SQL PROJECT
RENEWABLE ENERGY

Dario Giordano for start2impact University

start2impact
UNIVERSITY

INTRODUCTION

This presentation is based on the Renewable Energy World Wide dataset, which includes countries' renewable energy data from 1965 to 2022. The analysis will include a comparison of the main renewable energy sources worldwide, and it will be accomplished using the PostgreSQL language and the softwares pgAdmin 4 and Visual Studio Code.

```sql
--First of all, we create the necessary tables in which we will import the data from the dataset

CREATE TABLE IF NOT EXISTS renewable_share(
    "Entity" text COLLATE pg_catalog."default",
    "Code" text COLLATE pg_catalog."default",
    "Year" bigint,
    "Renewables (% equivalent primary energy)" double precision);
CREATE TABLE IF NOT EXISTS modern_renewable_energy_consumption(
    "Entity" text COLLATE pg_catalog."default",
    "Code" text COLLATE pg_catalog."default",
    "Year" bigint,
    "Geo Biomass Other - TWh" double precision,
    "Solar Generation - TWh" double precision,
    "Wind Generation - TWh" double precision,
    "Hydro Generation - TWh" double precision);
CREATE TABLE IF NOT EXISTS modern_renewable_production(
    "Entity" text COLLATE pg_catalog."default",
    "Code" text COLLATE pg_catalog."default",
    "Year" bigint,
    "Electricity from wind (TWh)" double precision,
    "Electricity from hydro (TWh)" double precision,
    "Electricity from solar (TWh)" double precision
    "Other renewables including bioenergy (TWh)" double precision);
CREATE TABLE IF NOT EXISTS hydro_share_energy(
    "Entity" text COLLATE pg_catalog."default",
    "Code" text COLLATE pg_catalog."default",
    "Year" bigint,
    "Hydro (% equivalent primary energy)" double precision);
CREATE TABLE IF NOT EXISTS wind_share_energy(
    "Entity" text COLLATE pg_catalog."default",
    "Code" text COLLATE pg_catalog."default",
    "Year" bigint,
    "Wind (% equivalent primary energy)" double precision);
CREATE TABLE IF NOT EXISTS solar_share_energy(
    "Entity" text COLLATE pg_catalog."default",
    "Code" text COLLATE pg_catalog."default",
    "Year" bigint,
    "Solar (% equivalent primary energy)" double precision);
-- We then import the data using the built-in pgAdmin 4 "Import/Export" tool
```

The first step of the analysis is the creation of the tables, in which we will import the csv files from the dataset using the built-in Import/Export tool in pg Admin 4.

DATA CLEANING

```
-- After importing the data in the tables, we clean it by deleting the rows which contain aggregate groups of countries
-- (e.g. North America), as our analysis will only be focused on individual countries.
-- To do so, we use the DELETE function and delete all the rows in which "Code" is null,
-- as country aggregates won't have a country code.

DELETE FROM hydro_share_energy
WHERE hydro_share_energy."Code" IS null;
DELETE FROM modern_renewable_energy_consumption
WHERE modern_renewable_energy_consumption."Code" IS null;
DELETE FROM modern_renewable_production
WHERE modern_renewable_production."Code" IS null;
DELETE FROM renewable_share
WHERE renewable_share."Code" IS null;
DELETE FROM solar_share_energy
WHERE solar_share_energy."Code" IS null;
DELETE FROM wind_share_energy
WHERE wind_share_energy."Code" IS null;
```

The original dataset includes 17 tables. Instead of importing all 17 and dropping the unnecessary tables, only the ones that will be used have been imported. From these, I deleted all the aggregate values (except for World) as the analysis will only be focused on individual countries.

DATA CLEANING

```
-- We create a new table "countries" in which we import only the code and entity of each country + World
-- We then set "code" as the primary key for the table, and add it as a foreign key in all other tables
-- to preserve the data in case it's needed further in the analysis

CREATE TABLE countries(
code text,
entity text,
PRIMARY KEY (country_id));
INSERT INTO countries
SELECT "Code"
FROM renewable_share
WHERE renewable_share."Year" = '2000';
UPDATE countries
SET entity = renewable_share."Entity"
FROM renewable_share
WHERE countries.country_id = renewable_share."Code";
-- Having given the condition Year = 2000, USSR was left out, as its data stops in 1984
-- We add it manually using a new INSERT
INSERT INTO countries (code,entity)
VALUES ('OWID_USS','USSR');

-- We can now set the Code column as foreign key in all the tables
ALTER TABLE hydro_share_energy
ADD CONSTRAINT "Code" FOREIGN KEY ("Code")
REFERENCES countries (code);
ALTER TABLE modern_renewable_energy_consumption
ADD CONSTRAINT "Code" FOREIGN KEY ("Code")
REFERENCES countries (code);
ALTER TABLE renewable_share
ADD CONSTRAINT "Code" FOREIGN KEY ("Code")
REFERENCES countries (code);
ALTER TABLE solar_share_energy
ADD CONSTRAINT "Code" FOREIGN KEY ("Code")
REFERENCES countries (code);
ALTER TABLE wind_share_energy
ADD CONSTRAINT "Code" FOREIGN KEY ("Code")
REFERENCES countries (code);
```

I then added a new table called countries, in which I imported the Code and Entity name from the renewable_share table, and added Code as the primary key in the countries table. It was then set as a foreign key in all the other tables.

DATA CLEANING

```sql
-- Lastly, we rename some of the columns in the tables for simpler queries
ALTER TABLE hydro_share_energy
RENAME COLUMN "Hydro (% equivalent primary energy)" TO hydro_percentage;

ALTER TABLE modern_renewable_energy_consumption
RENAME COLUMN "Geo Biomass Other - TWh" TO geo_twh;
ALTER TABLE modern_renewable_energy_consumption
RENAME COLUMN "Solar Generation - TWh" TO solar_twh;
ALTER TABLE modern_renewable_energy_consumption
RENAME COLUMN "Wind Generation - TWh" to wind_twh;
ALTER TABLE modern_renewable_energy_consumption
RENAME COLUMN "Hydro Generation - TWh" to hydro_twh;

ALTER TABLE modern_renewable_production
RENAME COLUMN "Electricity from hydro (TWh)" TO solar_prod_twh;
ALTER TABLE modern_renewable_production
RENAME COLUMN "Electricity from solar (TWh)" TO solar_prod_twh;
ALTER TABLE modern_renewable_production
RENAME COLUMN "Electricity from wind (TWh)" TO wind_prod_twh;
ALTER TABLE modern_renewable_production
RENAME COLUMN "Other renewables including bioenergy (TWh)" TO other_renewables_twh;

ALTER TABLE renewable_share
RENAME COLUMN "Renewables (% equivalent primary energy)" TO renewables_percentage;

ALTER TABLE solar_share_energy
RENAME COLUMN "Solar (% equivalent primary energy)" TO solar_percentage;

ALTER TABLE wind_share_energy
RENAME COLUMN "Wind (% equivalent primary energy)" TO wind_percentage;
```

Lastly, we rename some of the columns inside the tables for simpler queries

DATA ANALYSIS

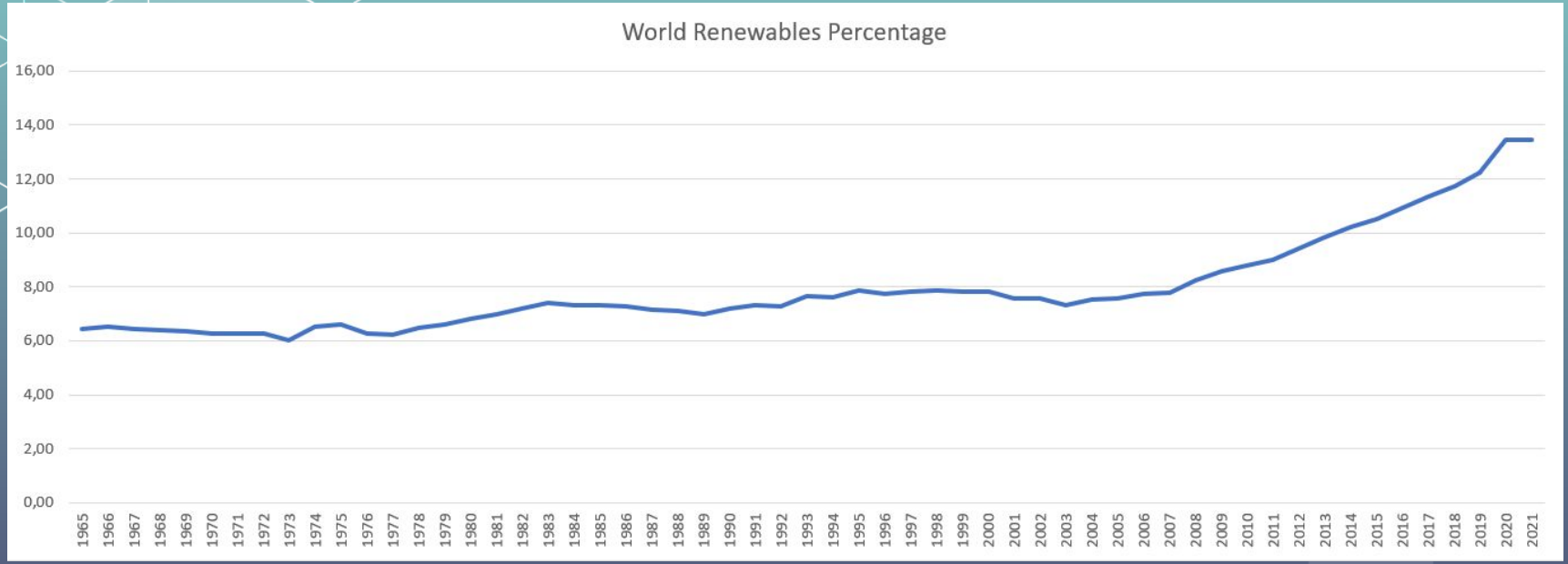
```sql
-- ANALYSIS
-- We start our analysis by looking at the 20 countries who have the best percentage
-- in renewables usage in the latest year available for our analysis

SELECT * FROM renewable_share
WHERE "Year" = (SELECT MAX("Year") FROM renewable_share)
ORDER BY renewables_percentage DESC
LIMIT 20;
```

Progetto SQL: SELECT * FROM r...

Entity	Code	Year	renewables_percentage
Filter...	Filter...	Filter...	Filter...
Iceland	ISL	2021	86.874535
Norway	NOR	2021	71.558365
Sweden	SWE	2021	50.924007
Brazil	BRA	2021	46.21975
New Zealand	NZL	2021	40.21865
Denmark	DNK	2021	39.24958
Austria	AUT	2021	37.481792
Switzerland	CHE	2021	36.716644
Finland	FIN	2021	34.61129
Colombia	COL	2021	33.02041
Portugal	PRT	2021	32.703953
Ecuador	ECU	2021	32.3542
Canada	CAN	2021	29.88844
Venezuela	VEN	2021	28.434155
Croatia	HRV	2021	28.271935
Peru	PER	2021	27.741009
Chile	CHL	2021	26.518875
Latvia	LVA	2021	23.917372
Vietnam	VNM	2021	22.734407
Spain	ESP	2021	22.341663

The first query aims at discovering the top 20 countries with the highest percentage in renewable energy usage, in the latest year we have available in the dataset. From the results we can see that countries in Northern Europe have a much higher share of renewables compared to most other countries, except Brazil and New Zealand, who also have a percentage higher than 40%. 7 countries (among which we find the rest of the Northern European countries) locate between 30% and 40%, whereas the rest of the world countries use less than 30% renewables in their energy mix.

DATA ANALYSIS



From this graph, we can see the percentage of renewable energy use in the world. We can immediately see that the percentage has been between 6% and 8% until 2007, when countries started becoming more aware of their non-renewable energy consumption, then jump to nearly 14% in the last few years.

DATA ANALYSIS

```sql
-- This query aims at discovering the countries with the highest consumption (in TWh) of renewable energy
-- In the modern_renewable_energy_consumption table, we have the total consumption for each renewable energy source
-- We sum them using the COALESCE() function, which will use '0' instead of 'null' in case a column is empty

SELECT "Entity", "Year",
COALESCE(geo_twh ,0) + COALESCE(solar_twh,0) + COALESCE(wind_twh, 0) + COALESCE (hydro_twh,0) AS tot_consumption_twh
FROM modern_renewable_energy_consumption
WHERE "Year" = (SELECT MAX("Year") FROM modern_renewable_energy_consumption)
AND "Entity" <> 'World'
ORDER BY tot_consumption_twh DESC;
```

The second query shows the countries with the highest energy consumption in the latest year available in the dataset. The total consumption is obtained with a sum of the various renewable energy sources.
Excluding “World”, the query returns the highest values for the countries which have the largest populations, as they will clearly have the highest general energy consumption compared to the others.

Entity	Year	tot_consumptio...
Filter...	Filter...	Filter...
China	2021	2452.53158
United States	2021	882.1402049999999
Brazil	2021	506.815918
Canada	2021	430.8209594
India	2021	332.199114
Germany	2021	236.70000000000002
Russia	2021	219.94892199999998
Japan	2021	207.91004700000002
Norway	2021	155.30786815
Spain	2021	125.341543
United Kingdom	2021	121.86582
France	2021	120.7320575
Turkey	2021	118.39397600000001
Italy	2021	114.49766399999999
Sweden	2021	113.543339
Vietnam	2021	104.2195441
Australia	2021	77.329906
Mexico	2021	74.4380835
Colombia	2021	63.056186675999996
Venezuela	2021	61.318909672
Indonesia	2021	56.20112
Austria	2021	56.12081810000001
Chile	2021	45.064792499999996

DATA ANALYSIS

```sql
-- In the same way, we can find the countries with the highest production (in TWh) of renewable energy
-- However, since the latest year available for this table is 2022 (with data from only 27 countries),
-- we will use the latest year available in the consumption table, in order to have an accurate comparison

SELECT "Entity","Year",
COALESCE(wind_prod_twh,0) + COALESCE(hydro_prod_twh,0) + COALESCE (solar_prod_twh,0) + COALESCE (other_renewables_twh,0)
AS tot_renewable_production_twh
FROM modern_renewable_production
WHERE "Year" = (SELECT MAX("Year") FROM modern_renewable_energy_consumption)
AND "Entity" <> 'World'
ORDER BY tot_renewable_production_twh DESC;
```

Using a similar query we can return the highest renewable energy producers in the same year. To have an accurate comparison, we used the latest year available in the consumption table, instead of the latest year available for the production table (which is 2022, but only includes data from 27 countries). We can immediately notice how China uses exactly the same amount of renewable energy as it produces, however we can run another query to quickly compare all other countries as well.

Entity	Year	tot_renewable_p...
Filter...	Filter...	Filter...
China	2021	2452.53158
United States	2021	861.5799999999999
Brazil	2021	508.67
Canada	2021	426.71000000000004
India	2021	332.203618
Germany	2021	230.8
Russia	2021	221.62
Japan	2021	216.73000000000002
Norway	2021	150.46999997999998
Spain	2021	125.74999999999999
France	2021	122.35999999999999
United Kingdom	2021	122.19999999999999
Turkey	2021	117.04
Italy	2021	116.32999999999998
Sweden	2021	115.74
Vietnam	2021	104.2180361
Mexico	2021	80.64999999999999
Australia	2021	71.98
Venezuela	2021	61.1
Colombia	2021	59.59
Indonesia	2021	56.22
Austria	2021	52.75
South Korea	2021	45.75

DATA ANALYSIS

```sql
-- We can now compare the results by using the JOIN function to have the two results close to each other.
-- We then create the renewables_surplus column, which is the difference between the total production and total consumption.
-- The countries which have a positive surplus use less renewable energy than what they produce, whereas those with a negative surplus
-- import part of their renewable energy consumption.

SELECT c."Entity",c."Year",
COALESCE(c.geo_twh ,0) + COALESCE(c.solar_twh,0) + COALESCE(c.wind_twh, 0) + COALESCE (c.hydro_twh,0)
  AS tot_consumption,
COALESCE(p.wind_prod_twh,0) + COALESCE(p.hydro_prod_twh,0) + COALESCE (p.solar_prod_twh,0) + COALESCE (p.other_renewables_twh,0)
  AS tot_production,
(COALESCE(p.wind_prod_twh,0) + COALESCE(p.hydro_prod_twh,0) + COALESCE (p.solar_prod_twh,0) + COALESCE (p.other_renewables_twh,0)) -
(COALESCE(c.geo_twh ,0) + COALESCE(c.solar_twh,0) + COALESCE(c.wind_twh, 0) + COALESCE (c.hydro_twh,0))
  AS renewables_surplus
FROM modern_renewable_energy_consumption c
JOIN modern_renewable_production p
ON c."Entity" = p."Entity"
  AND c."Code" = p."Code"
  AND c."Year" = p."Year"
WHERE c."Year"= (SELECT MAX ("Year") from modern_renewable_energy_consumption)
ORDER BY renewables_surplus ASC
```

With the next query, we join the two results (tot_consumption, tot_production) together to compare them side-by-side, and also generate a new column called renewables_surplus which counts the difference between the total production and total consumption. If this column has a positive number, it means the country doesn't use all the renewable energy it produces; if it has a negative number, then the country imports part of the renewable energy it uses.

DATA ANALYSIS

The resulting table shows how the United States, more than any other country in the World, goes out of its way to use renewable energy, having imported more than 20 TWh in all of 2021. Most other countries, on the other hand, tend to consume more or less the same amount of renewable energy as they consume.

Entity	Year	tot_consumption	tot_production	renewables_surp...
Filter...	Filter...	Filter...	Filter...	Filter...
United States	2021	882.1402049999999	861.5799999999999	-20.560204999999996
Germany	2021	236.70000000000002	230.8	-5.900000000000006
Australia	2021	77.329906	71.98	-5.34990599999999
Norway	2021	155.30786815	150.46999997999998	-4.837868170000007
Canada	2021	430.8209594	426.71000000000004	-4.110959399999956
Colombia	2021	63.056186675999996	59.59	-3.4661866759999924
Austria	2021	56.12081810000001	52.75	-3.370818100000008
Sri Lanka	2021	8.65986148	6.17	-2.48986148
Chile	2021	45.064792499999996	42.9337115	-2.1310809999999947
Egypt	2021	25.0944818	23.03	-2.0644817999999994
Malaysia	2021	35.497325200000006	33.9277085	-1.5696167000000045
Turkey	2021	118.39397600000001	117.04	-1.353976000000003
Switzerland	2021	41.43387792	40.11	-1.323877920000001
South Africa	2021	17.9194108	16.88	-1.0394108000000024
New Zealand	2021	36.0847451	35.05	-1.034745100000002
Israel	2021	5.67384672	4.96	-0.7138467200000003
Uzbekistan	2021	5.539050700000001	5.01	-0.5290507000000009
North Macedonia	2021	1.6340700000000001	1.1700000000000002	-0.46407
Estonia	2021	3.141190653	2.87	-0.2711906530000001
Venezuela	2021	61.318909672	61.1	-0.21890967199999523
Oman	2021	0.36474491000000003	0.16	-0.20474491000000003
Azerbaijan	2021	1.5205000000000002	1.4200000000000002	-0.10050000000000003
Qatar	2021	0.123208	0.04	-0.083208

DATA ANALYSIS

```sql
-- Now, we focus on the individual percentages from the tables we have available, and calculate the total
-- percentage for each renewable energy type as tot_percentage, then classify it as 'High' if it's more than 40%,
-- 'Moderate' if it's between 20% and 40%, and "Low" if it's less than 20%, and only take into account the last 3 years
-- available in the dataset.

SELECT w."Entity", w."Year", w.wind_percentage, s.solar_percentage, h.hydro_percentage,
w.wind_percentage + s.solar_percentage + h.hydro_percentage AS tot_percentage,
CASE
  WHEN w.wind_percentage + s.solar_percentage + h.hydro_percentage >= 40 THEN 'High'
  WHEN w.wind_percentage + s.solar_percentage + h.hydro_percentage >= 20 THEN 'Moderate'
  ELSE 'Low'
  END AS percentage_category
FROM wind_share_energy w
JOIN solar_share_energy s
  ON s."Entity" = w."Entity"
  AND s."Year" = w."Year"
JOIN hydro_share_energy h
  ON h."Entity" = w."Entity"
  AND h."Year" = w."Year"
WHERE w."Year" > '2019'
ORDER BY tot_percentage DESC;
```

We now look at the individual renewable energy types we have available in percentage points (solar, hydro, wind) and calculate the total percentage of these 3 renewables by summing the individual percentages in a column called “tot_percentage”. We then classify the amount in this column as “High” if it’s higher than 40%, “Moderate” if it’s between 20%and 40%, and “Low” if it’s less than 20%, in a new column called “percentage_category”.

DATA ANALYSIS

The resulting table confirms what we found with the earlier queries, and shows how Northern European countries (in particular Norway, Iceland and Sweden) have been in the lead for the past few years in renewable energy usage. In particular, this comes from their hydroelectric energy production, which is the leading renewable energy type for all 3 countries.

Entity	Year	wind_perce...	solar_pe...	hydro_p...	tot_percentage	percen...
Filter...	Filter	Filter...	Filter...	Filter...	Filter...	Filter...
Norway	2021	5.4197383	0.090746194	65.90561	71.41609449399999	High
Norway	2020	4.6556787	0.065970756	66.09646	70.81810945599999	High
Iceland	2021	0.027261456	0	61.955643	61.982904456	High
Iceland	2020	0.03064206	0	60.50321	60.53385206	High
Sweden	2020	11.692899	0.43966255	30.645115	42.77767655	High
Sweden	2021	11.262739	0.5976493	29.473484	41.333872299999996	High
Brazil	2020	4.4946084	0.84678334	31.22798	36.56937174	Moderate
Ecuador	2020	0.11104507	0.054384667	35.04654	35.211969737	Moderate
Switzerland	2021	0.12684327	2.6370564	31.890774	34.65467367	Moderate
Switzerland	2020	0.123253234	2.209208	31.844797	34.177258234	Moderate
Brazil	2021	5.4190493	1.2558651	27.199343	33.8742574	Moderate
Austria	2020	4.455559	1.3402594	27.552656	33.3484744	Moderate
Austria	2021	4.2896366	1.3500032	27.224585	32.8642248	Moderate
Ecuador	2021	0.07663507	0.045565866	31.606415	31.728615935999997	Moderate
New Zealand	2021	2.9579222	0.22934508	27.11879	30.30605728	Moderate
Peru	2020	1.6505098	0.7727318	27.75659	30.1798316	Moderate
Venezuela	2020	0.041011292	0.0015165611	30.09027	30.1327978531	Moderate
Colombia	2021	0.02931148	0.15540227	29.314417	29.49913075	Moderate
New Zealand	2020	2.530249	0.17590557	26.642075	29.348229569999997	Moderate
Canada	2020	2.4379308	0.29278252	26.439915	29.17062832	Moderate
Canada	2021	2.3737078	0.3489035	25.743002	28.4656133	Moderate
Venezuela	2021	0.03862083	0.0014274933	28.387844	28.4278923233	Moderate
Colombia	2020	0.0055207815	0.10406679	27.182451	27.2920385715	Moderate

DATA ANALYSIS

With this query, we find out how many countries classify as ‘High’, ‘Medium’, or ‘Low’ for their renewable energy usage from 2015 onwards.

From the results, while we can see that the high_count basically remains stable, we can easily notice how in the last few years more and more countries have become aware of their energy mix. In particular, between 2020 and 2021, 9 countries increased their renewable energy usage to more than 20%.

```sql
-- We now use this query to find the variation in the number of countries whose percentage classifies as 'High',
-- 'Moderate', and 'Low' using the COUNT() and CASE () functions. To find relevant results we expand the query to
-- include results from 2015 onwards.

SELECT w."Year",
  COUNT(CASE WHEN w.wind_percentage + s.solar_percentage + h.hydro_percentage >= 40 THEN 1 END) AS high_count,
  COUNT(CASE WHEN w.wind_percentage + s.solar_percentage + h.hydro_percentage >= 20
  AND w.wind_percentage + s.solar_percentage + h.hydro_percentage < 40 THEN 1 END) AS moderate_count,
  COUNT(CASE WHEN w.wind_percentage + s.solar_percentage + h.hydro_percentage < 20 THEN 1 END) AS low_count
FROM wind_share_energy w
JOIN solar_share_energy s
  ON s."Entity" = w."Entity"
  AND s."Year" = w."Year"
JOIN hydro_share_energy h
  ON h."Entity" = w."Entity"
  AND h."Year" = w."Year"
WHERE w."Year" > '2015'
GROUP BY w."Year"
ORDER BY w."Year";
```

Year	high_count	moderate_count	low_count
Filter...	Filter...	Filter...	Filter...
2016	2	11	67
2017	2	13	65
2018	2	13	65
2019	2	13	65
2020	3	14	63
2021	3	16	54

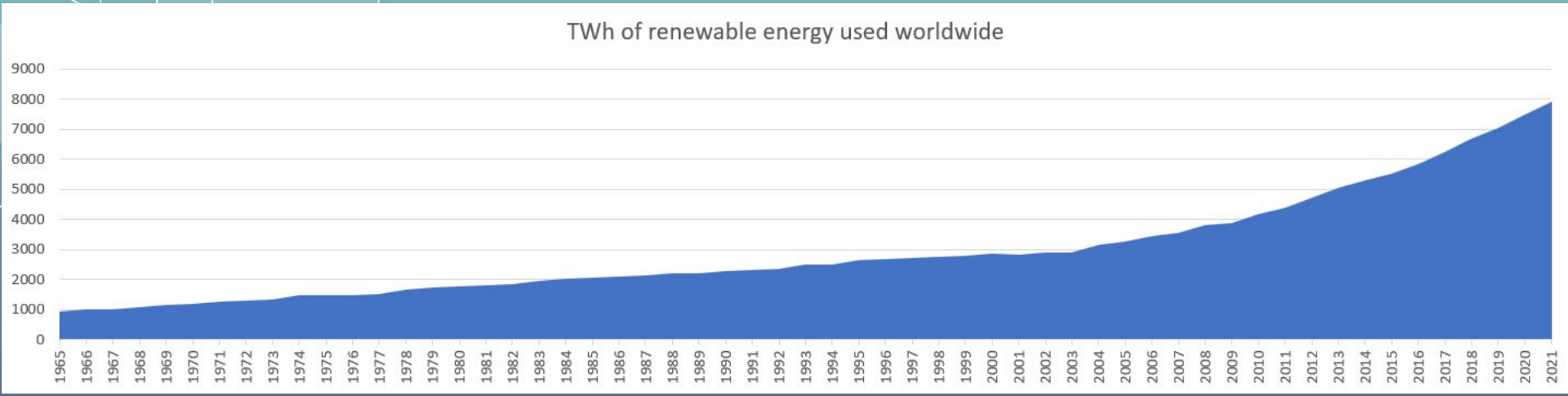
DATA ANALYSIS

```sql
-- Lastly, we will use this query to find the total amount of renewable energy consumed worldwide to see
-- its variation from 1965 to 2021. We use the CAST() function to convert the result of the SUM() function
-- from double precision to decimal, then we round it to the first 2 decimal numbers using the ROUND() function.

SELECT "Entity", "Year",
  ROUND(CAST(SUM(geo_twh + solar_twh + wind_twh + hydro_twh) AS decimal),2) AS renewables_twh
FROM modern_renewable_energy_consumption
WHERE "Entity" = 'World'
GROUP BY "Entity","Year"
ORDER BY "Year";
```

Entity	Year	renewables_twh
Filter...	Filter...	Filter...
World	1965	941.18
World	1966	1003.62
World	1967	1025.73
World	1968	1081.41
World	1969	1145.00
World	1970	1200.35
World	1971	1255.13
World	1972	1314.52
World	1973	1335.11
World	1974	1465.52
World	1975	1483.70
World	1976	1481.93
World	1977	1532.56
World	1978	1657.68
World	1979	1741.69
World	1980	1781.38
World	1981	1822.40
World	1982	1862.64
World	1983	1946.16
World	1984	2017.77
World	1985	2058.93
World	1986	2092.66
World	1987	2125.94
World	1988	2193.89
World	1989	2195.24
World	1990	2279.96
World	1991	2335.88

With the last query, we get a table with the total amount of renewable energy consumed worldwide from 1965 to 2021, using the CAST() and ROUND() function to prepare the data to be exported into a graph.

DATA ANALYSIS



From the graph, as per the previous one, we can easily notice how the renewable energy consumption worldwide remained approximately low until the early 2000s, then started spiking up after 2007. In fact, it took **19 years** to double from 1000 TWh to 2000 TWh, **24 years** to double from 2000 TWh to 4000 TWh and just **14 years**(!) to double from 4000 TWh to 8000 TWh.

CONCLUSION

From this analysis we found how countries are becoming more and more aware of their renewable energy production and consumption, and how in the last years they started making new efforts to try and use more sustainable energy sources. And while these efforts are clearly tangible from this analysis, in the real world, our planet still needs a lot more in order to recover from decades of abuse. It's clear how the many differences between countries make this process a lot more complicated than it sounds, however it's auspicable for the future of the younger generations that this effort continues to be made and keeps increasing over time.

THANKS

This presentation was created by Dario Giordano as the final project for the SQL course at start2impact University.

Dario Giordano