

Optional Question 1

Total Revenue, Total\_Orders,Total\_Stores,Total\_Customers,No\_Of\_Products

Query:

```
SELECT ROUND(SUM(SALES_VALUE),2) as TotalRevenue ,
       count(distinct BASKET_ID) as nooforders,
       count(distinct STORE_ID ) as No_of_Stores,
       count(distinct household_key) as No_of_customer,
       count(distinct PRODUCT_ID) AS No_of_products
FROM `Ecommerce.transaction`;
```

Output:

Row	TotalRevenue	nooforders	No_of_Stores	No_of_customer	No_of_products
1	4029338.41	233356	552	2500	71540

Insight's:

- Business has Total Revenue of 4,029,338 with Total Order of 233,356
- Business has total Store of 552 with 2500 Customer
- Business is catering 71,540 Type of Products

Optional Question 2

Revenue Trend Over all weeks

Query:

```
SELECT WEEK_NO, ROUND(SUM(SALES_VALUE)) AS TOTAL_REVENUE
FROM `Ecommerce.transaction`
GROUP BY WEEK_NO
ORDER BY WEEK_NO
```

Output:

Row	WEEK_NO	TOTAL_REVENUE
1	1	2454.0
2	2	5854.0
3	3	6823.0
4	4	7915.0
5	5	10039.0
6	6	12567.0
7	7	14042.0
8	8	15680.0
9	9	15654.0
10	10	21458.0

Row	WEEK_NO	TOTAL_REVENUE
1	102	36112.0
2	101	42201.0
3	100	45634.0
4	99	50361.0
5	98	49448.0
6	97	44594.0
7	96	43258.0
8	95	44113.0
9	94	49079.0
10	93	40455.0
11	92	57722.0

Insight's:

- Weekly Revenue Increase from 2454 to peak of 57722

Optional Question 3

-- Average Revenue per visit per store wise

Query:

```
SELECT STORE_ID, AVG(SALES_VALUE) AS AVERAGE_SALES_VISIT
FROM `Ecommerce.transaction`
GROUP BY STORE_ID
ORDER BY AVERAGE_SALES_VISIT DESC;
```

Output

:

Row	STORE_ID	AVERAGE_SALES_VISIT
1	3065	72.0
2	489	51.6
3	3098	43.62
4	3422	39.99
5	144	39.0
6	896	30.0
7	648	29.48
8	84	27.01
9	771	24.669999999999995
10	3090	24.4

Row	STORE_ID	AVERAGE_SALES_VI
1	1235	0.0
2	3932	0.25
3	681	0.4
4	564	0.41
5	2825	0.4350000000000...
6	751	0.585
7	472	0.74
8	1132	0.749565217391...
9	256	0.76
10	26	0.875

Insight's:

- Average Sale Values per Visit Ranges from 0.0 ~ 72
- There is definite need to measure the in-out merit for existence of some of the stores and close them for better resource management

## Optional Question 4

-- Average Revenue per customer , No of Orders , Revenue per customer

### Query:

```
SELECT household_key,ROUND(SUM(SALES_VALUE)/COUNT(DISTINCT BASKET_ID),2) AS
Avg_revenue_per_Customer
      ,COUNT(DISTINCT BASKET_ID) as No_of_order
      ,ROUND(SUM(SALES_VALUE),2) AS Total_Revenue_per_customer
FROM `Ecommerce.transaction`
group by 1
order by 2 desc;
```

### Output

Row	household_key	Avg_revenue_per_Cu	No_of_order	Total_Revenue_per_c
1	2042	89.97	26	2339.21
2	973	85.95	80	6875.89
3	1899	83.91	69	5789.59
4	1900	76.87	55	4227.72
5	1574	68.27	27	1843.3
6	1315	63.48	5	317.39
7	2479	62.65	111	6954.64
8	931	61.38	40	2455.29
9	1344	60.4	26	1570.37
10	248	58.32	53	3090.89

Row	household_key	Avg_revenue_per_Cu	No_of_order	Total_Revenue_per_c
1	2337	5.61	1021	5726.25
2	1510	6.32	917	5799.9
3	900	9.89	848	8386.39
4	1795	5.85	837	4896.15
5	2459	14.48	712	10307.55
6	1901	9.34	646	6030.94
7	1228	12.26	603	7394.22
8	1479	7.73	597	4617.22
9	1453	18.14	591	10720.72
10	762	7.45	590	4395.92

Row	household_key	Avg_revenue_per_Cu	No_of_order	Total_Revenue_per_c
1	1023	44.79	422	18901.09
2	1609	42.61	324	13804.38
3	2322	41.15	290	11934.66
4	1453	18.14	591	10720.72
5	2459	14.48	712	10307.55
6	1430	39.95	254	10147.21
7	718	19.83	483	9577.63
8	1111	33.6	284	9542.2
9	1653	20.79	458	9519.93
10	400	34.6	274	9481.19

### Insight's:

- Maximum Average Revenue per customer is Highest for household\_key 2042 with value 89.96 and total sales revenue 2339
- Maximum no of Orders placed by household\_key 2337 with total order 1021 with Total sales Revenue 5726
- But Maximum Sales Revenue was mde by household\_key 1023 with total sales Revenue of 18,901
- So No. of Orders & Average sales per Customer cannot be only indicators for Customer Segmentation , a collective approach need to set Marketing Strategies for each customer based on Recency Frequency Monetary Analysis

# Optional Question 5

## -- RFM Analysis and Customer Segmentation

Query:

```
SELECT RFM_Segment ,count(distinct household_key) as customer_count
from
(SELECT *,
CASE WHEN (Rrank = 5 AND FMrank = 5) OR (Rrank = 5 AND FMrank = 4) OR (Rrank = 4 AND FMrank = 5) THEN '1Champions'
      WHEN (Rrank = 5 AND FMrank =3) OR (Rrank = 4 AND FMrank = 4) OR (Rrank = 3 AND FMrank = 5) OR (Rrank = 3 AND
FMrank = 4) THEN '2Loyal Customers'
      WHEN (Rrank = 5 AND FMrank = 2) OR (Rrank = 4 AND FMrank = 2) OR (Rrank = 3 AND FMrank = 3) OR (Rrank = 4 AND
FMrank = 3) THEN '3Potential Loyalists'
      WHEN Rrank = 5 AND FMrank = 1 THEN '4Recent Customers'
      WHEN (Rrank = 4 AND FMrank = 1) OR (Rrank = 3 AND FMrank = 1) THEN '5Promising'
      WHEN (Rrank = 3 AND FMrank = 2) OR (Rrank = 2 AND FMrank = 3) OR (Rrank = 2 AND FMrank = 2) THEN '6Customers
Needing Attention'
      WHEN Rrank = 2 AND FMrank = 1 THEN '7About to Sleep'
      WHEN (Rrank = 2 AND FMrank = 5) OR (Rrank = 2 AND FMrank = 4) OR (Rrank = 1 AND FMrank = 3) THEN '8At Risk'
      WHEN (Rrank = 1 AND FMrank = 5) OR (Rrank = 1 AND FMrank = 4) THEN '90Cant Lose Them'
      WHEN Rrank = 1 AND FMrank = 2 THEN '91Hibernating'
      WHEN Rrank = 1 AND FMrank = 1 THEN '92Lost'
END AS RFM_Segment
from
(
SELECT household_key , ntile(5) over(order by Recency desc) as Rrank,
      ROUND((ntile(5) over(order by Monetary asc) + ntile(5) over(order by Frequency asc))/2,0) as FMrank
from
(
select household_key,COUNT(DISTINCT BASKET_ID) as Frequency,
      ROUND(SUM(SALES_VALUE),2) AS Monetary,
      (select max(DAY) FROM `Ecommerce.transaction`) - MAX(DAY) AS Recency
from `Ecommerce.transaction`
group by household_key)
)
group by RFM_Segment
order by RFM_Segment ;
```

### Output

Row	RFM_Segment	customer_count
1	1Champions	529
2	2Loyal Customers	467
3	3Potential Loyalists	341
4	4Recent Customers	22
5	5Promising	60
6	6Customers Needing Attention	367
7	7About to Sleep	67
8	8At Risk	216
9	90Cant Lose Them	63
10	91Hibernating	139
11	92Lost	229

### Insights:

- Business has Good Amount Champion Customer and Also Customer who need Major attention
- From RFM Segmentation, we can see How Customer are Categorized into Different segments
- With this Segmentation , Marketing Strategies can be fine-tuned to increase Revenue and Engagement for efficient use of Resources

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:

```
SELECT A.SALES_CATEGORY, COUNT(A.SALES_CATEGORY) AS COUNT
FROM
(
  SELECT BASKET_ID, CASE
    WHEN ROUND(SUM(SALES_VALUE)) BETWEEN 0 AND 10 THEN "SMALL"
    WHEN ROUND(SUM(SALES_VALUE)) BETWEEN 10 AND 20 THEN "MEDIUM"
    WHEN ROUND(SUM(SALES_VALUE)) > 20 THEN "LARGE"
  END AS SALES_CATEGORY
FROM `Ecommerce.transaction`
group by BASKET_ID
) A
GROUP BY A.SALES_CATEGORY
ORDER BY 2 DESC
```

Output:

Row	SALES_CATEGORY	COUNT
1	SMALL	119640
2	LARGE	65606
3	MEDIUM	48110

Insight's:

- Most of the order fall in Small Category where Small Category means order value between 0~10
- Next come Large with order value range > 20
- And then come Medium with order value range between 10~20

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:

```
SELECT A.SALES_CATEGORY, COUNT(A.SALES_CATEGORY) AS COUNT
FROM
(
  SELECT BASKET_ID, CASE
    WHEN ROUND(SUM(SALES_VALUE)) BETWEEN 0 AND 5 THEN "SMALL"
    WHEN ROUND(SUM(SALES_VALUE)) BETWEEN 5 AND 10 THEN "MEDIUM"
    WHEN ROUND(SUM(SALES_VALUE)) > 10 THEN "LARGE"
  END AS SALES_CATEGORY
FROM `Ecommerce.transaction`
group by BASKET_ID
) A
GROUP BY A.SALES_CATEGORY
ORDER BY 2 DESC
```

Output:

Row	SALES_CATEGORY	COUNT
1	LARGE	113716
2	SMALL	75476
3	MEDIUM	44164

Insight's:

- Most of the order fall in Large Category where Large Category means order value >10
- Next come Small Category with order value range 0~5
- And then come Medium with order value range between 5~10

### Question 3

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

#### Query:

```
SELECT A.WEEK_NO, MIN(CASE WHEN A.RANKING = 1 THEN A.STORE_ID END) AS HIGHEST,
      MIN(CASE WHEN A.RANKING = 2 THEN A.STORE_ID END) AS SECOND_HIGHEST,
      MIN(CASE WHEN A.RANKING = 3 THEN A.STORE_ID END) AS THRID_HIGHEST
FROM
(
SELECT WEEK_NO,STORE_ID,ROW_NUMBER() OVER (PARTITION BY WEEK_NO ORDER BY
COUNT(DISTINCT BASKET_ID)) AS RANKING
FROM `Ecommerce.transaction`
GROUP BY 1,2
) A
GROUP BY A.WEEK_NO
ORDER BY A.WEEK_NO ASC;
```

#### Output:

Row	WEEK_NO	HIGHEST	SECOND_HIGHEST	THRID_HIGHEST
1	1	337	286	3313
2	2	3313	352	563
3	3	3313	440	366
4	4	2915	3313	352
5	5	569	678	563
6	6	154	724	678
7	7	337	3313	154
8	8	3036	569	3313
9	9	563	46	2841
10	10	3257	3236	2804

#### Insight's:

- Foot fall varies across all store week wise
- By Checking the pattern of Foot fall, we can better efficiently deploy Human & Physical Resources



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

Query:

```
select A.household_key ,
       min(A.DAY) AS FIRSTENTRY,
       max(A.DAY) AS LASTENTRY,
       count(A.day) AS NOOFENTRIES,
       ROUND(avg(A.totalsales),2) AS AVGSALES,
       ROUND(sum(A.totalsales),2) AS TOTALSALES
from
(SELECT household_key , DAY ,sum(SALES_VALUE) as totalsales from
`Ecommerce.transaction` group by 1,2) A
group by A.household_key
ORDER BY AVGSALES DESC;
```

Output:

Row	household_key	FIRSTENTRY	LASTENTRY	NOOFENTRIES	AVGSALES	TOTALSALES
1	2042	52	683	24	97.47	2339.21
2	973	95	710	72	95.5	6875.89
3	1899	20	705	66	87.72	5789.59
4	948	52	372	2	83.74	167.48
5	1900	111	707	55	76.87	4227.72
6	1574	107	651	24	76.8	1843.3
7	931	94	668	36	68.2	2455.29
8	1864	103	710	130	65.67	8537.28
9	2479	111	706	108	64.39	6954.64
10	1315	60	624	5	63.48	317.39

Insight's:

- This Type Data can be Used for Customer Segmentation for Fine tuned Targeting

## 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:

```
SELECT L.* , R.totalspent
FROM `Ecommerce.demographic` L
inner join
( SELECT household_key , sum(SALES_VALUE) as totalspent
from `Ecommerce.transaction`
group by 1
ORDER BY totalspent desc
limit 10 ) R
on L.household_key = R.household_key
order by R.totalspent desc;
```

### Output:

Row	AGE_DESC	MARITAL_STATUS_CODE	INCOME_DESC	HOMEOWNER_DESC	HH_COMP_DESC	HOUSEHOLD_SIZE_DESC	KID_CATEGORY_DESC	household_key	totalspent
1	45-54	A	125-149K	Homeowner	2 Adults Kids	5+	3+	1609	13804.37999999...
2	45-54	U	175-199K	Homeowner	Single Male	1	None/Unknown	2322	11934.65999999...
3	45-54	A	125-149K	Homeowner	2 Adults Kids	3	1	1453	10720.71999999...
4	35-44	A	35-49K	Homeowner	2 Adults Kids	3	1	1430	10147.20999999...
5	45-54	A	25-34K	Homeowner	2 Adults Kids	5+	3+	718	9577.62999999...
6	35-44	B	Under 15K	Homeowner	Single Female	1	None/Unknown	1653	9519.92999999...
7	35-44	A	150-174K	Homeowner	2 Adults Kids	3	1	400	9481.18999999...

### Insight's:

- Customer Who are **HomeOwners** are among Most Spent in this business
- It seems HomeOwner\_Desc has high correlatios with total amount spent in Customer Demographics Compared to marital Status, income,Household size,No of Kids

### 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)

### Query:

```
WITH TABLE1 AS
(
SELECT L.BASKET_ID,L.PRODUCT_ID, R.SUB_COMMODITY_DESC
FROM `Ecommerce.transaction` L
INNER JOIN `Ecommerce.product` R
ON L.PRODUCT_ID = R.PRODUCT_ID),
TABLE2 AS
(
SELECT L.BASKET_ID,L.SUB_COMMODITY_DESC AS PRODUCTNAME1,R.SUB_COMMODITY_DESC AS
PRODUCTNAME2
FROM TABLE1 L
JOIN TABLE1 R
ON L.BASKET_ID = R.BASKET_ID AND L.SUB_COMMODITY_DESC < R.SUB_COMMODITY_DESC
)
SELECT PRODUCTNAME1,PRODUCTNAME2,COUNT(BASKET_ID) AS FREQUENCY
FROM TABLE2
GROUP BY PRODUCTNAME1,PRODUCTNAME2
ORDER BY FREQUENCY DESC
```

### Output:

Row	PRODUCTNAME1	PRODUCTNAME2	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

### Insight's:

- Market basket Analysis is done with SQL
- Association is High for “FLUID MILK WHITE ONLY” & “YOGURT NOTMULTI-PACKS”

Optional Question 6

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

Query:

```
SELECT R.SUB_COMMODITY_DESC,SUM(L.QUANTITY) AS TOTAL_QUANTITY
FROM `Ecommerce.transaction` L
INNER JOIN `Ecommerce.product` R
ON L.PRODUCT_ID = R.PRODUCT_ID
GROUP BY R.SUB_COMMODITY_DESC
ORDER BY TOTAL_QUANTITY DESC
```

Output:

Row	SUB_COMMODITY_DESC	TOTAL_QUANTITY
1	GASOLINE-REG UNLEADED	128146567
2	FLUID MILK WHITE ONLY	45613
3	YOGURT NOT MULTI-PACKS	32845
4	SOFT DRINKS 12/18&15PK CA...	27527
5	SFT DRNK 2 LITER BTL CARB I...	27216
6	CANDY BARS (SINGLES)(INCL...	23222
7	SHREDDED CHEESE	17843
8	MAINSTREAM WHITE BREAD	17437
9	CONDENSED SOUP	16601
10	BANANAS	15841
11	SS ECONOMY ENTREES/DINN...	15788

Insight's:

- Gasoline is the Most Sold product

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:

```
SELECT household_key,WEEK_NO ,ROUND((customerwiseweeklysales -  
LAG(customerwiseweeklysales,1) OVER (PARTITION BY household_key ORDER BY  
WEEK_NO)),2) AS weeklyRevenuechange  
from  
(SELECT household_key,WEEK_NO,SUM(SALES_VALUE) as customerwiseweeklysales  
from `Ecommerce.transaction`  
group by household_key,WEEK_NO)  
ORDER BY household_key,WEEK_NO
```

Output:

Row	household_key	WEEK_NO	weeklyRevenuechange
1	1	8	null
2	1	10	-28.57
3	1	13	0.02
4	1	14	11.68
5	1	15	-14.73
6	1	16	-1.89
7	1	17	4.89
8	1	19	33.37
9	1	20	-15.58
10	1	22	7.21

Insight's:

- No Monotonic Increase or Decrease observe in Weekly Revenue chag, the change is kind of sinusoidal
- Looking at the Weekly change in Revenue, Promotions can be send in the Dull Sales Week

# Actionable Insights

- Business has Total Revenue of 4,029,338 with Total Order of 233,356
- Business has total Store of 552 with 2500 Customer
- Business is catering 71,540 Type of Products
- Weekly Revenue Increase from 2454 to peak of 57722
- Average Sale Values per Visit Ranges from 0.0 ~ 72
- Business has Good Amount Champion Customer and Also Customer who need Major attention
- Most of the order fall in Large Category where Large Category means order value >10
- Next come Small Category with order value range 0~5
- And then come Medium with order value range between 5~10
- Foot fall varies across all store week wise
- Customer Who are **HomeOwners** are among Most Spent in this business
- It seems HomeOwner\_Desc has high correlatios with total amount spent in Customer Demographics Compared to marital Status, income,Household size,No of Kids
- Association is High for "FLUID MILK WHITE ONLY" & "YOGURT NOTMULTI-PACKS"
- However Gasoline is the Most Sold product
- No Monotonic Increase or Decrease observe in Weekly Revenue change, the change is kind of sinusoidal

# Recommendations

- ❖ Revenue Growth is Good ,So Business needs to be ready for Future Growth by better managing of all operations by stepping up Inventory, Logistics, Human Resources & Infrastructures
- ❖ There is definite need to measure the in-out merit for existence of some of the stores and close them for better resource management as they are generating close to Zero Avg Revenue
- ❖ No. of Orders & Average sales per Customer cannot be only indicators for Customer Segmentation , a collective approach needed to set Marketing Strategies for each customer based on Recency, Frequency, Monetary Analysis
- ❖ By Checking the pattern of Foot fall, we can better efficiently deploy Human & Physical Resources in Required Stores
- ❖ Customer Demographics can be Analyzed to check their Buying Pattern to further Fine tune marketing /upselling /Discount Offer
- ❖ Association Rules can be Found to Design the Place of Products for Ease of Customer & for increasing upselling