

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.

This project aims to leverage the power of e-commerce data (sales, product, and demographic) analyzed through SQL to unlock actionable insights driving profitable growth. By delving into customer behavior, product trends, and sales patterns, we will uncover hidden value that can inform key business decisions. You need to find these patterns and calculate various metrics and KPIs that suit the data and the goal.

Down below there are are 20 questions, each accompanied by the corresponding SQL code and its results.

Note: For convenience some results from the code have been further converted into customized big query tables in addition to these three tables to derive much more complex insights.

Demographic table

Variable	Description
HOUSEHOLD_KEY	Uniquely identifies each household
AGE_DESC	Estimated age range
MARITAL_STATUS_CODE	Marital Status (A - Married, B- Single, U - Unknown)
INCOME_DESC	Household income
HOMEOWNER_DESC	Homeowner, renter, etc.
HH_COMP_DESC	Household composition
HOUSEHOLD_SIZE_DESC	Size of household up to 5+
KID_CATEGORY_DESC	Number of children present up to 3+

Transactions table

Variable	Description
HOUSEHOLD_KEY	Uniquely identifies each household
BASKET_ID	Uniquely identifies a purchase occasion
DAY	Day when transaction occurred
PRODUCT_ID	Uniquely identifies each product
QUANTITY	Number of the products purchased during the trip
SALES_VALUE	Amount of dollars retailer receives from sale
STORE_ID	Identifies unique stores
COUPON_MATCH_DISC	Discount applied due to retailer's match of manufacturer coupon
COUPON_DISC	Discount applied due to manufacturer coupon
RETAIL_DISC	Discount applied due to retailer's loyalty card program
TRANS_TIME	Time of day when the transaction occurred
WEEK_NO	Week of the transaction. Ranges 1 - 102

Products table:

Variable	Description
PRODUCT_ID	Number that uniquely identifies each product
DEPARTMENT	Groups similar products together
COMMODITY_DESC	Groups similar products together at a lower level
SUB_COMMODITY_DESC	Groups similar products together at the lowest level
MANUFACTURER	Code that links products with same manufacturer together
BRAND	Indicates Private or National label brand
CURR_SIZE_OF_PRODUCT	Indicates package size (not available for all products)

SQL code for combining all the three tables to produce a final table called as “combined table” to extract all the essential insights.

```
select *
from `Ecommerce_company.transactions_data` as t
left join `Ecommerce_company.Products_table` as p on t.product_id = p.product_id
left join `Ecommerce_company.demographics_table` as d on t.household_key = d.household_key;
```

Q1. Find the number of orders that have small, medium or large order value (small:0-10 dollars, medium:10-20 dollars, large:20+)

```
select sum(case when total_order_value > 0 and total_order_value <=10 then 1 end) as small_orders,
      sum(case when total_order_value > 10 and total_order_value <= 20 then 1 end) as medium_orders,
      sum(case when total_order_value > 20 then 1 end) as large_orders
from (select BASKET_ID, sum(SALES_VALUE) as total_order_value
from `Ecommerce_company.combined_table`
group by BASKET_ID) as aggregated_orders;
```

Row	small_orders	medium_orders	large_orders
1	115493	49630	67311

Q2. Find the number of orders that are small, medium or large order value(small:0-5 dollars, medium:5-10 dollars, large:10+)

```
select sum(case when total_order_value > 0 and total_order_value <=5 then 1 end) as small_orders,
      sum(case when total_order_value > 5 and total_order_value <= 10 then 1 end) as medium_orders,
      sum(case when total_order_value > 10 then 1 end) as large_orders
from (select BASKET_ID, sum(SALES_VALUE) as total_order_value
from `Ecommerce_company.combined_table`
group by BASKET_ID) as aggregated_orders;
```

Row	small_orders	medium_orders	large_orders
1	69920	45573	116941

Q3. Find top 3 stores with highest foot traffic for each week (Foot traffic: number of customers transacting)

```
select x.WEEK_NO, x.STORE_ID, x.number_of_customers
from (select WEEK_NO, STORE_ID, count(distinct household_key) as number_of_customers,
dense_rank() over(partition by WEEK_NO order by count(distinct household_key) desc) as rank_of_stores
from `Ecommerce_company.combined_table`
group by WEEK_NO, STORE_ID) as x
where x.rank_of_stores <= 3
order by WEEK_NO asc, rank_of_stores asc;
```

Row	WEEK_NO	STORE_ID	number_of_customers
1	1	32004	5
2	1	324	3
3	1	367	3
4	1	396	3
5	1	446	3
6	1	358	2
7	1	634	2
8	1	306	2
9	1	359	2
10	1	400	2

Q4. 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(DAY) as first_visit_day, max(DAY) as last_visit_day,
count(distinct DAY) as number_of_visits,
round(avg(SALES_VALUE), 2) as avg_money_spent, round(sum(SALES_VALUE), 2) as total_monety_spent
from `Ecommerce_company.combined_table`
group by household_key
order by avg_money_spent desc;
```

Row	household_key	first_visit_day	last_visit_day	number_of_visits	avg_money_spent	total_monety_spent
1	1730	34	707	70	16.73	1656.76
2	1727	109	118	2	12.72	114.51
3	2163	51	674	13	10.54	221.32
4	1339	52	701	6	10.42	187.53
5	991	44	665	16	10.26	451.6
6	2219	80	702	12	10.05	321.66
7	2428	67	702	13	10.0	180.0
8	755	36	709	178	9.48	5461.54
9	1023	107	710	303	8.58	18901.09
10	120	62	653	11	8.18	130.92

Q5. 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 top_spender as (
select household_key, round(sum(SALES_VALUE), 2) as total_spent
from `Ecommerce_company.combined_table`
where AGE_DESC is not null and MARITAL_STATUS_CODE is not null
and INCOME_DESC is not null and HOMEOWNER_DESC is not null
and HH_COMP_DESC is not null and HOUSEHOLD_SIZE_DESC is not null
and KID_CATEGORY_DESC is not null
group by household_key
order by total_spent desc
limit 1)

select c.household_key, total_spent, AGE_DESC, MARITAL_STATUS_CODE,
INCOME_DESC, HOMEOWNER_DESC, HH_COMP_DESC,
HOUSEHOLD_SIZE_DESC, KID_CATEGORY_DESC
from `Ecommerce_company.combined_table` as c, top_spender
where c.household_key = (select household_key from top_spender)
limit 1;
```

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

HOMEOWNER_DESC	HH_COMP_DESC	HOUSEHOLD_SIZE_DESC	KID_CATEGORY_DESC
Homeowner	2 Adults Kids	5+	3+

Q6. 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 product_pairs AS
(SELECT t1.DAY,
      LEAST(t1.SUB_COMMODITY_DESC, t2.SUB_COMMODITY_DESC) AS product_1,
      GREATEST(t1.SUB_COMMODITY_DESC, t2.SUB_COMMODITY_DESC) AS product_2
FROM `Ecommerce_company.Transaction_product_table` t1
JOIN `Ecommerce_company.Transaction_product_table` t2
ON t1.DAY = t2.DAY AND t1.SUB_COMMODITY_DESC != t2.SUB_COMMODITY_DESC)

SELECT product_1, product_2, COUNT(*) AS count_of_combination
FROM product_pairs
GROUP BY product_1, product_2
having count(*) > 300000
ORDER BY count_of_combination DESC;
```

Row	product_1	product_2	count_of_combination
1	FLUID MILK WHITE ONLY	SOFT DRINKS 12/18&15PK CA...	1962204
2	FLUID MILK WHITE ONLY	YOGURT NOT MULTI-PACKS	1913850
3	FLUID MILK WHITE ONLY	SFT DRNK 2 LITER BTL CARB I...	1883974
4	BANANAS	FLUID MILK WHITE ONLY	1544424
5	FLUID MILK WHITE ONLY	SHREDDED CHEESE	1422400
6	FLUID MILK WHITE ONLY	MAINSTREAM WHITE BREAD	1410630
7	CANDY BARS (SINGLES)(INCL...	FLUID MILK WHITE ONLY	1296934
8	FLUID MILK WHITE ONLY	GASOLINE-REG UNLEADED	1241884
9	FLUID MILK WHITE ONLY	POTATO CHIPS	1141796
10	SOFT DRINKS 12/18&15PK CA...	YOGURT NOT MULTI-PACKS	1126486

Q7. Find the weekly change in Revenue Per Account (RPA) (difference in spending by each customer compared to last week)(use lag function)

```
select household_key, WEEK_NO, total_spent_week, previous_week, total_spent_pre_week,
round((total_spent_week - total_spent_pre_week), 2) as difference
from
(select household_key, WEEK_NO, round(sum(SALES_VALUE), 2) as total_spent_week,
lag(WEEK_NO) over(partition by household_key order by WEEK_NO) as previous_week,
lag(round(sum(SALES_VALUE), 2)) over(partition by household_key order by WEEK_NO) as total_spent_pre_week,
from `Ecommerce_company.combined_table`
group by WEEK_NO, household_key) as x
order by household_key, WEEK_NO;
```

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

Q8. "Which top 10 households received the least total discounts, but have correspondingly highest total amounts spent, and average amounts spent?"

```
select household_key, round(sum(Total_discount), 2) as total_discount_given,
round(avg(Total_discount), 2) as avg_discount, round(sum(SALES_VALUE), 2) as total_amt_spent,
round(avg(SALES_VALUE), 2) as avg_spent
from `Ecommerce_company.sales_table`
group by household_key
order by total_discount_given desc, total_amt_spent desc
limit 10;
```

Row	household_key	total_discount_given	avg_discount	total_amt_spent	avg_spent
1	1035	0.0	0.0	109.43	6.08
2	494	0.0	0.0	6.43	3.21
3	522	-0.29	-0.07	15.58	3.9
4	1504	-0.5	-0.13	11.45	2.86
5	1897	-0.98	-0.33	16.49	5.5
6	379	-1.07	-0.09	28.94	2.41
7	657	-1.08	-0.05	61.77	2.81
8	1626	-1.27	-0.32	10.59	2.65
9	252	-1.47	-0.1	54.1	3.61
10	1876	-1.5	-0.21	36.95	5.28

Q9. Which households have an average order value (AOV) of \$40 or more, and how does their income level relate to their total spending and AOV?

```
select household_key, round(sum(SALES_VALUE), 2) as total_amt_spent,
round(sum(sales_value)/count(distinct BASKET_ID),2) as AOV_per_household, INCOME_DESC
from `Ecommerce_company.combined_table`
group by household_key, INCOME_DESC
having AOV_per_household >= 40
order by AOV_per_household desc;
```

Row	household_key	total_amt_spent	AOV_per_household	INCOME_DESC
1	2042	2339.21	89.97	null
2	973	6875.89	85.95	50-74K
3	1899	5789.59	83.91	25-34K
4	1900	4227.72	76.87	null
5	1574	1843.3	68.27	null
6	1315	317.39	63.48	null
7	2479	6954.64	62.65	150-174K
8	931	2455.29	61.38	null
9	1344	1570.37	60.4	null
10	248	3090.89	58.32	125-149K
11	688	1558.95	57.74	125-149K

Q10. Which sub-commodities have been purchased by at least 40% of households, and how many unique buyers contributed to this percentage?"

```
select SUB_COMMODITY_DESC, count(distinct household_key) as unique_number_of_buyers
((count(distinct household_key))*100/2500) as percentage_contribution
from `Ecommerce_company.combined_table`
group by SUB_COMMODITY_DESC
having percentage_contribution >= 40
order by unique_number_of_buyers desc;
```

Row	SUB_COMMODITY_DESC	unique_number_of_buyers	percentage_contribution
1	FLUID MILK WHITE ONLY	2291	91.64
2	SOFT DRINKS 12/18&15PK CA...	1886	75.44
3	BANANAS	1882	75.28
4	SHREDDED CHEESE	1868	74.72
5	POTATO CHIPS	1819	72.76
6	SFT DRNK 2 LITER BTL CARB I...	1798	71.92
7	MAINSTREAM WHITE BREAD	1744	69.76
8	PREMIUM	1711	68.44
9	CANDY BARS (SINGLES)(INCL...	1684	67.36
10	EGGS - LARGE	1636	65.44
11	TORTILLA/NACHO CHIPS	1608	64.32
12	DAIRY CASE 100% PURE JUICE...	1601	64.04

Q11. What is the total amount spent and the average order value (AOV) for each income level, and how do these income levels rank by AOV?

```
select INCOME_DESC, round(sum(SALES_VALUE),2) as Total_amt_spent,
round(sum(sales_value)/count(distinct BASKET_ID),2) as AOV_per_Incomelevel
from `Ecommerce_company.combined_table`
where INCOME_DESC is not null
group by INCOME_DESC
order by AOV_per_Incomelevel desc;
```

Row	INCOME_DESC	Total_amt_spent	AOV_per_Incomelevel
1	150-174K	126501.69	29.6
2	200-249K	14279.43	27.46
3	250K+	58935.02	26.86
4	175-199K	47268.08	25.1
5	125-149K	150464.7	23.89
6	75-99K	279738.22	22.53
7	50-74K	547139.05	19.92
8	100-124K	100931.47	18.68
9	35-49K	414471.79	16.62
10	Under 15K	169160.19	15.79

Q12. What is the total amount spent and the average order value (AOV) for each age group, and how do these age groups rank by AOV?

```
select AGE_DESC, round(sum(SALES_VALUE),2) as Total_amt_spent,
round(sum(sales_value)/count(distinct BASKET_ID),2) as AOV_per_Age
from `Ecommerce_company.combined_table`
where AGE_DESC is not null
group by AGE_DESC
order by AOV_per_Age desc;
```

Row	AGE_DESC	Total_amt_spent	AOV_per_Age
1	35-44	622164.35	20.4
2	25-34	389545.17	20.12
3	45-54	827984.9	19.04
4	55-64	150371.27	17.4
5	19-24	108404.35	15.67
6	65+	151606.81	15.08

Q13. How do the two brands contribute to overall sales and discounts, and what are their performance metrics in terms of commodities offered, revenue, average order value (AOV), and store presence?

```
with brand_categorization as
(select BRAND, count(distinct COMMODITY_DESC) as L1_commodities,
count(distinct SUB_COMMODITY_DESC) as L2_commodities,
round(sum(SALES_VALUE), 2) as total_revenue,
round(sum(COUPON_DISC + COUPON_MATCH_DISC + RETAIL_DISC), 2) as Total_discount,
round(sum(sales_value)/count(distinct BASKET_ID),2) as AOV_by_brand,
count(distinct STORE_ID) as stores_operated
from `Ecommerce_company.combined_table`)
group by BRAND)

select brand, L1_commodities, L2_commodities, total_revenue,
round((total_revenue*100/4029338.41), 2) as sales_percent_contribution,
Total_discount, round((Total_discount*100/-724362.85), 2) as Disc_percent_contribution,
AOV_by_brand, stores_operated
from brand_categorization;
```

Row	brand	L1_commodities	L2_commodities	total_revenue	sales_percent_contri	Total_discount
1	Private	222	918	1119659.89	27.79	-171028.33
2	National	302	2213	2909678.52	72.21	-553334.52

Disc_percent_contribution	AOV_by_brand	stores_operated
23.61	7.69	459
76.39	14.41	519

Q14. How does the monthly revenue fluctuate across different departments, and what are the month-to-month differences in total spending?

```
with departmentRevenue_monthly_fluc as
(select Department, ceil(DAY/30.0) as month, round(sum(SALES_VALUE), 2) as total_spent_month
from `Ecommerce_company.combined_table`
where Department is not null
group by Department, month)

select Department, month, total_spent_month,
lag(month) over(partition by Department order by month) as previous_month,
lag(total_spent_month) over(partition by Department order by month) as total_spent_pre_month,
round((total_spent_month - coalesce(lag(total_spent_month) over(partition by Department order by month),0)), 2) as
difference
from departmentRevenue_monthly_fluc;
```

Row	Department	month	total_spent_month	previous_month	total_spent_pre_mon	difference
1	COUP/STR & MFG	1.0	0.0	null	null	0.0
2	COUP/STR & MFG	2.0	0.0	1.0	0.0	0.0
3	COUP/STR & MFG	3.0	4.5	2.0	0.0	4.5
4	COUP/STR & MFG	4.0	26.26	3.0	4.5	21.76
5	COUP/STR & MFG	5.0	25.97	4.0	26.26	-0.29
6	COUP/STR & MFG	6.0	16.07	5.0	25.97	-9.9
7	COUP/STR & MFG	7.0	13.72	6.0	16.07	-2.35
8	COUP/STR & MFG	8.0	20.77	7.0	13.72	7.05

Q15. How does spending and order frequency for different sub-commodities vary by time of day?

```
select SUB_COMMODITY_DESC,
case when formatted_time between time(05, 00, 00) and time(11, 59, 59) then 'Morning'
when formatted_time between time(12, 00, 00) and time(16, 59, 59) then 'Afternoon'
when formatted_time between time(17, 00, 00) and time(20, 59, 59) then 'Evening'
else 'Night' end as time_of_day,
count(distinct BASKET_ID) as number_of_orders,
round(sum(sales_value), 2) as total_spent
from `Ecommerce_company.combined_table`
group by SUB_COMMODITY_DESC, time_of_day
order by SUB_COMMODITY_DESC, total_spent desc;
```

Row	SUB_COMMODITY_DESC	time_of_day	number_of_orders	total_spent
1	*ATTERIES:CAMERA/FLASH/W...	Evening	1	6.99
2	*ATTERIES:CAMERA/FLASH/W...	Afternoon	1	6.99
3	*BOYS/GIRLS MISC TOYS	Afternoon	6	32.45
4	*BOYS/GIRLS MISC TOYS	Night	2	9.69
5	*MISC. LOBBY ITEMS	Morning	1	4.99
6	*MISC. LOBBY ITEMS	Afternoon	1	4.99
7	*PURSES UMBRELLAS	Afternoon	4	47.97
8	*PURSES UMBRELLAS	Evening	3	29.97
9	*PURSES UMBRELLAS	Night	1	14.99
10	*PURSES UMBRELLAS	Morning	1	14.99

Q16. How does the total spending and order frequency for different departments vary by time of day?

```
select DEPARTMENT, case when formatted_time between time(05, 00, 00) and time(11, 59, 59) then 'Morning'
                        when formatted_time between time(12, 00, 00) and time(16, 59, 59) then 'Afternoon'
                        when formatted_time between time(17, 00, 00) and time(20, 59, 59) then 'Evening'
                        else 'Night' end as time_of_day,
count(distinct BASKET_ID) as number_of_orders,
round(sum(sales_value), 2) as total_spent,
from `Ecommerce_company.combined_table`
group by DEPARTMENT, time_of_day
order by DEPARTMENT, total_spent desc;
```

Row	DEPARTMENT	time_of_day	number_of_orders	total_spent
1	AUTOMOTIVE	Evening	18	108.12
2	AUTOMOTIVE	Afternoon	13	94.57
3	AUTOMOTIVE	Night	3	22.27
4	AUTOMOTIVE	Morning	2	13.58
5	CHARITABLE CONT	Night	1	1.0
6	CHEF SHOPPE	Afternoon	170	518.03
7	CHEF SHOPPE	Evening	144	459.39
8	CHEF SHOPPE	Morning	57	157.17
9	CHEF SHOPPE	Night	24	62.34
10	CNTRL/STORE SUP	Evening	5	8.0
11	CNTRL/STORE SUP	Afternoon	6	4.1

Q17. How does total spending and average order value (AOV) vary across different household sizes and income levels?

```
select HOUSEHOLD_SIZE_DESC, INCOME_DESC, round(sum(SALES_VALUE), 2) as total_spent,
round(sum(sales_value)/count(distinct BASKET_ID),2) as AOV_size_income
from `Ecommerce_company.combined_table`
where HOUSEHOLD_SIZE_DESC is not null or INCOME_DESC is not null
group by HOUSEHOLD_SIZE_DESC, INCOME_DESC
order by HOUSEHOLD_SIZE_DESC desc, total_spent desc;
```

Row	HOUSEHOLD_SIZE_DESC	INCOME_DESC	total_spent	AOV_size_income
1	5+	125-149K	48295.73	30.05
2	5+	50-74K	36234.72	26.86
3	5+	35-49K	30208.25	18.08
4	5+	75-99K	27827.42	27.52
5	5+	25-34K	17362.14	17.63
6	5+	100-124K	16674.09	22.81
7	5+	150-174K	16556.77	36.31
8	5+	Under 15K	7918.0	17.63
9	5+	250K+	6203.67	36.07
10	5+	15-24K	5561.79	18.73
11	5+	175-199K	5524.22	20.09

Q18. How do total discounts and total spending vary by income level, and what influence does discounts actually have on overall spending among different income levels?

```
with income_discount_categorization as
(select INCOME_DESC, round(sum(COUPON_DISC + COUPON_MATCH_DISC + RETAIL_DISC), 2) as Total_discount,
round(sum(SALES_VALUE), 2) as total_spent
from `Ecommerce_company.combined_table`
where INCOME_DESC is not null
group by INCOME_DESC)

select INCOME_DESC, Total_Discount,
round((Total_Discount*100/-724362.85), 2) as disc_taken_percent,
total_spent, round((total_spent*100/4029338.41), 2) as sales_percent_contribution
from income_discount_categorization
order by sales_percent_contribution desc, disc_taken_percent desc;
```

Row	INCOME_DESC	Total_Discount	disc_taken_percent	total_spent	sales_percent_contribution
1	50-74K	-96709.17	13.35	547139.05	13.58
2	35-49K	-78624.87	10.85	414471.79	10.29
3	75-99K	-48229.03	6.66	279738.22	6.94
4	25-34K	-37646.35	5.2	189846.93	4.71
5	Under 15K	-35396.42	4.89	169160.19	4.2
6	15-24K	-30110.69	4.16	151340.28	3.76
7	125-149K	-22986.3	3.17	150464.7	3.73
8	150-174K	-18085.61	2.5	126501.69	3.14
9	100-124K	-16383.59	2.26	100931.47	2.5
10	250K+	-7455.96	1.03	58935.02	1.46
11	175-199K	-5622.36	0.78	47268.08	1.17
12	200-249K	-1724.78	0.24	14279.43	0.35

Q19. What are the top 10 most purchased sub-commodities for each age group, and how many purchases were made for each of these top sub-commodities?

```
select AGE_DESC, SUB_COMMODITY_DESC, count_of_purchases
from(
select AGE_DESC, SUB_COMMODITY_DESC, count(*) as count_of_purchases,
dense_rank() over(partition by age_desc order by count(*) desc) as rank_of_orders
from `Ecommerce_company.combined_table`
where age_desc is not null
group by Age_desc, SUB_COMMODITY_DESC) as x
where x.rank_of_orders <=10
order by Age_desc, count_of_purchases desc;
```

Row	AGE_DESC	SUB_COMMODITY_DESC	count_of_purchases
1	19-24	SFT DRNK 2 LITER BTL CARB I...	1007
2	19-24	FLUID MILK WHITE ONLY	929
3	19-24	SOFT DRINKS 12/18&15PK CA...	709
4	19-24	YOGURT NOT MULTI-PACKS	498
5	19-24	MAINSTREAM WHITE BREAD	491
6	19-24	SHREDDED CHEESE	447
7	19-24	SS ECONOMY ENTREES/DINN...	398
8	19-24	CANDY BARS (SINGLES)(INCL...	381
9	19-24	KIDS CEREAL	363
10	19-24	POTATO CHIPS	348
11	25-34	FLUID MILK WHITE ONLY	3219
12	25-34	YOGURT NOT MULTI-PACKS	2349

Q20. What is the percentage contribution of total spending by time of day for each store, and how does the number of unique customers vary throughout the day?

```
with store_pending as (  
  select STORE_ID, case when formatted_time between time(05, 00, 00) and time(11, 59, 59) then 'Morning'  
    when formatted_time between time(12, 00, 00) and time(16, 59, 59) then 'Afternoon'  
    when formatted_time between time(17, 00, 00) and time(20, 59, 59) then 'Evening'  
    else 'Night' end as time_of_day,  
    count(distinct household_key) as number_of_unique_customers,  
    round(sum(sales_value), 2) as total_spent  
  from `Ecommerce_company.combined_table`  
  group by STORE_ID, time_of_day  
  order by STORE_ID)  
  
select store_id, time_of_day, number_of_unique_customers, total_spent,  
  round((total_spent*100/nullif(sum(total_spent) over(partition by store_id), 0)), 2) as percent_contribution  
from store_pending  
order by STORE_ID;
```

Row	store_id	time_of_day	number_of_unique_customers	total_spent	percent_contribution
1	1	Morning	1	15.0	85.76
2	1	Afternoon	1	2.49	14.24
3	2	Evening	1	5.78	29.24
4	2	Afternoon	1	13.99	70.76
5	12	Night	1	10.48	100.0
6	19	Evening	1	10.0	68.97
7	19	Afternoon	1	4.5	31.03
8	20	Afternoon	1	10.0	100.0
9	21	Morning	1	0.8	39.02
10	21	Afternoon	1	1.25	60.98
11	22	Night	1	9.76	100.0
12	26	Afternoon	1	1.75	100.0

Key Insights:

- The majority of orders (approximately 115,493) fall within the \$0 to \$10 range, indicating a high frequency of small orders. Large orders (greater than \$20) rank second, with 67,311 orders. Medium-sized orders (\$10 to \$20) are the least common, totalling 49,630 orders.
- However, if we redefine "large orders" to be those greater than \$10, the count of large orders increases significantly to 116,941.
- ‘367’ was the only store_id which was among the top one for a total of 50 weeks in terms of foot traffic and sale value.
- Households who are visiting more than 200 times are contributing more towards total monthly spent, however most of the household’s single time purchase for one visit is less than \$4, except for the case of household ‘1023’ whose average purchase stood at \$8.58 for every visit.

- Household number '1609' is the top purchaser among customers for whom we have complete demographic details, spending a substantial total of \$13,804.38. This individual is middle-aged (45-54 years old), belongs to the income bracket of \$125,000-\$149,000, and has a large family with a household size of 5 or more which includes 3 or more kids.
- The highest positive sales difference, when comparing the current week to the previous week (48), was recorded in week 49, with an increase of +\$944.87 for household number '1609'. Interestingly, the highest negative sales difference was also observed for the same household in consecutive week 50, with a decrease of -\$915.12 compared to week 49.
- Households '1035' and '494' are the only ones who didn't take any discounts, but their purchase totals were not that significant—\$109.43 and \$6.08, respectively. Additionally, it is evident that households that purchase more frequently and in larger amounts tend to take advantage of major discounts than those who don't.
- Households '2042', '973', and '1899' have a very high Average Order Value (AOV), each exceeding \$80. However, their total purchases were relatively low, not even surpassing \$6,000, indicating that they are not repeat customers. On the other hand, repeat customers who tend to make more purchases, although these are spread across many visits, which impacts their AOV.
- 'Fluid milk white' is one of our top level-2 commodities, with a buyer percentage of 91.64%. This means that out of every 100 customers who visit our stores, nearly 92 of them purchase it. Following closely are soft drinks (75.44%), bananas (75.28%), and shredded cheese (75.28%) etc. An interesting insight is that all products with a buyer percentage greater than 40% belong to regular groceries and food items.
- Households with income levels ranging from \$150k to \$174k have the highest average order value (AOV), amounting to approximately \$29.60. The second-highest AOV is from those with an income level of \$200k to \$249k, with an AOV of \$27.46, followed by the \$250k+ income group at \$26.86. Interestingly, those earning \$175k to \$199k rank fourth with an AOV of \$25.10, while the \$100k to \$124k income group is in the eighth position, even below the (\$75k - \$99k & \$50k - \$74k) brackets, which occupy the sixth and seventh positions, respectively.
- The 35-44 and 25-34 age groups have nearly identical AOVs, with \$20.4 and \$20.12, respectively. However, the total amount spent by the 35-44 group (\$6.22 L) is almost twice that of the 25-34 group (\$3.89 L). The 65+ group is the lowest-spending demographic, with a total spend of less than 1.5L and an AOV of \$15.08.
- National Brand accounted for nearly three-quarters of sales (72.21%) due to the large number of level-2 commodities listed under it—2,213 compared to just 918 under Private Brand. Interestingly, National Brand also received a larger share of total discounts, surpassing its sales percentage contribution. A significant 76.39% of all discounts were given to National Brand commodities, a difference of more than 4% compared to its sales percentage contribution.
- The highest negative sales difference (-\$34.8K) was observed for Grocery department for the month of 24 when compared to that of 23rd week and in the same manner the highest positive difference (27K) in sales was also observed in the same department in the month of 4 when compared to that of 3rd month.
- The Grocery department recorded the highest negative sales difference of -\$34.8K in the 24th week compared to the 23rd week. Similarly, the highest positive sales difference of +\$27K was also observed in the same department during the 4th month compared to the 3rd month.

- Nearly all of the top 50 highest-selling sub-commodities were purchased in the afternoons and evenings, leaving minimal sales occurring in the mornings and nights.
- Out of all 41 departments, the top five are Grocery, Drug GM, Produce, Meat, and KIOSK-Gas. Conversely, the bottom five are Charitable_count, Gro_Bakery, Toys, HBC, and Meat-WHSE.
- The largest spending group is households with a size of 2 and an income level of \$50-70K, with a total sales value of 2.136L. Additionally, the top 10 highest spenders all belong to households with a size of 3 or fewer and an income level of \$50-74K or below. And majority of lowest spenders are categorized under household size of 3+ and surprisingly they also have high income levels.
- Income level groups of \$50-74K, \$35-49K, \$75-99K, \$25-34K and under 15k are the top spenders in terms of sales, in that order. And they also receive the highest percentage of discounts in total (41% of total discounts)
- Households with higher incomes and smaller family sizes do not make frequent purchases, resulting in a higher Average Order Value (AOV). In contrast, lower-income groups with larger families purchase smaller quantities but make more frequent visits, contributing to a larger share of overall sales, but low AOV.
- Store '367' stood at top in terms of number of customers & sales and surprisingly the total sales were divided almost equally for both evening and afternoon.

Recommendations:

- Since households making higher-value purchases tend to benefit more from discounts, tailor discount strategies to encourage frequent purchases among high-value customers. Consider offering targeted promotions to these groups to boost sales further.
- Given that products like 'Fluid milk white,' soft drinks, bananas, and shredded cheese have high purchase rates, ensure these are well-stocked and prominently displayed along with all other major fast-moving groceries. Consider bundling these popular items with less frequently purchased products to increase overall sales. Expand the variety of level-2 commodities in high-demand categories to capture a larger share of the market.
- Target marketing and product recommendations based on household size and income level. For example, high-income households with smaller family sizes have higher AOV, so personalized and premium product recommendations could be beneficial. Develop strategies to retain high-value customers, such as loyalty programs or exclusive offers.
- Since the majority of top-selling items are purchased in the afternoon and evening, schedule promotions and stock replenishments accordingly. Consider extending store hours or increasing staffing during peak times to better serve customers and capture additional sales.
- The high AOV for specific high-income groups but low total purchases indicate a need for strategies to convert these premium one-time buyers into repeat customers. Implement loyalty programs or subscription models to increase customer retention and frequency of visits. And also, do not neglect the existing base of lower-income groups with larger families who makes more frequent but smaller purchases. Consider value-based purchasing offers or psychological pricing options to increase their AOV.
- Strengthen the performance of bottom-performing departments by analyzing their sales patterns and customer preferences and assessing what's going wrong. Implement targeted marketing and promotional activities to improve their rankings. Invest in enhancing the visibility and appeal of top departments, such as Grocery and Drug GM, to maintain and grow their performance.