1. *Show the* ***total cumulative cases per day*** *on a* ***national level***

select

  date(N.date) date

  , N.new\_total\_confirmed\_cases

from `bigquery-public-data.covid19\_italy.national\_trends` N;

1. *How many* ***distinct provinces per region*** *are recorded? Show the* ***code, name, and abbreviation*** *per province as well as the* ***region name***

select

  distinct (P.province\_code)

  , P.province\_name

  , P.province\_abbreviation

  , P.name region\_name

from `bigquery-public-data.covid19\_italy.data\_by\_province` P;

1. *Show the* ***increase*** *in* ***current cases******per day*** *in* ***region code = 6.***

select

  date(R.date) date

  , R.region\_code

  , R.region\_name

  , R.new\_current\_confirmed\_cases

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  R.region\_code='6';

1. *Show the* ***positivity rate*** *and* ***fatality rate*** *for each day, on a national level. Take note that* ***positivity rate = total cases/total performed****, while* ***fatality rate = total deaths/total cases***

select

  date(N.date) date

  , N.country

  , round(N.total\_confirmed\_cases/N.tests\_performed,5) positivity\_rate

  , round(N.deaths/N.total\_confirmed\_cases,5) fatality\_rate

from `bigquery-public-data.covid19\_italy.national\_trends` N;

1. *Which* ***region and days***  *recorded more than 10000 in* ***current hospitalized patients?*** *Show the split between* ***hospitalized with symptoms and hospitalized in intensive care***

select

  date(R.date) date

  , R.region\_code

  , R.region\_name

  , R.hospitalized\_patients\_symptoms

  , R.hospitalized\_patients\_intensive\_care

  , R.total\_hospitalized\_patients

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  R.total\_hospitalized\_patients>10000;

1. *What is the* ***hospitalization rate*** *on a national level, for* ***December 25, 2020****? Note the ff: Hospitalization rate = total current hospitalized/total current cases. To filter by date, you have to convert the* ***date*** *column into a date. It’s currently a* ***timestamp****. To do that, just add the function DATE(). For ex: Date(date) will return just the date (without the time) of the date column*

select

  date(N.date) date

  , N.country

  , round(N.total\_hospitalized\_patients/N.total\_current\_confirmed\_cases,5) hospitalization\_rate

from `bigquery-public-data.covid19\_italy.national\_trends` N

where

  date(N.date)='2020-12-25';

1. *Get the daily positivity rate for each region. Show data only for October 20, 2020 – October 25, 2020 and order results by date then by region name.   
   positivity rate = total cases/ total tests*

select

  date(R.date) date

  , R.region\_code

  , R.region\_name

  , round(R.total\_confirmed\_cases/R.tests\_performed,5) positivity\_rate

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  date(R.date) between ('2020-10-20') and ('2020-10-25')

order by

  1 asc, 3 asc;

1. *From your answer in 1 (i.e. using the same columns and conditions), which region and date registered the highest positivity rate? Adjust your query accordingly*

select

  date(R.date) date

  , R.region\_code

  , R.region\_name

  , round(R.total\_confirmed\_cases/R.tests\_performed,5) positivity\_rate

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  date(R.date) between ('2020-10-20') and ('2020-10-25')

order by

  4 desc

limit

  1;

1. *Show the total confirmed cases for each province in Veneto (5), for Nov 30, 2020, Dec 31, 2020 and Jan 31, 2021 only*

select

  date(P.date) date

  , P.province\_code

  , P.province\_name

  , P.name

  , P.confirmed\_cases

from `bigquery-public-data.covid19\_italy.data\_by\_province` P

where

  P.region\_code='5'

  and (date(P.date) in ('2020-11-30', '2020-12-31', '2021-01-31'))

order by

  1;

1. *Which day had the highest positivity rate in the month of December 2020, for* Toscana (9)

select

  date(R.date) date

  , R.region\_name

  , R.region\_code

  , round(R.total\_confirmed\_cases/R.tests\_performed,5) positivity\_rate

from `bigquery-public-data.covid19\_italy.data\_by\_region`R

where

  R.region\_code='9'

  and (date(R.date) between ('2020-12-01') and ('2020-12-31'))

order by

  4 desc

limit

  1;

1. *Which days registered the highest increase of current confirmed cases Consider June 2020, October 2020, December 2020, and March 2021, in Sicilia (19) and Basilicata (17) only.*

select

  date(R.date) date

  , R.region\_code

  , R.region\_name

  , R.new\_current\_confirmed\_cases

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  R.region\_code in ('19', '17')

  and (date(R.date) between ('2020-06-01') and ('2020-06-30') --June

  or date(R.date) between ('2020-10-01') and ('2020-10-30') --October

  or date(R.date) between ('2020-12-01') and ('2020-12-31') --December

  or date(R.date) between ('2021-03-01') and ('2021-03-31'))  --March

order by

  4 desc;

1. *Which days and regions had a negative increase in current confirmed cases? Consider only region names that start with the letter ‘P’, and January 01 – October 25 for both years 2020 and 2021 only. Sort output by date.*

select

  date(R.date) date

  , R.region\_code

  , R.region\_name

  , R.new\_current\_confirmed\_cases

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  R.new\_current\_confirmed\_cases<0

  and (R.region\_name like 'P%'

  and (date(R.date) between ('2020-01-01') and ('2020-10-25') --2020

  or date(R.date) between ('2021-01-01') and ('2021-10-25'))) --2021

order by

  1 asc;

1. *From your answer in the previous question (i.e. using the same columns and conditions), which day and region registered the lowest case count? Adjust your query accordingly.*

select

  date(R.date) date

  , R.region\_code

  , R.region\_name

  , R.new\_current\_confirmed\_cases

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  R.new\_current\_confirmed\_cases<0

  and (R.region\_name like 'P%'

  and (date(R.date) between ('2020-01-01') and ('2020-10-25') --2020

  or date(R.date) between ('2021-01-01') and ('2021-10-25'))) --2021

order by

  4 asc

limit

  1;

1. *Which regions have listed province names as ‘Provincia Autonoma’ (Autonomous province)? Show the distinct regions only, and order by region name ascending.*

select

  distinct (P.region\_code)

  , P.name

  , P.province\_name

from `bigquery-public-data.covid19\_italy.data\_by\_province` P

where

  P.province\_name like '%Provincia Autonoma%'

order by

  2 asc;

1. *Show all dates and provinces where there is a note recorded in the ‘note’ column. Consider only regions 20 and 8, and the second quarter for 2020 and 2021.  
   Note: to filter for if a column is null, you can use [column] IS NOT NULL in the WHERE clause.*

select

  date(P.date) date

  , P.province\_code

  , P.province\_name

  , P.note

from `bigquery-public-data.covid19\_italy.data\_by\_province` P

where

  P.note is not null

  and (P.region\_code in ('20','8')

  and date(P.date) between ('2020-04-01') and ('2020-06-30') -- 2nd quarter 2020

  or date(P.date) between ('2021-04-01') and ('2021-06-30')); -- 2nd quarter 2021

1. *Which dates and regions registered more than 2000 current cases in the months of July 2021 and August 2021? Sort results by hospitalization rate (total hospitalized/total current cases).*

select

  date(R.date) date

  , R.region\_code

  , R.region\_name

  , round(R.total\_hospitalized\_patients/R.total\_current\_confirmed\_cases,5) hosp\_rate

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  R.total\_current\_confirmed\_cases>2000

  and (date(R.date) between ('2021-07-01') and ('2021-08-31'))

order by

  4 desc;

1. *Which days have the highest ratio between home confinement and total hospitalized patients? Show the recovery rate (total recovered/total confirmed cases). Consider only the 15th of every month.   
   Note: to get a certain ‘part’ of a date or timestamp, you can use EXTRACT([part] FROM [date]) (for ex: to get the day of the month, use EXTRACT(DAY FROM date), where date is the column in the database)*

select

  date(N.date) date

  , round(N.home\_confinement\_cases/N.total\_hospitalized\_patients,5) ratio

  , round(N.recovered/N.total\_confirmed\_cases,5) recovery\_rate

from `bigquery-public-data.covid19\_italy.national\_trends` N

where

  extract(day from N.date)=15;

1. *What is the first and latest date recorded, in the national trends table? Show both in one query.*

select

  min(date(N.date)) first\_date

  , max(date(N.date)) last\_date

from `bigquery-public-data.covid19\_italy.national\_trends` N;

1. *What is the national average total current cases across all dates?*

select

  round(avg(N.total\_current\_confirmed\_cases),2) avg\_cases

from `bigquery-public-data.covid19\_italy.national\_trends` N;

1. *Show the average total current cases per month? Sort results per month, ascending   
   Note: you can aggregate on the whole month by using DATE\_TRUNC([date or timestamp column], PART). For ex: DATE\_TRUNC(date, MONTH) will return the first day of the month for each date in the column.*

select

  date\_trunc(date(N.date), month) month

  , round(avg(N.total\_current\_confirmed\_cases),2) avg\_cases

from `bigquery-public-data.covid19\_italy.national\_trends` N

group by

  1

order by

  1 asc;

1. *Based on the previous question, which months registered an average of more than 100000 current cases? Sort results by date ascending.*

select

  date\_trunc(date(N.date), month) month

  , round(avg(N.total\_current\_confirmed\_cases),2) avg\_cases

from `bigquery-public-data.covid19\_italy.national\_trends` N

group by

  1

having

  avg\_cases>100000

order by

  1 asc;

1. *Show the monthly increase of cases per regions per months of July 2021 to September 2021. Sort results by region name, then by date.*

select

  date\_trunc(date(R.date), month) month

  , R.region\_code

  , R.region\_name

  , round(sum(R.new\_total\_confirmed\_cases),2) total\_cases

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  date(R.date) between ('2021-07-01') and ('2021-09-30')

group by

  1,2,3

order by

  2,1;

1. *From the previous question, which regions and month registered an increase of more than 10000?*

select

  date\_trunc(date(R.date), month) month

  , R.region\_code

  , R.region\_name

  , round(sum(R.new\_total\_confirmed\_cases),2) total\_cases

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  date(R.date) between ('2021-07-01') and ('2021-09-30')

group by

  1,2,3

having

  total\_cases>10000

order by

  2,1;

1. *Which regions have an average fatality rate of less than 5%? Consider only days where total cases > 0, and sort results from highest fatality rate to lowest.*

select

  R.region\_code

  , R.region\_name

  , round(avg(R.deaths/R.total\_confirmed\_cases),5) fatality\_rate

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  R.total\_confirmed\_cases>0

group by

  1,2

having

  fatality\_rate<.05

order by

  3 desc;

1. *Which 3 dates registered the highest case increase, on a national level?*

select

  date(N.date) date

  , N.new\_total\_confirmed\_cases

from `bigquery-public-data.covid19\_italy.national\_trends` N

order by

  2 desc

limit

  3;

1. *Which month registered the highest average current case count, on a national level?*

select

  date\_trunc(date(N.date), month) month

  , round(avg(N.total\_current\_confirmed\_cases),2) avg\_current\_cases

from `bigquery-public-data.covid19\_italy.national\_trends` N

group by

  1

order by

  2 desc

limit

  1;

1. *Write the query that will show the count of unique provinces.*

select

  count(distinct (P.province\_code)) province\_count

from `bigquery-public-data.covid19\_italy.data\_by\_province` P

1. *How many days registered a case increase of more than 10000, on a national level?*

select

  count(date(N.date)) count\_days

from `bigquery-public-data.covid19\_italy.national\_trends` N

where

  N.total\_current\_confirmed\_cases>10000;

1. *Write the query that will show all unique regions. Order alphabetically.*

select

  distinct(R.region\_code)

  , R.region\_name

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

order by

  2 asc;

1. *Which dates and region(s) have the highest deaths to tests ratio? Consider only August 2020 and December 2020. Limit the output to show only 5 rows.*

select

  date(R.date) date

  , R.region\_code

  , R.region\_name

  , R.deaths/R.tests\_performed death\_to\_test\_ratio

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  date(R.date) between ('2020-08-01') and ('2020-08-31') --august

  or  (date(R.date) between ('2020-12-01') and ('2020-12-31')) --december

order by

  4 desc

limit

  5;

1. *What is the total case increase for each region, for December 2020?*

select

  R.region\_code

  , R.region\_name

  , sum(R.new\_total\_confirmed\_cases) total\_case\_increase

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  date(R.date) between ('2020-12-01') and ('2020-12-31') --december

group by

  1,2;

1. *What is the average number of new cases per month for regions 1,4,7, and 10? Show only months and regions where average new cases > 100*

select

  date\_trunc(date(R.date),month) month

  , R.region\_name

  , round(avg(R.new\_total\_confirmed\_cases),2) avg\_case\_increase

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  R.region\_code in ('1','4','7','10')

group by

  1,2

having

  avg\_case\_increase>100;

1. *Which region registered the highest average hospitalization rate, for 2021?*

select

  R.region\_name

  , round(avg(R.total\_hospitalized\_patients/R.total\_current\_confirmed\_cases),5) hospitalization\_rate

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  extract(year from R.date)=2021

group by

  1

order by

  2 desc

limit

  1;

1. *How many unique regions had a daily increase of more than 1000, on any day of a month? Consider only Feb 2021, May 2021, and August 2021. (truncate monthly, then get unique region count, with the specified filter).*

select

  date\_trunc(date(R.date), month) month

  , count(distinct (R.region\_name)) region\_count

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  R.new\_total\_confirmed\_cases>1000

  and (date(R.date) between ('2021-02-01') and ('2021-02-28') --feb

  or date(R.date) between ('2021-05-01') and ('2021-05-31') --may

  or date(R.date) between ('2021-08-01') and ('2021-08-31')) --aug

group by

  1;

1. *Combine the regional and national tables. Show both regional and national total confirmed cases, and rename columns as needed.*

select

  date(R.date) date

  , R.region\_code

  , R.region\_name

  , R.total\_confirmed\_cases region\_total

  , N.total\_confirmed\_cases national\_total

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

  inner join `bigquery-public-data.covid19\_italy.national\_trends` N

    on R.date=N.date;

1. *How much did region 5 contribute in total testing? (i.e. total regional testing / total national testing)? Show on a daily basis for August 2020.*

select

  date(R.date) date

  , R.region\_code

  , R.region\_name

  , round(R.tests\_performed/N.tests\_performed,5) test\_pct\_contribution

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

  left join `bigquery-public-data.covid19\_italy.national\_trends` N

    on R.date=N.date

where

  R.region\_code='5'

  and (date(R.date) between ('2020-08-01') and ('2020-08-31'));

1. *Show region 3’s contribution to total national cases, as of August 01, 2021.*

select

  date(R.date) date

  , R.region\_code

  , R.region\_name

  , round(R.total\_confirmed\_cases/N.total\_confirmed\_cases,5) pct\_contribution

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

  left join `bigquery-public-data.covid19\_italy.national\_trends` N

    on R.date=N.date

where

  R.region\_code='3'

  and (date(R.date)='2021-08-01');

1. *Which region had the highest contribution in total current national cases, as of October 31, 2021?*

select

  date(R.date) date

  , R.region\_code

  , R.region\_name

  , round(R.total\_current\_confirmed\_cases/N.total\_current\_confirmed\_cases,5) pct\_contribution

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

  left join `bigquery-public-data.covid19\_italy.national\_trends` N

    on R.date=N.date

where

  date(R.date)='2021-10-31'

order by

  4 desc

limit

  1;

1. *Show the contribution of each province to the total regional cases, for each day. Only include days where total regional confirmed cases > 0.*

select

  date(P.date) date

  , P.province\_code

  , P.province\_name

  , R.region\_code

  , R.region\_name

  , round(P.confirmed\_cases/R.total\_confirmed\_cases,5) pct\_contribution

from `bigquery-public-data.covid19\_italy.data\_by\_province` P

  left join `bigquery-public-data.covid19\_italy.data\_by\_region` R

    on P.date=R.date and P.region\_code=R.region\_code

where

  R.total\_confirmed\_cases>0;

1. *Which province had the highest contribution to total national cases, for November 01, 2021?*

select

  date(P.date) date

  , P.province\_code

  , P.province\_name

  , round(P.confirmed\_cases/N.total\_confirmed\_cases,5) pct\_contribution

from `bigquery-public-data.covid19\_italy.data\_by\_province` P

  left join `bigquery-public-data.covid19\_italy.national\_trends` N

    on P.date=N.date

where

  date(P.date)='2021-11-01'

order by

  4 desc

limit

  1;

1. *For October 10,2021, show the contribution of each province to regional and national total cases. Exclude any instances where there is no match in region.*

select

  date(P.date) date

  , P.province\_code

  , P.province\_name

, P.region\_code

, P.name region\_name

  , round(P.confirmed\_cases/R.total\_confirmed\_cases,5) regional\_pct\_cont

  , round(P.confirmed\_cases/N.total\_confirmed\_cases,5) national\_pct\_cont

from `bigquery-public-data.covid19\_italy.data\_by\_province` P

  left join `bigquery-public-data.covid19\_italy.data\_by\_region` R

    on P.date=R.date and P.region\_code=R.region\_code

  left join `bigquery-public-data.covid19\_italy.national\_trends` N

    on P.date=N.date

where

  date(P.date)='2021-10-10';

1. *Show the daily new current confirmed cases per region, and its corresponding level (0-500 = low, 501 – 1000 = Medium, >1000 = High; else negative). Consider only August 2021 to October 2021.*

select

  date(R.date) date

  , R.region\_code

  , R.region\_name

  , R.new\_current\_confirmed\_cases

  , case

      when R.new\_current\_confirmed\_cases between 0 and 500 then 'low'

      when R.new\_current\_confirmed\_cases between 501 and 1000 then 'medium'

      when R.new\_current\_confirmed\_cases>1000 then 'high'

      else 'negative'

    end level

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  date(R.date) between ('2021-08-01') and ('2021-10-31');

1. *What is the average positivity rate (total confirmed/tests performed) for each region and level (as defined in the previous question), per month? Consider only August 2021 to October 2021*

select

  date\_trunc(date(R.date),month) month

  , R.region\_code

  , R.region\_name

  , round(avg(R.total\_confirmed\_cases/R.tests\_performed),5) positivity\_rate

  , case

      when R.new\_current\_confirmed\_cases between 0 and 500 then 'low'

      when R.new\_current\_confirmed\_cases between 501 and 1000 then 'medium'

      when R.new\_current\_confirmed\_cases>1000 then 'high'

      else 'negative'

    end level

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  date(R.date) between ('2021-08-01') and ('2021-10-31')

group by

  1,2,3,5;

1. *Get the national positivity rate and hospitalization rate (current hospitalized/current cases) as of the most recent date available.*

select

  date(N.date) latest\_date

  , round(N.total\_confirmed\_cases/N.tests\_performed,5) positivity\_rate

  , round(N.total\_hospitalized\_patients/N.total\_current\_confirmed\_cases,5) hospitalization\_rate

from `bigquery-public-data.covid19\_italy.national\_trends` N

where

  date(N.date)=(select max(date(N.date))

  from `bigquery-public-data.covid19\_italy.national\_trends` N);

1. *Compare the daily sum of new current cases of all regions to the national totals, using a subquery.*

*Using left join only*

select

  date(R.date) date

  , sum(R.new\_current\_confirmed\_cases) region\_current\_cases

  , N.new\_current\_confirmed\_cases national\_current\_cases

  , (sum(R.new\_current\_confirmed\_cases) - N.new\_current\_confirmed\_cases) difference

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

  left join `bigquery-public-data.covid19\_italy.national\_trends` N

    on R.date=N.date

group by

  1,3;

1. *What percentage of the total historical case increase did the increases from October 2020 to December 2020 make up (i.e. sum of October 2020 to December 2020 increase/sum of all increase). Show for each region.*

select

  distinct R.region\_code

  , R.region\_name

  , round(sum(R.new\_total\_confirmed\_cases)/

  (select sum(R.new\_total\_confirmed\_cases)

  from `bigquery-public-data.covid19\_italy.data\_by\_region` R),5) percentage

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  date(R.date) between ('2020-10-01') and ('2020-12-31')

group by

  1,2;

1. *Let Sector 1 = regions 1,2,3,4; Sector 2 = regions 5,6,7,8,9,10, Sector 3 = regions 11,12,13, Sector 4 = all other regions. Show each sector’s average increase in cases per month, from Jan 2021 to Oct 2021*

select

  date\_trunc(date(R.date), month) month

  , case

        when R.region\_code in ('1','2','3','4') then 'Sector 1'

        when R.region\_code in ('5','6','7','8','9','10') then 'Sector 2'

        when R.region\_code in ('11','12','13') then 'Sector 3'

        else 'Sector 4'

    end sector

  , round(avg(R.new\_total\_confirmed\_cases),2) avg\_monthly\_increase

from `bigquery-public-data.covid19\_italy.data\_by\_region` R

where

  date(R.date) between ('2021-01-01') and ('2021-10-31')

group by

  1,2

order by

  2 asc;

1. *Get total case increase (not the cumulative count) per province, per month.*

WITH MonthlyCases AS (

    SELECT

        P.province\_code,

        P.province\_name,

        DATE\_TRUNC(date(P.date),month) AS report\_month,

        MAX(P.confirmed\_cases) AS month\_end\_cases

    FROM `bigquery-public-data.covid19\_italy.data\_by\_province` P

    GROUP BY 1, 2, 3

)

SELECT

    province\_code,

    province\_name,

    report\_month,

    month\_end\_cases,

    LAG(month\_end\_cases) OVER (

        PARTITION BY province\_code ORDER BY report\_month

    ) AS prev\_month\_cases,

    (month\_end\_cases - LAG(month\_end\_cases) OVER (

        PARTITION BY province\_code ORDER BY report\_month

    )) AS case\_increase

FROM MonthlyCases

ORDER BY 1, 2;

1. *In the current query, is there a way for us to show it per month per region? Like the results will only show one row for August, Sept, Oct per region (3 rows lang for Lazio, Sicilia, etc)?*
2. *Identify regions with a hospitalization rate exceeding the national average hospitalization rate. Note that the problem statement requires us to evaluate each region's hospitalization rate in comparison with the national average hospitalization rate.*