CREATE TABLE weekly\_sales (

"week\_date" NVARCHAR(10),

"region" NVARCHAR(20),

"platform" NVARCHAR(20),

"segment" NVARCHAR(4),

"customer\_type" NVARCHAR(20),

"transactions" INTEGER,

"sales" INTEGER

)

select \* from weekly\_sales;

INSERT INTO weekly\_sales

(week\_date, region, platform, segment, customer\_type, transactions, sales)

VALUES

('31/8/20', 'ASIA', 'Retail', 'F1', 'New', '31574', '996575'),

('25/12/20', 'USA', 'Retail', 'null', 'Guest', '529151', '16509610'),

('01/4/20', 'INDIA', 'Retail', 'C1', 'New', '4517', '141942'),

('31/10/20', 'AFRICA', 'Retail', 'C2', 'New', '58046', '1758388'),

('16/3/20', 'CANADA', 'Shopify', 'F2', 'Existing', '1336', '243878'),

('07/9/20', 'AFRICA', 'Shopify', 'F3', 'Existing', '2514', '519502'),

('29/11/20', 'ASIA', 'Shopify', 'F1', 'Existing', '2158', '371417'),

('31/8/20', 'AFRICA', 'Shopify', 'F2', 'New', '318', '49557'),

('09/2/20', 'AFRICA', 'Retail', 'C3', 'New', '111032', '3888162'),

('13/8/20', 'USA', 'Shopify', 'F1', 'Existing', '1398', '260773');

QUESTIONS:

**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,

platform,

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,

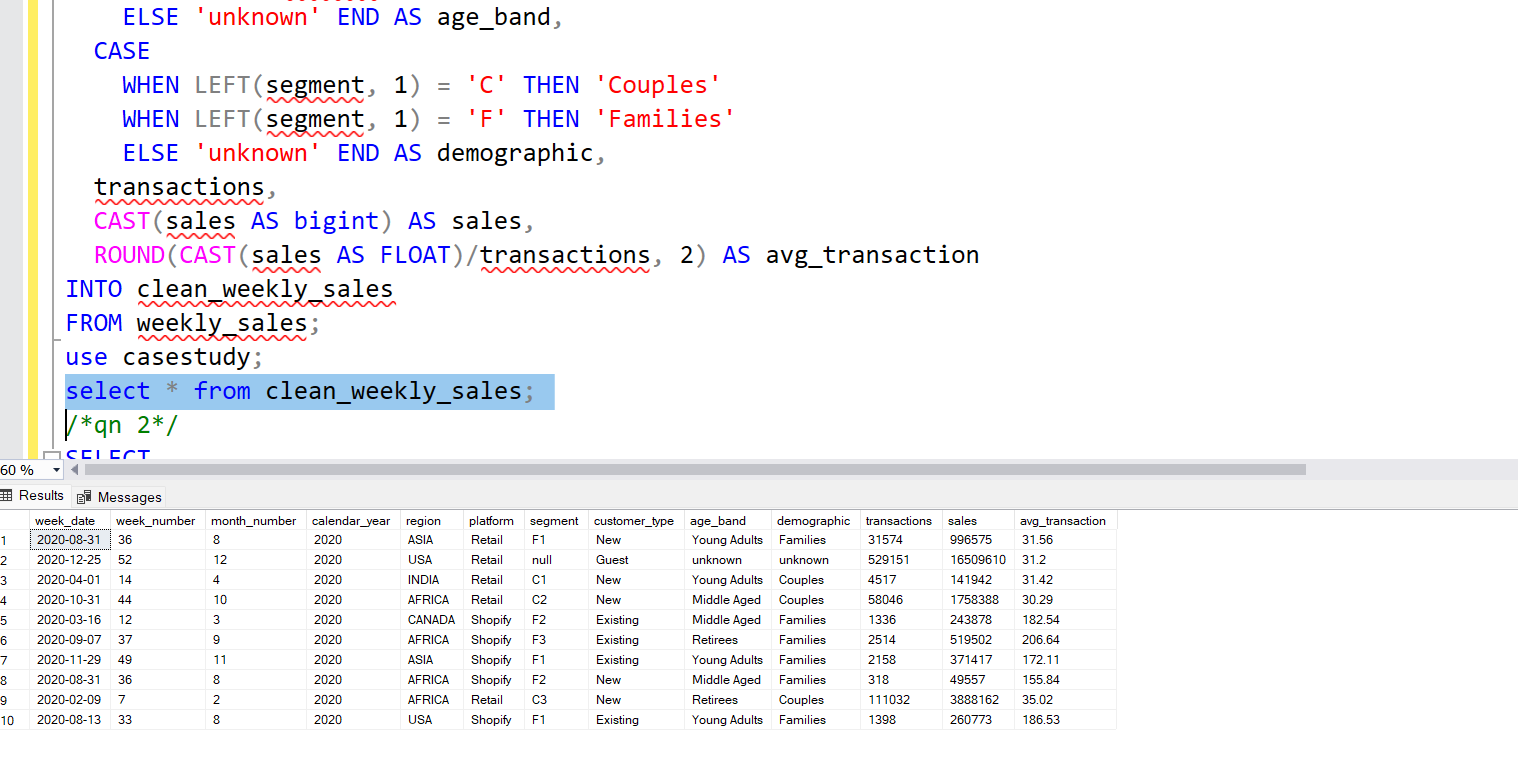
ROUND(CAST(sales AS FLOAT)/transactions, 2) AS avg\_transaction

INTO clean\_weekly\_sales

FROM weekly\_sales;

use casestudy;

select \* from clean\_weekly\_sales;

****

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

SELECT

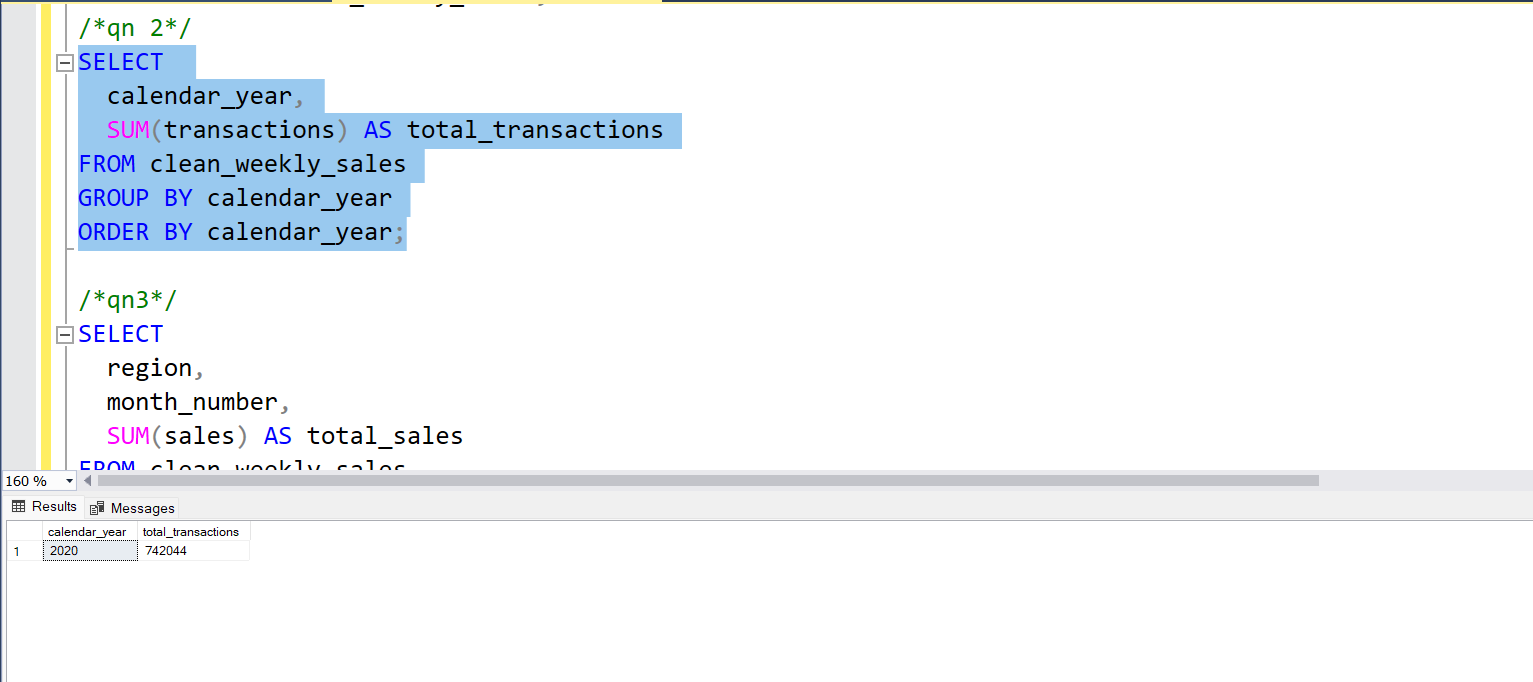
calendar\_year,

SUM(transactions) AS total\_transactions

FROM clean\_weekly\_sales

GROUP BY calendar\_year

ORDER BY calendar\_year;

****

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

SELECT

region,

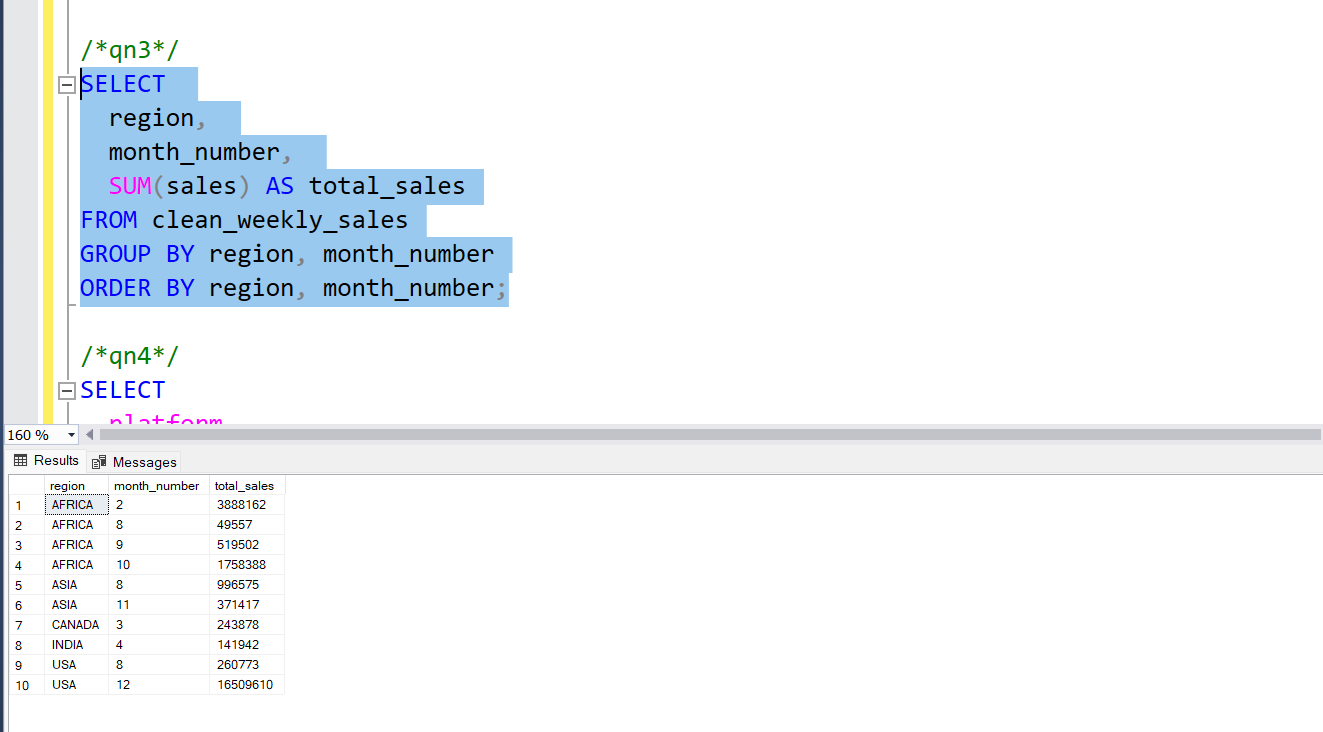
month\_number,

SUM(sales) AS total\_sales

FROM clean\_weekly\_sales

GROUP BY region, month\_number

ORDER BY region, month\_number;

****

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

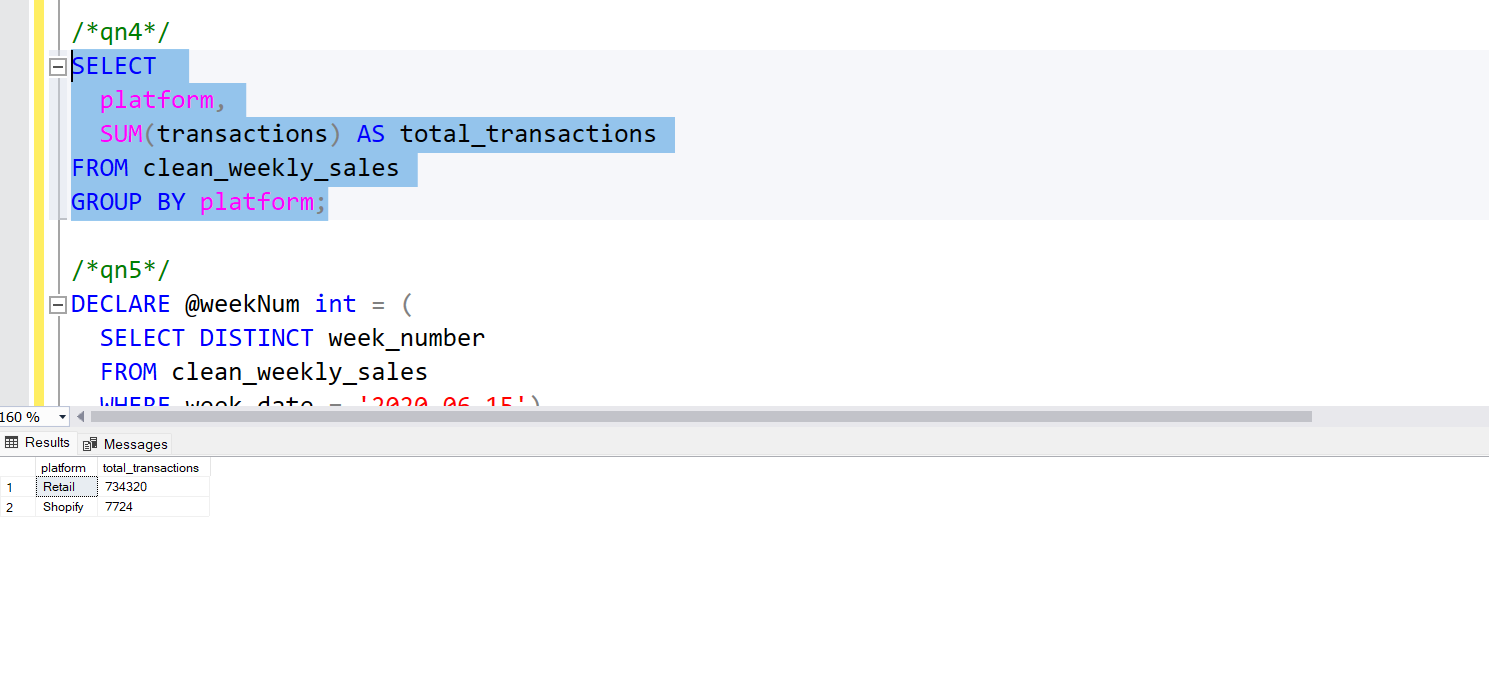
SELECT

platform,

SUM(transactions) AS total\_transactions

FROM clean\_weekly\_sales

GROUP BY platform;

****

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

DECLARE @weekNum int = (

SELECT DISTINCT week\_number

FROM clean\_weekly\_sales

WHERE week\_date = '2020-06-15')

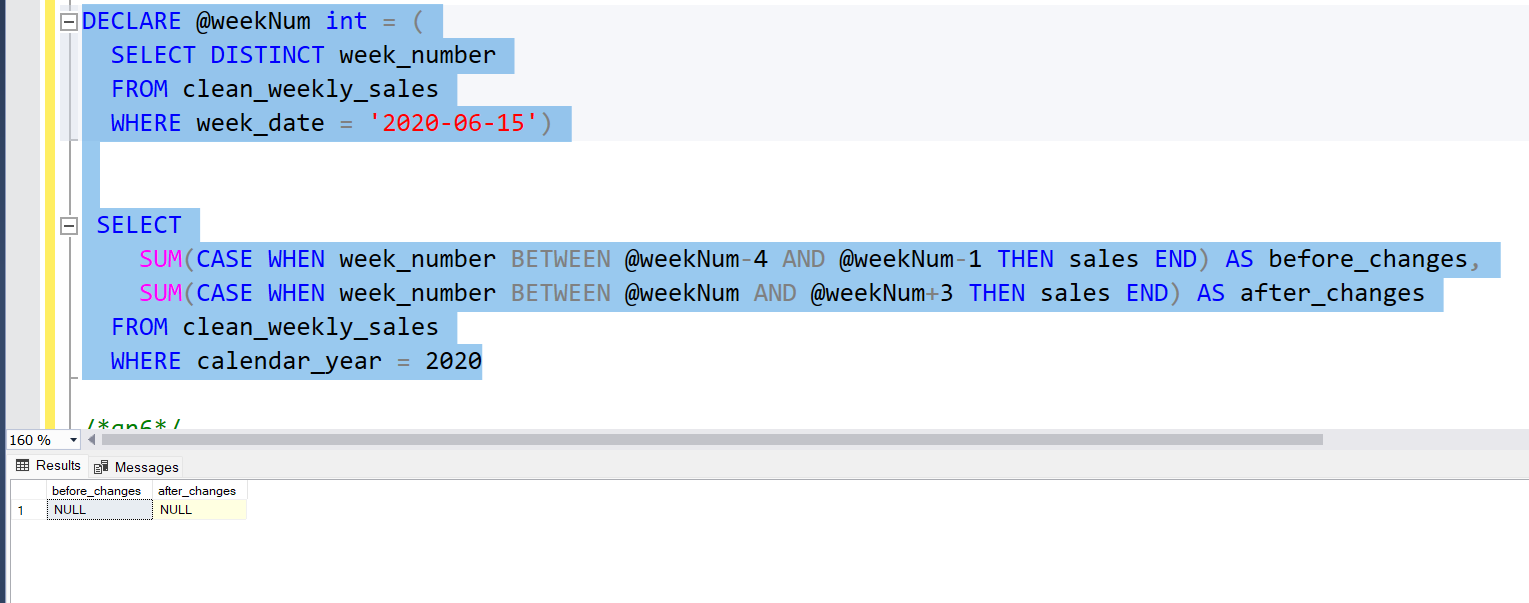
SELECT

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 clean\_weekly\_sales

WHERE calendar\_year = 2020

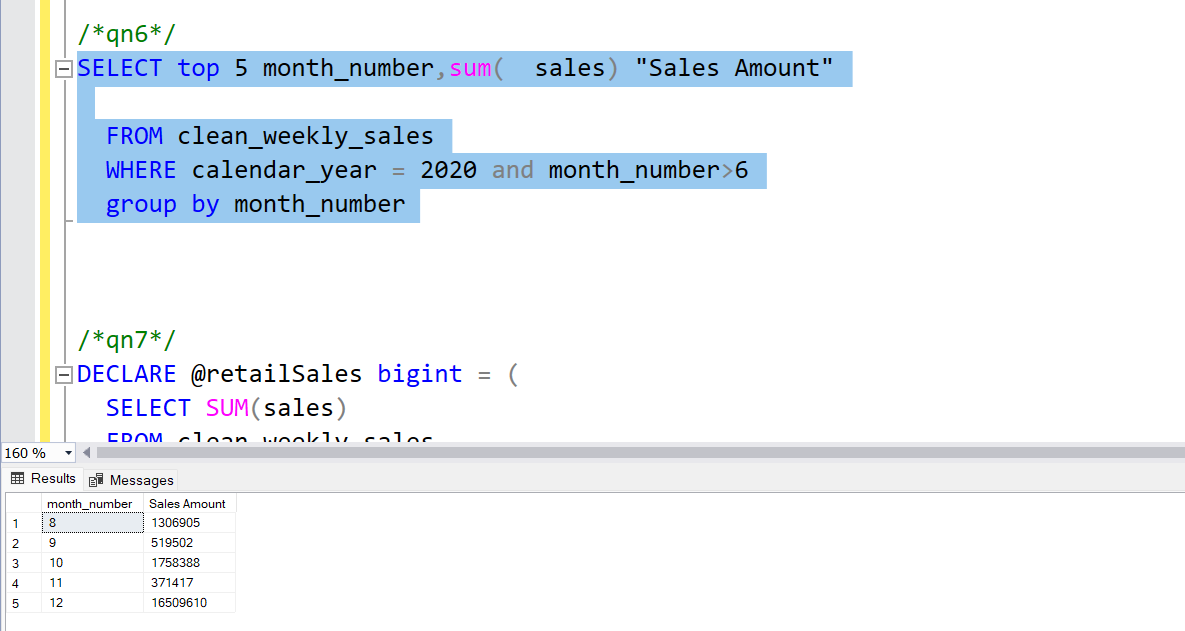


**6. Which areas of the business have the 5 highest negative impact in sales metrics performance in 2020 for the 12 weeks in the 2 nd quarter?**

SELECT top 5 month\_number,sum( sales) "Sales Amount"

FROM clean\_weekly\_sales

WHERE calendar\_year = 2020 and month\_number>6

group by month\_number

**7. Which “age\_band” and “demographic” values contribute the most to Retail sales?**

DECLARE @retailSales bigint = (

SELECT SUM(sales)

FROM clean\_weekly\_sales

WHERE platform = 'Retail')

SELECT

age\_band,

demographic,

SUM(sales) AS sales,

CAST(100.0 \* SUM(sales)/@retailSales AS decimal(5, 2)) AS contribution

FROM clean\_weekly\_sales

WHERE platform = 'Retail'

GROUP BY age\_band, demographic

ORDER BY contribution DESC;

