

# Esercitazione finale

## Analisi evolutiva sull'impatto della pandemia COVID-19 in Italia tra il 2020 e il 2022: numeri, trend e scenari

### Scenario di progetto:

Una società del settore salute (Salus S.p.A.) vuole comprendere l'effettiva evoluzione della pandemia negli anni 2020-2022 per ragionare su nuovi investimenti in servizi di monitoraggio e prevenzione sanitaria sul territorio. Il compito sarà realizzare un'analisi completa utilizzando dati pubblici ufficiali, tecniche di analisi dati, strumenti collaborativi e supporto da AI generativa per documentare ogni fase del lavoro e sintetizzare gli insight principali in una presentazione chiara.

### Obiettivi:

1. Analizzare l'evoluzione dei dati COVID-19 tra 2020-2022 in Italia.
2. Studiare trend di contagi, decessi, guariti, vaccinazioni per periodi e territori.
3. Creare indicatori personalizzati, es. rapporto vaccinati/contagiati.
4. Usare strumenti AI per integrare parte dell'analisi e velocizzare interpretazioni.
5. Lavorare in modo ordinato e documentato con strumenti collaborativi.
6. Realizzare una presentazione finale chiara e efficace.

### Dataset di riferimento:

- Dati COVID Protezione Civile – <https://github.com/pcm-dpc/COVID-19> (fino a dicembre 2022)
- Open data ISTAT per popolazione regionale
- Dati vaccinazioni Ministero Salute (fino a 2022)

### Tecnologie utilizzate:

- Notion per tracciare il progetto
- Google Sheets per pulizia e analisi preliminare
- Python per analisi avanzata

- ChatGPT o Gemini AI come supporto tecnico e interpretativo
- Google Drive come archivio unico
- Google Slides per presentazione finale

### **Individuazione e analisi dei dataset necessari per il progetto:**

- Dati COVID 19 Protezione Civile
- dpc-covid19-ita-regioni.csv
- popolazione-istat-regione-range.csv
- Vaccinazioni2020.csv
- Vaccinazioni2021.csv
- Vaccinazioni2022.csv

### **Pulizia Dataset Covid 19:**

1. Eliminazione colonne non necessarie
2. Riduzione dataset con eliminazione di record precedenti marzo 2020
3. Riduzione dataset con eliminazione di record successivi al 1 gennaio 2023
4. Inserimento del csv sulla Popolazione in un nuovo foglio
5. Eliminazione colonne non necessarie sulla Popolazione
6. Raggruppare i dati per regione
7. Inserimento dei csv vaccinazioni in un nuovo foglio
8. Eliminazione colonne non necessarie sulle Vaccinazioni
9. Raggruppare i dati per regione
10. Analisi complete unendo i tre dataset

### **Formule utilizzate in Google Sheets:**

1. =QUERY(Dati\_Originali!A:AD;"select A,C,D,G,H,I,J,K,N,O,R,S,T,W,X")
2. =QUERY(Pulizia\_Colonne!A:Z; "SELECT \* WHERE A>= '2020-03-01' AND A <= '2022-12-31'")

3. =QUERY(Dati\_Originali\_Popolazione!A:L; "select A,E,J,K,L")
4. =ARRAYFORMULA(SE(SORT(UNIQUE(Pulizia\_Colonne\_Popolazione!B2:B))="Bolzano";"P.A. Bolzano";SE(SORT(UNIQUE(Pulizia\_Colonne\_Popolazione!B2:B))="Trento";"P.A. Trento";SORT(UNIQUE(Pulizia\_Colonne\_Popolazione!B2:B)))))
5. =QUERY(Pulizia\_Colonne\_Popolazione!A:E; "select B, sum(C), sum(D), sum(E) group by B")
6. =QUERY(Filtro\_Record!A:O; "select C, sum(D), sum(E), sum(G), sum(H), sum(I), sum(L), sum(M), sum(F) group by C label C 'Regione'")
7. =CERCA.VERT(A3;Raggruppo\_per\_Regione!A:D;4)
8. =CERCA.VERT(A3;Raggruppo\_per\_Regione!A:D;2)
9. =CERCA.VERT(A3;Raggruppo\_per\_Regione!A:D;3)
10. =ARRAYFORMULA(SE(SORT(UNIQUE(Vaccinazioni2020!D2:D))="Provincia Autonoma Bolzano / Bozen";"P.A. Bolzano";SE(SORT(UNIQUE(Vaccinazioni2020!D2:D))="Provincia Autonoma Trento";"P.A. Trento";SE(SORT(UNIQUE(Vaccinazioni2020!D2:D))="Valle d'Aosta / Vallée d'Aoste";"Valle d'Aosta";SE(SORT(UNIQUE(Vaccinazioni2020!D2:D))="Friuli-Venezia Giulia"; "Friuli Venezia Giulia";SORT(UNIQUE(Vaccinazioni2020!D2:D)))))
11. =QUERY(Vaccinazioni2020!A:D; "select sum (B) + sum (C) group by D label sum (B) + sum (C) 'Totali Dosi Somministrate 2020'")
12. =CERCA.VERT(\$A3; Vaccinazioni\_Raggruppate!\$A:\$D;2; FALSO)
13. =QUERY(Filtro\_Record!A:L; "select C, max(J) where A >= '2020-03-01' and A < '2021-01-01' group by C label C 'Regione', max(J) 'Totale Deceduti 2020' ")
14. =QUERY(Filtro\_Record!A:L; "select C, max(J) where A >= '2020-03-01' and A < '2021-01-01' group by C label C 'Regione', max(J) 'Totale Deceduti 2020' ")
15. =QUERY(Filtro\_Record!A:L; "select max(J) where A >= '2022-01-01' and A < '2023-01-01' group by C label max(J) 'Totale Deceduti 2022' ")

### Indicatori personalizzati:

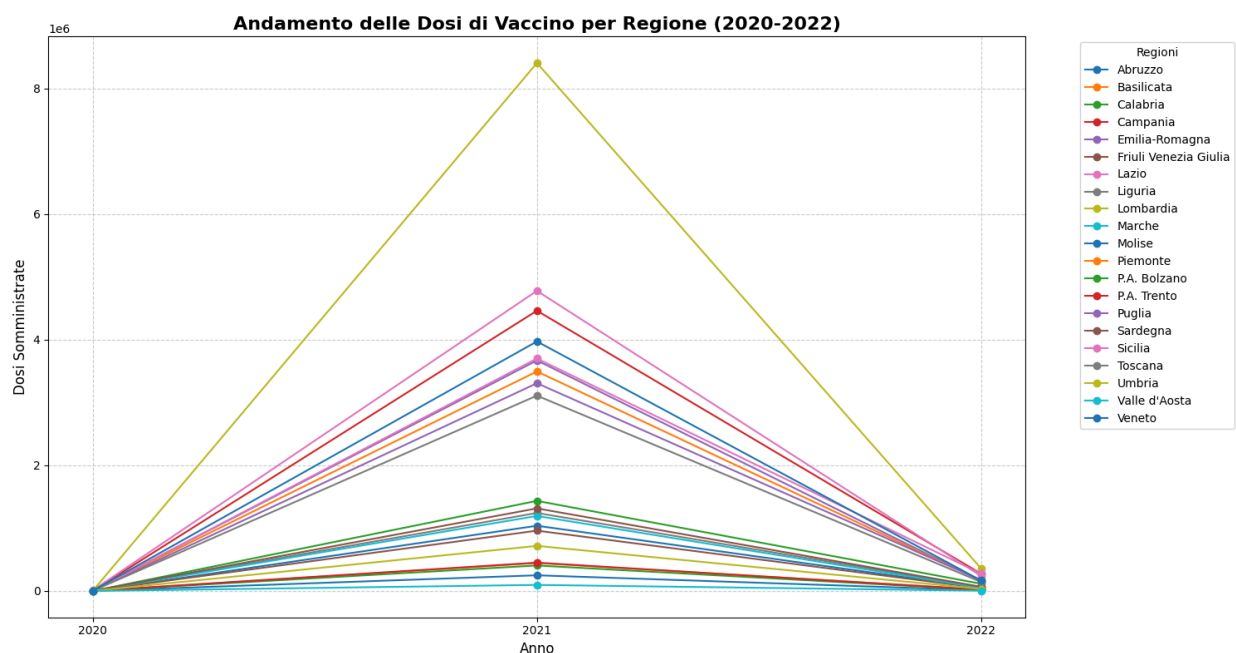
- Totali vaccinazioni per anno
- Numero di deceduti per anno
- Totale dimessi/guariti per anno
- Totale positivi in gruppi da 7 giorni
- Totale vaccinazioni in gruppi da 7 giorni

## Analisi indicatori personalizzati e grafici:

### Totali vaccinazioni per anno:

Andamento annuale delle vaccinazioni: questo indicatore fornisce una panoramica completa del progresso della campagna vaccinale nel tempo, aggregando il numero totale di dosi somministrate per ogni anno solare. Il grafico riportato di seguito mostra la progressione e la velocità di somministrazione delle vaccinazioni per anno.

- Grafico:



- Formule python creazione grafico:

```
file_path = 'IndicatoreVaccinazioni.csv'
df = pd.read_csv(file_path)

df = df.set_index('Regione')
df_transposed = df.transpose()

fig, ax = plt.subplots(figsize=(15, 8))
df_transposed.plot(kind='line', ax=ax, marker='o')
```

```

ax.set_title('Andamento delle Dosi di Vaccino per Regione (2020-2022)', fontsize=16, font
weight='bold')
ax.set_xlabel('Anno', fontsize=12)
ax.set_ylabel('Dosi Somministrate', fontsize=12)
ax.set_xticks(range(len(df_transposed.index)))
ax.set_xticklabels(df_transposed.index, rotation=0)

ax.legend(title='Regioni', bbox_to_anchor=(1.05, 1), loc='upper left')

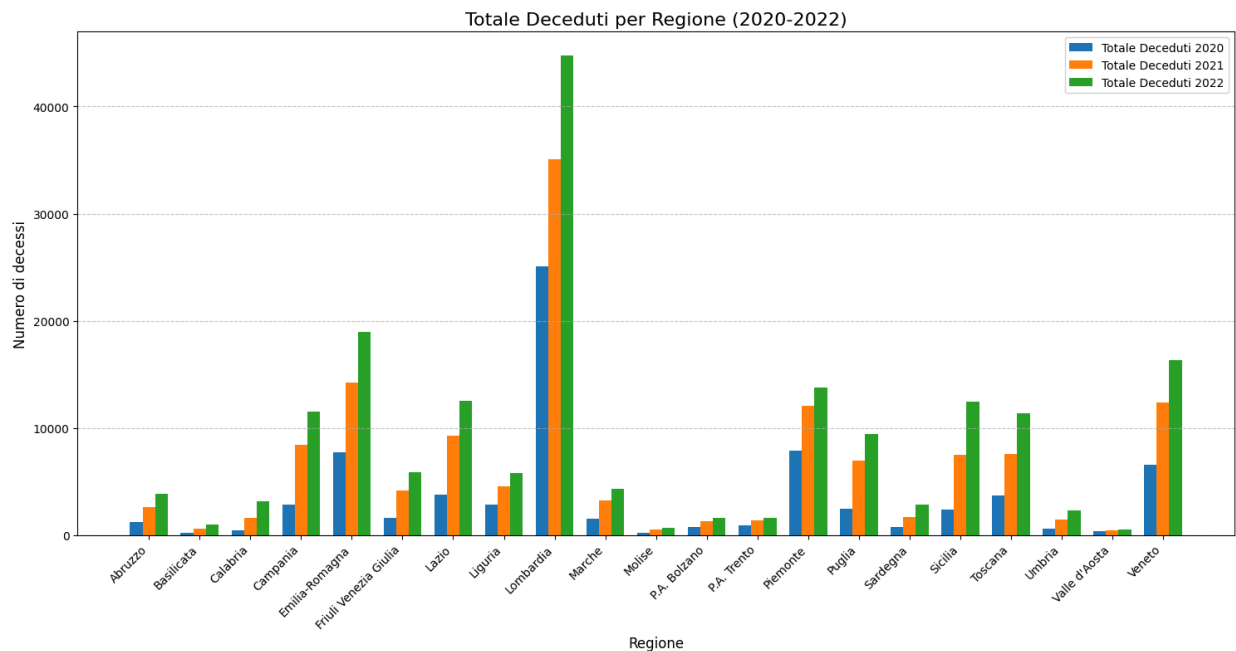
```

## Numero di deceduti per anno:

Dati aggregati dei decessi legati alla pandemia divisi per anno.

Questo indicatore analizza il numero di decessi per anno prendendo il valore massimo cumulativo disponibile per ciascun periodo. Questo metodo garantisce che si stia considerando il totale finale, permettendo di identificare i periodi di maggiore mortalità e di valutare l'impatto complessivo della pandemia. Il grafico di seguito riporta il bilancio delle vittime divise per anno. Si può dimostrare come la Lombardia sia la regione con più decessi in Italia.

- Grafico:



- Formule python creazione grafico:

```

df_decessi = pd.read_csv('IndicatoreDeceduti.csv')

regioni = df_decessi['Regione']
decessi_2020 = df_decessi['Totale Deceduti 2020']
decessi_2021 = df_decessi['Totale Deceduti 2021']
decessi_2022 = df_decessi['Totale Deceduti 2022']

x = np.arange(len(regioni))
width = 0.25

plt.figure(figsize=(15, 8))
bar1 = plt.bar(x - width, decessi_2020, width, label='Totale Deceduti 2020')
bar2 = plt.bar(x, decessi_2021, width, label='Totale Deceduti 2021')
bar3 = plt.bar(x + width, decessi_2022, width, label='Totale Deceduti 2022')

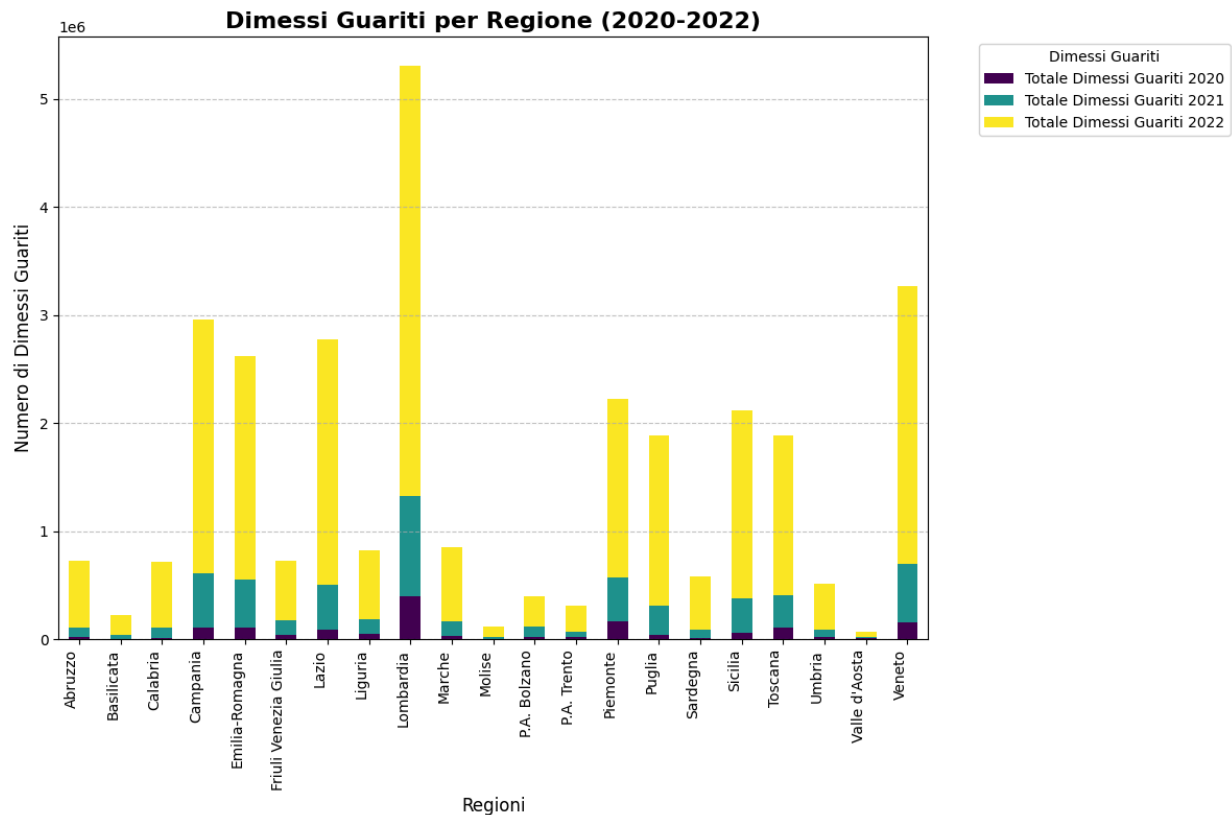
plt.title('Totale Deceduti per Regione (2020-2022)', fontsize=16)
plt.ylabel('Numero di decessi', fontsize=12)
plt.xlabel('Regione', fontsize=12)
plt.xticks(x, regioni, rotation=45, ha='right') # Aggiungi le etichette delle regioni
plt.legend()
plt.grid(axis='y', linestyle='--', alpha=0.7) # Aggiungi una griglia orizzontale
plt.tight_layout()

```

## Totale dimessi/guariti per anno:

Sintesi del numero di persone guarite e dimesse dagli ospedali, aggregata per anno. Questo indicatore traccia il numero totale di persone che si sono riprese dall'infezione ogni anno. Il calcolo si basa sul valore massimo cumulativo di dimessi/guariti, fornendo un dato finale e completo per ogni periodo. E' un dato chiave per comprendere la capacità di ripresa del sistema sanitario e l'evoluzione della malattia. Si può notare come nel grafico sottostante, la Lombardia sia la regione che presenta la barra più alta, indicando il numero più elevato di persone dimesse/guarite, specialmente si nota che il contributo maggiore al totale cumulativo proviene dall'anno 2022, questo evidenzia un aumento significativo e diffuso dei casi e delle successive guarigioni. Il grafico in sintesi mostra non solo l'impatto complessivo del virus in termini di guarigioni, ma anche come questo impatto si sia evoluto e distribuito in modo non uniforme tra le varie regioni italiane nel corso degli anni.

- Grafico:



- Formule python creazione grafico:

```
df = df.set_index('Regione')
fig, ax = plt.subplots(figsize=(12, 8))
df.plot(kind='bar', stacked=True, ax=ax, colormap='viridis')

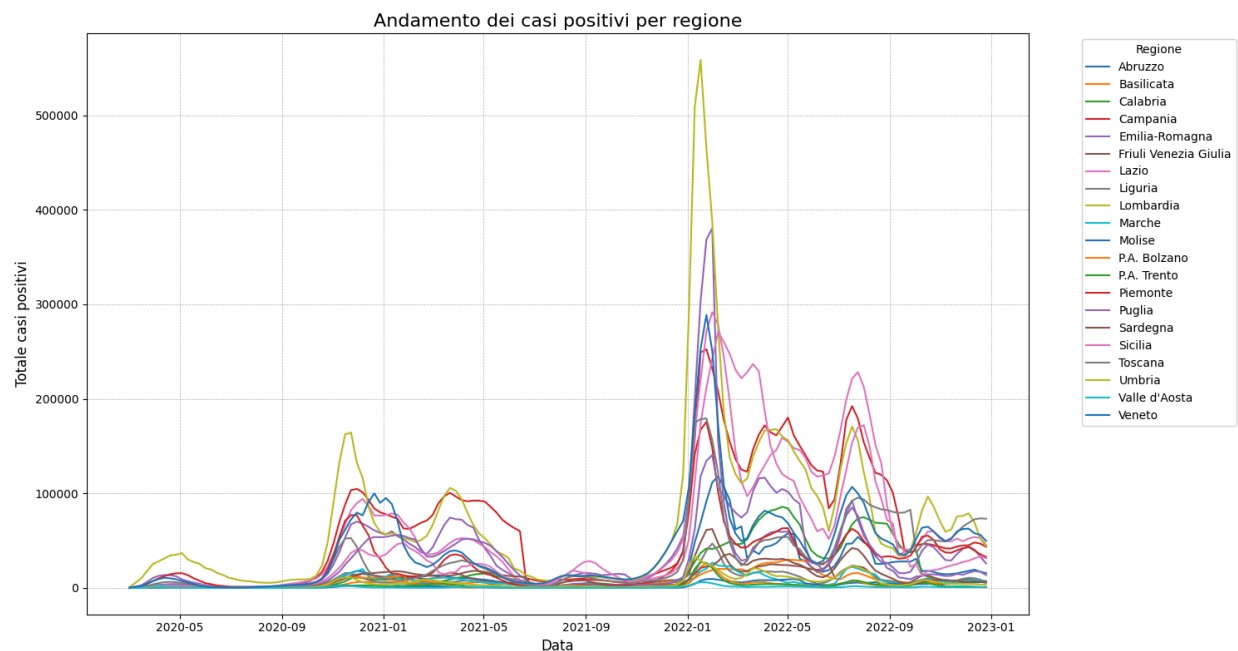
ax.set_title('Dimessi Guariti per Regione (2020-2022)', fontsize=16, fontweight='bold')
ax.set_xlabel('Regioni', fontsize=12)
ax.set_ylabel('Numero di Dimessi Guariti', fontsize=12)
ax.set_xticklabels(ax.get_xticklabels(), rotation=90, ha='right', fontsize=10)
ax.legend(title='Dimessi Guariti', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
```

## Totale positivi in gruppi da 7 giorni:

Analisi dell'evoluzione del contagio su un periodo di 7 giorni. Questo indicatore calcola la somma dei casi positivi su un intervallo di 7 giorni. Questa metodologia "appiattisce" le fluttuazioni giornaliere dovute, ad esempio, a ritardi nei test, fornendo una visione più chiara e

stabile della tendenza del contagio nel tempo. Nel grafico sottostante le linee mostrano chiaramente le diverse "onde" di contagio che si sono succedute nel tempo. Ogni linea regionale segue un percorso simile, con picchi e cali che corrispondono a periodi di alta e bassa incidenza. Il picco più elevato, visibile nella maggior parte delle linee verso la fine del periodo di analisi, suggerisce l'impatto di varianti virali più contagiose, che hanno portato a un aumento massiccio dei casi, ma con un impatto sulla gravità della malattia spesso diverso rispetto ai picchi iniziali.

- Grafico:



- Formule python creazione grafico:

```
df = df.sort_values(by=['data', 'denominazione_regione'])
plt.figure(figsize=(15, 8))

for regione in df['denominazione_regione'].unique():
    dati_regione = df[df['denominazione_regione'] == regione]
    plt.plot(dati_regione['data'], dati_regione['totale_positivi'], label=regione)

plt.title('Andamento dei casi positivi per regione', fontsize=16)
plt.xlabel('Data', fontsize=12)
plt.ylabel('Totale casi positivi', fontsize=12)
plt.legend(title='Regione', bbox_to_anchor=(1.05, 1), loc='upper left')
```

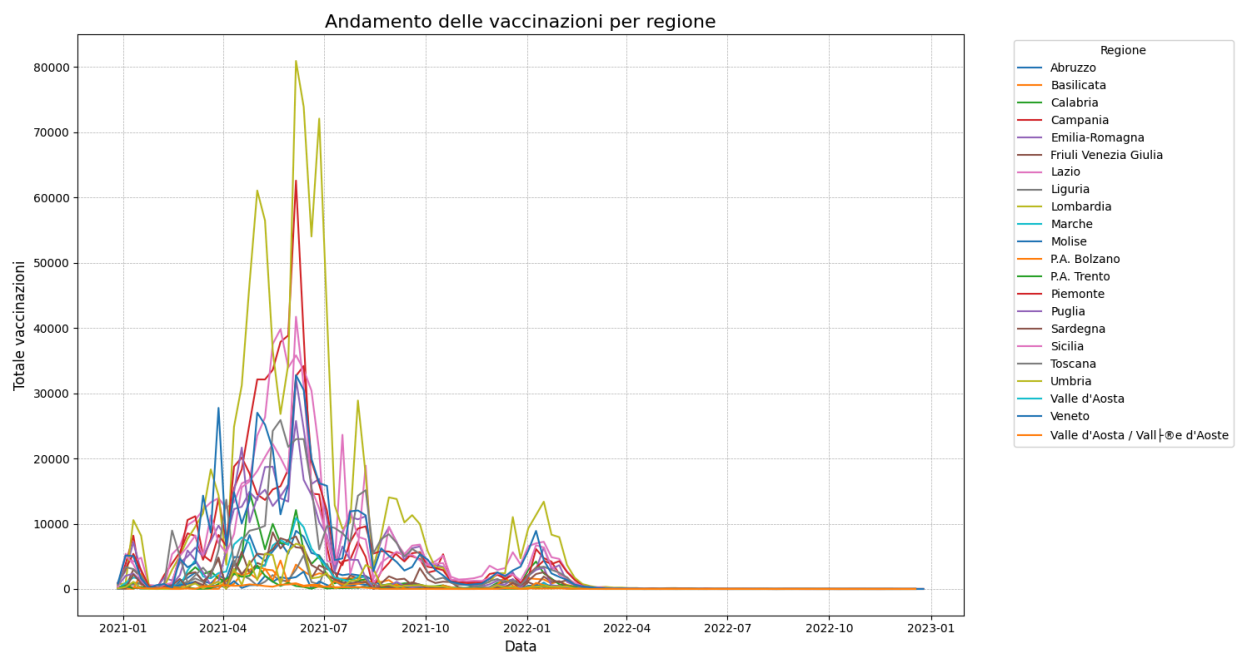


```
plt.grid(True, which='both', linestyle='--', linewidth=0.5)
plt.tight_layout()
```

## Totale vaccinazioni in gruppi da 7 giorni:

Analisi del numero di vaccinazioni somministrate ogni 7 giorni. Simile all'indicatore precedente, questo calcolo mostra il volume delle vaccinazioni su base settimanale. Permette di identificare il ritmo di somministrazione, di riconoscere i periodi di picco o di rallentamento e di valutare l'efficienza logistica della campagna. Dal grafico si evince che tutte le regioni mostrano un'iniziale fase di avvio lenta, con valori bassi nelle prime settimane della campagna vaccinale. Questo è un dato atteso, legato alla fase logistica iniziale e alla disponibilità limitata dei vaccini. La curva di tutte le regioni sale poi rapidamente, indicando un'accelerazione significativa nel ritmo di somministrazione. Verso la fine del periodo di analisi invece, le linee tendono a stabilizzarsi o a mostrare una crescita più moderata, suggerendo che la campagna è entrata in una fase di mantenimento dopo aver raggiunto una larga fetta della popolazione.

- Grafico:



- Formule python creazione grafico:

```
df = df.sort_values(by=['data', 'Regione'])
plt.figure(figsize=(15, 8))
```

```
for regione in df['Regione'].unique():
```

```
dati_regione = df[df['Regione'] == regione]
plt.plot(dati_regione['data'], dati_regione['Totale Vaccinazioni'], label=regione)

plt.title('Andamento delle vaccinazioni per regione', fontsize=16)
plt.xlabel('Data', fontsize=12)
plt.ylabel('Totale vaccinazioni', fontsize=12)
plt.legend(title='Regione', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.grid(True, which='both', linestyle='--', linewidth=0.5)
plt.tight_layout()
```

## Conclusioni: Analisi integrata della pandemia Covid-19 in Italia:

Questa dashboard offre una visione d'insieme dell'evoluzione della pandemia in Italia, basata su un'analisi multidimensionale che va oltre i semplici conteggi giornalieri. L'uso di indicatori personalizzati ha permesso di trasformare i dati grezzi in informazioni significative, offrendo una prospettiva più profonda sulle dinamiche del contagio, sulla risposta sanitaria e sull'impatto a lungo termine.

## Principali considerazioni emerse:

### Impatto e gestione regionale:

- L'analisi dei **deceduti e dimessi/guariti per anno** ha evidenziato una netta disparità tra le regioni. La Lombardia, pur con il maggior numero di decessi, mostra anche il numero più alto di guarigioni, riflettendo l'impatto iniziale devastante ma anche una robusta capacità di gestione sanitaria.
- I grafici a barre, sia raggruppati che in pila, hanno visualizzato chiaramente come la pandemia abbia colpito in modo diverso le regioni, con il Nord Italia che ha sostenuto il carico più pesante all'inizio, mentre le differenze regionali si sono mantenute costanti nel tempo.

### Dinamica del contagio e del contenimento:

- L'indicatore **positivi in gruppi da 7 giorni** si è dimostrato uno strumento essenziale per comprendere le "onde" epidemiche. La sua capacità di appianare le fluttuazioni giornaliere ha permesso di individuare con chiarezza i picchi di contagio e di correlarli a eventi specifici, come l'introduzione di nuove varianti o l'allentamento delle restrizioni. Questo dato ha permesso di valutare l'efficacia delle misure di contenimento nel tempo.

## **L'impatto della campagna vaccinale:**

- I grafici che mostrano l'andamento delle **vaccinazioni per anno e in gruppi da 7 giorni** narrano una storia di successo logistico e sanitario. Dopo un avvio graduale, il ritmo di somministrazione ha accelerato in modo significativo in tutte le regioni, portando a una copertura vaccinale capillare. La correlazione visiva tra l'aumento delle vaccinazioni e la stabilizzazione dei casi (o il minor impatto dei picchi successivi) suggerisce l'efficacia della campagna nel mitigare la gravità della malattia.

## **Conclusioni finali:**

In sintesi, l'analisi dei dati ha permesso di passare da una semplice lettura dei numeri a una comprensione olistica della pandemia. Il report dimostra che, sebbene il virus abbia colpito ogni regione in modo unico, la risposta collettiva, trainata da una rapida e massiccia campagna di vaccinazione, ha giocato un ruolo cruciale nel trasformare una crisi in una fase di gestione più controllabile. La dashboard su Notion funge da testimonianza visiva di questo percorso, offrendo uno strumento prezioso per la consultazione e l'analisi.