

PROGETTO SQL MICHELE CECCARINI

Importante:

Di seguito tutto il codice che ho utilizzato e gli appunti presi via via. Ho deciso di non riportare tutto nella presentazione PP in quanto alcune cose le ho cercate per pura curiosità (e divertimento con SQL), ma se le avessi inserite nella presentazione finale, questa sarebbe risultata dispersiva e confusionaria. Spero apprezziate la scelta.

Buon lavoro.

Michele

-- Creazione database con pgadmin tramite comando apposito--

--Creazione delle tabelle principalmente in VARCHAR() in quanto ho riscontrato difficoltà a caricare successivamente i file csv con i reali datatype--

```
CREATE TABLE IF NOT EXISTS world_data
(
country text COLLATE pg_catalog."default",
"Density (P/Km2)" varchar,
abbreviation text COLLATE pg_catalog."default",
agricultural_land varchar,
land_area varchar,
armed_forces_size varchar,
birth_rate double precision,
calling_code integer,
capital_major_city text COLLATE pg_catalog."default",
co2_emissions varchar,
cpi varchar,
cpi_change varchar,
currency_code text COLLATE pg_catalog."default",
fertility_rate double precision,
forested_area varchar,
gasoline_price varchar,
gdp varchar,
gross_primary_education_enrollment varchar,
gross_tertiary_education_enrollment varchar,
infant_mortality numeric,
largest_city text COLLATE pg_catalog."default",
life_expectancy double precision,
maternal_mortality_ratio integer,
```

minimum_wage varchar,
official_language text COLLATE pg_catalog."default",
out_of_pocket_health_expenditure varchar,
physicians_per_thousand double precision,
population varchar,
population_labor_force_participation varchar,
tax_revenue varchar,
total_tax_rate varchar,
unemployment_rate varchar,
urban_population varchar,
latitude numeric,
longitude numeric)

CREATE TABLE energy
(
Entity VARCHAR,
Year VARCHAR,
Access_to_electricity VARCHAR,
Access_to_clean_fuels_for_cooking VARCHAR,
Renewable_electricity_generating_capacity_per_capita VARCHAR,
Financial_flows_to_developing_countries VARCHAR,
Renewable_energy_share_in_total_final_energy_consumption VARCHAR,
Electricity_from_fossil_fuels VARCHAR,
Electricity_from_nuclear VARCHAR,
Electricity_from_renewables VARCHAR,
Low_carbon_electricity VARCHAR,
Primary_energy_consumption_per_capita VARCHAR,
Energy_intensity_level_of_primary_energy VARCHAR,
Value_co2_emissions VARCHAR,
Renewables VARCHAR,
GDP_growth VARCHAR,
GDP_per_capita VARCHAR,

```
Density VARCHAR,  
Land_Area VARCHAR,  
Latitude VARCHAR,  
Longitude VARCHAR  
)
```

```
CREATE TABLE missing_migrants  
(  
Incident_Type VARCHAR,  
Incident_Year VARCHAR,  
Reported_Month VARCHAR,  
Region_of_Origin VARCHAR,  
Region_of_Incident VARCHAR,  
Country_of_Origin VARCHAR,  
Number_of_Dead VARCHAR,  
Minimum_Estimated_Number_of_Missing VARCHAR,  
Total_Number_of_Dead_and_Missing VARCHAR,  
Number_of_Survivors VARCHAR,  
Number_of_Females VARCHAR,  
Number_of_Males VARCHAR,  
Number_of_Children VARCHAR,  
Cause_of_Death VARCHAR,  
Migration_Route VARCHAR,  
Location_of_Death VARCHAR,  
Information_Source VARCHAR,  
Coordinates VARCHAR,  
UNSD_Geographical_Grouping VARCHAR)
```

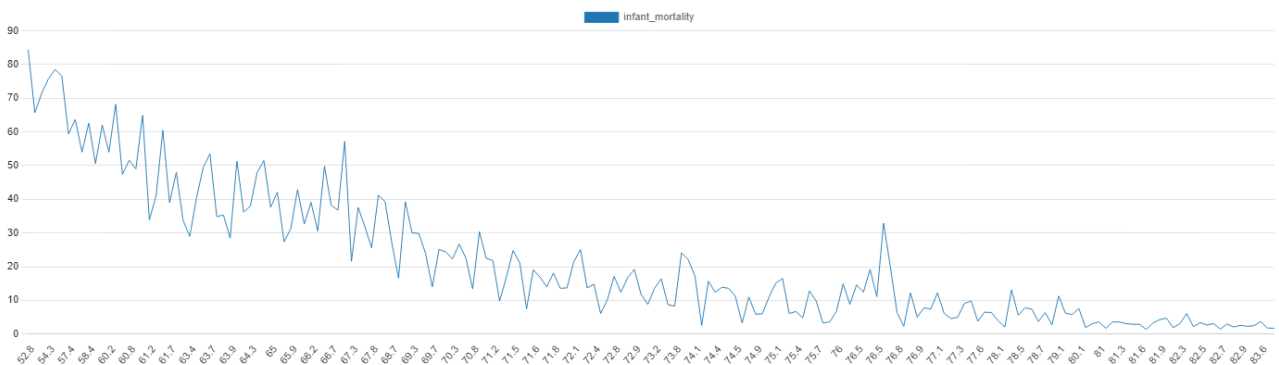
-- Import data csv e assegnazione PRIMARY KEY alla colonna "Country" nella tabella "World_data" tramite "proprietà tabella" tasto destro su tabella--

ANALISI PARAMETRI SOCIOSANITARI ED ENERGETICI

- Quali sono i paesi che hanno maggiori difficoltà dal punto di vista della mortalità?
- Quali sono i fattori che influenzano questi dati?
- Un'occhiata ai livelli di fertilità
- Più emissioni = maggiore crescita economica?
- Conclusioni

--Prime analisi--

```
SELECT
country,
population,
fertility_rate,
infant_mortality,
life_expectancy
FROM
world_data
WHERE
(country, population, fertility_rate, infant_mortality, life_expectancy) IS NOT NULL
ORDER BY
life_expectancy
```



-- Evidente correlazione inversa tra aspettativa di vita e mortalità infantile. Paesi dove l'aspettativa di vita è inferiore hanno la più alta concentrazione di mortalità infatile e il tasso di fertilità maggiore--

--Valutazione della presenza dei medici x 1000 abitanti nei paesi con più bassa aspettativa di vita--

WITH medici AS (

SELECT

country,

physicians_per_thousand

FROM

world_data

WHERE

physicians_per_thousand IS NOT NULL

ORDER BY

physicians_per_thousand DESC

)

SELECT

wd.country,

wd.fertility_rate,

wd.infant_mortality,

wd.life_expectancy,

m.physicians_per_thousand AS medici

FROM

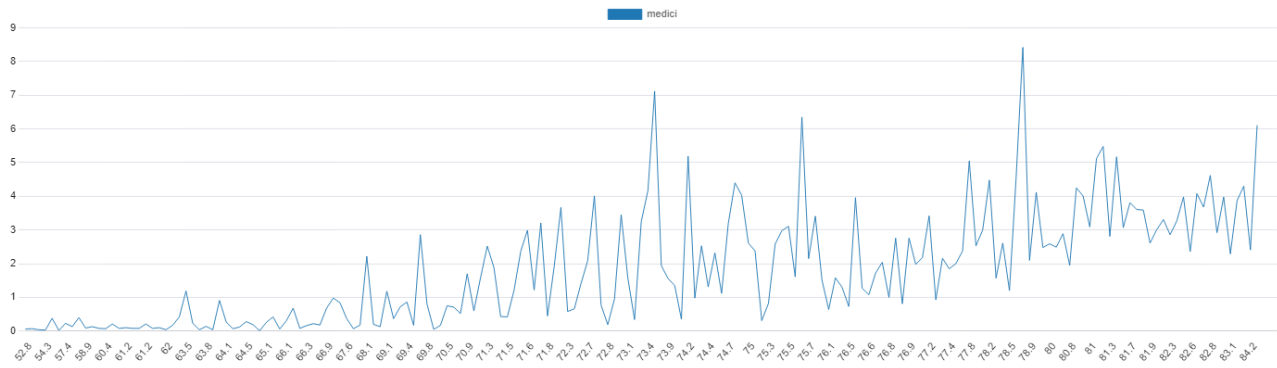
world_data wd

JOIN

medici m ON wd.country = m.country

ORDER BY

wd.life_expectancy;



-- Notiamo anche in questo caso una correlazione tra le variabili aspettativa di vita e presenza di medici, una maggiore presenza di medici è correlata ad una maggiore aspettativa di vita—

-- Da tenere presente che l'aspettativa di vita è fortemente influenzata dalla mortalità infantile che, essendo più alta, abbassa inevitabilmente il dato sull'aspettativa di vita che ne tiene conto--

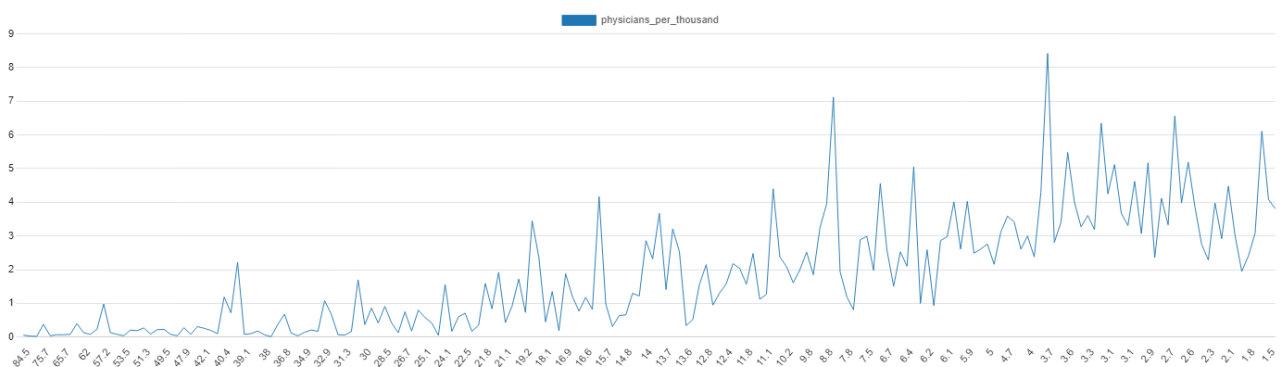
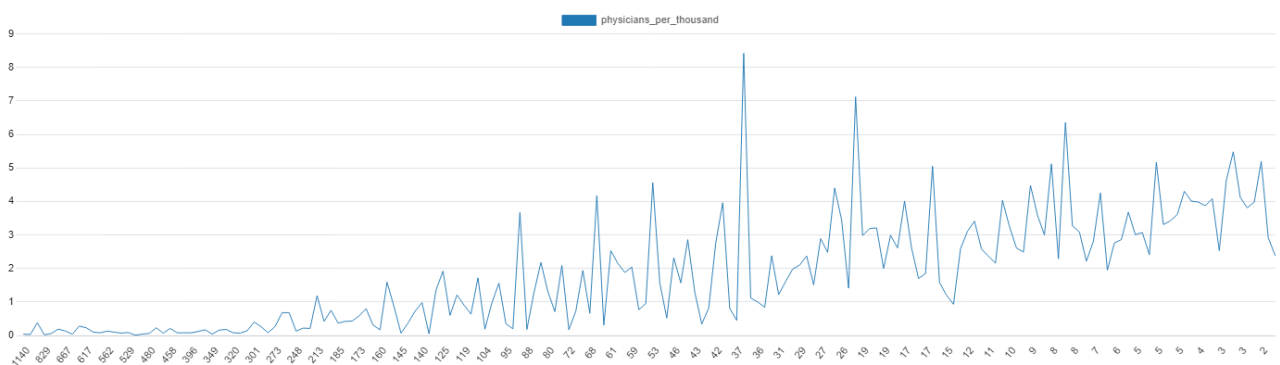
--E per quanto riguarda il tasso di mortalità delle madri?

```
SELECT
country,
maternal_mortality_ratio,
physicians_per_thousand
FROM
world_data
WHERE
(maternal_mortality_ratio, physicians_per_thousand) IS NOT NULL
ORDER BY
maternal_mortality_ratio DESC
```

```

SELECT
country,
infant_mortality,
physicians_per_thousand
FROM
world_data
WHERE
(infant_mortality, physicians_per_thousand) IS NOT NULL
ORDER BY
infant_mortality DESC

```

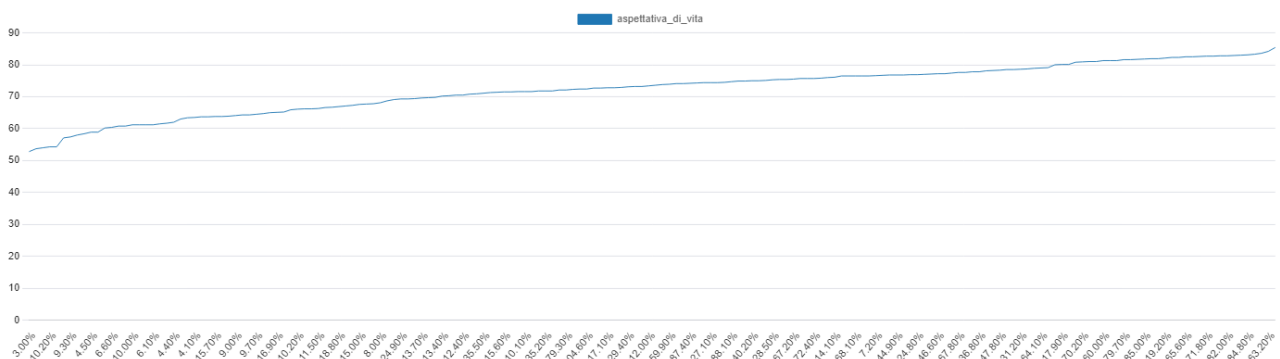


-- La presenza di medici aiuta a ridurre anche il tasso di mortalità materna, oltre che quella infantile. I dati sono praticamente sovrapponibili--

--Cos'altro può aiutare ad alzare l'aspettativa di vita e ridurre la mortalità?--

-- Prendiamo ad esempio il fattore istruzione--

```
SELECT
country,
life_expectancy AS aspettativa_di_vita,
gross_primary_education_enrollment as
scuola_primaria,
gross_tertiary_education_enrollment as scuola_secondaria
FROM
world_data
WHERE
(life_expectancy,gross_primary_education_enrollment,gross_tertiary_education_enrollment) IS
NOT NULL
GROUP BY
country
ORDER BY
aspettativa_di_vita
```



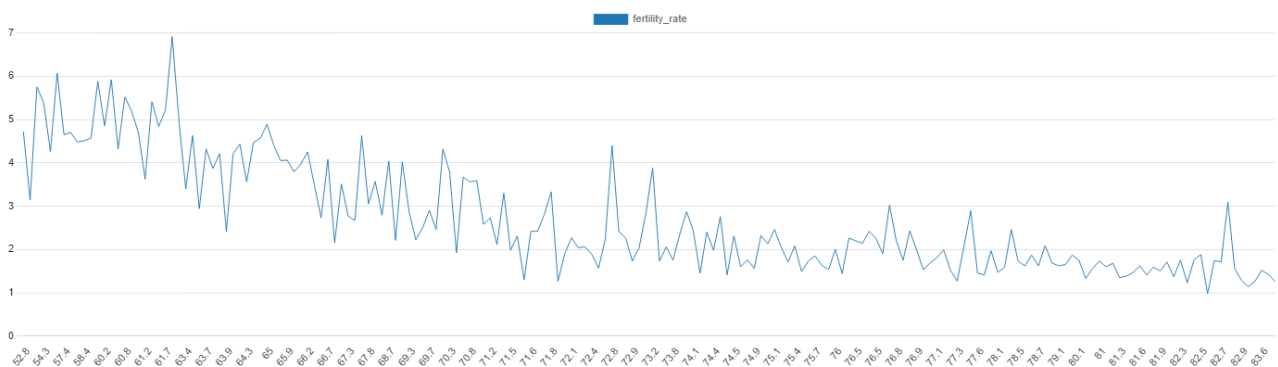
-- Un maggior livello di istruzione secondaria porta ad un lineare aumento dell'aspettativa di vita--

--Paesi con più bassa aspettativa di vita sono tendenzialmente caratterizzati da forti problematiche sociali, ambientali ed economiche. Dalla letteratura sappiamo che sono anche i paesi in cui si fanno più figli, a causa della necessità di manodopera agricola, estrema povertà

che limita l'aspettativa di vita e mancanza di una struttura welfare adeguata che promuova la contraccezione. Una condizione di scarsa emancipazione femminile contribuisce a creare una situazione estremamente difficoltosa per un sano sviluppo economico e sociale.

-- E' interessante notare che gli stessi paesi dove l'aspettativa di vita è inferiore, vi è anche il tasso di fertilità maggiore. Segue approfondimento.—

```
SELECT
    fertility_rate as tasso_di_fertilità,
    life_expectancy as aspettativa_di_vita
FROM
    world_data
ORDER BY
    aspettativa_di_vita
```



-- Come organizzare suddividere i vari tassi di fertilità? --

Molto Basso (0-1.5): Tasso di fertilità molto basso, indicativo di una bassa crescita demografica e possibili sfide legate all'invecchiamento della popolazione.

Basso (1.6-2.0): Tasso di fertilità basso, che può indicare una crescita demografica modesta, ma ancora inferiore al livello di sostituzione.

Moderato (2.1-2.5): Tasso di fertilità moderato, indicativo di una sostituzione generazionale e una crescita demografica stabile.

Alto (2.6-3.0): Tasso di fertilità alto, che può portare a una crescita demografica significativa e ad una popolazione giovane.

Molto Alto (oltre 3.0): Tasso di fertilità molto alto, che può indicare una crescita demografica rapida, con implicazioni sulla distribuzione demografica e delle risorse.

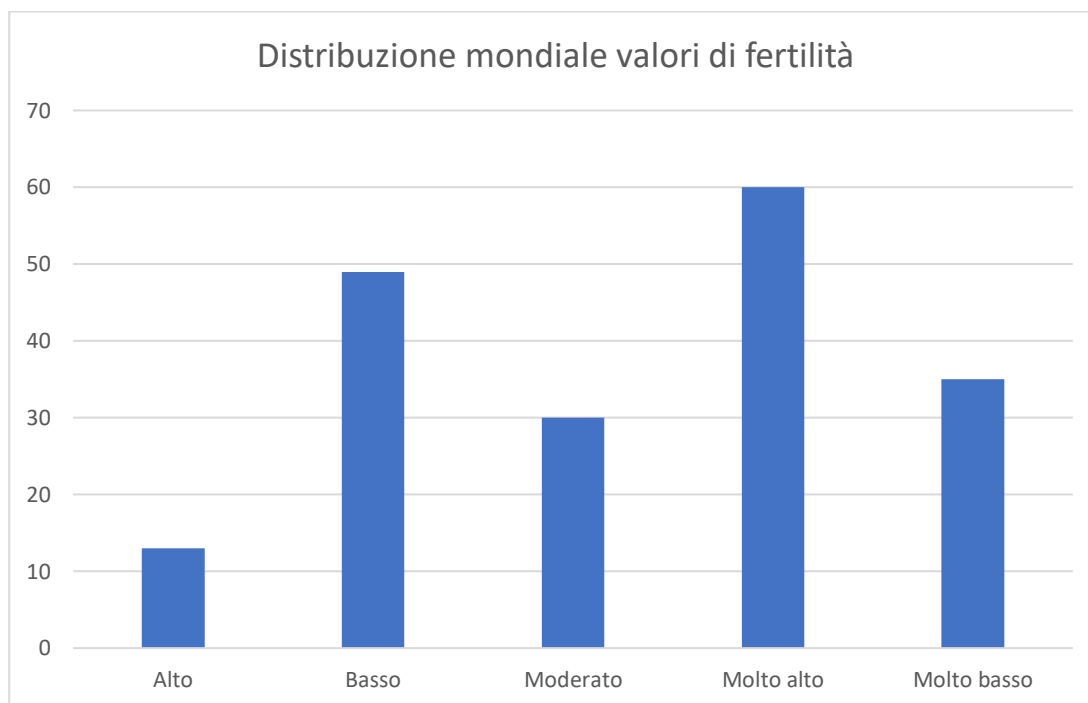
```
SELECT
country AS stato,
fertility_rate AS tasso_di_fertilità,
CASE
WHEN fertility_rate BETWEEN 0 AND 1.5 THEN 'Molto basso'
WHEN fertility_rate BETWEEN 1.6 AND 2 THEN 'Basso'
WHEN fertility_rate BETWEEN 2.1 AND 2.5 THEN 'Moderato'
WHEN fertility_rate BETWEEN 2.6 AND 3 THEN 'Alto'
WHEN fertility_rate > 3 THEN 'Molto alto'
ELSE 'NULL'
END AS valutazione_fertilità
FROM
world_data
ORDER BY tasso_di_fertilità;
```

-- Ricontrato errore in quanto valori come 2.05,2.06,2.07,2.08, 2.09 risultano null--

--Risolto di seguito--

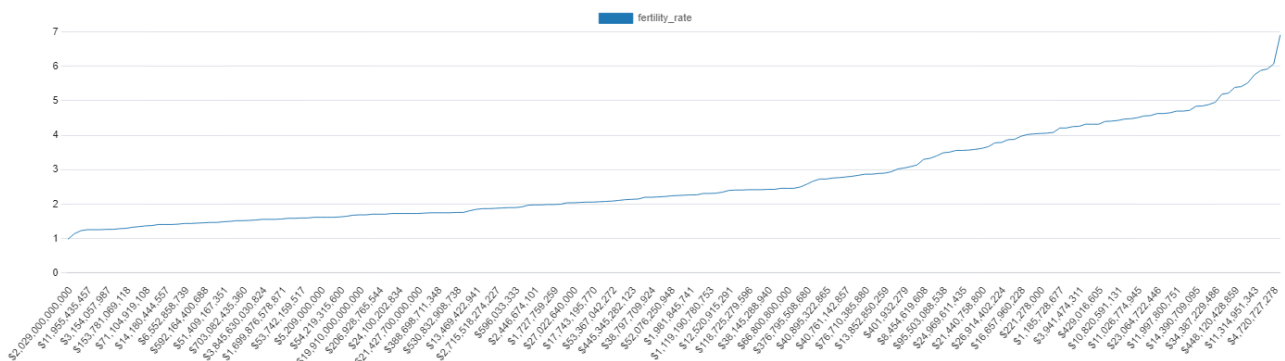
```
SELECT
country AS stato,
fertility_rate AS tasso_di_fertilità,
CASE
WHEN fertility_rate >= 0 AND fertility_rate < 1.6 THEN 'Molto basso'
WHEN fertility_rate >= 1.6 AND fertility_rate < 2.1 THEN 'Basso'
WHEN fertility_rate >= 2.1 AND fertility_rate < 2.6 THEN 'Moderato'
WHEN fertility_rate >= 2.6 AND fertility_rate < 3.0 THEN 'Alto'
```

```
WHEN fertility_rate >= 3.0 THEN 'Molto alto'
ELSE 'NULL'
END AS valutazione_fertilità
FROM
world_data
ORDER BY tasso_di_fertilità
```



--Tasso di fertilità comparato al GDP--

```
SELECT
country,
fertility_rate,
gdp
FROM
world_data
WHERE
fertility_rate IS NOT NULL
ORDER BY
fertility_rate
```



--Osservando i dati emerge che i paesi più sviluppati hanno anche un tasso di natalità piuttosto basso. Dagli studi di demografia emerge questa forte tendenza legata allo sviluppo economico e sociale di un paese in cui più migliorano le condizioni di vita e più scende la natalità. Questo è dovuto a tanti fattori, alcuni già descritti per la situazione opposta nella query che confronta aspettativa di vita e istruzione. Le implicazioni di queste osservazioni sono innumerevoli, prima tra tutte quella del veicolare lo sviluppo dei paesi meno sviluppati da parte di quelli più ricchi, fornendo innovazione tecnologica e risorse.—

-- A questo punto mi chiedo qual'è la media della mortalità infantile dei 20 paesi più emettitori di CO2?--

```
WITH PaesiConMassimeEmissioni AS (  
    SELECT  
        country,  
        AVG(CAST(REPLACE(co2_emissions, ',', '') AS DECIMAL(18,2))) AS Media_Emissioni  
    FROM  
        world_data  
    WHERE  
        infant_mortality IS NOT NULL AND co2_emissions IS NOT NULL  
    GROUP BY  
        country  
    ORDER BY  
        Media_Emissioni DESC  
    LIMIT 20  
)  
SELECT  
    AVG(wd.infant_mortality) AS Media_Mortalità_Infantile  
FROM  
    PaesiConMassimeEmissioni pcm  
JOIN  
    world_data wd ON pcm.country = wd.country;
```

--Risultato: 8.915--

-- E quella dei 20 paesi che emettono meno?--

```
WITH PaesiConMassimeEmissioni AS (  
    SELECT  
        country,  
        AVG(CAST(REPLACE(co2_emissions, ',', '') AS DECIMAL(18,2))) AS Media_Emissioni  
    FROM  
        world_data  
    WHERE  
        infant_mortality IS NOT NULL AND co2_emissions IS NOT NULL  
    GROUP BY  
        country  
    ORDER BY  
        Media_Emissioni ASC  
    LIMIT 20  
)  
SELECT  
    AVG(wd.infant_mortality) AS Media_Mortalità_Infantile  
FROM  
    PaesiConMassimeEmissioni pcm  
JOIN  
    world_data wd ON pcm.country = wd.country;
```

-- Risultato: 27.005--

--Questo risultato è veramente molto interessante e porta ad una considerazione, un abbassamento della mortalità infantile non può prescindere da un marcato sviluppo economico di un paese e quindi, purtroppo, ad un suo aumento delle emissioni di CO2--

--A questo punto mi chiedo qual'è stato il GDP 2018 dei primi 20 paesi che hanno emesso più co2?—

SELECT

Entity AS stato,

Year AS anno,

CAST(value_co2_emissions AS decimal(18,2)) AS emission_co2,

gdp_growth AS crescita_gdp

FROM

energy

WHERE (entity, value_co2_emissions,gdp_growth) IS NOT NULL AND year = '2018'

ORDER BY value_co2_emissions desc

LIMIT 20

	stato character varying 🔒	anno character varying 🔒	emissioni_co2 numeric (18,2) 🔒	crescita_gdp character varying 🔒
1	China	2018	10502929.69	6.749773832
2	United States	2018	4975310.06	2.996464352
3	India	2018	2451929.93	6.532989011
4	Japan	2018	1116150.02	0.558851275
5	Germany	2018	707700.01	1.086024514
6	Canada	2018	580090.03	2.777040554
7	Indonesia	2018	576989.99	5.17429154
8	Saudi Arabia	2018	521260.01	2.434110781
9	Mexico	2018	452570.01	2.194994725
10	South Africa	2018	434350.01	1.487617373
11	Brazil	2018	433989.99	1.783666761
12	Australia	2018	387070.01	2.87097248
13	United Kingdom	2018	360730.01	1.650925496
14	Italy	2018	324880.00	0.925810941
15	Poland	2018	311910.00	5.353703355
16	France	2018	307049.99	1.865066071
17	Thailand	2018	257049.99	4.189585496
18	Spain	2018	257040.01	2.288785676
19	Malaysia	2018	244410.00	4.84305661
20	Kazakhstan	2018	216600.01	4.1

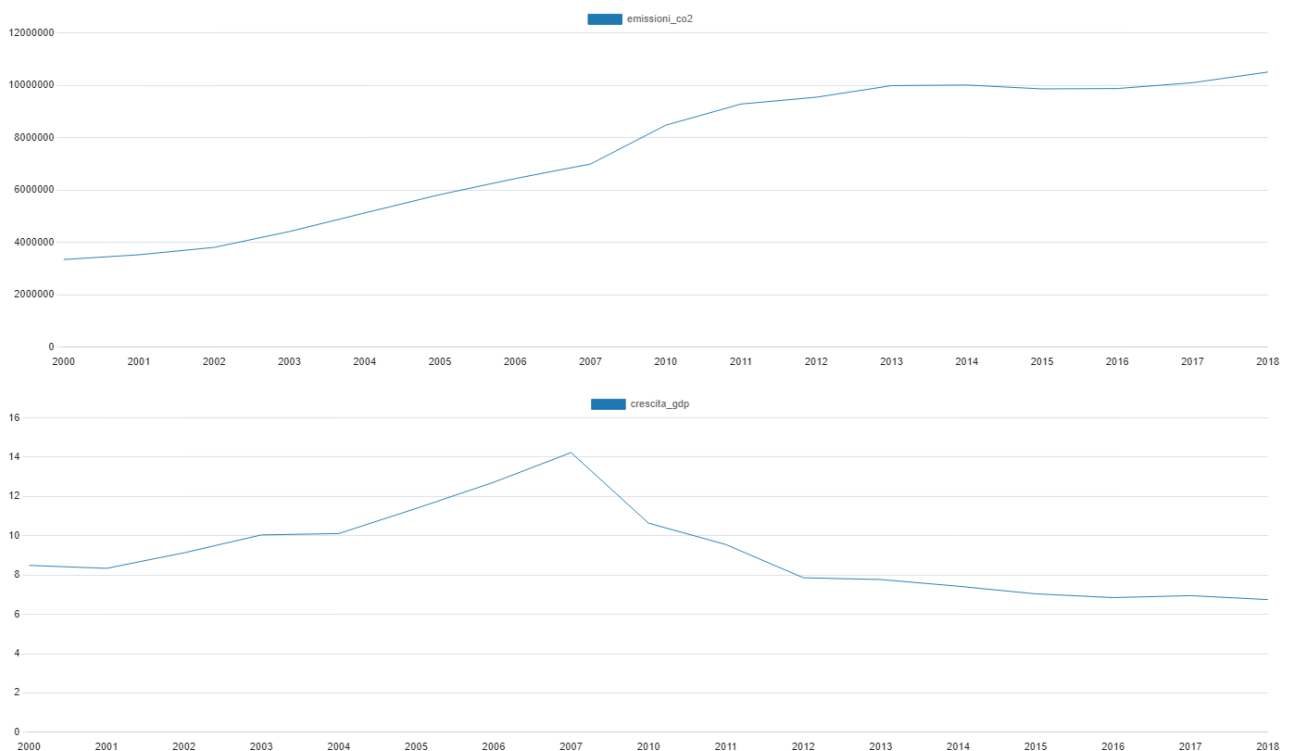
--Confronto tra i paesi che emettono di più--

```
SELECT
    entity AS stato,
    year as anno,
    CAST(value_co2_emissions AS decimal(18,2)) AS emissioni_CO2,
    CAST(gdp_growth AS decimal(18,2)) AS crescita_GDP
FROM
    energy
WHERE
    year != '2008'
    AND year != '2009'
    AND year != '2019'
    AND value_co2_emissions IS NOT NULL
    AND entity IN ('United States','China','India')
ORDER BY
    CAST(value_co2_emissions AS decimal(18,2)) DESC
```

--Ho deciso di escludere gli anni 2008,2009 e 2019 a causa dell'intervento delle crisi finanziarie e sanitarie che hanno interessato questi anni e che avrebbero potuto confondere l'interpretazione dei dati--

-- Sono incuriosito dai dati cinesi--

```
SELECT
    entity AS stato,
    year as anno,
    CAST(value_co2_emissions AS decimal(18,2)) AS emissioni_CO2,
    CAST(gdp_growth AS decimal(18,2)) AS crescita_GDP
FROM
    energy
WHERE
    AND value_co2_emissions IS NOT NULL
    AND entity IN ('China')
ORDER BY
    Year
```



-- La crescita del GDP cinese ha subito una battuta di arresto durante la crisi finanziaria del 2007-2008 e nonostante abbia mantenuto un livello crescente di emissioni, il valore del gdp non ha smesso di scendere--

-- A questo punto mi interessa confrontare con la situazione degli Stati Uniti—

SELECT

entity AS stato,

year as anno,

CAST(value_co2_emissions AS decimal(18,2)) AS emissioni_CO2,

CAST(gdp_growth AS decimal(18,2)) AS crescita_GDP

FROM

energy

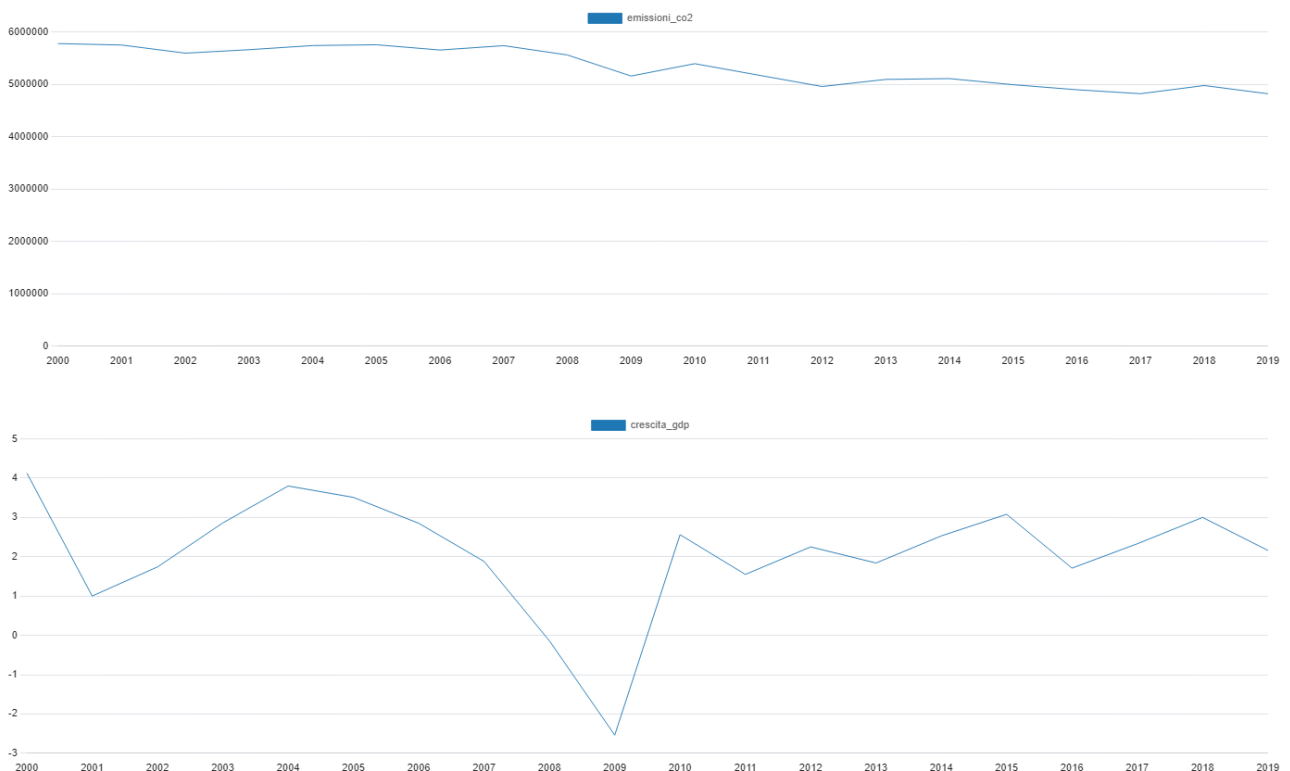
WHERE

value_co2_emissions IS NOT NULL

AND entity IN ('United States')

ORDER BY

year

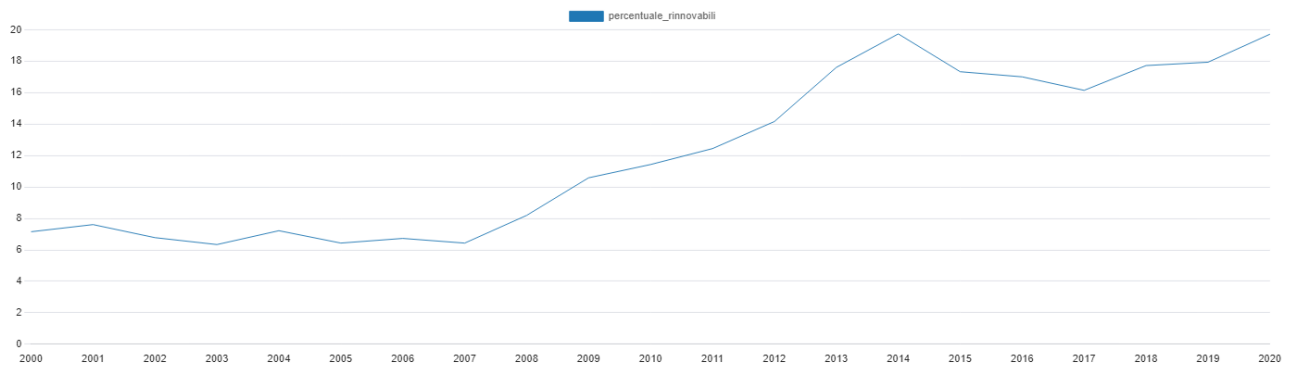


-- Notiamo come le emissioni siano in costante, anche se lieve, calo. Al contrario della Cina, gli USA, successivamente alla crisi del 2008, hanno mantenuto livelli di emissioni decrescenti, nonostante una ripresa del GDP—

-- E in italia?--

--Osservare l'utilizzo di fonti rinnovabili in Italia partendo dall'anno in cui ne abbiamo utilizzate maggiormente. Il 2014!--

```
SELECT
entity AS stato,
year AS Anno,
CAST(renewables AS decimal) as percentuale_rinnovabili
FROM
energy
WHERE
entity LIKE '%al%' AND entity LIKE 'I%'
ORDER BY
percentuale_rinnovabili DESC
```



--Analizzare la relazione tra aspettativa di vita, istruzione e impatto ambientale. Esaminare come le condizioni di vita influenzano la sostenibilità ambientale. --

--Utilizzo di una tabella temporanea che esclude i valori nulli da poter usare successivamente per una funzione JOIN con la tabella principale --

-- Ho preferito utilizzare JOIN in questo modo in quanto l'unica tabella in cui sono presenti valori univoci utilizzabili per una PRIMARY KEY è il csv world_data_2023, precisamente nella colonna country. Le altre tabelle contengono colonne con dati duplicati, perciò non mi è stato possibile impostare una PK –

WITH SaluteIstruzioneEmissioni AS (

SELECT

wd.Country,
wd.Life_Expectancy,
wd.Gross_Primary_Education_Enrollment,
wd.Gross_Tertiary_Education_Enrollment,
wd.CO2_Emissions,
wd.population

FROM

world_data wd

WHERE

wd.Life_Expectancy IS NOT NULL

AND wd.Gross_Primary_Education_Enrollment IS NOT NULL

AND wd.Gross_Tertiary_Education_Enrollment IS NOT NULL

AND wd.CO2_Emissions IS NOT NULL

AND wd.population IS NOT NULL

)

SELECT

sie.Country,
sie.Life_Expectancy,
sie.Gross_Primary_Education_Enrollment,
sie.Gross_Tertiary_Education_Enrollment,
sie.CO2_Emissions,
sie.population

FROM

SaluteIstruzioneEmissioni sie

JOIN

world_data wd ON sie.Country = wd.Country

ORDER BY

sie.Life_Expectancy DESC;

	country [PK] text	life_expectancy double precision	gross_primary_education_enrollment character varying	gross_tertiary_education_enrollment character varying	co2_emissions character varying	population character varying
1	Japan	84.2	98.80%	63.20%	1,135,886	126,226,568
2	Switzerland	83.6	105.20%	59.60%	34,477	8,574,832
3	Spain	83.3	102.70%	88.90%	244,002	47,076,781
4	Singapore	83.1	100.60%	84.80%	37,535	5,703,569
5	Liechtenstein	83	104.70%	35.60%	51	38,019
6	Italy	82.9	101.90%	61.90%	320,411	60,297,396
7	Norway	82.8	100.30%	82.00%	41,023	5,347,896
8	Israel	82.8	104.90%	63.40%	65,166	9,053,300
9	Iceland	82.7	100.40%	71.80%	2,065	361,313
10	Australia	82.7	100.30%	113.10%	375,908	25,766,605
11	South Korea	82.6	98.10%	94.30%	620,302	51,709,098
12	Sweden	82.5	126.60%	67.00%	43,252	10,285,453
13	France	82.5	102.50%	65.60%	303,276	67,059,887
14	Malta	82.3	105.00%	54.30%	1,342	502,653
15	Republic of Ireland	82.3	100.90%	77.80%	37,711	5,007,069
16	Luxembourg	82.1	102.30%	19.20%	8,988	645,397
17	New Zealand	81.9	100.00%	82.00%	34,382	4,841,000
18	Canada	81.9	100.90%	68.90%	544,894	36,991,981
19	Netherlands	81.8	104.20%	85.00%	170,780	17,332,850
20	Finland	81.7	100.20%	88.20%	45,871	5,520,314
21	Belgium	81.6	103.90%	79.70%	96,889	11,484,055
22	Austria	81.6	103.10%	85.10%	61,448	8,877,067
23	Portugal	81.3	106.20%	63.90%	48,742	10,269,417
24	United Kingdom	81.3	101.20%	60.00%	379,025	66,834,405

--Notiamo una forte correlazione tra numero di abitanti e quantità di emissioni, anche se vi sono alcune eccezioni. Penso ad esempio all'Australia le cui emissioni sono pari a quelle dell'Italia, nonostante quest'ultima abbia più del doppio degli abitanti.--

OFF TOPIC:

--Contare il numero di paesi che parlano una determinata lingua ufficiale --

```
SELECT  
official_language AS Lingua,  
COUNT(country) AS Stati  
FROM  
world_data  
GROUP BY  
official_language  
ORDER BY  
Stati DESC;
```

	lingua text	stati bigint
1	English	31
2	French	25
3	Spanish	19
4	Arabic	18
5	Portuguese	7
6	Russian	4
7	German	4
8	None	4
9	Swahili	4

-- Non sorprende vedere che l'inglese è la lingua parlata nel maggior numero di stati. Una possibile spiegazione della presenza di francese e spagnolo è da ricercarsi probabilmente nella forte impronta colonialista che nei secoli scorsi ha caratterizzato questi 3 paesi, Inghilterra, Francia e Spagna, soprattutto nel continente africano.--