**PROBLEM STATEMENT**

**EcoMart** is my latest venture, and I am seeking your assistance to analyze its sales and performance. In June 2020, we implemented significant supply chain changes at EcoMart, transitioning all our products to sustainable packaging methods from the farm to the customer.

I need your expertise to measure the impact of this change on EcoMart's sales performance and its various business segments.

**Case Study Questions**

**Data Cleansing 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 (e.g., any value from January 1st to 7th will be 1, January 8th to 14th will be 2, etc.).
2. Add a **month\_number** as the third column with the calendar month for each **week\_date** value.
3. Add a **calendar\_year** column as the fourth column containing values of either 2018, 2019, or 2020.
4. Add a new column called **age\_band** after the original **segment** column.

|  |  |
| --- | --- |
| **segment** | **age\_band** |
| **1** | Young Adults |
| **2** | Middle Aged |
| **3 or 4** | Retirees |

1. Add a new **demographic** column using the following mapping based on the first letter in the **segment** values:
   * **C** -> **Couples**
   * **F** -> **Families**
2. Replace all null string values in the original **segment** column, as well as in the new **age\_band** and **demographic** columns, with the string value "unknown".
3. Generate a new **avg\_transaction** column by dividing the **sales** value by **transactions** and rounding to 2 decimal places for each record.

**Data Exploration**

1. Identify which **week\_number** values are missing from the dataset.
2. Determine the total number of transactions for each year in the dataset.
3. Calculate the total sales for each region for each month.
4. Count the total number of transactions for each platform.
5. Calculate the percentage of sales for Retail versus Shopify for each month.
6. Determine the percentage of sales by demographic for each year in the dataset.
7. Identify which **age\_band** and **demographic** values contribute the most to Retail sales.