# Data Mart Analysis

## SQL Project: Data Mart Solutions

### Data Cleansing

```
CREATE TABLE clean_weekly_sales AS
SELECT
  week_date,
  week(week_date) AS week_number,
  month(week_date) AS month_number,
  year(week_date) AS calendar_year,
  region,
  platform,
  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,
  customer_type,
  transactions,
  sales,
  ROUND(
     sales / transactions,
     2
  ) AS avg_transaction
FROM weekly_sales;

select * from clean_weekly_sales limit 10;
```

## Data Exploration

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

```
create table seq100
(x int not null 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 distinct x as week_day from seq52 where x not in(select distinct week_number from
clean_weekly_sales);

select distinct week_number 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;
```

### 3.What are the total sales for each region for each month?

```
SELECT
 month_number,
 region,
 SUM(sales) AS total_sales
FROM clean_weekly_sales
GROUP BY month_number, region
ORDER BY month_number, region;
```

### 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 percentage of sales for Retail vs Shopify for each month?

```
WITH cte_monthly_platform_sales AS (
 SELECT
  month_number,calendar_year,
```

```
  platform,
  SUM(sales) AS monthly_sales
 FROM clean_weekly_sales
 GROUP BY month_number,calendar_year, platform
)
SELECT
 month_number,calendar_year,
 ROUND(
  100 * MAX(CASE WHEN platform = 'Retail' THEN monthly_sales ELSE NULL END) /
   SUM(monthly_sales),
  2
 ) AS retail_percentage,
 ROUND(
  100 * MAX(CASE WHEN platform = 'Shopify' THEN monthly_sales ELSE NULL END) /
   SUM(monthly_sales),
  2
 ) AS shopify_percentage
FROM cte_monthly_platform_sales
GROUP BY month_number,calendar_year
ORDER BY month_number,calendar_year;
```

### 6.What is the percentage of sales by demographic for each year in the dataset?

```
SELECT
 calendar_year,
 demographic,
 SUM(SALES) AS yearly_sales,
 ROUND(
  (
   100 * SUM(sales)/
    SUM(SUM(SALES)) OVER (PARTITION BY demographic)
  ),
  2
 ) AS percentage
FROM clean_weekly_sales
GROUP BY
 calendar_year,
 demographic
ORDER BY
 calendar_year,
 demographic;
```

### 7.Which age_band and demographic values contribute the most to Retail sales?

```
SELECT
 age_band,
```

```
  demographic,
  SUM(sales) AS total_sales
FROM clean_weekly_sales
WHERE platform = 'Retail'
GROUP BY age_band, demographic
ORDER BY total_sales DESC;
```