DATA VALIDATION PROCESS FOLLOWED:

SALES:

Customer 1:

First go and check in the old dashboard I made - This customer does not exist. Why?

```
#Sales for Customer 1
SELECT MP_SUP_KEY AS SELLER_ID ,
PURCHASE_DAY AS DATE_OF_PURCHASE,
SUM(ITEM_PRICE * QUANTITY) OVER (PARTITION BY PURCHASE_DAY,MP_SUP_KEY )
AS SALES,
ORDER_STATUS,
SUM(ITEM_PRICE * QUANTITY) OVER (PARTITION BY PURCHASE_DAY) AS
TOTAL_SALES
FROM `bigqueryexport-183608.amazon.customer_order_metrics`
WHERE EXTRACT(YEAR FROM purchase_date) IN (2022, 2023)  AND MP_SUP_KEY=
'd3727236-53ab-45fc-81a2-bba077c33074'
ORDER BY DATE_OF_PURCHASE DESC;
```

But there in the main table..

The sales channel, order status filter added prevents it from getting added in the dashboard…

What I have understood yet:

- I had added a filter on the order_status - completed and delivered to buyer in the old dashboard query. I do not think that is what Amazon dashboard gives. It is giving for all order_statuses. Asked Gregor , his response- they count all orders that you mentioned, only excluded are pending and cancelled. WIP for verification

```
#In general checking order history to manually look how sales is being calculated in the above
query
SELECT *
FROM `bigqueryexport-183608.amazon.customer_order_metrics`
WHERE EXTRACT(YEAR FROM purchase_date) IN (2022, 2023)  AND MP_SUP_KEY=
'd3727236-53ab-45fc-81a2-bba077c33074'
ORDER BY purchase_day ASC;
```

Analysis when I add amazon_order_id as well in my query:

111-9188815-7497035 - order id for customer 1 has 5 records.
Sales for that is 5 - let us see what is the sum - 183.75 , but in my data it gets added 5 times..
Should get added only once

The way to deal with this- order_id no more needed

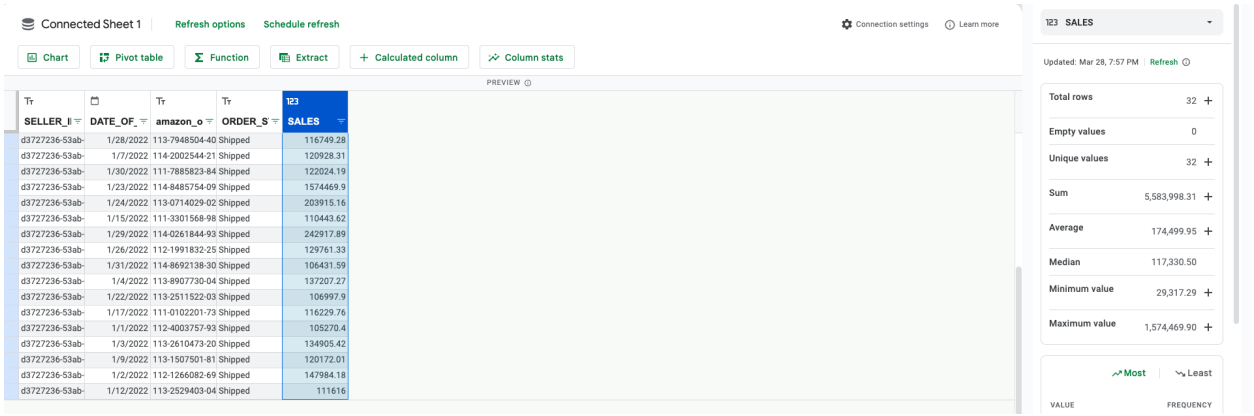
Also in the initial partition , what is happening is all records within the partition are coming
So we will have to select only one row from that

Running the query to find records day wise for the whole month

```
SELECT SELLER_ID ,DATE_OF_PURCHASE, amazon_order_id, ORDER_STATUS, SALES,
FROM
(SELECT MP_SUP_KEY AS SELLER_ID ,purchase_day AS
DATE_OF_PURCHASE,amazon_order_id, ORDER_STATUS,
SUM(ITEM_PRICE * QUANTITY) OVER (PARTITION BY PURCHASE_DAY, MP_SUP_KEY )
AS SALES,
ROW_NUMBER() OVER (PARTITION BY PURCHASE_DAY, MP_SUP_KEY) as rn
FROM `bigqueryexport-183608.amazon.customer_order_metrics`
WHERE EXTRACT(YEAR FROM purchase_date) IN (2022) AND EXTRACT(MONTH FROM
purchase_date) IN (1) AND MP_SUP_KEY= 'd3727236-53ab-45fc-81a2-bba077c33074' AND
ORDER_STATUS NOT IN ('Pending', 'Cancelled') )
WHERE rn=1
order by DATE_OF_PURCHASE;
```

Output is 32 records - so basically 32 days of the month it is showing

Checking the column statistics on Google sheet for the sum of sales



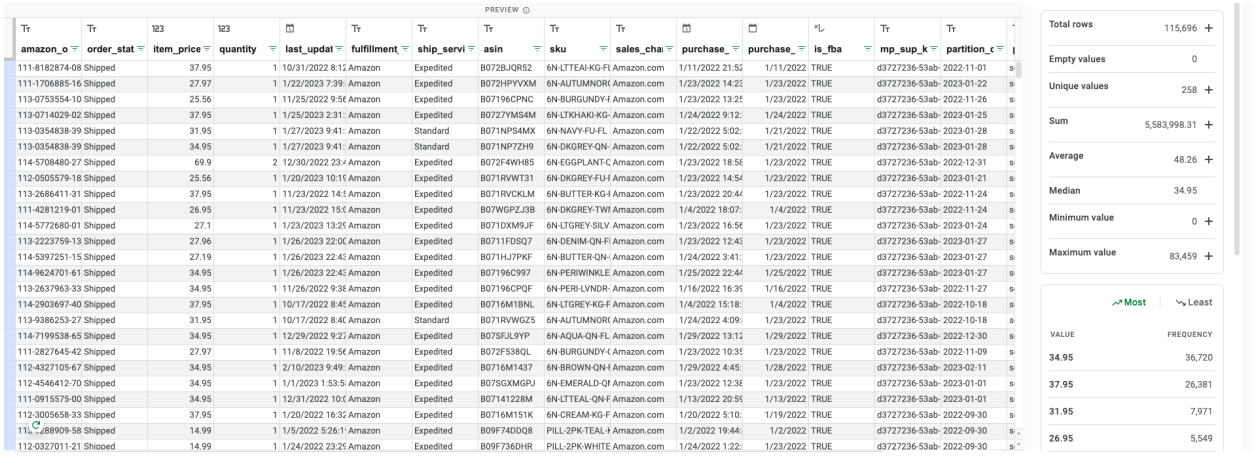
OBSERVE THAT SUM of SALES for the month of JAN - 5583998.31

Now to verify , let us select all records without group by partition for JAN 2022

Query:

```
SELECT * FROM `bigqueryexport-183608.amazon.customer_order_metrics`
WHERE EXTRACT(YEAR FROM purchase_date) IN (2022) AND EXTRACT(MONTH FROM
purchase_date) IN (1) AND MP_SUP_KEY= 'd3727236-53ab-45fc-81a2-bba077c33074' AND
ORDER_STATUS NOT IN ('Pending', 'Cancelled');
```

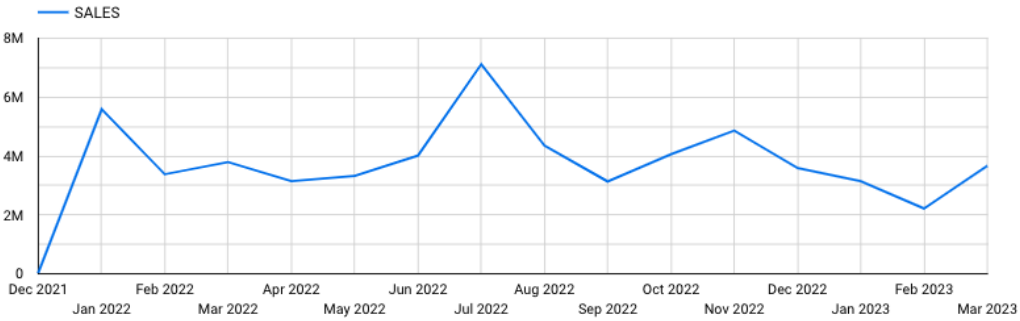
Now in google sheets first making the column SALES , and then checking the column stats:



OBSERVE THAT SUM of SALES for the month of JAN - 5583998.31

Here also the same. SO NO ISSUE IN QUERYING. That is what the data is.

customer_order_metrics



	Date	SALES
1.	Dec 2021	29,317.29
2.	Jan 2022	5,589,613.83
3.	Feb 2022	3,380,584.24
4.	Mar 2022	3,793,398.42
5.	Apr 2022	3,149,338.41
6.	May 2022	3,328,625.39
7.	Jun 2022	4,015,017.83
8.	Jul 2022	7,115,516.34
9.	Aug 2022	4,253,211.24

1 - 16 / 16

SAME ON MY CUSTOM DASHBOARD

SO CONCLUSION:

My query is correct for the data in table for this customer and the grouping by month dashboard too. But it is still different than the one shown in Amazon.

Also our outputs are NOW in the same range as the given sales dashboard (not as inflated as earlier) but still a little more

LET US ONCE CHECK IF ANY OTHER FILTERS EXIST. And try playing with those filters

Technically the only filter is of order_Status so it may be possible that it is also not considering other order statuses. Let us see which distinct order statuses are even there in this query -

Shipped

So that is also not an issue.

Changing purchase_Date to last_updated in the query–

For last_updated -62603 records
For purchase_Date - 62391 records

SO THERE IS A DIFFERENCE….. Let us maybe try plotting by that date maybe ? Maybe amazon uses that ? – NO

Diving deeper

Row	amazon_order_id	order_status	item_price	quantity	last_updated_date	fulfillment_channel	ship_servi
1	114-3391066-4321054	Shipped	26.95	1	2023-01-06 09:41:23 UTC	Amazon	Expedited
2	113-9942210-0173820	Shipped	31.95	1	2023-02-11 08:03:51 UTC	Amazon	Expedited
3	114-3391066-4321054	Shipped	26.95	1	2023-01-03 18:01:21 UTC	Amazon	Expedited

Record 1 and 3 same but just have last updated_date - so kind of a duplicate … BUT HOW TO WORK AROUND THIS NOW

Rownumber including amazon_orderid

SELLER_I	DATE_OF_	amazon_order_id	ORDER_S	SALES	rn
d3727236-53ab-	4/15/2022	114-3391066-4321054	Shipped	80.85	1
d3727236-53ab-	4/15/2022	114-3391066-4321054	Shipped	80.85	2
d3727236-53ab-	4/15/2022	114-3391066-4321054	Shipped	80.85	3

Let us add amazon_oder_id also in the partition and check so more duplicate values removed

```
SELECT SELLER_ID ,DATE_OF_PURCHASE, amazon_order_id, ORDER_STATUS, SALES
FROM
(SELECT MP_SUP_KEY AS SELLER_ID ,purchase_day AS DATE_OF_PURCHASE,
amazon_order_id, ORDER_STATUS,
SUM(ITEM_PRICE * QUANTITY) OVER (PARTITION BY PURCHASE_DAY, MP_SUP_KEY,
amazon_order_id ) AS SALES,
ROW_NUMBER() OVER (PARTITION BY PURCHASE_DAY, MP_SUP_KEY,
amazon_order_id) as rn
FROM `bigqueryexport-183608.amazon.customer_order_metrics`
WHERE EXTRACT(YEAR FROM purchase_date) IN (2022) AND EXTRACT(MONTH FROM
purchase_date) IN (4) AND MP_SUP_KEY= 'd3727236-53ab-45fc-81a2-bba077c33074' AND
ORDER_STATUS NOT IN ('Pending', 'Cancelled'))
WHERE rn=1 ;
```

Sales for April did go down to 3126920 … That means more duplicates removed!!

```
SELECT * FROM `bigqueryexport-183608.amazon.customer_order_metrics`
WHERE EXTRACT(YEAR FROM purchase_date) IN (2022) AND EXTRACT(MONTH FROM purchase_date) IN (4) AND MP_SUP_KEY= 'd3727236-53ab-45fc-81a2-bba077c33074' AND ORDER_STATUS NOT IN ('Pending', 'Cancelled')
and amazon_order_id= '111-0003254-7021861';
```

Press Alt+F1 for Accessibility Options

Query results

SAVE RESULTS | EXPLORE DATA

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH | PREVIEW

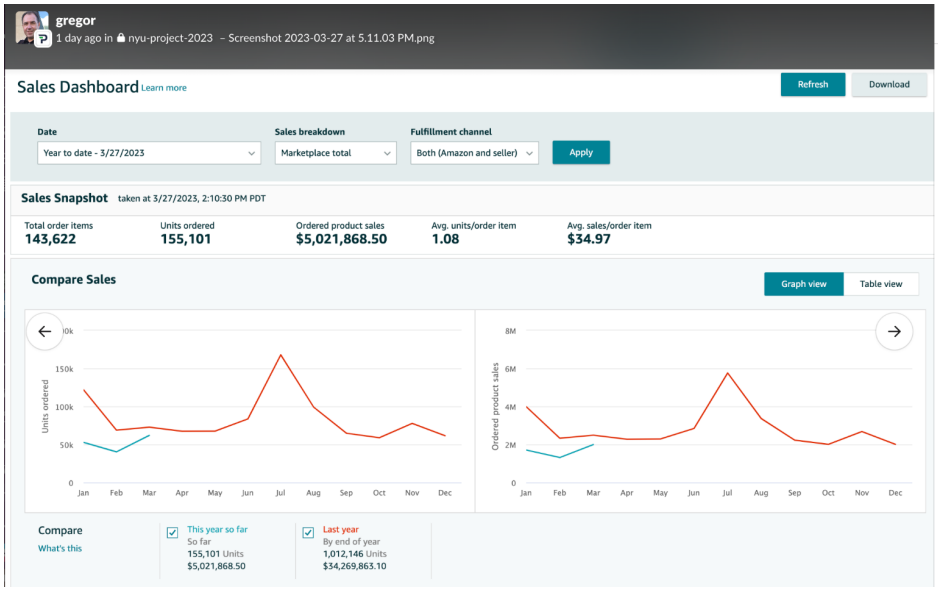
Row	amazon_order_id	order_status	item_price	quantity	last_updated_date	fulfillment_channel	ship_service_level	asin	sku
1	111-0003254-7021861	Shipped	26.95	1	2022-04-04 04:35:27 UTC	Amazon	Expedited	B07WHS1MKG	6N-LTPLUM-TWN-FL
2	111-0003254-7021861	Shipped	34.95	1	2022-04-04 04:35:27 UTC	Amazon	Expedited	B071HJ7PKF	6N-BUTTER-QN-FL
3	111-0003254-7021861	Shipped	34.95	1	2022-04-04 04:35:27 UTC	Amazon	Expedited	B07196BWNV	6N-LIME-QN-FL

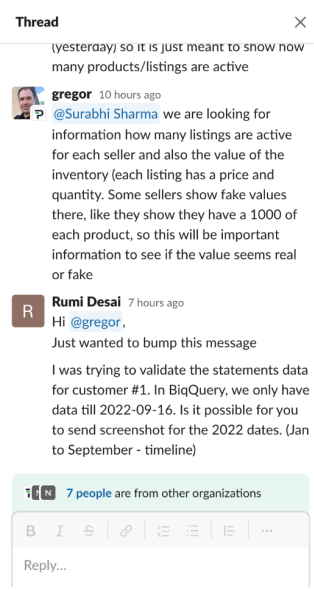
Randomly checking one order with three values if addition is correct with row. Correct as it shows

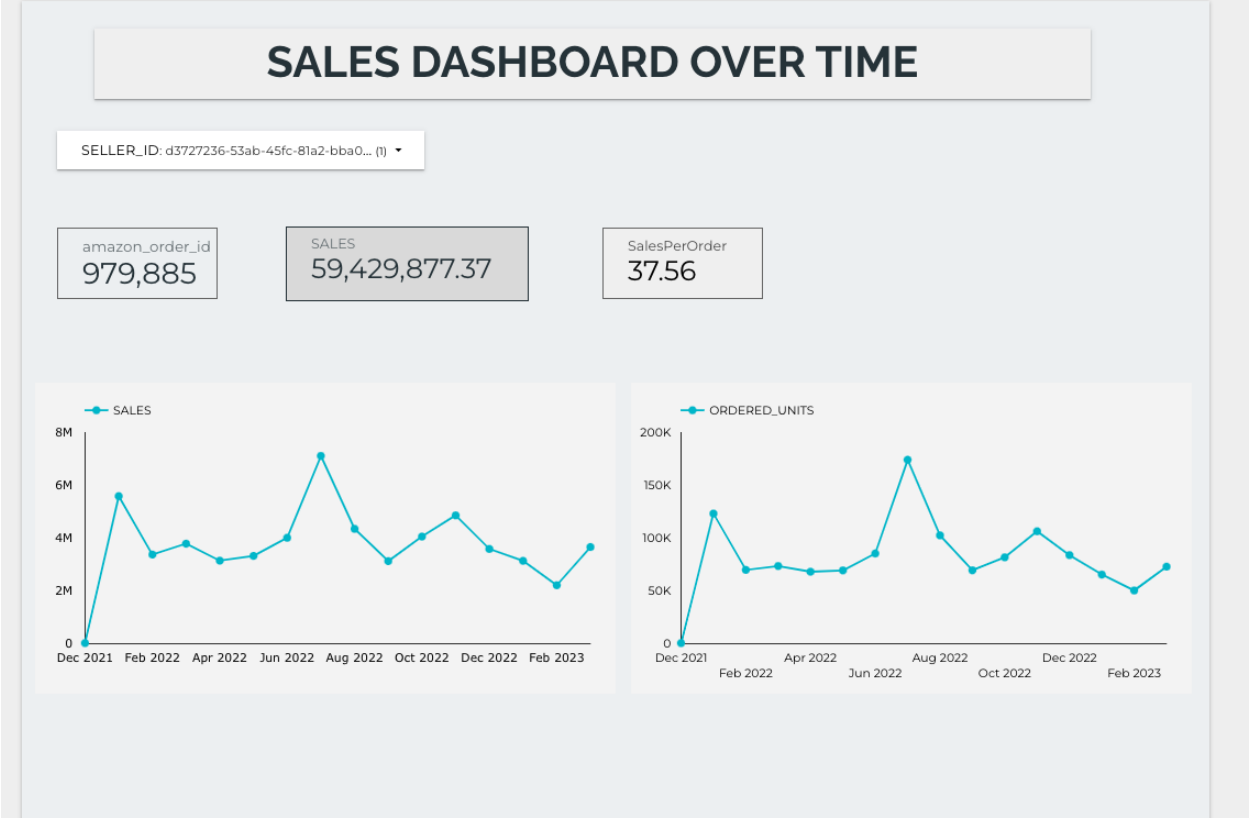
Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH PREVIEW

Row	SELLER_ID	DATE_OF_PURCH	amazon_order_id	ORDER_STATUS	SALES	rn
1	d3727236-53ab-45fc-81a2-bba...	2022-04-02	111-0003254-7021861	Shipped	96.85	1
2	d3727236-53ab-45fc-81a2-bba...	2022-04-02	111-0003254-7021861	Shipped	96.85	2
3	d3727236-53ab-45fc-81a2-bba...	2022-04-02	111-0003254-7021861	Shipped	96.85	3







At this point, the dashboard looks very similar to the one Gregor sent. Lots of data cleaned and trend patterns as you can see very similarly followed with much lesser error difference.
Still need to work on some data discrepancies and confirm with Gregor because I do not think more duplicate data exists and no cleaning is required further.

Adding more properties to the dashboard:

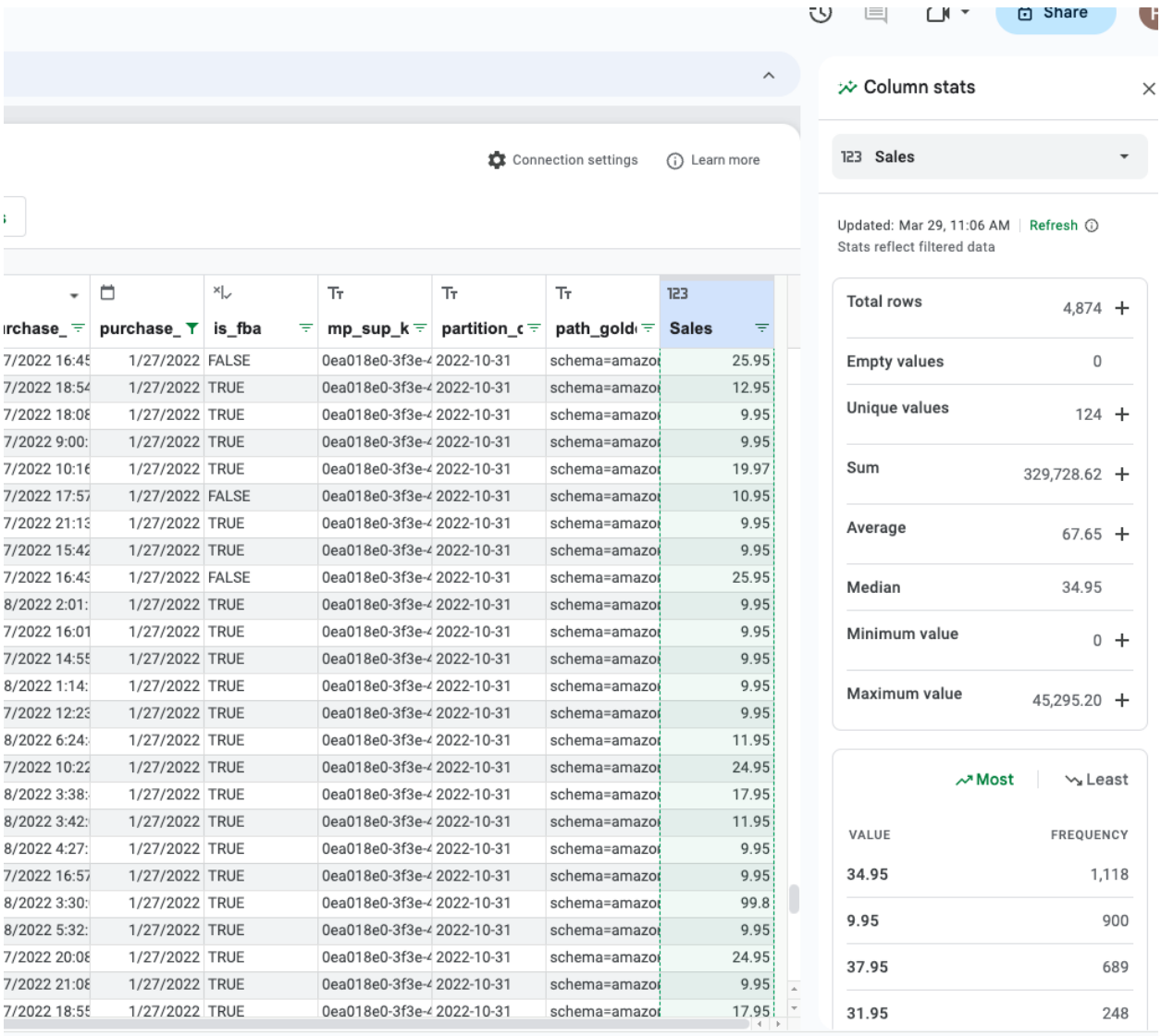
Last sprint I had also added the total sales + total orders to guage the performance of one seller with all the sellers. Adding that property in the query and checking for two sellers:

```
SELECT SELLER_ID ,DATE_OF_PURCHASE, amazon_order_id, ORDER_STATUS, SALES, TOTAL_SALES,
TOTAL_ORDERS, ORDERS
FROM
(SELECT MP_SUP_KEY AS SELLER_ID ,purchase_day AS DATE_OF_PURCHASE, amazon_order_id,
ORDER_STATUS,
SUM(ITEM_PRICE * QUANTITY) OVER (PARTITION BY PURCHASE_DAY, MP_SUP_KEY,
amazon_order_id ) AS SALES,
SUM(ITEM_PRICE * QUANTITY) OVER (PARTITION BY PURCHASE_DAY) AS TOTAL_SALES,
COUNT(amazon_order_id) OVER (PARTITION BY PURCHASE_DAY) AS TOTAL_ORDERS,
```

```
COUNT(amazon_order_id) OVER (PARTITION BY PURCHASE_DAY, MP_SUP_KEY , amazon_order_id )
AS ORDERS,
ROW_NUMBER() OVER (PARTITION BY PURCHASE_DAY, MP_SUP_KEY, amazon_order_id) as rn
FROM `bigqueryexport-183608.amazon.customer_order_metrics`
WHERE  EXTRACT(YEAR FROM purchase_date) IN (2022) AND EXTRACT(MONTH FROM
purchase_date) IN (1) AND MP_SUP_KEY IN ('d3727236-53ab-45fc-81a2-bba077c33074',
'0ea018e0-3f3e-4132-ae5b-cf119d480528') AND ORDER_STATUS NOT IN ('Pending',
'Cancelled'))
WHERE rn=1;
```

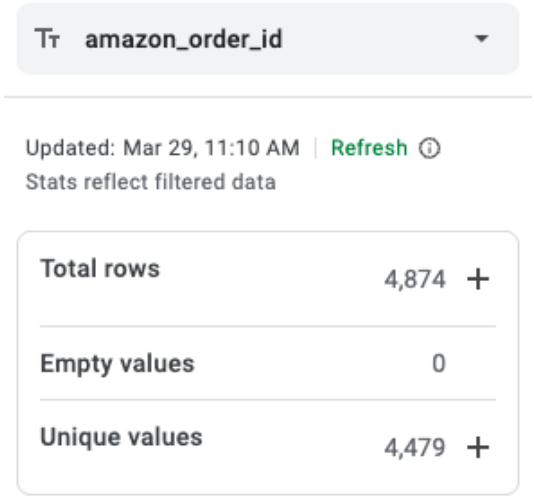
To verify the data, checking all records in the actual table and seeing how it is translating to the query

```
SELECT *
FROM `bigqueryexport-183608.amazon.customer_order_metrics`
WHERE EXTRACT(YEAR FROM purchase_date) IN (2022) AND EXTRACT(MONTH FROM purchase_date)
IN (1) AND MP_SUP_KEY IN ('d3727236-53ab-45fc-81a2-bba077c33074',
'0ea018e0-3f3e-4132-ae5b-cf119d480528') AND ORDER_STATUS NOT IN ('Pending',
'Cancelled');
```



Added filter of one day for both the sellers and checked the sales- as you can see 329728.62 and that is shown in the results above the first query too, so total_Sales is working.

Lets check orders.



Same as the one on the main query.

But the issue is that this will still come many times so we will have to add rowNum for this partition too

Ran the average one for this , added this data on BigQuery to add a table in the dashboard

Issue - only first two months coming - why ? - Forgot to remove month filter while adding data on Looker Studio. Fixed and now works for all months

<u>Statements Table:</u>

<u>Customer 1:</u>

```
#Statements for customer 1
SELECT financial_event_group_id, financial_event_group_start, financial_event_group_end, original_total_amount,processing_status,beginning_balance_amount
FROM `bigqueryexport-183608.amazon.statements`
WHERE EXTRACT(YEAR FROM financial_event_group_start) IN (2022, 2023) AND MP_SUP_KEY= 'd3727236-53ab-45fc-81a2-bba077c33074'
ORDER BY financial_event_group_start DESC, financial_event_group_end DESC;
```

Press Alt+F1 for Accessi

Query results SAVE RESULTS EXPLORE DATA

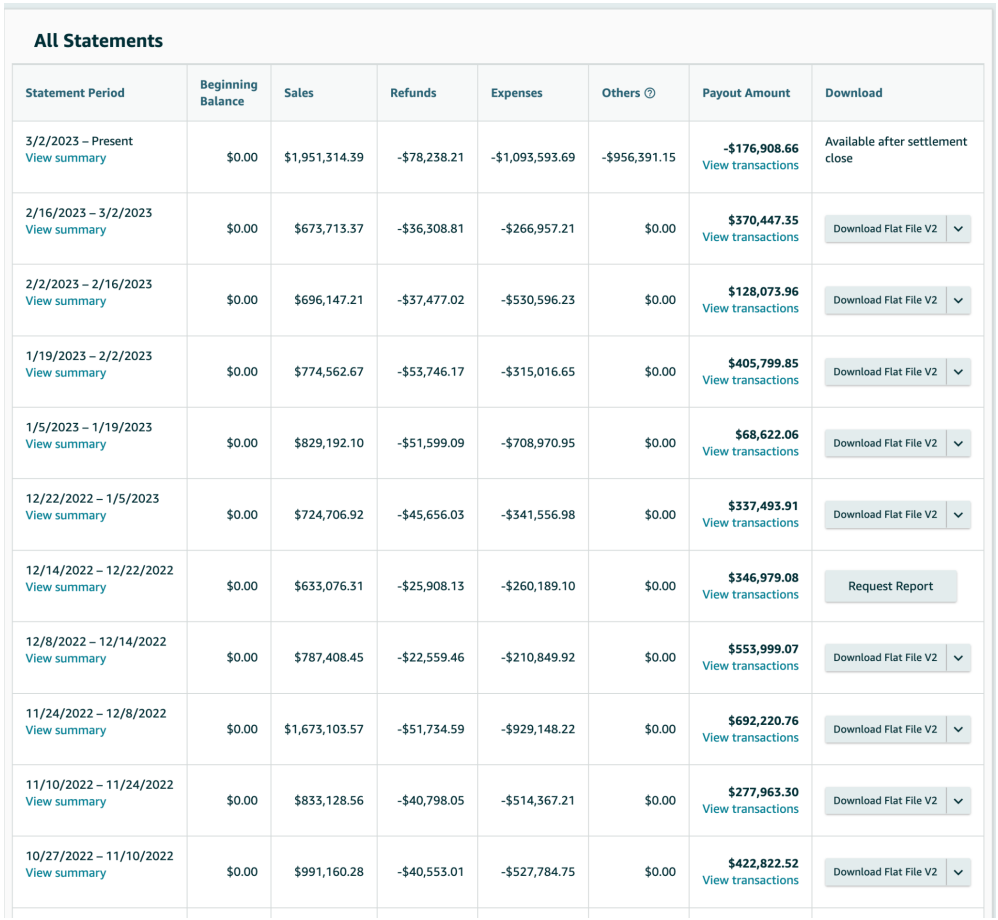
JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH PREVIEW

Row	financial_event_group_id	financial_event_group_start	financial_event_group_end	original_total_amount	processing_status	beginning_balan
1	j5JDfDxpo9erz3PNYM7Iy_CN7…	2022-09-16 06:01:26.563000 UTC	2022-09-16 19:25:46.005580 UTC	22.73	Open	0.0
2	9zMl5OW0w2FBWinfuuli3npH…	2022-09-15 22:12:29.851000 UTC	2022-09-16 19:25:46.005617 UTC	1512.51	Open	0.0
3	P2sWP-XfDq-8xLPYKeisc3quF…	2022-09-15 17:22:21.296000 UTC	2022-09-16 19:25:46.005627 UTC	49940.12	Open	0.0
4	TZ0RPKQOI9UHC1sEkkp7ChfP…	2022-09-13 15:11:39.381000 UTC	2022-09-15 17:22:21.296000 UTC	103881.52	Closed	0.0
5	VtqLkFEfSOs-VqzRZ_t887n5-U…	2022-09-07 17:32:45.236000 UTC	2022-09-16 19:25:46.005643 UTC	510.91	Open	0.0
6	H62HgaSE0DAQLamPJzXEiUs…	2022-09-02 06:01:26.538000 UTC	2022-09-16 06:01:26.563000 UTC	1315.27	Closed	0.0
7	js9vIpIh54J9YqdKfv8ZDqD4HE…	2022-09-01 22:12:30.080000 UTC	2022-09-15 22:12:29.851000 UTC	22804.34	Closed	0.0
8	UQkG7xkHi-RrTq_9kNe9-T34Y…	2022-09-01 17:22:21.711000 UTC	2022-09-13 15:11:39.381000 UTC	1056243.72	Closed	0.0
9	ZXD-aUv5gvlX6dZRLbaDPhOR…	2022-08-24 17:32:45.585000 UTC	2022-09-07 17:32:45.236000 UTC	655.3	Closed	0.0
10	ZQHLdDASDPQPic9euObqPAC…	2022-08-19 06:01:26.918000 UTC	2022-09-02 06:01:26.538000 UTC	1514.8	Closed	0.0
11	D1f81PvrbRfeeDpV2Vu8j5ivFjU…	2022-08-18 22:12:29.231000 UTC	2022-09-01 22:12:30.080000 UTC	24564.65	Closed	0.0
12	9P9VK5sBgWSvuVxFEzSd1kEy…	2022-08-18 17:22:21.681000 UTC	2022-09-01 17:22:21.711000 UTC	955434.64	Closed	0.0
13	grYyJR2Mum5G2Ctji8AZTt9hO…	2022-08-10 17:32:45.597000 UTC	2022-08-24 17:32:45.585000 UTC	0.0	Closed	0.0
14	tfMcMiola-7hQo2KlZYkMSRHU…	2022-08-05 06:01:26.053000 UTC	2022-08-19 06:01:26.918000 UTC	1351.24	Closed	0.0
15	rndfsdaaV4CjKCWD1kMK5WM…	2022-08-04 22:12:29.555000 UTC	2022-08-18 22:12:29.231000 UTC	15236.0	Closed	0.0
16	4sM60RmcttS7sFEt0EjIdh8DeI…	2022-08-04 17:22:21.903000 UTC	2022-08-18 17:22:21.681000 UTC	594803.35	Closed	0.0

Data shown on the dashboard is for different dates. Asked Gregor for the timeline I needed. No response yet.

Apart from this- statements dashboard given by Amazon is way different than what I was asked to prepare. Confirm in the upcoming meeting about the same.

Financial start and end date in the table is of one one day. Should we have fixed 15 day cycle and add all the financial start and end date that belong to that period? - Understand from Payability



All Statements

Statement Period	Beginning Balance	Sales	Refunds	Expenses	Others	Payout Amount	Download
3/2/2023 – Present View summary	$0.00	$1,951,314.39	-$78,238.21	-$1,093,593.69	-$956,391.15	-$176,908.66 View transactions	Available after settlement close
2/16/2023 – 3/2/2023 View summary	$0.00	$673,713.37	-$36,308.81	-$266,957.21	$0.00	$370,447.35 View transactions	Download Flat File V2
2/2/2023 – 2/16/2023 View summary	$0.00	$696,147.21	-$37,477.02	-$530,596.23	$0.00	$128,073.96 View transactions	Download Flat File V2
1/19/2023 – 2/2/2023 View summary	$0.00	$774,562.67	-$53,746.17	-$315,016.65	$0.00	$405,799.85 View transactions	Download Flat File V2
1/5/2023 – 1/19/2023 View summary	$0.00	$829,192.10	-$51,599.09	-$708,970.95	$0.00	$68,622.06 View transactions	Download Flat File V2
12/22/2022 – 1/5/2023 View summary	$0.00	$724,706.92	-$45,656.03	-$341,556.98	$0.00	$337,493.91 View transactions	Download Flat File V2
12/14/2022 – 12/22/2022 View summary	$0.00	$633,076.31	-$25,908.13	-$260,189.10	$0.00	$346,979.08 View transactions	Request Report
12/8/2022 – 12/14/2022 View summary	$0.00	$787,408.45	-$22,559.46	-$210,849.92	$0.00	$553,999.07 View transactions	Download Flat File V2
11/24/2022 – 12/8/2022 View summary	$0.00	$1,673,103.57	-$51,734.59	-$929,148.22	$0.00	$692,220.76 View transactions	Download Flat File V2
11/10/2022 – 11/24/2022 View summary	$0.00	$833,128.56	-$40,798.05	-$514,367.21	$0.00	$277,963.30 View transactions	Download Flat File V2
10/27/2022 – 11/10/2022 View summary	$0.00	$991,160.28	-$40,553.01	-$527,784.75	$0.00	$422,822.52 View transactions	Download Flat File V2