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
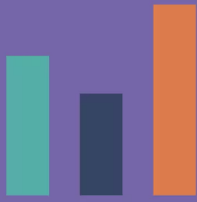
SELECT
    [ALIAS.column]
    , [ALIAS.column2]
    , …
FROM
    [left table ALIAS]
LEFT JOIN
    [right table ALIAS]
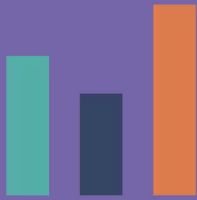   ON [left table].[key] = [right table].[key]
    AND …
WHERE
    [column] [condition]

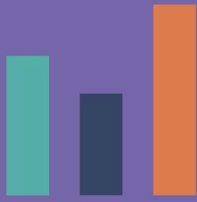*IV. Which region had the highest contribution in total current national cases, as of October 31, 2021?*

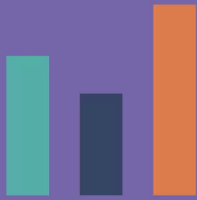| #	Question/Task	Query/Answer
1	Which table do I get the data **from?** (which will I use as my **LEFT** table?)	bigquery-public-data.covid19_italy.data_by_region
2	Which columns from my **LEFT** table do I need?	Date, region_name, total_confirmed_cases
3	Do I need to reference another table? If so, 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.	```/*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```

| | | ```sql
ON J.date = M.date
WHERE
DATE(J.date) = DATE('2021-10-31')
ORDER BY
3 DESC;
``` |
| 10 | Show a screenshot of the output. No need to show everything, just a sample will do. | |

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

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-data`.covid19_italy.data_by_province 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-data`.covid19_italy.data_by_region M
4	Which columns from the **RIGHT** table do I need?	Region_code, region_name,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,region_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?	Total_regional_confirmed_cases > 0
9	Show the query to get the data needed. You can type, copy paste, or paste an image.	```/*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```

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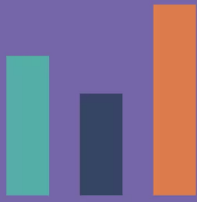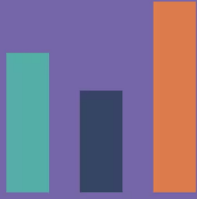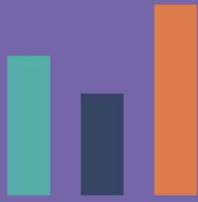| 10 | Show a screenshot of the output. No need to show everything, just a sample will do. | |

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

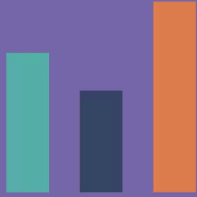| #	Question/Task	Query/Answer
1	Which table do I get the data **from?** (which will I use as my **LEFT** table?)	`bigquery-public-data`.covid19_italy.data_by_province
2	Which columns from my **LEFT** table do I need?	Date, province_name, province_code, confirmed_cases
3	Do I need to reference another 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/O.total_confirmed_cases AS province_contribution_total_national_cases`

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

Row	date	province_name	province_code	province_contribution_total_national_cases
1	2021-11-01 17:00:00 UTC	Milano	15	0.06038389598019428

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.	`/*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`
		`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`

<table>
<tr><td></td><td></td><td>

```
,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');
```

</td></tr>
<tr><td>10</td><td>Show a screenshot of the output. No need to show everything, just a sample will do.</td><td>

</td></tr>
</table>