Google Analytics ecommerce

Google Merchandise Store is an online store that sells Google-branded merchandise. The site uses Google Analytics's standard web ecommerce implementation along with enhanced measurement. The `ga4_obfuscated_sample_ecommerce` dataset available through the BigQuery Public Datasets program contains a sample of obfuscated BigQuery event export data for three months from 2020-11-01 to 2021-01-31.

1."Using the GA4 sample ecommerce dataset, write a SQL query to find the total number of events, the total number of unique users, and the total number of unique days recorded in the dataset."

```
SELECT
  COUNT(*) AS event_count,
  COUNT(DISTINCT user_pseudo_id) AS user_count,
  COUNT(DISTINCT event_date) AS day_count
FROM `bigquery-public-data.ga4_obfuscated_sample_ecommerce.events_*`
```

event_count	user_count	day_count
4295584	270154	92

Insight: The dataset spans **~3 months (92 days)**, capturing interactions from **270K unique users** with a high activity level (**~46 events per user on average**). This indicates a rich dataset suitable for funnel and behavioral analysis.

2."Using the GA4 sample ecommerce dataset, write a SQL query to analyze the e-commerce conversion funnel. Specifically, find the number of users who performed each of the following steps: viewed an item, added an item to cart, began checkout, and completed a purchase. Also report the total number of users in the dataset."

```sql
WITH funnel AS (
 SELECT
   user_pseudo_id,
   MAX(CASE WHEN event_name = 'view_item' THEN event_timestamp END) AS
view_item_time,
   MAX(CASE WHEN event_name = 'add_to_cart' THEN event_timestamp END) AS
add_to_cart_time,
   MAX(CASE WHEN event_name = 'begin_checkout' THEN event_timestamp END) AS
begin_checkout_time,
   MAX(CASE WHEN event_name = 'purchase' THEN event_timestamp END) AS purchase_time
FROM `bigquery-public-data.ga4_obfuscated_sample_ecommerce.events_*`
group by 1)

select
      count(distinct user_pseudo_id) as user_count,
      count(distinct case when view_item_time is not null then user_pseudo_id end)
as view_item_users,
      count(distinct case when add_to_cart_time is not null then user_pseudo_id
end) as add_cart_users,
      count(distinct case when begin_checkout_time is not null then user_pseudo_id
end) as begin_checkout_users,
      count(distinct case when purchase_time is not null then user_pseudo_id end)
as purchase_time
from funnel
```

user_count	view_item_users	add_cart_users	begin_checkout_users	purchase_time
270154	61252	12545	9715	4419

Insight: The largest drop-off is between **viewing an item and adding it to cart** (only ~20% continue). The **overall conversion rate** (purchase / total users) is **1.6%**, which is within typical e-commerce benchmarks but suggests scope to optimize product page design, promotions, or checkout experience.

3."Using the GA4 sample ecommerce dataset, write a SQL query to analyze the Top 10 items added to cart by most users".

```sql
SELECT item_id,
       item_name,
       count(distinct user_pseudo_id) as total_users
from `bigquery-public-data.ga4_obfuscated_sample_ecommerce.events_*`,UNNEST(items)
where event_name='add_to_cart'
group by item_id,item_name
order by total_users desc
limit 10
```

item_id	item_name	total_users
GGOEGAPH161899	Google Land & Sea Cotton Cap	3567
GGOEGXXX1344	Google Navy Speckled Tee	3526
GGOEGXXX1109	Google Zip Hoodie F/C	3519
GGOEGXXX1613	Super G Unisex Joggers	3247
GGOEYHPB121110	YouTube Leather Strap Hat Black	3190
GGOEGXXX1193	Google Campus Bike Eco Tee Navy	3061
GGOEGAEC134910	Google Speckled Beanie Navy	3057
GGOEGXXX1181	Google Women's Striped L/S	3023
GGOEGAED134810	Google Heathered Pom Beanie	2925
GGOEGHPC120810	Google Leather Strap Hat Blue	2924

Insight: Apparel and accessories dominate cart additions (caps, tees, hoodies, joggers). Google-branded lifestyle merchandise has strong appeal. These could be focus areas for **marketing campaigns** and **cross-sell strategies**.

4. "Using the GA4 sample ecommerce dataset, write a SQL query to analyze the Average number of pageviews by purchaser type" ?

```
with page_data as(
select user_pseudo_id,
       countif(event_name = "page_view") as total_page_views,
       countif((event_name IN ('in_app_purchase','purchase'))) as
purchase_event_count
from `bigquery-public-data.ga4_obfuscated_sample_ecommerce.events_*`
group by 1)

select (purchase_event_count > 0) as purchaser,
       count(*) as total_users,
       sum(total_page_views) / count(*) as avg_page_views
from page_data
group by 1
```

purchaser	total_users	avg_page_views
FALSE	265735	4.222627806
TRUE	4419	51.66960851

Insight: Purchasers are **12x more engaged**, browsing extensively before making a purchase. This suggests **deep product discovery behavior**. Strategies like **personalized recommendations** and **faster product findability** could encourage conversions among non-purchasers.

5."Using the GA4 sample ecommerce dataset, write a SQL query to analyze the Average number of transactions per purchaser.

```
select count(*) as total_transactions,
       count(distinct user_pseudo_id) as unique_customers,
       round(count(*)/count(distinct user_pseudo_id),2) as average_per_person
from `bigquery-public-data.ga4_obfuscated_sample_ecommerce.events_*`
where event_name IN ('in_app_purchase','purchase')
```

total_transactions	unique_customers	average_per_person
5692	4419	1.29

Insight: Purchasers mostly buy **once**, with limited repeat transactions. Customer retention strategies (loyalty programs, targeted remarketing, discounts for repeat buyers) could boost **customer lifetime value (CLV)**.

6."Using the GA4 sample ecommerce dataset, write a SQL query to get the 'Total User' count and 'New User' count" ?

```
with new_user as(
SELECT user_pseudo_id,
MAX(IF(event_name IN ('first_visit', 'first_open'), 1, 0)) AS is_new_user
FROM `bigquery-public-data.ga4_obfuscated_sample_ecommerce.events_*`
WHERE _TABLE_SUFFIX BETWEEN '20201101' AND '20201130'
group by 1)

select count(*) as total_users,
       sum(is_new_user) as new_users
from new_user
```

total_users	new_users
79421	71734

Insight: A very high **proportion of new users** suggests strong acquisition campaigns in this period. However, retention might be weak if most users don't return — **measuring repeat activity in later months is crucial**.

7."Using the GA4 sample ecommerce dataset, write a SQL query to get the individual timestamps and values for all 'purchase' events."

```
select TIMESTAMP_MICROS(event_timestamp) as time_stamp,
       COALESCE(value.int_value,value.float_value,value.double_value) as
revenue_value
FROM
`bigquery-public-data.ga4_obfuscated_sample_ecommerce.events_*`,UNNEST(event_params
)
WHERE event_name = 'purchase'
AND KEY = "value"
order by revenue_value desc
```

time_stamp	revenue_value
2020-12-14 14:52:48.275159 UTC	1,530.40
2021-01-19 01:13:12.172903 UTC	1,200
2020-11-13 00:11:31.451611 UTC	1,170
2020-11-02 17:39:12.669361 UTC	997.5
2021-01-14 15:31:46.589580 UTC	864.8
2020-11-30 22:59:51.669717 UTC	856.88
2020-12-10 10:25:37.529528 UTC	840
2020-11-26 01:48:35.656786 UTC	805.6
2020-12-14 04:26:15.292661 UTC	784.32
2020-11-08 14:00:45.369847 UTC	717.6
2020-12-21 16:00:44.743693 UTC	711.84
2020-12-16 21:19:30.442826 UTC	683.24
2020-12-08 19:44:57.617650 UTC	673.12
2020-12-07 07:43:16.693997 UTC	671.2
2020-12-23 12:36:35.986453 UTC	656
2020-12-31 00:44:23.654977 UTC	607.2
2020-11-11 06:48:44.165767 UTC	604.8
2020-12-16 00:04:24.380100 UTC	600
2020-11-25 04:43:07.967900 UTC	587.84

Insight: Purchases show **wide value variability**, with some **premium product sales**. This suggests opportunities for **tiered customer segmentation** (budget vs premium buyers).

8."Using the GA4 sample ecommerce dataset, write a SQL query to get the total event value for all 'purchase' events.(Total revenue in $)”

```
select SUM(COALESCE(value.int_value,value.float_value,value.double_value)) as
revenue_value
FROM
`bigquery-public-data.ga4_obfuscated_sample_ecommerce.events_*`,UNNEST(event_params
)
WHERE event_name = 'purchase'
AND KEY = "value"
```

revenue_value
362,203.66

Insight: Revenue over ~3 months indicates an **average of ~$120K/month**. When combined with **conversion rates** and **transaction frequency**, this provides a baseline for **growth forecasting**.

---

9. "Using the GA4 sample ecommerce dataset, write a SQL query to get the Products purchased by customers who purchased a specific product.

```
With PurchaseEvents AS (
SELECT user_pseudo_id,
      items
FROM `bigquery-public-data.ga4_obfuscated_sample_ecommerce.events_*`
WHERE event_name = 'purchase'),

specific_customer as (
select distinct user_pseudo_id
from PurchaseEvents,UNNEST(items) as i
where i.item_name = 'Google Navy Speckled Tee')

select i.item_name as item_name,
       sum(i.quantity) as item_quantity
from PurchaseEvents,
     UNNEST(items) as i
where user_pseudo_id in (select distinct user_pseudo_id from specific_customer)
AND i.item_name != 'Google Navy Speckled Tee'
group by 1
order by 2 desc
```

item_name	item_quantity
Google Red Speckled Tee	37
Google 24oz Ring Bottle Blue	32
Google Pen White	28
Google Metallic Notebook Set	23
Google Heather Green Speckled Tee	18
Google Unisex Eco Tee Black	16
Maze Pen	15
Google Clear Pen 4-Pack	15
Google Youth FC Tee Charcoal	15
Google Sunglasses	14
Google Leather Strap Hat Blue	14
Google Dino Game Tee	14
Android Small Trace Journal Black	14
YouTube Leather Strap Hat Black	13
Google Leather Strap Hat Black	12
Google F/C Longsleeve Ash	12

Google Toddler FC Tee Charcoal	12
YouTube Jotter Task Pad	11
Google Camp Mug Ivory	11

Insight:

- Customers who bought the *Google Navy Speckled Tee* often purchased **other T-shirts in similar styles (Red/Green Speckled Tees)**, suggesting strong **style/color substitution or bundling opportunity**.

- Non-apparel items like **bottles, pens, and notebooks** also feature, showing that these customers are **cross-shopping lifestyle merchandise**.

---

10 ."Using the GA4 sample ecommerce dataset, write a SQL query to get the Average amount of money spent per purchase session by user" ?

```
with final as (
select user_pseudo_id,
       event.value.int_value as session_id,

COALESCE(spend.value.int_value,spend.value.float_value,spend.value.double_value) as
expenditure
from
`bigquery-public-data.ga4_obfuscated_sample_ecommerce.events_*`,UNNEST(event_params
) as event,UNNEST(event_params) as spend
where event.key = "ga_session_id"
AND spend.key = "value"
AND event_name = 'purchase')
```

```sql
select user_pseudo_id,
       count(distinct session_id) as total_sessions,
       sum(final.expenditure)/count(distinct session_id) as avg_purchase_session
from final
where session_id is not null
group by 1
order by 3 desc
```

user_pseudo_id	total_sessions	average_purchase_session
81036435.12	1	1,530.40
15046526.46	1	1,424
7134559299	1	1,200
12306432.79	1	1,170
5645225.692	1	1,121.44
51248671.81	1	997.5
41474740.1	1	903.2
2566418.554	1	864.8
39388951.37	1	805.6
2278219.039	1	784.32
3833171.618	1	771.68
1621120.106	1	757.6
4484520.521	1	748.16
63177378.77	1	744.32
1033085.692	1	711.84
7019543.715	1	708
2768703.487	1	704
1532927.64	1	683.24
13141816.13	1	680
26064119.03	1	678
6822684.163	1	671.2
40810614.7	1	656
30220372.76	1	645.12
60695124.47	2	630
33534334.3	1	617.92
87100826.21	1	604.8
5766379.666	1	600

48879002.21	1	587.84
37930099.38	1	576

Insight:

- Spending distribution is **heavily skewed**: a few customers make very high-value purchases, while the majority have single, smaller transactions.

- **Premium customers** drive a disproportionate share of revenue. Identifying and retaining them is critical (loyalty perks, exclusive offers).

- Low repeat purchases reinforce earlier findings (from Query 5) — **retention and re-engagement strategies** are needed to improve average revenue per customer.