

Data Analytics Bootcamp

Data Analytics Bootcamp – Cohort 9
2023 February 20

Guide

- SELECT Column
- FROM Table
- LEFT/INNER JOIN ON Combining
- WHERE Filter
- ORDER BY Arranging
- LIMIT View

Syntax Guide

SELECT

[ALIAS.column]
, [ALIAS.column2]
, ...

FROM

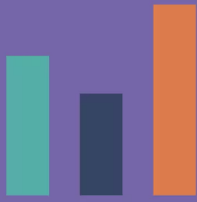
[left table ALIAS]

LEFT JOIN

[right table ALIAS]
ON [left table].[key] = [right table].[key]
AND ...

WHERE

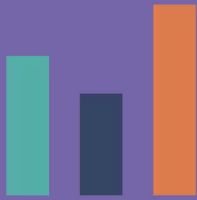
[column] [condition]



Data Analytics Bootcamp

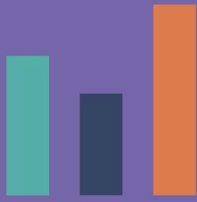
IV. Which region had the highest contribution in total current national cases, as of October 31, 2021?

#	Question/Task	Query/Answer
1	Which table do I get the data from ? (which will I use as my LEFT table?)	bigquery-public-data.covid19_italy.data_by_region
2	Which columns from my LEFT table do I need?	Date, region_name, total_confirmed_cases
3	Do I need to reference another table? If so, what will be my RIGHT table?	bigquery-public-data.covid19_italy.national_trends
4	Which columns from the RIGHT table do I need?	Total_current_confirmed_cases
5	Do I need to do any basic calculations from columns in LEFT and RIGHT table? Which columns, and what calculation is needed?	YES. J.total_confirmed_cases/M.total_current_confirmed_cases AS region_highest_contribution
6	Which columns are common in my LEFT and RIGHT table?	date
7	Do I need all the data from my LEFT table? Or do I just need the data where there is a common match?	No. The data needed from the left table are: Date, region_name, total_confirmed_cases
8	Do I need to filter for anything? If so, what filters do I need?	YES. date='2020-10-31'
9	Show the query to get the data needed. You can type, copy paste, or paste an image.	<pre>/*Which region had the highest contribution in total current national cases, as of October 31, 2021?*/ SELECT J.date ,J.region_name ,J.total_confirmed_cases/M.total_current_confirmed_cases AS region_highest_contribution FROM bigquery-public-data.covid19_italy.data_by_region J LEFT JOIN bigquery-public-data.covid19_italy.national_trends M</pre>



Data Analytics Bootcamp

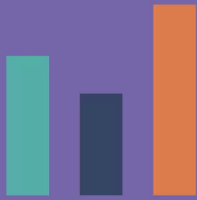
		<pre>ON J.date = M.date WHERE DATE(J.date) = DATE('2021-10-31') ORDER BY 3 DESC;</pre>																																
10	Show a screenshot of the output. No need to show everything, just a sample will do.	<div><div><</div><div>JOB INFORMATION</div><div>RESULTS</div><div>JSON</div><div>EXECUTION DETAILS</div></div> <table><tr><th>Row</th><th>date</th><th>region_name</th><th>region_highest_contribution</th></tr><tr><td>1</td><td>2021-10-31 17:00:00 UTC</td><td>Lombardia</td><td>10.850269260624879</td></tr><tr><td>2</td><td>2021-10-31 17:00:00 UTC</td><td>Veneto</td><td>5.8274306229380946</td></tr><tr><td>3</td><td>2021-10-31 17:00:00 UTC</td><td>Campania</td><td>5.6668202988550362</td></tr><tr><td>4</td><td>2021-10-31 17:00:00 UTC</td><td>Emilia-Romagna</td><td>5.2429895206675718</td></tr><tr><td>5</td><td>2021-10-31 17:00:00 UTC</td><td>Lazio</td><td>4.7926693188433918</td></tr><tr><td>6</td><td>2021-10-31 17:00:00 UTC</td><td>Piemonte</td><td>4.66920968367941</td></tr><tr><td>7</td><td>2021-10-31 17:00:00 UTC</td><td>Sicilia</td><td>3.7385018435862603</td></tr></table>	Row	date	region_name	region_highest_contribution	1	2021-10-31 17:00:00 UTC	Lombardia	10.850269260624879	2	2021-10-31 17:00:00 UTC	Veneto	5.8274306229380946	3	2021-10-31 17:00:00 UTC	Campania	5.6668202988550362	4	2021-10-31 17:00:00 UTC	Emilia-Romagna	5.2429895206675718	5	2021-10-31 17:00:00 UTC	Lazio	4.7926693188433918	6	2021-10-31 17:00:00 UTC	Piemonte	4.66920968367941	7	2021-10-31 17:00:00 UTC	Sicilia	3.7385018435862603
Row	date	region_name	region_highest_contribution																															
1	2021-10-31 17:00:00 UTC	Lombardia	10.850269260624879																															
2	2021-10-31 17:00:00 UTC	Veneto	5.8274306229380946																															
3	2021-10-31 17:00:00 UTC	Campania	5.6668202988550362																															
4	2021-10-31 17:00:00 UTC	Emilia-Romagna	5.2429895206675718																															
5	2021-10-31 17:00:00 UTC	Lazio	4.7926693188433918																															
6	2021-10-31 17:00:00 UTC	Piemonte	4.66920968367941																															
7	2021-10-31 17:00:00 UTC	Sicilia	3.7385018435862603																															



Data Analytics Bootcamp

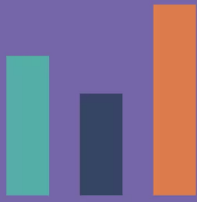
V. Show the contribution of each province to the total regional cases, for each day. Only include days where total regional confirmed cases > 0.

#	Question/Task	Query/Answer
1	Which table do I get the data from ? (which will I use as my LEFT table?)	bigquery-public- <code>data.covid19_italy.data_by_province</code> J
2	Which columns from my LEFT table do I need?	Date, region_code, name, confirmed_cases
3	Do I need to reference another table? If so, what will be my RIGHT table?	bigquery-public- <code>data.covid19_italy.data_by_region</code> M
4	Which columns from the RIGHT table do I need?	<code>Region_code, region_name, total_confirmed_cases</code>
5	Do I need to do any basic calculations from columns in LEFT and RIGHT table? Which columns, and what calculation is needed?	<code>confirmed_cases/total_confirmed_cases</code>
6	Which columns are common in my LEFT and RIGHT table?	<code>date, region_code</code>
7	Do I need all the data from my LEFT table? Or do I just need the data where there is a common match?	NO. Only the following: Date, region_code, name, confirmed_cases
8	Do I need to filter for anything? If so, what filters do I need?	<code>Total_regional_confirmed_cases > 0</code>
9	Show the query to get the data needed. You can type, copy paste, or paste an image.	<pre>/*Show the contribution of each province to the total regional cases, for each day. Only include days where total regional confirmed cases > 0.*/ SELECT J.date ,J.region_code ,J.name AS region_name ,M.region_code AS region_code_R ,M.region_name AS region_name_R ,J.confirmed_cases/M.total_confirmed_cases AS province_contribution_total_cases</pre>



Data Analytics Bootcamp

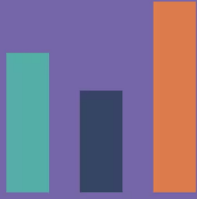
		<pre>FROM bigquery-public- data.covid19_italy.data_by_province J LEFT JOIN bigquery-public- data.covid19_italy.data_by_region M ON DATE(J.date) = DATE(M. date) AND J.region_code = J.region_code WHERE M.total_confirmed_cases > 0;</pre>																																																								
1 0	Show a screenshot of the output. No need to show everything, just a sample will do.	<table><tr><th colspan="2">JOB INFORMATION</th><th>RESULTS</th><th>JSON</th><th>EXECUTION DETAILS</th><th>EXECUTION GRAPH</th><th>PREVIEW</th></tr><tr><th>Row</th><th>date</th><th>region_code</th><th>region_name</th><th>region_code_R</th><th>region_name_R</th><th>province_contribution_total_cas</th></tr><tr><td>1</td><td>2020-06-25 17:00:00 UTC</td><td>13</td><td>Abruzzo</td><td>1</td><td>Piemonte</td><td>0.00079882413087934562</td></tr><tr><td>2</td><td>2020-06-26 17:00:00 UTC</td><td>13</td><td>Abruzzo</td><td>1</td><td>Piemonte</td><td>0.00079844144230462142</td></tr><tr><td>3</td><td>2020-06-27 17:00:00 UTC</td><td>13</td><td>Abruzzo</td><td>1</td><td>Piemonte</td><td>0.00079816103697081923</td></tr><tr><td>4</td><td>2020-06-28 17:00:00 UTC</td><td>13</td><td>Abruzzo</td><td>1</td><td>Piemonte</td><td>0.000797804442175134</td></tr><tr><td>5</td><td>2020-06-29 17:00:00 UTC</td><td>13</td><td>Abruzzo</td><td>1</td><td>Piemonte</td><td>0.00079775352607058521</td></tr><tr><td>6</td><td>2020-06-30 17:00:00 UTC</td><td>13</td><td>Abruzzo</td><td>1</td><td>Piemonte</td><td>0.00079747360362372006</td></tr></table> <div>Results per page: 50 1 - 50 of 3629074 1/6</div>	JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW	Row	date	region_code	region_name	region_code_R	region_name_R	province_contribution_total_cas	1	2020-06-25 17:00:00 UTC	13	Abruzzo	1	Piemonte	0.00079882413087934562	2	2020-06-26 17:00:00 UTC	13	Abruzzo	1	Piemonte	0.00079844144230462142	3	2020-06-27 17:00:00 UTC	13	Abruzzo	1	Piemonte	0.00079816103697081923	4	2020-06-28 17:00:00 UTC	13	Abruzzo	1	Piemonte	0.000797804442175134	5	2020-06-29 17:00:00 UTC	13	Abruzzo	1	Piemonte	0.00079775352607058521	6	2020-06-30 17:00:00 UTC	13	Abruzzo	1	Piemonte	0.00079747360362372006
JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW																																																				
Row	date	region_code	region_name	region_code_R	region_name_R	province_contribution_total_cas																																																				
1	2020-06-25 17:00:00 UTC	13	Abruzzo	1	Piemonte	0.00079882413087934562																																																				
2	2020-06-26 17:00:00 UTC	13	Abruzzo	1	Piemonte	0.00079844144230462142																																																				
3	2020-06-27 17:00:00 UTC	13	Abruzzo	1	Piemonte	0.00079816103697081923																																																				
4	2020-06-28 17:00:00 UTC	13	Abruzzo	1	Piemonte	0.000797804442175134																																																				
5	2020-06-29 17:00:00 UTC	13	Abruzzo	1	Piemonte	0.00079775352607058521																																																				
6	2020-06-30 17:00:00 UTC	13	Abruzzo	1	Piemonte	0.00079747360362372006																																																				



Data Analytics Bootcamp

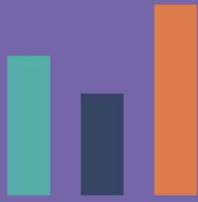
VI. Which province had the highest contribution to total national cases, for November 01, 2021?

#	Question/Task	Query/Answer
1	Which table do I get the data from ? (which will I use as my LEFT table?)	bigquery-public- data.covid19_italy.data_by_province
2	Which columns from my LEFT table do I need?	Date, province_name, province_code, confirmed_cases
3	Do I need to reference another table? If so, what will be my RIGHT table?	bigquery-public- data.covid19_italy.national_trends
4	Which columns from the RIGHT table do I need?	Total_confirmed_cases
5	Do I need to do any basic calculations from columns in LEFT and RIGHT table? Which columns, and what calculation is needed?	confirmed_cases/total_confirmed_cases
6	Which columns are common in my LEFT and RIGHT table?	Date, country
7	Do I need all the data from my LEFT table? Or do I just need the data where there is a common match?	No. the data needed are the following: Date, province_name, province_code, confirmed_cases
8	Do I need to filter for anything? If so, what filters do I need?	YES. DATE(J.date) = DATE('2021-11-01') ORDER BY province_contribution_total_national_cases DESC LIMIT 1;
9	Show the query to get the data needed. You can type, copy paste, or paste an image.	/*Which province had the highest contribution to total national cases, for November 01, 2021*/ SELECT J.date , J.province_name , J.province_code , J.confirmed_cases/0.total_confirmed_cases AS province_contribution_total_national_cases



Data Analytics Bootcamp

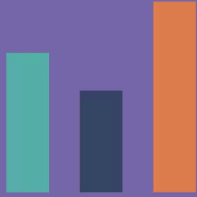
		<pre>FROM bigquery-public- data.covid19_italy.data_by_province J LEFT JOIN bigquery-public- data.covid19_italy.national_trends O ON DATE(J.date) = DATE (O.date) AND J.country = O.country WHERE DATE(J.date) = DATE('2021-11-01') ORDER BY province_contribution_total_national_cases DESC LIMIT 1;</pre>										
1 0	Show a screenshot of the output. No need to show everything, just a sample will do.	<table><tr><th>Row</th><th>date</th><th>province_name</th><th>province_code</th><th>province_contribution_total_national_cases</th></tr><tr><td>1</td><td>2021-11-01 17:00:00 UTC</td><td>Milano</td><td>15</td><td>0.06038389598019428</td></tr></table>	Row	date	province_name	province_code	province_contribution_total_national_cases	1	2021-11-01 17:00:00 UTC	Milano	15	0.06038389598019428
Row	date	province_name	province_code	province_contribution_total_national_cases								
1	2021-11-01 17:00:00 UTC	Milano	15	0.06038389598019428								



Data Analytics Bootcamp

VII. 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.

#	Question/Task	Query/Answer
1	Which table do I get the data from? (which will I use as my LEFT table?)	bigquery-public- data.covid19_italy.data_by_province J
2	Which columns from my LEFT table do I need?	Date, province_name, region_code, name, confirmed_cases
3	Do I need to reference another table? If so, what will be my RIGHT table?	bigquery-public- data.covid19_italy.data_by_region O AND bigquery-public- data.covid19_italy.national_trends M
4	Which columns from the RIGHT table do I need?	Total_confirmed_cases for both regional and national
5	Do I need to do any basic calculations from columns in LEFT and RIGHT table? Which columns, and what calculation is needed?	J.confirmed_cases/M.total_confirmed_cases, J.confirmed_cases/O.total_confirmed_cases
6	Which columns are common in my LEFT and RIGHT table?	Date
7	Do I need all the data from my LEFT table? Or do I just need the data where there is a common match?	No. the following the only the needed data: Date, province_name, region_code, name, confirmed_cases
8	Do I need to filter for anything? If so, what filters do I need?	YES. DATE(J.date) = DATE('2021-10-10') ;
9	Show the query to get the data needed. You can type, copy paste, or paste an image.	<i>/*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.*/</i> SELECT J.date ,J.province_name ,J.region_code AS region_code_P ,J.name AS region_name_P ,O.region_name AS region_name_R



Data Analytics Bootcamp

```
,O.region_code AS region_name_R
,J.confirmed_cases/O.total_confirmed_cases AS
province_contribution_regional_cases
,J.confirmed_cases/M.total_confirmed_cases AS
province_contribution_national_cases
FROM bigquery-public-
data.covid19_italy.data_by_province J
LEFT JOIN bigquery-public-
data.covid19_italy.data_by_region O
ON DATE(J.date) = DATE(O.date)
LEFT JOIN bigquery-public-
data.covid19_italy.national_trends M
ON DATE(J.date) = DATE(M.date)
WHERE
DATE(J.date) = DATE('2021-10-10');
```

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

Row	date	province_name	region_code	region_name_P	region_name_R	region_name	province_contribution_regional_cs	province_contribution_national_cas
1	2021-10-10 ..	Fuori Regione / Provincia ..	13	Abruzzo	Lombardia	3	0.00075439188008213736	0.0001423308560530824
2	2021-10-10 ..	Fuori Regione / Provincia ..	13	Abruzzo	Emilia-Romagna	8	0.001568842644379252	0.0001423308560530824
3	2021-10-10 ..	Fuori Regione / Provincia ..	13	Abruzzo	Lazio	12	0.0017277401726707145	0.0001423308560530824
4	2021-10-10 ..	Fuori Regione / Provincia ..	13	Abruzzo	Valle d'Aosta	2	0.054804820299827966	0.0001423308560530824
5	2021-10-10 ..	Fuori Regione / Provincia ..	13	Abruzzo	Basilicata	17	0.021981271562345982	0.0001423308560530824
6	2021-10-10 ..	Fuori Regione / Provincia ..	13	Abruzzo	Molise	14	0.045992025299051283	0.0001423308560530824
7	2021-10-10 ..	Fuori Regione / Provincia ..	13	Abruzzo	P.A. Trento	22	0.013753546317996792	0.0001423308560530824

Results per page: 50 1 - 50 of 3179