

Name- Shagun Wadhwa

E-commerce 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+)

Query:

```
with orders_value as (  
  Select BASKET_ID, sum(SALES_VALUE) as total  
  from `transaction.transaction` group by 1)  
  
  Select case when total >=0 and total <=10 then "Small"  
    when total >10 and total<=20 then "Medium"  
  else "Large" end as order_category, count(*) as order_count  
    from orders_value  
      group by 1  
      order by 2 desc
```

Row	order_category	order_count
1	Small	116415
2	Large	67311
3	Medium	49630

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+)

Query:

```
with orders_value as (  
  Select BASKET_ID, sum(SALES_VALUE) as total  
  from `transaction.transaction` group by 1)  
  
  Select case when total >=0 and total <=5 then "Small"  
    when total >5 and total<=10 then "Medium"  
  else "Large" end as order_category, count(*) as order_count
```

```
from orders_value
```

```
group by 1
```

```
order by 2 desc
```

Row	order_category ▼	order_count ▼	
1	Large	116941	
2	Small	70842	
3	Medium	45573	

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

```
Query: WITH cte AS(
```

```
SELECT WEEK_NO,STORE_ID,
```

```
COUNT(*) AS FootTraffic
```

```
FROM `transaction.transaction`
```

```
GROUP BY 1,2
```

```
order by 1,3 desc),
```

```
cte2 AS (
```

```
SELECT WEEK_NO, STORE_ID,
```

```
ROW_NUMBER() OVER (PARTITION BY WEEK_NO ORDER BY FootTraffic DESC) AS Rank
```

```
FROM cte
```

```
order by 1)
```

```
SELECT WEEK_NO,STORE_ID,
```

```
FROM cte2
```

```
WHERE Rank <= 3
```

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

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

```

Query: SELECT household_key,

        MIN(DAY) AS FirstVisit,

        MAX(DAY) AS LastVisit,

        COUNT(DISTINCT BASKET_ID) AS NumberOfVisits,

        round(AVG(SALES_VALUE),2) AS AvgMoneySpentPerVisit,

        round(SUM(SALES_VALUE),2) AS TotalMoneySpent

        FROM `transaction.transaction`

        GROUP BY household_key

        ORDER BY AvgMoneySpentPerVisit DESC;

```

Row	household_key	FirstVisit	LastVisit	NumberOfVisits	AvgMoneySpentPerV	TotalMoneySpent
1	2042	52	683	26	89.97	2339.21
2	973	95	710	80	85.95	6875.89
3	1899	20	705	69	83.91	5789.59
4	1900	111	707	55	76.87	4227.72
5	1574	107	651	27	68.27	1843.3
6	1315	60	624	5	63.48	317.39
7	2479	111	706	111	62.65	6954.64
8	931	94	668	40	61.38	2455.29

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)

Query: **WITH** CustomerSpending **AS** (

SELECT household_key,

round(**SUM**(SALES_VALUE),2) **AS** TotalSpent

FROM `transaction.transaction`

GROUP BY household_key)

SELECT d.*, cs.TotalSpent

FROM `hh_demographic.hh_demographic` d

JOIN CustomerSpending cs **ON** d.household_key = cs.household_key

ORDER BY cs.TotalSpent **DESC**

LIMIT 1;

Row	household_key	TotalSpent	AGE_DESC	MARITAL_STATUS_CODE	INCOME_DESC
1	1609	13804.38	45-54	A	125-149K

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, t2.SUB_COMMODITY_DESC, row_number()over() as id

from `transaction.transaction` as t1

left join `Product.Product` as t2

on t1.PRODUCT_ID=t2.PRODUCT_ID),

cte3 as(

Select t3.SUB_COMMODITY_DESC as item_1,

t4.SUB_COMMODITY_DESC as item_2,

count(*) as frequent_count from cte as t3

inner join cte as t4

on t3.BASKET_ID=t4.BASKET_ID

and t3.SUB_COMMODITY_DESC != t4.SUB_COMMODITY_DESC

and t4.id>t3.id

group by 1,2

order by 3 desc)

Select case when item_1>item_2 then item_1 else item_2 end as product_1,

case when item_1<item_2 then item_1 else item_2 end as product_2

, sum(frequent_count) as frequent_bought from cte3

group by 1,2 order by 3 desc
```

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS	E
Row	product_1 ▼	product_2 ▼	frequent_bought ▼			
1	YOGURT NOT MULTI-PACKS	FLUID MILK WHITE ONLY	5953			
2	FLUID MILK WHITE ONLY	BANANAS	4365			
3	SOFT DRINKS 12/18&15PK CA...	FLUID MILK WHITE ONLY	4326			
4	MAINSTREAM WHITE BREAD	FLUID MILK WHITE ONLY	3934			
5	YOGURT NOT MULTI-PACKS	BANANAS	3847			
6	SHREDDED CHEESE	FLUID MILK WHITE ONLY	3840			
7	SFT DRNK 2 LITER BTL CARB I...	FLUID MILK WHITE ONLY	3494			
8	YOGURT NOT MULTI-PACKS	FRZN SS PREMIUM ENTREES/...	3344			
9	BABY FOOD JUNIOR ALL BRAN...	BABY FOOD - BEGINNER	3290			
10	YOGURT NOT MULTI-PACKS	SHREDDED CHEESE	3189			
11	KIDS CEREAL	FLUID MILK WHITE ONLY	3134			
12	FLUID MILK WHITE ONLY	DAIRY CASE 100% PURE JUICE...	2881			
13	YOGURT NOT MULTI-PACKS	SOFT DRINKS 12/18&15PK CA...	2639			

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

Query: **WITH** WeeklyRevenue **AS** (

SELECT household_key, WEEK_NO,

round(**SUM**(SALES_VALUE),2) **AS** WeeklySpending

FROM `transaction.transaction`

GROUP BY household_key, WEEK_NO),

WeeklyChange **AS** (

SELECT household_key,WEEK_NO,WeeklySpending,

round(**LAG**(WeeklySpending, 1) **OVER** (**PARTITION BY** household_key
ORDER BY WEEK_NO),2) **AS** PreviousWeekSpending

FROM WeeklyRevenue)

SELECT household_key,WEEK_NO, WeeklySpending,

PreviousWeekSpending, round(WeeklySpending -
PreviousWeekSpending,2) AS ChangeInSpending

FROM WeeklyChange

WHERE PreviousWeekSpending IS NOT NULL;

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS		EXECUTION
Row	household_key	WEEK_NO	WeeklySpending	PreviousWeekSpending	ChangeInSpending		
1	37	15	36.32	17.68	18.64		
2	37	17	16.86	36.32	-19.46		
3	37	29	28.0	16.86	11.14		
4	37	31	55.0	28.0	27.0		
5	37	34	35.0	55.0	-20.0		
6	37	35	41.13	35.0	6.13		
7	37	36	9.5	41.13	-31.63		
8	37	39	19.85	9.5	10.35		
9	37	40	8.99	19.85	-10.86		
10	37	41	15.0	8.99	6.01		
11	37	44	61.34	15.0	46.34		
12	37	45	40.42	61.34	-20.92		
13	37	48	29.69	40.42	-10.73		

Additional Questions

Task 1: Find the top 5 most purchased products by revenue and their total revenue

Query: SELECT p.SUB_COMMODITY_DESC ,

round(SUM(t.SALES_VALUE),2) AS total_revenue

FROM `transaction.transaction` t

JOIN `Product.Product` p ON t.PRODUCT_ID = p.PRODUCT_ID

GROUP BY p.SUB_COMMODITY_DESC

ORDER BY total_revenue DESC

LIMIT 5;

Row	SUB_COMMODITY_DESC	total_revenue
1	GASOLINE-REG UNLEADED	315997.09
2	FLUID MILK WHITE ONLY	80754.44
3	SOFT DRINKS 12/18&15PK CA...	79214.44
4	BEERALEMALT LIQUORS	75036.18
5	CIGARETTES	48179.15

Task2 : Calculate the average order value per household and report top 10 household having highest average order value

```
Query: SELECT hh.household_key,
        round(avg(t.SALES_VALUE),2) AS average_order_value
FROM `transaction.transaction` t
JOIN `hh_demographic.hh_demographic` hh
ON t.household_key = hh.household_key
GROUP BY 1
ORDER BY 2 desc
LIMIT 10;
```


Row	household_key	average_order_value
1	755	9.48
2	1357	6.62
3	2162	6.23
4	101	6.23
5	2097	6.23
6	853	5.59
7	513	5.57
8	2203	5.49
9	13	5.47
10	193	5.35

Insights and Recommendations on above Queries:

Insights:

- 1. Distribution of Order Sizes:** By categorizing orders into small, medium, and large, we can understand the distribution of order sizes. If a majority of orders fall into the "small" category, it might indicate that customers prefer making frequent small purchases rather than fewer large ones.
- 2. Customer Spending Patterns:** This categorization helps in identifying the spending patterns of customers, which can be crucial for targeted marketing campaigns.
- 3. Detailed Spending Insights:** By refining the order value categories, we gain more granular insights into customer spending behaviors.
- 4. Behavioral Trends:** This detailed breakdown can reveal more precise trends in how customers are spending their money.

- 5. Store Popularity:** Identifying the top stores by foot traffic helps in understanding which locations are the most popular among customers.
- 6. Customer Preferences:** This analysis reveals customer preferences and the effectiveness of store locations.
- 7. Customer Loyalty:** Identifying the number of visits, first and last visit, and spending patterns helps in understanding customer loyalty.
- 8. Spending Habits:** Profiling customers based on their average and total spending can reveal insights into customer segments and their value to the business.
- 9. High-Value Customers:** Analyzing the most spending customers helps in identifying the key drivers of revenue.
- 10. Demographic Influence:** Understanding the demographic profile of high-spending customers can provide insights into target markets.
- 11. Product Affinities:** Understanding which products are frequently bought together can reveal product affinities and customer preferences.
- 12. Cross-Selling Opportunities:** This information can be used to identify cross-selling opportunities.
- 13. Customer Spending Trends:** Analyzing weekly changes in spending can reveal trends and patterns in customer behavior.
- 14. Seasonal Effects:** This analysis can help identify seasonal effects or external factors influencing customer spending.

Recommendations:

1. Promotions for Medium Orders: If medium-sized orders are less frequent, consider running promotions to encourage customers to increase their order size from small to medium.

2. Loyalty Programs: Implement loyalty programs that reward customers for reaching higher order value thresholds, encouraging them to make larger purchases.

3.Product Bundling: Bundle products together to encourage customers to increase their order size, thus moving from small to medium or large categories.

4.Targeted Discounts: Offer targeted discounts to encourage customers who make small orders to increase their spending to the medium category.

5.Enhanced User Experience: Improve the user experience on the platform to make it easier for customers to find and purchase higher-value items.

7.Marketing Strategies: Tailor marketing strategies to focus on customers in the medium order value range to encourage them to move to the large category.

8.Resource Allocation: Allocate more resources (staff, inventory) to the stores with the highest foot traffic to ensure optimal customer service.

9.Promotional Activities: Conduct promotional activities and events at the top stores to capitalize on their high foot traffic.

10.Replication of Success: Analyze the factors contributing to the success of the top stores and apply similar strategies to other locations to boost their performance.

11.Personalized Marketing: Use customer profiles to send personalized marketing messages and offers to high-value customers to increase their loyalty.

12.Customer Retention: Implement retention strategies for customers with high average spending but fewer visits, encouraging them to shop more frequently.

13.Customer Feedback: Collect feedback from top customers to understand their needs and preferences better, improving the overall shopping experience.

14.Exclusive Offers: Provide exclusive offers and benefits to high-spending customers to enhance their loyalty.

15.Customer Engagement: Engage with these customers through personalized communication, loyalty programs, and special events.

16.Market Expansion: Use demographic insights to identify and target similar customer segments for market expansion.

17.Product Bundling: Create product bundles based on frequently bought together items to increase the average order value.

18.Cross-Sell Promotions: Implement cross-sell promotions on the website and in-store to encourage customers to purchase related products.

19.Inventory Management: Ensure that frequently bought together products are stocked adequately to meet customer demand.

THANK YOU