Data Analytics Bootcamp – Cohort 9
2023 February 15

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]
```

Operators

Inequalities

<, >, >=, <=, <>

Ex: date > DATE('2021-01-01')

ate after jan 1, 2021

Text

LIKE, IS, NOT

% for wildcard in LIKE operators

Ex: text LIKE '%ext'

ext ends with 'ext'

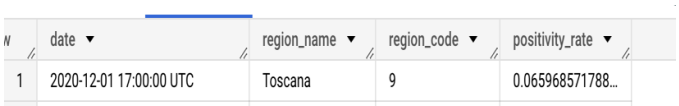
Specific Values

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

#	*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)	date, region_name, region_code, total_confirmed_cases, test_performed
3	Do I need to use any filter?	YES
4	If #3 is yes, what filters do I need?	`Month of December, region_code of 9 and name of Toscana`
5	Do I need to arrange my dataset? Which column? In ascending or descending order?	YES, DESC
6	Do I need to limit the results of my dataset? If so, to how many rows?	YES, LIMIT 1

7	Show the query to get the data needed. You can type, copy paste, or paste an image.	`SELECT`
		`O.date`
		`,O.region_name`
		`,O.region_code`
		`,O.total_confirmed_cases/O.tests_performed AS positivity_rate`
		`FROM bigquery-public-data.covid19_italy.data_by_region O`
		`WHERE`
		`O.region_code in ('9')`
		`AND (DATE(O.date) BETWEEN '2020-12-01' AND '2020-12-31')`
		`ORDER BY`
		`Positivity_rate DESC`
		`LIMIT 1;`

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

V. 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.

#	*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, new_current_confirmed_cases
3	Do I need to use any filter?	YES
4	If #3 is yes, what filters do I need?	`Region_code=('17', '19')`
5	Do I need to arrange my dataset? Which column? In ascending or descending order?	New_current_confirmed_cases DESC
6	Do I need to limit the results of my dataset? If so, to how many rows?	1

7	Show the query to get the data needed. You can type, copy paste, or paste an image.	`SELECT`
		`date`
		`,region_name`
		`,region_code`
		`,new_current_confirmed_cases`
		`FROM bigquery-public-data.covid19_italy.data_by_region`
		`WHERE`
		`region_code IN ('17','19')`
		`AND (DATE(date) BETWEEN '2020-06-01' AND '2020-06-30')`
		`OR (DATE(date) BETWEEN '2020-10-01' AND '2020-10-31')`
		`OR (DATE(date) BETWEEN '2020-12-01' AND '2020-12-31')`
		`OR (DATE(date) BETWEEN '2021-03-01' AND '2021-03-31')`
		`ORDER BY`
		`new_current_confirmed_cases DESC`
		`LIMIT 1;`

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

VI. 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

#	*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)	date, region_name, region_code, new_current_confirmed_cases
3	Do I need to use any filter?	YES
4	If #3 is yes, what filters do I need?	**Region_name starting in P** **Date BETWEEN 2020-01-01 AND 2020-10-25** **OR 2021-01-01 AND 2021-10-25**
5	Do I need to arrange my dataset? Which column? In ascending or descending order?	YES, Date 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.	`/*Days and regions had a negative increase in current confirmed cases. Consider only region that starts with the letter "P", and January 01-October 25 for both years 2020 and 2021 only. Sort output by date*/`
		
		`SELECT`
		`K.date`
		`, K.region_name`
		`, K.region_code`
		`, K.new_current_confirmed_cases`
		`FROM bigquery-public-data.covid19_italy.data_by_region K`
		`WHERE`
		`K.region_name LIKE 'P%'`
		`AND (DATE(K.date) BETWEEN '2020-01-01'AND '2020-10-25')`
		`OR (DATE(K.date) BETWEEN '2021-01-01' AND '2021-10-25')`

		AND K.new_current_confirmed_cases < 0 ORDER BY K.date ASC;
8	Show a screenshot of the output. No need to show everything, just a sample will do.	

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH PREVIEW

Row	date	region_name	region_code	new_current_confirmed_cases
1	2020-02-24 18:00:00 UTC	Piemonte	1	0
2	2020-02-24 18:00:00 UTC	Puglia	16	0
3	2020-02-24 18:00:00 UTC	P.A. Bolzano	21	0
4	2020-02-24 18:00:00 UTC	P.A. Trento	22	0
5	2020-02-25 18:00:00 UTC	Piemonte	1	0
6	2020-02-25 18:00:00 UTC	Puglia	16	0

VII. 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.

#	*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)	date, region_name, region_code, new_current_confirmed_cases
3	Do I need to use any filter?	YES

4	If #3 is yes, what filters do I need?	`Region_name starting in P`
		`Date BETWEEN 2020-01-01 AND 2020-10-25`
		`OR 2021-01-01 AND 2021-10-25`

5	Do I need to arrange my dataset? Which column? In ascending or descending order?	Date, region ASC
6	Do I need to limit the results of my dataset? If so, to how many rows?	LIMIT 1
7	Show the query to get the data needed. You can type, copy paste, or paste an image.	`/*Days and regions had a negative increase in current confirmed cases. Consider only region that starts with the letter "P" and January 01-October 25 for both years 2020 and 2021. Days and region registered the lowest case count*/`
8	Show a screenshot of the output. No need to show everything, just a sample will do.