**SQL Case Study 1: Data Mart Analysis**



**INTRODUCTION:**

Data Dart is my latest venture and I want your help to analyze the sales and performance of my venture. 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.

I need your help to quantify the impact of this change on the sales performance for Data Mart and its separate business areas.

**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 |

**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, 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 |

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

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

1. 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
2. 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?



Data cleaning: It is very important to clean the data and also I don’t understand what is c3 and f1 in data set

For example 1 to 8 as 1st week like that we want we will use week function

To find the week number

Step1: add a week number

Now we create a table clean\_weekly\_sales and we don’t need to add col separately why because we are adding data on which we perform data cleaning on weekly sales only whatever we got the data after cleaning we will store that data in clean\_weekly\_sales

Here we are applying three cases because we are having segment part here we are taking right(segment,1)because in segment I have example c3 in that I want only right side 3 value

Now data exploration part

Here we are using x+50 so that I can get the values added to the first and I get the remaining 50 values

But I want to consider only 52 weeks right now create table for 52 but we want to consider only 52 numbers right

I want to get only 52 so create table for 52 and limit the value for 52

First we create CTE it’s a temporary table I want to access table for someother query we have month number calendar year and monthly sales and platform

Over partition will apply for each row aggregation and do partition