**CODE**

CREATE DATABASE IF NOT EXISTS walmartDatabase;

USE walmartDatabase;

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(10) NOT NULL,

product\_line VARCHAR(50) NOT NULL,

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

quantity INT NOT NULL,

VAT FLOAT(6) NOT NULL,

total DECIMAL(12,4) NOT NULL,

date DATETIME NOT NULL,

time TIME NOT NULL,

payment\_method VARCHAR(15) NOT NULL,

cogs DECIMAL(10,2) NOT NULL,

gross\_margin\_pct FLOAT(11),

gross\_income DECIMAL(12,4) NOT NULL,

rating FLOAT(2) NOT NULL

);

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

-- ---------------------------------------------------------FEATURE ENGINEERING-------------------------------------------------------------------

-- time\_of\_day

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\_date

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

);

-- day\_name

SELECT date, dayname(date)

FROM sales;

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

UPDATE sales

SET day\_name = dayname(date);

-- month\_name

SELECT date, monthname(date)

FROM sales;

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

UPDATE sales

SET month\_name = monthname(date);

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

-- ----------------------------Generic questions------------------

-- What are the unique cities present in the data? count how many times it has occured?.-------------------

SELECT DISTINCT(city),count(city) AS count

FROM sales

GROUP BY city;

-- In which city is each branch?

SELECT DISTINCT(branch),city

FROM sales;

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

-- ----------------------------Product based questions--------------------------

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

SELECT COUNT(DISTINCT(product\_line))

FROM sales;

-- What are the unique values of product lines? and count each,

SELECT DISTINCT(product\_line), count(product\_line)

FROM sales

GROUP BY product\_line;

-- What is the most common payment method?

SELECT payment\_method, count(payment\_method) as payement\_count

FROM sales

GROUP BY payment\_method

ORDER BY payement\_count DESC

LIMIT 1;

-- What is the most selling product line?

SELECT product\_line, sum(quantity) as total\_qty\_sold

FROM sales

GROUP BY product\_line

ORDER BY total\_qty\_sold DESC

LIMIT 1;

-- What is the total revenue by month?

SELECT month\_name, sum(total) as total\_revenue

FROM sales

GROUP BY month\_name;

-- What month had the largest COGS?

SELECT month\_name, sum(cogs) as total\_cogs

FROM sales

GROUP BY month\_name

ORDER BY total\_cogs DESC

LIMIT 1;

-- 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

LIMIT 1;

-- What is the city with the largest revenue?

SELECT city, sum(total) as total\_revenue

FROM sales

GROUP BY city

ORDER BY total\_revenue DESC

LIMIT 1;

-- What product line had the largest VAT?

SELECT product\_line, sum(vat) as total\_vat

FROM sales

GROUP BY product\_line

ORDER BY total\_vat

LIMIT 1;

-- Fetch each product line and add a column to those product line showing "Good", "Bad". Good if its greater than average sales

SELECT product\_line, total,

CASE

WHEN total> (select avg(total) from sales) then "Good"

ELSE "Bad"

END AS good\_bad\_rating

FROM sales;

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

SELECT branch, sum(quantity) as total\_qty\_sold

FROM sales

GROUP BY branch

HAVING total\_qty\_sold > (SELECT avg(total\_qty\_sold) FROM (

SELECT sum(quantity) as total\_qty\_sold

FROM sales

GROUP BY branch) AS subquery);

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

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

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

-- ---------------------------------------Sales based questions------------------------------------------------

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

SELECT day\_name as weekday, time\_of\_day, sum(quantity)

FROM sales

GROUP BY weekday, time\_of\_day

ORDER BY field(weekday, 'Monday', 'Tuesday', 'Wednesday','Thursday', 'Friday', 'Saturday', 'Sunday'),

field(time\_of\_day, 'Morning','Afternoon','Evening');

-- 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

LIMIT 1;

-- Which city has the largest tax percent/ VAT (Value Added Tax)?

SELECT city, sum(vat) as total\_vat

FROM sales

GROUP BY city

ORDER BY total\_vat DESC

LIMIT 1;

-- Which customer type pays the most in VAT?

SELECT customer\_type, sum(vat) as total\_vat

FROM sales

GROUP BY customer\_type

ORDER BY total\_vat DESC

LIMIT 1;

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

-- ------------------------------------Customer based questions------------------------------------------------

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

SELECT count(DISTINCT customer\_type) as no\_of\_customer\_types

FROM sales;

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

SELECT count(DISTINCT payment\_method) as payment\_types

FROM sales;

-- What is the most common customer type?

SELECT customer\_type, count(customer\_type) as total\_customer

FROM sales

GROUP BY customer\_type

ORDER BY total\_customer DESC

LIMIT 1;

-- Which customer type buys the most?

SELECT customer\_type, count(quantity) as total\_buy

FROM sales

GROUP BY customer\_type

ORDER BY total\_buy DESC

LIMIT 1;

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

SELECT gender, count(gender) as total\_count

FROM sales

GROUP BY gender

ORDER BY gender DESC

LIMIT 1;

-- What is the gender distribution per branch?

SELECT branch, gender, count(gender)

FROM sales

GROUP BY branch, gender

ORDER BY branch;

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

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

-- Which day fo the week has the best avg ratings?

-- Which day of the week has the best average ratings per branch?