Answers to SQL Assessment

SELECT

  COUNT(DISTINCT organizer.organizer\_id) AS Number\_of\_organizers

FROM `st-project-353901.Interview\_Test.event\_table` AS Events

RIGHT JOIN `st-project-353901.Interview\_Test.order\_table` AS Orders

  ON(Events.event\_id = Orders.event\_id)

JOIN `st-project-353901.Interview\_Test.organizer\_table` AS organizer

  ON(organizer.organizer\_id=Events.organizer\_id)

WHERE   signup\_country="USA"

1. How many organizers who signed up in the USA had at least one order?

After analyzing the data, the answer that I received from writing the query above was 1,736 organizers that had signed up in the USA, had at least one order. The reason I wrote this query is because I needed information from three different tables in order to answer this question. I was able to combine the tables with a Right Join, Inner Join, and the standard FROM statement. I needed to combine the tables in this order that way none of the necessary data was lost, and so I could apply the appropriate filter. I also used the COUNT DISTINCT function to avoid receiving duplicate organizers, as one organizer could have multiple orders.

Graphical user interface, text, application, email

Description automatically generated

SELECT

ORGANIZER.organizer\_id,

COUNT(\*) AS Number\_of\_events

FROM `st-project-353901.Interview\_Test.event\_table` EVENTS

LEFT JOIN `st-project-353901.Interview\_Test.organizer\_table` ORGANIZER

ON(EVENTS.organizer\_id=ORGANIZER.organizer\_id)

WHERE event\_paid\_type="free event" AND NOT first\_event\_country="US" AND NOT event\_country="US"

GROUP BY ORGANIZER.organizer\_id

ORDER BY 2 DESC

LIMIT 10

1. What are the Top 10 organizers by the number of free events created outside the USA?

After analyzing the data, I have included a table below of the top organizer’s below, meeting the specific criteria. I structured this query to show me the organizers in the select function, as well as counting the total number of number of events. I then used the from and Join, to connect the two tables that had the necessary data to answer the question. I was then able to use the WHERE function to filter out any events that started in the US. I then wanted to group by the organizer id to connect those events with the organizers that hosted them. I then used order by to show the 2 columns we needed to view. Lastly, I used the LIMIT function to only return the top 10 organizers.

Table

Description automatically generated

1. What organizers are growing the fastest?

I believe there are many ways to measure growth of an organizer. Given more time, I would be able to perform a more in-depth analysis. However, for this example, I chose to measure the number of orders for each organizer from the last 5 years. This will give me a small insight as to how well each organizer has been doing. I composed the following query to get the table that follows.

--I want the orders for each org\_id within the last 5 years

SELECT organizer.organizer\_id,

  COUNT(orders.order\_id) AS Number\_of\_Orders,

  FROM `st-project-353901.Interview\_Test.event\_table` AS Events

  RIGHT JOIN `st-project-353901.Interview\_Test.order\_table` AS Orders

  ON(Events.event\_id = Orders.event\_id)

JOIN `st-project-353901.Interview\_Test.organizer\_table` AS organizer

  ON(organizer.organizer\_id=Events.organizer\_id)

  WHERE organizer.first\_paid\_publish\_date > "01/01/2016"

GROUP BY (organizer.organizer\_id)

ORDER BY 2 DESC

LIMIT 10

Table

Description automatically generated

1. An analyst is using the following query to understand how long it takes for an organizer to publish their first event after signup.

SELECT subquery.signup\_to\_publish, COUNT(subquery.organizer\_id) org\_count FROM ( SELECT organizer\_table.organizer\_id, DATE\_DIFF('day',organizer\_table.signup\_date,event\_table.event\_publish\_date) signup\_to\_publish FROM event\_table JOIN organizer\_table ON organizer\_table.organizer\_id = event\_table.organizer\_id GROUP BY 1 )subquery GROUP BY 1

a. Do you think this query is working as intended?

b. How, if at all, would you alter this query? Why?

I do not think that this query is working as intended. The structure of the query appears to be correct, however I do not see any specified dates that the analyst would be trying to collect data between to answer the question. The only thing I would change is to include the actual dates to find the gap between the signup and publish date.