**-- Create database**

CREATE DATABASE IF NOT EXISTS walmartSales;

USE walmartSales;

**-- Create table**

CREATE TABLE IF NOT EXISTS sales(

invoice\_id VARCHAR(30) NOT NULL PRIMARY KEY,

branch VARCHAR(5) NOT NULL,

city VARCHAR(30) NOT NULL,

customer\_type VARCHAR(30) NOT NULL,

gender VARCHAR(30) NOT NULL,

product\_line VARCHAR(100) NOT NULL,

unit\_price DECIMAL(10,2) NOT NULL,

quantity INT NOT NULL,

tax\_pct FLOAT(6,4) NOT NULL,

total DECIMAL(12, 4) NOT NULL,

date DATETIME NOT NULL,

time TIME NOT NULL,

payment VARCHAR(15) NOT NULL,

cogs DECIMAL(10,2) NOT NULL,

gross\_margin\_pct FLOAT(11,9),

gross\_income DECIMAL(12, 4),

rating FLOAT(2, 1)

);

**-- Data cleaning**

SELECT \* FROM sales;

**-- Add the time\_of\_day column**

SELECT

time,

(CASE

WHEN `time` BETWEEN "00:00:00" AND "12:00:00" THEN "Morning"

WHEN `time` BETWEEN "12:01:00" AND "16:00:00" THEN "Afternoon"

ELSE "Evening"

END) AS time\_of\_day

FROM sales;

ALTER TABLE sales ADD COLUMN time\_of\_day VARCHAR(20);

UPDATE sales

SET time\_of\_day = (

CASE

WHEN `time` BETWEEN "00:00:00" AND "12:00:00" THEN "Morning"

WHEN `time` BETWEEN "12:01:00" AND "16:00:00" THEN "Afternoon"

ELSE "Evening"

END

);

**-- Add day\_name column**

SELECT

date, DAYNAME(date)

FROM sales;

ALTER TABLE sales ADD COLUMN day\_name VARCHAR(10);

UPDATE sales

SET day\_name = DAYNAME(date);

**-- Add month\_name column**

SELECT

date,

MONTHNAME(date)

FROM sales;

ALTER TABLE sales ADD COLUMN month\_name VARCHAR(10);

UPDATE sales

SET month\_name = MONTHNAME(date);

**-- GENERIC QUESTIONS**

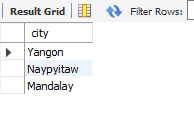
-- 1 How many unique cities does the data have?

SELECT

DISTINCT city

FROM sales;

RESULT:



-- 2 In which city is each branch?

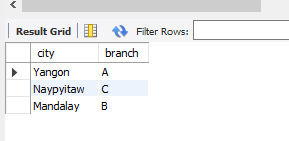
SELECT

DISTINCT city,

branch

FROM sales;

RESULT:



**-- PRODUCT QUESTIONS**

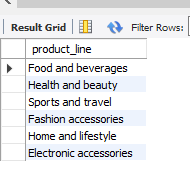
-- 1 How many unique product lines does the data have?

SELECT

DISTINCT product\_line

FROM sales;

RESULT:



-- 2 What is the most selling product line

SELECT

SUM(quantity) as qty,

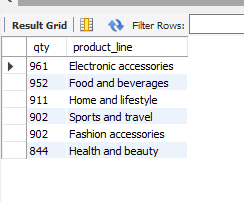
product\_line

FROM sales

GROUP BY product\_line

ORDER BY qty DESC;

RESULT:



-- 3 What is the total revenue by month

SELECT

month\_name AS month,

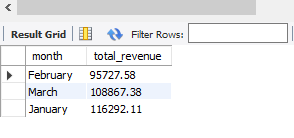
SUM(total) AS total\_revenue

FROM sales

GROUP BY month\_name

ORDER BY total\_revenue;

RESULT:



-- 4 What month had the largest COGS?

SELECT

month\_name AS month,

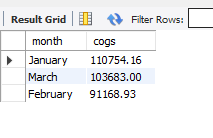
SUM(cogs) AS cogs

FROM sales

GROUP BY month\_name

ORDER BY cogs;

RESULT:



-- 5 What product line had the largest revenue?

SELECT

product\_line,

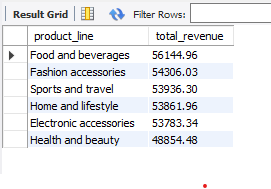
SUM(total) as total\_revenue

FROM sales

GROUP BY product\_line

ORDER BY total\_revenue DESC;

RESULT:



-- 6 What is the city with the largest revenue?

SELECT

branch,

city,

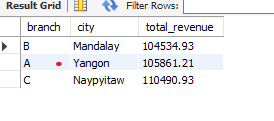
SUM(total) AS total\_revenue

FROM sales

GROUP BY city, branch

ORDER BY total\_revenue;

RESULT:



-- 7 What product line had the largest VAT?

SELECT

product\_line,

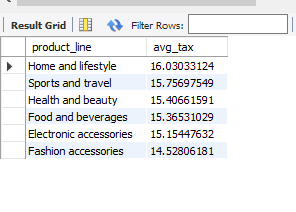
AVG(tax\_pct) as avg\_tax

FROM sales

GROUP BY product\_line

ORDER BY avg\_tax DESC;

RESULT:



-- 8 Fetch each product line and add a column to those product

-- line showing "Good", "Bad". Good if its greater than average sales

SELECT

AVG(quantity) AS avg\_qnty

FROM sales;

SELECT

product\_line,

CASE

WHEN AVG(quantity) > 6 THEN "Good"

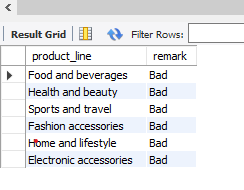
ELSE "Bad"

END AS remark

FROM sales

GROUP BY product\_line;

RESULT:



-- 9 Which branch sold more products than average product sold?

SELECT

branch,

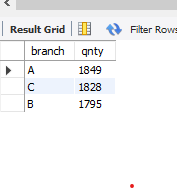
SUM(quantity) AS qnty

FROM sales

GROUP BY branch

HAVING SUM(quantity) > (SELECT AVG(quantity) FROM sales);

RESULT:



-- 10 What is the most common product line by gender

SELECT

gender,

product\_line,

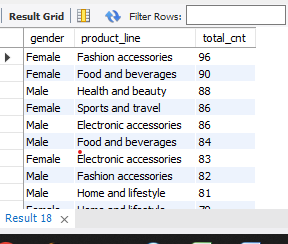
COUNT(gender) AS total\_cnt

FROM sales

GROUP BY gender, product\_line

ORDER BY total\_cnt DESC;

RESULT:



-- 11 What is the average rating of each product line

SELECT

ROUND(AVG(rating), 2) as avg\_rating,

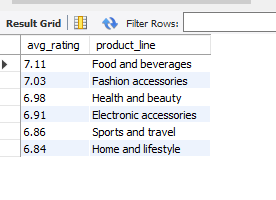
product\_line

FROM sales

GROUP BY product\_line

ORDER BY avg\_rating DESC;

RESULT:



-------------------------------------------------------------------------------------------------------------------

--CUSTOMER QUESTIONS

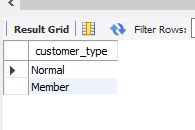
-- 1 How many unique customer types does the data have?

SELECT

DISTINCT customer\_type

FROM sales;

RESULT:



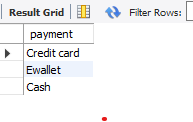
-- 2 How many unique payment methods does the data have?

SELECT

DISTINCT payment

FROM sales;

RESULT:



-- 3 What is the most common customer type?

SELECT

customer\_type,

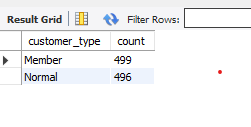
count(\*) as count

FROM sales

GROUP BY customer\_type

ORDER BY count DESC;

RESULT:



-- 5 Which customer type buys the most?

SELECT

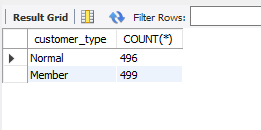
customer\_type,

COUNT(\*)

FROM sales

GROUP BY customer\_type;

RESULT:



-- 6 What is the gender of most of the customers?

SELECT

gender,

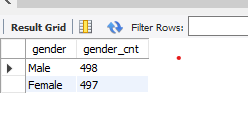
COUNT(\*) as gender\_cnt

FROM sales

GROUP BY gender

ORDER BY gender\_cnt DESC;

RESULT:



-- 7 What is the gender distribution per branch?

SELECT

branch,gender,

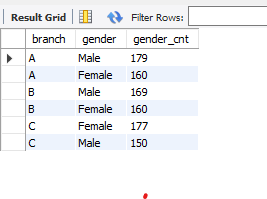
COUNT(\*) as gender\_cnt

FROM sales

GROUP BY branch,gender

ORDER BY branch,gender\_cnt DESC;

RESULT:



Gender per branch is more or less the same hence, I don't think has an effect of the sales per branch and other factors

-- 8 Which time of the day do customers give most ratings?

SELECT

time\_of\_day,

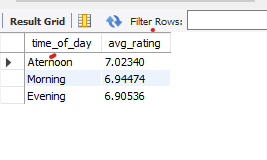
AVG(rating) AS avg\_rating

FROM sales

GROUP BY time\_of\_day

ORDER BY avg\_rating DESC;

RESULT:



-- Looks like time of the day does not really affect the rating, its

-- more or less the same rating each time of the day.

-- 9 Which time of the day do customers give most ratings per branch?

SELECT branch,

time\_of\_day,

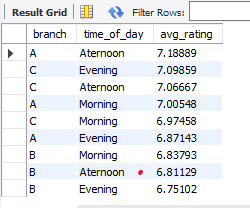
AVG(rating) AS avg\_rating

FROM sales

GROUP BY branch,time\_of\_day

ORDER BY avg\_rating DESC;

RESULT:



-- Branch A and C are doing well in ratings, branch B needs to do a

-- little more to get better ratings.

-- 10 Which day fo the week has the best avg ratings?how many sales are made on these days?

SELECT

day\_name,

AVG(rating) AS avg\_rating,

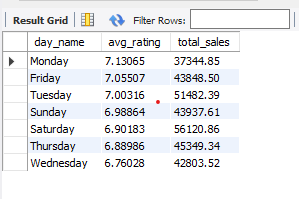
SUM(total) as total\_sales

FROM sales

GROUP BY day\_name

ORDER BY avg\_rating DESC;

RESULT:



-- Mon, Tue and Friday are the top best days for good ratings

-- 11 Which day of the week has the best average ratings to branch C?

SELECT

day\_name,

COUNT(day\_name) total\_sales

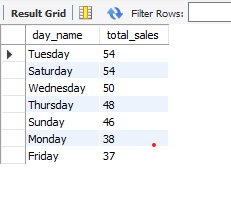
FROM sales

WHERE branch = "C"

GROUP BY day\_name

ORDER BY total\_sales DESC;

RESULT:



**-- SALES QUESTION**

-- 1 Number of sales made in each time of the day per weekday.

SELECT

day\_name as week\_day,time\_of\_day,

COUNT(\*) AS total\_sales

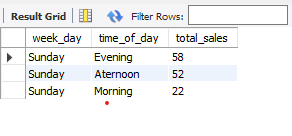
FROM sales

WHERE day\_name = "Sunday"

GROUP BY time\_of\_day

ORDER BY total\_sales DESC;

RESULT:



-- Evenings experience most sales, the stores are

-- filled during the evening hours

-- 2 Which of the customer types brings the most revenue?

SELECT

customer\_type,

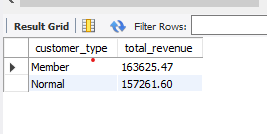
SUM(total) AS total\_revenue

FROM sales

GROUP BY customer\_type

ORDER BY total\_revenue DESC;

RESULT:



-- 3 Which city has the largest tax/VAT percent?

SELECT

city,

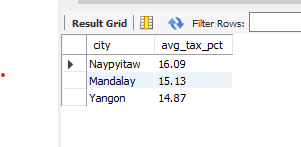
ROUND(AVG(tax\_pct), 2) AS avg\_tax\_pct

FROM sales

GROUP BY city

ORDER BY avg\_tax\_pct DESC;

RESULT:



-- 4 Which customer type pays the most in VAT?

SELECT

customer\_type,

AVG(tax\_pct) AS total\_tax

FROM sales

GROUP BY customer\_type

ORDER BY total\_tax;

RESULT:

