**Feature Engineering:**

**-- ------------------------ TIME\_OF\_DAY --------------------------------**

ALTER TABLE sales

ADD COLUMN time\_of\_day VARCHAR(10) DEFAULT NULL;

SET SQL\_SAFE\_UPDATES=0;

UPDATE sales

SET time\_of\_day = CASE

WHEN HOUR(Time) < 12 THEN 'Morning'

WHEN HOUR(Time) < 18 THEN 'Afternoon'

ELSE 'Evening'

END;

SELECT \* FROM sales;

**-- --------------------- DAY\_NAME --------------------------------------**

ALTER TABLE sales

ADD COLUMN day\_name VARCHAR(10) DEFAULT NULL;

UPDATE sales

SET day\_name = DAYNAME(STR\_TO\_DATE(Date, '%m/%d/%Y'));

**-- ----------------------- MONTH\_NAME ----------------------------------**

ALTER TABLE sales

ADD COLUMN month\_name VARCHAR(10) DEFAULT NULL;

UPDATE sales

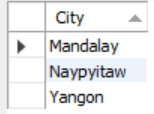
SET month\_name = MONTHNAME(STR\_TO\_DATE(Date, '%m/%d/%Y'));

**GENERIC QUESTIONS:**

**1. How many unique cities does the data have?**

SELECT DISTINCT City

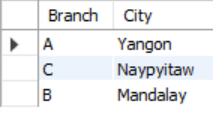
FROM sales;



**2. In which city is each branch?**

SELECT DISTINCT Branch,City

FROM sales;



**PRODUCT BASED QUESTIONS:**

SELECT \* FROM sales;

**1.How many unique product lines does the data have?**

SELECT DISTINCT Product\_line AS unique\_product\_line

FROM sales;



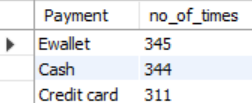
**2.What is the most common payment method?**

SELECT Payment,COUNT(Payment) AS no\_of\_times

FROM sales

GROUP BY Payment

ORDER BY no\_of\_times DESC;



**3. What is the most selling product line?**

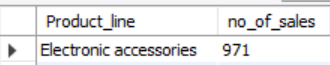
SELECT Product\_line,SUM(Quantity) AS no\_of\_sales

FROM sales

GROUP BY Product\_line

ORDER BY no\_of\_sales DESC

LIMIT 1;



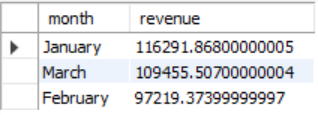
**4. What is the total revenue by month?**

SELECT month\_name AS month, SUM(Total) AS revenue

FROM sales

GROUP BY month

ORDER BY revenue desc;



**5. What month had the largest COGS?**

SELECT

month\_name AS month,

SUM(cogs) AS cogs

FROM sales

GROUP BY month

ORDER BY cogs DESC

LIMIT 1;

**6. What product line had the largest revenue?**

SELECT

Product\_line,

SUM(Total) AS highest\_revenue

FROM sales

GROUP BY Product\_line

ORDER BY highest\_revenue DESC;

**7. What is the city with the largest revenue?**

SELECT

City,

SUM(Total) AS highest\_revenue

FROM sales

GROUP BY City

ORDER BY highest\_revenue DESC;

**8.What product line had the largest VAT?**

SELECT

Product\_line,

SUM(VAT) AS VAT

FROM sales

GROUP BY Product\_line

ORDER BY VAT DESC

LIMIT 1;

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

ALTER TABLE sales

ADD COLUMN Product\_status VARCHAR(10) DEFAULT NULL;

ALTER TABLE sales

CHANGE Product\_status Product\_status VARCHAR(10) AFTER Total;

WITH avg\_sales AS(

SELECT Product\_line,

AVG(Total) AS avg\_sale

FROM sales

GROUP BY Product\_line)

UPDATE sales s

SET Product\_status = CASE

WHEN Total > (SELECT avg\_sale FROM avg\_sales a

WHERE s.product\_line = a.product\_line

) THEN 'Good'

ELSE 'Bad'

END;

select \* from sales;

**10. Which branch sold more products than average product sold?**

WITH product\_sold AS (SELECT

Branch,

SUM(Quantity) AS products\_sold

FROM sales

GROUP BY Branch),

product\_avg AS (SELECT AVG(products\_sold) AS avg\_product FROM product\_sold)

SELECT Branch,Products\_sold AS more\_than\_avg

FROM product\_sold ps,product\_avg pa

WHERE ps.Products\_sold > pa.avg\_Product;

**11. What is the most common product line by gender?**

WITH ranked\_product\_lines AS (

SELECT

gender,

product\_line,

SUM(Quantity) AS products\_sold,

ROW\_NUMBER() OVER (PARTITION BY gender ORDER BY SUM(Quantity) DESC) AS rn

FROM sales

GROUP BY gender, product\_line

)

SELECT gender, product\_line,products\_sold

FROM ranked\_product\_lines

WHERE rn = 1;

**12. What is the average rating of each product line?**

SELECT

Product\_line,

AVG(Rating) AS avg\_rating

FROM sales

GROUP BY Product\_line

ORDER BY avg\_rating DESC;

**SALES BASED QUESTIONS:**

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

SELECT

time\_of\_day,

COUNT(\*) AS total\_sales

FROM sales

WHERE day\_name='Monday'

GROUP BY time\_of\_day

ORDER BY total\_sales DESC;

**2. Which of the customer types brings the most revenue?**

SELECT

Customer\_type,

SUM(Total) AS revenue

FROM sales

GROUP BY Customer\_type

ORDER BY revenue DESC;

**3. Which city has the largest tax percent/ VAT (Value Added Tax)?**

SELECT

City,

MAX(VAT) AS largest\_tax

FROM sales

GROUP BY City

ORDER BY largest\_tax DESC;

**4. Which customer type pays the most in VAT?**

SELECT

Customer\_type,

MAX(VAT) AS largest\_tax

FROM sales

GROUP BY Customer\_type

ORDER BY largest\_tax DESC;

**5.Which day in the week has highest sales?**

SELECT day\_name,SUM(Quantity) AS no\_of\_sales

FROM sales

GROUP BY day\_name

ORDER BY no\_of\_sales DESC

LIMIT 1;

**6.Which day in the week has lowest sales?**

SELECT day\_name,SUM(Quantity) AS no\_of\_sales

FROM sales

GROUP BY day\_name

ORDER BY no\_of\_sales

LIMIT 1;

**CUSTOMER BASED QUESTIONS:**

**1. How many unique customer types does the data have?**

SELECT DISTINCT Customer\_type

FROM sales;

**2. How many unique payment methods does the data have?**

SELECT DISTINCT Payment

FROM sales;

**3. What is the most common customer type?**

SELECT Customer\_type,COUNT(\*) AS no\_of\_customers

FROM sales

GROUP BY Customer\_type;

**4. Which customer type buys the most?**

SELECT

Customer\_type,

COUNT(\*) AS no\_of\_customers

FROM sales

GROUP BY Customer\_type;

**5. What is the gender of most of the customers?**

SELECT Gender,COUNT(\*) AS num\_of\_customers

FROM sales

GROUP BY Gender;

**6. What is the gender distribution per branch?**

SELECT Branch,Gender, COUNT(\*) AS count\_per\_gender

FROM sales

GROUP BY Branch, Gender

ORDER BY Branch;

**7. Which time of the day do customers give most ratings?**

SELECT time\_of\_day,COUNT(Rating) AS no\_of\_ratings

FROM sales

GROUP BY time\_of\_day

ORDER BY no\_of\_ratings DESC;

**8. Which time of the day do customers give most ratings per branch?**

SELECT Branch,time\_of\_day,COUNT(Rating) AS no\_of\_ratings

FROM sales

WHERE Branch='A'

GROUP BY time\_of\_day

ORDER BY no\_of\_ratings DESC;

**9. Which day of the week has the best avg ratings?**

SELECT

day\_name,

AVG(Rating) AS avg\_ratings

FROM sales

GROUP BY day\_name

ORDER BY avg\_ratings DESC;

**10. Which day of the week has the best average ratings per branch?**

SELECT Branch,day\_name,AVG(Rating) AS no\_of\_ratings

FROM sales

WHERE Branch='A'

GROUP BY day\_name

ORDER BY no\_of\_ratings DESC;

**11.Which day in the week has highest sales?**

SELECT day\_name,SUM(Quantity) AS no\_of\_sales

FROM sales

GROUP BY day\_name

ORDER BY no\_of\_sales DESC

LIMIT 1;

**12.Which day in the week has lowest sales?**

SELECT day\_name,SUM(Quantity) AS no\_of\_sales

FROM sales

GROUP BY day\_name

ORDER BY no\_of\_sales

LIMIT 1;