

Ecommerce SQL Analysis

- **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+)

```
with sidcte as(
select BASKET_ID,
case
when sales_value between 0 and 10 then
'small'
when sales_value between 11 and 20 then
'medium'
else 'Large'
end as order_size
from ecommerce.transaction_data
)
select count(distinct(BASKET_ID)) as count,
order_size
from sidcte
group by 2;
```

Query results				
JOB INFORMATION		RESULTS	CHART	JSON
Row	count	order_size		
1	219473	small		
2	16416	Large		
3	20349	medium		

- **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+)

```
with sidcte as(
select BASKET_ID,
case
when sales_value between 0 and 5 then
'small'
when sales_value between 6 and 10 then
'medium'
else 'Large'
end as order_size
from ecommerce.transaction_data
)
select count(distinct(BASKET_ID)) as
count, order_size
from sidcte
group by 2;
```

Query results				
JOB INFORMATION		RESULTS	CHART	JSON
Row	count	order_size		
1	210259	small		
2	62239	Large		
3	53461	medium		

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

```
with sidcte as(
select week_no, store_id,
dense_rank() over(partition by week_no order by count(BASKET_ID) desc) as rnk
from ecommerce.transaction_data
group by 1,2
)
select week_no, store_id
from sidcte
where rnk <= 3
order by 1;
```

Query results					
JOB INFORMATION		RESULTS		CHART	JSON
Row	week_no	store_id	cnt		
1	1	324	80		
2	1	321	68		
3	1	32004	67		
4	2	375	99		
5	2	292	86		
6	2	315	74		
7	3	367	169		
8	3	375	158		
9	3	356	96		
10	4	367	249		

- **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_week,
max(WEEK_NO) as last_visit_week, round(avg(SALES_VALUE),2) as avg_spent,
round(sum(SALES_VALUE),2) as total_spent
from ecommerce.transaction_data
group by household_key
order by avg_spent desc;
```

Query results

JOB INFORMATION	RESULTS	CHART	JSON	EXECUTION DETAILS	EXECUTION GRA
Row	household_key	first_visit_week	last_visit_week	avg_spent	total_spent
1	1730	6	102	16.73	1656.76
2	1727	16	18	12.72	114.51
3	2163	8	97	10.54	221.32
4	1339	8	101	10.42	187.53
5	991	7	96	10.26	451.6
6	2219	12	101	10.05	321.66
7	2428	10	101	10.0	180.0
8	755	6	102	9.48	5461.54
9	1023	16	102	8.58	18901.09
10	120	10	94	8.18	130.92

- **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 sidcte as(
select household_key, round(sum(SALES_VALUE),2) as total_spent
from `ecommerce.transaction_data` join `ecommerce.hh_demographic`
using(household_key)
group by 1
order by total_spent desc
limit 1
)
select * from `ecommerce.hh_demographic` join sidcte using(household_key);
```

Query results

[SAVE RESULTS](#) [EXPLORE DATA](#)

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS		EXECUTION GRAPH		
Row	household_key	AGE_DESC	MARITAL_STATUS_CODE	INCOME_DESC	HOMEOWNER_DESC	HH_COMP_DESC	HOUSEHOLD_SIZE_DESC	KID_CATEGORY_DESC	total_spent
1	1609	45-54	A	125-149K	Homeowner	2 Adults Kids	5+	3+	13804.38

- **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 sidcte as(
select t1.BASKET_ID, t1.PRODUCT_ID, p.SUB_COMMODITY_DESC from
`ecommerce.transaction_data` t1
join `ecommerce.product` p using(product_id)
),
sidcte2 as(
select s1.SUB_COMMODITY_DESC as sub1, s2.SUB_COMMODITY_DESC as sub2 from
sidcte s1 join sidcte 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 sub1, sub2, count(*) as cnt
from sidcte2
group by 1,2
order by cnt desc
```

Query results

JOB INFORMATION					RESULTS	CHART	JSON	EXECUTION DETAILS	EXECUTION GRAPH
Row	sub1	sub2	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 sidcte as(
select household_key, week_no, round(sum(SALES_VALUE),2) as sale
from `ecommerce.transaction_data`
group by 1,2
order by 1,2
)
select *,
round(sale - lag(sale) over(partition by household_key order by week_no),2)
as RPA
from sidcte
```

Query results					
JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS
Row	household_key	week_no	sale	RPA	
1	155	16	22.81	null	
2	155	17	33.54	10.73	
3	155	18	31.8	-1.74	
4	155	19	9.58	-22.22	
5	155	21	19.32	9.74	
6	155	22	13.28	-6.04	
7	155	25	26.69	13.41	
8	155	26	6.58	-20.11	
9	155	28	26.08	19.5	
10	155	29	16.05	-10.03	

- **Question 8:** Do household income can be factor for the sale?

```
select INCOME_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;
```

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXEC
Row	INCOME_DESC	sale			
1	50-74K	547139.05			
2	35-49K	414471.79			
3	75-99K	279738.22			
4	25-34K	189846.93			
5	Under 15K	169160.19			

- Question 9: 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;
```

Query results

JOB INFORMATION		RESULTS	CHART	JSON
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		

- Question 10: 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;
```

Query results

JOB INFORMATION		RESULTS	CHART	JSON	EXEC
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 11: 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;
```

Query results			
JOB INFORMATION		RESULTS	CHART
Row	Brand	sale	JSON
1	National	2909678.52	
2	Private	1119659.89	

Analysis:

- Most orders fall within the \$0 to \$10 range, with the majority between \$0 and \$5.
- Household keys 1730 and 1727 have the highest average order values.
- Household key 1609 recorded the highest total sales over the period, amounting to \$13k. This household was present in both the demographic and transaction tables.
- The most frequently bought-together products were White Milk and Yogurt, as well as Bananas and Milk.
- The income range of \$35k to \$75k contributes the majority of sales for the company.
- The 44 to 54 age group generates the highest number of sales.
- Families without children contribute more to revenue compared to singles or families with children.
- National brands generate more revenue than private brands.

Recommendations:

- To boost revenue, sales promotions should target families with children.
- Products frequently bought together should be placed next to each other, and stock levels should be monitored to ensure availability when needed.
- The age group between 20-40 should be prioritized with special offers and discounts, as they have the potential to significantly increase revenue.
- Customers with the highest total spending over the year should be rewarded with loyalty gifts to encourage repeat business and strengthen customer loyalty.
- Since most orders are placed within the \$0 to \$10 range, products priced over \$10 should be displayed in prominent, eye-catching areas with "heavy discount" labels to attract attention and raise the average order value.