

Project “Population” in MySQL

By Ravan Mustafayev, August 2025.

This is a small project that I completed in a few hours. I did a small amount of data cleaning and EDA and answered 5 short questions. This project helped me strengthen my knowledge of MySQL. You can check out the raw and cleaned datasets on my GitHub account. I will be able to use the results of this project in my future work, such as data visualization, building dashboards, and similar projects.

The Start

The dataset is about populations in different regions.

Columns:

1. Country name
2. Year
3. Population
4. Population of children under the age of 1
5. Population of children under the age of 5
6. Population of children under the age of 15
7. Population under the age of 25
8. Population aged 15 to 64 years
9. Population older than 15 years
10. Population older than 18 years
11. Population at age 1
12. Population aged 1 to 4 years
13. Population aged 5 to 9 years
14. Population aged 10 to 14 years
15. Population aged 15 to 19 years
16. Population aged 20 to 29 years
17. Population aged 30 to 39 years
18. Population aged 40 to 49 years
19. Population aged 50 to 59 years
20. Population aged 60 to 69 years
21. Population aged 70 to 79 years
22. Population aged 80 to 89 years
23. Population aged 90 to 99 years
24. Population older than 100 years

Raw Dataset contains **17,952 rows**.

Part 1-4 - data cleaning, EDA.

Part 5-9 - some questions

Questions that are answered:

1. What is the population of people aged 90 and above 90 for each country in the latest year?
2. Which countries had the highest population growth in the last year?
3. Which single country had the highest population decline in the last year?
4. What are the top 10 countries with the highest population growth in the last 10 years?
5. Which countries had the highest percentage growth since the first recorded year?

Part 1–4: Data Cleaning, EDA

1. First, we created the population schema.

2. Then we imported the raw dataset from a CSV file into the schema.

- Right-click schema → Import Data Wizard → choose file → import.
- Dataset contains **17,952 rows**.

Initial Analysis:

3.1 What years does the dataset cover? → **1950–2021**

```
SELECT MIN(Year), MAX(Year)
FROM pop_and_demog;
```

3.2 We created a second table (pop_and_demog2) to work on, so the raw dataset remains unchanged.

```
CREATE TABLE pop_and_demog2 LIKE pop_and_demog;
INSERT INTO pop_and_demog2
SELECT * FROM pop_and_demog;
```

4. Some rows were not actual country names (e.g., Asia (UN), High developed countries).

- We added a new column **record_type** to classify them.
- Values: Continent, Category, Country.

```
select distinct `Country name`
from pop_and_demog2
group by `Country name`
order by `Country name` asc;
```

```
alter table pop_and_demog2
add column record_type Varchar(100);
```

```
update pop_and_demog2
set record_type = 'Continent'
where `Country name` like '%(UN)%';
```

– changing other rows too. Total 960 rows

```
UPDATE pop_and_demog2
SET record_type = 'Category'
WHERE `Country name` IN (
'High-income countries',
'Land-locked developing countries (LLDC)',
'Least developed countries',
'Less developed regions',
'Less developed regions, excluding China',
'Less developed regions, excluding least developed countries',
'Low-income countries',
'Lower-middle-income countries',
'More developed regions',
'Small island developing states (SIDS)',
'Upper-middle-income countries',
'World'
);
```

-- the rest are countries. Total 16992 rows

```
select record_type, `Country name`
from pop_and_demog2
where record_type is null;
```

```
update pop_and_demog2
set record_type = 'Country'
where record_type is null;
```

Part 5-9: answer some questions

5. What is the population of people aged 90 and above 90 for each country in the latest year?

```
select `Country name`, `Year`, `Population aged 90 to 99 years`
as `Between 90 and 99`,
```

```

`Population older than 100 years` as `Older than 100`,
`Population aged 90 to 99 years` + `Population older than 100 years`
as 'Everybody above 90'
from pop_and_demog2
where record_type like 'Country' and `Year` = 2021
order by `Country name` asc;

```

6. Which countries have the highest population growth in the last year?

```

select `Country name`,
population_2020,
population_2021,
population_2021 - population_2020 as 'growth'
from (
    select
    `Country name`,
    (
        select p1.population
        from pop_and_demog2 p1
        where p1.`Country name` = p.`Country name`
        and `Year` = 2020
    ) as population_2020,

    (
        select p2.population
        from pop_and_demog2 p2
        where p2.`Country name` = p.`Country name`
        and `Year` = 2021
    ) as population_2021
    from pop_and_demog2 p
    where record_type like 'Country' and `Year` = 2021
) as s;

```

-----SIMPLIER CODE-----

```

SELECT
    `Country name`,
    SUM(CASE WHEN `Year` = 2020 THEN population END) AS population_2020,
    SUM(CASE WHEN `Year` = 2021 THEN population END) AS population_2021,
    (SUM(CASE WHEN `Year` = 2021 THEN population END)
    - SUM(CASE WHEN `Year` = 2020 THEN population END)) AS growth,
    (SUM(CASE WHEN `Year` = 2021 THEN population END)

```

- SUM(CASE WHEN `Year` = 2020 THEN population END))
 / SUM(CASE WHEN `Year` = 2020 THEN population END) as groth_percent
 FROM pop_and_demog27. which single country has the highest population decline in the last year?

using CTEs

```
with decline as ( SELECT
  `Country name`,
  SUM(CASE WHEN `Year` = 2020 THEN population END) AS population_2020,
  SUM(CASE WHEN `Year` = 2021 THEN population END) AS population_2021,
  (SUM(CASE WHEN `Year` = 2021 THEN population END)
  - SUM(CASE WHEN `Year` = 2020 THEN population END)) AS growth,
  (SUM(CASE WHEN `Year` = 2021 THEN population END)
  - SUM(CASE WHEN `Year` = 2020 THEN population END))
  / SUM(CASE WHEN `Year` = 2020 THEN population END) as growth_percent
FROM pop_and_demog2
WHERE record_type = 'Country' AND `Year` IN (2020, 2021)
GROUP BY `Country name`
ORDER BY `Country name`

)
select growth, growth_percent, `Country name`,
population_2020, population_2021
from decline
group by `Country name`
order by growth asc
limit 1;
```

Output >>> Japan

8. what are the top 10 countries with the highest population growth in the last 10 years?

```
with growths as ( SELECT
  `Country name`,
  SUM(CASE WHEN `Year` = 2011 THEN population END) AS population_2011,
  SUM(CASE WHEN `Year` = 2021 THEN population END) AS population_2021,
  (SUM(CASE WHEN `Year` = 2021 THEN population END)
  - SUM(CASE WHEN `Year` = 2011 THEN population END)) AS growth,
  (SUM(CASE WHEN `Year` = 2021 THEN population END)
  - SUM(CASE WHEN `Year` = 2011 THEN population END))
  / SUM(CASE WHEN `Year` = 2011 THEN population END) as growth_percent
FROM pop_and_demog2
```

```
WHERE record_type = 'Country'
GROUP BY `Country name`
ORDER BY `Country name`
```

```
)
select growth, growth_percent, `Country name`,
population_2020, population_2021
from decline
group by `Country name`
order by growth desc
limit 10;
```

9. Which countries have the highest percentage growth since the first year recorded?

```
with most_growth_percent as ( SELECT
  `Country name`,
  SUM(CASE WHEN `Year` = 1950 THEN population END) AS population_1950,
  SUM(CASE WHEN `Year` = 2021 THEN population END) AS population_2021,
  (SUM(CASE WHEN `Year` = 2021 THEN population END)
  - SUM(CASE WHEN `Year` = 1950 THEN population END)) AS growth,
  (SUM(CASE WHEN `Year` = 2021 THEN population END)
  - SUM(CASE WHEN `Year` = 1950 THEN population END))
  / SUM(CASE WHEN `Year` = 1950 THEN population END) *100 as growth_percent
FROM pop_and_demog2
WHERE record_type = 'Country'
GROUP BY `Country name`
ORDER BY `Country name`

)
select growth, growth_percent, `Country name`,
population_1950, population_2021
from most_growth_percent
group by `Country name`
order by growth_percent desc
limit 10;
```

```
WHERE record_type = 'Country' AND `Year` IN (2020, 2021)
GROUP BY `Country name`
ORDER BY `Country name`;
```

The End