Martlytics (DataMart Exploration)

INTRODUCTION:

Data Dart is my latest venture, and I'm seeking your support in analyzing its sales and performance. In June 2020, **Data Mart** implemented significant supply chain changes, transitioning all its products to sustainable packaging—covering every step from farm to customer delivery.

I'd like your help in quantifying the impact of this sustainability initiative on overall sales performance, both for Data Mart as a whole and across its individual business segments.

SCHEMA USED: WEEKLY_SALES TABLE

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

ANALYSIS QUESTIONS

A. Data Cleaning Steps

In a single query, perform the following operations and generate a new table in the data_mart schema named clean_weekly_sales:

1. 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.
2. Add a month_number with the calendar month for each week_date value as the 3rd column
3. Add a calendar_year column as the 4th column containing either 2018, 2019 or 2020 values
4. 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

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

   segment | demographic |
   C | Couples |
   F | Families |

6. Ensure all null string values with an "unknown" string value in the original segment column as well as the new age_band and demographic columns

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

B. Data Exploration

1. Which week numbers are missing from the dataset?
2. How many total transactions were there for each year in the dataset
3. What are the total sales for each region for each month?
4. What is the total count of transactions for each platform
5. What is the percentage of sales for Retail vs Shopify for each month?
6. What is the percentage of sales by demographic for each year in the dataset?
7. Which age_band and demographic values contribute the most to Retail sales?


THANK YOU