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:

```
SELECT
  CASE
    WHEN SALES_VALUE BETWEEN 0 AND 10 THEN 'Small'
    WHEN SALES_VALUE BETWEEN 10 AND 20 THEN 'Medium'
    WHEN SALES_VALUE > 20 THEN 'Large'
  END AS Order_Size,
  COUNT(*) AS Order_Count
FROM ecommerce_store.transactions
GROUP BY Order_Size
ORDER BY Order_Size DESC;
```

Results:

Row	Order_Size	Order_Count
1	Small	1259081
2	Medium	26869
3	Large	12536

Insights:

- Almost 96% of the orders has sales value below 10 USD.

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

Query:

```
SELECT
  CASE
    WHEN SALES_VALUE BETWEEN 0 AND 5 THEN 'Small'
    WHEN SALES_VALUE BETWEEN 5 AND 10 THEN 'Medium'
    WHEN SALES_VALUE > 10 THEN 'Large'
  END AS Order_Size,
  COUNT(*) AS Order_Count
FROM ecommerce_store.transactions
GROUP BY Order_Size
ORDER BY Order_Size DESC;
```

Results:

Row	Order_Size	Order_Count
1	Small	1145982
2	Medium	113099
3	Large	39405

Insights:

- Almost 88% of the orders has sales value below 5 USD.

Question 3: Find top 3 stores with highest foot traffic for each week.

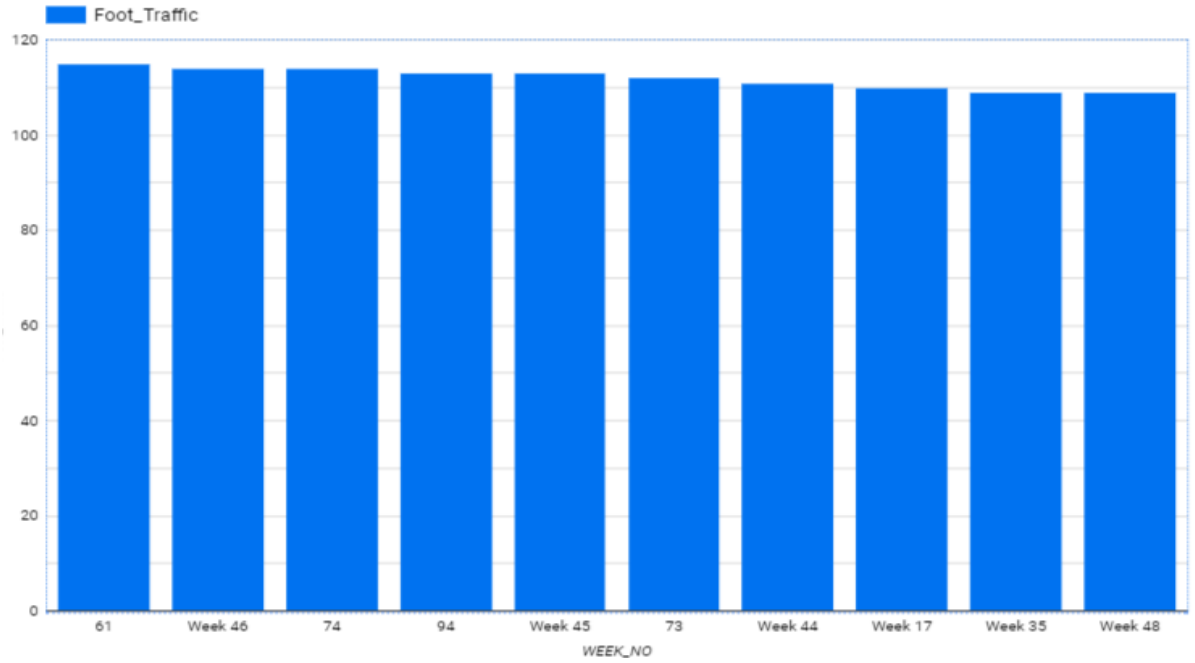
Query:

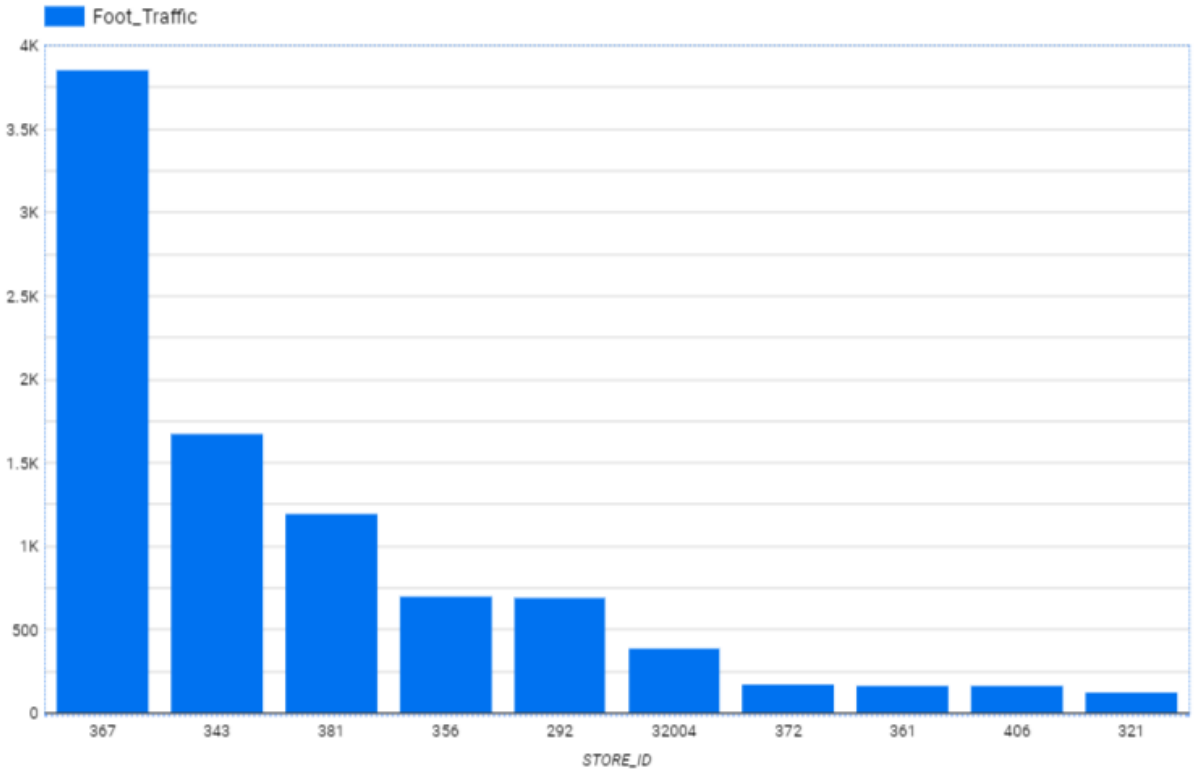
```
WITH StoreFootTraffic AS (
  SELECT WEEK_NO, STORE_ID, COUNT(DISTINCT household_key) AS Foot_Traffic
  FROM ecommerce_store.transactions
  GROUP BY STORE_ID, WEEK_NO
)
SELECT WEEK_NO,STORE_ID, Foot_Traffic
FROM ( SELECT STORE_ID, WEEK_NO, Foot_Traffic,
        ROW_NUMBER() OVER (PARTITION BY WEEK_NO ORDER BY Foot_Traffic DESC) AS rank
        FROM  StoreFootTraffic )
WHERE rank <= 3
ORDER BY WEEK_NO, Foot_Traffic DESC, STORE_ID;
```

Results:

Row	WEEK_NO	STORE_ID	Foot_Traffic
1	1	32004	5
2	1	324	3
3	1	367	3
4	2	32004	7
5	2	313	6
6	2	367	5
7	3	367	10
8	3	32004	9
9	3	356	8
10	4	367	17

Charts:





Insights:

- There's a general upward trend in foot traffic over the weeks.
- Certain weeks (e.g., weeks 61, 46, 74, 94, 45, 73) experienced significantly higher foot traffic.
- Store 367 consistently attracts the highest foot traffic.

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 average money.

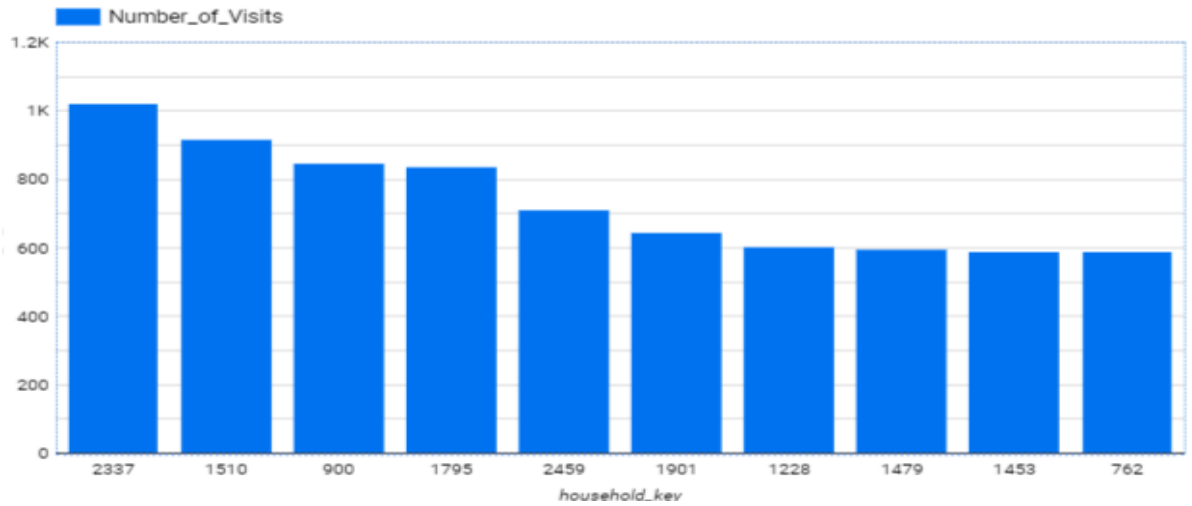
Query:

```
SELECT
  household_key,
  MIN(DAY) AS First_Visit,
  MAX(DAY) AS Last_Visit,
  COUNT(DISTINCT BASKET_ID) AS Number_of_Visits,
  SUM(SALES_VALUE) / COUNT(DISTINCT BASKET_ID) AS Avg_Money_Spent,
  SUM(SALES_VALUE) AS Total_Money_Spent
FROM
  ecommerce_store.transactions
GROUP BY
  household_key
ORDER BY
  Avg_Money_Spent DESC;
```

Results:

Row	household_key	First_Visit	Last_Visit	Number_of_Visits	Avg_Money_Spent	Total_Money_Spent
1	2042	52	683	26	89.96961538461...	2339.209999999...
2	973	95	710	80	85.94862499999...	6875.8899999999
3	1899	20	705	69	83.90710144927...	5789.589999999...
4	1900	111	707	55	76.86763636363...	4227.719999999...
5	1574	107	651	27	68.27037037037...	1843.300000000...
6	1315	60	624	5	63.47800000000...	317.3900000000...
7	2479	111	706	111	62.65441441441...	6954.639999999...
8	931	94	668	40	61.38225000000...	2455.290000000...
9	1344	87	691	26	60.39884615384...	1570.370000000...
10	248	29	704	53	58.31867924528...	3090.889999999...

Chart:



Insights:

- Houses 2337, 1510, 900, 1795 and 2459 are the top 5 in most number of visits.
- Houses 2042, 973, 1899, 1900 and 1574 are the top 5 houses in average money spent.

Question 5: Do a single customer analysis selecting most spending customer for whom we have demographic information

Query:

```
WITH CustomerSpend AS (
  SELECT household_key, SUM(SALES_VALUE) AS Total_Spent
  FROM ecommerce_store.transactions
  GROUP BY household_key
  ORDER BY Total_Spent DESC
)

SELECT d.*, c.Total_Spent
FROM CustomerSpend c JOIN ecommerce_store.demographics d
ON c.household_key = d.household_key
ORDER BY Total_Spent DESC
LIMIT 1;
```

Result:

Row	AGE_DESC	MARITAL_STATUS	INCOME_DESC	HOMEOWNER_DESC	HH_COMP_DESC	HOUSEHOLD_	KID_CATE	household	Total_Spent
1	45-54	A	125-149K	Homeowner	2 Adults Kids	5+	3+	1609	13804.37999999...

Insights:

- Customers from house 1609 are the highest spending customer with the total spent over 13K USD having household income of 125-149K USD.

Question 6: Find products which are most frequently bought together and the count of each combination bought together.

Query:

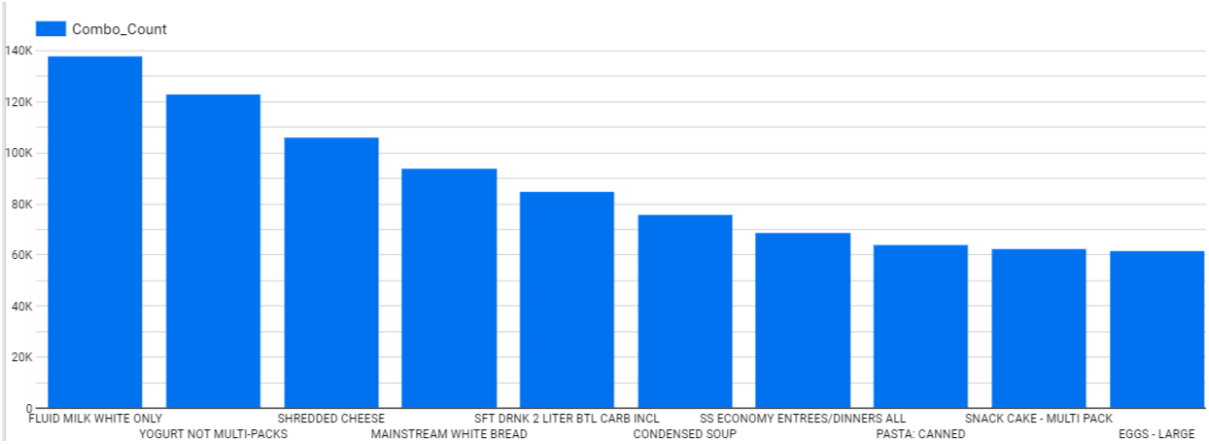
```
WITH BasketCombinations AS (
  SELECT t1.BASKET_ID,t1.PRODUCT_ID AS Product_A, t2.PRODUCT_ID AS Product_B
  FROM ecommerce_store.transactions t1 JOIN ecommerce_store.transactions t2
    ON t1.BASKET_ID = t2.BASKET_ID  AND t1.PRODUCT_ID < t2.PRODUCT_ID
)

SELECT p1.SUB_COMMODITY_DESC AS Product_A, p2.SUB_COMMODITY_DESC AS Product_B, COUNT(*) AS Combo_Count
FROM BasketCombinations bc
  JOIN ecommerce_store.products p1 ON bc.Product_A = p1.PRODUCT_ID
  JOIN ecommerce_store.products p2 ON bc.Product_B = p2.PRODUCT_ID
GROUP BY Product_A, Product_B
ORDER BY Combo_Count DESC;
```

Results:

Row	Product_A	Product_B	Combo_Count
1	YOGURT NOT MULTI-PACKS	YOGURT NOT MULTI-PACKS	15947
2	BABY FOOD - BEGINNER	BABY FOOD - BEGINNER	10080
3	SS ECONOMY ENTREES/DINN...	SS ECONOMY ENTREES/DINN...	6633
4	SOFT DRINK POWDER POUCHES	SOFT DRINK POWDER POUCHES	6375
5	FRZN SS PREMIUM ENTREES/...	FRZN SS PREMIUM ENTREES/...	6340
6	SFT DRNK 2 LITER BTL CARB I...	SFT DRNK 2 LITER BTL CARB I...	5459
7	SOFT DRINKS 12/18&15PK CA...	SOFT DRINKS 12/18&15PK CA...	5173
8	CANDY BARS (SINGLES)(INCL...	CANDY BARS (SINGLES)(INCL...	4194
9	BABY FOOD JUNIOR ALL BRAN...	BABY FOOD JUNIOR ALL BRAN...	3751
10	FLUID MILK WHITE ONLY	SOFT DRINKS 12/18&15PK CA...	3580

Chart:



Insights:

- Some combinations like "YOGURT NOT MULTI-PACKS", "BABY FOOD - BEGINNER" and "BABY FOOD JUNIOR ALL BRANDS", suggest complementary products within the same category.
- The products that appear in multiple combinations, like "FLUID MILK WHITE ONLY", "YOGURT NOT MULTI-PACKS" and "SHREDDED CHEESE" are likely more popular overall.

Question 7: Find the weekly change in Revenue Per Account (RPA)

Query:

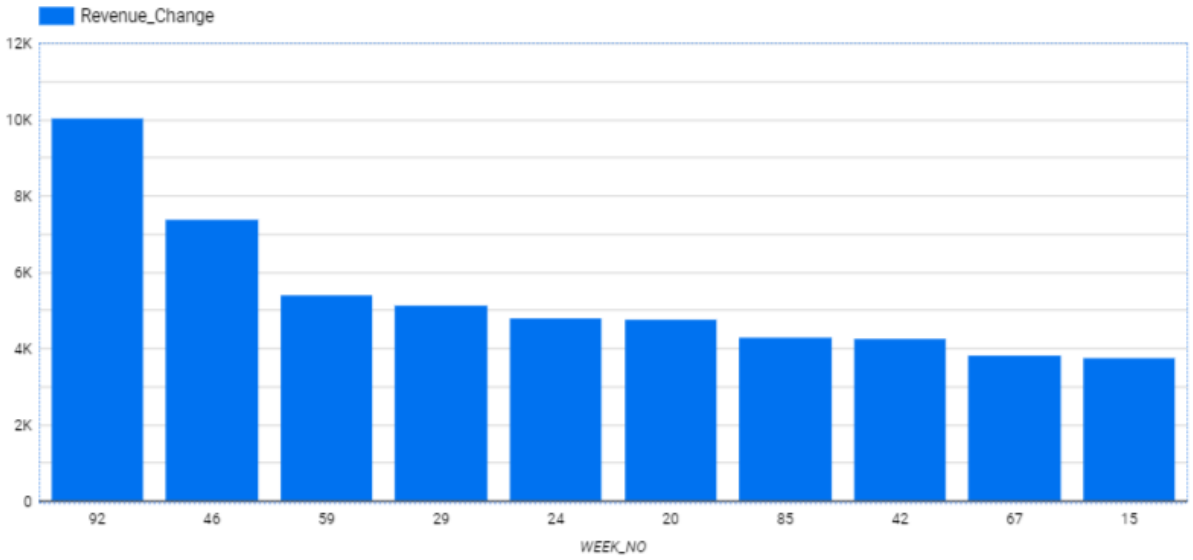
```
WITH WeeklyRevenue AS (
  SELECT household_key, WEEK_NO, SUM(SALES_VALUE) AS Weekly_Revenue
  FROM ecommerce_store.transactions
  GROUP BY household_key, WEEK_NO
)

SELECT household_key, WEEK_NO, Weekly_Revenue,
  LAG(Weekly_Revenue) OVER (PARTITION BY household_key ORDER BY WEEK_NO) AS Previous_Week_Revenue,
  (Weekly_Revenue - LAG(Weekly_Revenue) OVER (PARTITION BY household_key ORDER BY WEEK_NO)) AS
Revenue_Change
FROM WeeklyRevenue
ORDER BY household_key;
```

Results:

Row	household_key	WEEK_NO	Weekly_Revenue	Previous_Week_Reve	Revenue_Change
1	1	8	42.58	null	null
2	1	10	14.01	42.58	-28.57
3	1	13	14.03000000000...	14.01	0.020000000000...
4	1	14	25.71	14.03000000000...	11.68
5	1	15	10.98	25.71	-14.73
6	1	16	9.09	10.98	-1.89000000000...
7	1	17	13.98	9.09	4.890000000000...
8	1	19	47.35000000000...	13.98	33.37000000000...
9	1	20	31.77	47.35000000000...	-15.5800000000...
10	1	22	38.98000000000...	31.77	7.210000000000...

Chart:



Insights:

- Week 92 receives high revenue increase from the previous week with revenue change of 10K USD followed by 46 with 7.6K USD.

Question 8: Identify the top 5 most purchased products by total sales value.

Query:

```
SELECT  p.SUB_COMMODITY_DESC AS Product, SUM(t.SALES_VALUE) AS Total_Sales
FROM ecommerce_store.transactions t JOIN ecommerce_store.products p
  ON t.PRODUCT_ID = p.PRODUCT_ID
GROUP BY  p.SUB_COMMODITY_DESC
ORDER BY  Total_Sales DESC
LIMIT 5;
```

Results:

Row	Product	Total_Sales
1	GASOLINE-REG UNLEADED	315997.0900000…
2	FLUID MILK WHITE ONLY	80754.44000000…
3	SOFT DRINKS 12/18&15PK CA…	79214.43999999…
4	BEERALEMALT LIQUORS	75036.18000000…
5	CIGARETTES	48179.15000000…

Insights:

- GASOLINE-REG UNLEADED, FLUID MILK WHITE ONLY, SOFT DRINKS 12/18&15PK CAN CAR, BEERALEMALT LIQUORS and CIGARETTES are the top 5 products in total sales.

Question 9: Identify the top 5 stores with the highest revenue.

Query:

```
SELECT STORE_ID, SUM(SALES_VALUE) AS Total_Sales
FROM ecommerce_store.transactions
GROUP BY STORE_ID
ORDER BY Total_Sales DESC
LIMIT 5;
```

Results:

Row	STORE_ID	Total_Sales
1	367	134105.4600000…
2	406	108814.6600000…
3	361	72493.71999999…
4	429	70752.60999999…
5	343	70265.59999999…

Insights:

- 365 is the highest revenue generated store with 134K USD, 406 is the second highest with 100K USD, followed by 361, 429 and 343 having around 70K USD total sales.

Question 10: Find the most popular shopping times of the day.

Query:

```
SELECT
  CASE
    WHEN TRANS_TIME BETWEEN 0 AND 599 THEN 'Midnight to 6 AM'
    WHEN TRANS_TIME BETWEEN 600 AND 1199 THEN '6 AM to Noon'
    WHEN TRANS_TIME BETWEEN 1200 AND 1799 THEN 'Noon to 6 PM'
    WHEN TRANS_TIME BETWEEN 1800 AND 2359 THEN '6 PM to Midnight'
  END AS Time_Period,
  COUNT(*) AS Transaction_Count
FROM ecommerce_store.transactions
GROUP BY Time_Period
ORDER BY Transaction_Count DESC;
```

Results:

Row	Time_Period	Transaction_Count
1	Noon to 6 PM	661324
2	6 PM to Midnight	418605
3	6 AM to Noon	197897
4	Midnight to 6 AM	20660

Insights:

- The majority of people shop between 12 PM and 6 PM, with the next busiest period being from 6 PM to midnight.

Insights

- Almost 96% of the orders has sales value below 10 USD.
- Almost 88% of the orders has sales value below 5 USD.
- There's a general upward trend in foot traffic over the weeks.
- Certain weeks (e.g., weeks 61, 46, 74, 94, 45, 73) experienced significantly higher foot traffic.
- Store 367 consistently attracts the highest foot traffic.
- Houses 2337, 1510, 900, 1795 and 2459 are the top 5 in most number of visits.
- Houses 2042, 973, 1899, 1900 and 1574 are the top 5 houses in average money spent.
- Customers from house 1609 are the highest spending customer with the total spent over 13K USD having household income of 125-149K USD.
- Some combinations like "YOGURT NOT MULTI-PACKS", "BABY FOOD - BEGINNER" and "BABY FOOD JUNIOR ALL BRANDS", suggest complementary products within the same category.
- The products that appear in multiple combinations, like "FLUID MILK WHITE ONLY", "YOGURT NOT MULTI-PACKS" and "SHREDDED CHEESE" are likely more popular overall.

- Week 92 receives high revenue increase from the previous week with revenue change of 10K USD followed by 46 with 7.6K USD.
- GASOLINE-REG UNLEADED, FLUID MILK WHITE ONLY, SOFT DRINKS 12/18&15PK CAN CAR, BEERALEMALT LIQUORS and CIGARETTES are the top 5 products in total sales.
- 365 is the highest revenue generated store with 134K USD, 406 is the second highest with 100K USD, followed by 361, 429 and 343 having around 70K USD total sales.
- The majority of people shop between 12 PM and 6 PM, with the next busiest period being from 6 PM to midnight.

Recommendations

- **Increase Order Value:** Introduce value bundles or combo offers to encourage customers to spend above $10. Offer discounts on purchases exceeding $5 to boost sales volume.
- **Leverage High-Traffic Weeks:** Capitalize on weeks with high foot traffic by launching targeted promotions, special events, or flash sales to maximize revenue during these peak periods.
- **Boost Low-Traffic Stores:** Increase local marketing efforts and run special promotions or exclusive deals at low-traffic stores. Consider in-store experiences or partnerships to attract customers.
- **Reward High-Spending Households**: Develop loyalty programs for high-spending households (like 1609) and offer personalized rewards to encourage repeat purchases. Offer exclusive deals on premium products for these segments.
- **Promote Complementary Products:** Highlight complementary product combinations (e.g., baby food, yogurt) through in-store displays or online recommendations to increase basket size.
- **Capitalize on Popular Products:** Prioritize stocking and promoting top-selling products (like gasoline, milk, and soft drinks) with dedicated shelf space, special discounts, or combo offers to maintain high sales.
- **Target Prime Shopping Hours:** Run timed promotions or flash sales between 12 PM and midnight, aligning offers with peak shopping times to increase store visits and transactions.
- **Maximize Revenue at High-Traffic Stores:** Focus on improving the customer experience and product offerings at stores with consistently high traffic (like Store 367) to further boost sales and customer retention.
- **Implement Customer Feedback Loops:** Use surveys or feedback forms to identify areas for improvement and adjust offerings based on customer preferences, which can lead to increased satisfaction and sales.
- **Utilize Data Analytics for Personalized Marketing:** Leverage customer purchase data to create personalized marketing campaigns. Tailor promotions and recommendations based on individual shopping habits, increasing the likelihood of conversion and fostering customer loyalty.