CASE STUDY VASANTHABALAN M

Introduction

Data Mart is Danny's latest venture and after running international operations for his online supermarket that specialises in fresh produce - Danny is asking for your support to analyse his sales performance.

In June 2020 - large scale supply changes were made at Data Mart. All Data Mart products now use sustainable packaging methods in every single step from the farm all the way to the customer.

Danny needs your help to quantify the impact of this change on the sales performance for Data Mart and it's separate business areas.

The key business question he wants you to help him answer are the following:

- What was the quantifiable impact of the changes introduced in June 2020?
- Which platform, region, segment and customer types were the most impacted by this change?
- What can we do about future introduction of similar sustainability updates to the business to minimise impact on sales?

The basic schema would consist of 3 entities

data_mart.weekly_sales

Data Mart has international operations using a multi-region strategy

Data Mart has both, a retail and online platform in the form of a Shopify store front to serve their customers

Customer segment and customer_type data relates to personal age and demographics information that is shared with Data Mart

transactions is the count of unique purchases made through Data Mart and sales is the actual dollar amount of purchases

Each record in the dataset is related to a specific aggregated slice of the underlying sales data rolled up into a week_date value which represents the start of the sales week.

ANSWER THE FOLLOWING QUERIES

1. In a single query, perform the following operations and generate a new table in the data_mart schema named clean_weekly_sales:
   a. Convert the week_date to a DATE format
   b. Add a week_number as the second column for each week_date value, for example any value from the 1st of January to 7th of January will be 1, 8th to 14th will be 2 etc

c. Add a month_number with the calendar month for each week_date value as the 3rd column
d. Add a calendar_year column as the 4th column containing either 2018, 2019 or 2020 values
e. Add a new column called age_band after the original segment column using the following mapping on the number inside the segment value

segment	age_band
1	Young Adults
2	Middle Aged
3 or 4	Retirees

f. Add a new demographic column using the following mapping for the first letter in the segment values:

segment	demographic
C	Couples
F	Families

g. Generate a new avg_transaction column as the sales value divided by transactions rounded to 2 decimal places for each record

```
SELECT
  CONVERT(date, week_date, 3) AS week_date,
  DATEPART(week, CONVERT(date, week_date, 3)) AS week_number,
  DATEPART(month, CONVERT(date, week_date, 3)) AS month_number,
  DATEPART(year, CONVERT(date, week_date, 3)) AS calendar_year,
  region,
  platforms,
  segment,
  customer_type,
  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,
  transactions,
  CAST(sales AS bigint) AS sales,
  ROUND(CAST(sales AS FLOAT)/transactions, 2) AS avg_transaction
INTO cleaned_weekly_sales
FROM weakly_sales;
```

Results | Messages

	week_date	week_number	month_number	calendar_year	region	platforms	segment	customer_type	age_band	demographic	transactions	sales	avg_transaction
1	2020-01-12	3	1	2020	India	retail	C1	new	Young Adults	Couples	1321	98298	74.41
2	2020-03-21	12	3	2020	Austrlia	shopify	F3	Guest	Retirees	Families	2726	73322	26.9
3	2020-05-12	20	5	2020	Antartica	retail	F1	Guest	Young Adults	Families	38923	8937	0.23
4	2020-01-12	3	1	2020	India	retail	C2	New	Middle Aged	Couples	783673	8937832	11.41
5	2020-08-30	36	8	2020	Asia	shopify	F4	Guest	Retirees	Families	88421	783733	8.86
6	2020-01-12	3	1	2020	India	retail	C4	new	Retirees	Couples	783723	2783	0
7	2020-07-29	31	7	2020	Europe	shopify	C3	Guest	Retirees	Couples	8323	89723	10.78

2. How many total transactions were there for each year in the dataset?

```
select calendar_year, sum(transactions)as "total" from cleaned_weekly_sales group
by calendar_year order by calendar_year;
```

Results | Messages

	calendar_year	total
1	2020	1707110

3. What is the total sales for each region for each month?

```
SELECT
  region,
  month_number,
  SUM(sales) AS total_sales
FROM cleaned_weekly_sales
GROUP BY region, month_number
ORDER BY region, month_number;
```

Results | Messages

	region	month_number	total_sales
1	Antartica	5	8937
2	Asia	8	783733
3	Austrlia	3	73322
4	Europe	7	89723
5	India	1	9038913

4. What is the total count of transactions for each platform

```
SELECT
  platforms,
  SUM(transactions) AS "total"
FROM cleaned_weekly_sales
GROUP BY platforms;
```

Results | Messages

	platforms	total
1	retail	1607640
2	shopify	99470

5. What is the total sales for the 4 weeks before and after 2020-06-15?

```
DECLARE @weekNum int = (
  SELECT DISTINCT week_number
  FROM cleaned_weekly_sales
  WHERE week_date = '2020-06-15')


 SELECT
    SUM(CASE WHEN week_number BETWEEN @weekNum-4 AND @weekNum-1 THEN sales END) AS
before_changes,
    SUM(CASE WHEN week_number BETWEEN @weekNum AND @weekNum+3 THEN sales END) AS
after_changes
  FROM cleaned_weekly_sales
  WHERE calendar_year = 2020
```

Results | Messages

	before_changes	after_changes
1	NULL	NULL

6. Which areas of the business have the 5 highest negative impact in sales metrics performance in 2020 for the 12 weeks in the 2nd quarter?

```
SELECT top 5 month_number,sum( sales) "Sales Amount"

  FROM cleaned_weekly_sales
  WHERE calendar_year = 2020 and month_number>6
  group by month_number
```

Results | Messages

	month_number	Sales Amount
1	7	89723
2	8	783733

7. Which “age_band” and “demographic” values contribute the most to Retail sales?

```
DECLARE @retailSales bigint = (
SELECT SUM(sales)
FROM cleaned_weekly_sales
WHERE platforms = 'Retail')

SELECT
age_band,
demographic,
SUM(sales) AS sales,
CAST(100.0 * SUM(sales)/@retailSales AS decimal(5, 2)) AS contribution
FROM cleaned_weekly_sales
WHERE platforms = 'Retail'
GROUP BY age_band, demographic
ORDER BY contribution DESC;
```

Results | Messages

	age_band	demographic	sales	contribution
1	Middle Aged	Couples	8937832	98.78
2	Young Adults	Couples	98298	1.09
3	Young Adults	Families	8937	0.10
4	Retirees	Couples	2783	0.03