GIADA ALESSI

Nel file excel le colonne/tabelle con le risposte agli esercizi sono sottolineate e in grassetto.

Le altre sono servite per arrivare alla soluzione.

In questo file: Spiegazione passaggi in azzurro

ESERCIZIO

Aprire un nuovo file excel e creare un modello dati utilizzando i due file nella cartella (COMUNI.xlsx di tipo Excel e RIPARTIZIONE-GEOGRAFICA.TXT di tipo TESTO) seguendo le seguenti istruzioni/richieste:

●      Importare il foglio presente nel file COMUNI.XLSX e tramite POWER QUERY applicare tutti i passaggi per pulire/preparare i dati (righe/colonne vuote, ultime righe, maiuscole/minuscole, intestazione, tipologia dei dati, ecc.)

Fatto con PowerQuery (puliti da colonne vuote, righe vuote, da spazi e caratteri problematici con strumenti Taglio e Pulisci, rimossi duplicati nelle colonne in cui non dovevano esserci (es: Regione), usato l’apposita funzione per far diventare ogni nome scritto con l’iniziale maiuscola, ecc.).

Controllato che il tipo di dati fosse corretto.

Una volta caricata la tabella, ricalcolato il totale (con funzione “Somma”) poiché non corrispondeva a quello fatto in seguito con altri calcoli.

●      Creare una nuova colonna per calcolare la percentuale POPOLAZIONE sul TOTALE per ogni COMUNE

Calcolata nel foglio COMUNI, dividendo il valore della popolazione del comune per la popolazione totale.

●      Creare una nuova colonna dove inserire la percentuale POPOLAZIONE sul TOTALE DI OGNI REGIONE per ogni COMUNE (n.b. vi è più di un modo di risolvere questo punto e non è richiesto di farlo in un unico passaggio)

Prima ho calcolato il totale di ogni regione con la funzione “Somma.se” (nel foglio “Totale\_Popolazione\_Per\_Regione”), controllando i risultati anche con una Pivot.

Poi ho creato, usando la funzione “cerca.x”, una nuova colonna che per ogni riga riportasse il totale di regione, in base alla regione di appartenenza del comune. Poi ho diviso la popolazione di ogni comune per il totale di regione. (Foglio COMUNI).

●      Creare una colonna dal nome MAGGIORE dove scrivere MAGGIORE solo se POPOLAZIONE è maggiore di 300.000 altrimenti non scrivere nulla

Usata la funzione “SE” (foglio COMUNI).

●      Importare il file testo RIPARTIZIONE-GEOGRAFICA.TXT e preparare i dati tramite POWER QUERY se necessario.

Usata la funzione “Recupera dati” e creato un nuovo foglio chiamato “Ripartizione-Geografica”.

Con la funzione “Cerca.x” ho aggiunto alla tabella anche il totale della popolazione di regione (per poter poi creare facilmente una Pivot che mi restituisse il totale dei casi in base all’area geografica).

●      Inserire la tabella nel modello dati tramite POWER PIVOT (se si ha lo strumento) e creare la relazione con la tabella comuni tramite il campo REGIONE. In alternativa, aggiungere la colonna RIPARTIZIONE-GEOGRAFICA alla tabella precedente

Fatto dopo aver controllato che ci fosse corrispondenza tra i nomi delle regioni nelle due tabelle (intestazione inclusa).

●      Importare i dati presenti nel pagina Web <https://www.curaitalia.it/covid-19-regioni> TRAMITE IMPORTA DA WEB e pulirli, se necessario. (Se non si riesce a scaricare i dati, usare i dati covid19\_italy\_region)

Importato i dati dal sito web e inseriti nel nuovo foglio “Covid”.

●      Mantenere solo i campi (colonne): REGIONE, TOTALE, GUARITI, DECEDUTI

Tolte le colonne non necessarie con PowerQuery.

●      Verificare che le Regioni siano scritte come nelle basi dati precedenti

Modificate su PowerQuery.

●      SE SI HA POWER PIVOT: Importare nel modello dati e creare le relazioni con le tabelle COMUNI e RIPARTIZIONE GEOGRAFICA tramite il campo REGIONE. In alternativa, aggiungere le colonne necessarie alla tabella comuni.xlsx

Creato il collegamento con PowerPivot.

●      Calcolare: totale deceduti, totale guariti. Rapporto con la popolazione per area e regione. Creare una colonna per concatenare REGIONE e AREA.

Calcolati i totali nel foglio “Covid” come totali di colonna di “deceduti” e “guariti”.

Sempre nel foglio “Covid”, trovato il rapporto fra:

I casi di covid per ogni regione e il totale dei casi;

Il rapporto tra guariti per ogni regione e ammalati totali;

Il rapporto tra deceduti per ogni regione e ammalati totali;

Dopo aver calcolato con una Pivot i totali di casi in base alla partizione geografica (area) ho calcolato il rapporto tra i casi per ogni area e il totale dei casi (tabella Azzurra nel foglio “Covid”; la stessa tabella è presente, per la fretta, anche nel foglio “Ripartizione-geografica”, accompagnata da una Pivot con cui mi sono aiutata per i calcoli).

Tutte queste misure sono a mio parere interessanti in quanto permettono di capire qual è la regione/area geografica con più ammalati (e quindi capire se per esempio in Piemonte ci sono stati più o meno casi che in Sicilia, o più in generale se al nord ci sono stati più casi che al sud, se le regioni con più persone sono anche quelle con più casi perché più persone=più possibilità di contagiati potenziali, ecc.).

In base ai risultati delle analisi si possono poi fare delle ipotesi (es: le Isole hanno meno ammalati del resto d’Italia. Forse essendo isole, e quindi difficili da raggiungere/meno collegate con il resto d’Italia, il virus è arrivato con più difficoltà?).

Concatenati area e regione con funzione Concat nel foglio “Covid” (anche nel foglio “Ripartizione-geografica” per la fretta).

Ho usato la funzione “cerca.x” per aggiungere facilmente alla tabella le varie aree geografiche per ogni regione, in modo da usare agevolmente la concatenazione.

●      Creare una tabella per riassumere tutte le misure dal nome MISURE.

Creata inserendo le misure più interessanti con la funzione Cerca.x (foglio Misure).

Prendere la tabella covid19\_italy\_province e inserirla nel modello dei dati/aggiungere le colonne alla tabella comuni.

Inserita con PowerQuery, cambiati i nomi delle regioni in modo che corrispondessero a quelle delle precedenti tabelle (intestazione inclusa) e aggiunta al modello dati, creando un collegamento con “Covid” basato appunto sulla regione.

Effettuare delle analisi ritenute interessanti con tutti i dati a disposizione. Consegnare un foglio docx in cui si propongono le analisi suddette corredate da screen di tabelle riepilogative o grafici. Il documento presentato sarà valutato in base alla sua coerenza, profondità e ampiezza delle analisi proposte, originalità.