Analisi e riassunto misure

**N.B. mancanza Power Pivot**

Foglio di lavoro “Comuni”:

* Foglio di lavoro importato tramite Power Query, lavorato in modo tale da pulire il file da dati in eccesso e/o non corretti;
* Creazione colonna “%TOTALE”: calcolata la percentuale della popolazione divisa per comune sul totale della stessa; Formattazione della colonna in %;
* Creazione Tabella Pivot per estrapolare il dato del totale popolazione suddiviso per regione;
* Creazione colonna “%TOT.REGIONE” in cui viene calcolata la percentuale della popolazione sul totale della popolazione in base alla regione (dato generato dalla tabella pivot)
* Creazione colonna “MAGGIORE” in cui viene indicato ‘Maggiore’ in corrispondenza del comune con una popolazione maggiore/uguale a 300.000 persone.

Foglio di lavoro “Ripartizione\_geografica”

* Foglio di lavoro importato tramite Power Query, lavorato in modo tale da pulire il file da dati in eccesso e/o non corretti;

Foglio di lavoro “Merge\_comuni\_ripartizioni”

* Tramite Power Query creazione di un Merge tra i due fogli di lavoro precedenti, creando la relazione tramite il campo ‘Regione’ , quindi creazione di una nuova colonna “ripartizione geografica”, in modo che per ogni regione risulta la corrispondete area geografica.

Foglio di lavoro “Covid\_19\_Italy”

* Foglio di lavoro importato tramite Power Query modalità WEB, lavorato in modo tale da pulire il file da dati in eccesso e/o non corretti;
* Eliminati i campi e le colonne non necessarie;
* Verifica e modifica nomi delle regione tramite comando “Sostituisci” in Query;
* Creazione di un Merge con il foglio di lavoro “Merge\_comuni\_ripartizioni” tramite il campo Regione;
* Creazione Tabella Pivot per estrapolare il dato della media del totale suddiviso per aree;
* Creazione colonna “%TOT.AREA” in cui viene calcolata la percentuale della popolazione sul totale in base all’area(dato generato dalla tabella pivot);
* Creazione colonna “Regione-Area” utilizzando la funzione “concatena” tra i valori corrispondenti alle colonne “ Regione – Ripartizione geografica”;

ANALISI DATA\_SET “covid19\_italy\_province”

Il primo passaggio è stato importare tramite Power Query il file csv datoci in consegna, lo stesso è stato ripulito dalle colonne e dai valori ritenuti non necessari al fine dello svolgimento delle analisi che riporto in seguito.

La criticità che si nota subito è che nonostante sia un dataset impostato su dati delle province, risultano nella stessa colonna “ProvinceName” molti valori nulli.

Come prima analisi si è calcolato il totale complessivo annuo, in questo caso del 2020, dei casi di positività al Covid suddiviso anche in regioni.

I dati risultati sono suddivisi di seguito, con un complessivo di 103058478.

|  |  |
| --- | --- |
| Lombardia | 32943004 |
| Piemonte | 11661450 |
| Emilia-Romagna | 10087931 |
| Veneto | 8862185 |
| Campania | 6167917 |
| Lazio | 5731966 |
| Toscana | 5683045 |
| Liguria | 3950192 |
| Sicilia | 2754860 |
| Puglia | 2660799 |
| Marche | 2446508 |
| Friuli Venezia Giulia | 1631995 |
| Abruzzo | 1572466 |
| P.A. Trento | 1552471 |
| P.A. Bolzano | 1311720 |
| Umbria | 1142485 |
| Sardegna | 1099868 |
| Calabria | 758798 |
| Valle d'Aosta | 473043 |
| Basilicata | 329039 |
| Molise | 236736 |
|  |  |

Vediamo quindi in ordine crescente come la regione Lombardia si piazza prima per il maggior numero di casi positivi, caso contrario il Molise ultimo con il minor numero.

Come seconda analisi invece vediamo l’andamento della positività sempre suddiviso per regione ma su una scala dell’anno divisa in quattro trimestri.

Dal precedente grafico a linee vediamo come in modo molto lineare durante il primo trimestre i casi di positività sono molto bassi tranne in modo più evidente per la regione Lombardia.

Infatti sulla la stessa regione si nota che già al secondo trimestre l’aumento impatta molto, fino ad arrivare all’ultimo trimestre dove raggiunge la soglia di positività più alta di tutta l’Italia.

Inoltre vediamo anche come nell’ultimo trimestre risulta una crescita molto esponenziale su molte regioni.

Come ultima analisi si è calcolata la media mobile di 7 giorni.

**N.B. in fase di ultima verifica mi sono accorto che per la regione “Trentino” i dati sono suddivisi per Bolzano e Trento”**