



# Data Analytics Bootcamp

Data Analytics Bootcamp

## Fundamental Queries I

- SELECT Column
- FROM Table
- WHERE Filter
- ORDER BY Arranging
- LIMIT View

## Syntax Guide

```
SELECT
  [ALIAS.column]
  , [ALIAS.column2]
  , ...
FROM
  [table ALIAS]
WHERE
  [column] [condition]
ORDER BY
  [column] [ASC/DESC]
LIMIT
  [number of rows]
```



# Data Analytics Bootcamp

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

#	Question/Task	Query/Answer
1	Which table do I get the data from?	<code>bigquery-public-data.covid19_italy.data_by_province</code>
2	Which columns do I need to select. Show any relevant calculations as well as intended output column name if applicable (i.e. column1 + column2 AS added_columns)	<code>Region_code, name, province_code, province_name</code>
3	Do I need to use any filter?	YES
4	If #3 is yes, what filters do I need?	<code>Province_name LIKE '%Provincia Autonoma%'</code>
5	Do I need to arrange my dataset? Which column? In ascending or descending order?	YES. ASC
6	Do I need to limit the results of my dataset? If so, to how many rows?	No
7	Show the query to get the data needed. You can type, copy paste, or paste an image.	<pre>SELECT DISTINCT   P.region_code   , P.name   , P.province_code   , P.province_name FROM   `bigquery-public- data.covid19_italy.data_by_province` P WHERE   P.province_name LIKE '%Provincia Autonoma%' ORDER BY   2 ASC</pre>



## Data Analytics Bootcamp

8 Show a screenshot of the output. No need to show everything, just a sample will do.

Row	region_code	name	province_code	province_name
1	13	Abruzzo	879	Fuori Regione / Provincia Auto...
2	17	Basilicata	880	Fuori Regione / Provincia Auto...
3	18	Calabria	882	Fuori Regione / Provincia Auto...
4	15	Campania	883	Fuori Regione / Provincia Auto...
5	8	Emilia-Romagna	884	Fuori Regione / Provincia Auto...
6	6	Friuli Venezia Giulia	885	Fuori Regione / Provincia Auto...
7	12	Lazio	886	Fuori Regione / Provincia Auto...



## Data Analytics Bootcamp

11. 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.

#	Question/Task	Query/Answer
1	Which table do I get the data from?	<code>bigquery-public-data.covid19_italy.data_by_province</code>
2	Which columns do I need to select. Show any relevant calculations as well as intended output column name if applicable (i.e. column1 + column2 AS added_columns)	<code>Date, region_code, name, province_code, province_name, confirmed_cases, note</code>
3	Do I need to use any filter?	YES
4	If #3 is yes, what filters do I need?	<code>(DATE(J.date) BETWEEN DATE('2021-04-01') AND DATE('2021-06-30')) OR DATE(J.date) BETWEEN DATE('2020-04-01') AND DATE('2020-06-30')) AND J.note IS NOT NULL AND J.region_code IN ('20','8')</code>
5	Do I need to arrange my dataset? Which column? In ascending or descending order?	No
6	Do I need to limit the results of my dataset? If so, to how many rows?	No
7	Show the query to get the data needed. You can type, copy paste, or paste an image.	<code>/*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.</code>



## Data Analytics Bootcamp

Note: to filter for if a column is null, you can use [column] IS NOT NULL in the WHERE clause.\*/

```
SELECT
  J.date
  ,J.region_code
  ,J.name AS region_name
  ,J.province_code
  ,J.province_name
  ,J.confirmed_cases
  ,J.note
FROM
  `bigquery-public-
  data.covid19_italy.data_by_province` J
WHERE
  (DATE(J.date) BETWEEN DATE('2021-04-01') AND
  DATE('2021-06-30'))
  OR DATE(J.date) BETWEEN DATE('2020-04-01') AND
  DATE('2020-06-30'))
  AND J.note IS NOT NULL
  AND J.region_code IN ('20','8')
```

8 Show a screenshot of the output. No need to show everything, just a sample will do.

Row	date	region_code	region_name	province_code	province_name	confirmed_cases	note
1	2021-05-14 17:00:00	20	Sardegna	993	Fuori Regione / Provincia Auton.	15	Marina isolati sulla nave
2	2021-05-20 17:00:00	20	Sardegna	993	Fuori Regione / Provincia Auton.	14	Sui 15 casi segnalati Fuori Regione si verifica il dato: 2 in regione di accesso ospedaliere, 11 nuovo positivo tra i quarantantati
3	2021-05-17 17:00:00	20	Sardegna	95	Oristano	5069	1 NUOVA POSITIVA DELLA



## Data Analytics Bootcamp

III. 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).

#	Question/Task	Query/Answer
1	Which table do I get the data from?	<code>bigquery-public-data.covid19_italy.data_by_region</code>
2	Which columns do I need to select. Show any relevant calculations as well as intended output column name if applicable (i.e. column1 + column2 AS added_columns)	<code>Date, region_code, region_name, total_hospitalization_patients, total_current_confirmed_case, total_hospitalization_patients, total,current_confirmed_cases</code>
3	Do I need to use any filter?	YES
4	If #3 is yes, what filters do I need?	<code>J.total_current_confirmed_cases &gt; 2000 AND DATE(J.date) BETWEEN DATE('2021-07-01') AND DATE('2021-08-31')</code>
5	Do I need to arrange my dataset? Which column? In ascending or descending order?	YES. <code>hospitalization_rate DESC</code>
6	Do I need to limit the results of my dataset? If so, to how many rows?	No
7	Show the query to get the data needed. You can type, copy paste, or paste an image.	<pre>/*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   J.date   ,J.region_code   ,J.region_name   ,J.total_hospitalized_patients   ,J.total_current_confirmed_cases</pre>



## Data Analytics Bootcamp

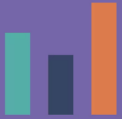
	<pre>,J.total_hospitalized_patients/J.total_current_confirmed_cases AS hospitalization_rate FROM `bigquery-public-data.covid19_italy.data_by_region` J WHERE J.total_current_confirmed_cases &gt; 2000 AND DATE(J.date) BETWEEN DATE('2021-07-01') AND DATE('2021-08-31') ORDER BY 6 DESC</pre>
8 Show a screenshot of the output. No need to show everything, just a sample will do.	



## Data Analytics Bootcamp

IV. 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 15<sup>th</sup> 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)

#	Question/Task	Query/Answer
1	Which table do I get the data from?	<code>bigquery-public-data.covid19_italy.national_trends</code>
2	Which columns do I need to select. Show any relevant calculations as well as intended output column name if applicable (i.e. column1 + column2 AS added_columns)	<code>date, home_confinement_cases, total_hospitalized_patients, home_confinement_cases, total_hospitalized_patients, recovered, total_confirmed_cases</code>
3	Do I need to use any filter?	YES
4	If #3 is yes, what filters do I need?	<code>EXTRACT(DAY FROM J.date) = 15</code>
5	Do I need to arrange my dataset? Which column? In ascending or descending order?	YES. <code>total_hospitalized_patients, home_confinement_case</code> . DESC
6	Do I need to limit the results of my dataset? If so, to how many rows?	No
7	Show the query to get the data needed. You can type, copy paste, or paste an image.	<pre>/*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,</pre>



# Data Analytics Bootcamp

```
use EXTRACT(DAY FROM date), where date is the
column in the database)*/
SELECT
  J.date
  ,J.home_confinement_cases
  ,J.total_hospitalized_patients

  ,J.home_confinement_cases/J.total_hospitalized
_patients AS home_confinement_ratio
  ,J.recovered/J.total_confirmed_cases AS
recovery_rate
FROM
`bigquery-public-
data.covid19_italy.national_trends` J
WHERE
  EXTRACT(DAY FROM J.date) = 15
ORDER BY
  3 DESC
```

8 Show a screenshot of the output. No need to show everything, just a sample will do.

Row	date	home_confinement_cases	total_hospitalized_patients	home_confinement_ratio	recovery_rate
1	2020-11-15 17:00:00 UTC	677021	35469	19.087682201358934	0.357063763386...
2	2020-04-15 17:00:00 UTC	74696	30772	2.4271507105510662	0.220640940404...
3	2020-12-15 17:00:00 UTC	632968	30345	20.85054529919261	0.610189588641...
4	2021-04-15 17:00:00 UTC	481019	29004	16.58457454144256	0.836399770422...
5	2021-03-15 17:00:00 UTC	991862	28495	17.612282856641517	0.804877207097...
6	2021-01-15 17:00:00 UTC	932705	23763	21.805233056026495	0.728168117450...
7	2021-02-15 17:00:00 UTC	517494	28004	18.32130760492919	0.915763405966...

Results per page: 50 1 - 39 of 3