E-commerce SQL Analysis

Problem Statement

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

Layout to solve the business case study:

To solve every business problem the following steps will be followed:

- The given business problem will be mentioned.
- Followed by the query which would help us derive the result will be mentioned.
- Followed by the result of the mentioned query.
- An analysis of the results will be followed by insights and appropriate recommendations.

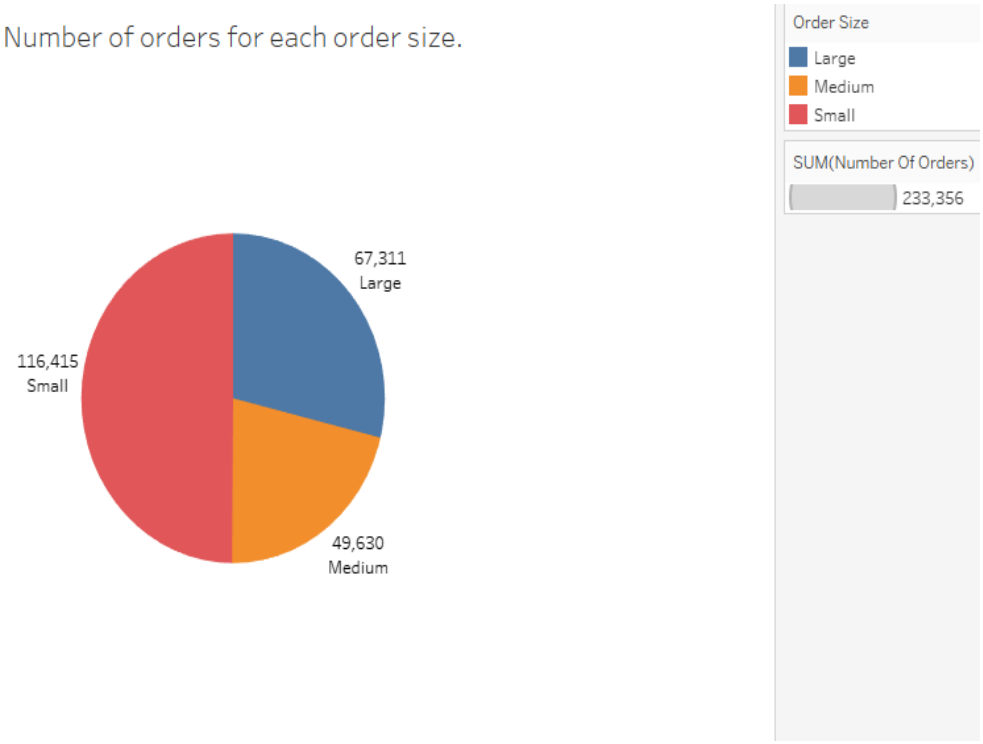
1. Find the number of orders that have small, medium, or large order values (small:0-10 dollars, medium:10-20 dollars, large:20+) ;

```
WITH BasketTotalAmounts AS (
    SELECT
        BASKET_ID,
        SUM(SALES_VALUE) AS TotalAmount
    FROM
        `xyz_ecom.transaction_data`
    GROUP BY
        BASKET_ID
)
```

```sql
SELECT
    CASE
        WHEN TotalAmount <= 10 THEN 'Small'
        WHEN TotalAmount > 10 AND TotalAmount <= 20 THEN 'Medium'
        WHEN TotalAmount > 20 THEN 'Large'
    END AS OrderSize,
    COUNT(BASKET_ID) AS NumberOfOrders
FROM
    BasketTotalAmounts
GROUP BY
    OrderSize
ORDER BY
    OrderSize;
```

Query Results:

Row	OrderSize	NumberOfOrders
1	Large	67311
2	Medium	49630
3	Small	116415



Insights

The query provides a breakdown of the distribution of order sizes based on sales values. The number of distinct basket IDs in each category indicates the popularity of different order sizes.

total sales value for each category provides insights into the contribution of different order sizes to overall revenue. It is clear from the above data that small order size has a higher contribution to the overall revenue. We can also conclude that there are no outliers because no sale is coming from an unknown category.

Recommendations:

Align inventory management strategies with the popularity of different order sizes. Ensure that there are adequate stock levels for products frequently included in specific order size categories.

Implementing or refining customer loyalty programs that reward customers for making larger orders. Loyalty points, exclusive discounts, or other benefits can encourage customers to increase their spending.

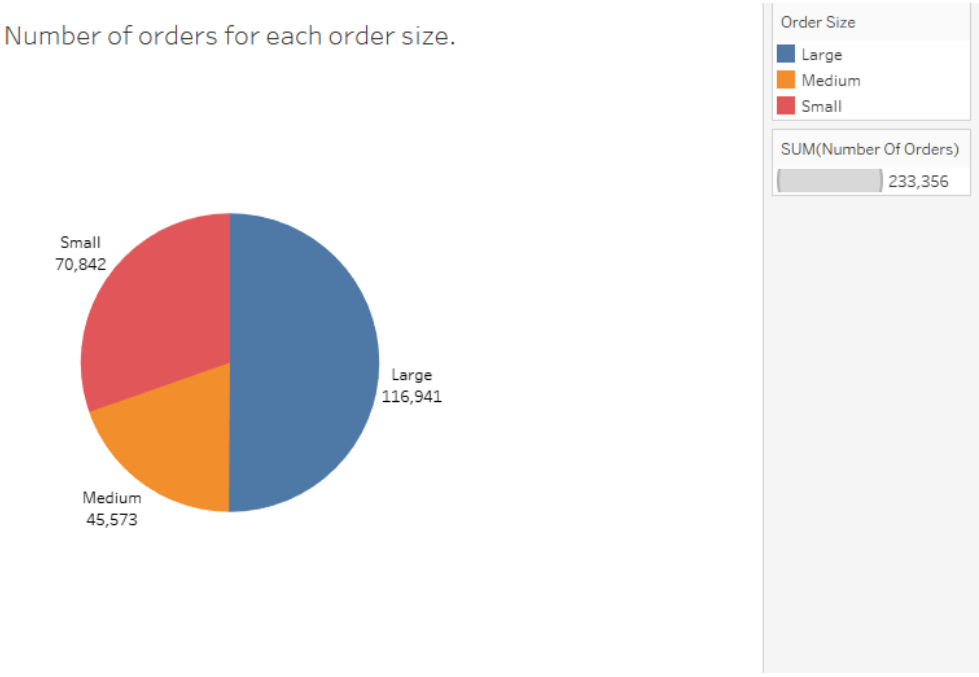
2. Find the number of orders that are small, medium, or large order value(small:0-5 dollars, medium:5-10 dollars, large:10+)

```
WITH BasketTotalAmounts AS (
    SELECT
        BASKET_ID,
        SUM(SALES_VALUE) AS TotalAmount
    FROM
        `xyz_ecom.transaction_data`
    GROUP BY
        BASKET_ID
)

SELECT
    CASE
        WHEN TotalAmount <= 5 THEN 'Small'
        WHEN TotalAmount > 5 AND TotalAmount <= 10 THEN 'Medium'
        WHEN TotalAmount > 10 THEN 'Large'
    END AS OrderSize,
    COUNT(BASKET_ID) AS NumberOfOrders
FROM
```

```
    BasketTotalAmounts
GROUP BY
    OrderSize
ORDER BY
    OrderSize;
```

Query Results:

Row	OrderSize	NumberOfOrders
1	Large	116941
2	Medium	45573
3	Small	70842



Insights and Recommendations:

As we have changed the range size more contribution can be seen from large followed by small and at last medium categories this grouping will help us with customer profiling and will confirm which segment of customers we need to target more and which segment to involve in our target campaigns.
From both queries, we can see that when the total amount is <= 10 that has more contribution toward the total revenue. Thus we should target customers who spend around this amount.

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

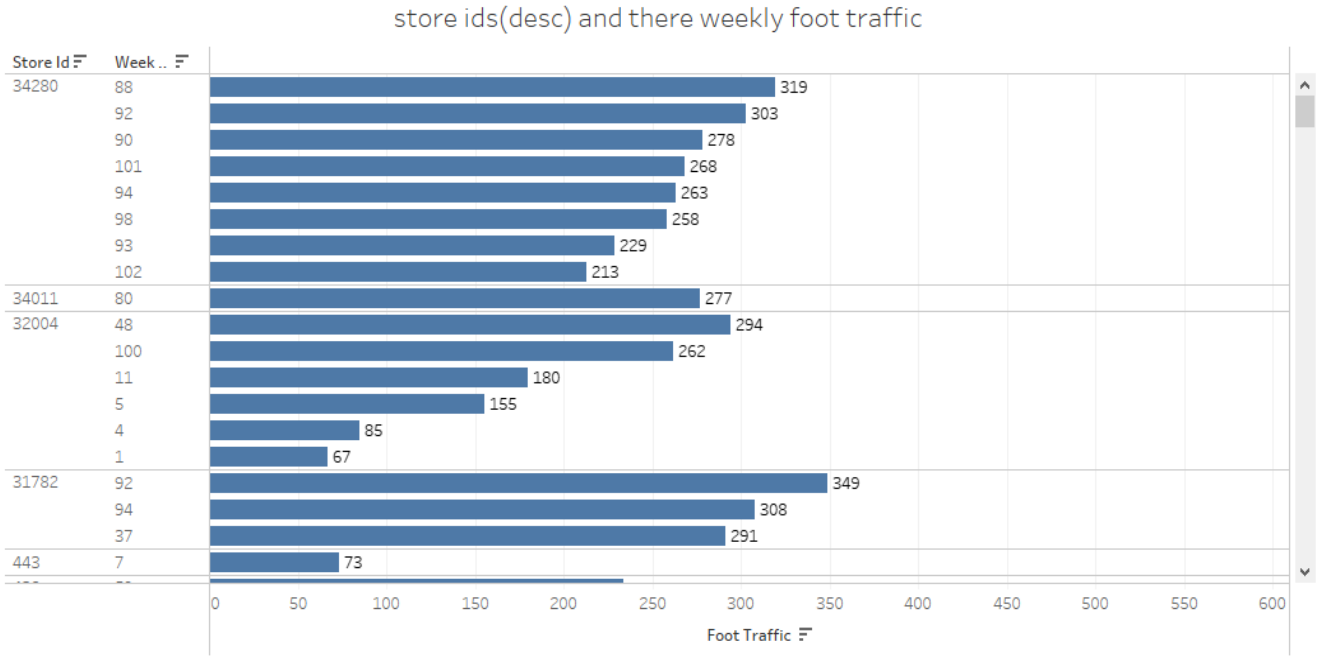
```
WITH RankedStores AS (
  SELECT
    WEEK_NO,
    STORE_ID,
    COUNT(SALES_VALUE) AS FootTraffic,
    ROW_NUMBER() OVER (PARTITION BY WEEK_NO ORDER BY SUM(SALES_VALUE)
DESC) AS Rank
  FROM
    `xyz_ecom.transaction_data`
  GROUP BY
    WEEK_NO, STORE_ID
)
SELECT
  WEEK_NO,
  STORE_ID,
  FootTraffic,Rank
FROM
  RankedStores
WHERE
  Rank <= 3
```

```
ORDER BY
    WEEK_NO, Rank;
```

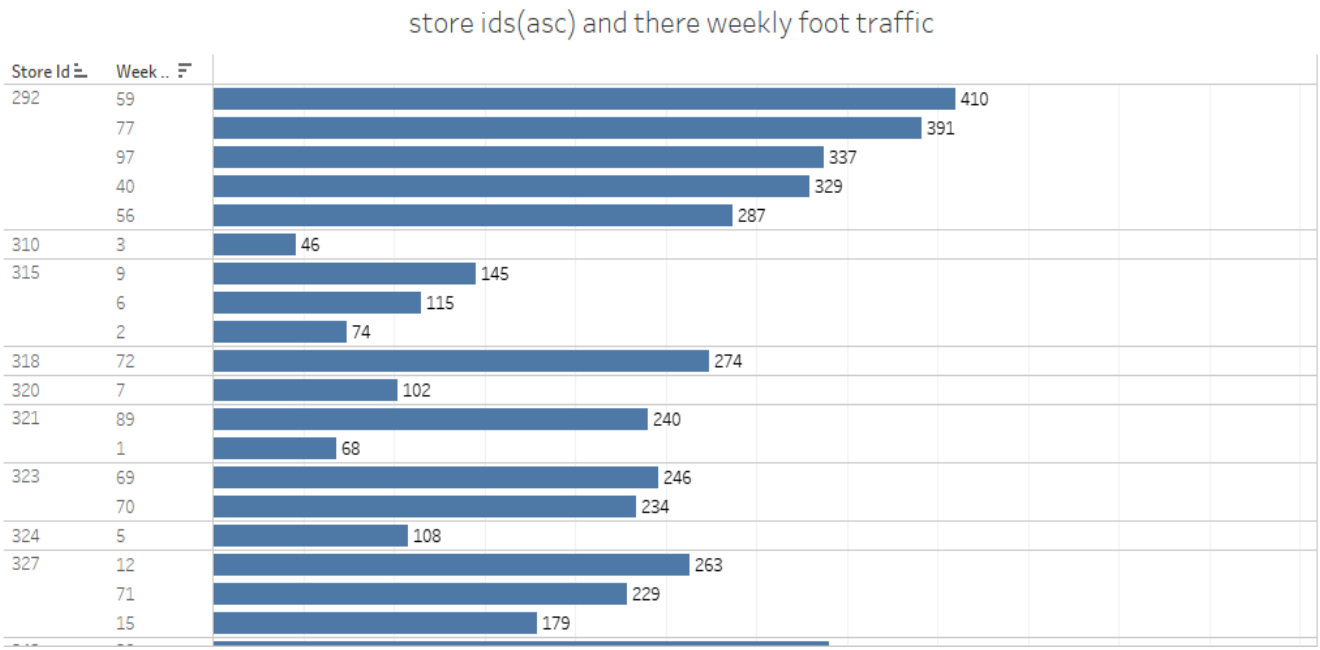
Query Results:

Row	WEEK_NO	STORE_ID	FootTraffic	Rank
1	1	321	68	1
2	1	32004	67	2
3	1	367	40	3
4	2	372	42	1
5	2	375	99	2
6	2	315	74	3
7	3	367	169	1
8	3	375	158	2
9	3	310	46	3
10	4	367	249	1
11	4	375	98	2

Store IDs with the highest weekly foot traffic(among those who are in the top 3)



Store IDs with low weekly foot traffic(among those who are in the top 3)



Insights:

We can see that on different weeks we have different stores in the top 3. On further querying the most commonly occurring store is 367.

Recommendations:

We can increase the loyalty programs in these stores and also identify what kind of locations these stores are in, and analyze what factors are present in these stores that are absent from other stores. Similarly, we can identify stores that are in the bottom three.

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

```
SELECT

    household_key,

    ROUND(avg(sales_value),2) avg_money_per_visit,

    min(WEEK_NO) as First_visit,

    max(WEEK_NO) as last_visit,

    count(DISTINCT BASKET_ID) as no_of_visits,

    ROUND(sum(sales_value),2) as Total_money_spent

FROM `xyz_ecom.transaction_data`

GROUP BY household_key

ORDER BY avg_money_per_visit desc
```

Query Results:

Row	household_key	avg_money_per_visit	First_visit	last_visit	no_of_visits	Total_money_spent
1	1730	16.73	6	102	83	1656.76
2	1727	12.72	16	18	2	114.51
3	2163	10.54	8	97	13	221.32
4	1339	10.42	8	101	6	187.53
5	991	10.26	7	96	18	451.6
6	2219	10.05	12	101	12	321.66
7	2428	10.0	10	101	14	180.0
8	755	9.48	6	102	201	5461.54
9	1023	8.58	16	102	422	18901.09
10	120	8.18	10	94	13	130.92

Insights:

The avg_money_per_visit metric represents the average amount spent by a household during a visit. The First_visit and Last_visit metrics indicate each household's earliest and latest transaction dates. The no_of_visits metric shows the total number of visits for each household. The Total_money_spent metric represents the cumulative amount spent by each household. We can also see that the household with the highest average money spent has made a good no of visits to the store and has continued coming to the store till the time we have data for the week no. The second-highest household key only came to the store for 3 weeks and then stopped coming.

Recommendations:

We need to regain our highest-spending customers who have stopped coming to stores like Household Key 1727 we can use email marketing to offer some special discounts for these kinds of customers and push notifications via SMS.

Household 1730 is the one with the high average spending per visit We can segment customers according to results we have obtained and can also offer loyalty and referral codes to households that have the highest average and total spending and can potentially target segments for premium products or personalized promotions.

We should also offer promos or referrals to those who visit frequently but spend less by giving them exclusive offers which will encourage them to buy high-value products.

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

```
WITH Demo_analysis as (

    SELECT

        household_key,ROUND(sum(sales_value),2) as Total_money_spent

    FROM `xyz_ecom.transaction_data`

    WHERE household_key in (SELECT household_key  FROM
`xyz_ecom.hh_demographic` )

    GROUP BY household_key

    ORDER BY Total_money_spent desc

    LIMIT 1

)

SELECT
d.household_key,d.AGE_DESC,d.MARITAL_STATUS_CODE,d.INCOME_DESC,d.HOM
EOWNER_DESC,d.HH_COMP_DESC,

    d.household_key,d.KID_CATEGORY_DESC

FROM demo_analysis da
```

```
JOIN `xyz_ecom.hh_demographic` d

ON da.household_key = d.household_key
```

Query Results:

Row	household_key	AGE_DESC	MARITAL_STATUS_CODE	INCOME_DESC	HOMEOWNER_DESC	HH_COMP_DESC	household_key_1	KID_CATEGORY_DESC
1	1609	45-54	A	125-149K	Homeowner	2 Adults Kids	1609	3+

Household key: 1609,

AGE_DESC: 45-54

MARITAL_STATUS_CODE:125-149k

INCOME_DESC: A

HOMEOWNER_DESC: Homeowner

HH_COMP_DESC: 2 Adult kids

KID_CATEGORY: 3+

Insights:

Household key 1609 is the most spending customer for whom we have demographic info.

Recommendation:

This one single customer can be sussed for customer profiling we should study the behavior of people with similar features and accordingly target customers and roll out offers. We can also test whether people with 3 or more children are the ones who spend more and if that's true we can introduce more family offers.

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

Approach:

SUB_COMMODITY_DESC: Groups similar products together at the lowest level

Since the column is already grouped similar products we can directly use this to get the count of the most frequently bought product.

```
SELECT SUB_COMMODITY_DESC,count(*) as no_of_products

FROM `xyz_ecom.product`

GROUP BY SUB_COMMODITY_DESC

ORDER BY no_of_products desc
```

Query Results:

Row	SUB_COMMODITY_DESC	no_of_products
1	CARDS EVERYDAY	1005
2	BEERALEMALT LIQUORS	833
3	SPICES & SEASONINGS	629
4	GIFT-WRAP EVERYDAY	547
5	POTATO CHIPS	531
6	MAYBELLINE	525
7	SHAMPOO	518
8	COVERGIRL	517
9	YOGURT NOT MULTI-PACKS	512
10	PREMIUM	495

Insights:

This data helps us with finding combos that are most frequently purchased.

Recommendations:

In order to increase sales, we can introduce offers such as buy this and get this product for a long period of time so that people become habitual of buying them, and then drop this offer after some time, so that people become accustomed to purchasing these products together.

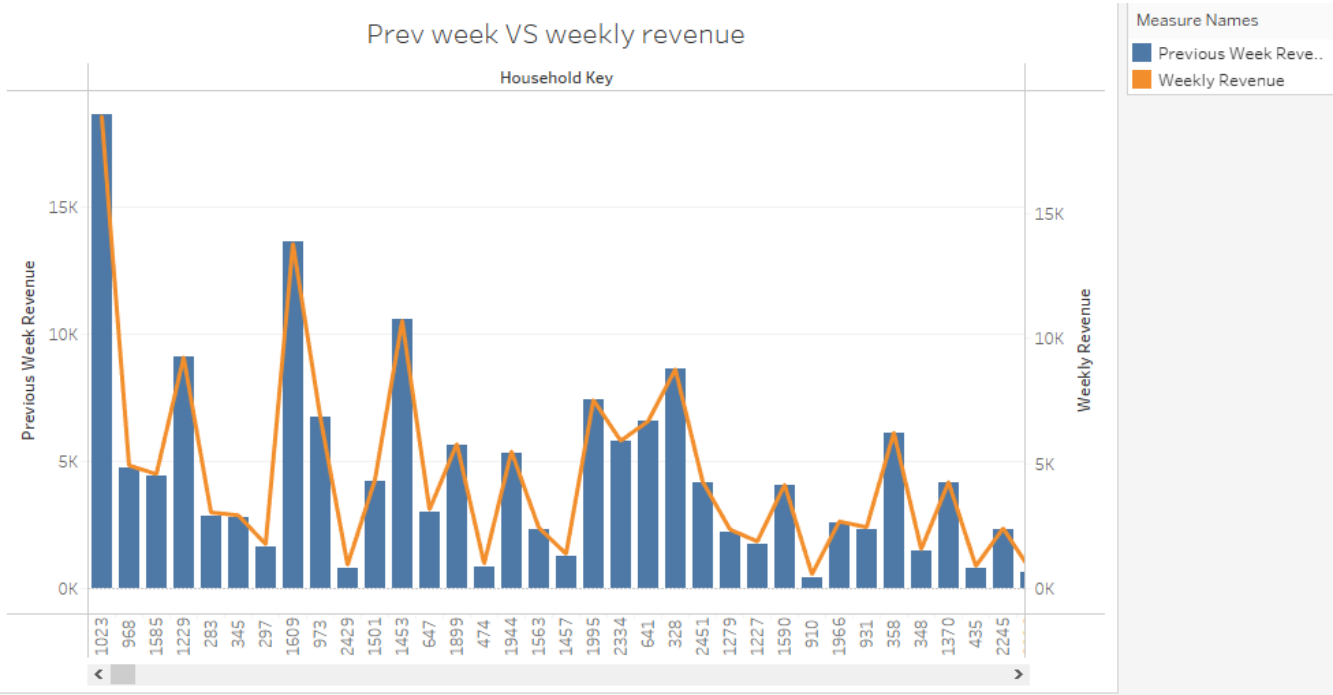
7. Find the weekly change in Revenue Per Account (RPA) (difference in spending by each customer compared to last week)(use lag function)

```
WITH WeeklyRPA AS (
    SELECT
        HOUSEHOLD_KEY,
        WEEK_NO,
        ROUND(SUM(SALES_VALUE),2) AS WeeklyRevenue,
        ROUND(LAG(SUM(SALES_VALUE)) OVER (PARTITION BY HOUSEHOLD_KEY
ORDER BY WEEK_NO),2) AS PreviousWeekRevenue
    FROM
        `xyz_ecom.transaction_data`
    GROUP BY
        HOUSEHOLD_KEY, WEEK_NO
)
SELECT
    HOUSEHOLD_KEY,
    WEEK_NO,
    WeeklyRevenue,
    COALESCE(PreviousWeekRevenue, 0) AS PreviousWeekRevenue,
    ROUND(WeeklyRevenue - COALESCE(PreviousWeekRevenue, 0),2) AS
WeeklyChangeInRPA
FROM WeeklyRPA
```

```
ORDER BY HOUSEHOLD_KEY, WEEK_NO;
```

Query Results:

Row	HOUSEHOLD_KEY	WEEK_NO	WeeklyRevenue	PreviousWeekReven	WeeklyChangeInRPA
1	1	8	42.58	0.0	42.58
2	1	10	14.01	42.58	-28.57
3	1	13	14.03	14.01	0.02
4	1	14	25.71	14.03	11.68
5	1	15	10.98	25.71	-14.73
6	1	16	9.09	10.98	-1.89
7	1	17	13.98	9.09	4.89
8	1	19	47.35	13.98	33.37
9	1	20	31.77	47.35	-15.58
10	1	22	38.98	31.77	7.21



Insights:

Throughout the data, we observe that there isn't much difference between weekly revenues this can be seen in the above visual representation.

Recommendation:

In such cases, we can introduce certain price drops, sales offers, or other exclusive offers and see the impact of the weekly sales WRT to the previous week's sale and in addition to that we can also see which offer is more beneficial for us.

ADDITIONAL QUESTIONS:

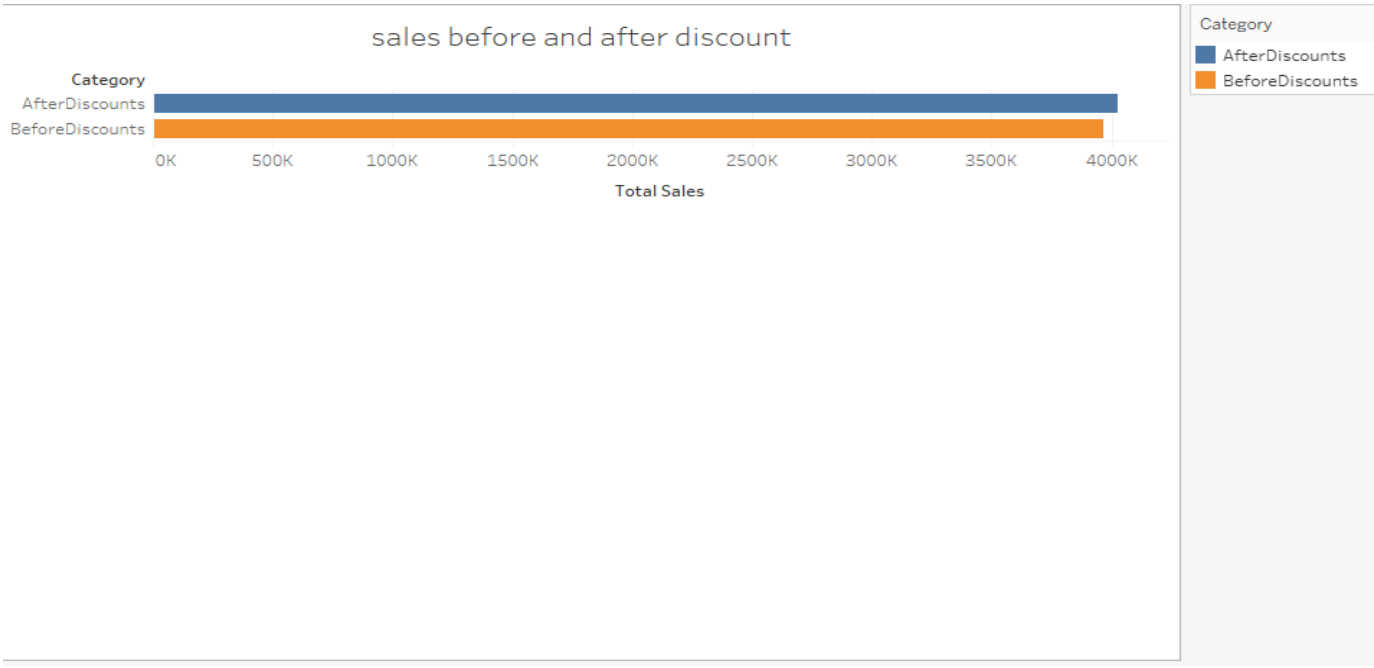
1. Sales before and after discount:

```
SELECT

    'TotalSalesBeforeDiscounts' AS Category,

    SUM(SALES_VALUE) AS TotalSales

FROM

    `xyz_ecom.transaction_data`

WHERE

    COUPON_DISC =0 AND COUPON_MATCH_DISC =0 -- Exclude transactions
with discounts


UNION ALL


SELECT

    'TotalSalesAfterDiscounts' AS Category,
```

```
    SUM(SALES_VALUE) AS TotalSales
FROM
    `xyz_ecom.transaction_data`
```

Query Results:

Row	Category	TotalSales
1	TotalSalesBeforeDiscounts	3965194.160021...
2	TotalSalesAfterDiscounts	4029338.410022...



Insights:

There is little impact on sales after applying discounts. This can be that either the discounts offered are very low or are offered only on a few products or awareness of discounts is not known.

Recommendations:

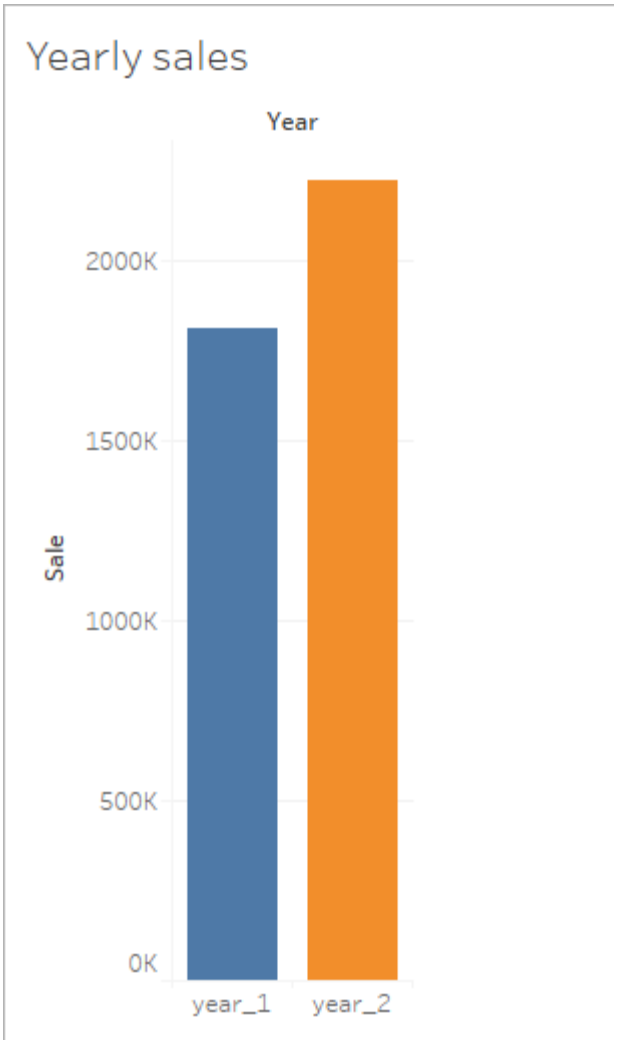
Based on the scenario we can increase the ales with the help of discounts for instance if the discount offered is very low we can increase the discount or if there is less awareness about the discount we can also increase awareness through various channels.

2. Yearly sales trend:

```
WITH yearly_sales as (
  SELECT *,
    CASE
      WHEN WEEK_NO between 1 and 52 THEN 'year_1'
      ELSE 'year_2'
    END AS Year
  FROM `xyz_ecom.transaction_data`

)
SELECT Year,round(sum(sales_value),2) as sale
FROM yearly_sales
GROUP BY Year
```

Query results:

Row	Year	sale
1	year_1	1809125.37
2	year_2	2220213.04



Insights:

The sales have increased as compared to last year, sales show an increasing trend. This can be due to better awareness of the e-commerce company.

Recommendations:

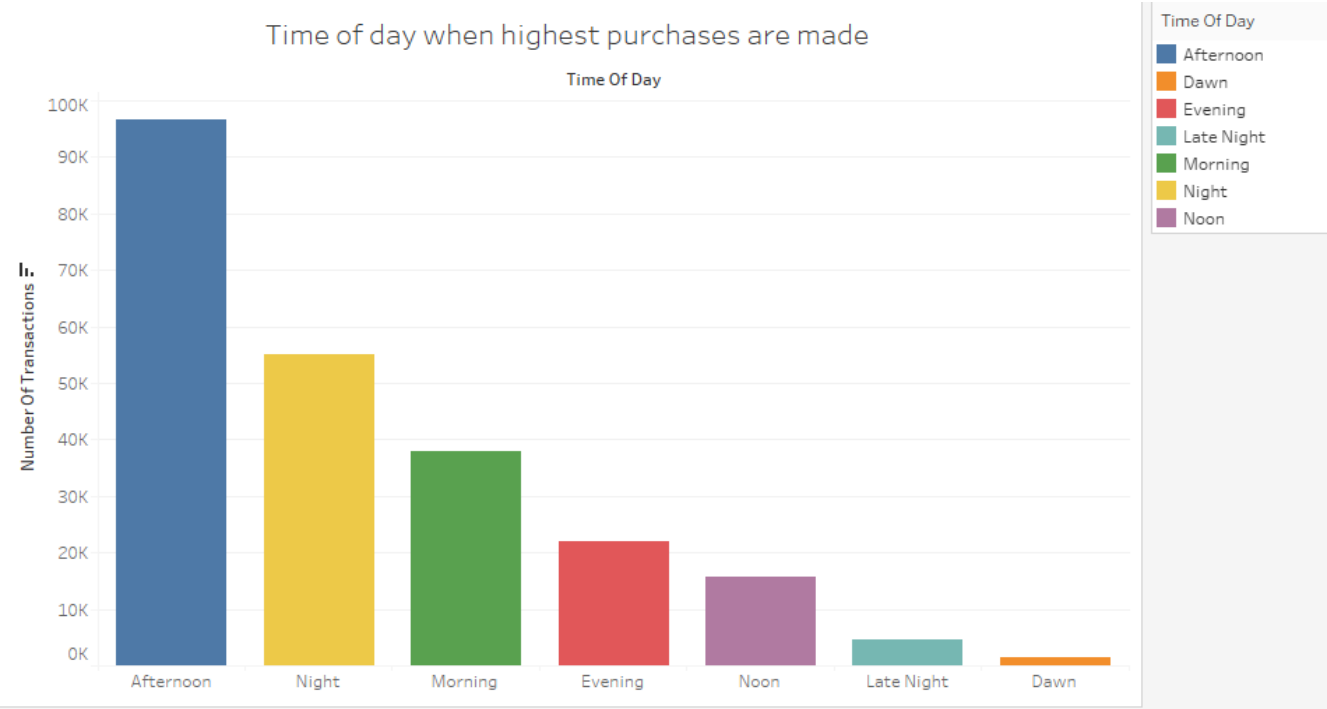
Analyze further on the most common factors which might have led to an increase in sales of the company.

3. Time of day when more purchases are made

```
SELECT
    CASE
        WHEN TRANS_TIME >= 0 AND TRANS_TIME < 600 THEN 'Late Night'
        WHEN TRANS_TIME >= 600 AND TRANS_TIME <= 700 THEN 'Dawn'
        WHEN TRANS_TIME > 700 AND TRANS_TIME <= 1200 THEN 'Morning'
        WHEN TRANS_TIME > 1200 AND TRANS_TIME <= 1300 THEN 'Noon'
        WHEN TRANS_TIME > 1300 AND TRANS_TIME <= 1800 THEN
'Afternoon'
        WHEN TRANS_TIME > 1800 AND TRANS_TIME <= 1900 THEN 'Evening'
        WHEN TRANS_TIME > 1900 AND TRANS_TIME <= 2400 THEN 'Night'
        ELSE 'Unknown'
    END AS TimeOfDay,
    COUNT(DISTINCT BASKET_ID) AS NumberOfTransactions
FROM
    `xyz_ecom.transaction_data`
GROUP BY
    TimeOfDay
ORDER BY
    TimeOfDay;
```

Query Results:

TimeOfDay	NumberOfTransactions
Afternoon	96588
Dawn	1417
Evening	21948
Late Night	4625
Morning	37999
Night	55079
Noon	15700



Insights:

After querying the above data we can conclude that the customers prefer to shop in the afternoon.

Recommendation:

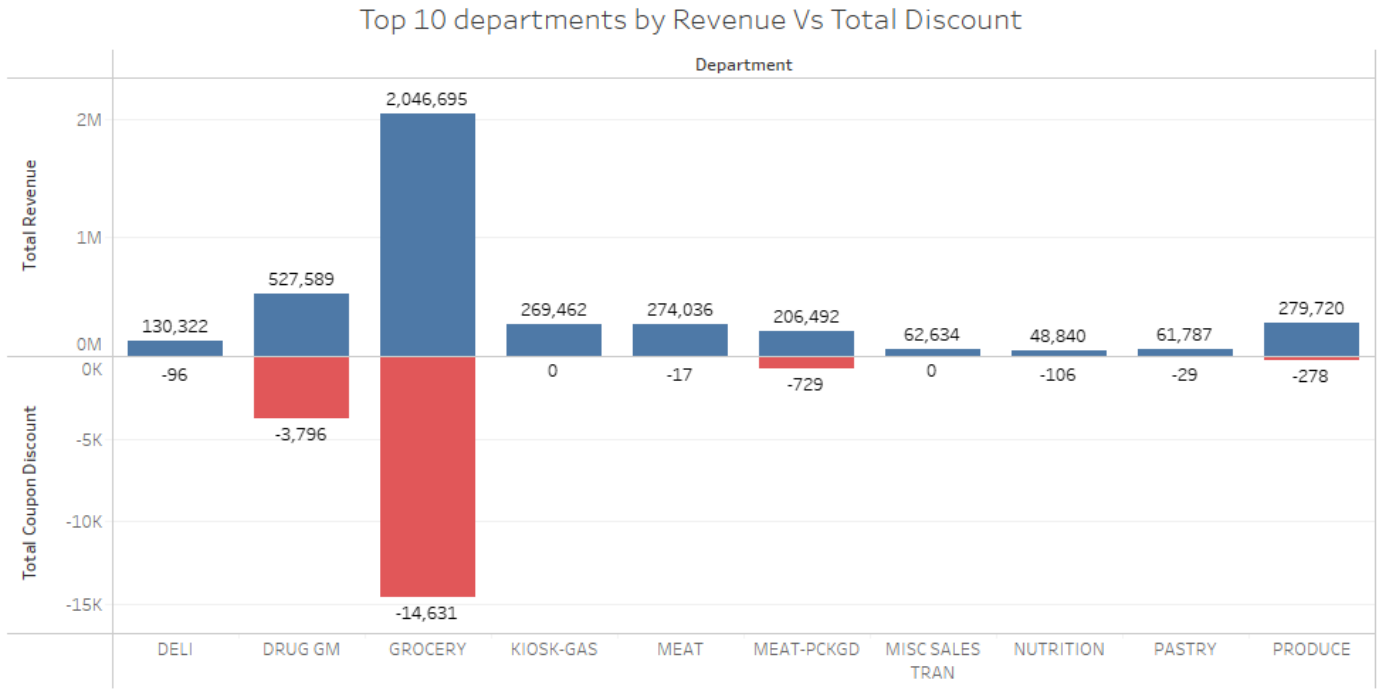
To help the customers have a good experience we can increase the floor staff during the afternoon hours. The maintenance and restocking of inventory can be done during dawn hours which is when fewer customers prefer to shop, also the items with the highest sales or the currently trending items should be restocked before the afternoon. This will help avoid inconvenience to the customers and give them a better experience.

4. Top 10 departments by revenue and coupon discount.

```
SELECT

    p.DEPARTMENT,

    ROUND(SUM(t.SALES_VALUE),2) AS TotalRevenue,

    ROUND(SUM(t.COUPON_MATCH_DISC + t.COUPON_DISC),2) AS
TotalCouponDiscount

FROM

    `xyz_ecom.transaction_data`t

JOIN

    `xyz_ecom.product` p ON t.PRODUCT_ID = p.PRODUCT_ID

GROUP BY

    p.DEPARTMENT

ORDER BY

    TotalRevenue DESC, TotalCouponDiscount DESC

LIMIT 10;
```

Query Results:

Row	DEPARTMENT	TotalRevenue	TotalCouponDiscount
1	GROCERY	2046695.13	-14631.18
2	DRUG GM	527588.65	-3796.01
3	PRODUCE	279720.39	-277.86
4	MEAT	274036.32	-16.8
5	KIOSK-GAS	269461.67	0.0
6	MEAT-PCKGD	206491.71	-728.7
7	DELI	130322.26	-96.2
8	MISC SALES TRAN	62633.95	0.0
9	PASTRY	61786.56	-28.62
10	NUTRITION	48840.39	-105.8



Insights:

These departments are our highest-performing product categories in terms of sales. We can see that departments with 0 discounts like MISC sales tran and Kisk Gas have given 0 discounts which may indicate strong sales without heavy reliance on discounts.

5. Demographic Analysis:

```
SELECT d.AGE_DESC,KID_CATEGORY_DESC, AVG(t.SALES_VALUE*t.QUANTITY)
as amount_spent

FROM `xyz_ecom.hh_demographic` d

JOIN `xyz_ecom.transaction_data` t

ON d.household_key = t.household_key

GROUP BY d.AGE_DESC,KID_CATEGORY_DESC

ORDER BY 3 desc
```

Query Results:

AGE_DESC	KID_CATEGORY_DESC	amount_spent
25-34	2	6608.265077992...
35-44	3+	5637.026256191...
45-54	3+	4950.167910655...
55-64	None/Unknown	4882.317354058...
19-24	2	4399.759266375...
35-44	2	4162.293024863...
35-44	None/Unknown	4102.375676306...
45-54	2	3945.529256215...
25-34	None/Unknown	3872.314862540...
35-44	1	3743.217631733...

Insights:

Customers in the age group "35-44" with "3+" kids and "45-54" with "3+" kids are the top spending segments. They have the highest average amounts spent, indicating a higher purchasing power, possibly driven by family-related expenses.

Recommendations:

Develop targeted marketing campaigns and promotions specifically tailored to families, emphasizing family-friendly products and services. Highlighting discounts or bundled deals for households with "3+" kids can attract and retain high-spending customers.

OVERALL INSIGHTS AND RECOMMENDATIONS:

Total Sales Before and After Discounts:

The analysis shows the total sales before and after applying discounts. The comparison suggests that there isn't a significant impact on sales after applying discounts. This could be due to various reasons such as the magnitude of discounts, product coverage, or customer awareness. Further analysis of the effectiveness of specific discounts and their impact on customer behavior could provide valuable insights.

Yearly Sales Trend:

The query provides insights into the yearly sales trend, categorizing data into two years. The results show an increasing trend in sales, indicating positive growth for the e-commerce company. This could be attributed to improved marketing strategies, increased product offerings, or enhanced customer engagement. Continuing to monitor and analyze yearly trends will help in adapting strategies to sustain or accelerate growth.

Time of Day Analysis:

The time-of-day analysis reveals that customers prefer shopping in the afternoon. This insight can guide operational decisions, such as staffing levels and inventory management. Allocating more staff during peak afternoon hours and strategically planning inventory restocking during less busy periods can enhance the overall customer experience.

Top 10 Departments by Revenue and Coupon Discount:

The query identifies the top 10 departments by total revenue and total coupon discount. This information is valuable for strategic decision-making. Departments with high revenue and coupon discounts may indicate successful promotional strategies, while those with high revenue and low coupon discounts might highlight product popularity without heavy discount reliance. Further analysis of the performance of individual departments can guide marketing and pricing strategies.

Customer Profiling:

The customer profiling query provides insights into the spending behavior of households, including average money spent per visit, first and last visit dates, the number of visits, and total money spent. Understanding customer segments based on spending patterns can inform personalized marketing campaigns, loyalty programs, and product recommendations to maximize customer satisfaction and loyalty.

Single Customer Analysis:

The analysis of the most spending customer with demographic information offers a deeper understanding of a high-value customer. Leveraging demographic data can help tailor marketing strategies and promotions to specific customer segments, enhancing engagement and loyalty. Identifying similar customer profiles can guide targeted marketing efforts.

Weekly Change in Revenue Per Account (RPA):

The query calculates the weekly change in Revenue Per Account (RPA) for each household, comparing spending to the previous week. Understanding fluctuations in customer spending can help identify trends, seasonality, or the impact of specific promotions. This insight can inform dynamic pricing strategies and promotional planning.

Product Associations:

The analysis of products bought together provides insights into product associations. Identifying frequently co-purchased products can guide recommendations, bundling strategies, and targeted promotions to boost cross-selling opportunities.

Demographic Analysis based on Age and No of kids:

Implement personalized offers for different age groups based on spending behaviors. For example, exclusive discounts or loyalty programs for the "65+" age group might enhance customer loyalty and satisfaction.

Customer Engagement Strategies:

Engage with customers in the "55-64" age group to understand their varied spending patterns. Conduct surveys or gather feedback to tailor marketing strategies to meet the diverse preferences within this age segment.

Promotions for the "19-24" Age Group:

Design promotions and marketing strategies to attract and retain customers in the "19-24" age group. Consider offering discounts, exclusive products, or loyalty programs to cater to the preferences of this age demographic.

Overall, these insights and recommendations aim to guide strategic decisions across various aspects of the e-commerce business, from marketing and promotions to customer experience and operational efficiency. Ongoing monitoring and analysis will enable the adaptation of strategies based on changing market dynamics and customer behaviors. It would have been great if the reference point for dates had been provided so that we could combine the three columns i.e. day,trans_time, and week_no to get a better analysis.