# 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;