

SPRINT 8 : PRODUCTS

1.Data Dictionary for Products

2. EDA

```
SELECT * FROM `bigqueryexport-183608.amazon.products`  
LIMIT 10;
```

```
SELECT DISTINCT id_type FROM `bigqueryexport-183608.amazon.products`; (ASIN , )  
SELECT DISTINCT PRODUCT_GROUP FROM `bigqueryexport-183608.amazon.products`;
```

- H categories. Kitchen, Home, Baby Product , Movie, Beauty etc

```
SELECT DISTINCT product_type_name FROM `bigqueryexport-183608.amazon.products`;
```

- 1894 categories - backpack , sneakers , suitcase etc.

ID_TYPE - ASIN

ID - ASIN NUMBER

Can check product group + product type name

Things to check in general:

1. Listing Amount

2. Path_golden contains seller key ? if yes then no need to join maybe with the COM table

```
SELECT id,product_group, path_golden FROM `bigqueryexport-183608.amazon.products`  
LIMIT 10;
```

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW
Row	id	product_group	path_golden			
1	B07T53FMZM	Home Improvement	schema=amazon/table=products/date=2022-03-28/B07T53FMZM.csv			
2	0615244548	Book	schema=amazon/table=products/date=2022-03-28/0615244548.csv			
3	B092MXBYMF	Sports	schema=amazon/table=products/date=2022-03-28/B092MXBYMF.csv			
4	B07Z1K17L3	Toy	schema=amazon/table=products/date=2022-03-28/B07Z1K17L3.csv			
5	B07TKP8FST	Automotive Parts and Access...	schema=amazon/table=products/date=2022-03-28/B07TKP8FST.csv			
6	B071JS2TNV	Jewelry	schema=amazon/table=products/date=2022-03-28/B071JS2TNV.csv			
7	B07VPLJ9YM	Furniture	schema=amazon/table=products/date=2022-03-28/B07VPLJ9YM.csv			
8	B007AM1PXW	Wireless	schema=amazon/table=products/date=2022-03-28/B007AM1PXW.csv			
9	B009313LZI	Beauty	schema=amazon/table=products/date=2022-03-28/B009313LZI.csv			
10	B0874LJ8GZ	Shoes	schema=amazon/table=products/date=2022-03-28/B0874LJ8GZ.csv			

Hene, path_golden does not contain seller information. Join with COM table to understand

```
SELECT DISTINCT EXTRACT(DATE FROM partition_date) as date_partitioned FROM
`bigqueryexport-183608.amazon.products`
ORDER BY date_partitioned DESC;
```

Row	date_partitioned
1	2022-09-28
2	2022-09-27
3	2022-09-26
4	2022-09-25
5	2022-09-24
6	2022-09-23
7	2022-09-22
8	2022-09-21
9	2022-09-20
10	2022-09-19
11	2022-09-18
12	2022-09-17
13	2022-09-16
14	2022-09-15

Joining with COM to see sales wrt product categories:

```
SELECT SELLER_ID ,DATE_OF_PURCHASE, AMAZON_ORDER_ID, SALES.ASIN, PRODUCT_GROUP,
PRODUCT_TYPE_NAME, SALES_FULFILLMENT_CHANNEL, ORDER_STATUS, SALES, ORDERED_UNITS,
FROM (SELECT SELLER_ID ,DATE_OF_PURCHASE, AMAZON_ORDER_ID, ASIN,
SALES_FULFILLMENT_CHANNEL, ORDER_STATUS, SALES, ORDERED_UNITS,
FROM
(SELECT MP_SUP_KEY AS SELLER_ID ,purchase_day AS DATE_OF_PURCHASE, amazon_order_id AS
AMAZON_ORDER_ID, fulfillment_channel as SALES_FULFILLMENT_CHANNEL, ORDER_STATUS, ASIN,
SUM(ITEM_PRICE * QUANTITY) OVER (PARTITION BY PURCHASE_DAY, MP_SUP_KEY,
amazon_order_id, partition_date ) AS SALES,
SUM(QUANTITY) OVER (PARTITION BY PURCHASE_DAY, MP_SUP_KEY , amazon_order_id,
partition_date ) AS ORDERED_UNITS,
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 (2022,2023) AND ORDER_STATUS NOT IN
('Pending', 'Cancelled'))
WHERE rn=1
) SALES
INNER JOIN
(SELECT ID_TYPE, ID AS ASIN, PRODUCT_GROUP, PRODUCT_TYPE_NAME FROM
`bigqueryexport-183608.amazon.products`
WHERE id_type='ASIN') PRODUCTS
ON SALES.ASIN = PRODUCTS.ASIN;

```

customer_order_metrics

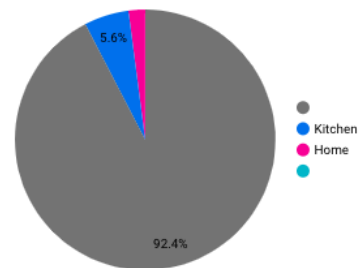
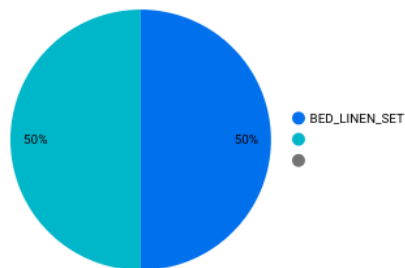
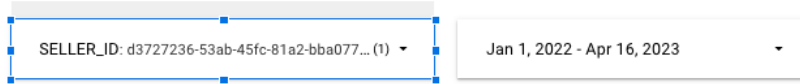


1. Validating why is the record count number different in this chart wrt to the built dashboard , even the sales value is different when comparing
2. Validating to understand why am I getting null values in the chart.

Changing INNER JOIN —> LEFT JOIN.

Record count same for d3727.. User. What about other users?

customer_order_metrics



Validating the pie chart values in explore sheets:

```
SELECT SELLER_ID ,DATE_OF_PURCHASE, AMAZON_ORDER_ID, SALES.ASIN, PRODUCT_GROUP,
PRODUCT_TYPE_NAME, SALES_FULFILLMENT_CHANNEL, ORDER_STATUS, SALES, ORDERED_UNITS,
FROM (SELECT SELLER_ID ,DATE_OF_PURCHASE, AMAZON_ORDER_ID, ASIN,
SALES_FULFILLMENT_CHANNEL, ORDER_STATUS, SALES, ORDERED_UNITS,
FROM
(SELECT MP_SUP_KEY AS SELLER_ID ,purchase_day AS DATE_OF_PURCHASE, amazon_order_id AS
AMAZON_ORDER_ID, fulfillment_channel as SALES_FULFILLMENT_CHANNEL, ORDER_STATUS, ASIN,
SUM(ITEM_PRICE * QUANTITY) OVER (PARTITION BY PURCHASE_DAY, MP_SUP_KEY,
amazon_order_id, partition_date ) AS SALES,
SUM(QUANTITY) OVER (PARTITION BY PURCHASE_DAY, MP_SUP_KEY , amazon_order_id,
partition_date ) AS ORDERED_UNITS,
```


```
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 (2022,2023) AND ORDER_STATUS NOT IN
('Pending', 'Cancelled'))
WHERE rn=1
) SALES
LEFT JOIN
(SELECT ID_TYPE, ID AS ASIN, PRODUCT_GROUP, PRODUCT_TYPE_NAME FROM
`bigqueryexport-183608.amazon.products`
WHERE id_type='ASIN') PRODUCTS
ON SALES.ASIN = PRODUCTS.ASIN
WHERE SELLER_ID= 'd3727236-53ab-45fc-81a2-bba077c33074' ;
```

RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW		
	DATE_OF_PURCH	AMAZON_ORDER_ID	ASIN	PRODUCT_GROUP	PRODUCT_TYPE_NAME	SALES_FULFILLMENT_CHANNEL
›ba...	2022-01-08	114-8126753-3601830	B09F75FRCV	null	null	Amazon
›ba...	2022-01-15	113-6641187-1578650	B07WHR6R2L	null	null	Amazon
›ba...	2022-04-10	113-0079247-9397032	B071YWT7B5	null	null	Amazon
›ba...	2022-04-11	111-2745156-5183419	B071YWT7B5	null	null	Amazon
›ba...	2022-05-06	114-5231183-8735456	B07WHR6R2L	null	null	Amazon
›ba...	2022-06-26	111-5715492-0220227	B09F714995	null	null	Amazon
›ba...	2022-07-01	113-6752083-8286625	B09F73T3LS	null	null	Amazon
›ba...	2022-07-01	114-2596490-3509846	B071YWT7B5	null	null	Amazon
›ba...	2022-07-13	114-6294630-4369015	B07WHR6R2L	null	null	Amazon
›ba...	2022-07-18	112-5574555-2584239	B09F714995	null	null	Amazon
›ba...	2022-07-23	112-7018486-1185065	B09F714995	null	null	Amazon
›ba...	2022-08-31	111-2851078-2203435	B071YWT7B5	null	null	Amazon
›ba...	2022-09-04	113-5145713-8297040	B071YWT7B5	null	null	Amazon
›ba...	2022-10-13	701-2806764-0835414	B071YWT7B5	null	null	Amazon

```
1 SELECT * FROM `bigqueryexport-183608.amazon.products`
2 WHERE id='B09F75FRCV' ;
```

Query results

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

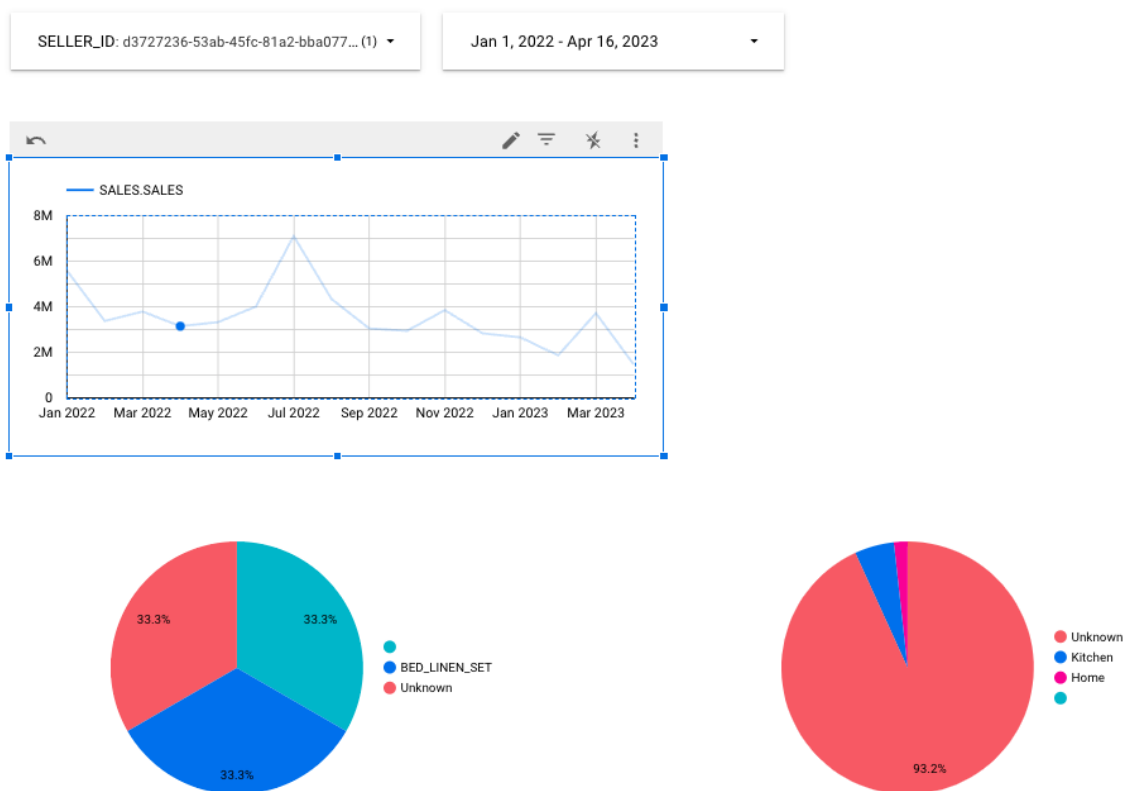
 There is no data to display.

Which is correct. THE ASIN data is basically not available in the products table. Handle it in the query.

Using IFNULL for both of them so NULL values become a category kind of that is Unknown.

Handled in the sense, we do get another category called Unknown but for some cases we still have something still. Let us check.

customer_order_metrics



If you add percentage value- adds to 100. So checking in style

By adding ASIN, is it causing issue with number of records?

No. inner query runs same.

REFUNDS:

Three tables:

1. Refund_event - No data
2. Refund_event_historical - No seller key
3. Vbq_refund_event_historical- Seller key that matches with the path_golden in Refund_event_historical so using this table for analysis

To identify the number of refunds and the amount charged wrt seller:

```
SELECT MP_SUP_KEY AS SELLER_ID, DATE(posted_date) as POSTED_DATE, amazon_order_id
AS AMAZON_ORDER_ID,
SUM(ITEM_CHARGE_AMOUNT) OVER (PARTITION BY MP_SUP_KEY, DATE(posted_date)) AS
CHARGED_AMOUNT,
COUNT(amazon_order_id) OVER (PARTITION BY MP_SUP_KEY, DATE(posted_date)) AS
NUMBER_OF_REFUNDS,
FROM `bigqueryexport-183608.amazon.vbq_refundevent_historical`
WHERE EXTRACT(YEAR FROM posted_date) IN (2022, 2023);
```

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW
Row	SELLER_ID	POSTED_DATE	AMAZON_ORDER_ID	CHARGED_AMO	NUMBER_OF_REFUNDS	
1	003a8c69-233f-4451-98db-7ea5a952efc6	2022-04-06	112-9225728-8063417	-94.04	5	
2	003a8c69-233f-4451-98db-7ea5a952efc6	2022-04-06	112-9225728-8063417	-94.04	5	
3	003a8c69-233f-4451-98db-7ea5a952efc6	2022-04-06	112-9225728-8063417	-94.04	5	
4	003a8c69-233f-4451-98db-7ea5a952efc6	2022-04-06	112-9225728-8063417	-94.04	5	
5	003a8c69-233f-4451-98db-7ea5a952efc6	2022-04-06	112-9225728-8063417	-94.04	5	
6	0058840c-9aa0-4ba6-a29e-eed6af93adc8	2022-05-24	113-8338805-7029020	-65.1	9	
7	0058840c-9aa0-4ba6-a29e-eed6af93adc8	2022-05-24	113-8338805-7029020	-65.1	9	
8	0058840c-9aa0-4ba6-a29e-eed6af93adc8	2022-05-24	112-8726257-2347402	-65.1	9	
9	0058840c-9aa0-4ba6-a29e-eed6af93adc8	2022-05-24	112-8726257-2347402	-65.1	9	
10	0058840c-9aa0-4ba6-a29e-eed6af93adc8	2022-05-24	113-8338805-7029020	-65.1	9	
11	0058840c-9aa0-4ba6-a29e-eed6af93adc8	2022-05-24	113-8338805-7029020	-65.1	9	

Does have duplidates. Why?

```
SELECT * FROM `bigqueryexport-183608.amazon.vbq_refundevent_historical`
WHERE MP_SUP_KEY= '003a8c69-233f-4451-98db-7ea5a952efc6' AND
DATE(posted_date)='2022-04-06';
```

JOB INFORMATION		RESULTS	JSON	EXECUTIO
Row	amazon_order_id	item_charge_am	item_charg	
1	112-9225728-8063417	16.03	Commissic	
2	112-9225728-8063417	-106.86	Principal	
3	112-9225728-8063417	10.95	Marketplac	
4	112-9225728-8063417	-3.21	RefundCon	
5	112-9225728-8063417	-10.95	Tax	

The values are different. Order item id is the same. Seller order id is the same. Why? Understand. All values are the same but what do they mean?

PREVIEW												
Tr	amazon_order_id	item_chan	item_charge_type	marketplace	order_item_id	posted_date	quantity	seller_order_id	seller_sku	path_golden	time_stamp	mp_sup_key
	112-9225728-8063417	16.03	Commission	Amazon.com	02260211633218	4/6/2022 1:36:41	1	112-9225728-8063417	42.06_8004X1FQXS_34522_20220308_106	schema-amazon/earni/1650518872		003a8c69-233f-4451-98db-7ea
	112-9225728-8063417	-106.86	Principal	Amazon.com	02260211633218	4/6/2022 1:36:41	1	112-9225728-8063417	42.06_8004X1FQXS_34522_20220308_106	schema-amazon/earni/1650518872		003a8c69-233f-4451-98db-7ea
	112-9225728-8063417	-3.21	RefundCommission	Amazon.com	02260211633218	4/6/2022 1:36:41	1	112-9225728-8063417	42.06_8004X1FQXS_34522_20220308_106	schema-amazon/earni/1650518872		003a8c69-233f-4451-98db-7ea
	112-9225728-8063417	10.95	MarketplaceFacilitatorTax-Principal	Amazon.com	02260211633218	4/6/2022 1:36:41	1	112-9225728-8063417	42.06_8004X1FQXS_34522_20220308_106	schema-amazon/earni/1650518872		003a8c69-233f-4451-98db-7ea
	112-9225728-8063417	-10.95	Tax	Amazon.com	02260211633218	4/6/2022 1:36:41	1	112-9225728-8063417	42.06_8004X1FQXS_34522_20220308_106	schema-amazon/earni/1650518872		003a8c69-233f-4451-98db-7ea

End of preview

As of now, added the row number so that it comes once only.

Also understand if they need any analysis from Item_Charge_type?

Adding Row_number:

```
SELECT SELLER_ID, POSTED_DATE, AMAZON_ORDER_ID, CHARGED_AMOUNT, NUMBER_OF_REFUNDS
FROM
(SELECT MP_SUP_KEY AS SELLER_ID, DATE(posted_date) as POSTED_DATE, amazon_order_id AS
AMAZON_ORDER_ID,
SUM(ITEM_CHARGE_AMOUNT) OVER (PARTITION BY MP_SUP_KEY, DATE(posted_date)) AS
CHARGED_AMOUNT,
COUNT(amazon_order_id) OVER (PARTITION BY MP_SUP_KEY, DATE(posted_date)) AS
NUMBER_OF_REFUNDS,
ROW_NUMBER() OVER (PARTITION BY MP_SUP_KEY, DATE(posted_date), AMAZON_ORDER_ID) as rn
FROM `bigqueryexport-183608.amazon.vbq_refundevent_historical`
WHERE EXTRACT(YEAR FROM posted_date) IN (2022, 2023))
where rn=1 AND SELLER_ID= '003a8c69-233f-4451-98db-7ea5a952efc6' AND
POSTED_DATE='2022-04-06';
```

```
1 FROM bigqueryexport-183608.amazon.vbq_refundevent_historical
8 WHERE EXTRACT(YEAR FROM posted_date) IN (2022, 2023))
9 where rn=1 AND SELLER_ID= '003a8c69-233f-4451-98db-7ea5a952efc6' AND POSTED_DATE='2022-04-06';
10
```

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW
Row	SELLER_ID	POSTED_DATE	AMAZON_ORDER_ID	CHARGED_AMO	NUMBER_OF_RE	
1	003a8c69-233f-4451-98db-7ea...	2022-04-06	112-9225728-8063417	-94.04	5	

As we can see, duplication removed.

But number_of_refunds still 5. Should be 1. MAX(quantity) since quantity should be same for a given order_id

```
SELECT SELLER_ID, POSTED_DATE, AMAZON_ORDER_ID, CHARGED_AMOUNT, NUMBER_OF_REFUNDS
FROM
(SELECT MP_SUP_KEY AS SELLER_ID, DATE(posted_date) as POSTED_DATE, amazon_order_id AS
AMAZON_ORDER_ID,
SUM(ITEM_CHARGE_AMOUNT) OVER (PARTITION BY MP_SUP_KEY, DATE(posted_date)) AS
CHARGED_AMOUNT,
```

```

MAX(quantity) OVER (PARTITION BY MP_SUP_KEY, DATE(posted_date), AMAZON_ORDER_ID ) AS
NUMBER_OF_REFUNDS,
ROW_NUMBER() OVER (PARTITION BY MP_SUP_KEY, DATE(posted_date), AMAZON_ORDER_ID) as rn
FROM `bigqueryexport-183608.amazon.vbq_refundevent_historical`
WHERE EXTRACT(YEAR FROM posted_date) IN (2022, 2023))
where rn=1 and SELLER_ID= '003a8c69-233f-4451-98db-7ea5a952efc6' AND
POSTED_DATE='2022-04-06';

```

Will have to update the sum also in case of edge cases.

```

SELECT SELLER_ID, POSTED_DATE, AMAZON_ORDER_ID, CHARGED_AMOUNT, NUMBER_OF_REFUNDS
FROM
(SELECT MP_SUP_KEY AS SELLER_ID, DATE(posted_date) as POSTED_DATE, amazon_order_id AS
AMAZON_ORDER_ID,
SUM(ITEM_CHARGE_AMOUNT) OVER (PARTITION BY MP_SUP_KEY, DATE(posted_date),
AMAZON_ORDER_ID ) AS CHARGED_AMOUNT,
MAX(CAST(quantity AS INT64)) OVER (PARTITION BY MP_SUP_KEY, DATE(posted_date),
AMAZON_ORDER_ID ) AS NUMBER_OF_REFUNDS,
ROW_NUMBER() OVER (PARTITION BY MP_SUP_KEY, DATE(posted_date), AMAZON_ORDER_ID) as rn
FROM `bigqueryexport-183608.amazon.vbq_refundevent_historical`
WHERE EXTRACT(YEAR FROM posted_date) IN (2022, 2023))
where rn=1;

```

Questions:

- a. Item_charge_amount - what exactly does it denote. What breakdown should we consider for our dashboard?
- a. Item_Charge_Type - is there some significant value for which we may need analytics?

BUSINESS CONTEXT FOR THESE COLUMN VALUES

For example consider order X.

In Customer order metrics:

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW
Row	amazon_order_id	order_status	item_price	quantity	last_updated_date	
1	112-9225728-8063417	Shipped	106.86	1	2022-03-31 20:09:01 UTC	
2	112-9225728-8063417	Shipped	106.86	1	2022-03-31 20:09:01 UTC	

In refunds:

Tr	123	Tr
amazon_order_id	item_charge_amount	item_charge_type
112-9225728-8063417	16.03	Commission
112-9225728-8063417	-106.86	Principal
112-9225728-8063417	-3.21	RefundCommission
112-9225728-8063417	10.95	MarketplaceFacilitatorTax-Principal
112-9225728-8063417	-10.95	Tax