Dataset Covid_19_geographic_distribution_worldwide

Step 1: Exploring the dataset

1. What columns are there and what is the data type?

Field name	Type
date	DATE
day	INTEGER
month	INTEGER
year	INTEGER
daily_confirmed_cases	INTEGER
daily_deaths	INTEGER
confirmed_cases	INTEGER
deaths	INTEGER
countries_and_territories	STRING
geo_id	STRING
country_territory_code	STRING
pop_data_2019	INTEGER

2. How does the table look like?

```
SELECT
    *
FROM
 `bigquery-public-data.covid19_ecdc.covid_19_geographic_distribution_worldwide`
```

3. Get a list of all countries available in the dataset

```
SELECT
  DISTINCT countries_and_territories
FROM
  `bigquery-public-data.covid19_ecdc.covid_19_geographic_distribution_worldwide`
ORDER BY
  countries_and_territories DESC;
```

→ *There are 210 different countries.*

Step 2. Business Questions

1. What is the total number of cases of COVID-19 today globally?

```
SELECT
  SUM(confirmed_cases)
FROM
  `bigquery-public-data.covid19_ecdc.covid_19_geographic_distribution_worldwide`
WHERE
  date = '2020-09-03';
```

2. What are the top countries with the highest number of cases at the moment?

```
SELECT
 countries_and_territories,
 confirmed_cases
FROM
  `bigquery-public-data.covid19_ecdc.covid_19_geographic_distribution_worldwide`
WHERE
  date = '2020-09-03'
ORDER BY
  confirmed_cases DESC
LIMIT
  20;
```

3. What countries are most effected with covid-19 compared to their total population? (%)

```
SELECT
  countries_and_territories,
  pop_data_2019,
  confirmed_cases,
  confirmed_cases / pop_data_2019 * 100 AS cases_as_percentage_of_population
FROM
  `bigquery-public-data.covid19_ecdc.covid_19_geographic_distribution_worldwide`
WHERE
  date = '2020-09-03'
ORDER BY
    cases_as_percentage_of_population DESC
LIMIT
  20;
```

4. which countries grew the most in number of cases last month?

```
SELECT
  countries_and_territories,
  SUM(daily_confirmed_cases) AS total_growth,
FROM
  `bigquery-public-data.covid19_ecdc.covid_19_geographic_distribution_worldwide`
WHERE
  year = 2020
  AND month = 8
GROUP BY
  countries_and_territories
ORDER BY
  total_growth DESC
LIMIT
  20;
```

5. Are the daily cases in the Netherlands in the May 2020 higher or lower than its monthly average?

```
SELECT
  date,
  daily_confirmed_cases,
  (SELECT
    AVG(daily_confirmed_cases)
  FROM
    `bigquery-public-data.covid19_ecdc.covid_19_geographic_distribution_worldwide`
  WHERE
    countries_and_territories = 'Netherlands'
    AND year = 2020
    AND month = 5) AS average_daily_cases
FROM
  `bigquery-public-data.covid19_ecdc.covid_19_geographic_distribution_worldwide`
WHERE
  countries_and_territories = 'Netherlands'
  AND year = 2020
  AND month = 5
ORDER BY
  date;
```