**DATA MART ASSIGNMENT SOLUTION**

## 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

calender\_year,

SUM(transactions) AS total\_transactions

FROM clean\_weekly\_sales group by calender\_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?**

SELECT

month\_number,

calender\_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 (

SELECT

month\_number,

calender\_year,

platform,

SUM(sales) AS monthly\_sales

FROM clean\_weekly\_sales

GROUP BY month\_number, calender\_year, platform

) AS monthly\_platform\_sales

GROUP BY month\_number, calender\_year

ORDER BY month\_number, calender\_year;

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

select calender\_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 calender\_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;