**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.

**Done By : L. ANUSH BHARATHWAJ**

**[LINK](https://drive.google.com/drive/folders/1-dulIJKUcYr8jP4K3hGkcPVH5NO-xKj6?usp=sharing)**

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.**

**create table clean\_weekly\_sales as**

**select**

**week\_date,**

**week(week\_date) as week\_number ,**

**month(week\_date) as month\_number,**

**monthname(week\_date) as month\_name,**

**year(week\_date) as calendar\_year,**

**case**

**when segment='null' then 'Unknown'**

**else segment**

**end as 'segment',**

**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',**

**platform,**

**region,**

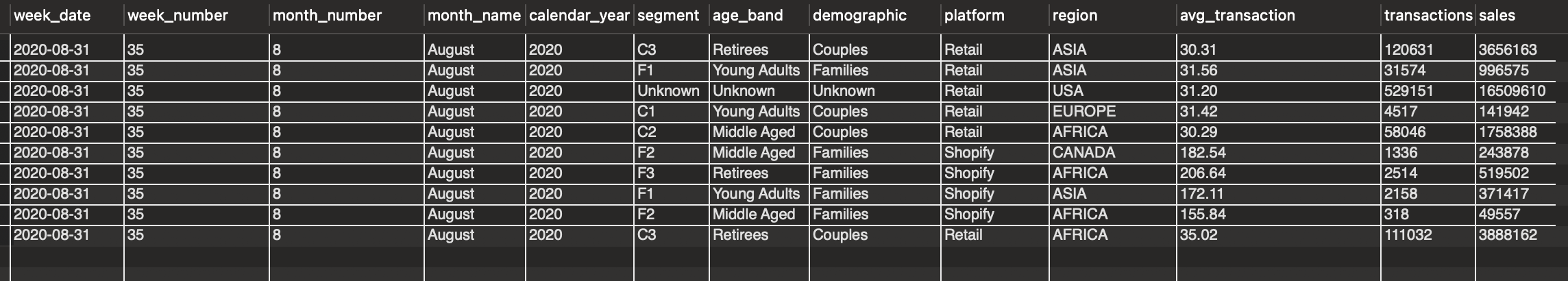
**Round(sales/transactions,2) as 'avg\_transaction',**

**transactions,**

**sales**

**from weekly\_sales;**

**select \* from clean\_weekly\_sales limit 10;**

****

## **B. Data Exploration**

1. **Which week numbers are missing from the dataset?**

**create table seq100(x int auto\_increment primary key);**

**insert into seq100 values (),(),(),(),(),(),(),(),(),();**

**insert into seq100 values (),(),(),(),(),(),(),(),(),();**

**insert into seq100 values (),(),(),(),(),(),(),(),(),();**

**insert into seq100 values (),(),(),(),(),(),(),(),(),();**

**insert into seq100 values (),(),(),(),(),(),(),(),(),();**

**insert into seq100 select x+50 from seq100;**

**select \* from seq100;**

**create table seq52 as select x from seq100 limit 52;**

**select \* from seq52;**

**select x as 'Miss\_week\_numbers' from seq52 where x not in (select distinct week\_number from clean\_weekly\_sales);**

**A black and white rectangular object with white lines

Description automatically generated with medium confidence**

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

**select calendar\_year as 'Year' ,sum(avg\_transaction) as 'Total Transactions' from clean\_weekly\_sales group by Year;**

A screenshot of a computer

Description automatically generated

1. **What are the total sales for each region for each month?**

**select month\_name , month\_number ,region ,sum(sales) from clean\_weekly\_sales group by month\_name, month\_number,region;**

A black and white table with numbers and names

Description automatically generated

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

**select platform , count(transactions) from clean\_weekly\_sales group by platform;**

**A black and white screen with numbers

Description automatically generated**

**select platform , sum(transactions) from clean\_weekly\_sales group by platform;**

**A screenshot of a phone

Description automatically generated**

1. **What is the percentage of sales for Retail vs Shopify for each month?**

**with cte\_montly\_sales as**

**( select month\_number , calendar\_year , platform , SUM(sales) as saless from clean\_weekly\_sales group by month\_number , calendar\_year , platform )**

**select month\_number , calendar\_year ,**

**ROUND( 100 \* MAX(case when platform = 'Retail' then saless else NULL end) / SUM(saless),2) as 'retail\_perc',**

**ROUND( 100 \* MAX(case when platform = 'Shopify' then saless else NULL end) / SUM(saless),2) as 'shopify\_perc'**

**from cte\_montly\_sales**

**group by month\_number , calendar\_year**

**order by month\_number , calendar\_year ;**

A black and white table with numbers

Description automatically generated

1. **What is the percentage of sales by demographic for each year in the dataset?**

**select calendar\_year, demographic , sum(sales) as year\_sales, round( 100 \* sum(sales)/sum(sum(sales)) over(partition by demographic),2)as perc from clean\_weekly\_sales**

**group by calendar\_year, demographic**

**order by calendar\_year, demographic;**

A black and white table with numbers

Description automatically generated

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

**select age\_band , demographic , sum(sales) as Sales from clean\_weekly\_sales where platform = 'Retail' group by age\_band , demographic order by Sales desc;**

**A screenshot of a black screen

Description automatically generated**