

MODULO 1 – ESERCITAZIONE 3, “OPEN DATA”

di Gabriele Pellegrino

1. Dataset “Report Nazionale Casi Covid-19 da inizio pandemia” - Ministero della Salute

URL

<https://github.com/pcm-dpc/COVID-19/blob/cae4442f4dd4e157fe18677a372d7be1c915e083/dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale.csv>

DESCRIZIONE

Il dataset indica la situazione Covid-19 da inizio pandemia aggiornata al 2/2/2023, “con chiave primaria” sotto il campo “Data”.

I campi utilizzati per analizzare i dati sono:

- la data
- lo stato
- n° pazienti ospedalizzati totali suddivisi poi in ricoverati con sintomi, ricoverati in terapia intensiva e in isolamento domiciliare;
- n° totale positivi con relativa variazione e nuovi positivi
- n° dimessi e guariti
- n° deceduti
- n° totale casi
- n° tamponi e casi testati e n° totale positivi al test molecolare e antigenico
- n° tamponi molecolari e antigenici positivi

CLASSIFICAZIONE (★★★) - Dataset in .csv accessibile ad una specifica repository e URL, codificato e aperto.

METADATI: Non ci sono metadati specifici. Unica informazione è quella della fonte, cioè il Ministero della Salute

2. Dataset “Report Regionale Casi Covid-19 da inizio pandemia” - Ministero della Salute

URL

<https://github.com/pcm-dpc/COVID-19/blob/cae4442f4dd4e157fe18677a372d7be1c915e083/dati-regioni/dpc-covid19-ita-regioni.csv>

DESCRIZIONE

Il dataset indica la situazione Covid-19 su scala regionale da inizio pandemia aggiornata al 2/2/2023, “con chiave primaria” sotto il campo “Regione” o “Codice regione”. I campi utilizzati per analizzare i dati sono:

- la data
- lo stato, il codice e la denominazione regione, latitudine e longitudine
- n° pazienti ospedalizzati totali suddivisi poi in ricoverati con sintomi, ricoverati in terapia intensiva e in isolamento domiciliare;
- n° totale positivi con relativa variazione e nuovi positivi
- n° dimessi e guariti
- n° deceduti
- n° totale casi
- n° tamponi e casi testati e n° totale positivi al test molecolare e antigenico
- n° tamponi molecolari e antigenici positivi

CLASSIFICAZIONE (★★★) - Dataset in .csv accessibile ad una specifica repository e URL, codificato e aperto.

METADATI: Non ci sono metadati specifici. Unica informazione è quella della fonte, cioè il Ministero della Salute;

3. Dataset “Report Provinciale Casi Covid-19 da inizio pandemia - Ministero della Salute”

URL

<https://github.com/pcm-dpc/COVID-19/blob/cae4442f4dd4e157fe18677a372d7be1c915e083/dati-province/dpc-covid19-ita-province.csv>

DESCRIZIONE

Il dataset indica la situazione Covid-19 su scala provinciale da inizio pandemia aggiornata al 2/2/2023, “con chiave primaria” sotto il campo “Denominazione Regione”, “Codice regione” o “Codice provincia”. I campi utilizzati per analizzare i dati sono:

- la data
- lo stato, il codice e la denominazione regione, la denominazione, la sigla e il codice provincia, latitudine e longitudine
- n° pazienti ospedalizzati totali suddivisi poi in ricoverati con sintomi, ricoverati in terapia intensiva e in isolamento domiciliare;
- n° totale casi

CLASSIFICAZIONE (★★★) - Dataset in .csv accessibile ad una specifica repository e URL, codificato e aperto.

METADATI: Non ci sono metadati specifici. Unica informazione è quella della fonte, cioè il Ministero della Salute

CONFRONTO SITUAZIONE COVID-19 STATO-CLASSE

PRESENTAZIONE

L’analisi è stata effettuata prendendo in considerazione principalmente 3 dataset:

- Dataset “Covid19-Classe Data Analyst-0123 Epicode”
(<https://github.com/gabropel/EPICODE-DA0123/blob/main/ESERCITAZIONE%20%20-%20COVID%20CLASSE%20EPICODE%20MOD.xlsx>)
- Dataset “Report Nazionale Casi Covid-19 da inizio pandemia” - Ministero della Salute (agg. al 2/2/2023)”
- Dataset “Report Regionale Casi Covid-19 da inizio pandemia - Ministero della Salute (agg. al 2/2/2023)”

L’analisi verte in modo specifico nel rapportare i dati “Classe DA-0123”, composta da 30 persone provenienti da 10 regioni italiane diverse, rispettivamente 4 del Nord, 2 del Centro e 4 del Sud e Isole, con i dati riguardanti le “Regioni Classe” su scala regionale e nazionale utilizzati dai dataset sopra citati, suddivisi in 3 macroaree (“Nord”, “Centro” e “Sud e Isole”), mettendo a confronto i dati riguardanti “Positivi”, “Ospedalizzazioni” e “Decessi” nelle aree citate.

I campi di analisi della “Classe DA-0123” sono: “Nome”, “Cognome”, “Città”, “Positività”, “Eventuale ricovero” e “Giorni di degenza”, con il campo “ID” in “chiave primaria”.

ANALISI

Dall’analisi sui dati Classe, si evince che il 60% di essa ha incontrato l’infezione da Covid-19 almeno una volta, con il restante 40% mai contagiato (18 almeno una volta positivi, 12 negativi), con una media di 14 giorni di degenza o positività per chi ha contratto il virus. Non vi sono tuttavia stati riscontrati episodi di ospedalizzazione o ricovero.

Passando poi ad un’analisi più approfondita sulla base delle 3 macroaree sotto osservazione, si evince che:

- Nella Macroarea “Nord”, nelle 4 regioni presenti in classe (Lombardia, Veneto, Piemonte e Emilia-Romagna), è stata riscontrata una percentuale di positività in aula del 33,3% (6 persone almeno una volta positive), in linea con la percentuale totale nazionale delle “Regioni Classe - Macroarea Nord”, proprio ferma al 33,3%. Per quanto riguarda l’analisi sui dati “Ospedalizzati-Dimessi/Guariti” delle “Regioni Classe – Macroarea Nord”, il 41,7% della popolazione delle regioni presenti in classe è uscito positivamente dalle cure ospedaliere rispetto al totale degli ospedalizzati in Italia (8.359.555 su 25.072.909), e di un 50,7% di “Deceduti” nella Macroarea Nord rispetto al totale dei morti in Italia (94.959 su 187.272);

- Nella Macroarea “Centro”, nelle 2 regioni presenti in classe (Lazio e Toscana), è stata riscontrata una percentuale di positività in aula del 22,2% (4 persone almeno una volta positive), più alta rispetto alla percentuale totale nazionale delle “Regioni Classe - Macroarea Centro”, ferma al 15,5%. Per quanto riguarda l’analisi sui dati “Ospedalizzati-Dimessi/Guariti” delle “Regioni Classe – Macroarea Centro”, il 15,3% della popolazione delle regioni presenti in classe è uscito positivamente dalle cure ospedaliere rispetto al totale degli ospedalizzati in Italia (3.851.382 su 25.072.909), e di un 12,9% di “Deceduti” nella “Macroarea Centro” rispetto al totale dei morti in Italia (24.178 su 187.272);
- Nella Macroarea “Sud e Isole”, nelle 4 regioni presenti in classe (Calabria, Puglia, Sicilia e Sardegna), è stata riscontrata una percentuale di positività in aula del 44,4% (8 persone almeno una volta positive), più alta rispetto alla percentuale totale nazionale delle “Regioni Classe - Macroarea Sud e Isole”, ferma al 17,9%. Per quanto riguarda l’analisi sui dati “Ospedalizzati-Dimessi/Guariti” delle “Regioni Classe – Macroarea Sud e Isole”, il 17,9% della popolazione delle regioni presenti in classe è uscito positivamente dalle cure ospedaliere rispetto al totale degli ospedalizzati in Italia (4.506.334 su 25.072.909), e di un 15,2% di “Deceduti” nella “Macroarea Sud e Isole” rispetto al totale dei morti in Italia (28.396 su 187.272);

CONCLUSIONI

Rispetto alla Classe, il Covid-19 ha avuto un impatto di positività maggiore nella Macroarea “Sud e Isole”, in tendenza contraria rispetto alle percentuali nazionali delle “Regioni-Classe” prese in esame, 17,9% rispetto al 44,4% dell’aula. Percentuali leggermente più alte per la “Macroarea Centro” (22,2% in classe rispetto alle regioni centrali rappresentate contro il 15,5%) a fronte invece di un’identica percentuale della “Macroarea Nord” rispetto alla percentuale nazionale delle regioni presenti, cioè il 33,3%. Per quanto riguarda invece l’analisi su “Ospedalizzati-Dimessi/Guariti” e “Deceduti” delle regioni rappresentate in Classe, i dati maggiori provengono dalla “Macroarea Nord”, dove vi sono evidentemente stati registrati più casi rispetto al resto d’Italia 41,7% di ospedalizzati guariti e il 50,7% dei deceduti rispetto ai rispettivi totali di Macroarea. Nelle Macroaree “Centro” e “Sud e Isole”, le percentuali invece differiscono di poco, stando a significare che evidentemente l’impatto della pandemia è stato decisamente più negativo e mortale nelle “Regioni-Classe” prese in esame nella macroarea settentrionale, pur segnalando che il virus, nella maggior parte dei casi ospedalieri e non, ha registrato comunque numeri di guarigione esponenzialmente più alti in tutti i casi.

ANALISI DEI RUOLI

DATA ENGINEER

Sono intervenuto in questo ruolo nella fase di ricerca dei dataset “Situazione Covid” in formato .csv sul web (in repository o cloud) e nella fase di sistemazione dei dati su Excel in un formato più comprensibile ed analizzabile;

DATA ANALYST

Sono intervenuto in questo ruolo nella fase di analisi dei dati del file Excel citato sopra, analizzando tutti i campi e relativi dati combinandoli per riassumere la situazione “Covid Classe”. Successivamente ho incrociato alcuni di questi con altri dataset riguardanti la “Situazione Covid Nazionale e Regionale” per trarre alcune analisi incrociate con i dati delle regioni presenti in Classe.

DATA SCIENTIST

Sono intervenuto parzialmente in questo ruolo nella fase delle conclusioni dell’analisi “Covid Classe” rispetto ad una parte dei dati regionali sui positivi nelle regioni presenti in classe, traendo alcune conclusioni analitiche. Si sarebbe potuto completare questo ruolo (alla presenza di dati più approfonditi e un lavoro più corposo con dati maggiori) per prendere ad esempio delle possibili decisioni di carattere organizzativo e medico rispetto alla gestione Covid in alcune zone d’Italia da parte delle autorità statali (es. Ministero della Salute) delle Macroaree prese in esame nelle regioni di riferimento.

DATA JOURNALIST

Sono ipoteticamente intervenuto in questo ruolo nella fase di analisi e stesura testuale dei dati, utilizzabile ad esempio per una possibile “inchiesta” sull’incidenza del Covid nelle Macroaree delle regioni di riferimento, per trarre delle possibili conclusioni riguardanti il miglioramento di alcune situazioni che potrebbero essere ipoteticamente migliorate dalle autorità statali e competenti.