Data Mart Case Study

By Rohit Patil

Introduction

- **Venture :** Data Mart

- **Sustainability Initiative:** In June 2020, Data Mart adopted sustainable packaging for all products, from farm to customer.

- **Goal:** Analyze the impact of this change on sales performance.

- **Scope:** Quantify the effect on overall sales and separate business areas.

Schema

Column name	Data type
week_date	date
region	varchar(20)
platform	varchar(20)
segment	varchar(10)
customer	varchar(20)
transactions	int
sales	int

Data Cleaning

```sql
--Data Cleaning Query
Select week_date,
datepart(week, week_date) as WeeK_number,
month(week_date) as month,
year(week_date) as year,
region,platform,isnull(segment,'Unknown') as segment,
Case
when right(segment,1) = '1' then 'Young Adults'
when right(segment,1) = '2' then 'Middle Aged'
when right(segment,1) in ( '3' , '4' ) then 'Retirees'
else 'Unknown'
end as Age_band,
Case
when left(segment,1) = 'C' then 'Couples'
when left(segment,1) = 'F' then 'Families'
else 'Unknown'
end as Demographic,
customer_type,
transactions,
sales,
round((sales/transactions),2) as Avg_transaction
into clean_weekly_sales
from weekly_sales
```

Data Cleaning Criteria mentioned in PDF

Questions

- **Question 1: How many total transactions were there for each year in the dataset?**

```
56  SELECT Sum(transactions) AS yearly_transactions,
57         year
58  FROM   clean_weekly_sales
59  GROUP  BY year
```

Question 2: What are the total sales for each region for each month?

```sql
SELECT month,
       region,
       Sum(Cast(sales AS BIGINT)) AS monthly_sales
FROM   clean_weekly_sales
GROUP  BY month,
          region
ORDER  BY month,
          region
```

Question 3: What is the total count of transactions for each platform

```sql
SELECT platform,
       Sum(transactions) count_of_transactions
FROM   clean_weekly_sales
GROUP  BY platform
```

Question 4: Which week numbers are missing from the dataset?

```
Insert into Seq52  (week_number) values (1), (2), (3), (4), (5), (6), (7), (8), (9), (10),
                                        (11), (12), (13), (14), (15), (16), (17), (18), (19), (20),
                                        (21), (22), (23), (24), (25), (26), (27), (28), (29), (30),
                                        (31), (32), (33), (34), (35), (36), (37), (38), (39), (40),
                                        (41), (42), (43), (44), (45), (46), (47), (48), (49), (50),
                                        (51), (52);


SELECT week_number
FROM   seq52
WHERE  week_number NOT IN (SELECT DISTINCT week_number
                           FROM   clean_weekly_sales)
```

Question 5: Which age_band and demographic values contribute the most to Retail sales?

```sql
SELECT age_band,
       demographic,
       Sum(Cast(sales AS BIGINT)) AS Max_sales
FROM   clean_weekly_sales
WHERE  platform = 'Retail'
GROUP  BY age_band,
          demographic
ORDER  BY max_sales DESC
```

Question 6: What is the percentage of sales for Retail vs Shopify for each month?

```sql
WITH cte_monthly_platform_sales AS (
  SELECT
    month,
    year,
    platform,
    SUM(CAST(sales AS BIGINT)) AS monthly_sales
  FROM clean_weekly_sales
  GROUP BY month, year, platform
)
SELECT
  month,
  year,
  ROUND(
    100.0 * MAX(CASE WHEN platform = 'Retail' THEN monthly_sales ELSE NULL END) /
      CAST(SUM(monthly_sales) AS DECIMAL(18, 2)),
    2
  ) AS retail_percentage,
  ROUND(
    100.0 * MAX(CASE WHEN platform = 'Shopify' THEN monthly_sales ELSE NULL END) /
      CAST(SUM(monthly_sales) AS DECIMAL(18, 2)),2
  ) AS shopify_percentage
FROM cte_monthly_platform_sales
GROUP BY month, year
ORDER BY year, month;
```

Question 7: What is the percentage of sales by demographic for each year in the dataset?

```sql
SELECT
  year,
  demographic,
  SUM(cast(sales as bigint)) AS yearly_sales,
  ROUND(
    (
      100.0 * CAST(SUM(cast(sales as bigint)) AS DECIMAL(18, 2)) /
      CAST(SUM(SUM(cast(sales as bigint))) OVER (PARTITION BY demographic) AS DECIMAL(18, 2))
    ),
    2
  ) AS percentage
FROM clean_weekly_sales
GROUP BY
  year,
  demographic
ORDER BY
  year,
  demographic;
```

Thank You