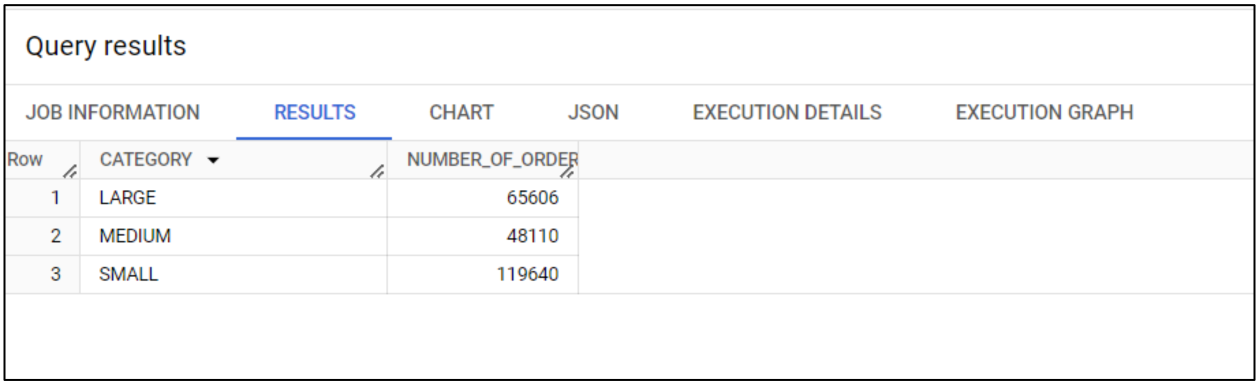
E-commerce SQL Analysis

Problem Statement

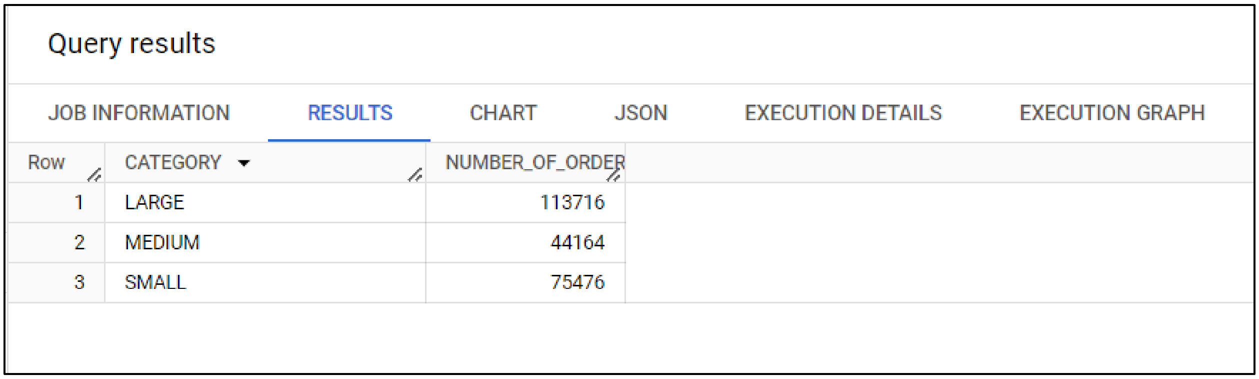
Analyzing the sales, product, and customer data for an e-commerce company. getting various insights and calculating various KPI and data with SQL in Big Query.

```
##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 ORDERVALUE AS
(SELECT BASKET_ID,ROUND(SUM(SALES_VALUE)) AS ORDER_VALUE,
from `Domain_Analytics_Commerce_SQL_Analysis.transaction_data`
GROUP BY 1)

SELECT CATEGORY, COUNT(*) AS NUMBER_OF_ORDERS
FROM (SELECT*,
             CASE WHEN ORDER_VALUE <=10 THEN 'SMALL'
             WHEN ORDER_VALUE <=20 AND ORDER_VALUE >10 THEN 'MEDIUM'
             WHEN ORDER_VALUE >20 THEN 'LARGE' END AS CATEGORY
FROM ORDERVALUE)
GROUP BY 1
ORDER BY 1
```



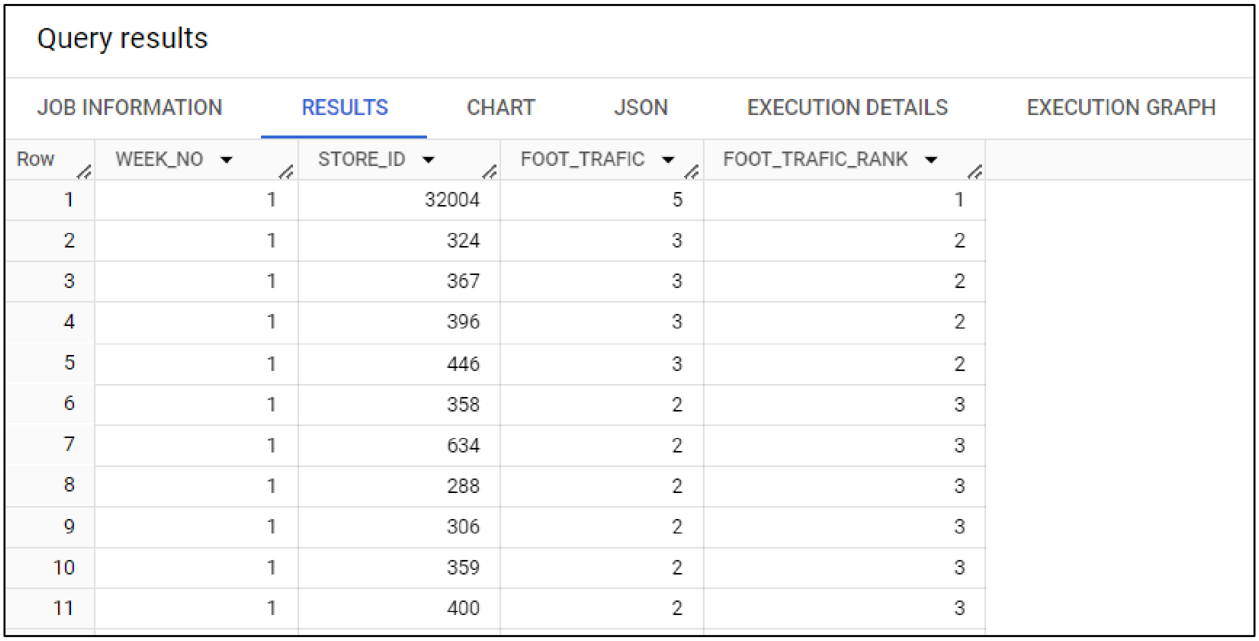
Row	CATEGORY	NUMBER_OF_ORDER
1	LARGE	65606
2	MEDIUM	48110
3	SMALL	119640

```
##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 ORDERVALUE AS
(SELECT BASKET_ID,ROUND(SUM(SALES_VALUE)) AS ORDER_VALUE,
FROM `Domain_Analytics_Commerce_SQL_Analysis.transaction_data`
GROUP BY 1)

SELECT CATEGORY, COUNT(*) AS NUMBER_OF_ORDERS
FROM (SELECT*,
             CASE WHEN ORDER_VALUE <=5 THEN 'SMALL'
             WHEN ORDER_VALUE <=10 AND ORDER_VALUE >5 THEN 'MEDIUM'
             WHEN ORDER_VALUE >10 THEN 'LARGE' END AS CATEGORY
FROM ORDERVALUE)
GROUP BY 1
ORDER BY 1
```



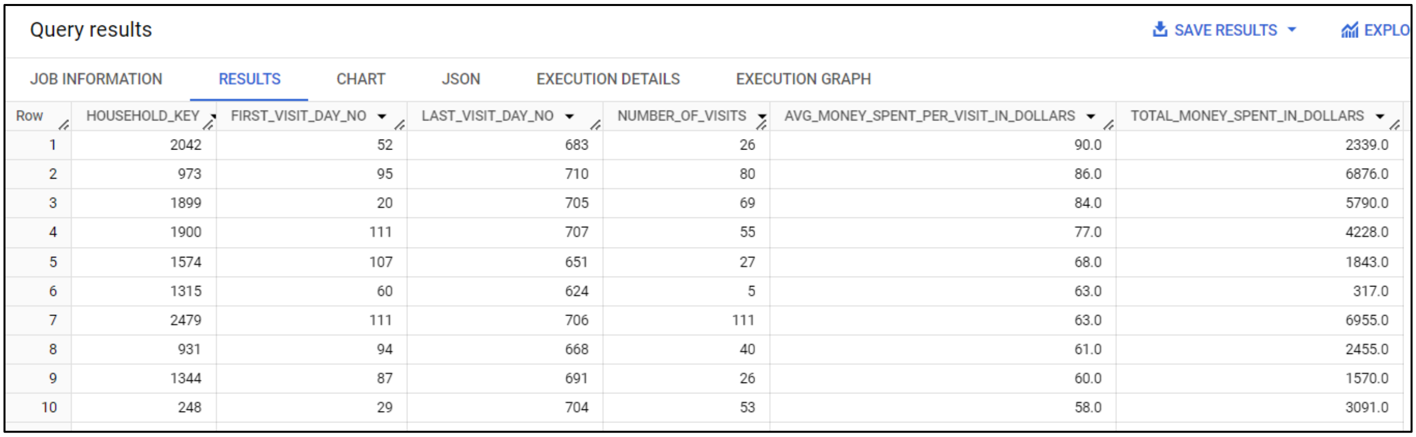
Row	CATEGORY	NUMBER_OF_ORDER
1	LARGE	113716
2	MEDIUM	44164
3	SMALL	75476

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

SELECT *
FROM (SELECT
      WEEK_NO,
      STORE_ID,
      COUNT(DISTINCT HOUSEHOLD_KEY) AS FOOT_TRAFIC,
      DENSE_RANK()OVER(PARTITION BY WEEK_NO ORDER BY (COUNT(DISTINCT HOUSEHOLD_KEY)) DESC)
AS FOOT_TRAFIC_RANK
      FROM `Domain_Analytics_Commerce_SQL_Analysis.transaction_data`
      GROUP BY 1,2) B
WHERE FOOT_TRAFIC_RANK <=3
ORDER BY 1,3 DESC
```



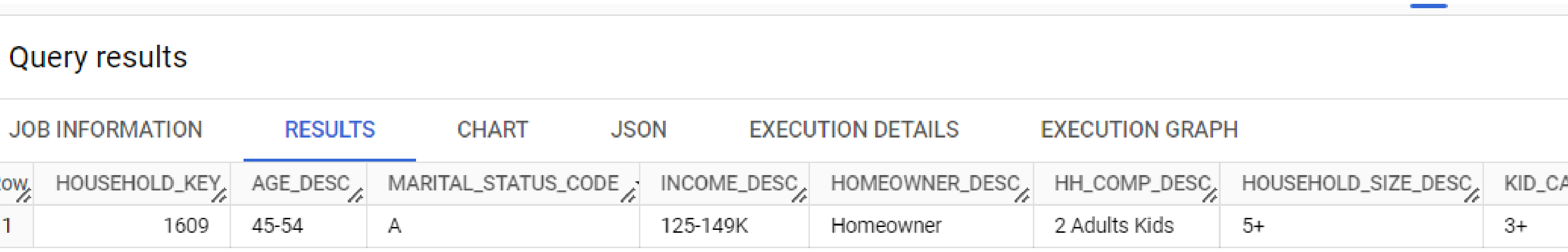
Row	WEEK_NO	STORE_ID	FOOT_TRAFIC	FOOT_TRAFIC_RANK
1	1	32004	5	1
2	1	324	3	2
3	1	367	3	2
4	1	396	3	2
5	1	446	3	2
6	1	358	2	3
7	1	634	2	3
8	1	288	2	3
9	1	306	2	3
10	1	359	2	3
11	1	400	2	3

```
## 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(DAY) AS FIRST_VISIT_DAY_NO,
    MAX(DAY) AS LAST_VISIT_DAY_NO,
    COUNT(DISTINCT BASKET_ID) AS NUMBER_OF_VISITS,
    ROUND(AVG(SUM(SALES_VALUE) / COUNT(DISTINCT BASKET_ID)) OVER (PARTITION BY
HOUSEHOLD_KEY)) AS AVG_MONEY_SPENT_PER_VISIT_IN_DOLLARS,
    ROUND(SUM(SALES_VALUE)) AS TOTAL_MONEY_SPENT_IN_DOLLARS,
FROM `Domain_Analytics_Commerce_SQL_Analysis.transaction_data`
GROUP BY 1
ORDER BY 5 DESC
```

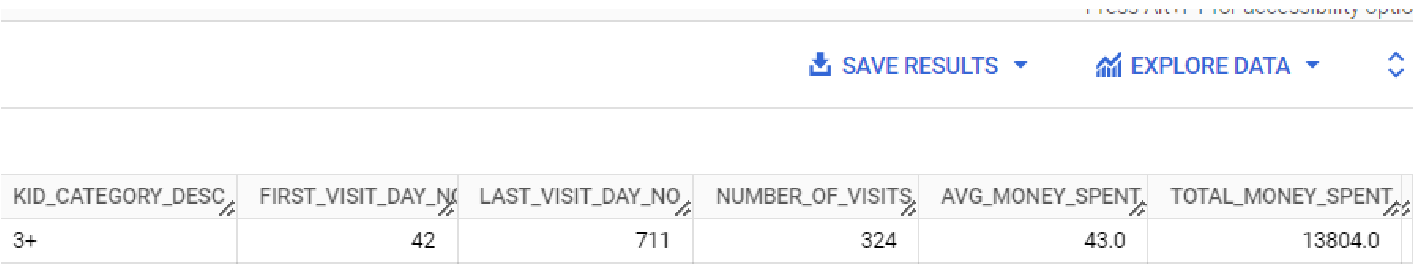


Row	HOUSEHOLD_KEY	FIRST_VISIT_DAY_NO	LAST_VISIT_DAY_NO	NUMBER_OF_VISITS	AVG_MONEY_SPENT_PER_VISIT_IN_DOLLARS	TOTAL_MONEY_SPENT_IN_DOLLARS
1	2042	52	683	26	90.0	2339.0
2	973	95	710	80	86.0	6876.0
3	1899	20	705	69	84.0	5790.0
4	1900	111	707	55	77.0	4228.0
5	1574	107	651	27	68.0	1843.0
6	1315	60	624	5	63.0	317.0
7	2479	111	706	111	63.0	6955.0
8	931	94	668	40	61.0	2455.0
9	1344	87	691	26	60.0	1570.0
10	248	29	704	53	58.0	3091.0

```
##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 SINGLE_CUSTOMER AS
(SELECT *
 FROM `Domain_Analytics_Commerce_SQL_Analysis.hh_demographic`
 LEFT JOIN `Domain_Analytics_Commerce_SQL_Analysis.transaction_data`
 USING (HOUSEHOLD_KEY))

SELECT
    HOUSEHOLD_KEY,
    AGE_DESC,
    MARITAL_STATUS_CODE,
    INCOME_DESC,
    HOMEOWNER_DESC,
    HH_COMP_DESC,
    HOUSEHOLD_SIZE_DESC,
    KID_CATEGORY_DESC,
    MIN(DAY) AS FIRST_VISIT_DAY_NO,
    MAX(DAY) AS LAST_VISIT_DAY_NO,
    COUNT(DISTINCT BASKET_ID) AS NUMBER_OF_VISITS,
    ROUND(AVG(SUM(SALES_VALUE) / COUNT(DISTINCT BASKET_ID)) OVER (PARTITION BY
HOUSEHOLD_KEY)) AS AVG_MONEY_SPENT_PER_VISIT_IN_DOLLARS,
    ROUND(SUM(SALES_VALUE)) AS TOTAL_MONEY_SPENT_IN_DOLLARS,
 FROM SINGLE_CUSTOMER
 GROUP BY 1,2,3,4,5,6,7,8
 ORDER BY 13 DESC
 LIMIT 1
```

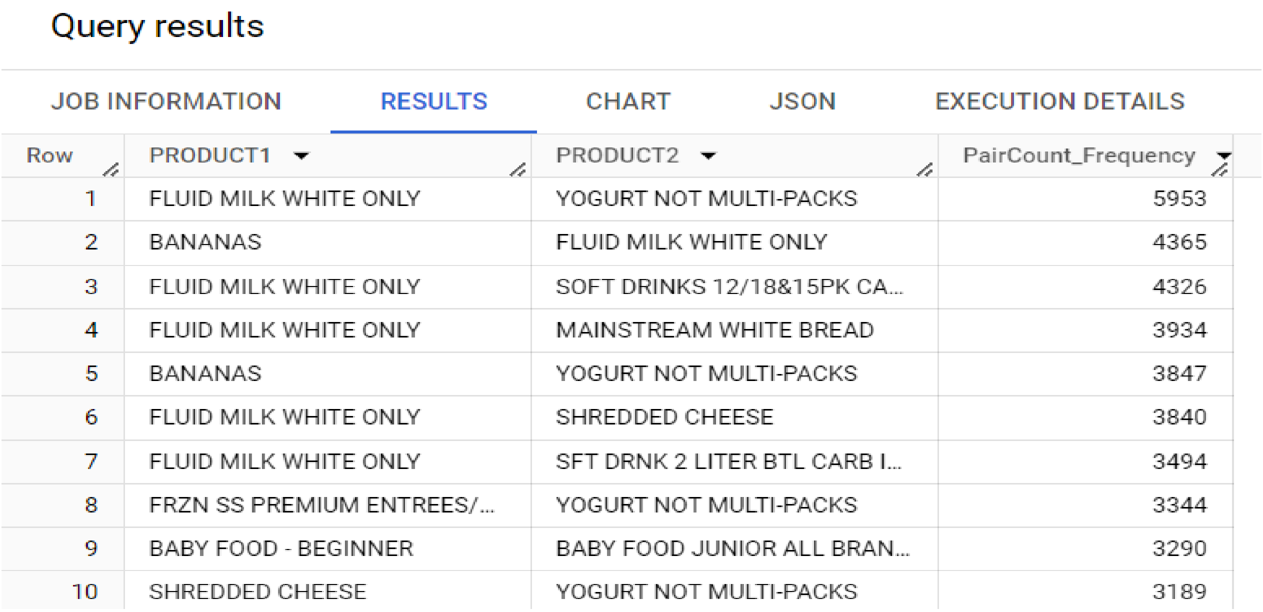


ow	HOUSEHOLD_KEY	AGE_DESC	MARITAL_STATUS_CODE	INCOME_DESC	HOMEOWNER_DESC	HH_COMP_DESC	HOUSEHOLD_SIZE_DESC	KID_CA
1	1609	45-54	A	125-149K	Homeowner	2 Adults Kids	5+	3+



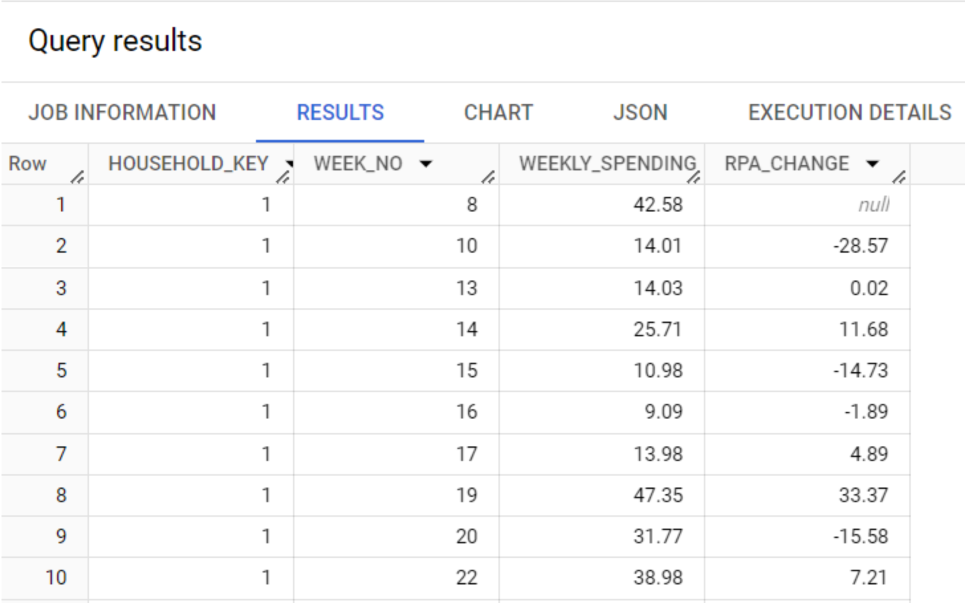
KID_CATEGORY_DESC	FIRST_VISIT_DAY_NO	LAST_VISIT_DAY_NO	NUMBER_OF_VISITS	AVG_MONEY_SPENT	TOTAL_MONEY_SPENT
3+	42	711	324	43.0	13804.0

```
##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 TransactionProducts AS
(SELECT
        t.BASKET_ID,
        p.SUB_COMMODITY_DESC
 FROM `Domain_Analytics_Commerce_SQL_Analysis.transaction_data` t
 JOIN `Domain_Analytics_Commerce_SQL_Analysis.product` p
 ON t.PRODUCT_ID = p.PRODUCT_ID),

ProductPairs AS
(SELECT
        tp1.SUB_COMMODITY_DESC AS PRODUCT1,
        tp2.SUB_COMMODITY_DESC AS PRODUCT2
 FROM TransactionProducts tp1
 JOIN TransactionProducts tp2
 ON tp1.BASKET_ID = tp2.BASKET_ID AND tp1.SUB_COMMODITY_DESC < tp2.SUB_COMMODITY_DESC),

PairCounts AS
(SELECT PRODUCT1, PRODUCT2, COUNT(*) AS PairCount_Frequency
 FROM ProductPairs
 GROUP BY 1,2)

SELECT PRODUCT1, PRODUCT2,PairCount_Frequency
FROM PairCounts
ORDER BY 3 DESC
```



Row	PRODUCT1	PRODUCT2	PairCount_Frequency
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

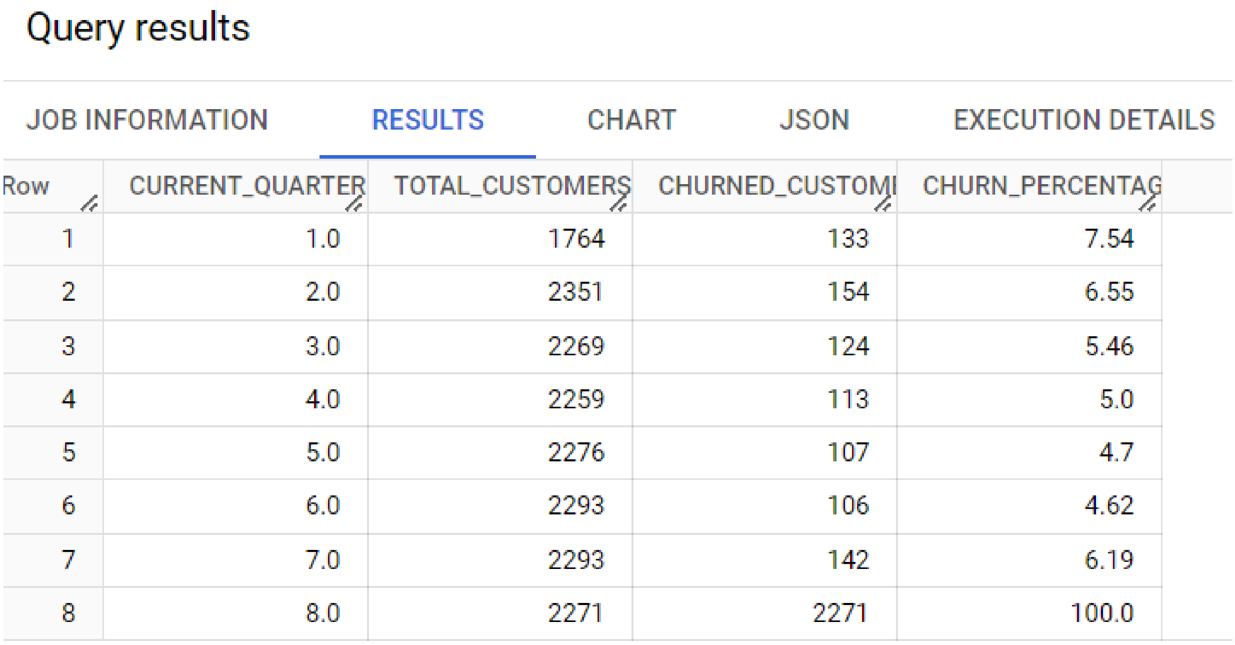
```
##Question 7:
##Find the weekly change in Revenue Per Account (RPA) (difference in spending by each
customer compared to last week)(use lag function)

SELECT
    HOUSEHOLD_KEY,
    WEEK_NO,
    ROUND(SUM(SALES_VALUE),2) AS WEEKLY_SPENDING,
    ROUND(SUM(SALES_VALUE) - LAG(SUM(SALES_VALUE)) OVER (PARTITION BY HOUSEHOLD_KEY ORDER BY
WEEK_NO),2) AS RPA_CHANGE
FROM `Domain_Analytics_Commerce_SQL_Analysis.transaction_data`
GROUP BY 1,2
ORDER BY 1,2
```



Query results

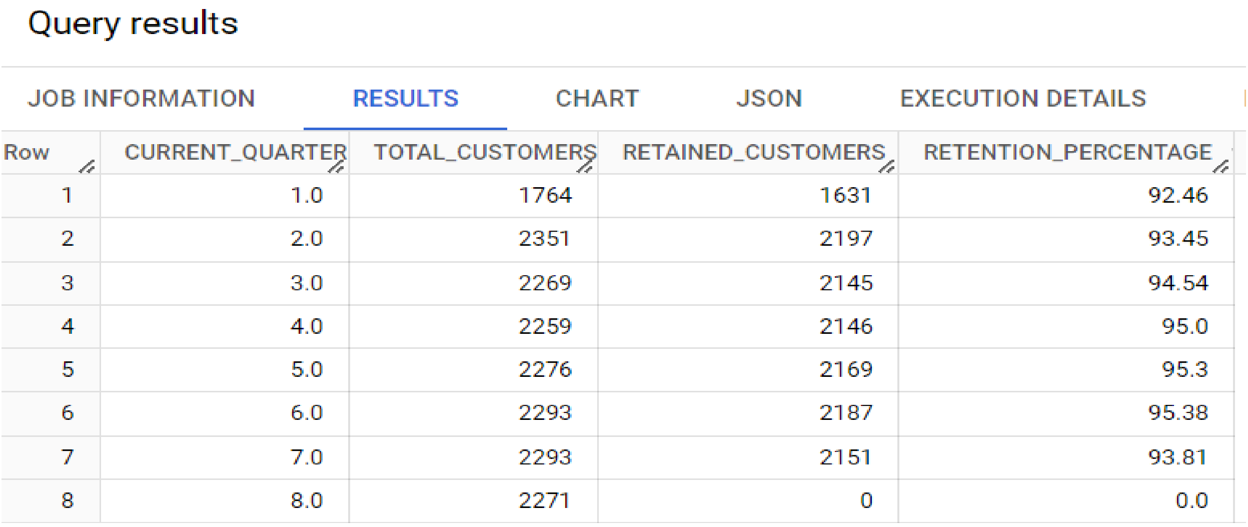
Row	HOUSEHOLD_KEY	WEEK_NO	WEEKLY_SPENDING	RPA_CHANGE
1	1	8	42.58	null
2	1	10	14.01	-28.57
3	1	13	14.03	0.02
4	1	14	25.71	11.68
5	1	15	10.98	-14.73
6	1	16	9.09	-1.89
7	1	17	13.98	4.89
8	1	19	47.35	33.37
9	1	20	31.77	-15.58
10	1	22	38.98	7.21

```
##Question 8:
## Quarter over quarter customer churn

WITH Quarterly_Customers AS
(SELECT
 HOUSEHOLD_KEY,
 CEIL(DAY / 91) AS QUARTER,
 COUNT(DISTINCT BASKET_ID) AS NUM_VISITS
 FROM `Domain_Analytics_Commerce_SQL_Analysis.transaction_data`
 GROUP BY 1,2),

Churn_Calculation AS (
SELECT
    A.HOUSEHOLD_KEY,
    A.QUARTER AS CURRENT_QUARTER,
    B.QUARTER AS NEXT_QUARTER
FROM Quarterly_Customers A
LEFT JOIN Quarterly_Customers B
ON A.HOUSEHOLD_KEY = B.HOUSEHOLD_KEY AND A.QUARTER + 1 = B.QUARTER)

SELECT CURRENT_QUARTER,
    COUNT(DISTINCT HOUSEHOLD_KEY) AS TOTAL_CUSTOMERS,
    COUNT(DISTINCT CASE WHEN NEXT_QUARTER IS NULL THEN HOUSEHOLD_KEY END) AS
CHURNED_CUSTOMERS,
    ROUND(COUNT(DISTINCT CASE WHEN NEXT_QUARTER IS NULL THEN HOUSEHOLD_KEY END) * 100.0 /
COUNT(DISTINCT HOUSEHOLD_KEY), 2) AS CHURN_PERCENTAGE
FROM Churn_Calculation
GROUP BY 1
ORDER BY 1
```



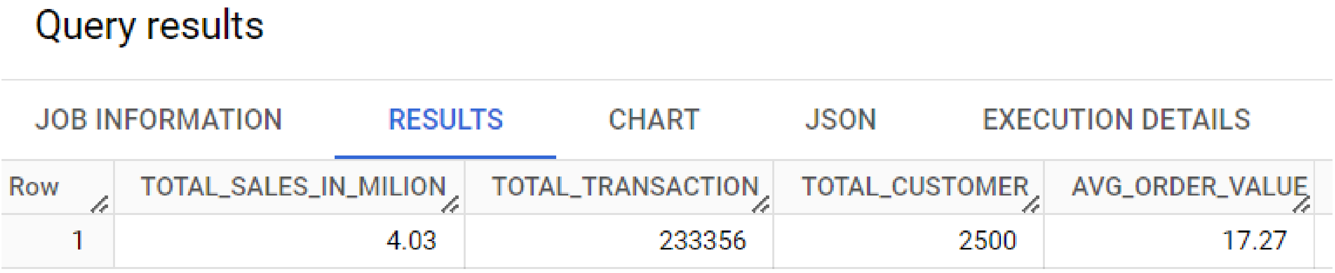
Row	CURRENT_QUARTER	TOTAL_CUSTOMERS	CHURNED_CUSTOMI	CHURN_PERCENTAG
1	1.0	1764	133	7.54
2	2.0	2351	154	6.55
3	3.0	2269	124	5.46
4	4.0	2259	113	5.0
5	5.0	2276	107	4.7
6	6.0	2293	106	4.62
7	7.0	2293	142	6.19
8	8.0	2271	2271	100.0

```
##Question 9:

## Q QoQ coustmer retained
WITH Quarterly_Customers AS (
SELECT
    HOUSEHOLD_KEY,
    CEIL(DAY / 91) AS QUARTER
FROM `Domain_Analytics_Commerce_SQL_Analysis.transaction_data`
GROUP BY 1,2),

Retention_Calculation AS (
SELECT
    A.HOUSEHOLD_KEY,
    A.QUARTER AS CURRENT_QUARTER,
    B.QUARTER AS NEXT_QUARTER
FROM Quarterly_Customers A
LEFT JOIN Quarterly_Customers B
ON A.HOUSEHOLD_KEY = B.HOUSEHOLD_KEY AND A.QUARTER + 1 = B.QUARTER)

SELECT
    CURRENT_QUARTER,
    COUNT(DISTINCT HOUSEHOLD_KEY) AS TOTAL_CUSTOMERS,
    COUNT(DISTINCT CASE WHEN NEXT_QUARTER IS NOT NULL THEN HOUSEHOLD_KEY END) AS
RETAINED_CUSTOMERS,
    ROUND(COUNT(DISTINCT CASE WHEN NEXT_QUARTER IS NOT NULL THEN HOUSEHOLD_KEY END) * 100.0
/ COUNT(DISTINCT HOUSEHOLD_KEY), 2) AS RETENTION_PERCENTAGE
FROM Retention_Calculation
GROUP BY 1
ORDER BY 1
```



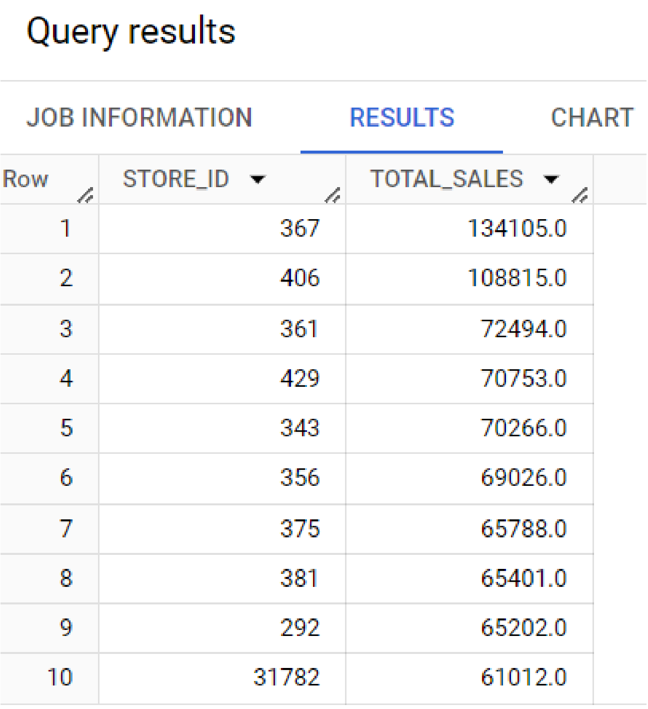
Row	CURRENT_QUARTER	TOTAL_CUSTOMERS	RETAINED_CUSTOMERS	RETENTION_PERCENTAGE
1	1.0	1764	1631	92.46
2	2.0	2351	2197	93.45
3	3.0	2269	2145	94.54
4	4.0	2259	2146	95.0
5	5.0	2276	2169	95.3
6	6.0	2293	2187	95.38
7	7.0	2293	2151	93.81
8	8.0	2271	0	0.0

```
##Question 10:
## total no of sales, total no of transaction, customer count, average order value

SELECT  ROUND(SUM(SALES_VALUE)/1000000,2) AS TOTAL_SALES_IN_MILION,
        COUNT(DISTINCT BASKET_ID) AS TOTAL_TRANSACTION,
        COUNT(DISTINCT HOUSEHOLD_KEY) AS TOTAL_CUSTOMER,
        ROUND(SUM(SALES_VALUE)/COUNT(DISTINCT BASKET_ID),2) AS AVG_ORDER_VALUE
FROM `Domain_Analytics_Commerce_SQL_Analysis.transaction_data`
```



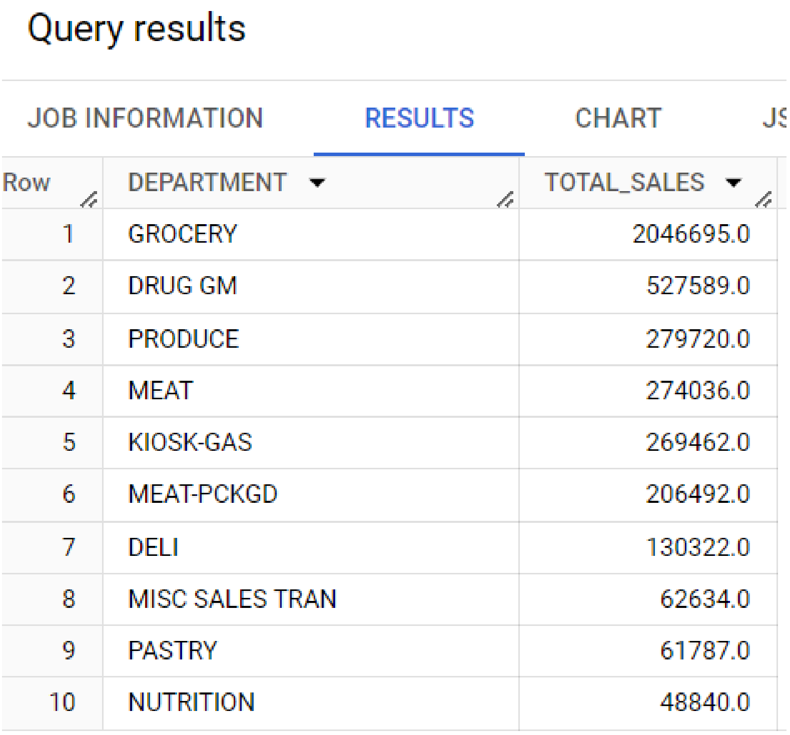
Row	TOTAL_SALES_IN_MILION	TOTAL_TRANSACTION	TOTAL_CUSTOMER	AVG_ORDER_VALUE
1	4.03	233356	2500	17.27

```
##Question 11:
## Top 10 Stores in the terms of sale

SELECT STORE_ID, ROUND(SUM(SALES_VALUE)) AS TOTAL_SALES
FROM `Domain_Analytics_Commerce_SQL_Analysis.transaction_data`
GROUP BY 1
ORDER BY 2 DESC
LIMIT 10
```



Row	STORE_ID	TOTAL_SALES
1	367	134105.0
2	406	108815.0
3	361	72494.0
4	429	70753.0
5	343	70266.0
6	356	69026.0
7	375	65788.0
8	381	65401.0
9	292	65202.0
10	31782	61012.0

```
##Question 12:
## Top 10 Department in the terms of sale

SELECT DEPARTMENT, ROUND(SUM(SALES_VALUE))
FROM `Domain_Analytics_Commerce_SQL_Analysis.transaction_data` T
LEFT JOIN `Domain_Analytics_Commerce_SQL_Analysis.product` P
ON T.PRODUCT_ID=P.PRODUCT_ID
GROUP BY 1
ORDER BY 2 DESC
LIMIT 10
```



Row	DEPARTMENT	TOTAL_SALES
1	GROCERY	2046695.0
2	DRUG GM	527589.0
3	PRODUCE	279720.0
4	MEAT	274036.0
5	KIOSK-GAS	269462.0
6	MEAT-PCKGD	206492.0
7	DELI	130322.0
8	MISC SALES TRAN	62634.0
9	PASTRY	61787.0
10	NUTRITION	48840.0

Insights / Recommendation

1. Total sales value is 4.03 M
2. Total no of products 72 thousand
3. Total no of customer is 2500
4. Over 2 lakh transaction done.
5. Avg order value per transaction is 17
6. More medium range of transaction has been done in the store
7. Most transaction performed by customer having age group 45-54 with income band in between 50-74k
8. Grocery department having highest sales.
9. Customer retention is above 90% across all quarter and customer churn rate averaged around 6-8 % in all quarter.
10. In week 50-100 there is decline in revenue from customers.
11. Milk and yoghurt is highest selling combination followed by fruit and beverage category
12. Highest Avg spending from a customer is 90.

By
Smruti prakash