**REPORT COVID-19**

**(PERIODO 24/02/2020-06/12/2020)**

by GOT GROUP

**Analisi dei dati relativi al COVID-19 in Italia**



***Cos’è il COVID-19?***

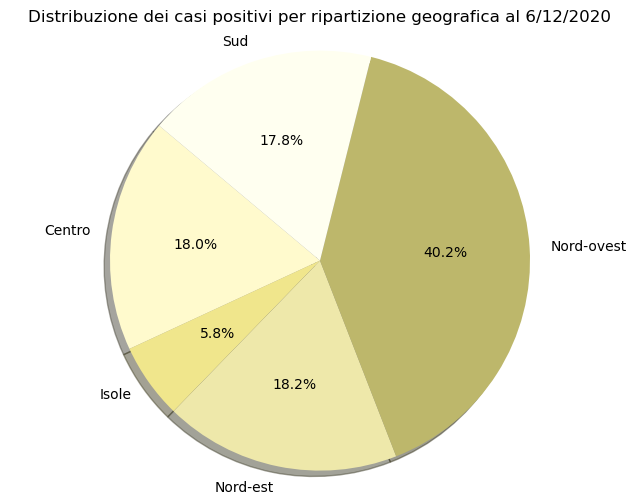
Il COVID-19 è la malattia causata da un coronavirus chiamato SARS-CoV-2. L'OMS è venuta a conoscenza di questo nuovo virus per la prima volta il 31 dicembre 2019, in seguito alla segnalazione di un gruppo di casi di polmonite virale a Wuhan, Repubblica popolare cinese.

Nel dettaglio, i coronavirus sono una grande famiglia di virus che possono causare malattie negli animali o nell'uomo. Negli esseri umani, diversi coronavirus causano infezioni respiratorie che vanno dal raffreddore comune a malattie più gravi come la sindrome respiratoria del Medio Oriente (MERS) e la sindrome respiratoria acuta grave (SARS). Il coronavirus più recentemente scoperto causa la malattia coronavirus COVID-19.

Questo virus si è diffuso rapidamente in tutto il mondo e l'Italia è stato uno dei paesi più colpiti. L'8 marzo 2020 - Il primo ministro italiano ha annunciato una totale quarantena a causa del coronavirus, limitando gli spostamenti di circa un quarto della popolazione del paese nel tentativo di limitare i contagi nell'epicentro dell'epidemia in Europa.

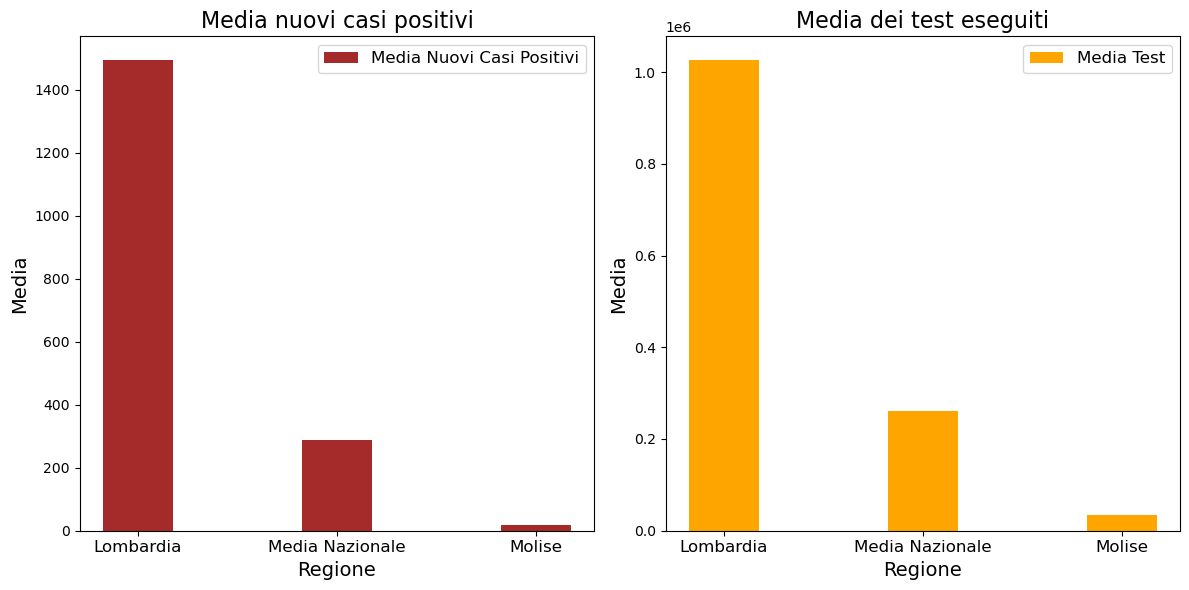
***Andamento della pandemia nelle regioni italiane***

L’epidemia in Italia ha avuto inizio da Codogno, un paesino Lombardo e da lì ha travolto tutta la penisola. Possiamo osservare dal grafico sottostante come ci sia una maggioranza di pazienti risultati positivi nel nord-ovest italiano (40,2%). Le altre zone d’Italia sono più omogenee tra loro (variano dal 17,8% al 18,2%), con le isole che risultano le meno toccate grazie appunto alla loro posizione geografica (5,8%).

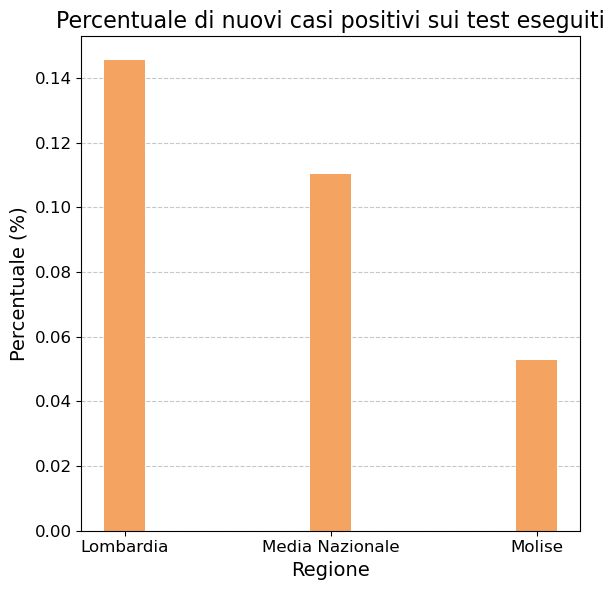


La seguente figura mostra due grafici a barre affiancati: uno per la media dei nuovi casi positivi e l’altro per la media dei test eseguiti. In particolare, vengono attenzionati la Lombardia e il Molise, e i loro risultati vengono messi a confronto con la media nazionale. Com’è possibile osservare dal grafico “Media nuovi casi positivi” la Lombardia risulta avere una media di nuovi casi positivi di gran lunga superiore alla media nazionale, e ancora superiore alla regione Molise.

La scelta di affiancare un grafico che specifichi la media dei test eseguiti deriva dal fatto che i tamponi effettuati variano notevolmente da regione a regione. Prendiamo a tal proposito, come esempi, la Lombardia e il Molise. In Lombardia durante il periodo considerato (24/02/2020-06/12/2020) venivano processati in media 1.026.507 tamponi, in Molise invece 34.124, contro la media nazionale che si attestava sui 260.081 tamponi eseguiti.

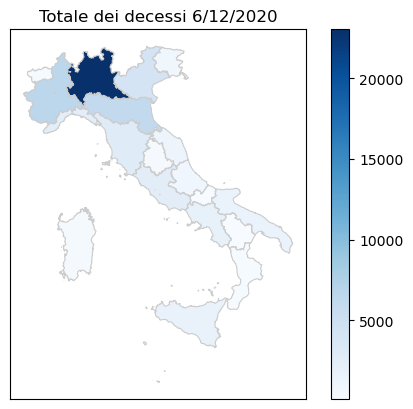
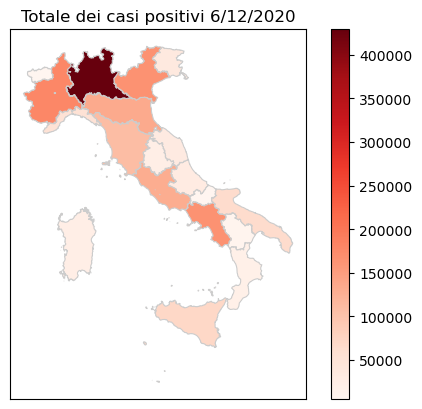


Per conoscere in maniera oggettiva la percentuale di nuovi casi positivi sui test eseguiti è stato necessario produrre un nuovo grafico che mettesse in relazione le due grandezze e restituisse questa percentuale. Come possiamo notare anche in questo caso alla Lombardia è associato un valore più elevato, pur essendo percentuali molto basse, ciò suggerisce che il numero di tamponi effettuati sia adeguato al monitoraggio dei soggetti positivi.

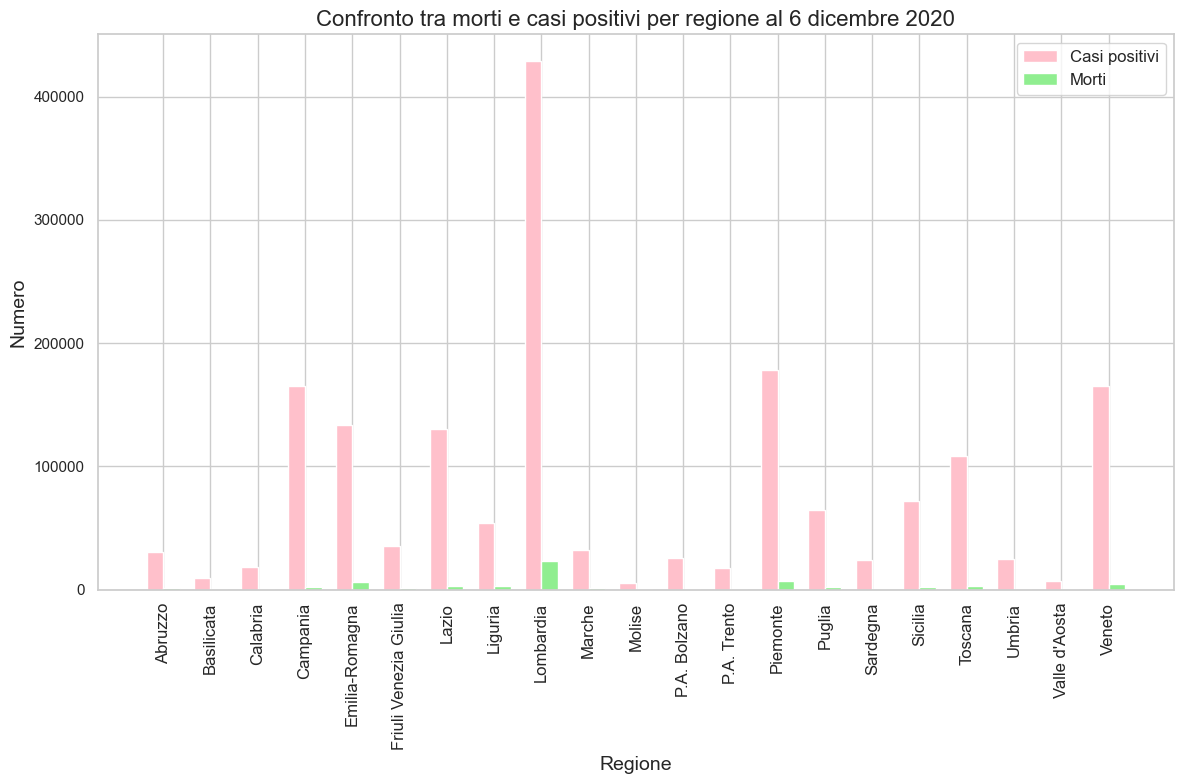


Come già detto, l’epicentro della pandemia è stato in Lombardia, inoltre, è stata anche la regione più colpita a livello nazionale, sia per quanto riguarda il totale positivi, sia per il totale dei decessi (entrambi i valori aggiornati fino al 6/12/2020).

I grafici sotto riportati forniscono un’idea più chiara della diffusione del COVID-19 e della sua mortalità nella penisola.



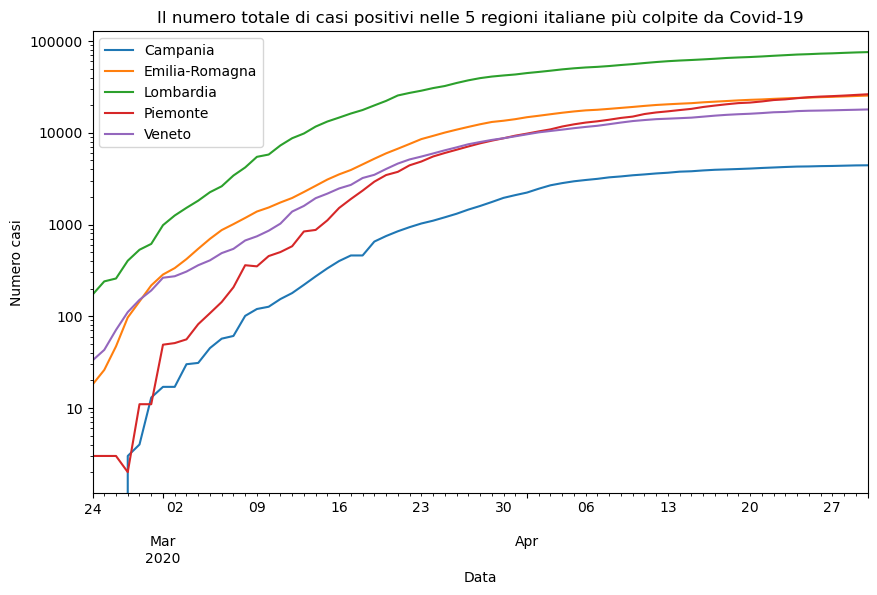
Dal grafico “Confronto tra morti e casi positivi per regione al 6 dicembre 2020” è possibile notare l’elevato divario tra i casi positivi e i decessi al giorno 6/12/2020, a suggerire una bassa mortalità del virus.



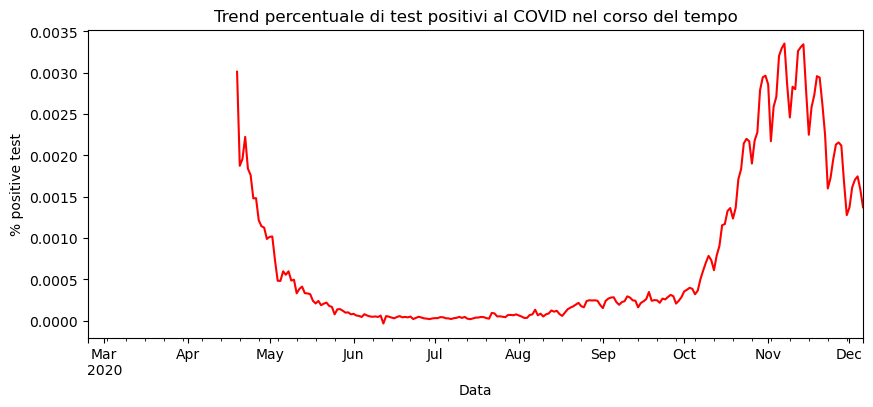
Dai precedenti grafici abbiamo notato che le regioni più colpite sono state: Lombardia, Piemonte, Emilia-Romagna, Veneto e Campania; per questo motivo, è stato attenzionato lo sviluppo della pandemia in queste regioni nei mesi più critici della pandemia, cioè quelli iniziali.

Il grafico risultante mostra l'andamento giornaliero del numero totale di casi positivi per ciascuna delle cinque regioni selezionate, evidenziando le tendenze nel corso dei primi mesi della pandemia. La scala logaritmica sull'asse y è stata utilizzata per evidenziare meglio le variazioni dei dati su un ampio intervallo.

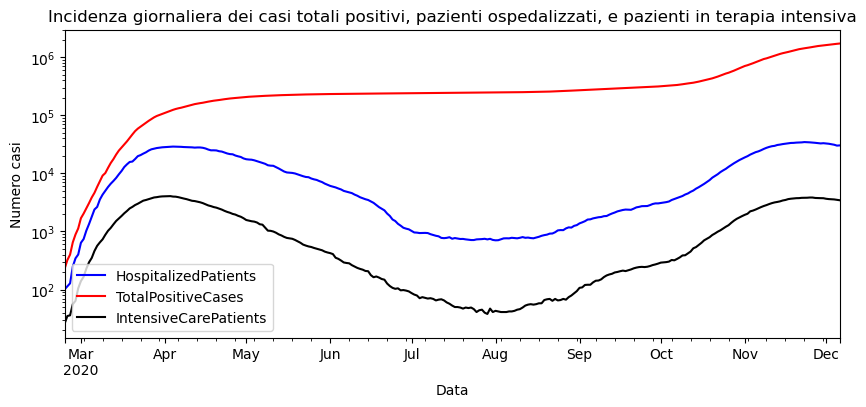
È possibile notare come in tutte le regioni ci sia stato un aumento repentino dei casi positivi tra Febbraio e Marzo, mentre a partire da Aprile l’andamento dei casi positivi si è progressivamente stabilizzato.



La figura sottostante permette di comprendere come la percentuale di test positivi al COVID-19 sia variata nel corso del tempo, fornendo un'indicazione del tasso di diffusione del virus in base ai test eseguiti. La scala temporale è suddivisa in giorni e la media giornaliera della percentuale di test positivi sul totale dei test eseguiti è rappresentata mediante una linea rossa. Malgrado i dati disponibili sui tamponi effettuati partano da metà Aprile è possibile distinguere chiaramente la prima e la seconda ondata della pandemia.

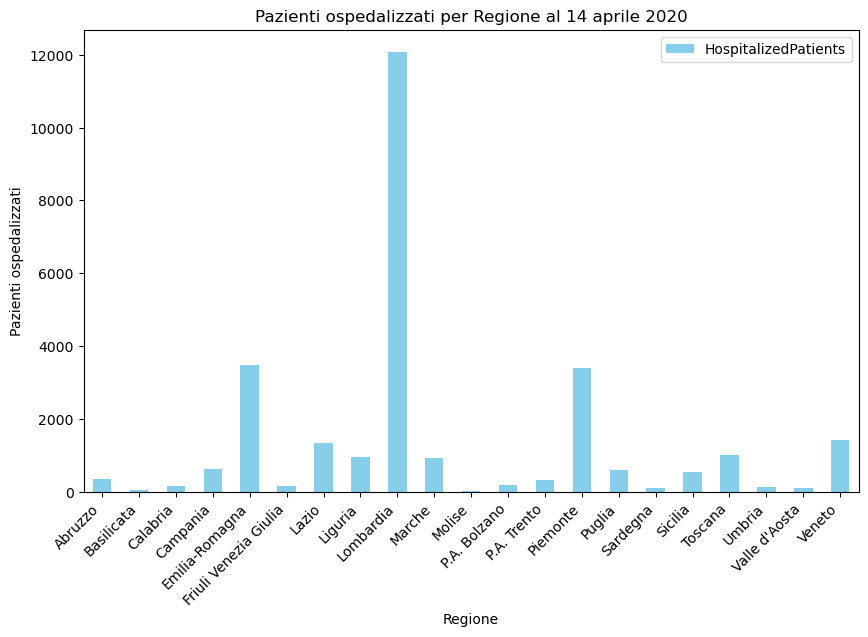


Il seguente grafico a linee mostra l'andamento giornaliero dei casi totali positivi, dei pazienti ospedalizzati e dei pazienti in terapia intensiva da fine Febbraio a inizio Dicembre. La scala logaritmica sull'asse y è stata utilizzata per evidenziare meglio le variazioni dei dati su un ampio intervallo. Osservando la figura è possibile distinguere chiaramente le due ondate del COVID-19 che nel 2020 hanno investito l’Italia. Nel dettaglio, possiamo osservare la velocità con cui il virus si è diffuso tra Febbraio e Marzo, e la diminuzione dei contagi nei mesi estivi. Osservando, invece, il secondo semestre dell’anno è possibile notare un aumento dei casi positivi superiore al picco registrato ad Aprile, ma un andamento piuttosto in linea con quello del semestre precedente relativamente ai pazienti ospedalizzati e a quelli in terapia intensiva.



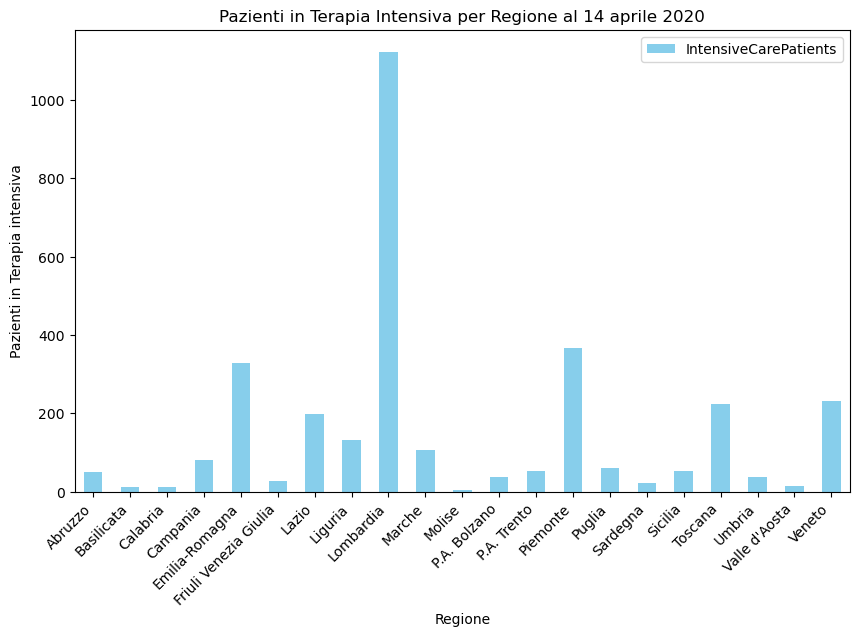
Per realizzare il grafico sottostante è stato preso in considerazione il giorno in cui in Lombardia (regione più colpita dal COVID-19) sono stati registrati il maggior numero di pazienti ospedalizzati, cioè il 14/04/2020. È interessante osservare il quantitativo di pazienti ricoverati nel resto delle regioni italiane, in relazione al numero più elevato relativo alla Lombardia (circa 12.000 persone).

Dall’osservazione del grafico notiamo che le altre regioni più colpite risultano essere Emilia-Romagna e Piemonte. Questi numeri ovviamente non sono riconducibili al solo fatto che la pandemia si sia sviluppata a partire dal nord, ma anche dal fatto che queste tre regioni sono tra le più popolate in Italia.



Oltre a considerare i pazienti ospedalizzati per malattie respiratorie lievi, è importante ricordare l’elevato numero di persone ricoverate nelle terapie intensive di tutta Italia attraverso specifici grafici relativi a questi dati.

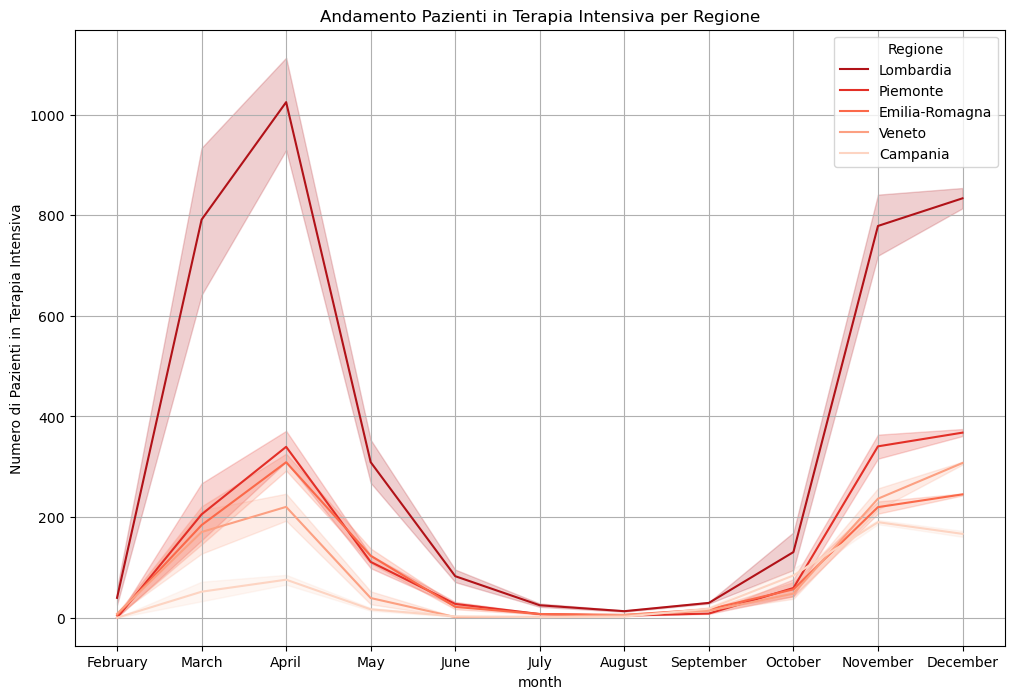
Il grafico a barre sottostante evidenzia il numero di pazienti in terapia intensiva registrati per ciascuna regione italiana al giorno 14 aprile 2020. Anche da questa figura, come per le precedenti, si può notare che le regioni maggiormente colpite sono state quelle dell’Italia settentrionale. La colonna più alta corrisponde alla regione Lombardia (1.122 persone) seguita, in ordine decrescente, dal Piemonte (367 persone), dall’Emilia-Romagna (328 persone), dal Veneto e dalla Toscana (entrambe circa 250 persone), evidenziando la gravità dell'epidemia in queste aree.



Relativamente al grafico successivo, è stata condotta un'ulteriore analisi per valutare l'andamento nel tempo della variazione del numero di pazienti ricoverati in terapia intensiva per le prime 5 regioni italiane maggiormente colpite dalla pandemia di COVID-19: Lombardia, Piemonte, Emilia-Romagna, Veneto e Campania.

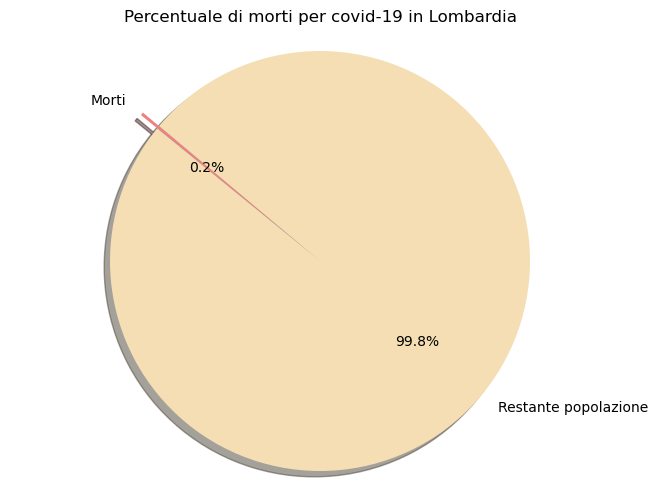
Dal grafico emerge che nel periodo preso in esame (febbraio-dicembre 2020), sono state osservate due ondate significative:

* la prima, che va da Febbraio e raggiunge il picco massimo ad Aprile, per poi subire una rapida decrescita del numero di persone ricoverate in terapia intensiva nei mesi estivi;
* la seconda, iniziata a partire da settembre, con una crescita rapida nei mesi successivi fino a Novembre, seguita da un rallentamento che ha portato ad una stabilizzazione della situazione pandemica.

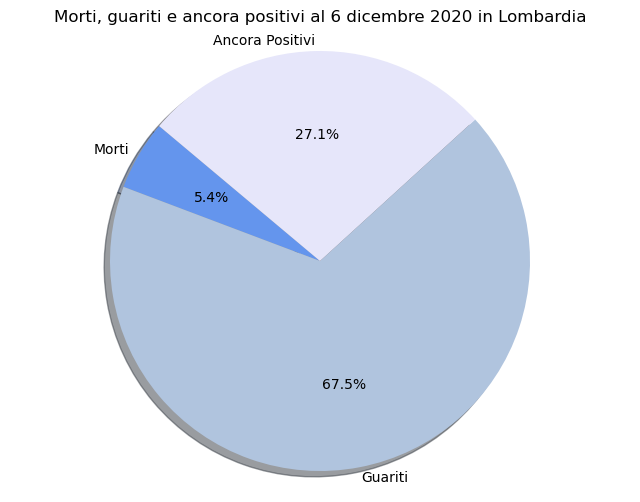


***Il caso Lombardia***

Il seguente grafico visualizza il confronto tra il numero totale di casi positivi e decessi per la regione Lombardia alla fine del periodo considerato. Nel dettaglio, i decessi al 6/12/2020 si attestavano intorno ai 23.024 individui, su un totale di 9.704.151 di abitanti.

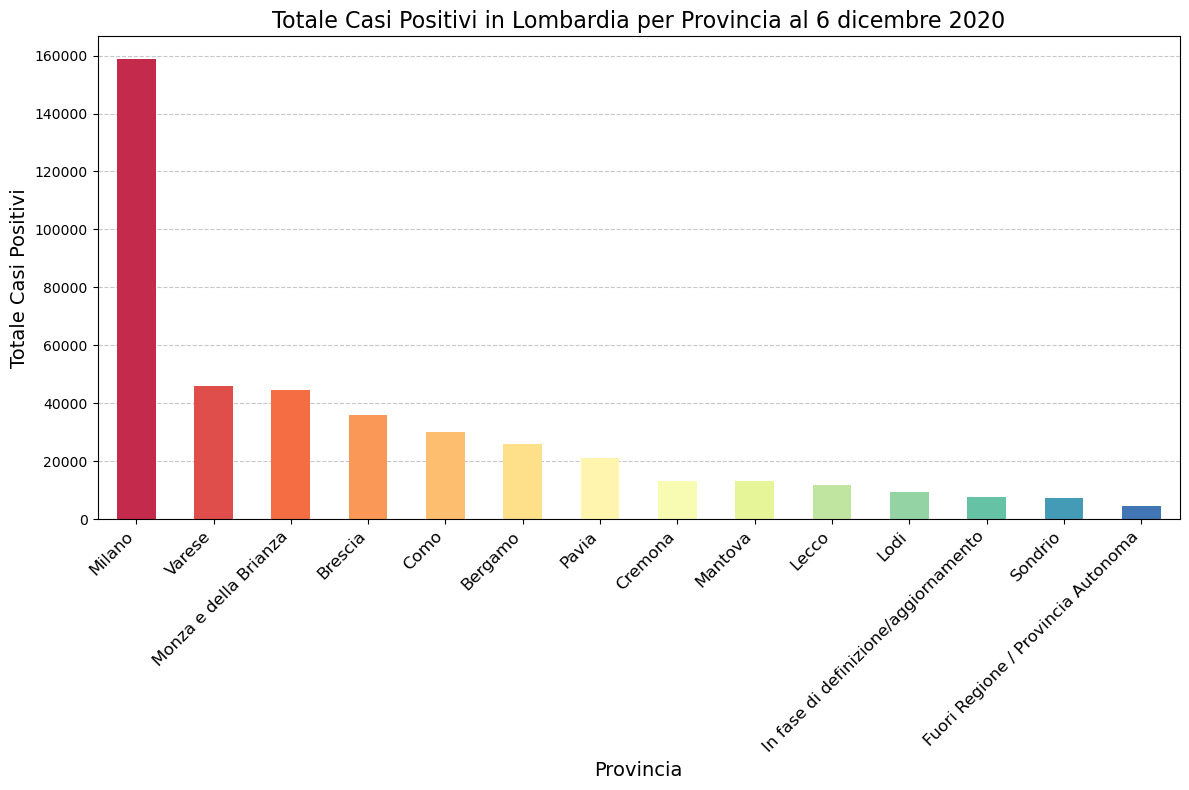


Grazie al grafico sotto riportato è possibile osservare chiaramente la distribuzione percentuale dei morti, dei guariti e dei positivi nella regione Lombardia al 6 dicembre 2020. Il grafico è stato calcolato su un totale di 429.109 casi positivi e mostra un’elevata percentuale di guariti corrispondenti a 289.706 persone.



Il seguente grafico mostra i dati raccolti per il numero totale di persone risultate positive fino al giorno 6 dicembre 2020 nelle province della Lombardia. Alcuni dati, ancora in fase di definizione e aggiornamento, sono stati opportunamente raccolti sotto la voce “in fase di definizione/aggiornamento”, mentre altre informazioni riguardano i dati provenienti da “fuori regione” oppure dalla provincia autonoma di Trento.

Il grafico evidenzia con il colore rosso la colonna che conta il numero più alto di casi positivi registrati nella provincia di Milano, seguita da un’ampia decrescita nel numero dei contagi registrati nelle province di Varese e di Monza e della Brianza. Le restanti province della Lombardia seguono in successione, con colori diversi, indicando un minor numero di contagi.



***Analisi della correlazione tra l'infezione da COVID e diversi fattori***

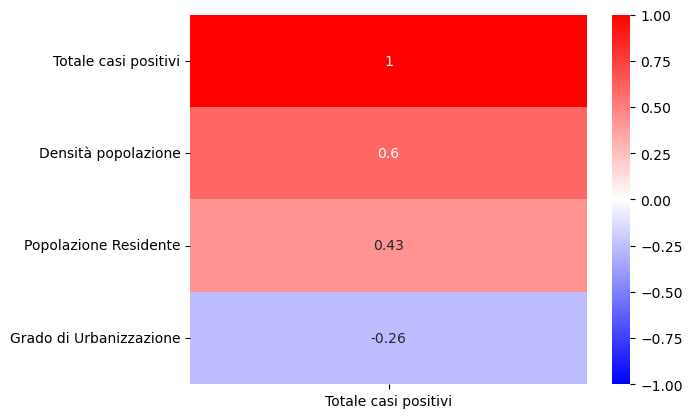
Dopo aver eseguito un'analisi sui dati sull'andamento del COVID-19 in Italia, si è deciso di cercare la correlazione tra i fattori che possono provocare un'alta predisposizione all'infezione e un'alta mortalità, ad esempio, la densità di popolazione e l'inquinamento dell'aria.

***COVID-19 e indicatori demografici***

Supponendo che una maggiore densità di popolazione nei comuni possa favorire la diffusione più rapida del coronavirus e portare a un'incidenza maggiore di COVID-19 in tali aree, è stata calcolata la correlazione tra la densità di popolazione e l'incidenza delle infezioni al picco della pandemia di COVID-19 (rappresentato per semplicità dal giorno 14 aprile 2020).

Attraverso il sito dell'Istituto Nazionale di Statistica (ISTAT), è stato importato il dataset relativo alla superficie territoriale e alla popolazione residente di ogni comune. Utilizzando tali dati, è stata calcolata la densità di popolazione e si è cercato di stabilire una correlazione con l'incidenza di COVID-19 durante il periodo critico, utilizzando anche altri due parametri disponibili: la popolazione residenziale e il grado di urbanizzazione.

Tuttavia, non sono state riscontrate correlazioni significative tra questi fattori e l'incidenza del COVID-19. La massima correlazione osservata è stata tra l'infezione da COVID-19 e la densità di popolazione, che si attesta a 0.6.



***COVID-19 e inquinamento atmosferico***

È stato ipotizzato che l'inquinamento atmosferico potrebbe correlarsi con malattie respiratorie e peggiori esiti del COVID-19. Pertanto, è stata analizzata l'incidenza del COVID-19 in relazione all'inquinamento atmosferico. Il dataset relativo alle emissioni annuali dei diversi inquinanti per la regione Lombardia nell'anno 2019 è stato importato dall’INEMAR (Inventario emissioni aria - regione Lombardia).

(Fonte: https://inemar.arpalombardia.it/inemar/webdata/elab\_standard\_reg.seam?cid=4995 file 2019\_RL2\_risultati\_prov\_inq.xls.).

In sostanza, la heat map evidenzia graficamente se e come le emissioni di diversi inquinanti dell’aria possano essere correlati al numero di casi positivi di COVID-19 in Lombardia, aiutando a individuare eventuali relazioni tra questi fattori. Dall’osservazione è emersa una correlazione positiva tra il numero totale di casi positivi di COVID e le emissioni di ossidi di azoto (NOx, correlazione 0.86), i composti organici volatili (COV, correlazione 0.88) e i precursori dell'ozono (precurs. O3, correlazione 0.9). Tuttavia, per confermare tali correlazioni, sarebbero necessari ulteriori studi a livello nazionale o anche globale.

