**CHALLENGE 2:**

1. What proportion of orders contained a product with category “Sheet Sets” and size “King”?

WITH order\_products AS (

SELECT \_id, COUNT(\*) AS num\_products

FROM order\_product\_mapping

),

sheet\_sets AS (

SELECT id, COUNT(\*) AS num\_sheet\_sets

FROM order\_product\_mapping

WHERE category = 'Sheet Sets' AND product\_size = 'King'

GROUP BY order\_id

)

SELECT 100.0 \* SUM(num\_sheet\_sets) / SUM(num\_products) AS proportion

FROM order\_products

JOIN sheet\_sets

ON order\_products.\_id = sheet\_sets.\_id;

1. Which product SKU generated the most gross revenue?

with total\_gross\_revenue as (\_

select (total - line\_items.line\_total\_discount) as gross\_revenue, \_id

from `ae\_data\_challenge\_v1.orders`

Group by 2

),

Select t.gross\_revenue, p.product\_sku

from total\_gross\_revenue t join order\_product\_mapping p on t.\_id=p.\_id

Order by 1 desc

Limit 1

1. Which date had the highest average order units?

I was confused as to which date field I could use, but according to me, created\_date made the most sense.

with total\_order\_units as (

select \_id, sum (line\_items[SAFE\_OFFSET(0)].quantity) as total\_units

from `ae\_data\_challenge\_v1.orders`

group by 1

)

select

AVG(t.total\_units), o.created\_date

from total\_order\_units t join `ae\_data\_challenge\_v1.orders` o

on t.\_id= o.\_id

group by 1

order by 2 desc

1. What was the conversion rate of web sessions where the user added a “Plush Bath Towel Set” product to their cart?

With conversion as ( SELECT

100.0 \* (COUNT(DISTINCT CASE WHEN w.event\_name = 'order\_completed' THEN n.session\_id END) /

COUNT(DISTINCT n.session\_id) ) AS conversion\_rate,

w.event\_name,

w.timestamp

FROM new\_web\_sessions n

LEFT JOIN `ae\_data\_challenge\_v1.web\_events` w

ON n.session\_id = w.session\_id)

Select c.conversion\_rate from conversion c

join order\_product\_mapping p on p.created\_at = c.timestamp

where c.event\_name = ‘product\_added’ and p.title = ‘Plush Bath Towel Set’

1. Among `page` events only, what are the top five most common page URLs that immediately preceded a user’s navigation to the “checkout.bollandbranch.com” domain during their session?

WITH checkout\_sessions AS (

SELECT session\_id, MIN(timestamp) AS checkout\_time

FROM new\_web\_sessions

WHERE event\_URL LIKE '%checkout.bollandbranch.com%'

GROUP BY session\_id

),

preceding\_pages AS (

SELECT session\_id, event\_URL, timestamp

FROM new\_web\_sessions

WHERE event\_name ilike '%page%'

AND timestamp < (

SELECT checkout\_time

FROM checkout\_sessions

WHERE new\_web\_sessions.session\_id = checkout\_sessions.session\_id

)

ORDER BY timestamp DESC

LIMIT 1

)

SELECT event\_URL, COUNT(\*) AS num\_checkouts

FROM preceding\_pages

GROUP BY 1

ORDER BY num\_checkouts DESC

LIMIT 5;

1. What are the top five non-null session campaigns which garnered the most web users?

WITH user\_sessions AS (

SELECT n.utm\_campaign, COUNT(DISTINCT a.customer\_id) AS num\_users

FROM new\_web\_sessions n join `ae\_data\_challenge\_v1.web\_events` a

On n.cookie\_id= a.cookie\_id

WHERE utm\_campaign IS NOT NULL

GROUP BY 1

)

SELECT utm\_campaign, num\_users

FROM user\_sessions

ORDER BY num\_users DESC

LIMIT 5;

1. What are the top five non-null session sources which garnered the most gross revenue?

with total\_gross\_revenue as (\_

select (total - line\_items.line\_total\_discount) as gross\_revenue, created\_at

from `ae\_data\_challenge\_v1.orders`

)

SELECT n.utm\_source, t.total\_revenue

FROM new\_web\_sessions n join total\_gross\_revenue t

On n.timestamp = t.created\_at

ORDER BY total\_revenue DESC

LIMIT 5;