della pandemia all'interno della regione; essi riassumono, infatti, i dati relativi ai decessi riconducibili al Covid-19 avvenuti all'interno del territorio regionale, ai tamponi effettuati all'interno della regione, ai casi positivi individuati, alla situazione dei ricoverati nelle varie strutture ospedaliere della regione, ai dimessi e ai guariti. Inoltre, sono stati realizzati anche dei cruscotti per l'analisi dei dati relativi al solo personale sanitario che, essendo stato in prima linea durante l'espandersi della pandemia, è stato particolarmente colpito dal virus. In realtà, questa fase di analisi è stata articolata in due sottoparti: nella prima sono stati analizzati i decessi avvenuti all'interno della regione, mentre nella seconda sono stati affrontati tutti gli altri aspetti sopra descritti. Si è deciso di trattare a parte i decessi perché i dati a nostra disposizione, provenienti dal GORES, contengono, anche, le informazioni relative alle singole persone decedute, e ciò ha permesso di analizzare nel dettaglio le caratteristiche di tali persone, quali età, sesso e patologie pregresse.

Terminata la fase di analisi descrittiva, il nostro studio è proseguito con una fase di analisi predittiva. Quest'ultima si occupa di esaminare i dati a disposizione con l'obiettivo di effettuare previsioni su ciò che accadrà nel futuro. Anche questa fase è stata svolta direttamente all'interno di Power BI, sfruttando la sua integrazione con il linguaggio R e la possibilità di importare oggetti visivi aggiuntivi rispetto a quelli predefiniti nel tool. Si è deciso di realizzare un'attività di forecasting, mediante la creazione di grafici basati sul modello predittivo ARIMA. A tal fine, è stato importato in Power BI un grafico basato su uno script di R e sul modello ARIMA. Il forecasting consente di effettuare una vera e propria predizione cercando di ipotizzare l'andamento delle variabili entro determinati livelli di confidenza. Nel nostro

caso di studio, esso è stato applicato ai dati relativi a decessi, ai casi positivi ed ai ricoveri.

Nell'ultima fase del nostro studio, si è cercato di tirare le somme di tutto il lavoro svolto, evidenziando la sua importanza nel futuro come base di partenza per ulteriori studi e analisi nell'ambito delle diffusione di pandemie o, anche, epidemie.

La presente tesi è strutturata come di seguito specificato:

- Nel Capitolo 1 verrà introdotta la pandemia da Covid-19, con particolare riferimento alle tappe della sua diffusione nel territorio italiano e nella Regione Marche. Successivamente, si analizzerà l'importanza del trattamento dei dati nell'analisi di una 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 verranno descritti nel dettaglio i dati utilizzati nella presenti tesi. Successivamente, verranno illustrate le attività di ETL effettuate al fine di importare i dati all'interno di Power BI Desktop in un formato pulito e adatto per i nostri studi.
- Nel Capitolo 4 verrà trattata la prima fase di un'analisi di dati, ovvero l'analisi descrittiva. All'interno del capitolo verranno realizzati cruscotti e grafici per illustrare un particolare aspetto dei dati a nostra disposizione, ovvero quello relativo ai decessi da Covid-19 all'interno della Regione Marche.
- Nel Capitolo 5 si proseguirà la fase di analisi descrittiva, mediante la realizzazione di ulteriori dashboard, riguardanti i dati relativi a casi positivi, ricoveri e guariti.
- Nel Capitolo 6 verrà affrontata la fase di analisi predittiva. In particolare, verranno svolte attività di forecasting mediante l'utilizzo di grafici basati sul modello ARIMA.
- Nel Capitolo 7 verranno analizzate le lezioni apprese durante lo sviluppo del presente elaborato.
- Nel Capitolo 8 verranno tratte delle conclusioni in merito al lavoro svolto e si analizzeranno alcuni possibili sviluppi futuri.

Pandemia da Covid-19

In questo capitolo introduciamo la pandemia da Covid-19, che verrà ricordata nella storia come l'evento più grave che si sia verificato dal dopoguerra ad oggi, e con il più profondo impatto sulle vite di tutti. In realtà, si parla diffusamente di Coronavirus da circa 20 anni, da quando, nel 2003, si verificò la prima epidemia di casi di una severa forma di polmonite causata da questa famiglia di virus. Oggi, nel 2020, la stessa specie virale si è resa protagonista di una vera e propria pandemia, ossia "una malattia epidemica che si espande rapidamente diffondendosi in più aree geografiche del mondo, unitamente ad un'elevata mortalità".

1.1 Il virus e la pandemia

Prima di approfondire nel dettaglio la malattia, è importante fare chiarezza sui termini in uso:

- *Coronavirus*: il termine indica un genere di virus isolati per la prima volta nel 1960 e noti per essere causa di malattie di gravità variabile in uccelli e mammiferi, tra cui l'uomo. Generalmente, i virus responsabili d'infezione in una specie animale non sono in grado di infettare anche specie diverse, ma, nella storia recente, si sono verificate alcune eccezioni. Il virus responsabile dell'attuale epidemia è stato chiamato Sars-CoV-2 (Sindrome respiratoria acuta grave Coronavirus 2).
- *Covid-19*: questa sigla è il nome internazionale ufficiale della malattia dato dall'Organizzazione Mondiale della Sanità (nel seguito OMS) (Figura 1.1); il termine è un acronimo di COrona VIRus Disease 2019 (anno di scoperta).

Si tratta del settimo Coronavirus riconosciuto in grado di infettare esseri umani. Il virus è stato sequenziato genomicamente dopo un test di acido nucleico effettuato su un campione prelevato da un paziente colpito da una polmonite, di cui non si conosceva la causa, all'inizio della pandemia del 2019-2020 a Wuhan. Questo virus, al momento ancora in fase di studio, è paragonato alla Sars e ha un tasso di mortalità ancora incerto a causa della fase transitoria di contagio ancora in corso. Non sono ancora ben chiare molte sue caratteristiche e, sebbene sia stata accertata la sua capacità di trasmettersi da una persona a un'altra, permangono ancora delle incertezze sulle esatte modalità di trasmissione e sulla sua patogenicità.

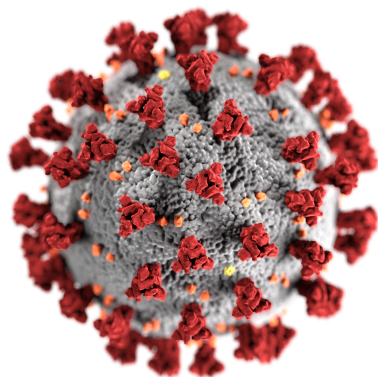


Figura 1.1. Modello 3D del Covid-19

1.1.1 Origine del virus

Mentre a fine Dicembre 2019 e all’inizio di Gennaio 2020 pensavamo ai buoni propositi per l’anno nuovo, ed eravamo del tutto ignari dell’emergenza sanitaria che si sarebbe creata, un nuovo virus altamente contagioso e completamente sconosciuto al nostro sistema immunitario aveva iniziato a circolare in una regione remota del globo. Non avremmo mai pensato, all’epoca, che questo virus apparentemente così lontano avrebbe potuto diffondersi e causare tanti problemi a livello individuale e collettivo, per la salute, per i sistemi sanitari ed economici. Ma in poco più di due mesi lo scenario globale è cambiato radicalmente e abbiamo dovuto adattarci e far fronte alle nuove esigenze.

Il “paziente 0” della pandemia di Covid-19 si è infettato tra la metà di settembre e l’inizio di dicembre dell’anno 2019. A datare il primo caso umano di Coronavirus Sars-CoV-2 è uno studio pubblicato su “Pnas” da ricercatori dell’università di Cambridge nel Regno Unito e da colleghi tedeschi che, analizzando genomi virali completi sequenziati nel mondo dopo essere stati isolati da malati Covid, ha ricostruito i primi passi dell’epidemia. Anche secondo le ipotesi di uno studio italiano, già a Novembre, e forse anche a Ottobre, il nuovo Coronavirus Sars-CoV-2 aveva iniziato a circolare in Cina, in particolare a Wuhan, la città più popolata della parte orientale, perno per il commercio e gli scambi. All’inizio, però, non si sapeva che si trattava di un nuovo virus, ciò che iniziò ad essere registrato è un certo numero di polmoniti anomale, dalle cause non ascrivibili ad altri patogeni.

La prima data ufficiale in cui inizia la storia del nuovo Coronavirus è il 31 Dicembre 2019, in cui le autorità sanitarie locali dettero notizia di questi casi insoliti. All’inizio di Gennaio 2020 la città aveva riscontrato decine di casi e centinaia di persone erano sotto osservazione. Dalle prime indagini infatti, era emerso che i contagiati erano frequentatori assidui del mercato “Huanan Seafood Wholesale Market” a Wuhan (Figura 1.2), che è stato chiuso dal 1 gennaio 2020; da qui l’ipotesi che il contagio possa essere stato causato da qualche prodotto di origine animale venduto nel mercato, anche se tuttora non si hanno informazioni dettagliate sulla reale causa del virus.



Figura 1.2. Il mercato “Huanan Seafood Wholesale Market” di Wuhan

1.1.2 La pandemia

La diffusione del virus, a partire dai primi casi rilevati in Cina, prosegue con il passare del tempo a grandi velocità, arrivando a colpire quasi tutti i paesi del mondo. Perciò, a causa della costante crescita di casi accertati di Coronavirus a livello globale, l’11 Marzo 2020 Tedros Adhanom Ghebreyesus, direttore generale dell’ Organizzazione Mondiale della Sanità (Figura 1.3) annuncia che Covid-19 “può essere caratterizzato come una situazione pandemica”, dichiarando, di fatto, la pandemia. Ma che cosa si intende precisamente con questo termine? E cosa lo distingue dalla cosiddetta “epidemia”? (Figura 1.4)



Figura 1.3. Logo dell’Organizzazione Mondiale della Sanità (OMS/WHO)

Cos’è una epidemia

L’epidemia è una patologia che si diffonde fino a colpire un gran numero di persone in un territorio più o meno vasto. Si sviluppa con andamento che varia a seconda dei casi e si estingue dopo una durata anche variabile. Quanto alla provenienza, le epidemie possono essere importate e autoctone. Le epidemie si propagano per contagio diretto (malato, convalescente o portatore sano) o indiretto (oggetti infetti, pulviscolo atmosferico, acqua, insetti, etc.).

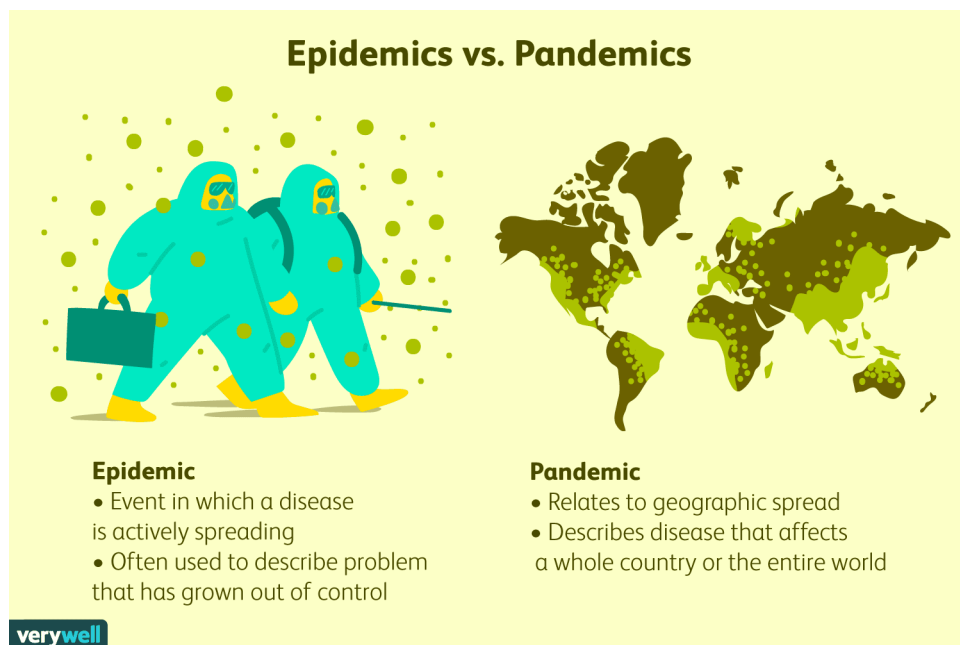


Figura 1.4. Epidemia VS Pandemia

Cos'è una pandemia

La pandemia è, invece, collegata a un agente infettivo che si diffonde in una zona molto più vasta in diverse aree del mondo. Di solito una malattia supera lo stadio epidemico, divenendo pandemica, quando ha un'elevata trasmissibilità nella specie umana e viene a contatto con popolazioni che precedentemente non avevano contratto quell'infezione. Quando si verifica quest'ultima condizione, la pandemia può essere estremamente grave per la mancanza di difese immunitarie adeguate negli individui.

1.2 Tappe della diffusione

Di seguito è riportata la trama della diffusione del Coronavirus sintetizzata nelle principali tappe temporali dell'epidemia di Covid-19, tappe che seguono il momento dell'individuazione del virus in Cina.

Fra il 9 e il 12 Gennaio 2020: l'annuncio del Coronavirus

Il 9 Gennaio 2020 le autorità cinesi avevano dichiarato ai media locali che il patogeno responsabile delle polmoniti anomale è un nuovo ceppo di Coronavirus, della stessa famiglia dei Coronavirus responsabili della Sars e della Mers ma anche di banali raffreddori, e, tuttavia, diverso da tutti questi. L'OMS divulgava la notizia il 10 Gennaio 2020, fornendo tutte le istruzioni del caso (evitare contatto con persone con

sintomi, etc.) e dichiarando che non era raccomandata alcuna restrizione ai viaggi per e dalla Cina. Tutti i casi, ancora molto pochi, erano concentrati a Wuhan e non si conosceva la contagiosità di questo virus. Il 7 Gennaio il virus veniva isolato e, pochi giorni dopo, il 12 Gennaio, veniva sequenziato, e la Cina condivideva la sequenza genetica. Questo è stato il primo passo importante, in termini di ricerca, anche per poter sviluppare e diffondere i test (i kit) diagnostici che serviranno a molti altri paesi.

21 Gennaio 2020: il virus si trasmette fra esseri umani

Il 21 Gennaio 2020 le autorità sanitarie locali e l'OMS annunciavano che il nuovo Coronavirus, passato probabilmente dall'animale all'essere umano (un salto di specie, in gergo tecnico), si trasmette anche da uomo a uomo. Ma ancora gli esperti non sapevano (e tuttora l'argomento è discusso) quanto facilmente questo possa avvenire. Il Ministero della Salute ha iniziato a raccomandare di non andare in Cina, salvo stretta necessità. Nel frattempo, Wuhan diventava una città isolata, e i festeggiamenti per il capodanno cinese venivano annullati lì e in altre città cinesi, come Pechino e Macao. In Italia i casi erano pochissimi e tutti provenienti dalla Cina: a partire dal 29 Gennaio c'erano due turisti cinesi di Wuhan contagiati, ricoverati allo Spallanzani, uno degli ospedali italiani che saranno protagonisti (loro malgrado) della vicenda del Coronavirus. C'era, poi, un ricercatore italiano positivo al virus e proveniente dalla Cina e un diciassettenne, rimasto bloccato a lungo a Wuhan a causa di sintomi simil-influenzali, non positivo al Coronavirus, ma ugualmente tenuto sotto osservazione e ricoverato allo Spallanzani. Tutte queste persone sono guarite e sono state dimesse nel mese di Febbraio, per ultima la paziente cinese della coppia malata, il 26 Febbraio. I contagi fuori dalla Cina erano ancora molto circoscritti e limitati, con focolai per ogni paese di un manipolo di persone.

30 Gennaio 2020: l'OMS dichiara lo stato di emergenza globale

Alla fine di Gennaio il rischio che l'epidemia si diffondesse passava da moderato ad alto e il 27 Gennaio 2020 l'OMS scriveva che era molto alto per la Cina e alto a livello regionale e globale. Tanto che, nella serata del 30 Gennaio 2020, l'OMS dichiarava l'emergenza sanitaria pubblica di interesse internazionale e l'Italia bloccava i voli da e per la Cina, unica in Europa. Ma la situazione in Cina stava già migliorando: pochi giorni dopo, alla data dell'8 Febbraio, l'OMS scriveva che i contagi in Cina si stavano stabilizzando, ovvero che il numero di nuovi casi giornalieri sembrava andare progressivamente calando.

Febbraio 2020: dare un nome alle cose

L'11 Febbraio 2020 è arrivato il nome della nuova malattia causata dal Coronavirus. Il nome, scelto dall'OMS, è Covid-19: "Co" e "vi" per indicare la famiglia dei Coronavirus, "d" per indicare la malattia ("disease" in inglese) e, infine, 19 per sottolineare che è stata scoperta nel 2019. Questo per quanto riguarda la malattia, mentre il virus cambia nome e non si chiama più 2019-nCoV, ma Sars-CoV-2, perché il patogeno è parente del Coronavirus responsabile della Sars (che, però, era molto

più letale, anche se meno contagiosa). All'epidemia di Covid-19 si affianca quella dell'informazione, con notizie non sempre veritiere (molte sono fake news). Tanto che, ai primi di Febbraio, proprio l'OMS parla, per la prima volta, di "infodemia", termine nuovo con cui si indica il sovraccarico di aggiornamenti e news non sempre attendibili (Figura 1.5).



Figura 1.5. “Infondemia” da Coronavirus

21 Febbraio 2020: primi casi in Italia

Venerdì 21 Febbraio 2020 è una data centrale per la vicenda italiana legata al nuovo Coronavirus. In questa data sono emersi diversi casi di Coronavirus nel lodigiano, in Lombardia; si tratta di persone non provenienti dalla Cina, un nuovo focolaio di cui non si conosce ancora l'estensione. Alcuni dei paesi colpiti (Codogno, Castiglione d'Adda, Casalpusterlengo ed altri) sono stati, di fatto, chiusi. Fuori dalla Cina, il numero di contagiati è molto alto in Italia, Iran e Corea del Sud, anche se, per l'OMS, Covid-19 non è ancora pandemia. Tuttavia, fra la fine di Febbraio e i primi giorni di Marzo 2020, dopo l'Italia, anche in altri stati (europei e non solo) vengono rilevati un numero crescente di casi e un'epidemia.

4, 8 e 9 Marzo 2020: le tre date chiave dei provvedimenti in Italia

Il contagio si è diffuso nel nostro paese, soprattutto nel nord, ma inizia anche in altre regioni. Per questo, mercoledì 4 Marzo 2020 il governo ha dato il via libera alla chiusura di scuole e università in tutta Italia fino al 15 marzo. Alla data del 4, stando ai dati della Protezione Civile, i positivi sono circa 2.700 e già c'è qualche caso (decine o qualche unità) in tutte le regioni. Domenica 8 Marzo 2020 arriva il decreto che prevede l'isolamento della Lombardia, in assoluto la regione più colpita, e di altre 14 province, che diventano “zona rossa”.

Infine si arriva ad un'altra importante data per l'Italia: quella di lunedì 9 Marzo 2020. In questa giornata, intorno alle 22, il Presidente del Consiglio Giuseppe Conte annuncia in televisione di aver esteso a tutto il paese le misure già prese per la Lombardia e per le altre 14 province, tanto che tutta l'Italia diventerà “zona protetta”. Le nuove norme sono contenute nel nuovo Dpcm 9 Marzo 2020, entrato

poi in vigore il 10 Marzo. Di fatto la regola è contenuta nell’hashtag “iorestoacasa”: si può uscire solo per comprovate ragioni di necessità, come per fare la spesa, per esigenze lavorative, per l’acquisto di farmaci o per altri motivi di salute.

11 Marzo 2020: l’OMS dichiara la pandemia

Mentre l’Italia si sta muovendo, per prima in Europa, con il plauso dell’OMS, per contenere il contagio, anche a livello globale sta succedendo qualcosa. L’11 marzo 2020 Tedros Adhanom Ghebreyesus, direttore generale dell’OMS, ha annunciato nel briefing da Ginevra sull’epidemia di Coronavirus che Covid-19 “può essere caratterizzato come una situazione pandemica”, dichiarando, di fatto, la pandemia.

La rapida diffusione del Sars-CoV-2 ha destato l’attenzione globale e l’OMS ha dichiarato il Covid-19 un’emergenza pubblica di interesse internazionale (PHEIC). Dichiarare un’emergenza pubblica di interesse internazionale è un appello per la comunità mondiale con lo scopo di attuare uno sforzo coordinato globale per fermare l’epidemia, che richiede un impegno sanitario, politico ed economico. Nonostante ciò, la diffusione mondiale, l’epidemiologia e le manifestazioni cliniche del Covid-19 rimangono largamente non chiare, soprattutto nei bambini (Figura 1.6).

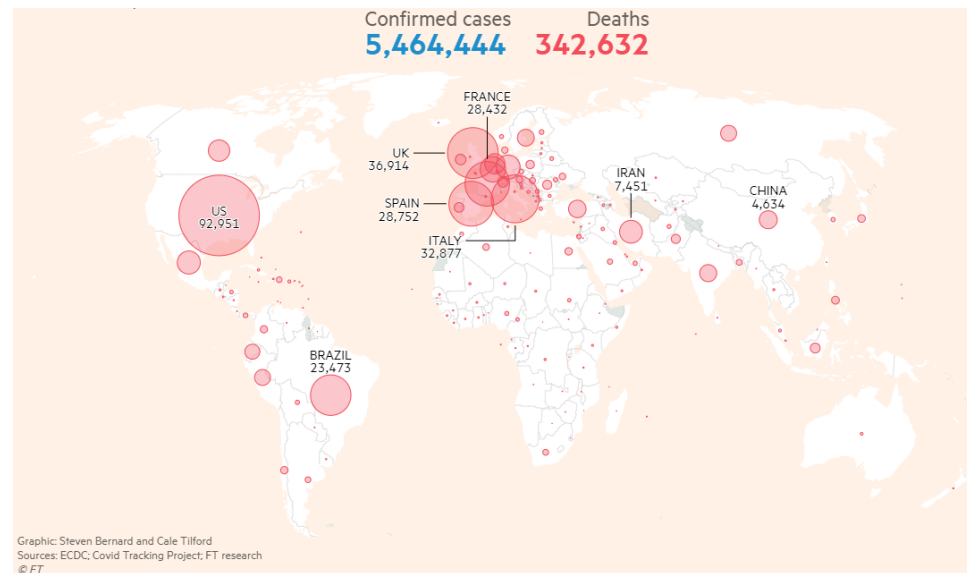


Figura 1.6. Decessi per Covid-19 al 27 Maggio 2020

1.2.1 La situazione italiana

La pandemia di Covid-19 del 2020, in Italia, ha avuto le sue manifestazioni epidemiche iniziali il 31 Gennaio 2020, quando due turisti provenienti dalla Cina sono risultati positivi al virus Sars-CoV-2 a Roma. Un focolaio di infezioni di Covid-19 è

stato successivamente rilevato il, 21 Febbraio 2020, a partire da 16 casi confermati in Lombardia, a Codogno, in provincia di Lodi, aumentati a 60 il giorno successivo, con i primi decessi segnalati negli stessi giorni. Alla data del 27 Maggio 2020 (Figura 1.7) sono stati registrati 231.139 casi positivi, tra cui 147.101 persone dimesse e guarite e 33.072 persone decedute, e sono stati effettuati oltre 3.5 milioni di tamponi per il virus, rendendo l'Italia il sesto paese al mondo per numero di casi totali, dopo gli Stati Uniti d'America, il Brasile, la Russia, la Spagna e il Regno Unito, e il terzo al mondo per numero di decessi.

Regione	AGGIORNAMENTO 27/05/2020 ORE 17.00										
	POSITIVI AL nCoV				DIMESSI/ GUARITI	DECEDUTI	CASI TOTALI	INCREMENTO CASI TOTALI (rispetto al giorno precedente)	TAMPONI	CASI TESTATI	Incremento tamponi
	Ricoverati con sintomi	Terapia intensiva	Isolamento domiciliare	Totale attualmente positivi							
Lombardia	3.626	175	20.236	24.037	47.810	15.954	87.801	+ 384	697.561	413.483	12.503
Piemonte	1.118	68	5.278	6.464	20.095	3.828	30.387	+ 73	298.571	196.790	5.098
Emilia Romagna	459	78	3.461	3.998	19.546	4.083	27.627	+ 16	301.568	186.933	3.714
Veneto	168	9	2.110	2.287	14.931	1.895	19.113	+ 8	616.691	317.690	14.439
Toscana	132	31	1.297	1.460	7.595	1.027	10.082	+ 12	238.667	170.802	3.871
Liguria	205	14	1.050	1.269	6.882	1.438	9.589	+ 39	98.835	55.899	2.157
Lazio	1.062	65	2.361	3.488	3.483	701	7.672	+ 11	242.287	195.001	3.320
Marche	90	13	1.347	1.450	4.272	996	6.718	0	98.535	62.815	1.250
Campania	275	6	865	1.146	3.221	406	4.773	+ 6	185.724	88.874	5.879
Puglia	170	17	1.326	1.513	2.471	495	4.479	+ 10	109.499	74.686	2.626
Trento	19	2	456	477	3.483	462	4.422	+ 7	82.269	46.350	1.120
Sicilia	73	10	1.235	1.318	1.845	272	3.435	+ 5	140.295	122.130	2.613
Friuli V.G.	51	1	304	356	2.568	331	3.255	+ 4	123.063	74.332	1.651
Abruzzo	138	3	725	866	1.969	400	3.235	+ 5	70.473	48.112	1.740
Bolzano	27	5	139	171	2.131	291	2.593	0	62.247	28.809	783
Umbria	13	2	17	32	1.324	75	1.431	0	66.219	46.975	1.206
Sardegna	41	2	172	215	1.010	130	1.355	+ 1	53.294	45.745	1.326
Valle d'Aosta	17	1	9	27	1.011	143	1.181	0	14.459	11.322	197
Calabria	35	1	154	190	872	96	1.158	+ 1	65.631	63.372	1.117
Molise	3	2	161	166	246	22	434	+ 2	13.629	13.087	177
Basilicata	7	0	29	36	336	27	399	0	27.734	27.344	537
TOTALE	7.729	505	42.732	50.966	147.101	33.072	231.139	+ 584	3.607.251	2.290.551	67.324

Figura 1.7. Situazione generale in Italia al 27 Maggio 2020

Per quanto riguarda le misure di contenimento dell'infezione, l'Italia è stata il primo paese europeo a sospendere tutti i voli diretti da e verso la Cina, con una delle misure più drastiche nell'Unione Europea. Una volta scoperto il primo focolaio interno, tra le prime misure adottate vi è stata la messa in quarantena di 11 comuni dell'Italia settentrionale (in Lombardia e in Veneto). Il 23 Febbraio il Consiglio dei ministri emana il decreto legge n. 6, che sancisce la chiusura totale dei comuni con focolai attivi e la sospensione di manifestazioni ed eventi sugli stessi comuni. Nei giorni successivi il premier Giuseppe Conte emana una serie di decreti attuativi (DPCM) in cui le misure di restrizione si fanno progressivamente più ferree ed estese via via all'intero territorio nazionale: DPCM del 25 Febbraio, del 1^o, 4, 8, 11 e 22 Marzo e del 1^o, 10 e 26 Aprile.

Con il DPCM del 16 Maggio 2020 il premier Giuseppe Conte annuncia al Paese l'inizio della fase 2, dal 18 Maggio sino al 14 Giugno 2020. In questa fase riprendono molte attività commerciali al dettaglio, inclusi bar, ristoranti e parrucchieri, e si annullano alcune restrizioni, quali isolamento sociale e spostamento regionale. Viene eliminata definitivamente anche l'autocertificazione che si sarebbe dovuta esibire alle autorità competenti. Lo spostamento interregionale sarà concesso dal 3 Giugno 2020 con un ulteriore smorzamento delle restrizioni.

1.2.2 Il caso delle Marche

Essendo oggetto di studio di questa tesi, riportiamo con maggior dettaglio informazioni aggiuntive sull'andamento del contagio e del virus nella regione Marche. Lo sviluppo del virus nella regione Marche segue quello generale di tutta l'Italia. Di seguito sono riportate alcune delle principali tappe.

Il 25 Febbraio 2020 il Presidente della Regione Marche Luca Ceriscioli dispone, con ordinanza, la chiusura di tutte le scuole e le università nel territorio regionale, nonostante l'assenza di contagi. Tale scelta, motivata dalla necessità di prevenire i contagi provenienti dal vicino comune di Cattolica, è stata criticata dal Governo, che ha espresso l'intenzione di ottenere l'annullamento dell'ordinanza. Il 27 Febbraio 2020 il TAR Marche, in sede cautelare, dispone la sospensione dell'ordinanza regionale, riaprendo, di fatto, le scuole e le università in tutto il territorio delle Marche. Tale sospensione viene motivata sulla circostanza che, al momento dell'emissione, non vi erano contagi nell'area marchigiana; la presenza di persone infette sul territorio regionale, infatti, avrebbe legittimato l'adozione di misure straordinarie.

Preso atto della sospensione, il Presidente della Regione emette, dopo poche ore, una nuova ordinanza identica alla precedente nei contenuti, ma con durata minore (due giorni, fino al 29 febbraio), basata sulla presenza di sei casi di contagi da Coronavirus che, nel frattempo, si erano verificati nelle Marche. Inizia, così, la "storia" del Covid-19 anche nella regione Marche, oggetto di trattazione in questa tesi.

Nelle Marche la preparazione all'allerta per il nuovo Coronavirus passa per le competenze tecniche del GORES, Gruppo Operativo Regionale per le Emergenze Sanitarie, tavolo tecnico che la Regione Marche ha costituito da tempo in modo strutturale per dare risposte tempestive alle emergenze sanitarie, tramite l'interfaccia tra operatori sanitari esperti e Protezione Civile Regionale. La popolazione è costantemente tenuta al corrente, sin dal primo giorno dell'allerta, sia attraverso la stampa, sia attraverso apposite comunicazioni nei punti sensibili.

Il GORES è attivo sul Coronavirus sin dal 27 Gennaio 2020, per strutturare il lavoro in base alle principali esigenze poste dalla specifica situazione, per tracciare e trasmettere le indicazioni operative, definendo le modalità di presa in carico di un potenziale caso sospetto di nuovo Coronavirus, le relative modalità di isolamento, diagnosi e trattamento e l'attuazione delle più idonee misure di sanità pubblica per contenere l'infezione ed evitare eventuali ulteriori casi.

La Regione e la Protezione Civile Regionale sono in continuo contatto con il Comitato Operativo Nazionale istituito presso il dipartimento della Protezione Civile a seguito della dichiarazione dello stato di emergenza del 31 Gennaio 2020.

Il 21 Febbraio il GORES si è riunito per seguire lo sviluppo della situazione a livello nazionale e per verificare lo stato delle attività da mettere in atto.

Inoltre, a differenza di altre regioni, le Marche, anche e soprattutto grazie al GORES, hanno cercato di mantenere fin da subito massima chiarezza e trasparenza sui dati riguardanti tamponi, ricoveri e decessi, scendendo ad un dettaglio di informazione molto più elevato rispetto a quello proposto dalla Protezione Civile (Figura 1.8). Queste informazioni, pubblicate sul sito della regione, vengono rilasciate giornalmente in 3 diversi report (blu, giallo, arancio) che verranno descritti in seguito.

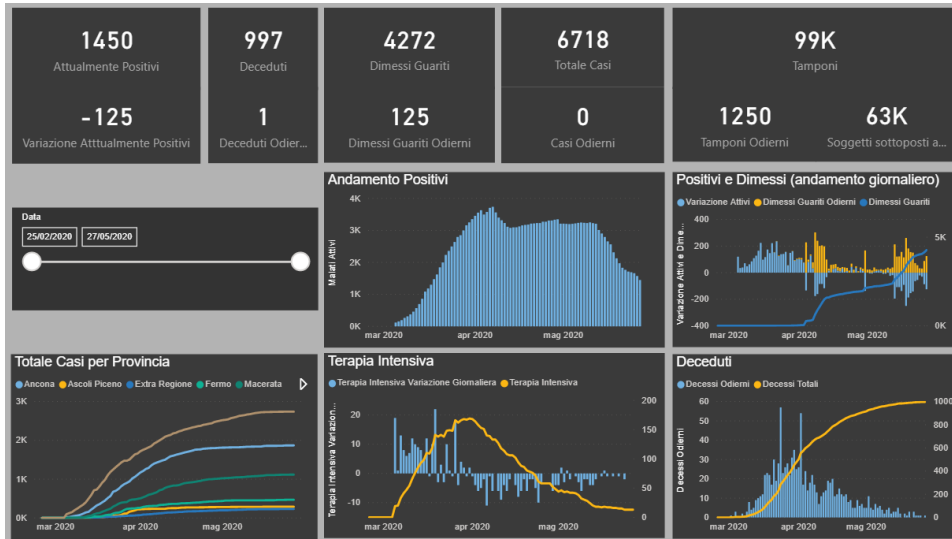


Figura 1.8. Situazione generale nella Regione Marche al 27 Maggio 2020

1.3 L'importanza dei dati nell'analisi della diffusione del virus

Fin dalle origini dell'epidemia, le varie nazioni, regioni ed organizzazioni mondiali si sono impegnate fortemente nella diffusione di dati riguardanti l'espansione e la situazione epidemica nei vari territori. I dati, in una situazione straordinaria come quella di una pandemia, hanno un ruolo fondamentale nelle decisioni politiche, economiche, sanitarie, ma soprattutto per tenere sotto controllo l'espansione ed i danni del virus, e per effettuare previsioni ed analisi dettagliate sulla sua evoluzione.

In principio è stata la Johns Hopkins University a pubblicare i primi dati di rilevanza mondiale in formato open (Figura 1.9). L'ateneo, con sede a Baltimora, nel Maryland, ha ricoperto per settimane il ruolo di principale fonte di informazioni per chi volesse monitorare l'andamento della diffusione del nuovo Coronavirus.

Quando l'infezione è arrivata in Italia, con i primi casi registrati nel lodigiano e nel padovano, è cambiata anche la percezione; da una parte c'era, ovviamente, un maggior interesse da parte del pubblico nel capire i numeri legati all'epidemia, dall'altra una difficoltà a reperirli. Per colmarla, prima che ci pensasse la Protezione Civile, sono venuti in aiuto privati ma anche qualche civic hacker. Le istituzioni hanno compreso l'importanza di fornire non solo i dati giornalieri sull'evoluzione del Coronavirus ma anche di renderli pubblici con licenza open, così da consentirne il riuso per monitorare l'emergenza, fare previsioni o solo corretta informazione. Nella giornata del 7 Marzo 2020, il Dipartimento della Protezione Civile, con un comunicato stampa, ha annunciato non solo il rilascio di una dashboard (Figura 1.10) per il monitoraggio dell'emergenza Coronavirus, ma che tutti i dati sarebbero stati resi pubblici con licenza open all'interno di un apposito repository Github.

Dal 28 Febbraio 2020 anche l'Istituto Superiore di Sanità (ISS) coordina un sistema di sorveglianza che integra a livello individuale i dati microbiologici ed epidemio-

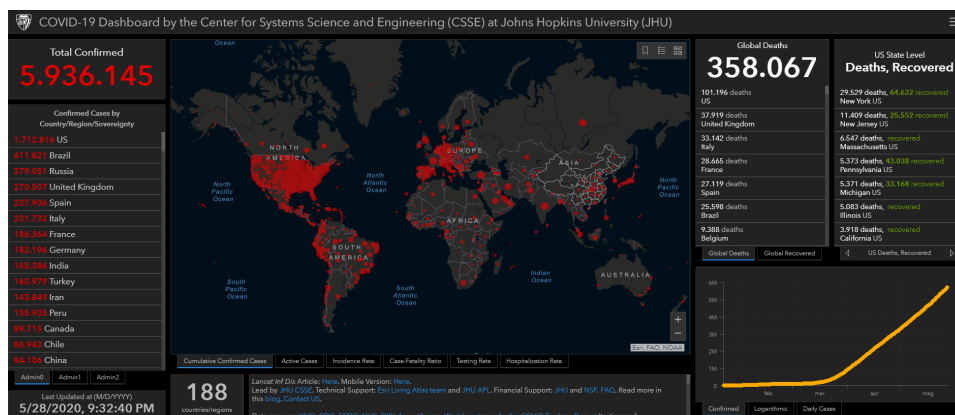


Figura 1.9. Dashboard Johns Hopkins University

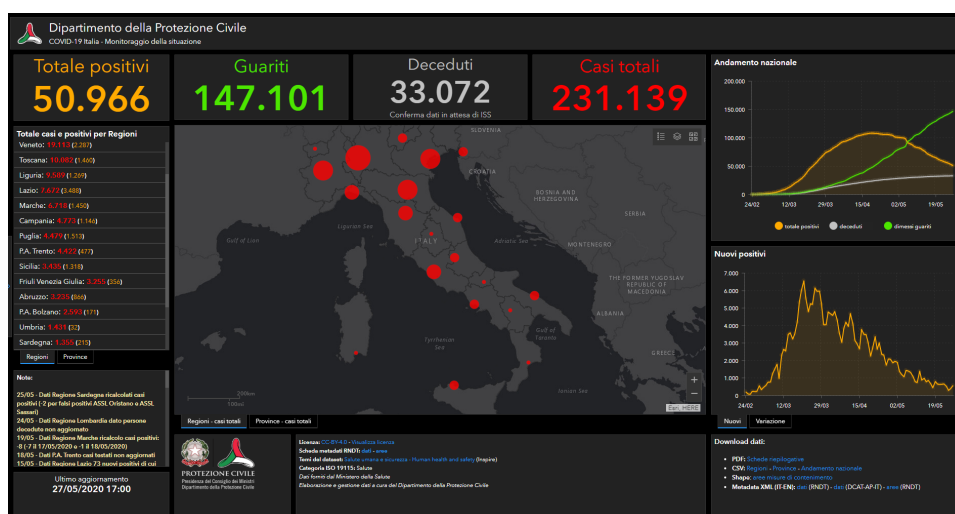


Figura 1.10. Dashboard Protezione Civile

logici forniti dalle Regioni e Province Autonome (PA) e dal Laboratorio nazionale di riferimento per Sars-CoV-2 dell'ISS. I risultati di questa sorveglianza vengono pubblicati regolarmente. Ogni giorno viene prodotta un'infografica dedicata (disponibile anche in inglese) che riporta, con grafici, mappe e tabelle, una descrizione della diffusione nel tempo e nello spazio dell'epidemia di Covid-19 in Italia, nonché una descrizione delle caratteristiche delle persone affette. Anche le singole regioni, ovviamente, fin dagli inizi dell'espansione del virus si sono impegnate, anche in modalità diverse, a diffondere i dati di loro pertinenza.

Tutti questi aspetti hanno fatto in modo che i dati potessero essere riutati da giornalisti, data scientist, centri di ricerca, professionisti in Italia e all'estero per monitorare l'emergenza, fare previsioni o, più semplicemente, fare corretta informazione. All'inizio dell'emergenza, invece, il Dipartimento della Protezione Civile aveva cominciato a rendere disponibili a fine giornata alcuni report riassuntivi sul

numero dei contagiati, i decessi, i guariti. Tutti rigorosamente sotto forma di file PDF. Ora, se c'è una cosa importante che 10 anni di open data sono riusciti certamente a fare, è quella di aver contribuito a creare e a far crescere la conoscenza e la cultura del dato in tanta parte della popolazione. Sono sempre di più, ad esempio, giornalisti che non si limitano soltanto a riportare nei loro articoli dati letti, ad esempio nei comunicati stampa, ma realizzano con quei dati visualizzazioni interattive ed infografiche che aggiungono nuova informazione e valore ai propri pezzi. Quando sono disponibili, i dati vengono poi utilizzati, ad esempio, da studenti, dai centri di ricerca, da professionisti, da chi fa analisi sui dati. E poiché i dati non hanno confini, quando sono disponibili possono essere utilizzati da chiunque ne intraveda un possibile utilizzo, sia all'interno dei confini nazionali che all'estero. Per fare tutto ciò diventa, però, necessario che i dati siano disponibili in formato aperto e rilasciati sotto una licenza d'uso open. Questo vuol dire non solo dare la possibilità a chiunque di poter utilizzare i dati, ma anche di poterlo fare avvalendosi di tool automatici, programmi software e algoritmi.

Inoltre, per fare analisi accurate, si necessita di dati in tempo reale, aggiornati giornalmente, ben disaggregati geograficamente e per caratteristiche demografiche. Per esempio, avere una stima accurata del numero di infetti e del loro tasso di crescita nel paese implica la presenza di dati campionari e dati sulla distribuzione dei test per età e per stato di salute. Senza una stima del numero di infetti e del loro tasso di crescita siamo in totale balia dell'epidemia e non possiamo nemmeno iniziare a valutare l'efficacia dei vari interventi di contenimento. E senza stima degli infetti, non abbiamo una stima del tasso di letalità nel paese, dove gli infetti compaiono al denominatore.

Ma dati accurati sono necessari non solo per studiare l'evoluzione dell'epidemia e l'efficacia degli interventi già messi in atto, ma anche per pianificare le successive fasi di interventi. Ogni analisi epidemiologica esistente, su dati da Cina, Corea del Sud, Singapore, ma anche da Vo' e dal Veneto, suggerisce l'efficacia di una politica di test di massa ed isolamento di coloro che risultino positivi (e dei loro contatti). Questa strategia richiede un enorme sforzo di raccolta ed analisi di dati in tempo reale. Essa richiede la geolocalizzazione di milioni di persone (positivi, lavoratori in servizi essenziali, anziani bisognosi di aiuto), la tracciatura dei movimenti di molti di questi, la comunicazione tempestiva di informazioni precise ed accurate alla popolazione.

La battaglia contro questa epidemia non si combatte solo negli ospedali e nei laboratori, dove si studiano vaccini e cure, ma anche con accurate analisi epidemiologiche e statistiche, giorno per giorno.

1.3.1 La corretta interpretazione di dati e statistiche

Mai come in questi giorni l'opinione pubblica italiana si è interessata di informazione medica, di numeri e statistiche, di procedure di accesso alle cure, di diagnosi, terapie e prognosi. Ci vengono quotidianamente riversate enormi quantità di dati, di stime, di valutazioni, di percentuali, di frequenze relative, assieme a pareri e opinioni di esperti, a volte concordi, non di rado in contrasto tra di loro. Questa attenzione è certamente un bene ed è un segno di quanto seriamente venga percepita e vissuta questa inedita condizione nazionale causata dal Coronavirus.

Tuttavia, proprio per via della serietà della situazione, è necessario fare un passo in avanti e chiedersi quanta parte di questa enorme mole di dati e notizie veramente si trasforma in informazione. Il numero dei nuovi contagiati, per esempio, rappresenta un dato, ma diventa informazione rilevante solo se messo a confronto con il dato dei giorni precedenti, perché solo in questo modo possiamo capire se l'epidemia è in una fase espansiva oppure si sta riducendo. Anche il numero totale dei contagiati può dirci poco se non accompagnato dal numero di coloro che sono guariti e, purtroppo, di quelli che non ce l'hanno fatta. In questo modo il dato può trasformarsi in informazione, in materiale per ragionare ed agire in maniera consapevole ed efficace. Non sempre, quindi, una grande disponibilità di dati equivale ad una opinione pubblica informata.

Quando, poi, i dati sono stati contestualizzati, rimane sempre il problema della loro corretta comprensione. Soprattutto coi numeri questo rappresenta un problema tutt'altro che trascurabile. La nostra visione del mondo, il modo in cui processiamo ed utilizziamo i dati che attingiamo dalla realtà che ci circonda sono fortemente influenzati dal modo in cui funzionano i nostri schemi cognitivi. Se è vero che tutti i numeri sono uguali, non è altrettanto vero che le rappresentazioni alternative che degli stessi numeri si possono dare siano neutrali. Uno stesso fenomeno può essere, per esempio, rappresentato con una percentuale, oppure sotto forma di probabilità o come una frequenza naturale. Il dato è lo stesso, ma l'informazione che siamo indotti a ricavarne varia e, spesso, in modo significativo. Il punto è, proprio, che lo stesso dato può veicolare informazioni differenti in relazione al modo in cui esso viene presentato. In sostanza quello che si vuole dire è che la disponibilità di dati pubblici e aperti è una prerogativa delle democrazie più avanzate, ma non basta fornire i dati: occorre agevolare il più possibile il passaggio dal dato all'informazione. In momenti di crisi come quello che stiamo vivendo in queste settimane avere dati affidabili è indispensabile, ma lo è ancora di più consentirne una corretta comprensione. La corretta interpretazione di queste informazioni può, per esempio, aiutare la popolazione a coordinarsi verso comportamenti virtuosi che producono, poi, benefici diffusi. La giusta informazione può, inoltre, funzionare come feedback verso i comportamenti individuali. Sapere che la proporzione dei contagiati, dopo l'attivazione delle misure di contenimento, sta scendendo, può giustificare e rinforzare la scelta che ciascuno di noi ha fatto di stare a casa, nonostante le innegabili difficoltà.

È indispensabile, quindi, non solo che la Protezione civile renda tempestivamente disponibili i dati che ha a disposizione, ma che questi vengano presentati ai non esperti, ai cittadini e, soprattutto, alla stampa, con modalità immediatamente informative: non solo dati, dunque, ma vera informazione.

Power BI

In questo capitolo introduciamo il tool Power BI, un'applicazione di Business Intelligence sviluppata da Microsoft. Si tratta di un software dotato di funzionalità avanzate per la creazione di report. Il suo obiettivo è, infatti, quello di supportare la Business Intelligence attraverso visualizzazioni interattive e con un'interfaccia semplice, in modo da dare all'utente la possibilità di creare report e dashboard.

2.1 Architettura e descrizione dell'applicazione

Power BI (Figura 2.1) è una raccolta di servizi software, app e connettori che interagiscono per trasformare le origini dei dati non correlate in un insieme di informazioni coerenti, visivamente accattivanti e interattive. I dati potrebbero essere, ad esempio, un foglio di calcolo Excel, una raccolta di data warehouse basata sul cloud, o anche ibridi locali. Power BI consente di connettersi facilmente alle origini dati, visualizzare e scoprire le informazioni importanti e condividerle con tutti gli utenti o con quelli necessari.



Figura 2.1. Logo di Power BI

Power BI è, inoltre, un software di recente sviluppo che, grazie ai grossi investimenti da parte di Microsoft, è riuscito ad imporsi in breve tempo sui già presenti leader del settore, quali QlikSense e Tableau.

Una delle peculiarità che mette Power BI al pari dei suoi competitor è, sicuramente, quella di poter modellare i dati a proprio piacimento, grazie ad una lunga collezione di diagrammi messi a disposizione dal tool, o che, eventualmente, è possibile importare, come vedremo successivamente. La cosa che colpisce di questo software è, soprattutto, l'integrazione di un editor di query molto potente e il supporto al linguaggio R (Figura 2.2).

R è un software statistico, disponibile gratuitamente sotto i vincoli della GPL (General Public Licence). Si tratta di uno dei linguaggi di programmazione maggiormente usati da statistici, data scientist e business analyst; esso è strutturato come un insieme di macro, librerie e oggetti, che possono essere utilizzati per la gestione, l'analisi dei dati e la produzione di grafici.



Figura 2.2. Logo del linguaggio R

Grazie agli aggiornamenti rilasciati con cadenza mensile e, quindi, alla sua predisposizione a migliorarsi, nel 2020, Power BI si è rivelato un vero e proprio leader nel Magic Quadrant di Gartner relativo ai tool di analisi dei dati (Figura 2.3).



Figura 2.3. Il Magic Quadrant di Gartner relativo ai tool di analisi dei dati

Power BI può essere considerato come un vero e proprio ecosistema, essendo il software disponibile in diverse varianti (Figura 2.4).



Figura 2.4. L'ecosistema di Power BI

Le componenti principali di questo ecosistema sono:

- **Power BI Desktop:** l'applicazione desktop che consente la creazione, la modifica, la personalizzazione e la visualizzazione di fogli di lavoro;
- **Power BI Service:** il SaaS basato sul servizio online che permette la visualizzazione e la condivisione delle dashboard aggiornate in tempo reale;
- **Power BI Mobile:** le applicazioni per dispositivi Android e iOS che permettono di monitorare e accedere ai dati aziendali creati dalla componente Desktop da qualsiasi dispositivo mobile.

Il modo di usare Power BI e le sue componenti può variare in base al ruolo ricoperto all'interno di un progetto o di un team, cioè può capitare che i singoli elementi di Power BI vengano usati in momenti diversi a seconda degli obiettivi da raggiungere o del ruolo ricoperto in un determinato progetto. Ad esempio, può capitare che un utente usi principalmente il servizio Power BI per visualizzare report e dashboard, e che i suoi colleghi addetti a calcoli e report aziendali usino Power BI Desktop per creare i report e pubblicarli nel servizio Power BI, dove l'utente iniziale può visualizzarli. Un altro collega che si occupa ad esempio di vendite potrebbe sfruttare in misura maggiore l'app per telefoni di Power BI, per monitorare l'avanzamento delle quote di vendita e accedere ai dettagli sui nuovi clienti potenziali.

Tutte le parti di Power BI sono a disposizione ed è questo il motivo per cui tale strumento è così flessibile ed efficace.

Il nostro caso di studio sarà incentrato su Power BI Desktop (Figura 2.5), grazie al quale possiamo importare tantissime tipologie di sorgenti di dati per estrapolare quelli di nostro interesse, elaborarli ed estrarre da essi informazioni e conoscenza.

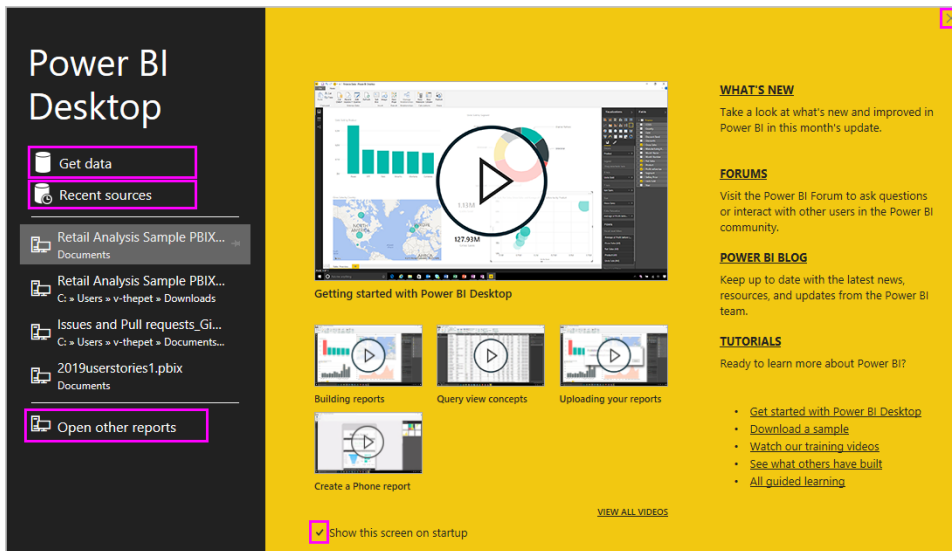


Figura 2.5. Schermata di avvio di Power BI Desktop

Power BI Desktop e Power BI Service: differenze e analogie

Le caratteristiche e le funzionalità di Power BI Desktop e Power BI Service sono piuttosto simili tra di loro e potrebbero essere facilmente confuse. Per tale motivo, cerchiamo di descrivere in maniera più approfondita tali componenti, in modo tale da mettere in evidenza eventuali similitudini e discrepanze.

- 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, in cui è possibile connettersi a molte sorgenti di dati diverse e combinare i dati (attività spesso denominata modellazione) in un modello di dati. 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.

La maggior parte di coloro che si occupano della progettazione di report in progetti di Business Intelligence usa Power BI Desktop per creare i report, e quindi il servizio Power BI per distribuirli ad altri. Le differenze e i punti comuni dei due componenti dell'ecosistema di Power BI sono ben descritte nel diagramma di Venn riportato in Figura 2.6 in cui, in particolare, l'area al centro indica le funzionalità disponibili in entrambe le soluzioni. I due lati del diagramma di Venn, invece, illustrano le funzionalità specifiche dell'applicazione e del servizio.

2.1.1 Il flusso di lavoro

Un flusso di lavoro comune in Power BI inizia con la connessione alle origini dati e la creazione di un report in Power BI Desktop (Figure 2.7). Il report viene, suc-



Figura 2.6. Power BI Desktop e Power BI Service

cessivamente, pubblicato da Power BI Desktop nel servizio Power BI e condiviso, in modo che gli utenti finali del servizio Power BI e dei dispositivi mobili possano visualizzare e interagire con il report. Questo flusso di lavoro è comune e illustra le interazioni e la complementarietà dei tre elementi principali di Power BI. Di seguito riportiamo il dettaglio di questo flusso di azioni:

- connettersi ai dati;
- trasformare e pulire i dati per creare un modello di dati;
- 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.

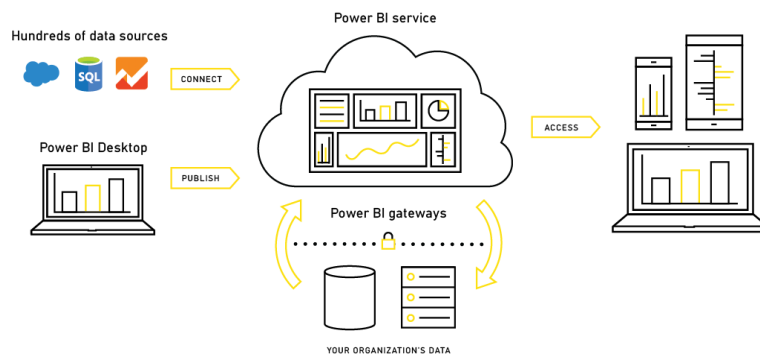


Figura 2.7. Componenti principali del flusso di lavoro in Power BI

2.2 Connessione e importazione dei dati

Concentriamoci ora in dettaglio su Power BI Desktop e sulle varie fasi del flusso di azioni appena descritto. Come illustrato in precedenza, il primo passaggio per iniziare a lavorare con Power BI consiste nel connettersi ai dati. Per fare ciò, è necessario selezionare semplicemente la barra multifunzione “Home” e, quindi, la voce “Scarica Dati”. La Figura 2.8 illustra la finestra “Recupera dati” che viene aperta a seguito di ciò; dalla figura si evince come Power BI sia in grado interfacciarsi a diversi tipi di sistemi tra cui:

- 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.

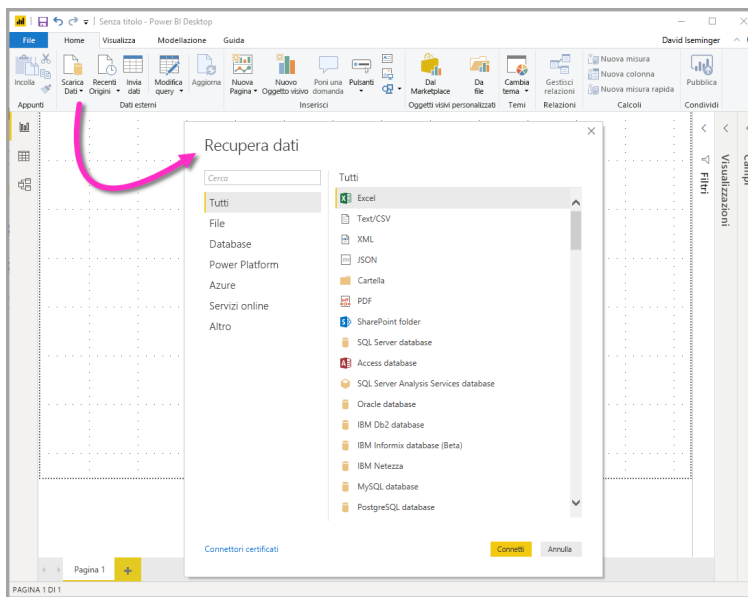


Figura 2.8. Finestra di recupero dei dati Power in BI Desktop

2.2.1 Editor di query

Power BI Desktop include, anche, la componente Editor di query, uno strumento che si rivela fondamentale nella fase di manipolazione dei dati che, grazie a questo editor, è resa indubbiamente più efficace. L’Editor di query viene aperto in una finestra separata (Figura 2.9); esso consente di creare query e trasformare i dati, nonché di caricare il modello di dati ottimizzato in Power BI Desktop.

L’Editor di query permette, cioè, di realizzare operazioni di ETL sui dati: attraverso tali operazioni, ad esempio, si possono rimuovere colonne o combinare i

dati da più origini. È possibile partire da un blocco di dati di grandi dimensioni e rimuovere o aggiungere porzioni di dati in base alle necessità, finché i dati non assumono esattamente la forma desiderata.

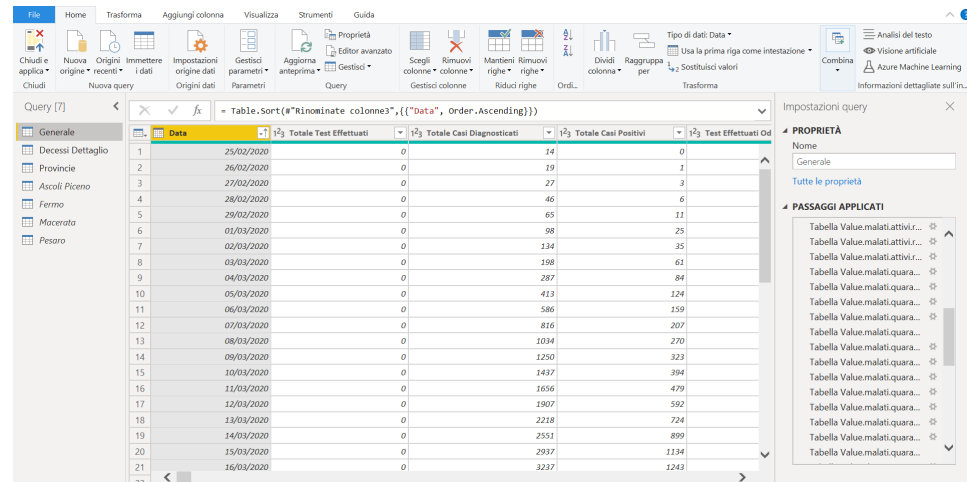


Figura 2.9. Finestra dell'Editor di query

Data shaping

Per “data shaping” si intende il processo di trasformazione dei dati, ad esempio rinominando colonne o tabelle, cambiando il tipo del dato, rimuovendo righe, impostando la prima riga come intestazione e così via. Esso interviene, anche, nel processo di combinazione dei dati, cioè nella connessione di due o più sorgenti dei dati, in cui il data shaping è necessario perché porta ad un consolidamento dei dati in un'unica query utile.

È importante specificare che, durante la modellazione dei dati, la sorgente dei dati di partenza non subisce alcuna modifica: viene modificata, o sottoposta a data shaping, solo questa vista specifica dei dati, che verrà salvata in formato **.pbix**.

Quando si eseguono operazioni di data shaping nell'Editor di query, si forniscono istruzioni dettagliate che verranno eseguite automaticamente. Ciò permette di modificare i dati durante la fase di caricamento e di presentazione. I passaggi specificati (ad esempio, rinominare una tabella, trasformare un tipo di dato o eliminare colonne) vengono registrati dall'Editor di query ed eseguiti ogni volta che questa query si connette alla sorgente dei dati, in modo che questi ultimi abbiano sempre la forma specificata. Questi passaggi vengono inoltre acquisiti, in sequenza, nel riquadro “Impostazioni query” in “Passaggi applicati”. La Figura 2.9 sopra riportata illustra l'aspetto della finestra dell'Editor di query, e, sul lato destro della schermata, il riquadro “Impostazioni query” con i “Passaggi Applicati”.

2.3 Creazione di report

Dopo avere creato un modello di dati (importando questi ultimi ed effettuando operazioni di ETL tramite l'Editor di query, integrato in Power BI Desktop), 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 Desktop, viene chiamata report. Un report può avere una o più pagine.

Con Power BI Desktop è possibile creare report complessi e visivamente dettagliati usando dati provenienti da più sorgenti, tali report possono essere, poi, condivisi con altri utenti. Nella Figura 2.10 viene visualizzata la prima pagina di un esempio di report, denominato "Introduzione" (come indicato nella scheda presente nella parte inferiore dell'immagine).

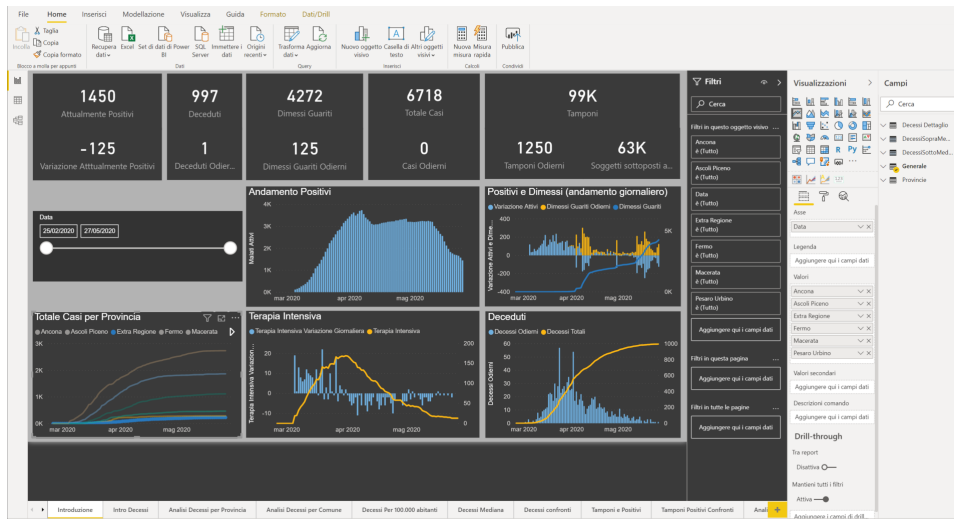


Figura 2.10. Esempio di pagina di un report

2.3.1 Il foglio di lavoro

Prima di iniziare il vero e proprio processo di creazione dei report, nonché la creazione di grafici, analizziamo la struttura del tipico foglio di lavoro dell'applicazione Power BI Desktop. Una volta avviata l'applicazione, dopo la schermata di avvio precedentemente riportata (Figura 2.5), ci troviamo davanti ad un foglio con delle aree di lavoro ben definite (Figura 2.11).

Le icone poste nell'estrema sinistra del foglio (Figura 2.12) rappresentano le tre visualizzazioni che Power BI mette a disposizione:

- **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.

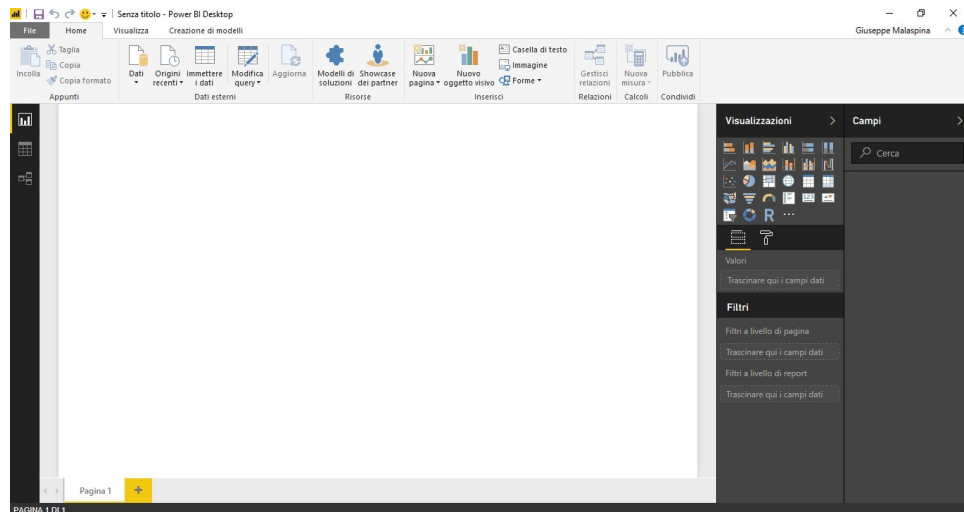


Figura 2.11. Foglio di lavoro in Power BI Desktop

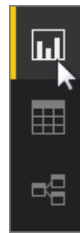


Figura 2.12. Icone delle viste di Power BI

- *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 di dati, e di gestirle o modificarle in base alle esigenze.

Nella parte alta troviamo tutte le voci del menù di navigazione e tutte le varie opzioni ad esso collegate. Nella parte centrale troviamo la pagina in cui verranno elaborati e visualizzati i grafici dei report. Sulla destra troviamo, invece, il menù “Visualizzazioni”, con cui possiamo scegliere quale grafico utilizzare per visualizzare i dati, e il menù “Filtri”, che consente di definire opportunamente i valori del grafico per migliorare la visualizzazione.

Visualizzazione Report

Nel foglio di lavoro di Power BI appena descritto, la pagina iniziale, riportata in Figura 2.11, è preimpostata nella visualizzazione *Report*. Ovviamente, quando si caricano i dati in Power BI Desktop per la prima volta, la visualizzazione *Report*

contiene un'area di disegno vuota. Da questa vista è possibile iniziare la costruzione di grafici nell'area di disegno stessa. In questa pagina iniziale sono, infatti, visibili tre riquadri: “Filtri”, “Visualizzazioni” e “Campi” (Figura 2.13). I primi due riquadri a sinistra, “Filtri” e “Visualizzazioni”, controllano l'aspetto delle visualizzazioni: tipo, colori, filtraggio e formattazione. L'ultimo riquadro a destra, “Campi”, gestisce i dati sottostanti utilizzati nelle visualizzazioni. Il contenuto visualizzato nell'editor di report varia a seconda delle selezioni effettuate nell'area di visualizzazione dei report.

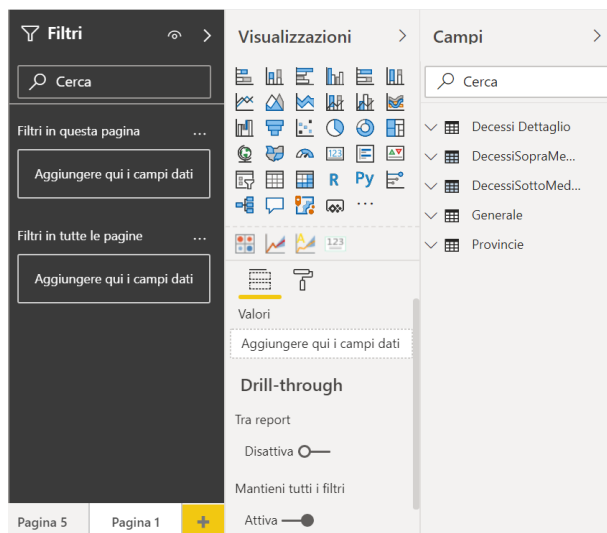


Figura 2.13. Riquadri laterali in visualizzazione *Report*

Visualizzazione *Dati*

La visualizzazione *Dati* (Figura 2.14) consente di esaminare, esplorare ed interpretare i dati nel modello di Power BI. 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.

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 al livello di riga. 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*

Una volta caricati i dati all'interno del foglio di lavoro è possibile, cliccando su “Gestisci Relazioni”, definire opportunamente le relazioni tra le nostre tabelle, impostare la cardinalità desiderata ed eventuali vincoli referenziali, come si evince

File Home Guida Strumenti tabella Strumenti colonna

Nome Id decesso Formato Numero intero Esecuzione del riepilogo Somma

Tipologia di dati Tipo di dati Numero intero Categoria di dati Senza categoria Ordina per colonna Gruppi di dati Gestisci relazioni Nuova colonna

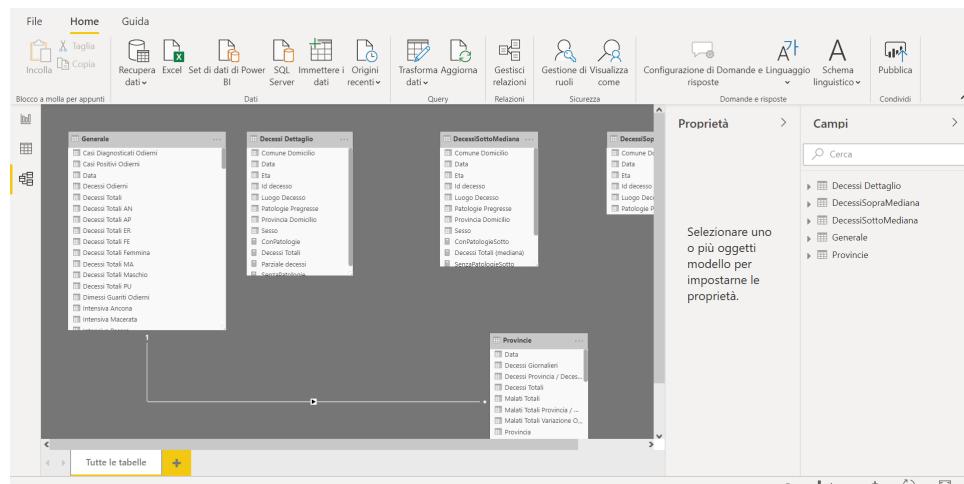
Struttura Formattazione Proprietà Ordina Gruppi Relazioni Calcoli

Id decesso	Data	Luogo Decesso	Sesso	Eta	Comune Domicilio	Provincia Domicilio	Patologie Progresso
1	lunedì 2 marzo 2020	TI FANO	M	87	FANO	Pesaro	1
2	mercoledì 4 marzo 2020	MN	M	61	PESARO	Pesaro	1
3	mercoledì 4 marzo 2020	INRCA PPI	M	84	ANCONA	Ancona	1
4	mercoledì 4 marzo 2020	UTIC PS	M	75	FANO	Pesaro	1
5	sabato 7 marzo 2020	TI FANO	M	77	PESARO	Pesaro	1
6	sabato 7 marzo 2020	ADAN	M	59	ANCONA	Ancona	1
7	domenica 8 marzo 2020	MI MN	M	83	PESARO	Pesaro	1
8	lunedì 9 marzo 2020	URB MED	F	98	PESARO	Pesaro	1
9	lunedì 9 marzo 2020	MI MN	F	72	TERRE ROVERESCHE	Pesaro	1
10	lunedì 9 marzo 2020	RIA MN	M	86	PESARO	Pesaro	1
11	martedì 10 marzo 2020	RIA MN	M	87	PESARO	Pesaro	1
12	martedì 10 marzo 2020	MI MN	F	80	FANO	Pesaro	1
13	martedì 10 marzo 2020	MI MN	M	93	CARTOCETO	Pesaro	1
14	martedì 10 marzo 2020	MI FERMO	F	84	MONTE CERIGNONE	Pesaro	1
15	martedì 10 marzo 2020	OSP SENIGALLIA	M	67	PESARO	Pesaro	1
16	martedì 10 marzo 2020	OSP SENIGALLIA	M	84	GIADARA	Pesaro	1
17	martedì 10 marzo 2020	OSP SENIGALLIA	M	85	PESARO	Pesaro	1
18	martedì 10 marzo 2020	MI FERMO	F	83	PORTO RECANATI	Macerata	1
19	mercoledì 11 marzo 2020	RIA MN	M	76	MONTELABBATE	Pesaro	1
20	mercoledì 11 marzo 2020	URBINO MED	F	87	PESARO	Pesaro	1
21	mercoledì 11 marzo 2020	UD AI MN	M	88	PESARO	Pesaro	1

TABELLA: Decessi Dettaglio (997 righe) COLONNA: Id decesso (997 valori distinti)

Figura 2.14. Visualizzazione *Dati*

dalla Figura 2.15. Ciò può essere particolarmente utile quando un modello contiene relazioni complesse tra molte tabelle. È importante notare, però, che Power BI Desktop semplifica la creazione di queste relazioni. Nella maggior parte dei casi, infatti, tutte le operazioni vengono eseguite automaticamente dalla funzionalità *Rilevamento automatico*: se si eseguono query in due o più tabelle alla volta, Power BI Desktop tenta di trovare e creare automaticamente le relazioni quando i dati vengono caricati. Talvolta, tuttavia, può essere necessario creare manualmente le relazioni, oppure apportarvi alcune modifiche.

Figura 2.15. Visualizzazione *Relazioni*

2.3.2 Tipi di visualizzazione supportati

Una quantità di dati molto elevata è molto lontana dall'extrapolazione dei dati utili, in quanto le informazioni sono disperse in una quantità molto vasta di dati. La visualizzazione grafica risulta essere molto utile per la comprensione immediata dei dati.

La visualizzazione *Report*, illustrata precedentemente, permette all'utente di visualizzare l'area di disegno dove vengono create e disposte le visualizzazioni grafiche. Selezionando i campi di nostro interesse dal riquadro "Campi", possiamo scegliere il tipo di visualizzazione grafica.

Power BI offre 24 visualizzazioni già preinstallate (Figura 2.16), ma molte altre sono reperibili dal Windows Store, come vedremo in seguito.

Vediamo, ora, una carrellata dei grafici più usati e più funzionali per questa tesi.

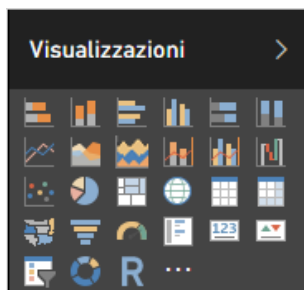


Figura 2.16. Grafici prefiniti in Power BI

L'oggetto visivo "Scheda"

Molto spesso in un dashboard o in un report di Power BI si vuole visualizzare semplicemente un numero, ad esempio il numero di persone decedute a causa del Coronavirus, oppure il numero di persone attualmente positive. Questo tipo di visualizzazione è denominato "Scheda" (Figura 2.17).

Come quasi tutte le visualizzazioni native di Power BI, è possibile creare le "Schede" tramite l'editor di report.

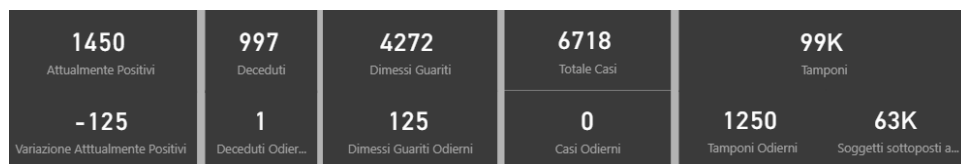


Figura 2.17. Esempio di oggetto visivo "Scheda"

Gli istogrammi

Gli istogrammi consentono una rappresentazione grafica di una funzione, nota solo per un certo numero di valori della variabile (Figura 2.18). Dunque, sono utilizzabili per analisi di dati che risultano essere definiti in intervalli discreti di valori.

Un istogramma 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.

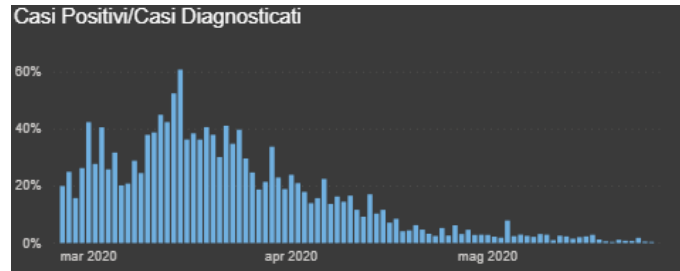


Figura 2.18. Esempio di istogramma

Il grafico a linee

Il grafico a linee evidenzia la forma complessiva di un'intera serie di valori, in genere nel tempo. Esso permette, anche, di rappresentare l'andamento temporale di più serie di valori; si hanno cioè più linee in un'unica rappresentazione (Figura 2.19).

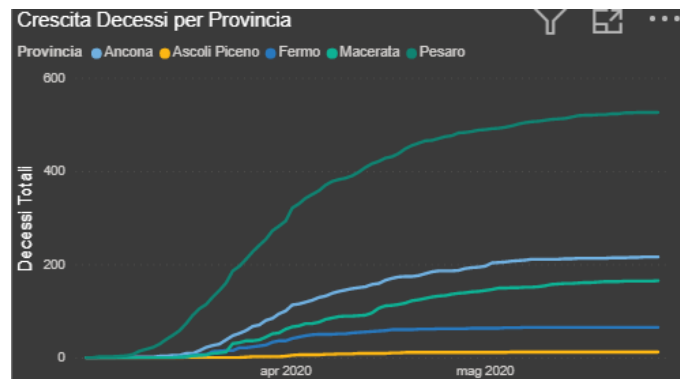


Figura 2.19. Esempio di grafico a linee

Grafici combinati

I grafici combinati rappresentano il tipo di grafico che utilizzeremo di più nell'ambito della presente tesi. L'importanza di tale grafico per il nostro studio è dovuto alle

sue caratteristiche intrinseche. Tale grafico, infatti, racchiude un istogramma e un grafico a linee. Inoltre, la combinazione dei due grafici in uno permette di confrontare i dati in modo più rapido.

Poiché i grafici combinati possono avere uno o due assi Y, è necessario esaminarli con attenzione (Figura 2.20).

I grafici combinati rappresentano un'ottima scelta, soprattutto 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 tra due misure in un unico oggetto visivo;
- per verificare se una misura raggiunge la destinazione definita da un'altra misura;
- per risparmiare spazio nell'area di disegno.

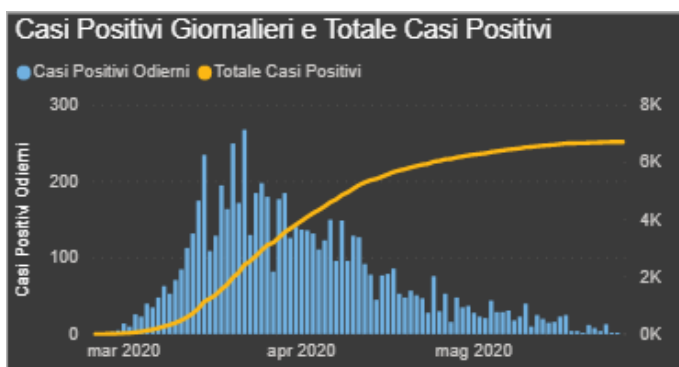


Figura 2.20. Esempio di grafico combinato

Il grafico ad area

Il grafico ad area si basa sul grafico a linee; la differenza è che l'area compresa tra l'asse e la linea viene riempita (Figura 2.21).

I grafici ad aree enfatizzano l'entità del cambiamento nel tempo e possono essere usati per attirare l'attenzione sul valore totale in una tendenza; proprio per questa caratteristica, tale tipologia di grafico riveste un ruolo importante nel nostro caso di studio.

Il grafico a torta

Il grafico a torta indica la relazione tra un intero e le sue parti. In questo grafico i dati vengono rappresentati sotto forma di un cerchio a spicchi, con le relative percentuali (Figura 2.22).

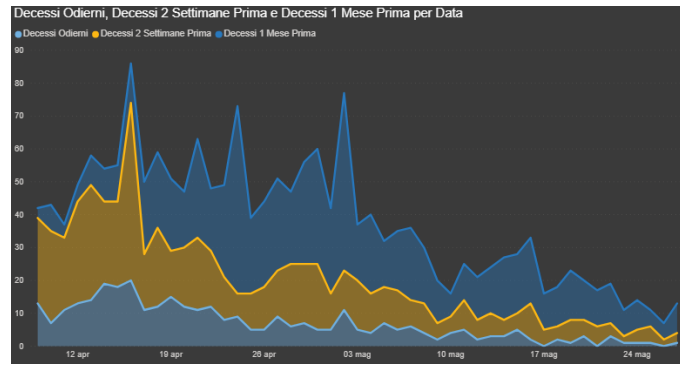


Figura 2.21. Esempio di grafico ad area

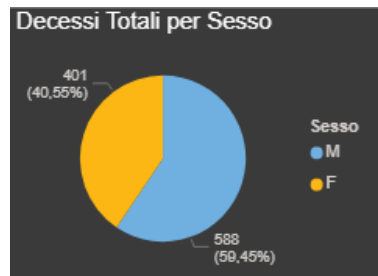


Figura 2.22. Esempio di grafico a torta

Le mappe

Questi particolari grafici vengono usati per associare informazioni categoriche e quantitative a posizioni nello spazio. Power BI dispone, in realtà, di due diversi tipi di visualizzazioni della mappa: la mappa a bolle, che posiziona una bolla su un punto geografico (Figura 2.23), e la mappa di forme, che illustra la struttura dell'area da visualizzare. In particolare Power BI si integra con Bing Maps per fornire coordinate della mappa predefinite (un processo denominato geocodifica) e poter creare le mappe.

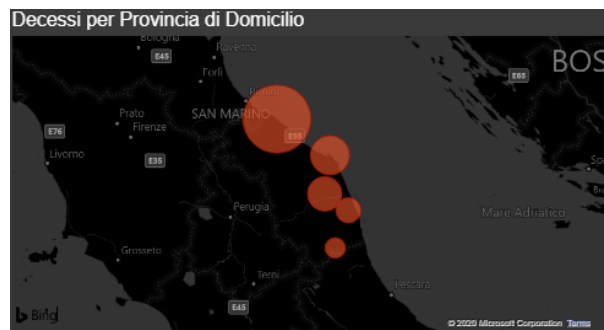


Figura 2.23. Esempio di mappa a bolle

Il grafico a dispersione

Un grafico a dispersione ha sempre due assi di valori per mostrare un set di dati numerici lungo un asse orizzontale e un altro set di valori numerici lungo un asse verticale. Nel grafico vengono visualizzati i punti in corrispondenza dell'intersezione di un valore numerico x e un valore numerico y , combinando questi valori in punti dati singoli. Power BI può distribuire questi punti dati uniformemente, o in maniera non uniforme, sull'asse orizzontale, a seconda dei dati rappresentati nel grafico. Un grafico a dispersione mostra, dunque, la relazione tra due valori numerici.

In particolare, ci sono i grafici a dispersione a bolle (Figura 2.24), che sostituiscono i punti dati con bolle, le cui dimensioni rappresentano una terza dimensione aggiuntiva dei dati.

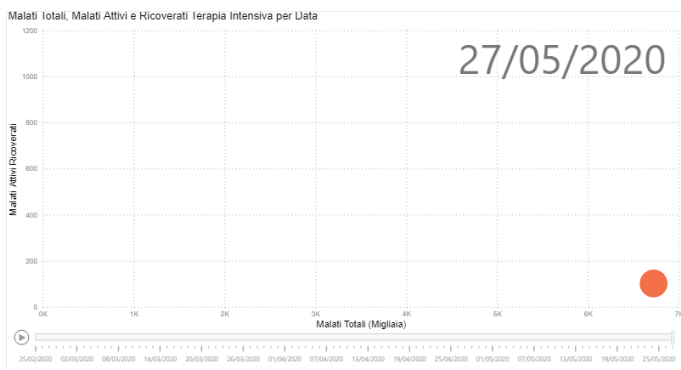


Figura 2.24. Esempio di grafico a dispersione a bolle

2.3.3 Oggetti visivi con script in linguaggio R

Nella parte introduttiva del software Power BI abbiamo evidenziato come uno dei suoi maggiori punti di forza sia il supporto al linguaggio R, un potente linguaggio di programmazione usato da numerosi esperti di statistica, data scientist e analisti di dati.

L'utilizzo del solo R, però, implica una serie di limitazioni: l'interfaccia non è graficamente moderna e gli utenti possono avere difficoltà nel produrre buoni risultati. Il maggiore inconveniente di tali limitazioni è che R non è particolarmente potente ed efficace nella visualizzazione dei dati e nella condivisione/collaborazione con il resto dell'organizzazione.

Per superare tutte queste limitazioni, Power BI ha introdotto lo scripting R all'interno del suo ambiente. Power BI si integra, dunque, perfettamente con R, e questa integrazione tra i due può aiutare a sviluppare contenuto significativo utilizzando R e le abilità di visualizzazione di Power BI.

Tuttavia, Power BI Desktop non include, distribuisce o installa il motore R. Per eseguire gli script R in Power BI Desktop è necessario installare R nel computer locale.

Ottenere oggetti visivi R da AppSource e da Power BI Desktop

L'aspetto fondamentale è che in Power BI Desktop e nel servizio Power BI è possibile usare oggetti visivi di Power BI basati su R senza conoscere R e senza dover creare script R. Ciò consente di sfruttare la potenza analitica degli oggetti visivi di R, nonché gli script di R, senza dover necessariamente conoscere tale linguaggio o dover programmare manualmente.

Per usare gli oggetti visivi di Power BI basati su R è sufficiente scaricare ogni oggetto visivo dal sito “AppSource” online; in alternativa, è possibile esplorare e scaricare tali oggetti da Power BI Desktop, per poi usarli come qualsiasi altro tipo di oggetti visivi in Power BI Desktop.

Per effettuare quanto sopra specificato è necessario, in Power BI Desktop, fare clic con il pulsante destro del mouse sui puntini di sospensione (...) nel riquadro “Visualizzazioni” e scegliere “Importa dal Marketplace”. La finestra di dialogo “Oggetti visivi di Power BI” (Figura 2.25) permette di scorrere gli oggetti visivi di Power BI disponibili e di selezionare quello desiderato. Una volta individuato l'oggetto visivo desiderato, è necessario fare clic su “Aggiungi” per aggiungerlo a Power BI Desktop.

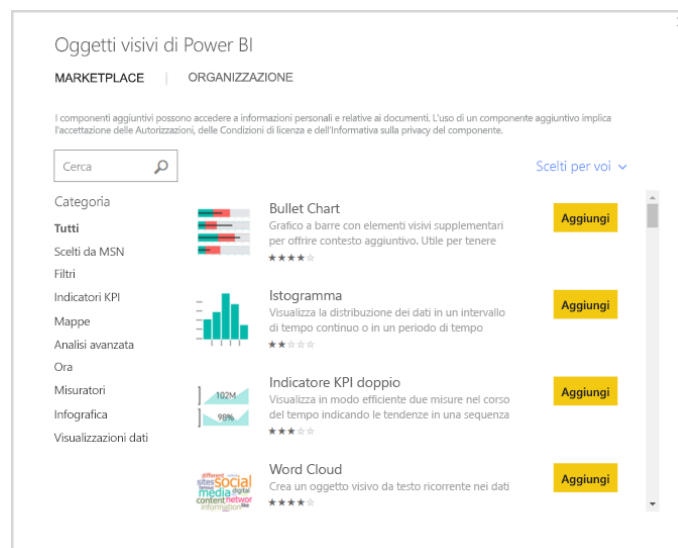


Figura 2.25. Finestra “Oggetti visivi di Power BI”

E poiché gli oggetti visivi R sono esattamente come qualsiasi altro oggetto visivo in Power BI, è possibile interagire con essi e stabilire connessioni con altri oggetti visivi nell'area di disegno, senza la necessità di modificare lo script R. Notare, dunque, come la semplificazione introdotta da Power BI sia notevole.

In particolare, all'interno della presente tesi, nella parte dedicata all'analisi predittiva, utilizzeremo oggetti visivi R per attività di forecasting basata sul modello ARIMA (Autoregressive Integrated Moving Average - Figura 2.26), che analizzeremo in seguito.

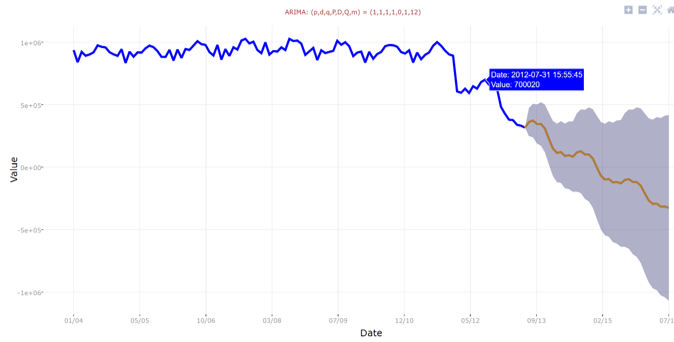


Figura 2.26. Esempio di grafico basato sul modello Arima

2.3.4 Filtri

Una funzionalità molto utile in Power BI Desktop (presente anche in Power BI Service) è, sicuramente, il filtraggio dei dati. Nella visualizzazione *Report*, nel riquadro “Filtri” (Figura 2.27), è possibile impostare diversi filtri ai nostri oggetti visivi ed alle nostre pagine.



Figura 2.27. Il riquadro “Filtri” nella visualizzazione *Report*

In particolare, possiamo vedere come Power BI sia dotato di tre tipologie di filtri che ci permettono di analizzare al meglio i nostri grafici. Tali tipologie sono:

- *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 usando un campo che non è già presente in una delle immagini del report.

Ogni singolo filtro ha dei sottofiltri, che servono a selezionare ulteriormente il dato in analisi:

- *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, giorno);
- *Filtro primi N*: utile nel caso in cui bisogna fare delle analisi su alcuni valori specifici.

2.3.5 Formule DAX

A volte i dati non includono tutto ciò che occorre per rispondere ad alcune domande importanti. A questo scopo possono essere create nuove misure, cioè 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; si tratta di un linguaggio funzionale, ovvero tutto il codice eseguito è contenuto all’interno di una funzione. DAX include una libreria con oltre 200 funzioni, operatori e costrutti. Questa libreria offre un’enorme flessibilità per la creazione di misure di calcolo dei risultati per quasi tutte le esigenze di analisi dei dati. In DAX le funzioni possono contenere altre funzioni annidate, istruzioni condizionali e riferimenti a valori. L’esecuzione in DAX inizia dalla funzione o dal parametro più interni e procede verso l’esterno. Esiste nell’ambiente DAX, in particolare, un tipo di dato che si userà probabilmente spesso in Power BI, ovvero DateTime. Questo tipo di dato può essere usato in modo accurato per calcoli; esso verrà utilizzato di frequente nella presente tesi. È importante sottolineare che i risultati calcolati delle misure DAX vengono aggiornati costantemente in base alle operazioni dell’utente nei report e consentono un’esplorazione dei dati ad hoc, rapida e dinamica.

Le misure vengono create e visualizzate in visualizzazione *Report* o in vista *Dati*. Tali misure vengono mostrate nell’elenco di campi con un’icona a forma di calcolatrice (Figura 2.28). È possibile assegnare qualsiasi nome alle misure e aggiungerle a una visualizzazione nuova o esistente con le normali procedure usate per gli altri campi.

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

- Colonne calcolate;
- Tabelle calcolate.

La Figura 2.29 mostra la formula DAX utilizzata per la creazione di una misura calcolata; si noti come gli elementi necessari per una misura calcolata siano:

- un nuovo nome di misura;
- almeno una funzione o un’espressione.

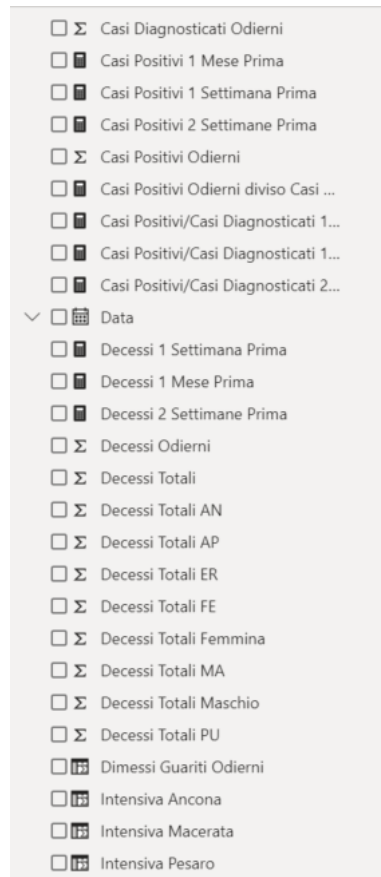


Figura 2.28. Il riquadro “Campi” dell’area *Report* con misure calcolate

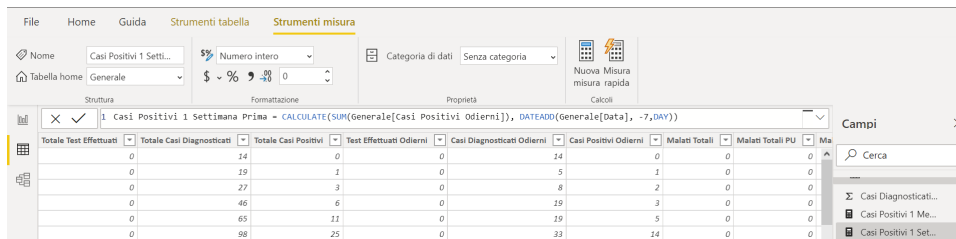


Figura 2.29. Esempio di misura calcolata con DAX

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 una sorgente di dati, viene creata una formula DAX che definisce i valori della colonna, la quale si basa, dunque, sui dati già caricati nel modello (Figura 2.30).

In Power BI, le colonne calcolate vengono create usando la funzionalità “Nuova

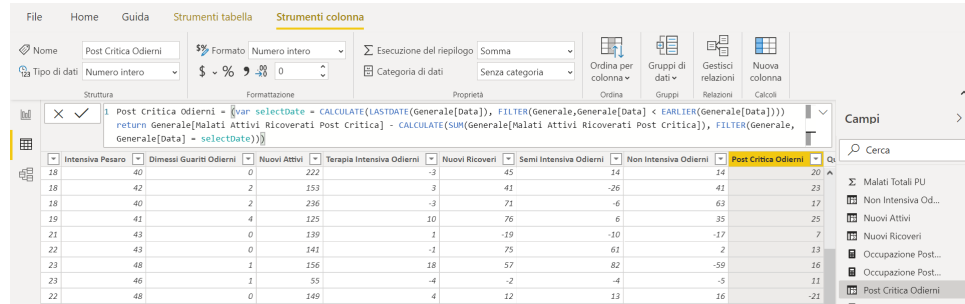


Figura 2.30. Esempio di colonna calcolata con DAX

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

Tabelle calcolate

Le tabelle calcolate consentono di aggiungere nuove tabelle basate sui dati già caricati nel modello; infatti, invece di eseguire query e caricare i valori nelle colonne della nuova tabella da un’origine dati, viene creata una formula DAX che definisce i valori della tabella, la quale si basa, dunque, sui dati già caricati nel modello (Figura 2.31).

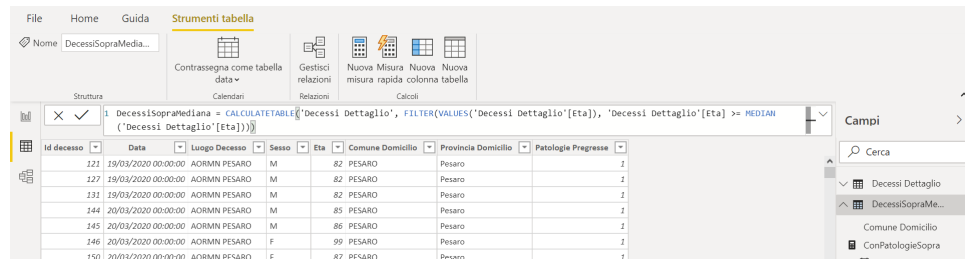


Figura 2.31. Esempio di tabella calcolata con DAX

Perciò le tabelle calcolate sono, in genere, preferibili per i dati e i calcoli intermedi da archiviare come parte del modello, anziché da calcolare automaticamente o come parte di una query.

In Power BI, le tabelle calcolate vengono create usando la funzionalità “Nuova tabella” in visualizzazione *Report* o in vista *Dati*.

Analogamente alle normali tabelle, le tabelle calcolate possono avere relazioni con altre tabelle. Le colonne nella tabella calcolata contengono i tipi di dati, la formattazione e possono appartenere a una categoria di dati. Le tabelle calcolate vengono, inoltre, ricalcolate se una delle tabelle da cui vengono estratti i dati viene in qualche modo aggiornata.

2.4 Pubblicazione e condivisione di dashboard

Una volta ultimata la fase di analisi dei dati all'interno del foglio di lavoro, il nostro report è pronto per essere condiviso con altri utenti; siamo, dunque, giunti all'ultima fase del flusso di lavoro in Power BI.

A tal proposito, una delle funzionalità più importanti che Power BI mette a disposizione è la possibilità di creare aree di lavoro di gruppo; queste rappresentano uno spazio eccezionale per collaborare con i colleghi nei dashboard, nei report e nei set di dati. Se si vuole solo distribuire un dashboard completato ai colleghi, è consigliabile, quindi, condividerlo con loro o creare un pacchetto di contenuto aziendale.

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.

In Power BI Desktop è possibile usare il comando “Pubblica”, posizionato nel menù “Home” in alto a destra, e successivamente, specificando una destinazione e l'area di lavoro di gruppo, si permette la pubblicazione del report nel servizio Power BI rendendolo disponibile agli altri utenti dell'organizzazione che dispongono di una licenza per Power BI.

Una volta completata la condivisione del file, non ci resta che autenticarci su Power BI Online con i nostri dati di accesso ed accedere all'area di lavoro, dentro la quale compariranno tutte le dashboard, come si evince dalla Figura 2.32 .

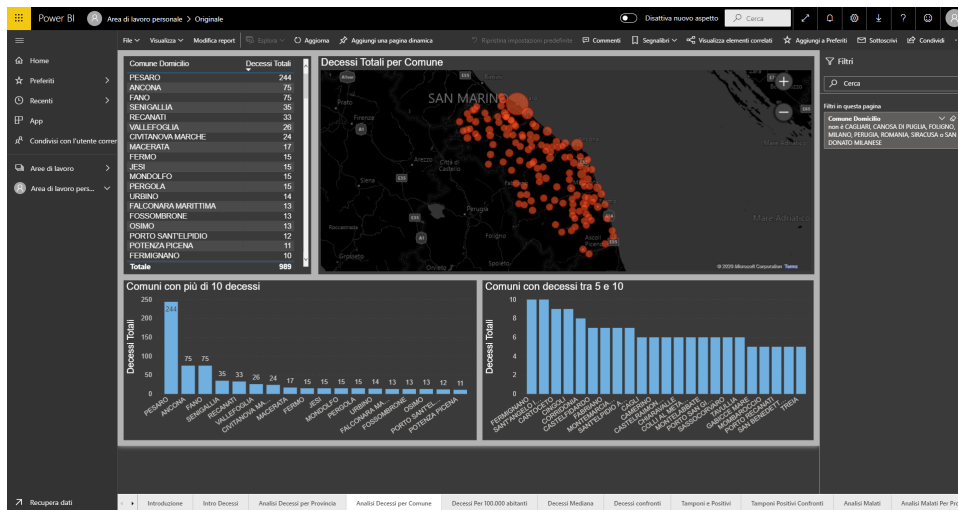


Figura 2.32. Esempio di dashboard pubblicata in Power BI Service

Acquisizione ed elaborazione dei dati

In questo capitolo verranno presentati ed analizzati i dati usati nella presente tesi. Nella prima parte del capitolo, in particolare, saranno descritte la sorgente da cui i dati provengono e la struttura dei dati stessi. Nella seconda parte, invece, si illustreranno le operazioni eseguite per l'importazione dei dati e per la creazione del modello dei dati in Power BI Desktop.

3.1 Origine dei dati

L'obiettivo del nostro studio è quello di analizzare i dati relativi al Covid-19 nella Regione Marche, al fine di poter descrivere l'andamento della diffusione del virus all'interno della regione e di poter fare eventuali analisi circa i futuri sviluppi della pandemia nel territorio marchigiano.

La sorgente di dati utilizzata per realizzare le nostre 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>).

Come sottolineato anche in precedenza nel Capitolo 1, la Regione Marche, grazie al GORES (“Gruppo Operativo Regionale Emergenza Sanitaria”), si è fin da subito impegnata nel fornire, con grande chiarezza e, soprattutto, con trasparenza, i dati riguardanti la diffusione del Coronavirus all'interno delle provincie della regione. I dati forniti dal GORES con cadenza giornaliera riguardano, in particolare, gli aspetti relativi ai tamponi, ai ricoveri e ai decessi legati allo sviluppo della pandemia all'interno della regione. Il repository di nostro interesse “raccolge” proprio questi dati forniti dal GORES (diffusi dal GORES stesso soltanto in formato PDF) in un formato utile ai fini di analisi e statistiche.

3.1.1 Dati forniti dalla Regione Marche

Cerchiamo, ora, di analizzare nel dettaglio i dati messi a disposizione dalla Regione Marche, dati che poi sono stati raccolti all'interno del repository GitHub di nostro interesse.

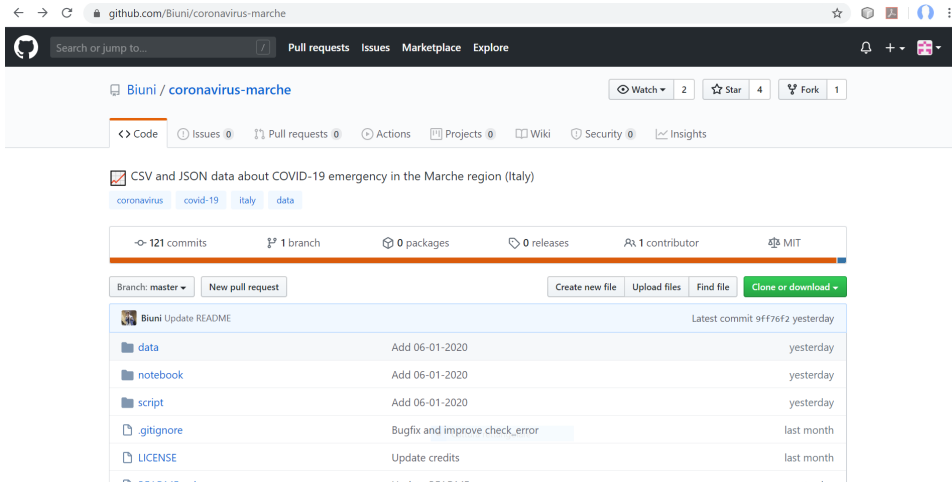


Figura 3.1. Repository di GitHub utilizzato per l'importazione dei dati

La Regione Marche, a differenza di altre regioni, ha dato grande peso e importanza alla comunicazione tempestiva dei dati riguardanti tamponi, ricoveri e decessi legati al Coronavirus.

I dati, raccolti ed elaborati dal GORES, sono stati rilasciati giornalmente in tre diversi report (*Blu*, *Giallo* e *Arancio*). Tali report sono stati resi disponibili in un formato PDF (poco adatto alle nostre analisi) e sono stati pubblicati direttamente sul sito della Regione stessa, in una sezione completamente dedicata al Coronavirus (Figura 3.2).

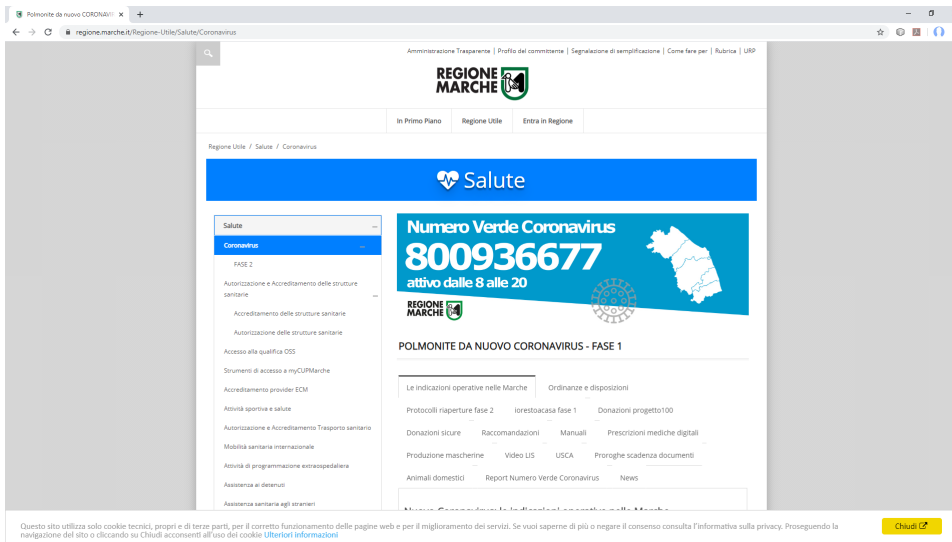


Figura 3.2. La pagina del sito della Regione Marche dedicata al Coronavirus

I tre report pubblicati giornalmente contengono, in particolare, le seguenti informazioni:

- *Report Blu*: documento nel quale viene rilasciato 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 guarti, etc.).
- *Report Arancio*: documento nel quale sono rilasciati i numeri e i dettagli sui decessi avvenuti nelle ultime 24 ore.

Esaminiamo, ora, più da vicino questi tre report.

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'altra tabella che, sempre con riferimento ad ogni singolo giorno, riporta i *Campioni Positivi totali: progressione* e i *Campioni testati totali: progressione*.

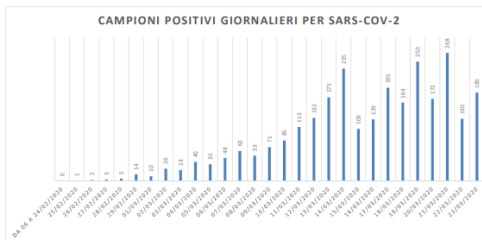
In questo report vengono, altresì, riportati due grafici relativi ai *Campioni Positivi giornalieri per SARS-Cov-2* e ai *Campioni Positivi totali: progressione* (Figura 3.3).

Report Giallo

Il *report Giallo* (Figura 3.4) si concentra, in particolare, sullo stato dei positivi al Coronavirus ed è organizzato in 7 tabelle:

- La prima riassume le informazioni relative al totale dei test effettuati e al totale dei test positivi.
- La seconda suddivide i malati in base alla criticità della loro situazione. In particolare, vengono riportati i malati suddivisi per tipo di ricovero, i dimessi e i guariti, i malati in isolamento domiciliare e il totale dei deceduti.
- La terza illustra i casi positivi provincia per provincia;
- La quarta espone i ricoveri in terapia intensiva e quelli in terapia semintensiva, suddivisi per struttura ospedaliera
- La quinta illustra i ricoveri in reparti non intensivi, suddivisi per struttura.
- La sesta 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.
- La settima riporta, invece, i ricoveri in degenze post critiche, suddivisi per struttura di ricovero.

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	197
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
23/03/2020	185	465



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	84	267
04/03/2020	124	413
05/03/2020	159	586
06/03/2020	207	816
07/03/2020	270	1014
08/03/2020	323	1252
09/03/2020	394	1437
10/03/2020	479	1656
11/03/2020	597	1907
12/03/2020	724	2218
13/03/2020	898	2551
14/03/2020	1133	2937
15/03/2020	1442	3237
16/03/2020	1371	3572
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
23/03/2020	2736	7229

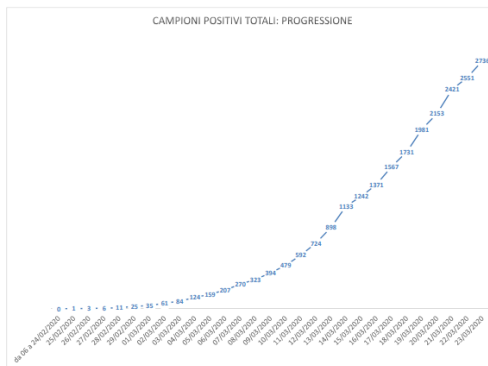


Figura 3.3. Il report Blu pubblicato giornalmente dalla Regione Marche

Report Arancio

Il report Arancio si concentra esclusivamente 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.

3.1.2 Struttura dei dati

Il repository utilizzato come sorgente dei dati del nostro studio si propone esattamente di raccogliere i tre report pubblicati giornalmente dalla Regione Marche all'interno di un unico "file", con lo scopo di trasformare i PDF forniti dal GORES in un formato utilizzabile per l'analisi.

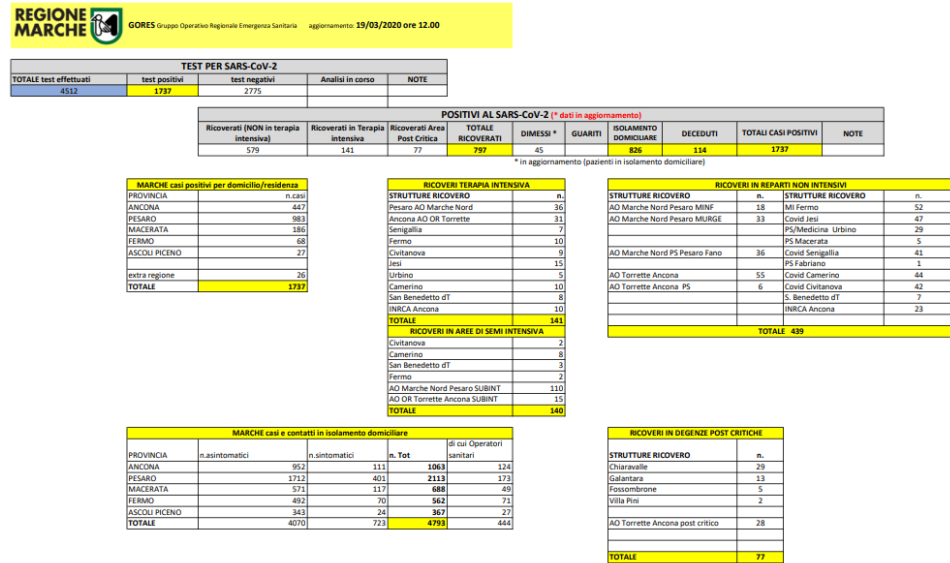


Figura 3.4. Il report Giallo pubblicato giornalmente dalla Regione Marche

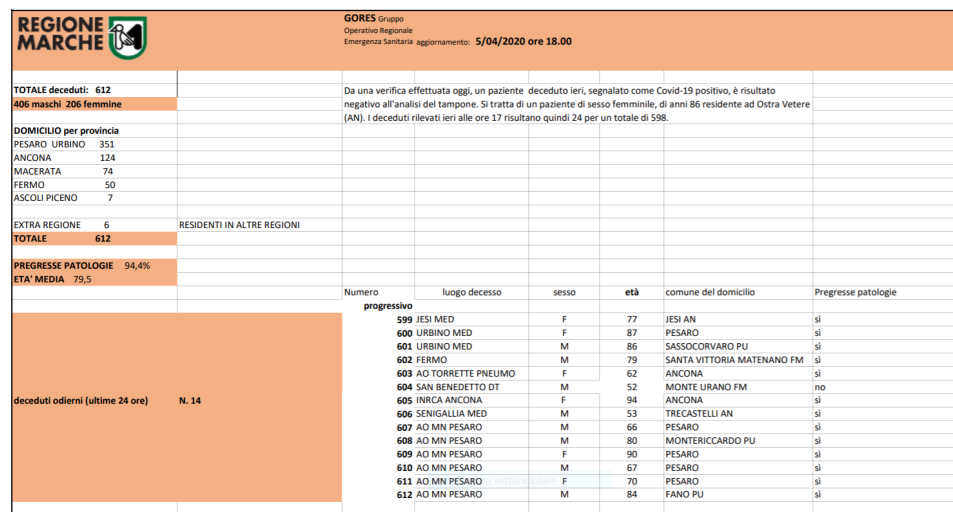


Figura 3.5. Il report Arancio pubblicato giornalmente dalla Regione Marche

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": "05-27-2020",
  "file_pdf": {
    "locale": {
      "blu": "./data/PDF/GORES_Blu/05-27-2020.pdf",
      "giallo": "./data/PDF/GORES_Giallo/05-27-2020.pdf",
    }
  }
}
```

```

    "arancio": "../data/PDF/GORES_Arancio/05-27-2020.pdf"
  },
  "online": {
    "blu": "https://www.regione.marche.it/portals/0/Salute/CORONAVIRUS/DatiGORES/
GORESblu27052020_9.pdf",
    "giallo": "https://www.regione.marche.it/portals/0/Salute/CORONAVIRUS/
goresgiallo27maggio.pdf",
    "arancio": "https://www.regione.marche.it/portals/0/Salute/CORONAVIRUS/DatiGORES/
Report_sint_decessi_COVID19_agg_27maggio2020_ore18.pdf"
  }
},
"tamponi": {
  "totali": {
    "test_effettuati": 98535,
    "casi_diagnosticati": 62815,
    "casi_positivi": 6718
  },
  "odierni": {
    "test_effettuati": 1250,
    "casi_diagnosticati": 892,
    "casi_positivi": 0
  }
},
"malati": {
  "totali": {
    "persone": 6718,
    "provincia": {
      "pesaro_urbino": 2740,
      "ancona": 1870,
      "macerata": 1117,
      "fermo": 469,
      "ascoli_piceno": 290,
      "extra_regione": 232
    }
  },
  "dimessi_guariti": 4272,
  "attivi": {
    "persone": 1450,
    "isolamento_domiciliare": 1347,
    "ricoverati": {
      "totali": 103,
      "terapia_intensiva": {
        "totali": 13,
        "strutture": {
          "AO_Marche_Nord": 0,
          "AO_Ancona_Torrette": 4,
          "Senigallia": 0,
          "Fermo": 2,
          "Civitanova": 0,
          "Jesi": 1,
          "Urbino": 0,
          "Camerino": 3,
          "San_Benedetto_dT": 3,
          "Macerata": 0,
          "INRCA_Ancona": 0
        }
      },
      "semi_intensiva": {
        "totali": 14,
        "strutture": {
          "Civitanova": 0,
          "Camerino": 5,
          "San_Benedetto_dT": 0,
          "Jesi": 0,
          "Macerata": 0,
          "Urbino": 0,
          "Senigallia": 0,
          "Fermo": 0,
          "AO_Marche_Nord": 0,
          "AO_Ancona_Torrette": 9,
          "Villa_Pini": 0
        }
      }
    }
  },
  "post_critica": {
    "totali": 191,
    "strutture": {
      "Chiaravalle": 19,
      "Galantara": 74,
      "Fossombrone": 11,
      "Villa_Pini": 11,
      "RSA_Campofilone": 60,
      "AO_Ancona_Torrette": 0,
      "INRCA_Fermo": 16,
      "OspedaleDaCampo_Jesi": 0
    }
  }
},
"non_intensiva": {
  "totali": 37,
  "strutture": {
    "Fermo": 8,
    "Jesi": 1,
    "Urbino": 0,
    "Macerata": 7,
    "Senigallia": 0,
  }
}

```

```

        "Fabriano": 0,
        "Camerino": 12,
        "Civitanova": 0,
        "San_Benedetto_dT": 0,
        "INRCA_Ancona": 0,
        "AO_Marcho_Nord": 9,
        "AO_Ancona_Torrette": 0,
        "Villa_Pini": 0
    }
}
},
"quarantena_domiciliare": {
    "totali": 33114,
    "attivi": {
        "totali": 3335,
        "operatori_sanitari": 345,
        "sintomatici": 1287,
        "asintomatici": 2048
    },
    "provincia": {
        "pesaro_urbino": {
            "totali": 15409,
            "attivi": {
                "totali": 1899,
                "operatori_sanitari": 162,
                "sintomatici": 924,
                "asintomatici": 975
            }
        },
        "ancona": {
            "totali": 9951,
            "attivi": {
                "totali": 638,
                "operatori_sanitari": 66,
                "sintomatici": 153,
                "asintomatici": 485
            }
        },
        "macerata": {
            "totali": 3892,
            "attivi": {
                "totali": 557,
                "operatori_sanitari": 100,
                "sintomatici": 153,
                "asintomatici": 404
            }
        },
        "fermo": {
            "totali": 2300,
            "attivi": {
                "totali": 177,
                "operatori_sanitari": 13,
                "sintomatici": 48,
                "asintomatici": 129
            }
        },
        "ascoli_piceno": {
            "totali": 1562,
            "attivi": {
                "totali": 64,
                "operatori_sanitari": 4,
                "sintomatici": 9,
                "asintomatici": 55
            }
        }
    }
},
"decessi": {
    "totali": {
        "decessi": 997,
        "sesso": {
            "maschio": 595,
            "femmina": 402
        }
    },
    "provincia": {
        "pesaro_urbino": 527,
        "ancona": 217,
        "macerata": 166,
        "fermo": 66,
        "ascoli_piceno": 13,
        "extra_regione": 8
    }
},
"odierni": {
    "decessi": 1,
    "dettaglio": [
        {
            "id": 997,
            "luogo_decesso": "OSP._MACERATA",
            "sesso": "M",
            "eta": 98,
            "comune_domicilio": "RECANATI",

```

```

    }
  }
}
"provincia_domicilio": "MC",
"pregresse_patologie": 1
}

```

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

I file `json` dei singoli giorni vengono, successivamente, raccolti in un unico file, sempre di tipo `json`, che include i precedenti come elementi di una sorta di “array di report”; questo unico file `json`, cioè, non contiene altro che un array, dove ciascun elemento dell’array 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 `json` finale.

Quest’ultimo file `json` è proprio quello che è stato importato in Power BI Desktop per eseguire il nostro studio.

Dettaglio dei campi che compongono il `json`

Analizziamo, ora, in dettaglio, la struttura del file `json` che contiene i dati di nostro interesse. È importante sottolineare come questo sia uno dei passaggi chiave della suddetta analisi, poiché 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_postivi*: 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 Attività di ETL

In questa sezione ci soffermiamo su una delle operazioni fondamentali che precedono, in generale, una qualunque attività di *Data Analysis*, ovvero il concetto di ETL.

Vediamo, in particolare, come funziona tale processo e come è stato applicato al nostro caso di studio.

ETL (“Extract, Transform and Load” - Figura 3.6) è il processo utilizzato per trasformare dati grezzi in informazioni fruibili di Business Intelligence. Si tratta, cioè, di un processo di caricamento dei dati da un sistema sorgente (*dati grezzi*) in un sistema target. Il sistema sorgente può essere un database, un file di dati piatti o un’applicazione, mentre il sistema target può essere un altro database o qualche altro sistema di memorizzazione dei dati.

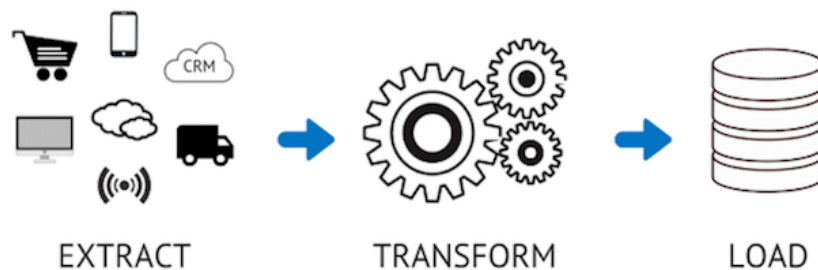


Figura 3.6. Il processo ETL e le tre fasi di *Data Preparation*

Il primo problema, in un processo ETL, è dunque quello di fondere dati provenienti dalle fonti disparate, e di inoltrarli tra i diversi ambienti presenti solitamente

in azienda o, eventualmente, all'esterno di essa, consentendo di orientare le scelte del management.

La sigla ETL indica i tre processi chiave della cosiddetta *Data Preparation*, ovvero:

- *Extract*: i dati grezzi vengono estratti da diverse fonti, quali database, registri di attività, report su anomalie, eventi di sicurezza e altre attività transazionali.
- *Transform*: in questa fase, i dati subiscono la trasformazione tramite l'applicazione di una serie di regole, solitamente definite a livello aziendale (standardizzazione, verifica, etc.). Attraverso la trasformazione, ciò che inizialmente era grezzo ed inutilizzabile viene plasmato in un insieme di dati, pronto per l'ultima fase che è quella del caricamento.
- *Load*: l'ultimo atto dell'intero processo di *Data Preparation* prevede il caricamento dei dati, estratti e trasformati, in una nuova destinazione.

Nelle sottosezioni seguenti descriveremo le operazioni più nello specifico.

3.2.1 Estrazione

La prima fase del processo di ETL, ovvero l'estrazione, consiste nel prelievo dei dati dai sistemi sorgente per prepararli alla successiva fase di trasformazione.

Durante questa fase di estrazione dei dati è possibile scegliere tra tre alternative:

- scrivere programmi ad hoc;
- utilizzare uno dei tanti tool di ETL presenti nel mercato;
- utilizzare entrambe le soluzioni.

Con particolare riferimento alla seconda soluzione, è importante sottolineare come sul mercato siano presenti degli applicativi che consentono di facilitare alcune operazioni comuni al processo di ETL, risparmiando agli utilizzatori molte ore di lavoro per scrivere righe di codice di infrastruttura in grado di far eseguire al sistema funzionalità quali l'apertura dei file o la lettura dei tracciati record; tutte queste attività, infatti, vengono svolte automaticamente dai tool di ETL in commercio. Ciò consente di ridurre al minimo il lavoro del personale incaricato della gestione dei dati.

3.2.2 Trasformazione

La trasformazione rappresenta la seconda fase del processo di ETL. Essa si articola in due passaggi:

- *Preparare e trasformare i dati*: in questo step tutti i dati vengono raccolti e aggregati secondo alcuni parametri, come tendenze, previsioni, e tanti altri. L'utilità di aggregare questi dati sta nel fatto di avere una maggiore comprensione e facilità di studio da parte dell'analista.
- *Filtrare e pulire*: dopo aver organizzato i dati in diversi aggregati, è possibile, al fine di affinare ulteriormente il dato analizzato, utilizzare dei filtri volti a ridurre al minimo lo spazio occupato dalle informazioni. La "pulitura", invece, ha lo scopo di eliminare dal sistema i dati non significativi o non consistenti rispetto al fenomeno da valutare.

3.2.3 Caricamento

L'ultima fase del processo di ETL è quella del caricamento, che consiste nell'inserire i dati ripuliti all'interno del sistema target, ovvero nella base di dati di destinazione.

Durante questa fase, i dati possono essere caricati sia periodicamente, cioè un po' alla volta, sia in tempo reale, cioè caricando tutte le informazioni in forma continuativa. In generale tutti gli applicativi sul mercato supportano entrambe le opzioni. Inoltre, l'utente può scegliere se inviare i dati al database sorgente già trasformati, se richiedere il caricamento dei dati passo passo, cioè man mano che il processo lo richiede, oppure se utilizzare un tipo di approccio misto.

3.3 Attività di ETL svolte in Power BI

Il repository di GitHub utilizzato come sorgente di dati e, in particolare, il file `json` contenente i report strutturati secondo i livelli di innestamento precedentemente illustrati, rendono disponibili i dati relativi alla situazione del Coronavirus nelle Marche in una forma pulita. Questo significa che i dati della sorgente sono già stati sottoposti a operazioni di ETL. Infatti, essendo il repository di nostro interesse un repository pubblico, e avendo come scopo quello di rendere i dati disponibili all'utenza in un formato semplice per le analisi, esso prevede una struttura dei dati stessi predefinita e ben precisa.

Sappiamo come Power BI sia in grado di connettersi a diverse tipologie di sorgenti di dati, tra cui anche i file `json`. Questo ha facilitato notevolmente le operazioni di importazione, anche se, a causa della complicata struttura del file `json`, sono comunque state fatte diverse operazioni per manipolare la forma iniziale dei dati.

In particolare, per quanto riguarda l'importazione dei dati, è stata inizialmente clonato il repository di GitHub sopra menzionato, e il file ottenuto da tale operazione è stato salvato nella macchina locale in una directory `coronavirus-marche` (Figura 3.7), all'interno della quale è, ovviamente, presente il file `covid-19-marche.json`, che verrà usato in PowerBI Desktop.

3.3.1 Importazione dei dati

Dopo aver effettuato il download del repository, il file `covid-19-marche.json` è stato sottoposto ad un'attenta attività di ispezione al fine di verificare la correttezza e l'integrità dei dati, nonché l'omogeneità e la corenza della struttura del formato `json` ivi rappresentato e, soprattutto, per evitare eventuali errori e problematiche nei successivi passaggi di importazione e di analisi dei dati nel modello in Power BI. Tale attività di verifica, sebbene sia di grande importanza, non ha comportato alcun intervento nel file `json` di interesse, essendo già questo ben strutturato dall'autore del repository stesso.

Il passaggio successivo, invece, è proprio quello del caricamento dei dati, che si effettua direttamente all'interno di Power BI Desktop, grazie alla funzionalità "Recupera dati" che permette di scegliere, tra le tante tipologie di sorgenti di dati, anche il tipo di dato `json`.

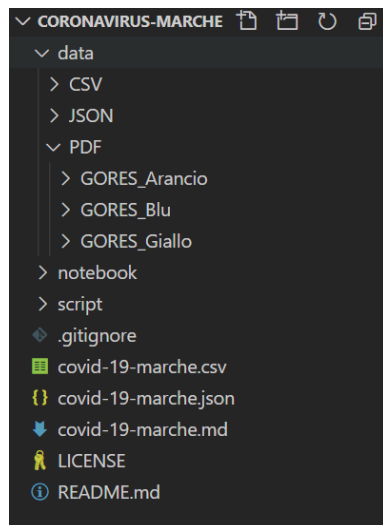


Figura 3.7. La directory `coronavirus-marche` con i file clonati dal repository di GitHub

Una volta selezionato il file di interesse, l’Editor di Power Query mostra la finestra riportata nella Figura 3.8, dove è possibile notare anche il primo passaggio della query, il quale descrive l’importazione del documento `json`.

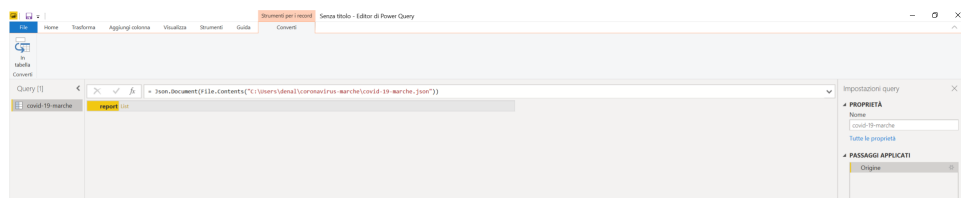


Figura 3.8. Caricamento del file `json` in Power BI Desktop

Come si evince da questa finestra, l’importazione di un file `json` in Power BI Desktop non è immediata. Il file di nostro interesse, come abbiamo già avuto modo di vedere, ha una struttura interna piuttosto complicata, essendo un array di “oggetti” complessi. Gli oggetti, infatti, rappresentano i report giornalieri, e le loro proprietà sono, a loro volta, degli oggetti. Proprio per tale motivo, appena caricato il file, l’Editor di Power Query mostra una sola riga composta dai campi “Report” e “List”. In particolare, “Report” rappresenta la radice del nostro file `json`, cioè indica il primo attributo all’interno di tale file, e, come sappiamo, il valore di tale attributo è un array di oggetti, che rappresentano i report dei singoli giorni.

A causa di tutto ciò, nell’Editor di Power Query, si è dovuta applicare inizialmente la funzionalità “In tabella”, presente nella barra superiore dell’editor stesso, che permette di convertire i record in tabella, potendo in questo modo “espandere” la “List” per ottenere una vera e propria tabella, che riportiamo in Figura 3.9 .

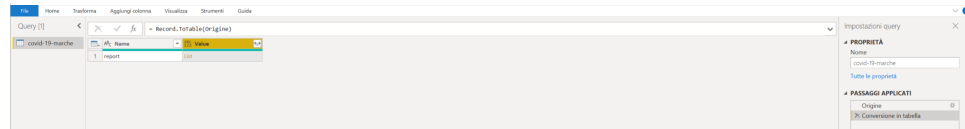


Figura 3.9. Funzionalità “In tabella” nell’Editor di Power Query

Ottenuta tale struttura, il passaggio successivo consiste nell’ “espandere” le singole colonne della tabella in più righe tramite la funzionalità “Espandi in nuove righe” dell’Editor di Power Query, applicabile alle colonne i cui valori sono di tipo complesso, proprio come nel nostro caso (Figura 3.10).

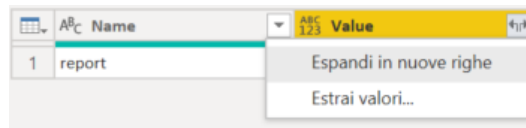


Figura 3.10. Funzionalità “Espandi in nuove righe” dell’Editor di Power Query

In Figura 3.11 riportiamo un esempio di applicazione di tale passaggio, che mostra la struttura che si ottiene a seguito dell’espansione della tabella illustrata in Figura 3.9.

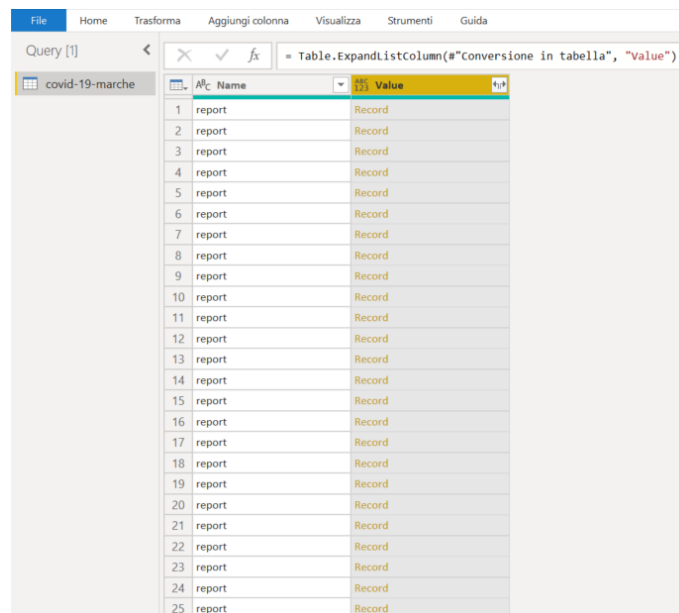


Figura 3.11. Esempio di espansione delle righe nell’Editor di Power Query

Questo passaggio è stato applicato ripetutamente fino ad espandere tutti i record

del file `json`, al fine di ottenere una tabella completa. La Figura 3.12 illustra la tabella finale risultante dall'applicazione di tutti i passaggi appena illustrati nell'Editor di Power Query.

Nome	Value.data	Value.tamponi.totali.test_effettuati	Value.tamponi.totali.casi_diagnosticati
1 report	02-25-2020	0	0
2 report	02-26-2020	0	0
3 report	02-27-2020	0	0
4 report	02-28-2020	0	0
5 report	02-29-2020	0	0
6 report	03-01-2020	0	0
7 report	03-02-2020	0	0
8 report	03-03-2020	0	0
9 report	03-04-2020	0	0
10 report	03-04-2020	0	0
11 report	03-04-2020	0	0
12 report	03-05-2020	0	0
13 report	03-06-2020	0	0
14 report	03-07-2020	0	0
15 report	03-07-2020	0	0
16 report	03-08-2020	0	0
17 report	03-09-2020	0	0
18 report	03-09-2020	0	0
19 report	03-09-2020	0	0
20 report	03-10-2020	0	0
21 report	03-10-2020	0	0
22 report	03-10-2020	0	0
23 report	03-10-2020	0	0
24 report	03-10-2020	0	0
25 report	03-10-2020	0	0
26			

Figura 3.12. Tabella risultante dall'espansione delle righe

Notiamo che fino a questo punto non è stata applicata alcuna operazione di modifica dei dati; si è semplicemente cercato di trasformare il file `json` in una struttura di tipo relazionale, adatta ai nostri studi successivi.

3.3.2 Modifica dei dati

La tabella riportata in Figura 3.12 non possiede ancora una struttura perfettamente adatta ad attività di analisi. Dalla figura si può infatti notare come tale tabella sia composta da ben 113 colonne, il che la rende piuttosto complicata da analizzare.

A questo punto, il successivo passaggio da applicare consiste nel rimuovere i dati che non risultano di nostro interesse. Ad esempio, sempre nella figura sopra riportata, si nota come i campi relativi a “Name” contengano informazioni completamente inutili al fine dell'analisi; pertanto, tale campo sarà rimosso.

Un aspetto molto importante in Power BI è la gestione dei tipi di dato delle colonne delle relazioni. La tabella sopra riportata presenta, per tutte le sue colonne, un tipo testuale (“Text”), anche se, in realtà, i valori di alcune colonne sono valori numerici di tipo intero. Power BI, infatti, in seguito al processo di importazione, ha automaticamente riconosciuto tali valori come di tipo testuale. È, però, importante sapere che il concetto di tipo di dato è fondamentale nella fase di analisi, per la costruzione di grafici, ma anche, e soprattutto, per la creazione di misure usando la sintassi DAX. Per risolvere questo problema di compatibilità, i formati di tutte le colonne relative a dati di tipo intero sono stati cambiati da “Text” a “Int”.

Molto importante in Power BI è, anche, la gestione delle date. È necessario infatti, per far leggere al tool una data, impostare questa da “Text” a “Date”, risolvendo in questo modo un altro problema di compatibilità.

Al fine di facilitare la lettura dei campi delle tabelle all’interno di Power BI, si è, inoltre, deciso di rinominare i nomi di tali campi con altri di più facile lettura e più intuitivi.

3.3.3 Creazione di nuove query

La tabella ottenuta in seguito a tutti i passaggi sopra descritti è sicuramente più semplice e più pulita rispetto a quella iniziale. Le nostre operazioni nell’Editor di Power Query, però, non sono ancora concluse.

A questo punto è, infatti, necessario semplificare ulteriormente la struttura di tale tabella (composta da ben 113 righe), creando ulteriori tabelle contenenti soltanto i dati relativi ad un particolare aspetto di interesse. Analizziamo ora nel dettaglio queste operazioni.

Tabella relativa ai decessi giornalieri

La Figura 3.13 mostra altri campi della nostra tabella iniziale, ovvero quella ottenuta dopo le operazioni di espansione dei record.

In particolare, si noti la presenza di numerosi valori di tipo “null”, ma anche di celle completamente vuote.

The screenshot shows the Power Query Editor interface. The main window displays a table with 26 rows and 3 columns. The columns are: 'comune_domicilio', 'provincia_domicilio', and 'pregresso_patolo'. The data is as follows:

comune_domicilio	provincia_domicilio	pregresso_patolo
1	null	null
2	null	null
3	null	null
4	null	null
5	null	null
6	null	null
7	PU	
8	null	null
9	PU	
10	AN	
11	PU	
12	null	null
13	null	null
14	PU	
15	AN	
16	PU	
17	PU	
18	CHE	
19	PU	
20	PU	
21	PU	
22	PU	
23	DNE	
24	PU	
25	PU	
26		

The right-hand pane shows the 'Impostazioni query' (Query Settings) for the current query. It includes sections for 'PROPRIETÀ' (Properties) and 'PASSAGGI APPLICATI' (Applied Steps). The 'PASSAGGI APPLICATI' list includes several steps such as 'Tabella Value.malati.quara...', 'Ridinate colonne', 'Rinominate colonne', 'Rimosse colonne1', 'Modificato tipo', 'Rimosse colonne2', 'Modificato tipo1', 'Rinominate colonne1', 'Rimosse colonne3', and 'Rinominate colonne2'.

Figura 3.13. Tabella con voci di tipo “null” e celle vuote

Nel nostro caso, tali valori fanno tutti riferimento alle informazioni relative ai decessi verificatisi nella Regione Marche a causa del Coronavirus. Per risolvere questo problema, ma anche per avere un modello più chiaro e semplice su cui poter

lavorare, si è deciso di creare una nuova query che permette di ottenere una nuova tabella contenente esclusivamente le informazioni relative ai decessi giornalieri. Tale query è stata semplicemente ottenuta ricopiando la query originale di partenza, ed eliminando tutte le colonne ad eccezione di quelle relative ai decessi. In seguito sono state eliminate tutte le voci di tipo “null”, inutili per il nostro studio, mediante la funzionalità di filtraggio delle righe dell’Editor di Power Query. Tale funzionalità ci ha, dunque, permesso di mantenere soltanto le righe con valori diversi da quello nullo, come si evince dalla Figura 3.14 che riporta la nuova tabella relativa ai dati dei decessi.

Id decesso	Data	Luogo Decesso	Sesso	Eta
1	02/03/2020	TI FANO	M	87
2	04/03/2020	MN	M	61
3	04/03/2020	INRCA PPI	M	84
4	04/03/2020	UTIC PS	M	75
5	07/03/2020	TI FANO	M	77
6	07/03/2020	AGAN	M	59
7	08/03/2020	MI MN	M	83
8	09/03/2020	URB MED	F	98
9	09/03/2020	MI MN	F	73
10	09/03/2020	RIA MN	M	86
11	10/03/2020	RIA MN	M	87
12	10/03/2020	MI MN	F	80
13	10/03/2020	MI MN	M	93
14	10/03/2020	MI FERMO	F	84
15	10/03/2020	OSP SENIGALLIA	M	67
16	10/03/2020	OSP SENIGALLIA	M	84
17	10/03/2020	OSP SENIGALLIA	M	85
18	10/03/2020	MI FERMO	F	83
19	11/03/2020	RIA MN	M	76
20	11/03/2020	URBINO MED	F	87
21	11/03/2020	UG AI MN	M	88
22	11/03/2020	RIA MN	M	73
23	12/03/2020	MI FERMO	M	70
24	12/03/2020	OSP SENIGALLIA	M	81
25	12/03/2020	OSP SENIGALLIA	M	56

Figura 3.14. Tabella relativa ai decessi da Covid-19 nella Regione Marche

In particolare, in questa seconda query, sono state applicate altre operazioni fondamentali per le nostre analisi successive. Tra queste, è importante sottolineare le operazioni di sostituzione dei valori dei campi relativi alla provincia di domicilio e al comune di domicilio delle persone decedute. Infatti, tali dati saranno utilizzati in seguito nelle nostre analisi per illustrare, mediante una visualizzazione basata su mappe di Power BI, i decessi per ogni provincia e per ogni comune della Regione Marche. La visualizzazione basata su mappe di Power BI, per poter geolocalizzare correttamente le provincie e i comuni, necessita delle notazioni ufficiali di quest’ultimi. I nomi delle provincie e dei comuni sono stati quindi sostituiti direttamente nell’Editor di Power Query con i nomi ufficiali, riconosciuti dalla visualizzazione stessa.

La Figura 3.15 riporta un esempio di passaggio applicato nell’Editor di Power Query per la sostituzione dei valori di una provincia.

Tali passaggi sono stati applicati sia per le provincie che per i comuni.

Questa seconda query ci ha consentito, in particolare, di separare in due relazioni diverse i dati a nostra disposizione. Tali relazioni sono:

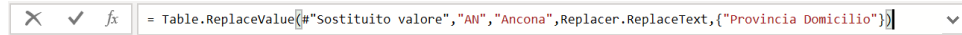


Figura 3.15. Esempio di sostituzione di valori nell’Editor di Power Query

- *Generale*: in questa tabella sono riportati i dati generali sulla situazione relativa al Coronavirus nella Regione Marche. Qui troviamo, infatti, le informazioni relative a tamponi, contagi, ricoveri e decessi nelle Marche, riportate giorno per giorno.
- *Decessi Dettaglio*: questa seconda tabella riporta, invece, esclusivamente i dati relativi alle persone decedute nella Regione Marche. In particolare vengono descritte le informazioni relative a ciascun decesso attraverso i campi “Id decesso”, “Data”, “Luogo Decesso”, “Sesso”, “Eta”, “Comune Domicilio”, “Provincia Domicilio” e “Patologie Pregresse”. È importante notare come in questa seconda tabella il campo “Data” non costituisce un valore di tipo “unique”, ovvero ci possono essere più tuple riportanti lo stesso valore in questo campo.

Tabella relative alle provincie

La tabella “Generale” contiene anche alcune informazioni che, però, sono specifiche per ogni singola provincia della Regione Marche: in tale tabella ci sono cioè, per alcune informazioni, come ad esempio quella sui malati totali, tante colonne per ogni provincia. Abbiamo, ad esempio, una colonna per i malati totali della provincia di Ancona, una per i malati totali della provincia di Macerata, etc.

Per semplificare ulteriormente tale tabella, ma anche per facilitare le analisi in Power BI che fanno riferimento ai dati provincia per provincia, si è deciso di creare una nuova query per ognuna delle provincie marchigiane, riportante i dati relativi esclusivamente alla singola provincia (Figura 3.16).

Le varie query così create, sono state, successivamente, “accodate” in una nuova query che definisce la relazione “Provincie”. Tale relazione, dunque, riunisce le relazioni sulle singole provincie in un’unica tabella, fondamentale per realizzare le analisi e i confronti in Power BI sui dati provincia per provincia.

In particolare, per realizzare questo passaggio, è stata utilizzata la funzionalità “Accoda query” presente nella scheda “Home” dell’Editor di Power Query.

Accodamento di query

Analizziamo ora brevemente la funzionalità di accodamento messa a disposizione dall’Editor di Power Query.

In tale editor 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.

In particolare, si possono eseguire due tipi di operazioni di accodamento:

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

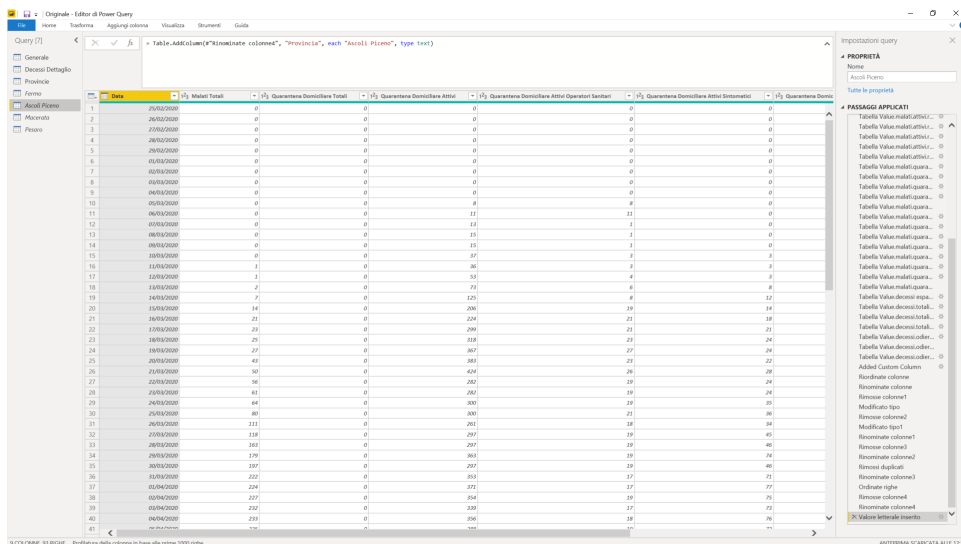


Figura 3.16. Tabella relativa alla provincia di Ascoli Piceno

Nel nostro caso è stata utilizzata la funzionalità di “Accodamento intermedio”. Essa si attiva dalla barra multifunzione dell’Editor di Power Query, semplicemente cliccando sull’apposita voce del menù. Questa operazione determina l’apertura della finestra di dialogo “Accoda” (Figura 3.17), dalla quale è possibile selezionare la prima e la seconda tabella (query) da accodare.

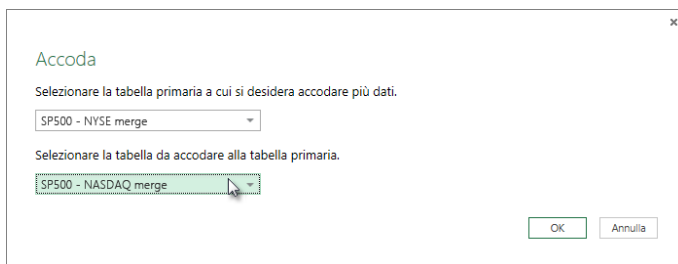


Figura 3.17. Finestra di dialogo “Accoda” nell’Editor di Power Query

3.3.4 Caricamento dei dati puliti

L’ultima operazione da effettuare è quella di salvataggio delle modifiche. Per farlo, è sufficiente cliccare su “Chiudi e applica” nel menù “Home” dell’Editor di Power Query, e l’operazione di modifica dei dati sarà così conclusa.

Le tabelle ottenute vengono, dunque, caricate in Power BI Desktop, e sono rese disponibili per l’utilizzo, cioè per ottenere tutti i grafici desiderati per la fase di studio.

È importante ribadire come Power BI permetta di modificare, in qualsiasi momento, le tabelle, semplicemente cliccando su “Trasforma dati” dalla barra “Home” del menù principale. In questo modo, si apre l’Editor di Power Query, in cui è anche possibile visualizzare lo storico delle modifiche.

Realizzazione di cruscotti relativi ai decessi

In questo capitolo entriamo nel dettaglio delle attività svolte in Power BI Desktop. In particolare, all'interno del capitolo visualizzeremo, analizzeremo e descriveremo le informazioni relative all'andamento dei decessi da Covid-19 all'interno della Regione Marche e nelle sue province. A tal fine faremo uso delle differenti tipologie di visualizzazione che Power BI mette a disposizione.

4.1 Dashboard introduttiva sui decessi

Iniziamo col fornire una panoramica generale sui decessi avvenuti nella Regione Marche e riconducibili alla pandemia da Covid-19. Ricordiamo come il modello dei dati ottenuto dalle operazioni descritte nel *Capitolo 3* contenga anche una tabella relativa al dettaglio dei decessi giornalieri. La Figura 4.1 illustra la struttura di tale tabella.

Si tratta, in particolare, di una tabella che riporta le caratteristiche di ogni persona deceduta. Si noti come il riquadro “Campi” di tale figura mostra non solo i campi della tabella, ma anche alcune misure ottenute utilizzando le formule DAX, facilmente riconoscibili, poiché rappresentate mediante il simbolo della calcolatrice.

Le misure calcolate ci sono servite per arricchire la nostra relazione, senza dover necessariamente realizzare query nell'Editor di Power Query, ma anche per fornire informazioni fondamentali ai fini delle nostre analisi.

In particolare, sono state create quattro misure legate alla relazione “Decessi Dettaglio”; tali misure sono:

- **Decessi Totali**: questa misura effettua, semplicemente, il conteggio delle righe della nostra relazione, al fine di ottenere il numero esatto di decessi avvenuti nella regione (Figura 4.2).
- **Parziale Decessi**: questa misura sfrutta il campo “Data” della nostra relazione, e permette di calcolare il numero dei decessi ad una certa data (Figura 4.3).
- **ConPatologie**: questa misura conta il numero di persone decedute nelle quali, al momento del decesso, sono state riscontrate una o più patologie. Anche questa misura effettua un semplice conteggio delle righe (Figura 4.4).

Id decesso	Data	Luogo Decesso	Sesso	Eta	Comune Domicilio	Provincia Domicilio	Patologie Pregresse
1	lunedì 2 marzo 2020	TI FANO	M	87	FANO	Pesaro	1
2	mercoledì 4 marzo 2020	MN	M	61	PESARO	Pesaro	1
3	mercoledì 4 marzo 2020	INRCA PPI	M	84	ANCONA	Ancona	1
4	mercoledì 4 marzo 2020	LITIC PS	M	75	FANO	Pesaro	1
5	sabato 7 marzo 2020	TI FANO	M	77	PESARO	Pesaro	1
6	sabato 7 marzo 2020	ADAN	M	59	ANCONA	Ancona	1
7	domenica 8 marzo 2020	MI MN	M	83	PESARO	Pesaro	1
8	lunedì 9 marzo 2020	URB MED	F	98	PESARO	Pesaro	1
9	lunedì 9 marzo 2020	MI MN	F	73	TERRE ROVERESCHE	Pesaro	1
10	lunedì 9 marzo 2020	RIA MN	M	86	PESARO	Pesaro	1
11	martedì 10 marzo 2020	RIA MN	M	87	PESARO	Pesaro	1
12	martedì 10 marzo 2020	MI MN	F	80	FANO	Pesaro	1
13	martedì 10 marzo 2020	MI MN	M	93	CARTOCETO	Pesaro	1
14	martedì 10 marzo 2020	MI FERMO	F	84	MONTE CERIGNONE	Pesaro	1
15	martedì 10 marzo 2020	OSP SENIGALLIA	M	67	PESARO	Pesaro	1
16	martedì 10 marzo 2020	OSP SENIGALLIA	M	84	GRADARA	Pesaro	1
17	martedì 10 marzo 2020	OSP SENIGALLIA	M	85	PESARO	Pesaro	1
18	martedì 10 marzo 2020	MI FERMO	F	83	PORTO RECANATI	Macerata	1
19	mercoledì 11 marzo 2020	RIA MN	M	76	MONTELABBATE	Pesaro	1
20	mercoledì 11 marzo 2020	URBINO MED	F	87	PESARO	Pesaro	1
21	mercoledì 11 marzo 2020	UO AI MN	M	88	PESARO	Pesaro	1

Figura 4.1. Tabella relativa al dettaglio dei decessi da Covid-19

- *SenzaPatologie*: questa misura esegue il calcolo complementare di quella precedente, facendo, cioè, riferimento alle persone decedute e senza patologie pregresse (Figura 4.5)

```
Decessi Totali = COUNT('Decessi Dettaglio'[Id decesso])
```

Figura 4.2. Misura calcolata “Decessi Totali”

```
1 Parziale decessi =
2 CALCULATE(
3     [Decessi Totali],
4     FILTER(
5         ALLSELECTED('Decessi Dettaglio'[Data]),
6         ISONORAFTER('Decessi Dettaglio'[Data], MAX('Decessi Dettaglio'[Data]), DESC)
7     )
8 )
```

Figura 4.3. Misura calcolata “Parziale Decessi”

```
1 ConPatologie = CALCULATE(COUNT('Decessi Dettaglio'[Id decesso]), 'Decessi Dettaglio'[Patologie Pregresse] = 1)
```

Figura 4.4. Misura calcolata “ConPatologie”


```
1 SenzaPatologie = CALCULATE(COUNT('Decessi Dettaglio'[Id decesso]), 'Decessi Dettaglio'[Patologie Pregresse] = 0)
```

Figura 4.5. Misura calcolata “SenzaPatologie”

Avendo a disposizione la relazione relativa al dettaglio dei decessi e le misure calcolate appena illustrate, si è costruito un primo cruscotto introduttivo sulla situazione dei decessi nelle Marche. Tale cruscotto riporta, infatti, i grafici e le schede che riassumono i campi della nostra relazione, semplicemente per dare una visione d’insieme sulla situazione dei decessi (Figura 4.6).

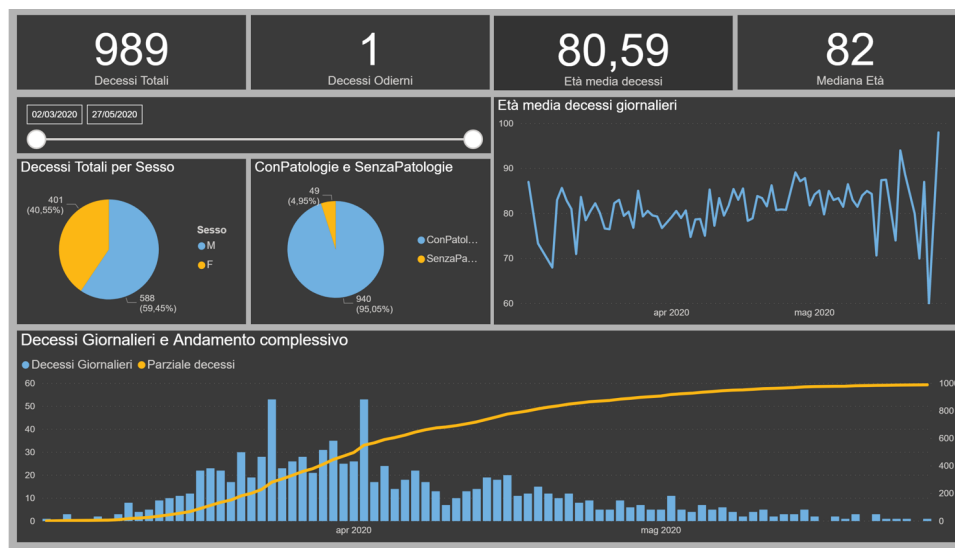


Figura 4.6. Dashboard introduttiva sui decessi

In tale cruscotto, così come anche in quelli che analizzeremo nelle sezioni successive, sono presenti, nella parte superiore, delle semplici “Schede”, che permettono di mostrare direttamente i numeri più importanti, in questo caso quelli legati ai decessi.

Tali schede riportano, infatti, il numero dei decessi totali, il numero dei decessi odierni, l’età media dei decessi e la mediana dell’età (nella figura riportata tali numeri si riferiscono al dato dei decessi della regione aggiornati al 27 Maggio 2020).

Al di sotto delle schede abbiamo, 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 patologie pregresse e quelle senza patologie pregresse.

Tutte queste informazioni verranno, comunque, illustrate nel dettaglio nei cruscotti che vedremo in seguito.

Un aspetto interessante da sottolineare, e che tornerà utile anche nei cruscotti successivi, è l’utilizzo di un filtro per visualizzare il dato relativo ai decessi del giorno corrente (“Decessi Odierni”). Per realizzare la scheda che riporta questo numero, infatti, è stata utilizzata la misura calcolata “Decessi Totali” e, all’oggetto visivo scheda, è stato applicato un filtro a livello di oggetto visivo per quanto riguarda il

campo “Data” della tabella (Figura 4.7). Si tratta di un filtro avanzato che seleziona soltanto il valore più recente del campo data, permettendo, così, di visualizzare esclusivamente i decessi relativi all’ultimo giorno disponibile, nel nostro caso il 27 Maggio 2020.

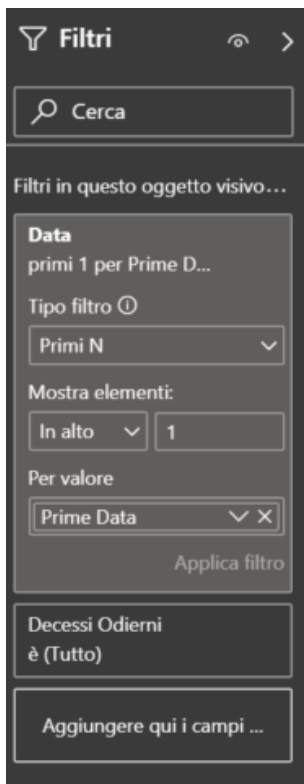


Figura 4.7. Il riquadro Filtri con i filtri applicati all’oggetto visivo scheda

Particolarmente significativo, in questo cruscotto introduttivo, 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, di ben 80,59 anni. Non approfondiamo, in questo momento, l’analisi sull’età dei deceduti, poiché questa verrà effettuata in seguito in un cruscotto realizzato appositamente per tale scopo.

Infine, in questo cruscotto, viene riportato, anche, un grafico misto, in cui il grafico a linee rappresenta la crescita dei decessi, mentre l’istogramma mostra i decessi giornalieri. Tale grafico evidenzia chiaramente come il picco più alto della pandemia all’interno della Regione Marche, per quanto riguarda il numero di morti giornalieri, si sia raggiunto tra il 24 Marzo 2020 e il 2 Aprile 2020 con 53 morti, corrispondenti ai due picchi ben evidenti dell’istogramma. La settimana precedente a quella relativa al picco, mostra, invece, una crescita graduale dei decessi giornalieri, mentre, a partire dalla settimana successiva, si nota un crollo evidente del numero

di decessi.

Anche il grafico a righe, riportante l'andamento complessivo, illustra una situazione analoga: la crescita dei decessi tra i due giorni del picco ha un andamento pressoché esponenziale, salvo poi appiattirsi sempre di più dopo il 2 Aprile 2020, circa 20 giorni dopo la data di inizio del lockdown imposto dal premier Giuseppe Conte a partire dal giorno 11 Marzo 2020.

4.2 Distribuzione geografica delle morti

Il primo approfondimento della nostra analisi sui decessi da Covid-19 avvenuti nella Regione Marche riguarda la distribuzione geografica delle morti.

Nel Capitolo 3 abbiamo evidenziato nel dettaglio le operazioni di sostituzione dei campi delle relazioni, realizzate nell'Editor di Power Query. Tale operazioni hanno, in particolare, coinvolto i campi i cui valori rappresentano i nomi delle province e dei comuni marchigiani. Sappiamo, infatti, che la visualizzazione basata sulle mappe di Power BI considera le notazioni ufficiale dei luoghi geografici. Queste operazioni risulteranno, perciò, particolarmente utili in questa sezione, in cui andremo ad analizzare le singole province e i singoli comuni delle Marche.

4.2.1 Distribuzione per provincia

La pandemia da Covid-19 ha colpito tutta la regione Marche provocando decessi in ognuna delle sue province. Il cruscotto riportato in Figura 4.8 illustra la situazione relativa ai decessi nelle province aggiornata al 27 Maggio 2020. Per realizzare tale cruscotto, e anche quelli successivi illustrati in questa sezione, è stata usata la relazione "Province", la cui struttura è stata analizzata nel capitolo precedente.

Osservando la figura emerge chiaramente come la provincia di Pesaro Urbino sia di gran lunga la più colpita dal virus. Dei 989 decessi avvenuti nelle Marche, infatti, più del 50% sono riconducibili proprio a tale provincia.

Il primo grafico a sinistra illustra la crescita dei decessi nelle 5 province delle Marche e, anche qui, si può facilmente notare il grande impatto di Pesaro nel numero di morti dell'intera regione.

Molto interessante è, anche, il grafico a righe sulla destra del cruscotto, riportante, per ogni provincia, il rapporto percentuale tra il numero di decessi della provincia stessa e il numero di decessi totale della regione. Quest'ultimo grafico evidenzia, ancora una volta, il grande impatto della provincia di Pesaro Urbino. Infatti, la percentuale dei decessi di questa provincia rispetto al totale delle Marche, si attesta a circa il 55%, comunque in decrescita rispetto alla situazione della prima metà del mese di Marzo 2020, in cui tale percentuale era addirittura superiore all'80%. Questo picco, che potrebbe sembrare discordante con l'andamento generale della curva, è dovuto al fatto che focolaio dell'epidemia all'interno delle Marche sia stato proprio nella provincia di Pesaro.

Per quanto riguarda quest'ultimo grafico, è stata utilizzata una particolare colonna calcolata, denominata "Decessi Provincia/Decessi Totali", che esegue proprio il calcolo percentuale sopra riportato per ognuna delle province e per ognuno dei giorni dall'inizio dell'epidemia (Figura 4.9)

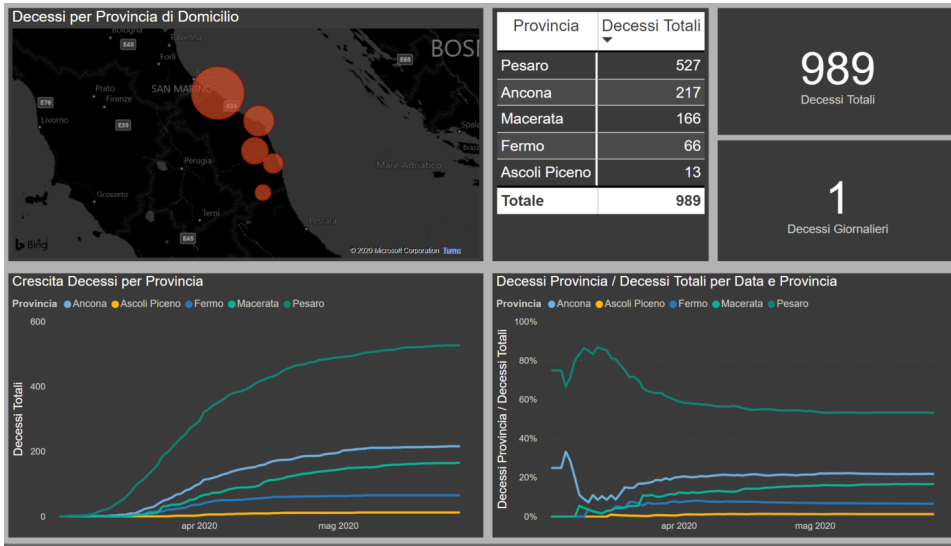


Figura 4.8. Cruscotto relativo al numero di decessi nelle province marchigiane

```

1 Decessi Provincia / Decessi Totali =
2 var Provincia = Provincie[Provincia]
3
4 var DateTime = Provincie[Data]
5
6
7 var PreviousDate = CALCULATE(MAX(Provincie[Data]), FILTER(Provincie, Provincie[Provincia] = Provincia && Provincie[Data] <= DateTime))
8 var TotaliProvincia = CALCULATE(MAX(Provincie[Decessi Totali]), FILTER(Provincie, Provincie[Provincia] = Provincia && Provincie[Data] <= PreviousDate))
9
10
11 var Tot = CALCULATE(SUM(Provincie[Decessi Totali]), FILTER(Provincie, Provincie[Data] = DateTime))
12 return TotaliProvincia / Tot

```

Figura 4.9. Formula DAX della colonna calcolata relativa all’impatto dei decessi di ciascuna provincia sul totale della regione

4.2.2 Distribuzione per comune

Analizziamo, ora, comune per comune la situazione relativa ai decessi avvenuti nelle Marche. Il cruscotto riportato in Figura 4.10 illustra, mediante una visualizzazione a mappa, la distribuzione geografica delle morti da Covid-19 nella regione. Si nota subito come siano molti i comuni marchigiani nei quali è stato segnalato almeno un residente deceduto a causa del Covid-19.

Il cruscotto sotto riportato contiene, anche, la tabella riportante il numero di decessi per ciascun comune. Tale tabella, in realtà, funge anche da filtro dei dati; cioè, se si seleziona un comune, tutti i grafici del cruscotto riportano immediatamente i dati relativi al solo comune selezionato. Tale comportamento, presente in tutti i cruscotti realizzati nel nostro studio in cui si analizzano province e regioni, è mostrato in Figura 4.11, dove, in particolare, si è selezionato il comune di Pesaro. Un comportamento analogo si ottiene anche selezionando il comune desiderato in uno qualsiasi degli altri grafici presenti all’interno del cruscotto.

Ritornando, ora, all’analisi del cruscotto riportato in Figura 4.10, vediamo come in questo ci siano altri due grafici a istogramma. L’istogramma alla sinistra mostra come ben 18 comuni della regione abbiano registrato un numero di decessi superiore

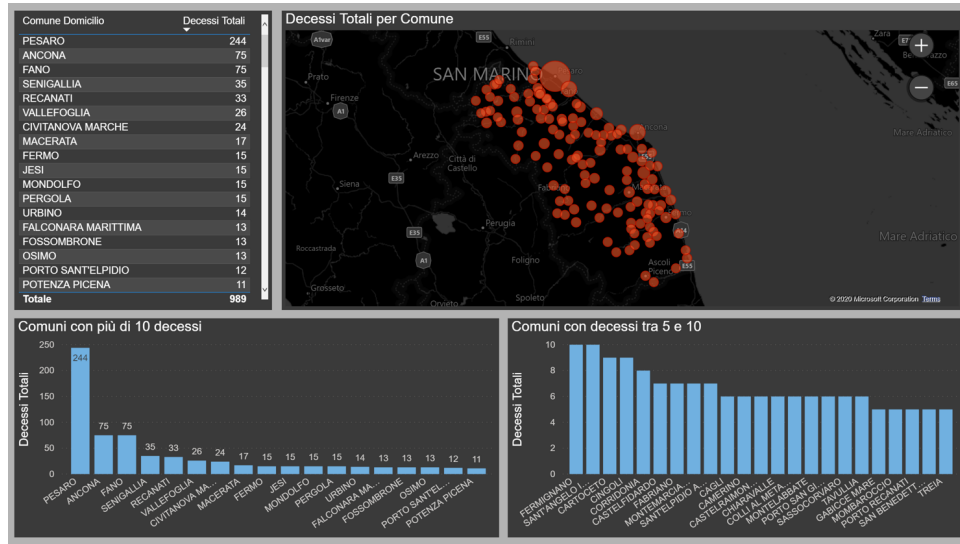


Figura 4.10. Cruscotto relativo al numero di decessi nei comuni marchigiani

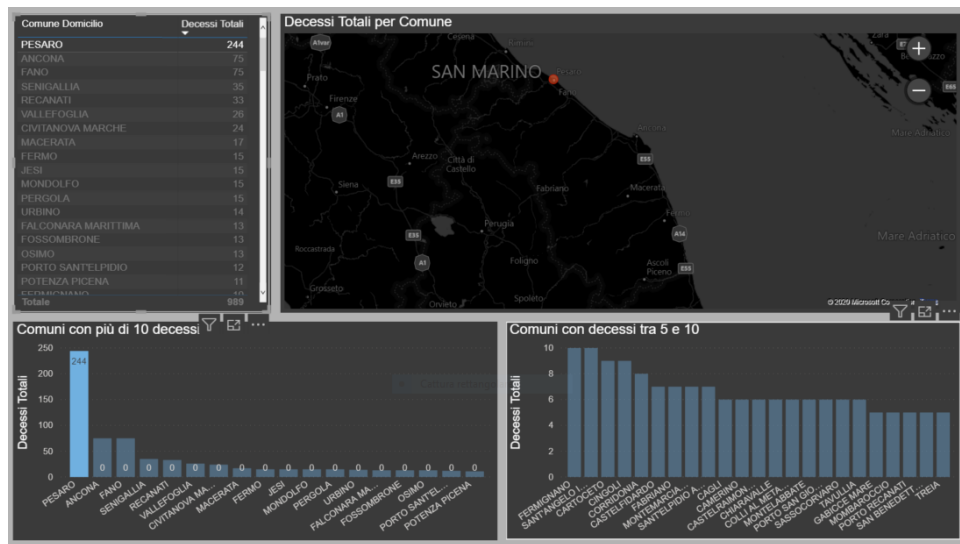


Figura 4.11. Esempio di cruscotto con filtro dei dati in base al comune selezionato

a 10. Il secondo istogramma mostra, invece, i comuni che hanno registrato un numero di decessi compreso tra 5 e 10; questi ultimi sono ben 23. Anche questi istogrammi mostrano sempre come la provincia di Pesaro Urbino, con i suoi comuni, sia stata la più colpita per quanto riguarda il numero di morti di persone positive al Covid-19.

4.3 Analisi sulla crescita dei decessi

Concentriamoci, ora, sull'aspetto relativo alla velocità e al modo con cui il numero dei decessi legati al Covid-19 sia cresciuto nel corso delle settimane, sia all'interno dell'intera Regione Marche, che nelle sue singole province.

4.3.1 Confronto temporale

In Figura 4.12 è riportato un cruscotto che, mediante un grafico ad area, mostra una sorta di confronto nel tempo del numero dei decessi. Questo cruscotto, cioè, permette di mettere a confronto, per ciascun giorno riportato nell'asse delle ascisse, il numero di decessi avvenuti quello stesso giorno, il numero di decessi avvenuti 14 giorni prima, e il numero dei decessi avvenuti 1 mese prima.

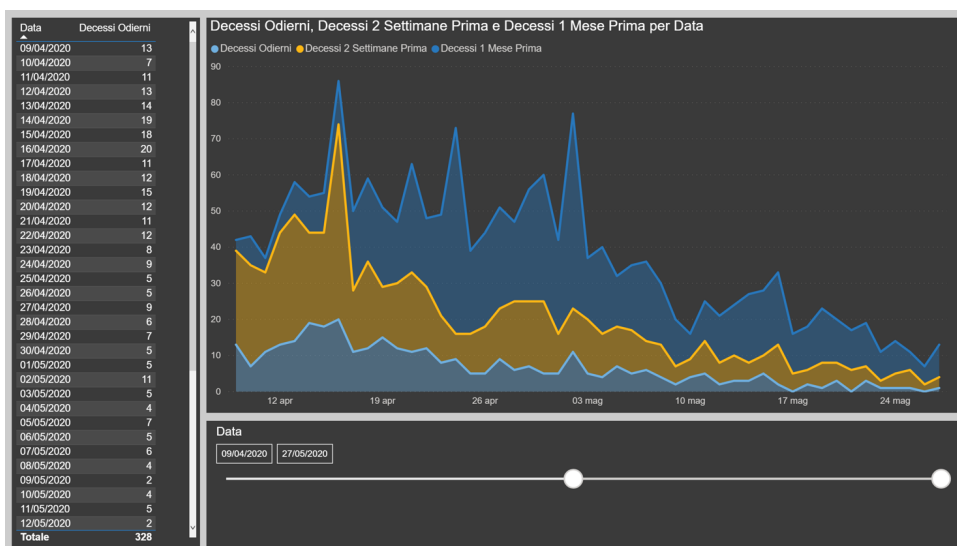


Figura 4.12. Cruscotto relativo al confronto temporale del numero dei decessi

Per ottenere questo particolare grafico, è stata utilizzata una particolare funzione DAX relativa al tipo di dato “Date”, che permette di ottenere il valore di una colonna anticipando la data di un certo numero di giorni, mesi, o anche anni.

In Figura 4.13 è riportata una delle misure realizzate usando tale funzione, che è la funzione DATEADD, molto utilizzata nell'ambiente DAX e in Power BI per eseguire confronti tra periodi diversi. In particolare, tale funzione restituisce una tabella che contiene una colonna di date, spostate in avanti o in indietro nel tempo in base al numero specificato di intervalli rispetto alle date nel contesto corrente. Infatti, il terzo parametro della funzione rappresenta proprio l'intervallo in base al quale spostare le date, e può assumere i valori “year”, “quarter” e “month”.

Ritornando al cruscotto in Figura 4.12, esso è interessante da analizzare poiché mette in evidenza il miglioramento generale dell'epidemia per quanto riguarda il

```
1 Decessi 1 Settimana Prima = CALCULATE(SUM(Generale[Decessi Odierni]), DATEADD(Generale[Data], -7, DAY))
```

Figura 4.13. Esempio di applicazione della funzione DAX DATEADD

numero di morti. In particolare, tale grafico permette, anche, di valutare l'efficacia delle misure di contenimento adottate dal governo per combattere la pandemia. Il grafico mostra con grande evidenza come i decessi del mese precedente siano di gran lunga superiori in numero rispetto a quelli odierni, e tale differenza è particolarmente marcata nei giorni compresi tra la seconda metà di Aprile e la prima metà di Maggio. Il grafico mostra, dunque, un trend perfettamente decrescente del numero dei decessi, che nel giro di sole due settimane ha portato ad una notevole diminuzione degli stessi.

4.3.2 Crescita dei decessi per provincia

Abbiamo in precedenza analizzato grafici relativi alla situazione delle province marchigiane, senza però tener conto del numero di abitanti di quest'ultime, parametro fondamentale per valutare l'incidenza del virus Covid-19. Ad esempio, abbiamo in precedenza visto come la provincia di Pesaro Urbino sia la prima per numero di decessi, davanti anche ad Ancona, anche se quest'ultima ha un numero di abitanti di gran lunga superiore (secondo i dati dell'ISTAT, Ancona ha circa 471.228 abitanti, mentre Pesaro ne ha circa 358.886). Risulta, perciò, importante effettuare un'analisi tenendo conto anche della popolazione delle singole province. A tale scopo, sono state effettuate una serie di operazioni in Power BI all'interno della relazione "province".

Inizialmente, si sono recuperati, direttamente dal sito dell'ISTAT (http://dati.istat.it/Index.aspx?DataSetCode=DCIS_POPRES1), i dati relativi alla popolazione residente nelle province della Regione Marche (il sito fornisce i dati aggiornati al 1 Gennaio 2020). Con tali dati sono state, poi, create delle misure rapide in Power BI (Figura 4.14), semplicemente per avere i dati relativi al numero di abitanti delle province direttamente tra i campi in visualizzazione *Report* nella nostra area di lavoro in Power BI Desktop.

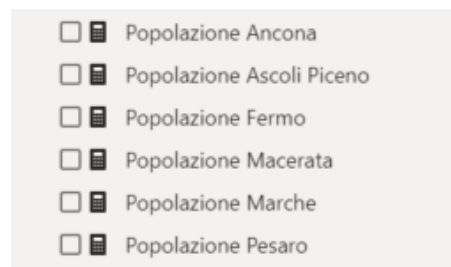


Figura 4.14. Misure calcolate relative agli abitanti delle province marchigiane

Successivamente, sono state realizzate altre 6 misure, una per ciascuna provincia marchigiana, per calcolare il numero di decessi ogni 100.000 abitanti di ciascuna provincia (Figura 4.15).

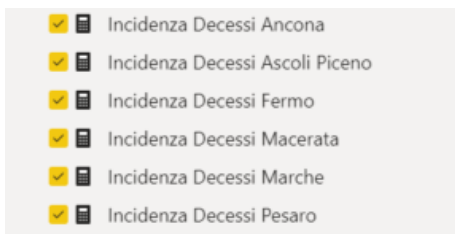


Figura 4.15. Misure calcolate sui decessi ogni 100.000 abitanti nelle province marchigiane

Queste misure rappresentano un indice che rappresenta il cosiddetto “tasso cumulativo grezzo di mortalità”, che si ottiene semplicemente facendo il rapporto tra il numero di decessi di ciascuna provincia e la sua popolazione residente, e moltiplicando tale risultato per 100.000. In Figura 4.16 è riportata la misura DAX effettuata per realizzare tale calcolo relativamente alla Provincia di Ancona; il calcolo per le altre province è del tutto analogo.

```
1 Incidenza Decessi Ancona = var sommaancona = CALCULATE(SUM(Province[Decessi Totali]), Province[Provincia] = "Ancona") return sommaancona/[Popolazione Ancona] * 100000
```

Figura 4.16. Calcolo del tasso cumulativo grezzo di mortalità per la provincia di Ancona

Una volta eseguite queste operazioni preliminari, si è realizzato il cruscotto riportato in Figura 4.17, che illustra l’andamento del tasso grezzo di mortalità per ciascuna provincia marchigiana e anche per l’intera regione.

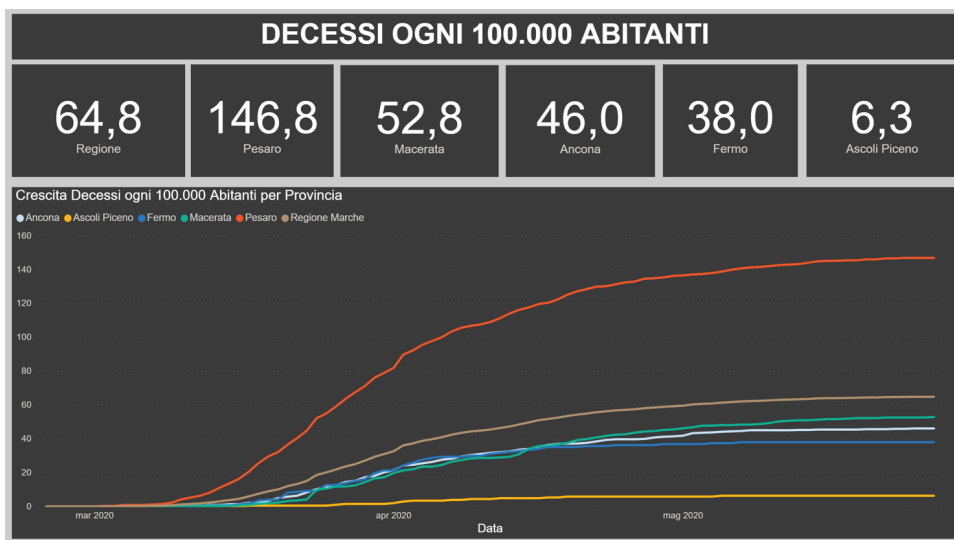


Figura 4.17. Cruscotto relativo al numero di decessi ogni 100.000 abitanti nelle province marchigiane

Nella parte superiore del cruscotto sono riportate le schede relative ai numeri esatti ed aggiornati all'ultimo giorno disponibile del tasso di mortalità per ciascuna provincia e, anche, per la regione.

Questi numeri ci fanno capire la gravità degli effetti dell'epidemia nella provincia di Pesaro Urbino, che riporta un tasso di mortalità di 146.8 morti ogni 100.000 abitanti, circa tre volte quello di Macerata, seconda, con un tasso di mortalità pari a 52.8.

La provincia di Pesaro Urbino rappresenta, come si può ben notare da questi dati, una delle zone italiane con il più alto numero di casi positivi ogni 100.000 abitanti, dato che la porta ad essere annoverata anche dai dati dell'ISTAT come una delle 38 province più colpite dall'epidemia da Covid-19, ed unica provincia a non essere nel Nord Italia.

Infine, la Regione Marche, con 64.8 morti ogni 100.000 abitanti, risulta essere una delle regioni più colpite a livello nazionale, come si evince anche dalla Figura 4.18, che riporta il tasso di mortalità grezzo in alcune delle regioni italiane maggiormente colpite dall'epidemia.

COVID-19 agg. 27 maggio														
	ITALIA	ABRUZZO	P.A. BOLZANO	EMILIA ROMAGNA	FRILSI VENEZIA GIULIA	LIGURIA	LOMBARDIA	MARCHE	PIEMONTE	TOSCANA	PROV. TRENTO	VALLE D'AGOSTA	VENETO	
Popolazione milioni	60,360	1,312	0,531	4,459	1,215	1,551	10,061	1,525	4,356	3,730	0,541	0,126	4,906	
Tasso di mortalità cumulativo grezzo (ogni 100.000 residenti)	54,79	30,50	54,78	91,56	27,24	92,74	158,58	65,30	87,87	27,54	85,38	113,79	38,63	
Numero totale di decessi	33072	400	291	4083	331	1438	15954	996	3828	1027	462	143	1895	
Tasso di mortalità giornaliero grezzo (ogni 100.000 residenti)	0,19	0,00	0,00	0,16	0,16	0,45	0,58	0,00	0,37	0,16	0,18	0,00	0,18	
Numero giornaliero di decessi	117	0	0	7	2	7	58	0	16	6	1	0	9	

Figura 4.18. Tabella con dati sui decessi delle regioni italiane più colpite dal Covid-19

4.4 Analisi su età, sesso e patologie dei decessi

Analizziamo, ora, nel dettaglio, altri aspetti relativi alle morti di persone risultate positive al Covid-19 nella Regione Marche. Per analizzare questa tematica è stato realizzato il cruscotto riportato in Figura 4.19, che presenta un'analisi basata sulla mediana dell'età dei decessi, riportata nella scheda con il testo colorato di rosso.

Tale cruscotto è stato appositamente suddiviso in tre sezioni verticali come di seguito specificato:

- La prima sezione illustra dati e grafici relativi al totale dei decessi. In particolare viene riportato il numero totale dei decessi della regione, il numero delle persone decedute con patologie pregresse e il numero di quelle senza patologie pregresse. Il primo grafico a torta rappresenta la ripartizione dei deceduti tra quelli con una o più patologie e quelli senza patologie. Il secondo grafico rappresenta la ripartizione dei deceduti tra maschi e femmine.
- La seconda sezione rappresenta sempre gli stessi dati della prima, facendo, però, riferimento alle sole persone decedute e con un'età inferiore alla mediana.

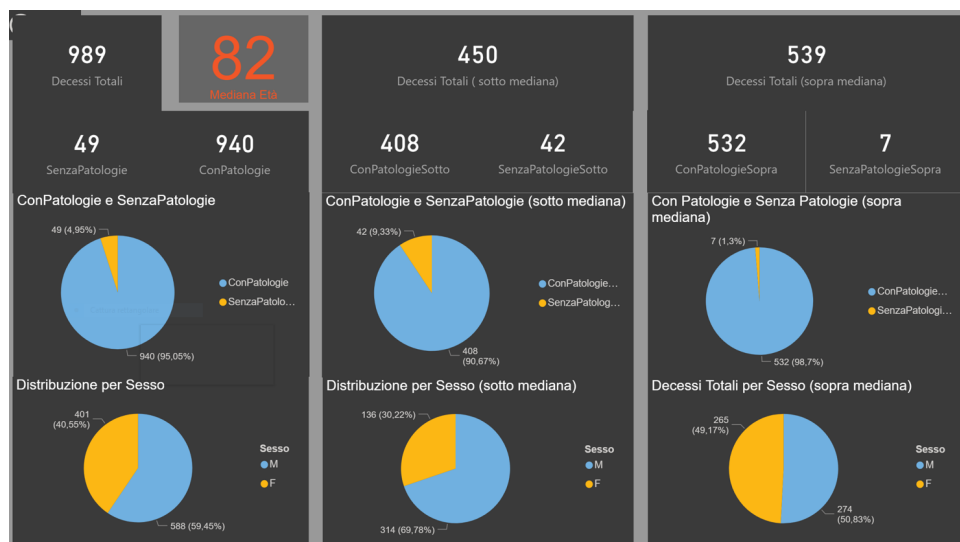


Figura 4.19. Cruscotto relativo all'analisi di età, sesso e patologie dei decessi

- La terza sezione, così come la seconda, rappresenta sempre le stesse informazioni della prima, considerando, però, le sole persone decedute con un'età superiore alla mediana.

Per realizzare questo cruscotto, sono state create due tabelle usando le formule DAX; tali tabelle sono state realizzate direttamente nel modello importato in Power BI Desktop e, quindi, senza la creazione di un'apposita query nell'Editor di Power Query. Le due tabelle sono "DecessiSottoMediana" e "DecessiSopraMediana" e si riferiscono, rispettivamente, ai dati delle persone decedute con un'età inferiore alla mediana, e ai dati dei decessi di persone con età superiore alla mediana. In Figura 4.20 riportiamo soltanto la prima delle due tabelle, avendo la seconda una struttura perfettamente analoga. Tale relazione, inoltre, contiene anche delle misure calcolate, le stesse viste nella tabella "Decessi Dettaglio", per quanto riguarda il conteggio delle persone decedute con patologie e quelle senza patologie. Queste misure, ovviamente, sono state effettuate in entrambe le relazioni che stiamo analizzando.

Ritornando al cruscotto di Figura 4.19, siamo ora in grado di analizzare nel dettaglio le tre sezioni in cui quest'ultimo è suddiviso. In particolare, vediamo come, con 989 morti (dato aggiornato al 27 Maggio 2020) la regione Marche si inserisce, sicuramente, tra le regioni italiane più colpite dall'epidemia, subito dopo le regioni più colpite del Nord Italia.

Inoltre, abbiamo in precedenza visto come l'età media dei deceduti nelle Marche sia molto alta, circa 80.59 anni. La mediana, come facilmente prevedibile, assume un valore ancora più alto, di ben 82 anni.

Questi dati sono perfettamente in linea con il dato nazionale: secondo un rapporto dell'ISTAT (<https://www.epicentro.iss.it/coronavirus/sars-cov-2-decessi-italia#1>), infatti, l'età media dei pazienti deceduti e positivi al Covid-19 è di 80 anni; la mediana, invece, è di 82 anni. Il dato conferma, dunque, come il virus sia letale soprattutto tra la popolazione più anziana.

MeseAnno	Data	Lungo Decesso	Sesso	Età	Comune Domestico	Provincia Domestico	Patologie Progresse
12/2019	18/02/2020	00:00:00	ADRIAN PESARO	F	27	PESARO	0
12/2019	18/02/2020	00:00:00	ADRIAN PESARO	M	50	PESARO	1
12/2019	18/02/2020	00:00:00	ADRIAN PESARO	F	61	PESARO	1
1/2020	20/03/2020	00:00:00	ADRIAN PESARO	M	75	PESARO	1
1/2020	20/03/2020	00:00:00	ADRIAN PESARO	M	66	PESARO	1
1/2020	20/03/2020	00:00:00	ADRIAN PESARO	M	67	PESARO	1
1/2020	20/03/2020	00:00:00	ADRIAN PESARO	M	78	PESARO	1
1/2020	22/03/2020	00:00:00	ADRIAN PESARO	F	67	PESARO	1
1/2020	22/03/2020	00:00:00	ADRIAN PESARO	F	74	PESARO	1
1/2020	23/03/2020	00:00:00	ADRIAN PESARO	M	79	PESARO	1
1/2020	23/03/2020	00:00:00	ADRIAN PESARO	F	77	PESARO	1
1/2020	24/03/2020	00:00:00	ADRIAN PESARO	M	57	PESARO	1
1/2020	24/03/2020	00:00:00	ADRIAN PESARO	M	78	PESARO	1
1/2020	24/03/2020	00:00:00	ADRIAN PESARO	M	79	PESARO	1
1/2020	30/03/2020	00:00:00	ADRIAN PESARO	M	74	PESARO	1
1/2020	30/03/2020	00:00:00	ADRIAN PESARO	M	64	PESARO	1
1/2020	30/03/2020	00:00:00	ADRIAN PESARO	M	75	PESARO	1
1/2020	31/03/2020	00:00:00	ADRIAN PESARO	F	78	PESARO	1
1/2020	31/03/2020	00:00:00	ADRIAN PESARO	M	67	PESARO	1
1/2020	31/03/2020	00:00:00	ADRIAN PESARO	M	78	PESARO	1
1/2020	31/03/2020	00:00:00	ADRIAN PESARO	M	80	PESARO	1
1/2020	31/03/2020	00:00:00	ADRIAN PESARO	M	80	PESARO	1
1/2020	02/04/2020	00:00:00	ADRIAN PESARO	F	79	PESARO	1
1/2020	02/04/2020	00:00:00	ADRIAN PESARO	M	80	PESARO	1
1/2020	02/04/2020	00:00:00	ADRIAN PESARO	M	81	PESARO	1
1/2020	02/04/2020	00:00:00	ADRIAN PESARO	M	81	PESARO	1
1/2020	02/04/2020	00:00:00	ADRIAN PESARO	M	81	PESARO	1
1/2020	02/04/2020	00:00:00	ADRIAN PESARO	M	74	PESARO	1
1/2020	02/04/2020	00:00:00	ADRIAN PESARO	M	66	PESARO	1
1/2020	02/04/2020	00:00:00	ADRIAN PESARO	M	75	PESARO	1
1/2020	02/04/2020	00:00:00	ADRIAN PESARO	M	60	PESARO	1
1/2020	02/04/2020	00:00:00	ADRIAN PESARO	M	72	PESARO	1
1/2020	02/04/2020	00:00:00	ADRIAN PESARO	F	80	PESARO	1
1/2020	02/04/2020	00:00:00	ADRIAN PESARO	F	77	PESARO	1
1/2020	04/05/2020	00:00:00	ADRIAN PESARO	F	72	PESARO	1
1/2020	04/05/2020	00:00:00	ADRIAN PESARO	F	79	PESARO	1
1/2020	11/05/2020	00:00:00	ADRIAN PESARO	M	78	PESARO	1
1/2020	14/05/2020	00:00:00	ADRIAN PESARO	M	64	PESARO	1
1/2020	14/05/2020	00:00:00	ADRIAN PESARO	F	67	PESARO	1

Figura 4.20. Tabella calcolata relativa ai decessi di persone di età inferiore alla mediana

Nonostante la maggior parte dei deceduti abbia un'età superiore ai 75 anni, nelle Marche ci sono ben 12 persone decedute con meno di 50 anni e, tra queste, una aveva appena 27 anni. Tale dato ci fa capire come la malattia, pur colpendo principalmente gli anziani, non risparmi assolutamente le persone più giovani, al contrario di quando spesso si legge o si sente.

Un'altra informazione significativa che possiamo trarre dal nostro cruscotto è quella relativa alle patologie pregresse riscontrate tra i deceduti della regione. A tal proposito, è importante sottolineare come i dati a nostra disposizione sulle singole persone decedute non forniscono il numero esatto di patologie pregresse di ciascun deceduto. Il report fornito dalla Regione Marche e dal GORES riporta, infatti, semplicemente, il numero intero 1 per indicare deceduti con una o più patologie pregresse e il numero 0 per indicare deceduti senza patologie pregresse. Avendo solo tale dato a disposizione siamo, comunque, in grado di evidenziare come il 95% delle persone decedute fosse affetta da patologie pregresse. In particolare, tale percentuale è ancora più alta tra i decessi sopra la mediana, raggiungendo, addirittura, un valore vicino al 99%.

Infine, concludiamo la nostra analisi prendendo in considerazione l'età delle persone decedute e analizzando la situazione delle Marche in rapporto a quella italiana. Dal grafico a torta riportato nella sezione sinistra del cruscotto di Figura 4.19 si nota facilmente come il virus sia più letale nel sesso maschile. Dei 989 morti delle Marche, infatti, 588 (59.45%) sono maschi e 401 (40.55%) sono femmine. Tale divario diventa, addirittura, più marcato tra le persone decedute con un'età inferiore alla mediana delle età dei deceduti totali, dato che fa capire come, tra i giovani, il virus sia significativamente più letale nei maschi. Tra i deceduti con età superiore alla mediana, invece, la differenza si affievolisce notevolmente, come si può notare dal grafico a torta della sezione destra, suddiviso in due metà quasi uguali.

Un'altra informazione interessante che possiamo trarre dal cruscotto a nostra

disposizione è che le donne decedute dopo aver contratto l'infezione da Covid-19 hanno un'età più alta rispetto agli uomini. Per quanto riguarda l'età media, per le donne è di 83.50 anni, mentre per i maschi di 78.60 anni. Invece, per quanto concerne la mediana, per i maschi questa assume il valore di 81 anni, mentre per le femmine il valore di 85 anni.

I dati dell'ISTAT sulla situazione italiana sono, ancora una volta, perfettamente in linea con quelli della Regione Marche. Riportiamo, in particolare, l'istogramma relativo al numero dei decessi in Italia per fasce di età di pazienti positivi all'infezione da Covid-19 (Figura 4.21), che conferma quanto finora detto.

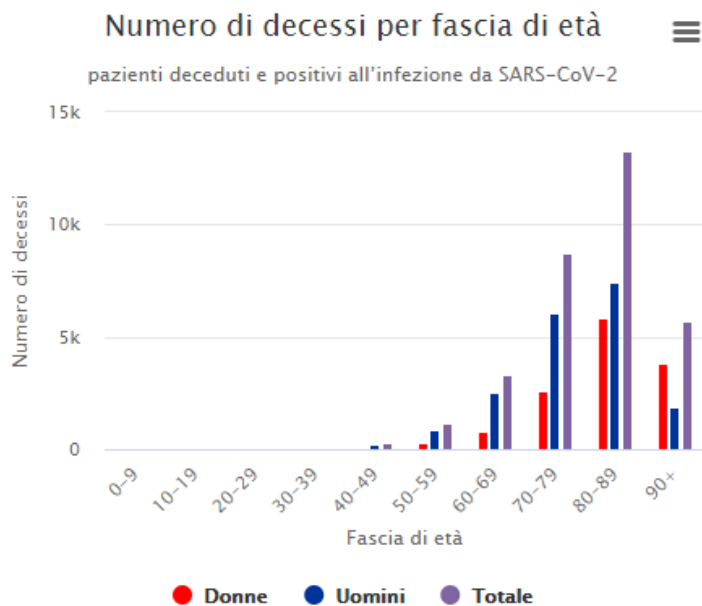


Figura 4.21. Iistogramma relativo ai decessi in Italia sulla base delle età

Realizzazione di cruscotti relativi a casi positivi, ricoveri e guariti

In questo capitolo proseguiamo le attività di analisi in Power BI Desktop. In particolare, al suo interno visualizzeremo, analizzeremo e descriveremo le informazioni relative ai tamponi ed ai casi positivi per quanto riguarda lo sviluppo della pandemia da Covid-19 all'interno della Regione Marche e nelle sue province.

5.1 Dashboard introduttiva su tamponi e positivi

Prima di iniziare ad analizzare nel dettaglio lo sviluppo della pandemia da Covid-19 all'interno della Regione Marche, in termini di tamponi e casi positivi, cerchiamo di fornire una panoramica generale riguardo a queste tematiche. A tal proposito, è stato realizzato un apposito cruscotto introduttivo che riassume tutte le informazioni presenti nel nostro modello di Power BI Desktop relative al dato dei tamponi e dei casi positivi rilevati all'interno delle Marche e delle sue province.

Inoltre, ricordiamo come, tra i dati forniti dalla Regione Marche attraverso il GORES, ci siano anche quelli di nostro interesse in questa sezione, ovvero quelli relativi proprio alle informazioni su tamponi eseguiti e casi positivi. Le Marche, fin dai primi sviluppi del Coronavirus all'interno della Regione, hanno, infatti, fornito con grande trasparenza questi dati. In particolare, è necessario ricordare come i report della Regione facciano una distinzione tra *Tamponi Totali* e *Casi Diagnosticati*.

Con *Tamponi Totali* si intende il numero totale di tamponi effettuati all'interno delle strutture regionali, includendo in tale numero anche i tamponi effettuati più volte sulla stessa persona. Infatti, come anche ribadito più volte dai diversi enti sanitari nazionali, un paziente affetto da Covid-19 si ritiene guarito se risolve i sintomi dell'infezione, ma, soprattutto, se risulta negativo in due test consecutivi, effettuati a distanza di 24 ore uno dall'altro. Tutto ciò significa che, in generale, se un paziente è sottoposto a un primo tampone e risulta positivo, in seguito può essere testato almeno altre due volte per verificare che sia negativo. Per un singolo paziente guarito, quindi, saranno conteggiati tre tamponi effettuati, e non uno solo.

Con *Casi Diagnosticati*, invece, la Regione Marche intende indicare il numero effettivo di persone che sono state sottoposte a tampone.

Dunque, è fondamentale saper riconoscere la differenza tra questi due dati che, ad un primo impatto, potrebbero sembrare identici.

Un'altro aspetto fondamentale da chiarire è, anche, lo stesso dato relativo ai *Casi Positivi*, che non è da confondere con quello dei *Malati Attivi*. La Regione Marche si esprime in maniera abbastanza chiara riguardo a questi due concetti, sottolineando come il dato sui *Casi Positivi* sia riferito al numero totale di persone risultate positive al tampone dall'inizio dell'epidemia; il dato relativo ai *Malati Attivi* rappresenta, invece, il numero totale dei pazienti ricoverati o in isolamento domiciliare, attualmente positivi.

Dopo questa premessa, possiamo ad analizzare la dashboard su tamponi e casi positivi precedentemente introdotta. Tale dashboard è riportata in Figura 5.1; notiamo come sia composta sia da semplici schede, che da grafici misti e istogrammi.

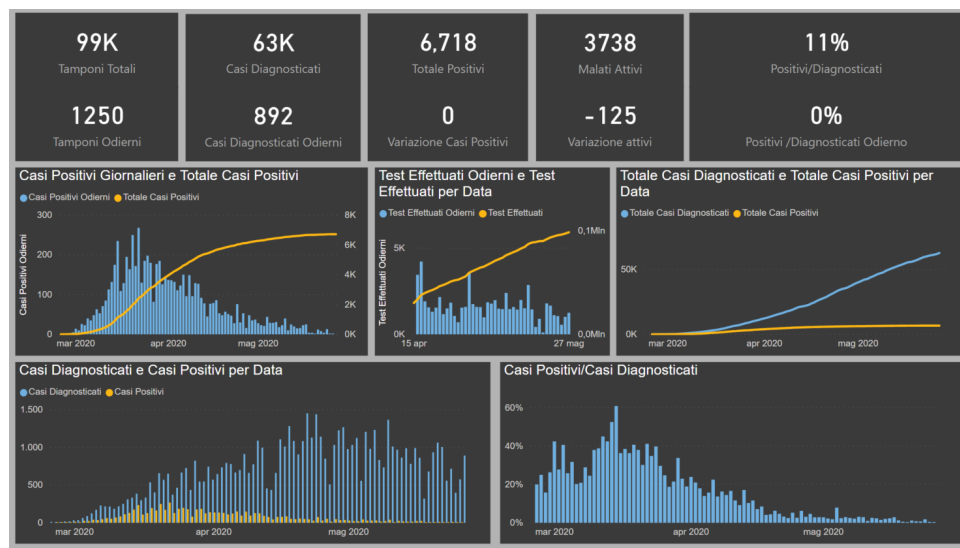


Figura 5.1. Dashboard introduttiva su tamponi e casi positivi

Un aspetto importante da sottolineare riguarda la disposizione delle schede: quelle immediatamente in prima riga riportano sempre informazioni riguardo a numeri che si riferiscono ad un valore cumulativo (“Tamponi Totali”, “Casi Diagnosticati”, “Totale Positivi”, “Malati Attivi”, “Positivi/Diagnosticati”); le schede in seconda riga riportano, invece, le informazioni relative a variazioni giornaliere (“Tamponi Odierni”, “Casi Diagnosticati Odierni”, “Variazione Casi Positivi”, “Variazione Attivi”, “Positivi/Diagnosticati Odierni”).

Si noti come tale cruscotto non riporti alcuna informazione relativa alle province, ma si riferisce ai dati dell'intera Regione. Perciò, per costruire i grafici al suo interno, sono stati utilizzati i soli campi della relazione “Generale”. Inoltre, sono state anche realizzate delle apposite misure DAX per poter ricavare delle particolari informazioni di interesse per quanto riguarda questa dashboard nonché le dashboard che analizzeremo successivamente.

Un dato particolarmente interessante di questo cruscotto è il rapporto percentuale tra casi positivi e casi diagnosticati, visualizzato nelle due schede in alto a destra. Tale dato è stato ricavato dalle informazioni già presenti nel nostro modello

in Power BI, applicando una semplice funzione DAX di divisione, ed esprimendo il risultato in notazione percentuale (Figura 5.2).

```

1 Casi Positivi Odierni diviso Casi Diagnosticati Odierni =
2 DIVIDE(
3     SUM('Generale'[Casi Positivi Odierni]),
4     SUM('Generale'[Casi Diagnosticati Odierni])
5 )

```

Figura 5.2. Misura DAX per il calcolo del rapporto percentuale tra casi positivi e casi diagnosticati

Questa misura rappresenta un indice abbastanza importante per valutare il fattore di crescita della pandemia, in quanto non indica un dato assoluto, ma permette di mettere a confronto il numero dei nuovi casi positivi di ciascun giorno con il numero dei casi diagnosticati relativi allo stesso giorno.

Si noti come, nelle Marche, tale rapporto sia pari all'11%, essendo 6.718 il totale dei positivi e 63.000 il numero dei casi diagnosticati (dati relativi al 27 Maggio 2020). Particolarmente significativo, a questo proposito, è l'istogramma riportato in basso a destra nel cruscotto, che riporta giornalmente questo rapporto percentuale. Tale grafico mostra una iniziale crescita della misura in questione per quasi tutta la durata del mese di Marzo, seguita da una sostanziale e progressiva decrescita, già a partire dall'ultima settimana di Marzo. Tale dato conferma, ancora una volta, l'efficacia delle misure di contenimento adottate dagli enti governativi, nonché una diminuzione della rapidità di diffusione del virus.

Il cruscotto presenta un altro grafico abbastanza interessante, che è quello riportante l'andamento complessivo e giornaliero dei casi positivi. Si tratta del grafico misto mostrato nella parte sinistra della fascia centrale del cruscotto. Il grafico a linee rappresenta il numero totale di casi positivi dal giorno d'inizio della raccolta dei dati da parte della Regione Marche e mostra come, dopo una crescita quasi esponenziale di tale numero durante tutto il mese di Marzo e la prima settimana di Aprile, si sia raggiunto un valore pressoché costante, la cui crescita, negli ultimi giorni di Maggio, si attesta sempre ad un numero inferiore ai cinque casi giornalieri.

Il cruscotto riporta, anche, un grafico che rappresenta il dato dei test effettuati dalla Regione Marche. Il numero totale di tamponi testati dalla Regione dall'inizio dell'epidemia è di circa 99.000, mentre le persone effettivamente sottoposte al tampone sono circa 63.000. Inoltre, la Regione mostra, anche, una certa costanza nel numero di test giornalieri effettuati che, salvo particolari picchi (dovuti, in realtà, al conteggio di test effettuati in giorni precedenti), si mantiene quasi sempre superiore ai 1000 tamponi giornalieri. Questo dato è ben evidente nel grafico a linee riportante il numero totale di test effettuati, che è molto simile a quello di una retta e che mostra, quindi, una crescita sempre costante.

Infine, l'ultimo grafico che ci resta da analizzare in questo cruscotto, è quello relativo all'andamento giornaliero di casi diagnosticati e di casi positivi, riportato nella zona in basso a destra del cruscotto stesso. Il grafico in questione è noto in Power BI come "Istogramma a colonne raggruppate", e permette di visualizzare un istogramma le cui colonne rappresentano i valori di due campi della nostra relazione. Tale grafico evidenzia la netta decrescita dei casi positivi giornalieri, già accennata precedentemente, a partire dalla prima settimana del mese di Aprile 2020. Inoltre,

evidenza, ancora una volta, la costanza del numero di casi giornalieri diagnosticati, dopo una iniziale fase di crescita di tale numero.

5.1.1 Confronti temporali

In questa sezione ci soffermiamo sulla crescita temporale del numero dei casi positivi effettuando, in particolare, un confronto settimanale e mensile. Per analizzare l'andamento della pandemia e l'efficacia delle misure di contenimento imposte dallo stato italiano all'intero territorio nazionale a partire dal mese di Marzo 2020, riteniamo sia indispensabile mettere a confronto la crescita dei casi positivi nel corso delle varie settimane della fase di espansione del virus.

Riportiamo in Figura 5.3 un cruscotto che, mediante due grafici ad area, mostra questo confronto nel tempo. Per quanto riguarda il grafico ad area nella parte superiore del cruscotto, esso permette di mettere a confronto, per ciascun giorno riportato nell'asse delle ascisse, il numero di casi positivi registrati in quello stesso giorno, il numero di casi positivi registrati 14 giorni prima, e il numero dei casi positivi registrati 1 mese prima. Il grafico nella parte inferiore riporta le stesse informazioni, riferite, però, al rapporto percentuale tra casi positivi e casi diagnosticati illustrato nella sezione precedente.

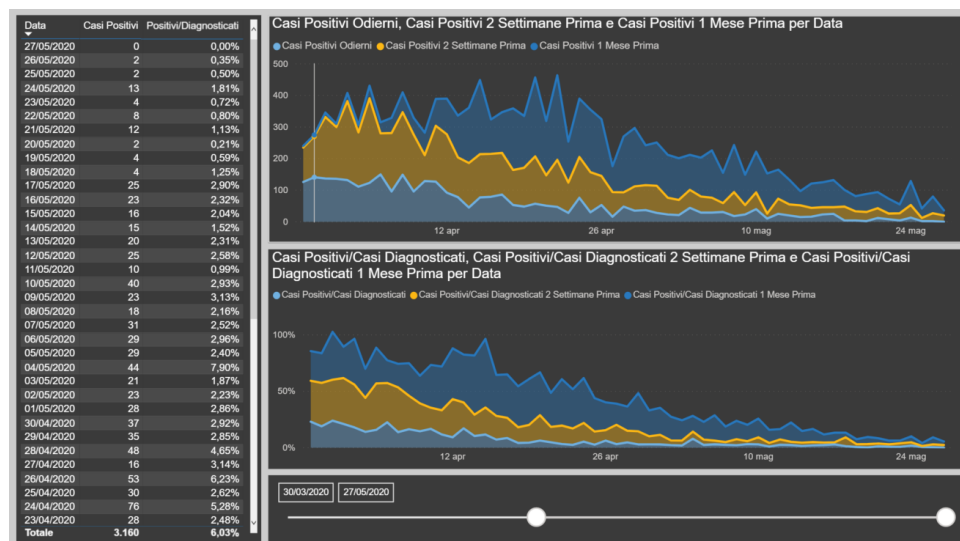


Figura 5.3. Cruscotto relativo al confronto temporale dei casi positivi

Abbiamo già visto, nel precedente capitolo, come questi ultimi grafici si basino sulla creazione di misure DAX calcolate facendo uso della funzione DATEADD, relativa al tipo di dato "Date". In questo caso abbiamo applicato tale funzione per ottenere le misure calcolate "Casi Positivi 1 Mese Prima" e "Casi Positivi 2 Settimane prima", che vengono utilizzate nel primo grafico, e le misure "Casi Positivi/Casi Diagnostici 1 Mese Prima" e "Casi Positivi/Casi Diagnostici 2 Settimane prima", che vengono, invece, utilizzate nel secondo grafico.

Osservando entrambi i grafici, con particolare riferimento al secondo, è facile notare un netto calo del numero di casi positivi e, soprattutto, del rapporto percentuale di quest'ultimo con il numero di casi diagnosticati. La differenza più alta si raggiunge intorno alla metà del mese di Aprile, come, ad esempio, il 15 Aprile, in cui il rapporto percentuale è dell'11.63%, molto più basso rispetto al valore di 23.95% di due settimane prima, e significativamente inferiore rispetto al 60.88% del mese precedente.

Per facilitare la lettura dei due grafici, nel cruscotto viene riportato anche un elemento grafico di tipo “Tabella” che mostra, per ciascun giorno, il numero di positivi e il rapporto percentuale tra casi positivi e casi diagnosticati.

5.2 Analisi dei casi positivi

Proseguiamo, ora, la nostra analisi concentrandoci esclusivamente sul dato dei positivi al virus. Analizzeremo, in particolare, la suddivisione dei casi positivi per gravità e la crescita del numero di positivi nell'intera Regione Marche e, anche, in tutte le sue province. Si vuole ribadire come, all'interno di questa sezione e in tutto il resto del capitolo, si utilizzeranno sia il termine “Casi Positivi” che il termine “Malati” per riferirsi alle persone risultate positive al virus.

Prima di tutto, cominciamo con l'analizzare un cruscotto introduttivo sulla situazione dei malati all'interno della Regione (Figura 5.4).

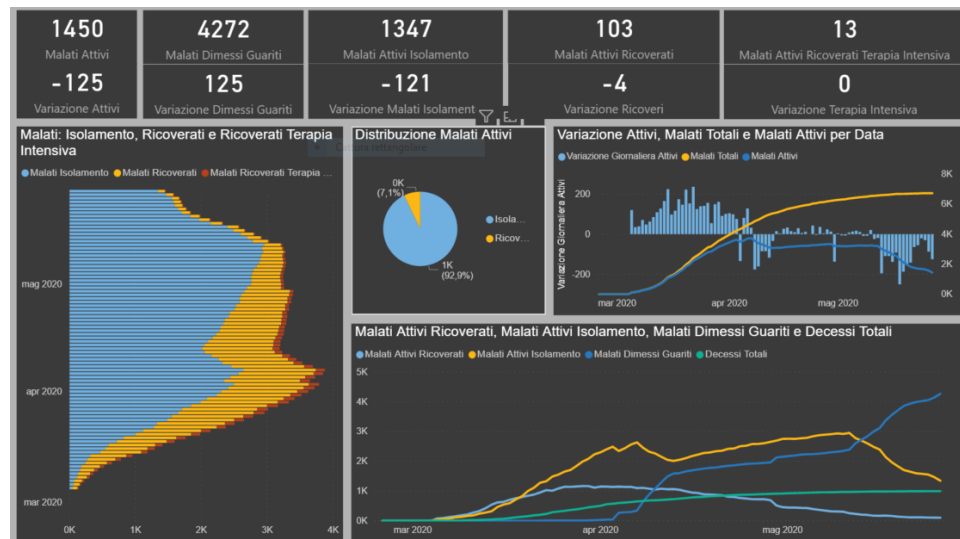


Figura 5.4. Cruscotto introduttivo sulla situazione dei malati

Il cruscotto sopra riportato presenta, così come quello visto nella sezione introduttiva di questo capitolo (Figura 5.1), una parte superiore caratterizzata dalla presenza di soli oggetti visivi di tipo “Scheda”. Tali schede riportano, in particolare, il dato esatto sul numero di “Malati Attivi”, “Malati Dimessi Guariti”, “Malati

Attivi Isolamento”, “Malati Attivi Ricoverati”, e “Malati Attivi Ricoverati Terapia Intensiva”. Questi dati, utilizzati anche nei grafici del cruscotto stesso, illustrano una sorta di suddivisione per gravità dei casi positivi.

Ricordiamo come con “Malati Attivi” si intende il numero di persone attualmente positive al virus. Con riferimento a tale dato (così come anche per gli altri), riportiamo anche la variazione giornaliera, indicata con “Variazione Attivi”. I dati forniti dal GORES non danno esplicitamente tale informazione, a nostro avviso particolarmente significativa. 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 (Figura 5.5).

```
Nuovi Attivi = (var selectDate = CALCULATE(LASTDATE(Generale[Data]), FILTER(Generale, Generale[Data] < EARLIER(Generale[Data])))
return Generale[Malati Attivi] - CALCULATE(SUM(Generale[Malati Attivi]), FILTER(Generale, Generale[Data] = selectDate)))
```

Figura 5.5. Colonna calcolata relativa alla variazione del numero degli attivi

Operazioni perfettamente analoghe a quella appena illustrata sono state effettuate, anche, per ottenere le seguenti colonne calcolate:

- *Dimessi Guariti Odierni*: colonna che rappresenta la variazione giornaliera dei dimessi e dei guariti.
- *Malati Isolamento Variazione Giornaliera*: colonna che rappresenta la variazione giornaliera dei casi positivi in isolamento domiciliare.
- *Nuovi Ricoveri*: colonna che rappresenta la variazione giornaliera dei malati ricoverati in strutture ospedaliere.
- *Terapia Intensiva Odierni*: colonna che rappresenta la variazione giornaliera dei ricoverati in terapia intensiva.

Inoltre, tutte queste misure saranno particolarmente utili anche nei cruscotti successivi.

Continuiamo, ora, la nostra analisi del cruscotto di Figura 5.5. Cominciamo con l’analizzare il grafico a torta che mostra la suddivisione degli attivi tra coloro che sono in isolamento domiciliare e quelli che, purtroppo, sono sotto ricovero. La percentuale dei ricoverati è significativamente più bassa rispetto a quella dei malati in isolamento. Tale dato ci fa capire come il contagio, nella maggior parte dei casi, si risolva in sintomi semplici, senza aggravare particolarmente le condizioni di salute della persona positiva.

Considerazioni perfettamente analoghe valgono anche per il “Grafico a barre in pila” presente nella parte sinistra del cruscotto che riporta, giorno per giorno, il dato relativo ai malati in isolamento, ai malati ricoverati, e ai malati ricoverati in terapia intensiva.

Il grafico misto presente in tale cruscotto, invece, permette di correlare tre campi presenti nel nostro modello, in particolare nella tabella “Generale”; infatti, tale grafico mostra la variazione giornaliera degli attivi, la crescita del totale dei malati dall’inizio dell’epidemia e l’andamento nel tempo dei malati attivi. In tale grafico è particolarmente significativo l’istogramma sulla variazione giornaliera degli attivi che mostra come fino al 7 Aprile si abbiano esclusivamente variazioni in positivo di tale numero, mentre, a partire da tale giorno, iniziano a comparire variazioni

negative, fino ad arrivare alla terza settimana di Maggio 2020 in cui il numero degli attivi è in continua decrescita e, quindi, con variazioni sempre negative. Stesse considerazioni si possono fare anche sul grafico a linee che mostra l'andamento complessivo degli attivi che, dopo una iniziale fase di crescita proseguita fino al 7 Aprile, mostra una fase di stallo fino al 12 Maggio, giorno a partire dal quale inizia una forte decrescita.

Il grafico in fondo a sinistra, invece, permette di avere un'ampia panoramica della situazione. Esso mette insieme i malati attivi (distinti tra quelli in isolamento domiciliare e quelli ricoverati in strutture ospedaliere), il totale delle persone dimesse e guarite ed il totale dei decessi. Tale grafico mostra un trend che possiamo considerare positivo; infatti, il numero dei dimessi guariti, dopo una fase di lenta crescita, subisce un forte rialzo verso la metà del mese di Maggio 2020, arrivando a sorpassare il numero dei malati attivi, sia ricoverati che in isolamento.

Infine, per concludere questa nostra analisi, cerchiamo di confrontare la situazione della Regione Marche con quella generale italiana.

In Figura 5.6 riportiamo un grafico tratto dalla dashboard messa a disposizione dalla Protezione Civile che mostra la variazione giornaliera del totale dei positivi in Italia. Si nota subito la forte somiglianza con il grafico relativo alle Marche, precedentemente illustrato nel nostro cruscotto. In entrambi i casi, infatti, si hanno variazioni in rialzo nella crescita del totale dei positivi fino alla fine del mese di Aprile 2020, a partire da cui si hanno, invece, solo variazioni negative di tale andamento.

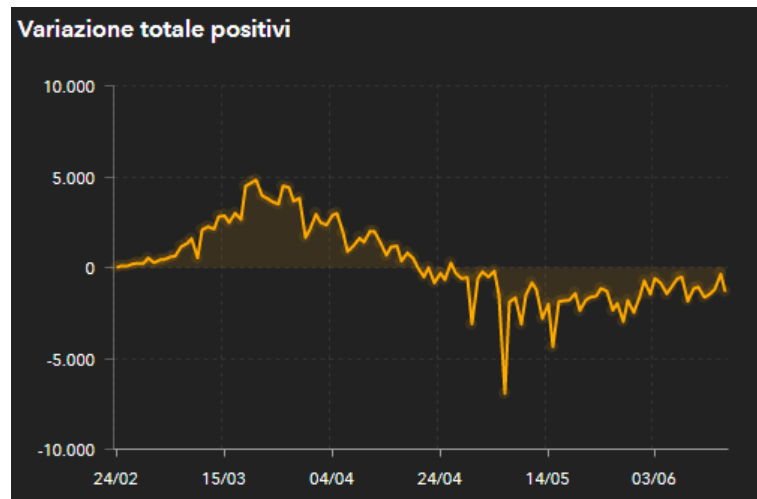


Figura 5.6. Grafico della Protezione Civile relativo alla variazione giornaliera degli attivi in Italia

In Figura 5.7 riportiamo, invece, un altro grafico della Protezione Civile, questa volta riguardante l'andamento del totale dei positivi, dei deceduti e dei dimessi guariti nell'intero territorio italiano. Si tratta delle stesse informazioni che abbiamo analizzato nell'ultimo grafico del nostro cruscotto e, ancora una volta, si nota la grande somiglianza tra i due grafici, sia nella crescita dei deceduti, ma soprattutto

nell'andamento dei dimessi guariti (che, in questo caso, ha un andamento più regolare e omogeneo) e del totale dei positivi, con quest'ultimo che assume la forma di una campana.

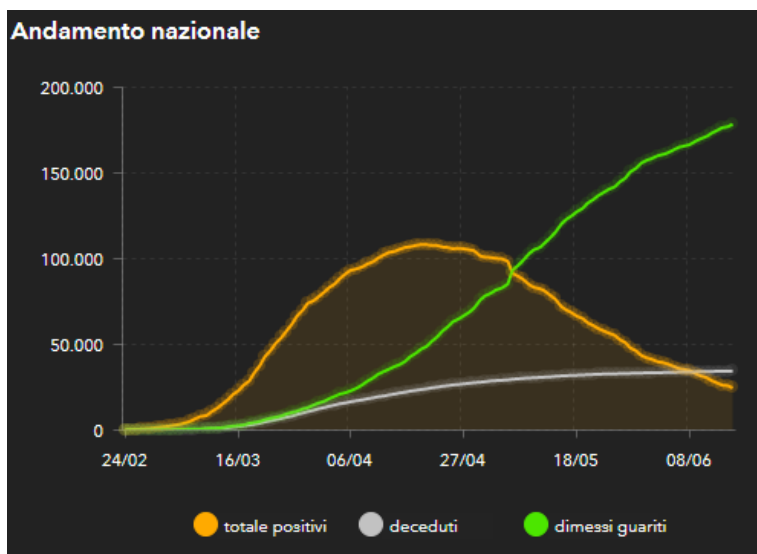


Figura 5.7. Grafico della Protezione Civile relativo all'andamento di casi positivi, decessi e guariti in Italia

5.2.1 Analisi per Provincia

Così come fatto anche nel capitolo precedente per quanto riguarda la situazione relativa ai decessi, passiamo, ora, ad analizzare, in dettaglio, le 5 province marchigiane.

Riportiamo, a tal proposito, un cruscotto relativo all'andamento dei casi positivi nelle province della Regione Marche, i cui dati sono aggiornati al 27 Maggio 2020. Ricordiamo come, per realizzare tutti i cruscotti relativi alle situazioni provinciali, sia stata adoperata la relazione "Province".

Il cruscotto in esame presenta, tramite un oggetto visivo "Scheda", il dato sul totale dei malati della regione e sulla variazione giornaliera di tale numero. Troviamo, all'interno del cruscotto, anche una semplice tabella riportante il numero dei malati per ciascuna provincia, e un "Filtro dei dati", che permette di filtrare i dati del cruscotto attraverso la selezione di una delle 5 province. Inoltre, è presente una visualizzazione a mappa che permette di visualizzare immediatamente la gravità della situazione nelle diverse province. La situazione che mostra questa mappa è particolarmente interessante: in alto è presente la provincia di Pesaro Urbino che, ancora una volta, si colloca come la provincia più colpita della Regione (e non solo, come vedremo più avanti), mentre, nella parte inferiore della mappa, è presente la provincia di Ascoli Piceno, la quale risulta essere, invece, la provincia meno colpita

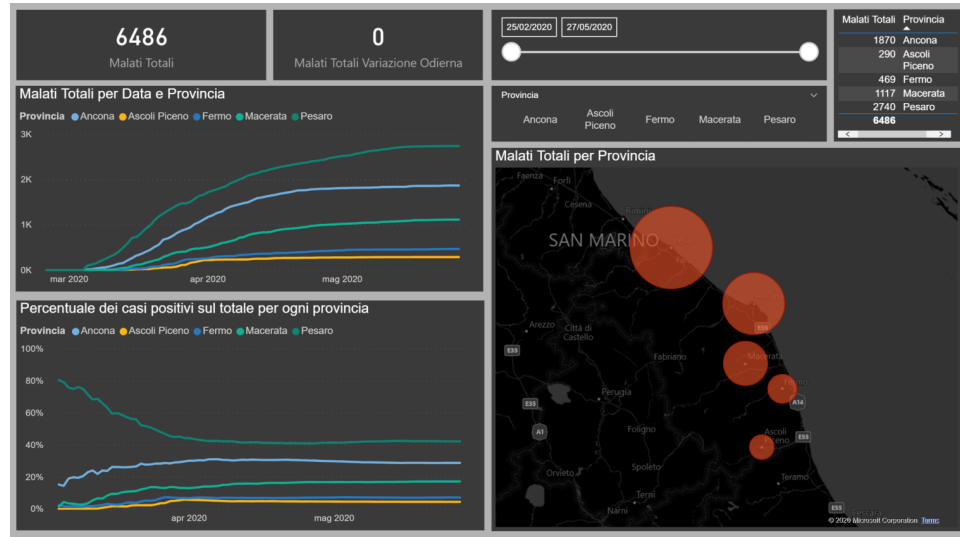


Figura 5.8. Cruscotto relativo all’andamento dei casi positivi nelle province marchigiane

in termini di casi positivi. Questo dato è interessante soprattutto perché mostra una sorta di perdita di forza del contagio scendendo da nord verso sud; si tratta, comunque, di un andamento che si riscontra anche a livello nazionale.

Il primo grafico a righe del cruscotto riporta, invece, la crescita giornaliera del numero dei positivi nelle province marchigiane mostrando, ancora una volta, le due conclusioni che abbiamo sempre tratto finora sulla situazione della Regione: la prima riguarda l’andamento delle curve che, dopo una iniziale fase di forte crescita, subiscono, intorno alla prima metà del mese di Aprile 2020, un appiattimento progressivo; la seconda riguarda, invece, il dato sulla provincia più colpita della Regione, ovvero Pesaro Urbino.

Un’altra interessante analisi a livello provinciale è quella che riguarda la percentuale dei casi positivi sul totale. Anche qui si vede bene come la provincia di Pesaro Urbino, che fino al 25 Marzo contava il 50% dei casi regionali, abbia influito molto anche sulle classifiche che si sono viste a livello nazionale. Anche la provincia di Ancona ha avuto un incremento abbastanza forte nelle prime fasi dell’epidemia e, ad oggi, si attese intorno al 30% dei casi totali. Ancora una volta, si conferma Ascoli Piceno la provincia meno colpita delle Marche.

Sottolineiamo come, per realizzare tale grafico, si sia realizzata una misura DAX per ottenere una colonna calcolata. La colonna in questione è “Malati Totali Provincia / Malati Totali” e la formula DAX utilizzata è riportata in Figura 5.9.

Proseguiamo la nostra analisi provinciale prendendo, ora, in considerazione anche il numero di abitanti delle province della Regione Marche. Il cruscotto precedentemente illustrato, infatti, non considera tale numero che, invece, è un parametro molto importante per valutare il reale impatto dell’epidemia nelle province, in quanto indica l’incidenza dei casi positivi sugli abitanti, nota anche come “Incidenza di contagio”.

```

1 Malati Totali Provincia / Malati Totali =
2 var Provincia = Provincie[Provincia]
3
4 var DateTime = Provincie[Data]
5
6
7 var PreviousDate = CALCULATE(MAX(Provincie[Data]), FILTER(Provincie, Provincie[Provincia] = Provincia && Provincie[Data] <= DateTime))
8 var TotaliProvincia = CALCULATE(MAX(Provincie[Malati Totali]), FILTER(Provincie, Provincie[Provincia] = Provincia && Provincie[Data] <= PreviousDate))
9
10
11 var Tot = CALCULATE(SUM(Provincie[Malati Totali]), FILTER(Provincie, Provincie[Data] = DateTime))
12 return TotaliProvincia / Tot

```

Figura 5.9. Formula DAX della colonna calcolata relativa alla percentuale dei casi positivi provinciali sul totale

Abbiamo illustrato, nel capitolo precedente, un ragionamento simile, quando abbiamo analizzato i decessi ogni 100.000 abitanti. Per realizzare il cruscotto di Figura 5.10 sono stati applicati gli stessi passaggi, sempre all’interno della relazione “Province” con riferimento, però, al numero di casi positivi. Pertanto, non riportiamo nel dettaglio tali passaggi, ma ci limitiamo a ribadire come, attraverso i dati dell’ISTAT sugli abitanti delle province marchigiane, abbiamo creato, in Power BI Desktop, le misure “Incidenza Malati Ancona”, “Incidenza Malati Ascoli Piceno”, “Incidenza Malati Fermo”, “Incidenza Malati Macerata” e “Incidenza Malati Pesaro” che esprimono il numero di casi positivi ogni 100.000 abitanti in ciascuna provincia Marchigiana. Inoltre, lo stesso calcolo è stato effettuato anche per l’intera Regione Marche.

Con tali dati a disposizione, abbiamo, dunque, realizzato il cruscotto in questione.

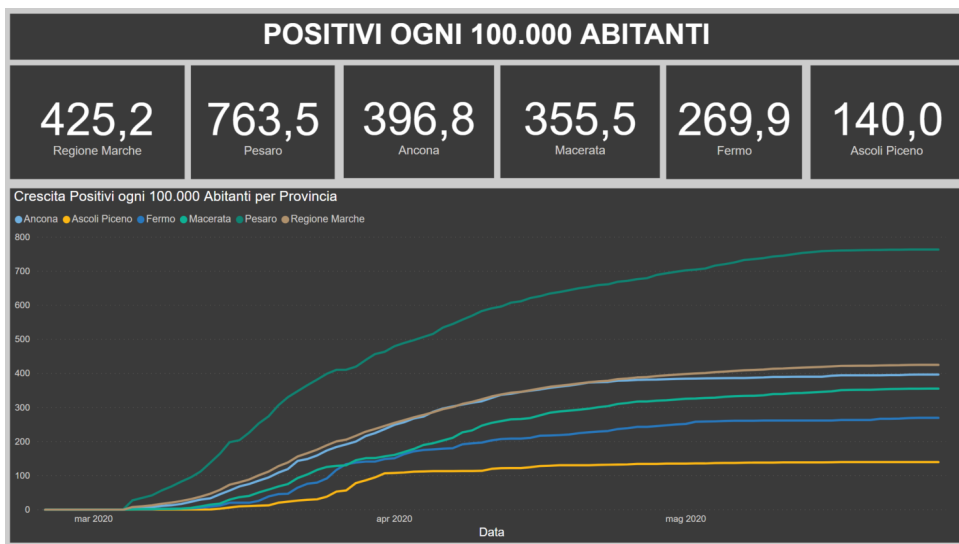


Figura 5.10. Cruscotto relativo al numero di positivi ogni 100.000 abitanti nelle province marchigiane

Esso illustra come Ascoli Piceno risulti essere la migliore provincia per quanto

riguarda il numero di positivi. Anche le altre province mostrano, tutto sommato, una situazione abbastanza positiva, fatta eccezione per Pesaro Urbino, caratterizzata da ben 763.5 positivi ogni 100.000 abitanti. L'intera Regione Marche presenta anch'essa un valore elevato di prevalenza di casi ogni 100.000 abitanti, che si attesta sul valore di 425.2, dato che la colloca subito dopo le regioni del Nord Italia e anche dell'Emilia-Romagna, ovvero le regioni più colpite dalla pandemia in Italia.

In particolare, il dato relativo alla provincia di Pesaro Urbino colloca quest'ultima direttamente tra le province in assoluto più colpite dal virus nell'intero suolo italiano. Questa situazione è particolarmente evidente nella Figura 5.11, riportante una mappa relativa all'incidenza di contagio nelle varie province italiane. Tale grafico si basa sui dati dell'ISTAT ed è aggiornato al 1 Maggio 2020; esso mostra come la Provincia di Pesaro Urbino sia l'unica provincia del Centro Italia con un numero di casi Covid-19 per 100.000 abitanti superiore a 600.

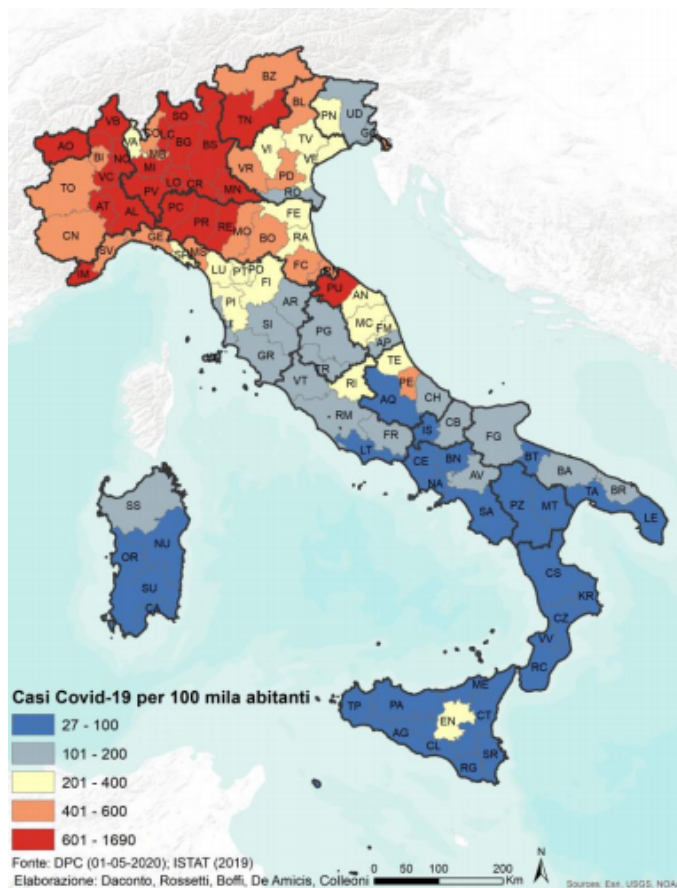


Figura 5.11. Mappa dei casi positivi per 100.000 abitanti in Italia

5.3 Analisi dei ricoveri

In questa sezione ci concentriamo su un altro aspetto fortemente negativo della pandemia da Covid-19 oltre a quello legato ai decessi, ovvero il dato relativo ai malati ricoverati all'interno delle strutture ospedaliere della regione Marche.

Come si è potuto constatare durante i giorni di sviluppo della pandemia nel territorio italiano, ma anche, in generale, nel mondo intero, il Covid-19, in alcuni casi, porta i malati ad essere sottoposti a ricoveri, spesso anche di terapia intensiva.

Fin dagli inizi della pandemia, lo stato italiano si è impegnato, in tutte le regioni, a fronteggiare la crescente richiesta di posti letto nei vari reparti ospedalieri. La forte crescita della pandemia, infatti, soprattutto nel corso del mese di Marzo 2020, ha portato ad un grande affollamento delle strutture ospedaliere italiane.

Anche le Marche hanno dovuto affrontare questo problema, soprattutto per quanto riguarda la domanda crescente di posti letto in reparti di terapia intensiva. L'amministrazione regionale, però, si è fatta cogliere abbastanza preparata, riuscendo ad aumentare, nel giro di pochi giorni, la capacità degli ospedali della regione, anche realizzando dei nuovi reparti appositamente pensati per ospitare i malati di Coronavirus.

Per avere una panoramica generale sulla situazione dei ricoveri della Regione è stato realizzato un apposito cruscotto introduttivo, riportato in Figura 5.12.

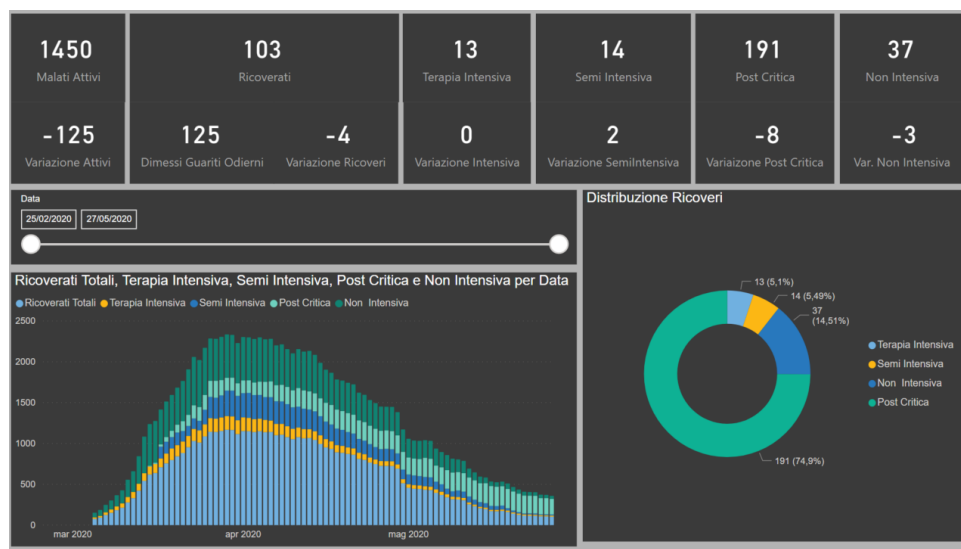


Figura 5.12. Cruscotto introduttivo sui ricoveri da Covid-19 nella regione Marche

Il cruscotto in questione fornisce, infatti, informazioni precise sul dato dei ricoveri da Covid-19 nei vari reparti ospedalieri delle Marche. Così come negli altri cruscotti finora analizzati, nella parte superiore troviamo solo oggetti visivi di tipo “Scheda” che riportano, in questo caso, i numeri esatti di malati e ricoverati nei reparti di terapia intensiva, non intensiva, semi-intensiva e post critica. Vengono,

inoltre, riportate, anche, la schede relative alle variazioni giornaliere di tali numeri (nel cruscotto di cui sopra tutte le schede sono relativa al giorno 27 Maggio 2020).

Insieme al dato dei ricoveri, il cruscotto presenta, anche, quello dei dimessi e guariti, ovvero il dato relativo sia alle persone risultate negative al doppio tampone di controllo, ma anche alle persone ancora positive che sono state dimesse dalle strutture ospedaliere.

Per ottenere i dati relativi alle variazioni giornaliere sono state create apposite colonne (direttamente in Power BI Desktop) utilizzando le formule DAX, secondo le stesse procedure che abbiamo analizzato nei cruscotti precedenti.

Riportiamo, a titolo di esempio, la formula utilizzata per il calcolo della colonna “Terapia Intensiva Odierni” che riporta, riga per riga, la variazione giornaliera del numero di ricoverati in strutture di terapia intensiva (Figura 5.13).

```
Terapia Intensiva Odierni = (var selectDate = CALCULATE(LASTDATE(Generale[Data]), FILTER(Generale, Generale[Data] < EARLIER(Generale[Data]))) return Generale[Malati Attivi Ricoverati Terapia Intensiva] - CALCULATE(SUM(Generale[Malati Attivi Ricoverati Terapia Intensiva]), FILTER(Generale, Generale[Data] = selectDate)))
```

Figura 5.13. Colonna calcolata relativa alla variazione giornaliera dei ricoverati in terapia intensiva

Il primo grafico, che andiamo ad analizzare in questo cruscotto, è il cosiddetto “Grafico ad anello” che ha una semantica analoga al grafico a torta. Esso riporta la distribuzione percentuale dei ricoveri nei vari reparti terapeutici evidenziando come i reparti di terapia post critica siano quelli maggiormente occupati, mentre quelli di terapia intensiva siano i meno affollati. Tale dato è sicuramente confortante ed è, naturalmente, riferito ad una fase di decrescita della pandemia all’interno delle Marche e di tutta Italia, sia dal punto di vista dei decessi, che dei ricoveri e degli attualmente positivi.

Più significativo, all’interno del cruscotto in analisi, è, sicuramente, il grafico a istogramma riportante l’andamento giornaliero dei ricoveri nelle diverse strutture. Per tutte le tipologie di reparti tale grafico riporta sempre un andamento a campana evidenziando come, al 27 Maggio 2020, la fase più acuta della pandemia sia stata superata. I giorni relativi alla fine del mese di Marzo, invece, rappresentano il momento culminante dello sviluppo del virus, anche in termini di occupazione di posti letto nelle strutture ospedaliere regionali. Si noti, in particolare, come il numero di ricoverati in reparti di terapia intensiva durante questi giorni sia pari, se non superiore, a quello dei ricoverati in reparti post-critici. Tale tendenza, fortunatamente, si è completamente rovesciata nel mese di Maggio, in cui il numero di ricoveri in terapia intensiva è andato costantemente a ridursi, mentre quello relativo ai ricoveri in post critica ha seguito un andamento opposto, segno che indica come si sia raggiunta una situazione di controllo della pandemia.

La traiettoria seguita da quest’ultimo istogramma è perfettamente analoga a quella relativa alla situazione generale dei ricoveri nell’intero territorio italiano, che riportiamo in Figura 5.14, in cui si può visualizzare anche l’andamento dei casi in isolamento domiciliare. Anche tale grafico, infatti, mostra una parte centrale, compresa tra la fine di Marzo e la prima settimana di Aprile, in cui il numero di ricoveri si mantiene stabile, salvo poi decrescere progressivamente.

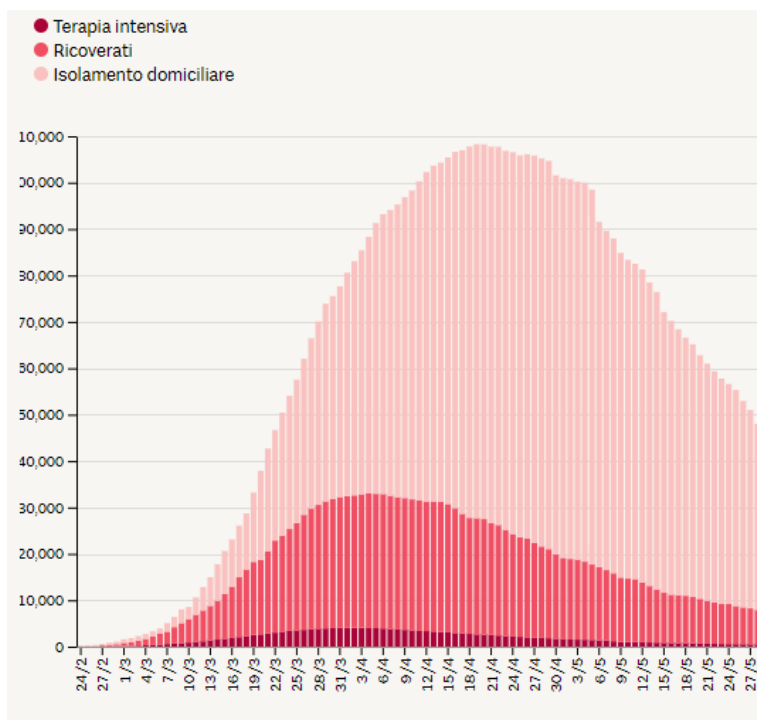


Figura 5.14. Grafico relativo all'andamento dei ricoveri da Covid-19 in Italia

5.3.1 Ricoveri per strutture

In questa sezione, mostriamo una serie di cruscotti relativi al dettaglio di come i ricoveri siano suddivisi nelle varie strutture della Regione Marche, ai vari livelli di gravità di terapia. Tali cruscotti sfruttano il report Giallo pubblicato giornalmente dal GORES, il quale riporta, giorno per giorno, il numero di ricoveri nei diversi ospedali della regione, suddivisi per gravità.

Ricoveri in reparti di terapia intensiva

In Figura 5.15 riportiamo il cruscotto relativo al dettaglio di come i ricoveri in terapia intensiva siano suddivisi nelle varie strutture del territorio. In particolare, utilizziamo un “Istogramma a colonne in pila” che riporta l'andamento giornaliero dei ricoveri in terapia intensiva per ciascuno degli ospedali della regione. Il grafico mostra come siano gli ospedali di Pesaro Urbino (“AO_Marche_Nord”) e quello di Ancona (“AO_Ancona_Torrette”) quelli sottoposti a maggiore stress. Anche qui, comunque, si nota sempre il trend decrescente, precedentemente visto, sul numero di ricoveri. Il cruscotto riporta, anche, nella parte superiore, una tabella che illustra il numero esatto di ricoverati per ciascuna struttura, aggiornato al 27 Maggio 2020.

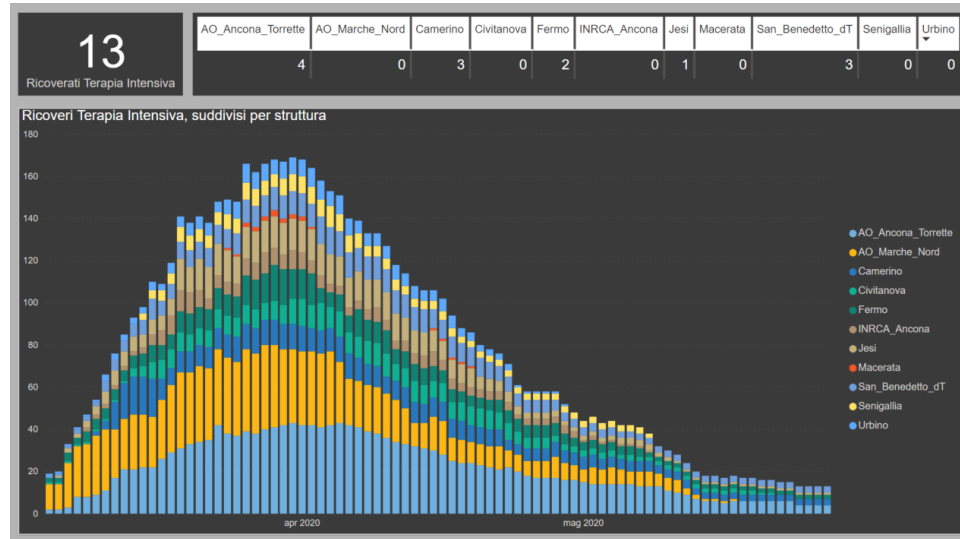


Figura 5.15. Cruscotto relativo all'andamento dei ricoveri in terapia intensiva per struttura

Ricoveri in reparti di terapia semi-intensiva

Riportiamo, in Figura 5.16, un cruscotto con la stessa struttura di quello appena analizzato, riguardante, però, la distribuzione dei ricoverati in reparti di terapia semi-intensiva nelle diverse strutture della Regione. Anche in questo caso si possono fare le stesse considerazioni di prima, ovvero che gli ospedali maggiormente affollati siano quelli di Ancona e Pesaro Urbino. Inoltre, si nota, ancora una volta, un trend decrescente nel numero di ospedalizzati.

Ricoveri in reparti post-critici

Nel cruscotto di Figura 5.17 riportiamo, secondo le stesse modalità viste nei due cruscotti precedenti, l'andamento dei ricoveri nei reparti post-critici delle varie strutture ospedaliere della regione Marche. A differenza dei casi precedenti, l'istogramma di questo cruscotto mostra una situazione costante del numero di ricoverati giornalieri. Ciò è dovuto, soprattutto, al fatto che tali reparti accolgono i dimessi dai reparti di terapia intensiva che, come abbiamo visto prima, sono in numero abbastanza elevato.

Ricoveri in altri reparti non intensivi

Infine, riportiamo in Figura 5.18 il cruscotto relativo all'andamento dei ricoverati in altri reparti non intensivi, suddivisi tra le varie strutture della Regione. L'istogramma di questo cruscotto mostra, ancora una volta, un andamento a campana, simile a quello visto per i ricoveri in terapia intensiva, con una grande crescita iniziale durante il mese di Marzo, seguita, poi, da una graduale fase di svuotamento del carico degli ospedali.

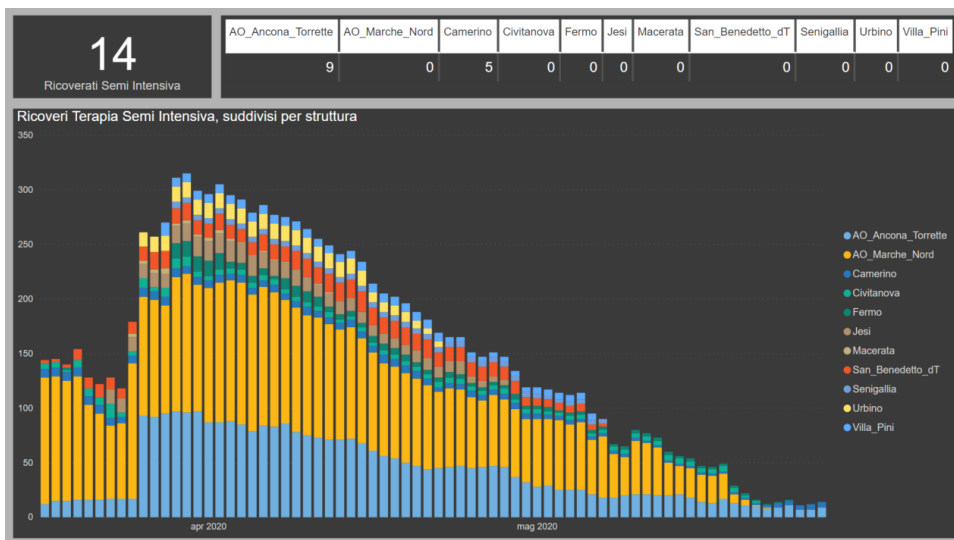


Figura 5.16. Cruscotto relativo all’andamento dei ricoveri in terapia semi-intensiva per struttura

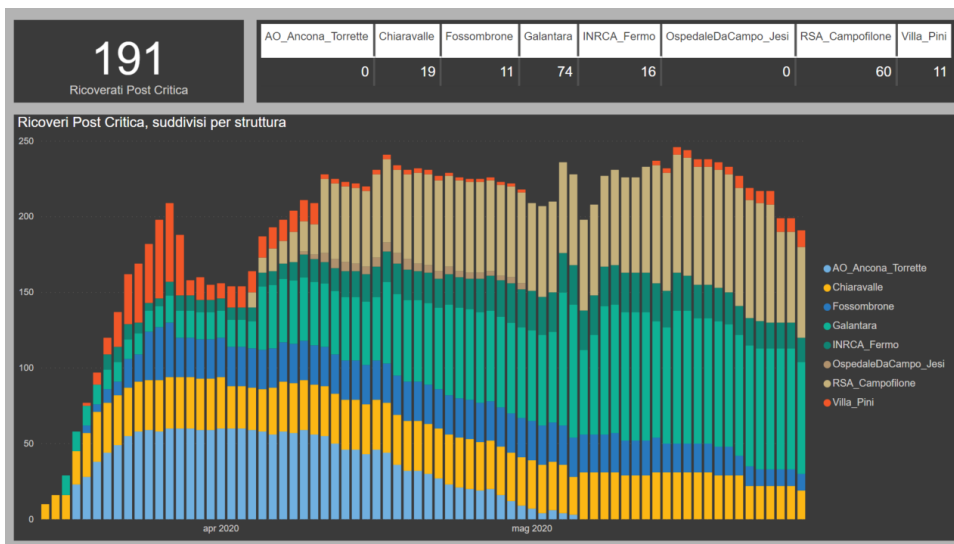


Figura 5.17. Cruscotto relativo all’andamento dei ricoverati in reparti post-critici, suddivisi per struttura

5.3.2 Focus sulle terapie intensive

Concentriamoci, ora, sul solo dato relativo all’andamento dei ricoveri in terapia intensiva all’interno della Regione. A tal proposito, abbiamo realizzato un cruscotto che presenta tutte le informazioni relative all’occupazione dei reparti di terapia intensiva prendendo in considerazione anche i dati relativi alle singole province (Fi-



Figura 5.18. Cruscotto relativo all’andamento dei ricoveri in reparti non intensivi, suddivisi per struttura

gura 5.19). Questi ultimi, in particolare, sono riportati nelle schede presenti nella parte superiore del cruscotto. Tali dati non erano inizialmente presenti nel nostro modello in Power BI Desktop; per tale motivo sono stati calcolati mediante la realizzazione di apposite colonne che, semplicemente, sommano il numero di ricoverati nei vari ospedali di ciascuna provincia.

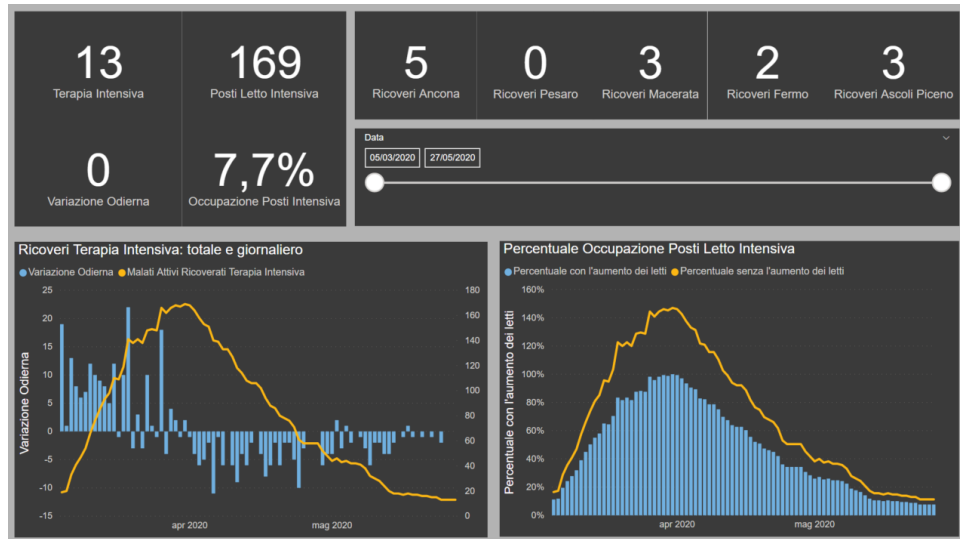


Figura 5.19. Cruscotto relativo all’analisi dei ricoveri in terapia intensiva nella regione Marche

Analizziamo, dapprima, il grafico misto presente nella parte sinistra del cruscotto. In esso, l'istogramma mostra la variazione giornaliera dei ricoverati in reparti di terapia intensiva, mentre il grafico a righe illustra l'andamento giornaliero del totale dei ricoverati. Abbiamo già avuto modo di commentare l'andamento del numero di ricoveri, perciò, ci limitiamo ad analizzare il dato sulle variazioni giornaliere. L'istogramma mostra come, per tutto il mese di Marzo (ad eccezione di soli sei giorni nella seconda metà del mese), ci siano state soltanto variazioni in positivo del numero di ricoveri; ciò ha spinto la Regione ad aumentare la capacità di posti letto in reparti di terapia intensiva degli ospedali che si trovano nel proprio territorio. A partire dal mese di Aprile, però, si nota una completa inversione del trend, con variazioni giornaliere quasi sempre negative, che hanno portato ad una continua diminuzione dei posti letto occupati fino ai 13 del 27 Maggio 2020.

Si noti come, in questo cruscotto, sia presente anche una scheda con etichetta "Occupazione Posti Intensiva", che riporta, cioè, la percentuale di occupazione dei posti letto disponibili per la terapia intensiva all'interno della Regione. Tale dato è stato ottenuto mediante una semplice operazione di divisione realizzata tramite una misura in Power BI Desktop (Figura 5.20).

```
1 Occupazione Posti Intensiva = DIVIDE(
  LASTNONBLANK(Generale[Malati Attivi Ricoverati Terapia Intensiva], 1 = 1),
  [Posti Letto Intensiva], 0
)
```

Figura 5.20. Misura calcolata relativa all'occupazione di posti letto di terapia intensiva

In particolare, per calcolare tale misura, è stato precedentemente recuperato il dato relativo al numero di posti letto in reparti di terapia intensiva all'interno degli ospedali della regione Marche, dato messo a disposizione dal Ministero della Salute (https://docs.google.com/spreadsheets/d/1EPoJ_T221tEy0K5QTkaZKjA-1Ho9VAEr6rFPJ_Qtds/edit#gid=0). Come precedentemente accennato, a causa della pandemia, la Regione ha aumentato la capienza di posti letto che nel 2018 era pari a 115. Nel momento più acuto dell'emergenza tali posti sono stati portati a 169 (tale incremento non tiene conto della realizzazione del "Covid Hospital" di Civitanova che ha portato all'aggiunta di ulteriori 84 posti letto di terapia intensiva verso la fine del mese di Maggio 2020), con un aumento del 47% dei posti letto disponibili.

Con questo dato a disposizione è stato realizzato l'interessante grafico misto riportato nella parte di destra del nostro cruscotto. In particolare, nel grafico a istogramma, viene mostrata la percentuale di occupazione di posti letto di terapia intensiva degli ospedali della regione dopo l'aumento dei posti letto, mentre nel grafico a righe viene riportata la percentuale di occupazione che si sarebbe registrata senza questo incremento. Notiamo, dunque, come questo aumento della capacità di posti letto sia stato fondamentale per evitare il superamento della soglia; infatti, il grafico a righe mostra che la percentuale di occupazione dei posti letto sarebbe andata addirittura sopra il 140%, considerato il precedente numero di posti letto.

5.4 Cruscotti relativi alla quarantena domiciliare

Tra le molte informazioni presenti nel nostro modello di dati c'è, anche, quella relativa al numero delle persone sottoposte alla quarantena domiciliare. Tale numero si riferisce sia alle persone positive in quarantena, ma anche a coloro che, pur non essendo positivi o non avendo fatto il tampone, hanno avuto contatti con casi positivi.

Tutte le informazioni relative a questo aspetto vengono riassunte nel cruscotto di Figura 5.21.

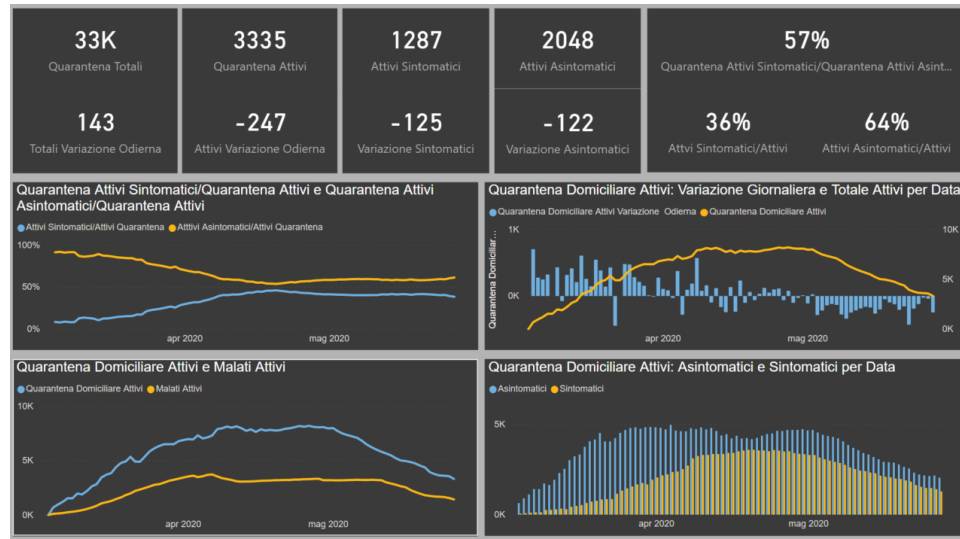


Figura 5.21. Cruscotto introduttivo sulle persone in quarantena domiciliare nella regione Marche

Nella parte superiore del cruscotto vengono riportate, sotto forma di schede, le seguenti informazioni:

- *Quarantena Totali*: numero totale di persone che, dall'inizio della pandemia, sono state in quarantena preventiva.
- *Quarantena Attivi*: numero totale di persone attualmente in isolamento domiciliare.
- *Attivi Sintomatici*: numero di persone con sintomi riconducibili al Coronavirus attualmente in isolamento domiciliare.
- *Attivi Asintomatici*: numero di persone senza alcun sintomo, attualmente in isolamento domiciliare.

Relativamente alle schede sopra esposte, sono riportati, anche, i dati relativi alle variazioni giornaliere ottenuti, per ciascuno dei campi, mediante colonne calcolate con formule DAX secondo procedure già viste nei cruscotti precedentemente analizzati.

Tra le schede nella parte superiore del cruscotto è presente, anche, il rapporto percentuale tra “Attivi Sintomatici” e “Attivi Asintomatici”, quello tra “Attivi Sintomatici” e “Quarantena Attivi” e, infine, quello tra “Attivi Asintomatici” e “Quarantena Attivi”. Tutte queste misure sono state ottenute mediante una semplice funzione DAX di divisione, il cui risultato è espresso in notazione percentuale.

I grafici presenti all’interno del cruscotto mostrano un’andamento a campana per quanto riguarda il numero di persone in quarantena domiciliare, con una fase di crescita di tale numero durante l’intero mese di Marzo 2020, seguita da una fase di stallo nel mese di Aprile, seguita, a sua volta, da una forte decrescita di tale numero nel corso del mese di Maggio. Anche i dati relativi alla variazione giornaliera del numero di persone in isolamento riflettono questo andamento, con variazioni soprattutto positive fino alla metà del mese di Aprile, e variazioni sempre negative per tutta la durata del mese di Maggio.

Un grafico interessante, all’interno del cruscotto, è quello mostrato in basso a destra, riportante l’andamento delle persone attualmente in quarantena messo a confronto con l’andamento dei casi positivi da Covid-19. Si noti come i due andamenti siano pressoché uguali, con il numero di casi positivi, però, sempre inferiore a quello dei casi in isolamento.

Infine, gli altri grafici del cruscotto mostrano il confronto tra casi sintomatici e casi asintomatici in isolamento domiciliare. Il dato che emerge è che, inizialmente, il numero di asintomatici era nettamente superiore a quello dei sintomatici; tale situazione però è cambiata nel corso del mese di Maggio 2020 in cui la diminuzione del numero di asintomatici in isolamento è stata abbastanza marcata, soprattutto con riferimento a quella dei sintomatici.

5.4.1 Situazione per provincia

Si è deciso di realizzare, anche, un cruscotto riportante i dati sulla quarantena domiciliare, particolarizzato alle singole province della regione. Il cruscotto è riportato in Figura 5.22 e mostra sempre la struttura dei cruscotti finora visti, con la presenza di schede nella parte superiore.

Nel cruscotto è presente una visualizzazione a mappa, riportante il numero di casi in isolamento nelle 5 province della regione. Inoltre, sono presenti tre grafici a righe che mostrano, per ciascuna provincia, l’andamento giornaliero del numero di persone in quarantena, del numero di casi sintomatici in quarantena e del numero di casi asintomatici in quarantena.

L’informazione che possiamo estrarre è, ancora una volta, come la provincia di Pesaro Urbino sia la più colpita anche da questo punto di vista, seguita sempre da Ancona. La provincia di Ascoli Piceno, invece, si dimostra ancora come la meno colpita dell’intera regione Marche.

5.4.2 Analisi sugli operatori sanitari in quarantena domiciliare

Concludiamo la nostra analisi prendendo in considerazione i dati relativi agli operatori sanitari in quarantena domiciliare. Infatti, tra le molte informazioni presenti nel nostro dataset c’è, anche, un parametro che tiene conto del numero di casi positivi

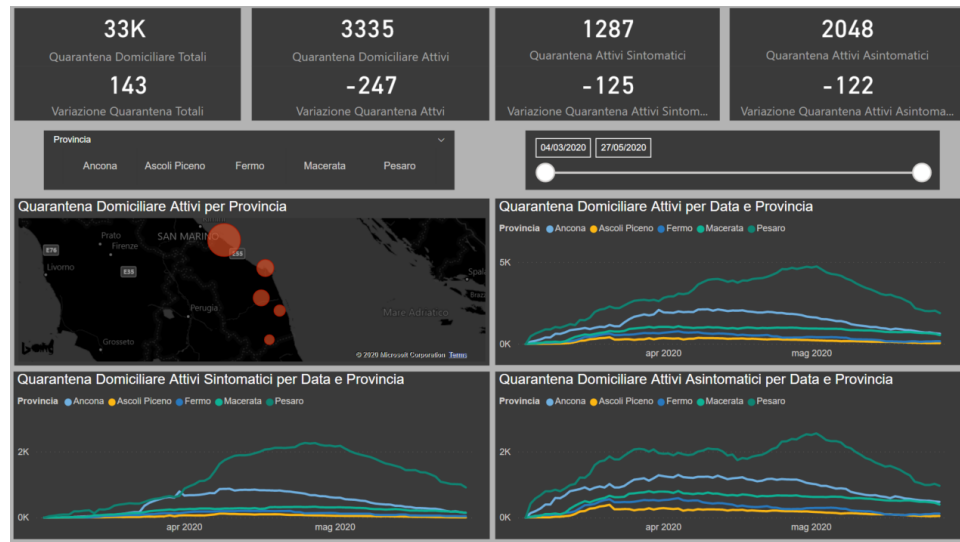


Figura 5.22. Cruscotto relativo all’andamento del numero delle persone in quarantena domiciliare casi nelle province marchigiane

e di persone che hanno avuto contatti con malati, limitatamente al solo personale sanitario.

Sappiamo come il personale sanitario sia stato, fin da subito, esposto in prima linea, e ovunque il numero di operatori sanitari contagiato è davvero molto elevato, non solo in Italia, ma anche nel resto del mondo. Il motivo alla base di tutto ciò, sicuramente, è dovuto al fatto per cui la pandemia, all’inizio della sua espansione, ci ha colto impreparati. Questa impreparazione si ricollega soprattutto al mancato e all’errato uso di mascherine e dei dispositivi di protezione individuale. Inoltre, anche il fatto di dover lavorare in reparti Covid-19, così come il permanere a lungo in aree dove hanno sostato molti pazienti positivi al Coronavirus, contribuisce al pericolo di contrarre l’infezione.

Ad esempio, secondo i dati forniti dall’Istituto Superiore di Sanità (“ISS”), alla data del 29 Aprile 2020, i contagi totali in Italia sono circa 210.000, fra cui quasi 22.000 operatori sanitari, una cifra corrispondente a circa il 10% dei contagiati totali.

Abbiamo, dunque, deciso di realizzare un apposito cruscotto, riguardante, in particolare, il dato relativo ai soli operatori sanitari in quarantena (Figura 5.23).

Un dato particolarmente interessante del cruscotto in esame è quello riguardante la percentuale degli operatori sanitari in quarantena rispetto al totale dei casi in isolamento; tale percentuale è del 10% e, sicuramente, rappresenta un valore piuttosto alto considerando, soprattutto, che si tratta di forza lavoro fondamentale per tutte le strutture ospedaliere della regione Marche, specialmente in un periodo di grande crisi sanitaria, come quella dovuta al Coronavirus.

Notiamo come il cruscotto contenga, anche, informazioni relative alle singole province della Regione. Infatti, per la sua realizzazione, è stata utilizzata la relazione “Province”, già analizzata in precedenza, che contiene anche dei campi relativi agli

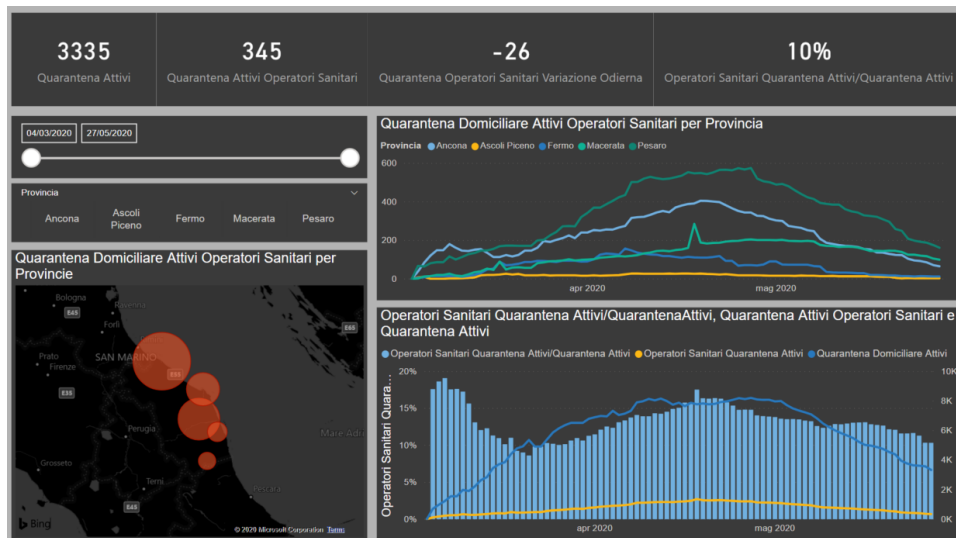


Figura 5.23. Cruscotto relativo agli operatori sanitari in quarantena domiciliare

operatori sanitari in quarantena per ciascuna delle province marchigiane.

Inoltre, per ottenere la percentuale sopra illustrata, è stata creata un'apposita misura DAX, poiché tale dato non era originariamente presente all'interno del nostro modello (Figura 5.24).

```

1 Operatori Sanitari Quarantena Attivi diviso Quarantena Domiciliare Attivi =
2 DIVIDE(
3     SUM('Province'[Quarantena Domiciliare Attivi Operatori Sanitari]),
4     SUM('Province'[Quarantena Domiciliare Attivi])

```

Figura 5.24. Misura calcolata relativa al rapporto tra gli operatori sanitari in quarantena domiciliare e il totale di casi in isolamento

I dati relativi alle province hanno permesso di realizzare il grafico di tipo mappa e il grafico a righe presenti nel cruscotto. Con particolare riferimento a quest'ultimo, il quale mostra gli operatori sanitari in quarantena per ciascuna provincia marchigiana, notiamo come sia ancora la provincia di Pesaro Urbino quella con i dati più preoccupanti, seguita sempre da Ancona, mentre, in ultima posizione, troviamo come provincia meno colpita, ancora una volta, quella di Ascoli Piceno. Il dato che emerge da questo grafico, come d'altronde in tutta Italia, è piuttosto preoccupante, almeno per quanto riguarda il periodo fino alla fine di Aprile 2020. Infatti, nonostante il primo caso positivo di Coronavirus nella Regione risalga al 25 Febbraio 2020, il trend è sempre crescente fino a metà Aprile. Addirittura, nella provincia di Pesaro Urbino, la crescita dura fino al 27 Aprile quando si è toccata la quota di 575 operatori sanitari in isolamento domiciliare. Come si nota chiaramente dal grafico, la situazione, a partire dal mese di Maggio 2020, si è completamente ribaltata, mostrando, fortunatamente, un andamento decrescente in tutte le province.

Per concludere questa nostra analisi sugli operatori sanitari, vediamo un grafico che mette in relazione il dato relativo ai decessi con quello dei ricoverati e, anche, con quello relativo agli operatori sanitari in quarantena domiciliare. Tale grafico è riportato in Figura 5.25.

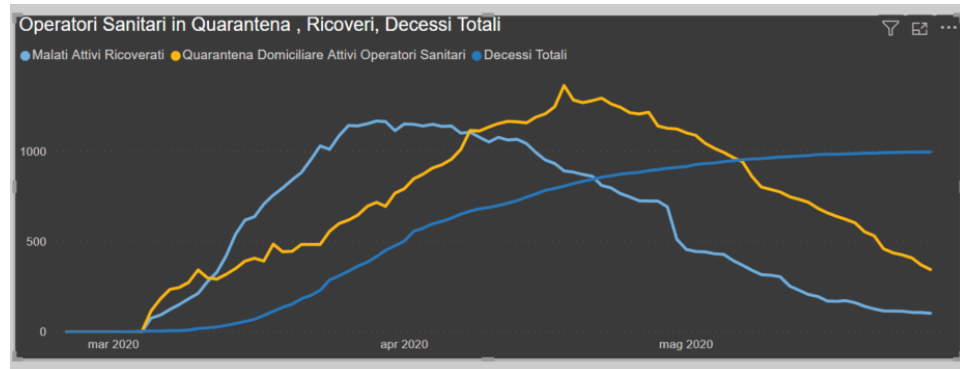


Figura 5.25. Grafico relativo a operatori sanitari in isolamento domiciliare, pazienti ricoverati e decessi totali

Lo scopo del grafico è quello di individuare una sorta di correlazione tra questi tre dati, per evidenziare la crisi degli ospedali della Regione durante il periodo più critico della pandemia. In particolare, si noti come, agli inizi di Aprile 2020, quando il numero di ricoverati negli ospedali raggiunge i valori massimi, così come anche quello relativo ai decessi, il numero degli operatori sanitari in isolamento raggiunge, anch'esso, valori molto elevati. Tutto ciò fa capire come, in tale periodo, ci sia stata una carenza di personale sanitario, soprattutto a fronte della situazione di gravità in cui si trovavano gli ospedali.

Realizzazione di cruscotti di tipo predittivo

In questo capitolo viene presentata un'altra fase del processo di analisi, ovvero l'analisi predittiva. In particolare, vedremo come Power BI Desktop permette di realizzare analisi di questo tipo mediante l'utilizzo di oggetti visivi del linguaggio R, basati sul modello ARIMA.

6.1 L'analisi predittiva

L'analisi predittiva (Figura 6.1), meglio nota come “Predictive Analytics”, è una forma di analisi avanzata che esamina i dati a disposizione per effettuare le previsioni di ciò che accadrà nel futuro. L'analisi predittiva, dunque, si nutre di dati storici per fare una previsione sugli eventi futuri; tali dati sono veri e propri ingredienti con i quali sviluppare un modello predittivo in grado di “imparare” i comportamenti, ovvero le regole più rilevanti del fenomeno osservato. Il passo successivo consiste nell'applicare queste regole ai dati correnti per prevedere gli eventi futuri, oppure suggerire nuove misure da adottare per arrivare a risultati ottimali.

L'analisi predittiva è fortemente utilizzata in ambito aziendale soprattutto per aumentare la competitività delle strategie di marketing digitale, per far emergere nuove opportunità di business e per ottimizzare le campagne. Ciò è possibile grazie alla raccolta, all'analisi e alla modellazione dei dati sul comportamento dei clienti e sullo storico delle performance delle varie campagne che vengono effettuate.

In particolare, in questo capitolo, l'analisi predittiva sarà svolta tramite l'uso del forecasting in Power BI.

6.1.1 Il forecasting

Nella moderna gestione del business, il forecasting (la traduzione del termine può essere “possibilità di interpretare le performance e le esigenze aziendali in divenire”) rientra sempre più frequentemente nelle attività di decision-making che caratterizzano le attività di un'azienda. Il vantaggio competitivo di un'azienda, infatti, si misura, anche, nella capacità di simulare situazioni future, interpretarle ed essere



Figura 6.1. Le caratteristiche dell'analisi predittiva

in grado di organizzarsi di conseguenza, valutando in anticipo le strategie più opportune e minimizzando le risorse impiegate. Metodi statistici e avanzati strumenti di analytics fanno sì che tutto possa essere oggetto di forecasting.

Rimanendo ancorati alla realtà del business (uno dei principali impieghi del forecasting), l'utilizzo più immediato del forecasting riguarda le previsioni dei consumi. Basti pensare, a titolo esemplificativo, alla distribuzione in ambito farmaceutico, in cui si ha l'esigenza di ottimizzare il livello delle giacenze, garantendo, al tempo stesso, l'assenza di stockout (onde evitare la mancata disponibilità di farmaci critici). Alla base di un'analisi previsionale di questo tipo c'è un principio semplice: utilizzare le informazioni del passato come strumento per la simulazione del futuro. Più nel dettaglio, si assume che il modo in cui le situazioni si sono evolute fino ad oggi continui anche nel domani e si cercano particolari modelli (pattern) nei dati storici, con l'obiettivo di proiettarli in avanti nel tempo.

Nel dettaglio, dunque, il forecasting è l'attività che si occupa della vera e propria predizione, con la caratteristica di mostrare, oltretutto, il "Forecast Value" ed una "fascia di confidenza", caratterizzata dal "Confidence High Bound" e dal "Confidence Low Bound", valori che indicano entro quale range, nel caso di limitati stravolgimenti, si troverà il dato nel futuro.

Introducendo strumenti di forecasting nel case study diventa, dunque, possibile limitare l'incertezza che caratterizza gli scenari futuri del dataset.

6.2 Forecasting in Power BI mediante l'utilizzo di oggetti visivi R

Nel Capitolo 2 abbiamo introdotto, tra le tante potenzialità messe a disposizione da Power BI, la sua integrazione con il linguaggio di programmazione R, che consente di

realizzare analisi e statistiche avanzate. In particolare, una importante funzionalità di R è che esso è un ottimo strumento per la creazione di modelli di predizione, ma anche per la visualizzazione di modelli esponenziali e modelli ARIMA.

Inoltre, Power BI consente di estendere le diverse visualizzazioni messe a disposizione attraverso il marketplace e, tra queste viste aggiuntive, ce ne sono diverse che sono ottenute tramite script di R.

Per quanto riguarda il forecasting, Power BI permette di scegliere, attraverso il marketplace, due oggetti visivi di R, uno basato sul modello ETS (“Error, Trend, Seasonality”), e l'altro sul modello ARIMA (“Autoregressive Integrated Moving Average”).

Nel nostro caso è stato importato in Power BI l'oggetto visivo “Forecasting with ARIMA” (Figura 6.2), che si basa sul concetto di “Time series forecasting”.

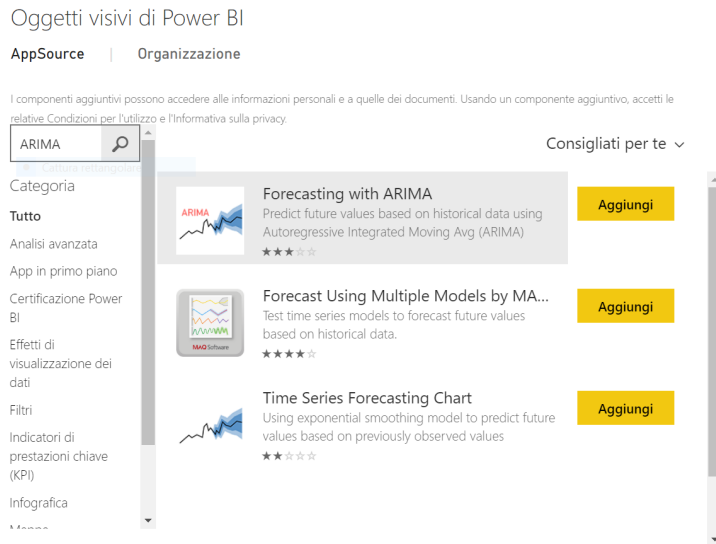


Figura 6.2. Importazione dell'oggetto visivo relativo al forecasting mediante ARIMA

Prima di analizzare tale modello, facciamo una breve introduzione sulle serie temporali, oggetto di studio del modello ARIMA stesso. La comprensione di tale concetto risulta, infatti, di fondamentale importanza per le analisi predittive che faremo nel prosieguo del capitolo.

6.2.1 Le serie temporali

Una serie temporale (Figura 6.3), anche nota come serie storica, è una sequenza di osservazioni relative ad una sola variabile registrate a intervalli di tempo regolari. Esempi di serie storiche sono: l'andamento dei prezzi delle materie prime, gli indici di borsa, lo spread, il tasso di disoccupazione, etc.

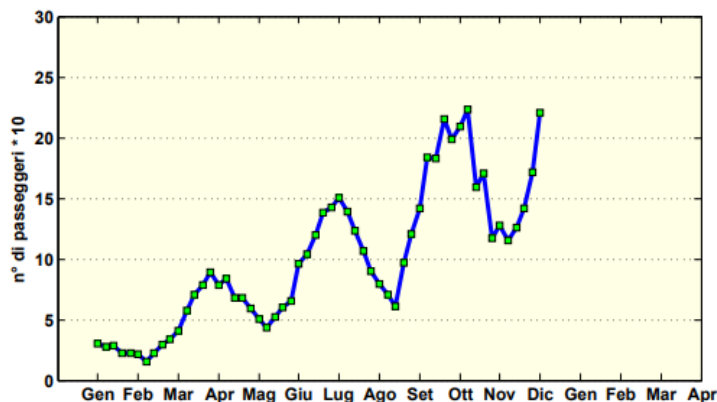


Figura 6.3. Esempio di serie temporale

In particolare, da questa serie di dati si vogliono estrarre informazioni per la caratterizzazione del fenomeno in osservazione e per la previsione di valori futuri della serie stessa.

A seconda della frequenza delle osservazioni, una serie temporale può essere oraria, giornaliera, mensile, trimestrale e annuale.

Qualsiasi serie temporale, inoltre, può essere suddivisa nei seguenti componenti: livello di base, trend, stagionalità ed errore. Un trend esiste nel momento in cui si può osservare una pendenza crescente o decrescente nella serie temporale. Esso indica, dunque, una variazione nel lungo termine. La stagionalità esiste, invece, quando si può osservare un pattern distinto ripetuto tra intervalli regolari, dovuto a fattori stagionali. Essa rappresenta, dunque, una componente periodica nel periodo osservato. Un aspetto da sottolineare è che non è obbligatorio che tutte le serie temporali debbano avere un trend e/o una stagionalità.

Una proprietà importante delle serie temporali è la stazionarietà. Una serie temporale si dice stazionaria se i suoi valori non dipendono dal tempo. In essa la media e la varianza, dunque, devono essere costanti nel tempo. Una serie stazionaria, inoltre, è priva di effetti stagionali.

Gran parte dei metodi di predizione statistica sono progettati per operare con serie temporali stazionarie. Pertanto, il primo passo del processo di predizione di una serie temporale è, tipicamente, quello di effettuare delle trasformazioni per convertire una serie non stazionaria in una stazionaria.

Per rendere stazionaria una serie, l'approccio più comune consiste nel differenziare la serie almeno una volta e, comunque, un numero sufficiente di volte per renderla approssimativamente stazionaria. Nel dettaglio, la differenziazione di una serie consiste nel sottrarre dal valore corrente il valore precedente. Inoltre, se la prima differenziazione di una serie non consente di avere una serie stazionaria, si procede con una seconda differenziazione, e così via fino a quando la serie non diventa stazionaria.

Grazie a queste nozioni preliminari siamo, ora, in grado di introdurre il modello ARIMA, con particolare riferimento alla sua implementazione all'interno di Power

BI.

6.2.2 Il modello ARIMA

Il modello ARIMA comprende, in realtà, una classe generale di modelli per la previsione di serie temporali che possono essere stazionarie, ovvero le cui proprietà non dipendono dal momento in cui la serie viene osservata. Abbiamo in precedenza introdotto anche l'oggetto visivo di Power BI basato su uno script di R e sul modello ETS che, a differenza di ARIMA, è un modello di "exponential smoothing" che si basa, cioè, su una descrizione dell'andamento e della stagionalità dei dati. I modelli ARIMA, invece, mirano a descrivere le autocorrelazioni nei dati.

In particolare, l'algoritmo alla base di ARIMA si basa sui seguenti parametri:

- Il parametro p , che rappresenta l'ordine della componente regressiva e si riferisce all'uso di valori passati nell'equazione di regressione.
- Il parametro d , che rappresenta il grado di differenziazione della componente integrata; esso si riferisce, cioè, alla sottrazione dei valori attuali e precedenti della componente integrata effettuata d volte, come precedentemente illustrato.
- Il parametro q , che rappresenta l'ordine della componente di media mobile e determina il numero di termini di errore da includere nel modello.

Oltre a questi parametri di base, ci sono anche i parametri P , D e Q , che servono a descrivere la componente stagionale di "m" periodi. Perciò, nel caso in cui la stagionalità non sia presente, tali parametri non saranno, a loro volta, presenti.

Inoltre, tutti i parametri appena presentati possono essere stabiliti automaticamente dall'algoritmo, oppure impostati manualmente.

6.3 Attività di forecasting svolte in Power BI

Entriamo, ora, nel dettaglio delle operazioni svolte all'interno di Power BI Desktop per quanto riguarda l'attività di analisi predittiva ed il forecasting.

In realtà, Power BI permette di eseguire l'attività di forecasting senza dover importare oggetti visivi di R; per fare ciò, è sufficiente selezionare un grafico a linee precedentemente creato nella finestra di visualizzazione report, cliccare sul menù "Analisi" e, successivamente, selezionare la voce "Previsione", come si evince dalla Figura 6.4.

In particolare, dal menù "Previsione" è possibile decidere la lunghezza del forecast (ad esempio giorni, mesi, trimestri, anni, etc.), ignorare gli ultimi punti, scegliere il livello di confidenza della previsione ed, eventualmente, la stagionalità. Nella Figura 6.4 sotto riportata è stata scelta, ad esempio, una previsione su 10 giorni con un livello di confidenza del 95%.

In Figura 6.5 riportiamo un cruscotto con tre grafici a linee, ai quali è stata applicata la funzionalità di previsione integrata in Power BI.

Per tutti i grafici si è scelta una lunghezza della previsione pari a 10 giorni (a partire dal 27 Maggio 2020) ed un intervallo di confidenza del 95%. I tre campi analizzati sono quelli relativi al totale dei decessi, al numero di malati positivi e al numero di decessi giornalieri.

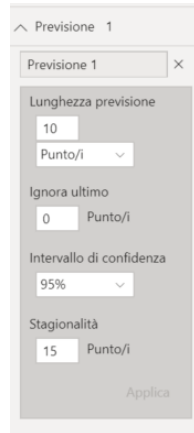


Figura 6.4. Il menù “Previsione” in Power BI

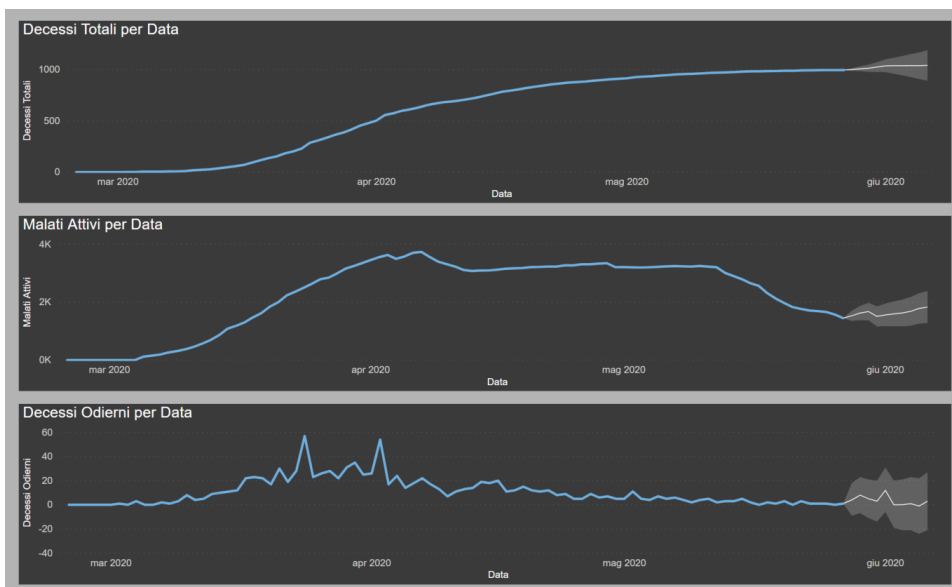


Figura 6.5. Esempi di previsioni in Power BI

Non esaminiamo nel dettaglio tali previsioni, poiché ci concentreremo sulle attività di analisi predittiva basate sul modello ARIMA, realizzate in Power BI tramite l’oggetto visivo “Forecasting with ARIMA”, le cui modalità di importazione sono state precedentemente descritte .

6.3.1 Decessi

L’analisi predittiva relativa ai decessi avvenuti nella Regione Marche e riconducibili al Covid-19 è stata svolta mediante la realizzazione di due grafici basati sul modello ARIMA. Riportiamo, in Figura 6.6, il cruscotto contenente tali grafici.

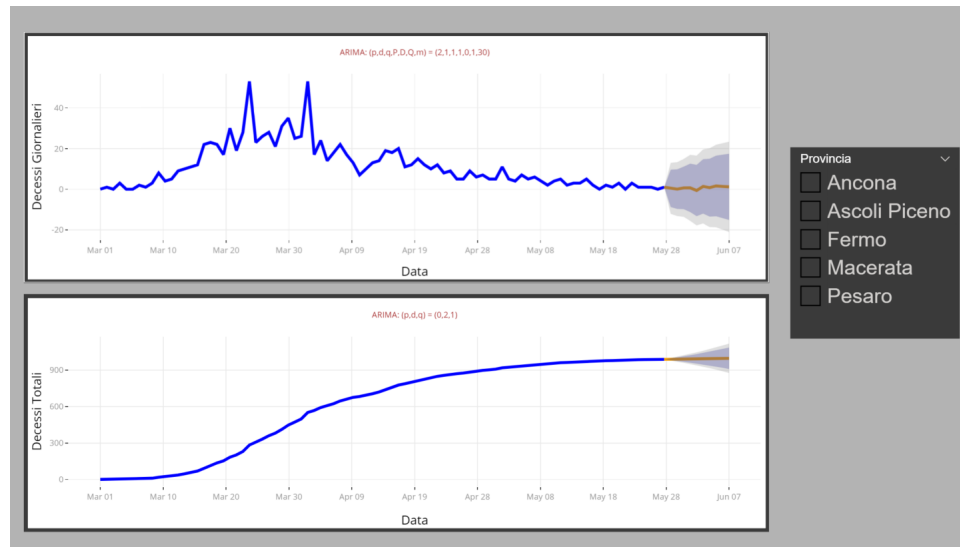


Figura 6.6. Cruscotto di tipo predittivo relativo ai decessi

I grafici riportano, rispettivamente, l'andamento dei decessi giornalieri e la crescita del totale dei decessi. Si tratta di due aspetti già analizzati in precedenza senza, però, l'applicazione del forecasting. In questo caso, entrambi i grafici sono stati realizzati mediante l'oggetto visivo basato sul modello ARIMA. In particolare, è stata impostata la durata della previsione a 10 giorni (a partire dal 27 Maggio 2020), un livello di confidenza base di 0.999 ed un livello di confidenza secondario pari a 0.95. Per quanto riguarda la stagionalità, invece, si è deciso di lasciare l'impostazione "automatic". Tali impostazioni sono riportate in Figura 6.7. Le stesse impostazioni, inoltre, verranno mantenute anche per i successivi grafici basati sul modello ARIMA.

Per quanto riguarda i decessi giornalieri, si noti come la previsione mostri un andamento pressoché costante di tale numero, anche se si è in presenza di valori di upper bound e lower bound piuttosto elevati, che indicano una bassa precisione della predizione.

Per quanto riguarda l'andamento dei decessi totali, invece, la previsione mostra un andamento quasi rettilineo, con una crescita davvero minimale di tale numero. In questo caso, inoltre, si può notare come il livello di approssimazione sia migliore, essendo più ravvicinati i valori di upper bound e di lower bound.

Il cruscotto in questione, inoltre, presenta, nella parte di destra, un oggetto visivo di tipo "Filtro dei dati" che consente di filtrare i dati in base alle 5 province marchigiane mediante una semplice selezione della provincia di interesse. Non analizziamo nel dettaglio le caratteristiche relative alle singole province, in quanto gli andamenti sono pressoché identici a quelli relativi all'intera Regione.

Forecasting settings

Forecast length
10

Confidence level
0.999

Confidence level #2
0.95

Ripristina valori predefi...

Seasonality At...

Target seasonal factor
automatic

Figura 6.7. Impostazioni relative al forecasting con ARIMA

6.3.2 Malati

Per quanto riguarda l'aspetto relativo ai malati, ovvero le persone positive al Covid-19 all'interno della Regione Marche, è stato realizzato un ulteriore cruscotto riportante due grafici a righe basati sul forecasting mediante il modello ARIMA (Figura 6.8).

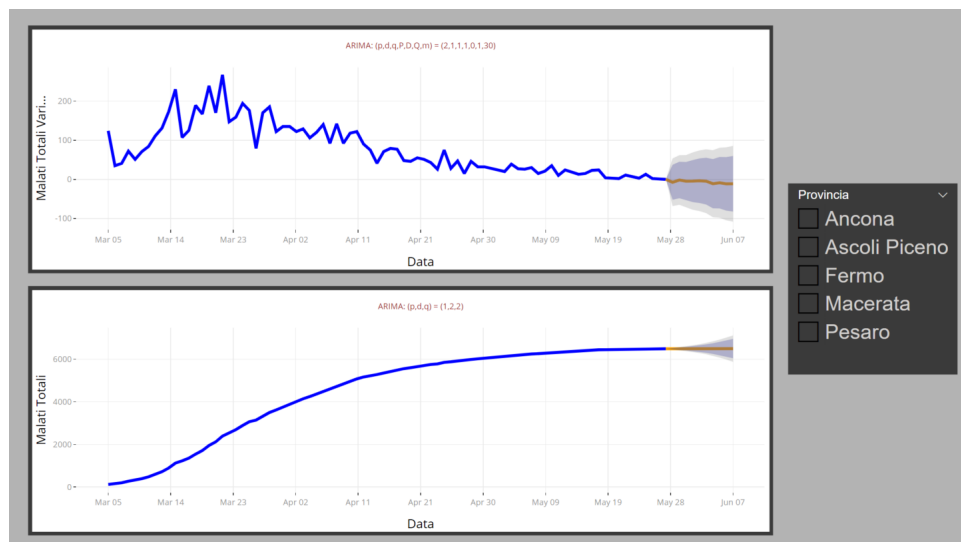


Figura 6.8. Cruscotto di tipo predittivo relativo ai malati totali

Anche in questo caso, si è deciso di analizzare la variazione giornaliera del numero dei malati totali, ma anche l'andamento del totale dei malati stessi. Entrambi

i forecast sono basati su una durata della previsione di 10 giorni (a partire dal 27 Maggio 2020), con un livello di confidenza primario pari a 0.999 e uno secondario di 0.95. Per quanto riguarda la stagionalità, invece, si è deciso di scegliere l'impostazione "automatic".

Relativamente alla variazione giornaliera dei malati totali, si nota un andamento costante che si attesta quasi sempre su valori di poco inferiori allo zero. I valori di upper bound e lower bound, però, sono abbastanza distanti l'uno dall'altro.

Anche questo cruscotto è interattivo, in quanto consente di selezionare la provincia di interesse per visualizzare la previsione ad essa relativa.

Per quanto riguarda l'aspetto relativo ai malati, inoltre, si è deciso di effettuare, anche, un'analisi del solo numero dei malati attualmente positivi. Anche in questo caso, sono stati realizzati due grafici basati sul modello ARIMA, utilizzando le stesse impostazioni di forecasting analizzate in precedenza (Figura 6.9).

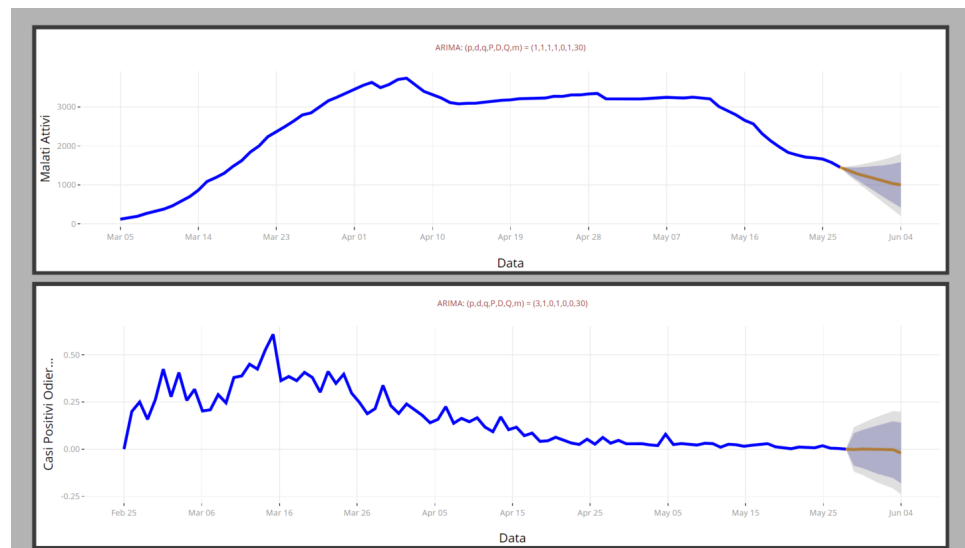


Figura 6.9. Cruscotto di tipo predittivo relativo ai malati attivi

Si noti come il numero di malati attivi, riportato nel primo grafico, sia destinato a decrescere ulteriormente nei 10 giorni successivi al 27 Maggio 2020. Per quanto riguarda i casi positivi giornalieri, invece, l'andamento si dimostra lievemente decrescente e tendente allo zero. In entrambi i grafici, comunque, l'intervallo di indecisione risulta accettabile, essendo i valori di upper bound e lower bound piuttosto ravvicinati.

6.3.3 Ricoveri

L'ultimo aspetto preso in considerazione nella nostra attività di analisi predittiva è quello riguardante l'andamento dei ricoveri riconducibili al Coronavirus, all'interno delle strutture ospedaliere della Regione Marche.

A tal proposito è stato realizzato un apposito cruscotto riportante due grafici relativi, rispettivamente, ai ricoveri in strutture di terapia intensiva e ai ricoveri in tutti gli ospedali della regione, senza distinzione tra i vari reparti terapeutici.

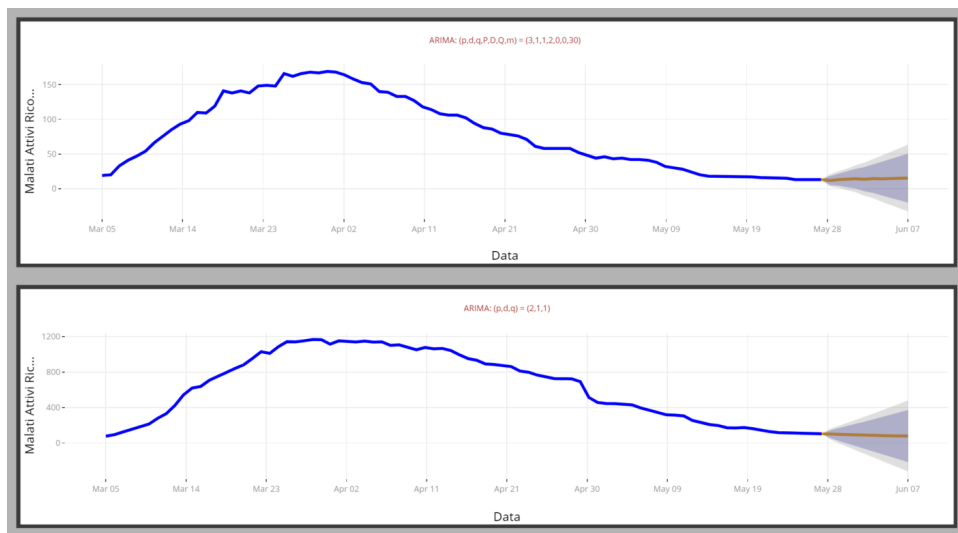


Figura 6.10. Cruscotto di tipo predittivo relativo ai malati ricoverati in terapia intensiva

Entrambi le previsioni sono basate sul modello ARIMA ed utilizzano le stesse impostazioni analizzate nei cruscotti precedenti. Si è deciso, cioè, di adottare, anche in questo caso, una durata della previsione pari a 10 giorni (a partire dal 27 Maggio 2020), un livello di confidenza primario di 0.999, un livello di confidenza secondario di 0.995 ed una stagionalità impostata ad “automatic”.

Si osservi come il risultato della previsione sia molto simile nei due grafici. Il forecast relativo al numero di malati ricoverati in terapia intensiva (grafico in alto) mostra un andamento in leggerissima crescita, mentre quello relativo ai ricoverati in tutti i reparti della Regione (grafico in basso) mostra una decrescita abbastanza marcata, che porta tale numero dai 103 ricoverati del 27 Maggio 2020, ai probabili 77 ricoverati al 6 Giugno 2020. Inoltre, in entrambi i grafici, la funzione di previsione risulta essere abbastanza precisa, in quanto si ha un intervallo di indecisione piuttosto ristretto.

Lezioni apprese

In questo capitolo parleremo delle lezioni apprese nella realizzazione della presente tesi, passando in rassegna le principali conoscenze acquisite.

7.1 Premessa

In questa tesi è stato studiato lo sviluppo della pandemia da Covid-19 all'interno della Regione Marche. Tale pandemia, sviluppatasi in Cina alla fine del 2019, ha avuto una forte espansione anche in Italia a partire dal mese di Febbraio 2020. La Regione Marche, in particolare, attraverso il cosiddetto GORES ("Gruppo Operativo Regionale per le Emergenze Sanitarie"), si è fin da subito attivata per diffondere i dati riguardanti la diffusione del virus all'interno della regione. Il nostro studio è basato proprio sull'analisi di questi dati, messi a disposizione direttamente nel sito della Regione Marche. Durante questo elaborato, per raggiungere i nostri risultati, abbiamo usato un tool di Business Intelligence, ovvero Power BI, che ci ha permesso di affrontare le fasi principali della data analysis.

In particolare, nel corso della suddetta tesi, sono state apprese ed approfondite le seguenti tematiche:

- l'analisi dei dati e tutte le sue fasi;
- la Business Intelligence;
- il tool Power BI;
- il formato JSON per lo scambio di informazioni.

Analizziamo, ora, una per una le lezioni apprese appena illustrate.

7.2 L'analisi dei dati

La nostra attività di studio ci ha permesso di entrare in contatto con il mondo dell'analisi dei dati, che riveste un ruolo fondamentale ai giorni d'oggi. Nel corso della tesi abbiamo affrontato le varie fasi principali che caratterizzano una qualunque attività di data analytics, a partire dalle attività di ETL, per proseguire con le attività

di analisi descrittiva, fino ad arrivare alle attività di analisi predittiva. Abbiamo, dunque, compreso l'importanza di tutti questi processi, inquadrati nell'ambito della diffusione della pandemia da Covid-19 all'interno della Regione Marche.

In realtà, il concetto di data analytics e, in generale, di Big Data, si estende in vari campi applicativi, come, ad esempio, i Social Media, che rappresentano uno degli ambiti di maggior utilizzo e diffusione della data analytics, oppure anche l'economia e la finanza, ambito in cui l'analisi dei dati risulta di fondamentale importanza per comprendere le esigenze ed i comportamenti dei clienti, nonché l'andamento delle vendite e della produzione a livello aziendale ed industriale.

Le conoscenze acquisite durante il nostro studio ci consentono, ora, di approfondire ancora di più nel dettaglio i processi di una campagna di analisi dei dati che, come sopra espresso, possono essere estesi ad un vasto numero di applicazioni. Riteniamo che tali conoscenze siano fondamentali nel mondo di oggi, caratterizzato da un crescente processo di digitalizzazione in ogni aspetto della vita quotidiana. Tutto ciò porta, infatti, alla generazione di un'enorme quantità di dati e alla necessità di analizzarli per estrarne contenuti informativi utili, al fine di trasformarli in una vera fonte di conoscenza.

7.3 La Business Intelligence

Le attività di data analytics e l'utilizzo di un tool come Power BI, inoltre, ci hanno permesso di entrare in contatto con un altro aspetto fondamentale del mondo dei dati, ossia la Business Intelligence (Figura 7.1).

La Business Intelligence, o semplicemente BI, infatti, è un insieme di processi e strumenti attraverso i quali un'azienda riesce a raccogliere dati di diversa natura per analizzarli e trarne decisioni strategiche. La Business Intelligence si realizza, in particolare, mediante vere e proprie applicazioni che permettono di trasformare grandi quantità di dati strutturati e de-strutturati, in informazioni strategiche per le attività di un'azienda. Il compito della BI è, infatti, quello di fornire report, statistiche, indicatori e grafici aggiornati utili a coloro che devono prendere le decisioni per effettuare scelte più corrette al fine della promozione del business e per rendere, quindi, più competitive le realtà in cui operano.

Nel corso del nostro caso di studio, abbiamo avuto la possibilità di realizzare cruscotti interattivi impiegando diverse tipologie di grafici fino ad ottenere, alla fine, un vero e proprio report relativo allo sviluppo della pandemia da Covid-19 all'interno della Regione Marche. Oltre a tutto ciò, abbiamo svolto altre attività tipiche della BI, come la raccolta, l'integrazione, la pulizia e la validazione dei dati. Abbiamo, anche, realizzato dei veri e propri calcoli analitici all'interno di Power BI e, inoltre, abbiamo effettuato anche diversi confronti temporali, che rappresentano uno degli aspetti fondamentali della BI, specialmente in ambito industriale.

Le conoscenze di BI acquisite possono, dunque, essere estese ad una molteplicità di aree di applicazione, soprattutto in ambito industriale, come, ad esempio, il *Controllo di gestione*, o la *Gestione delle vendite*, ma anche il campo delle *Human Resources*.



Figura 7.1. La Business Intelligence e le sue fasi

7.4 Il tool Power BI

Tutte le operazioni tipiche dell'analisi dei dati sono state eseguite sfruttando il tool Power BI, in particolare la sua versione Power BI Desktop. Power BI si configura come uno dei principali software di Business Intelligence.

In particolare, nel corso del nostro caso di studio, abbiamo avuto modo di approfondire e verificare personalmente le potenzialità di questo strumento.

L'aspetto sicuramente più interessante di Power BI è la possibilità di eseguire le operazioni di ETL direttamente dal software stesso in maniera veloce, efficace e, soprattutto, intuitiva. Tutto ciò grazie alla possibilità di modificare i dati tramite l'editor di Power Query, prima di caricare le cosiddette tabelle "pulite". Questo aspetto è stato fondamentale nel nostro caso, in quanto ha permesso di lavorare sul file JSON contenente i dati utilizzati per la nostra analisi, trasformandolo in una struttura di tipo relazionale effettuando, semplicemente, una sequenza di passaggi "elementari".

Un'altro aspetto di Power BI che abbiamo avuto modo di affrontare ed approfondire è quello legato alla realizzazione di cruscotti mediante i vari oggetti visivi messi a disposizione dallo stesso tool. A tale proposito, abbiamo avuto il modo di vedere un altro punto forte di questo software, ovvero la possibilità di usare script visivi creati con il linguaggio R, grazie ai quali è possibile importare oggetti visivi che, di base, non sono inclusi nel pacchetto di grafici standard offerto da Power BI. Nel nostro caso di studio, questi script, in particolare il grafico basato sul modello ARIMA, si sono rivelati fondamentali, specialmente per la fase di analisi predittiva.

Nel corso dei nostri studi, inoltre, abbiamo approfondito le conoscenze riguardo al linguaggio delle formule DAX, anch'esso rivelatosi fondamentale nel nostro caso di studio, poichè ci ha permesso di completare i dati a disposizione mediante la rea-

lizzazione di misure, colonne e tabelle calcolate, senza dover modificare direttamente il modello dei dati.

7.5 Il formato JSON per lo scambio di informazioni

La nostra attività di analisi all'interno di Power BI è iniziata con l'importazione dei dati di interesse. Tali dati, inizialmente, erano organizzati secondo una struttura piuttosto complessa all'interno di un file di tipo JSON, come abbiamo potuto analizzare nel dettaglio all'interno del Capitolo 3. Un'attività preliminare alla vera e propria analisi dei dati all'interno di Power BI ha riguardato l'analisi approfondita del file in formato JSON in questione, al fine di capirne la struttura e, anche, per individuare eventuali errori o problemi. Durante questa fase siamo entrati nel dettaglio delle caratteristiche di questo formato di scambio di dati, che illustriamo brevemente di seguito.

JSON (“JavaScript Object Notation”) è un semplice formato per lo scambio di dati. Per le persone è facile da leggere e scrivere, mentre per le macchine risulta facile generarlo e analizzarne la sintassi. Si basa sul 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 un formato di testo completamente indipendente dal linguaggio di programmazione, ma utilizza convenzioni conosciute dai programmatori di linguaggi della famiglia del C, come C, C++, C#, Java, JavaScript, Perl, Python, e molti altri. Questa caratteristica fa di JSON un linguaggio ideale per lo scambio di dati.

Nel dettaglio, JSON è basato su due strutture (Figura 7.2):

- *Un insieme di coppie nome/valore.* In diversi linguaggi, questo è realizzato come un oggetto, un record, uno 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.

Queste sono strutture di dati universali. Virtualmente tutti i linguaggi di programmazione moderni li supportano in entrambe le forme. È sensato che un formato di dati che è interscambiabile con linguaggi di programmazione debba essere basato su queste strutture.

Dunque, come appena visto, il formato JSON risulta estremamente efficace ed efficiente per permettere lo scambio di un qualunque tipo di informazione, e la sua conoscenza, ai giorni d'oggi, risulta di fondamentale importanza per qualsiasi applicazione in ambito informatico, e non solo.

```
{
  "employees": [
    { "firstName":"John" , "lastName":"Doe" , "age": 18 },
    { "firstName":"Anna" , "lastName":"Smith" , "age": 30},
    { "firstName":"Peter" , "lastName":"Jones" , "age": 41}
  ]
}
```

```
{
  "menu": {
    "id": "file",
    "value": "File",
    "popup": {
      "menuitem": [
        { "value": "New", "onclick": "CreateNewDoc()" },
        { "value": "Open", "onclick": "OpenDoc()" },
        { "value": "Close", "onclick": "CloseDoc()" }
      ]
    }
  }
}
```

Figura 7.2. La struttura di un file in formato JSON

Conclusioni

In questa tesi è stata studiata la pandemia da Covid-19 e, in particolare, la sua evoluzione all'interno della Regione Marche. Sono stati analizzati tutti i vari aspetti di questa pandemia, a partire dai decessi causati dal virus, fino ad arrivare ai dati relativi a tamponi, casi positivi, e dimessi e guariti dalla malattia. Sono stati utilizzati i dati forniti direttamente dalla Regione Marche attraverso il “Gruppo Operativo Regionale per le Emergenze Sanitarie”, noto anche come GORES. Tali dati fanno riferimento ai vari aspetti legati allo sviluppo della pandemia all'interno della sola regione e rappresentano la vera base su cui sono stati applicati gli strumenti di Business Intelligence e Data Analytics.

Per garantire una maggior chiarezza della tematica di studio, il primo passo affrontato è stato quello di introdurre le fasi dell'espansione e le caratteristiche principali della pandemia da Covid-19, diffusasi intorno alla fine del 2019 in Cina, e poi sviluppatasi in tutto il resto del mondo, Italia compresa, fin dagli inizi del 2020.

Successivamente a questa fase introduttiva, è stato studiato ed analizzato nei suoi vari aspetti il tool Power BI, per poter conoscere a fondo le sue caratteristiche principali e, soprattutto, le sue funzionalità. Tra i vari tool di analisi disponibili sul mercato che consentono di operare con grandi quantità di dati, Power BI rappresenta un vero e proprio leader del settore. Attraverso tale software, e grazie alle sue innumerevoli potenzialità, sono state effettuate analisi descrittive, nella prima parte dell'elaborato, e analisi predittive, nella parte finale.

Lo step successivo della tesi, prima di cominciare con i veri e propri processi di analisi, è consistito nell'attività di preparazione dei dati, al fine di renderli adatti e compatibili con il nostro studio e con il tool Power BI Desktop. È stato affrontato, quindi, il processo di ETL (“Extract, Transform and Load”) direttamente all'interno di Power BI. Questo processo è stato fondamentale per il prosieguo delle nostre attività, in quanto ci ha consentito di “pulire” i dati in nostro possesso, in modo tale da poter mantenere solo quelli utili per il nostro caso di studio e in un formato semplice per l'analisi.

Una volta terminato il processo di pulizia e formattazione dei dati (utilizzando, soprattutto, l'Editor di Power Query, integrato in Power BI Desktop), si è proceduto con la vera e propria analisi del nostro modello dei dati. Si è deciso di suddividere tale attività per tematiche:

- Nella prima parte sono stati realizzati dei cruscotti per illustrare ed analizzare i dati relativi ai soli decessi causati dal Coronavirus all'interno della Regione Marche.
- Nella seconda parte sono stati realizzati altri cruscotti relativi ai dati riguardanti i casi positivi al virus all'interno delle Marche, i ricoveri nelle varie strutture ospedaliere della Regione, e i malati dimessi e guariti dal virus.

Durante queste fasi sono stati analizzati i dati relativi all'intera regione in generale, ma anche quelli riguardanti le singole province. Si è, inoltre, cercato di allargare la panoramica del nostro studio relativo alla sola Regione Marche, facendo continui confronti con lo sviluppo della pandemia in tutto il territorio italiano, al fine di stabilire il grado di gravità della situazione delle Marche, collocandola nel contesto italiano.

Dopo aver completato l'analisi di tutti gli aspetti presenti nel nostro modello di dati in Power BI, si è deciso di affrontare un'ultima fase nel nostro studio, ovvero l'analisi predittiva. Quest'ultima attività, in particolare, è stata effettuata utilizzando un modello di analisi predittiva noto come ARIMA, che ci ha permesso di realizzare grafici ottenuti mediante script del linguaggio R, perfettamente integrato in Power BI. In particolare, durante questa fase, sono state effettuate attività di forecasting per quanto riguarda i dati dei decessi, dei casi positivi e dei ricoveri all'interno della sola Regione Marche.

Riteniamo che le analisi effettuate nel corso di questa tesi, così come tutto il processo di raccolta e pulizia dei dati, non debbano essere considerate come una semplice conclusione del nostro lavoro, ma, al contrario, devono costituire un punto di partenza per possibili analisi future. Una delle caratteristiche principali delle attività di analisi di dati e, in generale, dei Big Data, è, infatti, la possibilità di riutilizzo delle informazioni; questo significa che i nostri risultati possono essere usati come input per eventuali sviluppi futuri. Questo concetto diventa ancora più rilevante nel caso di una pandemia, come quella da noi analizzata. Sappiamo, infatti, come la storia dell'umanità sia caratterizzata da numerosi fenomeni epidemici e pandemici. Infatti, da quando l'essere umano ha iniziato a organizzarsi in società e a creare nuclei di persone che convivono insieme nello stesso spazio, le malattie contagiose hanno assunto un ruolo importante. Per tutti questi motivi, riteniamo che i nostri dati e, soprattutto, le nostre analisi, possano tornare molto utili in futuro, in caso di sviluppo di nuove pandemie o, comunque, di semplici epidemie.

Inoltre, i nostri cruscotti possono essere facilmente riutilizzati in caso di nuovi sviluppi della pandemia da Covid-19. Questo è un aspetto da sottolineare soprattutto in vista del fatto che, molto spesso, le epidemie e le pandemie sono caratterizzate dalle cosiddette "ondate di ritorno" che possono essere addirittura più violente della prima fase di espansione e diffusione del virus.

Ci teniamo a sottolineare, anche, come, nonostante i nostri cruscotti siano relativi alla sola Regione Marche, essi possono essere facilmente estesi non solo alle altre regioni italiane, ma anche ad intere nazioni e aree geografiche.

Infine, riteniamo che il nostro studio e le nostre analisi possano essere ulteriormente sviluppate ed approfondite nel futuro. Sebbene siano stati trattati molti aspetti legati allo sviluppo del Coronavirus all'interno della Regione Marche, pensiamo che possano essere introdotti nuovi grafici al fine di ricavare ulteriore conoscenza, anche aggiungendo ulteriori dati al nostro modello. In particolare, si potrebbero

sfruttare le potenzialità di Power BI e, soprattutto, la sua integrazione con R, per effettuare analisi predittive e prescrittive ad un maggiore livello di profondità.

Riferimenti bibliografici

1. 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.
2. Introduzione a Power BI Desktop. <https://docs.microsoft.com/it-it/power-bi/fundamentals/desktop-getting-started>, 2020.
3. 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.
4. Tracciamento Coronavirus, Financial Times. <https://www.ft.com/content/a26fbf7e-48f8-11ea-aeb3-955839e06441>, 2020.
5. L. Abenavoli, P. Cinaglia, F. Luzzza, I. Gentile, and L. Boccuto. Epidemiology of coronavirus disease outbreak: the Italian trends. *Rev Recent Clin Trials*, 15, 2020.
6. A. Aspin. *High Impact Data Visualization with Power View, Power Map, and Power BI*. Apress, 2014.
7. A. Aspin. *Pro Power BI Desktop*. Springer, 2016.
8. 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.
9. A. Ferrari and M. Russo. *The Definitive Guide to DAX: Business Intelligence with Microsoft Excel, SQL Server Analysis Services, and Power BI*. Microsoft Press, 2015.
10. A. Ferrari and M. Russo. *Introducing Microsoft Power BI*. Microsoft Press, 2016.
11. A. Ferrari and M. Russo. *Analyzing Data with Power BI and Power Pivot for Excel*. Microsoft Press, 2017.
12. S. Finlay. *Predictive analytics, data mining and big data: Myths, misconceptions and methods*. Springer, 2014.
13. A. Gandomi and M. Haider. Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2):137–144, 2015.
14. 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.
15. T. Lachev and E. Price. *Applied Microsoft Power BI (2nd Edition): Bring Your Data to Life!* Prologika Press, 2017.
16. E. Livingston and K. Bucher. Coronavirus disease 2019 (COVID-19) in Italy. *Jama*, 323(14):1335–1335, 2020.
17. S. Machiraju and S. Gaurav. *Power BI Data Analysis and Visualization*. De Gruyter, 2018.

18. P.N. Mahalle, N.P. Sable, N.P. Mahalle, and G.R. Shinde. Data analytics: COVID-19 prediction using multimodal data. *Preprints*, 2020.
19. F. Marinuzzi and M. Liciani. *Basi di dati e Big Data: come estrarre valore dai propri dati*. Youcanprint, 2016.
20. World Health Organization. Coronavirus disease 2019 (COVID-19): situation report, 72. 2020.
21. B. Powell. *Microsoft Power BI Cookbook: Creating Business Intelligence Solutions of Analytical Data Models, Reports, and Dashboards*. Packt Publishing Ltd, 2017.
22. 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.
23. A. Remuzzi and G. Remuzzi. COVID-19 and Italy: what next? *The Lancet*, 2020.
24. A. Rezzani. *Big Data Analytics. Il manuale del data scientist*. Apogeo Education, 2017.
25. P. Russom. Big data analytics. *TDWI best practices report, fourth quarter*, 19(4):1–34, 2011.
26. T. Singhal. A review of coronavirus disease-2019 (COVID-19). *The Indian Journal of Pediatrics*, pages 1–6, 2020.
27. C. Webb. *Power query for power BI and Excel*. Apress, 2014.

Ringraziamenti

Finalmente ho raggiunto il termine di questa prima parte del mio percorso universitario. Durante questi tre anni, nonostante le tante difficoltà, ho avuto la possibilità di conoscere ed imparare cose di cui sono veramente appassionato. È stato un percorso difficile, specialmente all'inizio: mi ricordo ancora le prime lezioni, i primi esami ed i tanti momenti di tensione, di ansia e preoccupazione. Le mie motivazioni, però, sono sempre state forti, fin dall'inizio, e questo mi ha permesso di andare sempre avanti, a testa alta, senza abbassarmi di fronte alle difficoltà, che in un percorso universitario sono sempre presenti. Durante questi anni, l'università e gli studi hanno occupato la maggior parte del mio tempo e, devo dire, che non è affatto tempo sprecato, anzi, ne è valsa veramente la pena!

Per tale motivo, mi sento di ringraziare tutti i professori che mi hanno accompagnato in questo percorso con grande professionalità; è grazie a loro se oggi ritengo di aver imparato moltissime cose sull'informatica, sulla matematica e, anche, sulla vita.

In particolare, vorrei ringraziare il Professore Domenico Ursino, per avermi supportato fin dall'inizio della presente tesi, con grandissima disponibilità, pazienza, e anche simpatia. Vorrei ringraziare anche i suoi collaboratori Gianluca Bonifazi e Luca Virgili, che sono stati fondamentali per me all'inizio dell'attività di tirocinio, dandomi consigli e un importante aiuto.

Inoltre, voglio fare un grande ringraziamento ai miei amici e compagni di corso, che, sicuramente, hanno contribuito a rendere più spensierati e divertenti questi anni di studi.

Infine, desidero ringraziare la mia Famiglia che, fin dall'inizio, mi ha sempre supportato ogni giorno, in ogni esame e difficoltà. I miei genitori, in questo periodo, non mi hanno veramente fatto mancare nulla e, a dire la verità, non mi hanno mai detto di no in qualunque cosa. Mio padre e mia madre si sono sempre sacrificati per me e per mio fratello, e sono veramente contento di averli resi ancora più fieri di me!