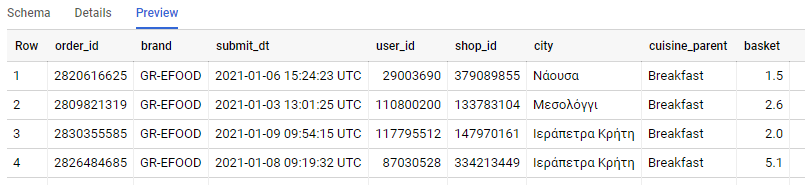
# SQL/Data Handling:

Below you will find a dummy snapshot of Orders data for January 2021 (excluding the big cities of Athens and Thessaloniki). For this period of time, we would like to focus on the "Breakfast" cuisine\_parent and see how it differs from other cuisines. "Breakfast" stands for restaurants that mainly offer coffee and beverages and create additional habits to efood users.



For every city that exceeds the 500 orders we would like to know:

- how many orders are placed in the "Breakfast" cuisine (Breakfast\_Orders)?

- how many users ordered from the "Breakfast" cuisine (Breakfast\_Users)?

- what is the Average Basket of the Breakfast\_Users that we defined before?

Order your findings, showing us the 10 cities with the most Breakfast Orders.

We are mainly using the [Standard SQL of BigQuery](https://cloud.google.com/bigquery/docs/reference/standard-sql/functions-and-operators).

You can find the table and the available SQL Editor here: [Link to BQ](https://console.cloud.google.com/bigquery?project=bi-2019-test&authuser=1&p=bi-2019-test&d=ad_hoc&t=orders_jan2021&page=table).

Please note that you need to be logged in with your **gmail** account in order to access it.

*------------------------------------MY ANSWER – Vasileios Anastasopoulos ----------------------------------*

*Please, paste here:*

1. *screenshot(s) from the editor’s Query Results, showing your answer.*
2. *your SQL query/queries*

*SQL queries & Screenshots of Results*

*a)* First of all we check that there is not duplicated records (distinct order\_id) and we verify there are not.Then we the below script we identify all cities with more than 500 orders, in descending order

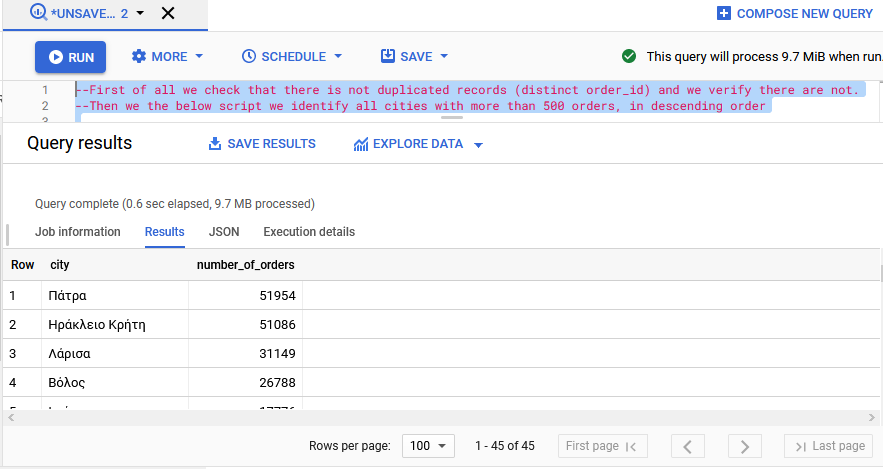
SELECT city, count (order\_id) number\_of\_orders

FROM `bi-2019-test.ad\_hoc.orders\_jan2021`

group by city

having count(order\_id)>500

order by count(order\_id) desc



*b)* how many orders are placed in the "Breakfast" cuisine (Breakfast\_Orders)?

select city, count(order\_id) as number\_of\_breakfast\_orders

from `bi-2019-test.ad\_hoc.orders\_jan2021`

where cuisine\_parent = 'Breakfast'

and city in

 (

 select distinct city from

    (

    SELECT city

    ,count (order\_id) number\_of\_orders

    FROM `bi-2019-test.ad\_hoc.orders\_jan2021`

    group by city

    having count(order\_id)>500

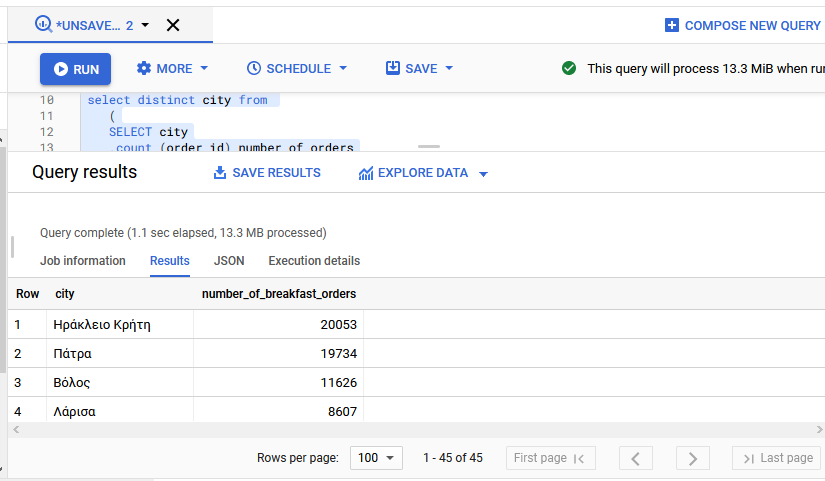
    )

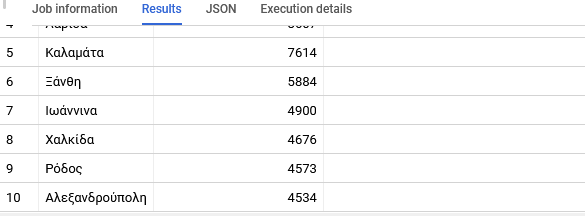
 )

 group by city

 order by count(order\_id) desc

limit 10 –-in order to show only top 10 cities with the most breakfast orders (of all the cities that have more than 500 total orders across cuisines)



**

*c)* how many users ordered from the "Breakfast" cuisine (Breakfast\_Users)?

--Updated Comment as part of Main Case Study: From all the cities with more than 500 orders, we have to include in the results only the top 10 cities with breakfast orders (same as we did in the first question)

select city, count(distinct user\_id) as number\_of\_breakfast\_users

from `bi-2019-test.ad\_hoc.orders\_jan2021`

where cuisine\_parent = 'Breakfast'

and city in

 (

 select distinct city from

    (

    SELECT city

    ,count (order\_id) number\_of\_orders

    FROM `bi-2019-test.ad\_hoc.orders\_jan2021`

    group by city

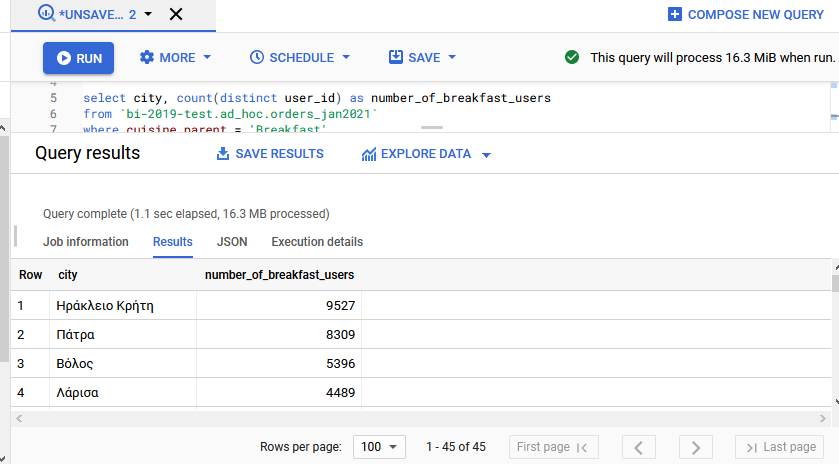
    having count(order\_id)>500

    )

 )

 group by city

 order by count(distinct user\_id) desc



*d) What* is the Average Basket of the Breakfast\_Users that we defined before?

--Updated Comment as part of Main Case Study: a) We take into account the top 10 cities (limit 10) with more than 500 orders, however we need to take into account the top 10 cities with Breakfast orders among all cities with more than 500 orders. After calculating the top 10 Breakfast cities, than we should find the average basket size of each user, for those top 10 cities.

select user\_id, avg(basket) as avg\_basket\_per\_user

from `bi-2019-test.ad\_hoc.orders\_jan2021`

where cuisine\_parent = 'Breakfast'

and city in

 (

 select distinct city from

    (

    SELECT city

    ,count (order\_id) number\_of\_orders

    FROM `bi-2019-test.ad\_hoc.orders\_jan2021`

    group by city

    having count(order\_id)>500

limit 10

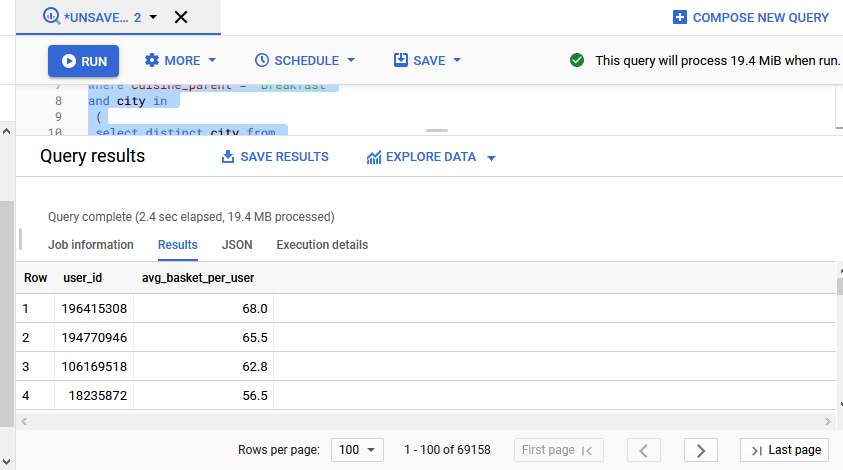
–-because we want only the 10 top cities with more than 500 orders

    )

 )

 group by user\_id

 order by avg\_basket\_per\_user desc



*--------------------------------------------------------- THE END -------------------------------------------------------*