

E-commerce SQL Analysis

Problem Statement

Analyzing the sales, product, and customer data for an e-commerce company. getting various insights and calculating various KPI and data with SQL in Big Query.

- **IMP KPIs**

No. of Transaction:

```
select count (BASKET_ID) as No_of_transaction from `Ecommerce.transaction_data` join  
`Ecommerce.hh_demographic` using(household_key)
```

Row	No_of_transaction
1	713857

No. of product:

```
select count(PRODUCT_ID) as No_of_product from `Ecommerce.product`
```

Row	No_of_product
1	92353

No of Customer:

```
select count(household_key) as No_of_cust from `Ecommerce.hh_demographic`
```

Row	No_of_cust
1	801

- **Question 1:** Find the number of orders that have small, medium or large order value (small:0-10 dollars, medium:10-20 dollars, large:20+)

```
SELECT
CASE
WHEN SALES_VALUE BETWEEN 0 AND 10 THEN 'SMALL'
WHEN SALES_VALUE BETWEEN 10.01 AND 20 THEN 'MEDIUM'
ELSE 'LARGE'
END AS Order_Size,
COUNT (DISTINCT BASKET_ID) AS Order_Count
FROM `Ecommerce`.`transaction_data`
GROUP BY Order_Size;
```

Row	Order_Size ▼	Order_Count ▼
1	SMALL	219473
2	LARGE	12268
3	MEDIUM	23985

Question 2: Find the number of orders that are small, medium or large order value (small:0-5 dollars, medium:5-10 dollars, large:10+)

```
SELECT
CASE
    WHEN SALES_VALUE BETWEEN 0 AND 5 THEN 'SMALL'
    WHEN SALES_VALUE BETWEEN 5 AND 10 THEN 'MEDIUM'
    ELSE 'LARGE'
END AS Order_Size,
COUNT(DISTINCT BASKET_ID) AS Order_Count
FROM `Ecommerce`.`transaction_data`
GROUP BY Order_Size;
```

Row	Order_Size ▼	Order_Count ▼
1	SMALL	210259
2	MEDIUM	73357
3	LARGE	35260

- **Question 3:** Find top 3 stores with highest foot traffic for each week (Foot traffic: number of customers transacting)

```
WITH CTE AS (
(
SELECT WEEK_NO,STORE_ID,DENSE_RANK() OVER (PARTITION BY WEEK_NO ORDER BY
COUNT(BASKET_ID) DESC) AS RANK
FROM Ecommerce.transaction_data
GROUP BY WEEK_NO, STORE_ID)
)
select WEEK_NO,STORE_IDFROM CTE
WHERE RANK<=3
ORDER BY 1;
```

Row	WEEK_NO	STORE_ID
1	1	324
2	1	321
3	1	32004
4	2	375
5	2	292
6	2	315
7	3	367
8	3	375
9	3	356
10	4	367

- **Question 4:** Create a basic customer profiling with first, last visit, number of visits, average money spent per visit and total money spent order by highest avg money

```

select
household_key,
min(WEEK_NO) as First_visit,
max(WEEK_NO) as Last_visit,
round(avg(SALES_VALUE),2) as Avg_spent,
round(sum(SALES_VALUE),2) as Total_money_spent
from
Ecommerce.transaction_data
group by 1
order by 1

```

Row	household_key	First_visit	Last_visit	Avg_spent	Total_money_spent
1	1	8	102	2.43	2100.14
2	2	15	96	2.97	1060.18
3	3	17	101	3.02	1392.33
4	4	16	90	4.17	630.38
5	5	13	101	3.73	414.39
6	6	18	102	3.28	2942.1
7	7	4	102	2.7	1733.88
8	8	10	102	2.89	2865.76

- **Question 5:** Do a single customer analysis selecting most spending customer for whom we have demographic information (because not all customers in transaction data are present in demographic table) (show the demographic as well as total spent)

with cte as(

select household_key,round(sum(SALES_VALUE),2) as total_amt_spent

from `Ecommerce.hh_demographic` join `Ecommerce.transaction_data` using(household_key)

group by 1

order by 2 desc

limit 1

)

select * from cte join `Ecommerce.hh_demographic` using(household_key)

Row	household_key	total_amt_spent	AGE_DESC	MARITAL_STATUS_CODE	INCOME_DESC	HOMEOWNER_DESC	HH_COMP_DESC	HOUSEHOLD_SIZE_DESC	KID_CATEGORY_DESC
1	1609	13804.38	45-54	A	125-149K	Homeowner	2 Adults Kids	5+	3+

- **Question 6:** Find products (product table: SUB_COMMODITY_DESC) which are most frequently bought together and the count of each combination bought together. do not print a combination twice (A-B / B-A)

```

with cte as(

select      t1.BASKET_ID,      t1.PRODUCT_ID,      p.SUB_COMMODITY_DESC      from
Ecommerce.transaction_data t1

join Ecommerce.product p using(product_id)

),

cte2 as(

select s1.SUB_COMMODITY_DESC as sub_com_1, s2.SUB_COMMODITY_DESC as sub_com_2
from cte s1 join cte s2

on s1.basket_id = s2.basket_id

and s1.product_id != s2.product_id

and s1.SUB_COMMODITY_DESC < s2.SUB_COMMODITY_DESC

)

select sub_com_1, sub_com_2, count(*) as cnt

from cte2

group by 1,2

order by cnt desc

```

Row	sub_com_1 ▼	sub_com_2 ▼	cnt ▼
1	FLUID MILK WHITE ONLY	YOGURT NOT MULTI-PACKS	5953
2	BANANAS	FLUID MILK WHITE ONLY	4365
3	FLUID MILK WHITE ONLY	SOFT DRINKS 12/18&15PK CA...	4326
4	FLUID MILK WHITE ONLY	MAINSTREAM WHITE BREAD	3934
5	BANANAS	YOGURT NOT MULTI-PACKS	3847
6	FLUID MILK WHITE ONLY	SHREDDED CHEESE	3840
7	FLUID MILK WHITE ONLY	SFT DRNK 2 LITER BTL CARB I...	3494
8	FRZN SS PREMIUM ENTREES/...	YOGURT NOT MULTI-PACKS	3344
9	BABY FOOD - BEGINNER	BABY FOOD JUNIOR ALL BRAN...	3290
10	SHREDDED CHEESE	YOGURT NOT MULTI-PACKS	3189

- **Question 7:** Find the weekly change in Revenue Per Account (RPA) (difference in spending by each customer compared to last week)(use lag function)

with cte as (

select

household_key,WEEK_NO,round(sum(SALES_VALUE),2) as total_amt_spent from
Ecommerce.transaction_data

group by 1,WEEK_NO

order by 1,2)

select *,

round(total_amt_spent-lag(total_amt_spent)over(partition by household_key order by week_no),2) as
RPA

from cte

order by household_key,week_no

Row	household_key	WEEK_NO	total_amt_spent	RPA
1	1	8	42.58	null
2	1	10	14.01	-28.57
3	1	13	14.03	0.02
4	1	14	25.71	11.68
5	1	15	10.98	-14.73
6	1	16	9.09	-1.89
7	1	17	13.98	4.89
8	1	19	47.35	33.37
9	1	20	31.77	-15.58
10	1	22	38.98	7.21

- **Question 8: Age group Vs Sale:**

```
select AGE_DESC, round(sum(SALES_VALUE),2) as sale from
'Ecommerce.transaction_data' join 'Ecommerce.hh_demographic'
using(household_key)
group by 1
order by 2 desc;
```

Row	AGE_DESC	sale
1	45-54	827984.9
2	35-44	622164.35
3	25-34	389545.17
4	65+	151606.81
5	55-64	150371.27
6	19-24	108404.35

- **Question 9: Number of Family Member Vs Sale**

```
select HOUSEHOLD_SIZE_DESC, round(sum(SALES_VALUE),2) as sale from
`ecommerce.transaction_data` join `ecommerce.hh_demographic`
```

```
using(household_key)
```

```
group by 1
```

```
order by 2 desc;
```

Row	HOUSEHOLD_SIZE_DESC	sale
1	2	880826.16
2	1	640187.6
3	3	332013.32
4	5+	221412.68
5	4	175637.09

- **Question 10: Which brand generates more revenue?**

```
select Brand, round(sum(sales_value),2) as sale from `Ecommerce.product`
```

```
join `Ecommerce.transaction_data` using(product_id)
```

```
group by 1
```

```
order by 2 desc;
```

Row	Brand	sale
1	National	2909678.52
2	Private	1119659.89

Analysis:

- The majority of orders fall within the \$0 to \$10 range, with most between \$0 and \$5.
- Households with keys 1730 and 1727 exhibit the highest average order values.
- Household key 1609 leads in total sales, contributing \$13k over the period, and is present in both the demographic and transaction data.
- The most commonly bought-together products are White Milk and Yogurt, as well as Bananas and Milk.
- The income bracket of \$35k to \$75k accounts for the largest share of the company's sales.
- Customers aged 44 to 54 generate the highest volume of sales.
- Families without children contribute more revenue than both singles and families with children.
- National brands drive more revenue compared to private label brands