



Data Analytics Bootcamp

Data Analytics Bootcamp 2023

Fundamental Queries I

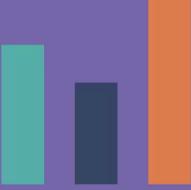
- SELECT Column
- FROM Table
- WHERE Filter

Syntax Guide

```
SELECT
  [ALIAS.column]
  , [ALIAS.column2]
  , ...
FROM
  [table ALIAS]
WHERE
  [ALIAS.column]
```

SAMPLE

```
SELECT
  NTO.date
  , N.deaths
  , N.total_confirmed_cases
FROM
  `bigquery-public-data.covid19_italy.national_trends` N
WHERE
  N.total_confirmed_cases > 100
```

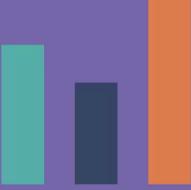


Data Analytics Bootcamp

[BATCH 2]

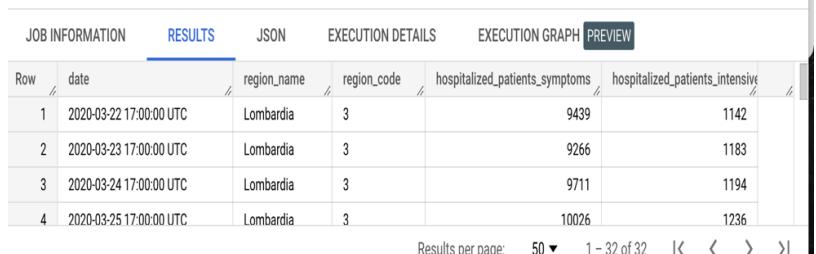
I. Show the **positivity rate** and **fatality rate** for each day, on a national level. Take note that **positivity rate** = **total cases/total performed**, while **fatality rate** = **total deaths/total cases**

#	Question/Task	Query/Answer																																										
1	Which table do I get the data from ?	bigquery-public-data.covid19_italy.national_trends																																										
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)	Data, total_confirmed_cases, deaths, test_performed,																																										
3	Do I need to use any filter?	No																																										
4	If #3 is yes, what filters do I need?	No																																										
5	Show the query to get the data needed. You can type, copy paste, or paste an image.	<pre>/* the positivity rate and fatality rate each day on a national level*/ SELECT JM. date , JM. tests_performed , JM. deaths AS total_death , JM. total_confirmed_cases/JM. tests_performed AS positivity_rate , JM. deaths/JM. total_confirmed_cases AS fatality_rate FROM bigquery-public-data.covid19_italy.national_trends JM;</pre>																																										
6	Show a screenshot of the output. No need to show everything, just a sample will do.	<p>JOBI INFORMATION RESULTS JSON EXECUTION DETAILS EXECUTION GRAPH PREVIEW</p> <table border="1"> <thead> <tr> <th>Row</th> <th>date</th> <th>tests_performed</th> <th>total_death</th> <th>positivity_rate</th> <th>fatality_rate</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>2022-08-20 17:00:00 UTC</td> <td>240836328</td> <td>174659</td> <td>0.089816175905...</td> <td>0.008074477192...</td> </tr> <tr> <td>2</td> <td>2022-11-25 17:00:00 UTC</td> <td>257255840</td> <td>181182</td> <td>0.094440056249...</td> <td>0.007457505284...</td> </tr> <tr> <td>3</td> <td>2021-11-19 17:00:00 UTC</td> <td>113437819</td> <td>133082</td> <td>0.043234620016...</td> <td>0.027134998667...</td> </tr> <tr> <td>4</td> <td>2021-11-20 17:00:00 UTC</td> <td>114012631</td> <td>133131</td> <td>0.043117862967...</td> <td>0.027081268214...</td> </tr> <tr> <td>5</td> <td>2020-10-28 17:00:00 UTC</td> <td>15152038</td> <td>37905</td> <td>0.038923212837...</td> <td>0.064271253344...</td> </tr> <tr> <td>6</td> <td>2021-09-22 17:00:00 UTC</td> <td>90166078</td> <td>130488</td> <td>0.051525508295...</td> <td>0.028086984241...</td> </tr> </tbody> </table> <p>Results per page: 50 ▾ 1 – 50 of 1173 < < > ></p>	Row	date	tests_performed	total_death	positivity_rate	fatality_rate	1	2022-08-20 17:00:00 UTC	240836328	174659	0.089816175905...	0.008074477192...	2	2022-11-25 17:00:00 UTC	257255840	181182	0.094440056249...	0.007457505284...	3	2021-11-19 17:00:00 UTC	113437819	133082	0.043234620016...	0.027134998667...	4	2021-11-20 17:00:00 UTC	114012631	133131	0.043117862967...	0.027081268214...	5	2020-10-28 17:00:00 UTC	15152038	37905	0.038923212837...	0.064271253344...	6	2021-09-22 17:00:00 UTC	90166078	130488	0.051525508295...	0.028086984241...
Row	date	tests_performed	total_death	positivity_rate	fatality_rate																																							
1	2022-08-20 17:00:00 UTC	240836328	174659	0.089816175905...	0.008074477192...																																							
2	2022-11-25 17:00:00 UTC	257255840	181182	0.094440056249...	0.007457505284...																																							
3	2021-11-19 17:00:00 UTC	113437819	133082	0.043234620016...	0.027134998667...																																							
4	2021-11-20 17:00:00 UTC	114012631	133131	0.043117862967...	0.027081268214...																																							
5	2020-10-28 17:00:00 UTC	15152038	37905	0.038923212837...	0.064271253344...																																							
6	2021-09-22 17:00:00 UTC	90166078	130488	0.051525508295...	0.028086984241...																																							



Data Analytics Bootcamp

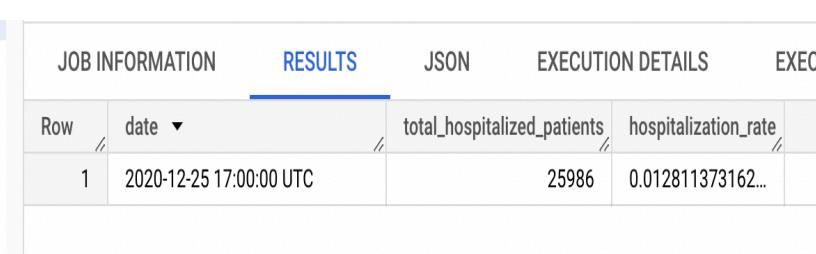
II. Which **region and days** recorded more than 10000 in **current hospitalized patients**? Show the split between **hospitalized with symptoms and hospitalized in intensive care**

#	Question/Task	Query/Answer
1	Which table do I get the data from ?	bigquery-public-data.covid19_italy.data_by_region.
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)	region_name, region_code, date, total_hospitalized_patient, hospitalized_patients_symptoms, hospitalized_patient_intensive_care
3	Do I need to use any filter?	Yes
4	If #3 is yes, what filters do I need?	< 10000 total_hospitalized_patients
5	Show the query to get the data needed. You can type, copy paste, or paste an image.	<pre>/* Region and days recorded more than 10000 in current, showing the split between hospitalized*/ SELECT RD. date , RD. region_name , RD. region_code , RD. hospitalized_patients_symptoms , RD. hospitalized_patients_intensive_care FROM bigquery-public-data.covid19_italy.data_by_region RD WHERE total_hospitalized_patients>10000</pre>
6	Show a screenshot of the output. No need to show everything, just a sample will do.	 <p>Results per page: 50 ▾ 1 - 32 of 32 < < > > </p>

Data Analytics Bootcamp

III. 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

#	Question/Task	Query/Answer								
1	Which table do I get the data from?	bigquery-public-data.covid19_italy.national_trends								
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)	date, total_hospitalized_patients, total_current_confirmed_cas								
3	Do I need to use any filter?	YES								
4	If #3 is yes, what filters do I need?	December 25, 2020								
5	Show the query to get the data needed. You can type, copy paste, or paste an image.	<pre>SELECT JA.date , JA.total_hospitalized_patients , JA.total_hospitalized_patients/JA.total_confirmed_cases AS hospitalization_rate FROM bigquery-public-data.covid19_italy.national_trends JA WHERE DATE(JA.date)='2020-12-25'</pre>								
6	Show a screenshot of the output. No need to show everything, just a sample will do.	 <table border="1"><thead><tr><th>Row</th><th>date</th><th>total_hospitalized_patients</th><th>hospitalization_rate</th></tr></thead><tbody><tr><td>1</td><td>2020-12-25 17:00:00 UTC</td><td>25986</td><td>0.012811373162..</td></tr></tbody></table>	Row	date	total_hospitalized_patients	hospitalization_rate	1	2020-12-25 17:00:00 UTC	25986	0.012811373162..
Row	date	total_hospitalized_patients	hospitalization_rate							
1	2020-12-25 17:00:00 UTC	25986	0.012811373162..							