Data Analytics Bootcamp – Cohort 9
2023 February 13

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'

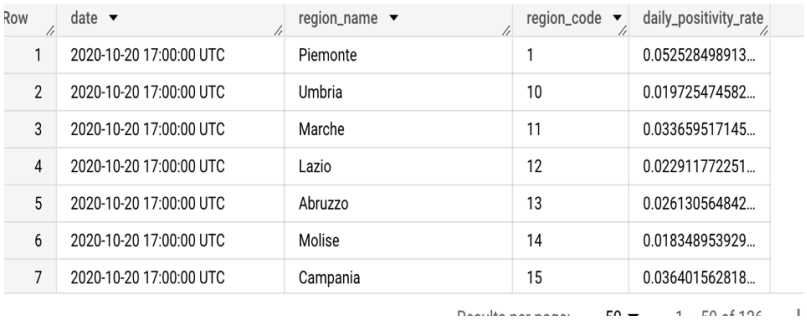
ext ends with 'ext'

Specific Values
BETWEEN, IN
Ex: values IN ('a','b','c')

alue = a or b or c

I. Get the daily positivity rate for each region. Show data only for October 20, 2020 – October 25, 2020 and order results by date then by region name.
positivity rate = total cases/ total tests

#	*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,total_confirmed_cases,tests_performed, region_code,region_name
3	Do I need to use any filter?	YES
4	If #3 is yes, what filters do I need?	Date(October 20, 2020-October 25, 2020)
5	Do I need to arrange my dataset? Which column? In ascending or descending order?	ORDER BY date ASC, region_code ASC
6	Do I need to limit the results of my dataset? If so, to how many rows?	BETWEEN October 20, 2020 AND October 25, 2020
7	Show the query to get the data needed. You can type, copy paste, or paste an image.	SELECT JOM.date , JOM.region_name , JOM.region_code , JOM.total_confirmed_cases/JOM.tests_performed AS daily_positivity_rate FROM bigquery-public-data.covid19_italy.data_by_region JOM WHERE DATE(JOM.date) BETWEEN DATE('2020-10-20') AND DATE ('2020-10-25') ORDER BY

		`JOM.date ASC, JOM.region_code ASC;`
8	Show a screenshot of the output. No need to show everything, just a sample will do.	

Row	date	region_name	region_code	daily_positivity_rate
1	2020-10-20 17:00:00 UTC	Piemonte	1	0.052528498913...
2	2020-10-20 17:00:00 UTC	Umbria	10	0.019725474582...
3	2020-10-20 17:00:00 UTC	Marche	11	0.033659517145...
4	2020-10-20 17:00:00 UTC	Lazio	12	0.022911772251...
5	2020-10-20 17:00:00 UTC	Abruzzo	13	0.026130564842...
6	2020-10-20 17:00:00 UTC	Molise	14	0.018348953929...
7	2020-10-20 17:00:00 UTC	Campania	15	0.036401562818...

II. From your answer in 1 (i.e. using the same columns and conditions), which region and date registered the highest positivity rate? 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,total_confirmed_cases,tests_performed, region_code,region_name
3	Do I need to use any filter?	YES
4	If #3 is yes, what filters do I need?	`Date(October 20, 2020-October 25, 2020)`
5	Do I need to arrange my dataset? Which column? In ascending or descending order?	ORDER BY date ASC, region_code ASC
6	Do I need to limit the results of my dataset? If so, to how many rows?	highest_daily_positivity_rate

7	Show the query to get the data needed. You can type, copy paste, or paste an image.	`SELECT`
		`JOM.date`
		`, JOM.region_name`
		`, JOM.region_code`
		`, JOM.total_confirmed_cases/JOM.tests_performed AS daily_positivity_rate`
		
		`FROM bigquery-public-data.covid19_italy.data_by_region JOM`
		`WHERE`
		`DATE(JOM.date) BETWEEN DATE('2020-10-20') AND DATE ('2020-10-25')`
		`ORDER BY`
		`daily_positivity_rate ASC`
		`LIMIT 1`

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