-- ADD THESE COLUMNS

- -- 1. timeofday Morning, Afternoon and Evening. (extract in time column)
- -- 2. dayname (Mon, Tue, Wed, Thur, Fri). (extract in date column)
- -- 3. monthname (Jan, Feb, Mar) (extrat in date column)

- -- Feature Engineering: This will help us generate some new columns from existing ones.

- ALTER TABLE AMAZON
- MODIFY Date Date;

- ALTER TABLE AMAZON
- ADD COLUMN TIME_OF_DAY VARCHAR(50);

- ALTER TABLE AMAZON
- ADD COLUMN DAY_NAME VARCHAR(50);

- ALTER TABLE AMAZON
- ADD COLUMN MONTH_NAME VARCHAR(50);

UPDATE COMMAND

- -- 1. timeofday Morning, Afternoon and Evening. (extract in time column)
- -- morning_range = range(6, 12) # 6:00 AM to 11:59 AM
- -- afternoon_range = range(12, 18) # 12:00 PM to 5:59 PM
- -- evening_range = range(18, 24) # 6:00 PM to 11:59 PM

- UPDATE AMAZON
- SET TIME_OF_DAY =
- CASE
- WHEN Time BETWEEN '06:00:00' AND '11:59:58.999' THEN 'Morning'
- WHEN Time BETWEEN '12:00:00' AND '17:59:58.999' THEN 'Afternoon'
- WHEN Time BETWEEN '18:00:00' AND '24:00:00' THEN 'Evening'
- ELSE 'NO'
- END ;

- -- 2. dayname (Mon, Tue, Wed, Thur, Fri). (extract in date column)

- UPDATE AMAZON
- SET DAY_NAME = DAYNAME(Date);

- -- 3. monthname (Jan, Feb, Mar) (extrat in date column)

- UPDATE AMAZON
- SET MONTH_NAME = MONTHNAME(Date);

-- 1. What is the count of distinct cities in the dataset?

- SELECT City,
- COUNT(DISTINCT City) AS Unique_City
- FROM AMAZON
- GROUP BY City;

- -- THERE ARE THREE UNIQUE CITY IN DATASET (MANDALAY, NAYPYITAW, YANGON)

-- 2. For each branch, what is the corresponding city?

- SELECT DISTINCT Branch, City
- FROM AMAZON;

- -- (A=YANGON), (B=MANDALAY), (C=NAYPYITAW) . THERE ARE THREE UNIQUE BRANCH.

-- 3. What is the count of distinct product lines in the dataset?

- SELECT Product_line,
- COUNT(DISTINCT Product_line) AS Unique_Products
- FROM AMAZON
- GROUP BY Product_line;

- -- THERE ARE SIX PRODUCTS IN DATASET (FASHION ACCESSORIES, FOOD AND BEVERAGE, HEALTH AND BEAUTY, HOME AND LIFESTYLE, SPORTS AND TRAVEL, ELECRONIC ACCESORIES)

-- 4. Which payment method occurs most frequently?

- SELECT Payment,
- COUNT(*) AS Used_Payment_Type
- FROM AMAZON
- GROUP BY Payment
- ORDER BY Used_Payment_Type DESC;

- -- EWALLET PAYMENT METHOD ARE USED MOST THAN OTHER.

-- 5. Which product line has the highest sales?

- SELECT Product_line,
- COUNT(*) AS Total_Sales
- FROM AMAZON
- GROUP BY Product_line
- ORDER BY Total_Sales DESC;

- -- FASHION ACCESSORIES ARE THE PRODUCT WHICH IS MORE SALES.

-- 6. How much revenue is generated each month?

```
SELECT MONTH_NAME,
       ROUND(SUM(Total),2) AS Total_Revenue
FROM AMAZON
GROUP BY MONTH_NAME
ORDER BY
  CASE MONTH_NAME
      WHEN 'January' THEN 1
      WHEN 'February' THEN 2
      WHEN 'March' THEN 3
  END ASC;
```

- -- TOTAL_REVENUE GENERATED BY MONTH ARE (JANUARY = 116291.87, FEBRUARY = 97219.37, MARCH = 109455.51), JANUARY IS HIGHEST .

-- 7. In which month did the cost of goods sold reach its peak?

```
SELECT MONTH_NAME,
       ROUND(SUM(cogs),2) AS Cost_Of_Goods
FROM AMAZON
GROUP BY MONTH_NAME
ORDER BY Cost_Of_Goods DESC;
```

-- JANUARY IS THE MONTH WHERE COST_OF_GOODS SOLD PEAK.

-- 8. Which product line generated the highest revenue?

- SELECT Product_line,
- ROUND(SUM(Total),2) AS Total_Revenue
- FROM AMAZON
- GROUP BY Product_line
- ORDER BY Total_Revenue DESC;

- -- FOOD AND BEVERAGES ARE THE PRODUCT WHO GENERATE MOST REVENUE.

-- 9. In which city was the highest revenue recorded?

- SELECT City,
- ROUND(SUM(Total),2) AS Total_Revenue
- FROM AMAZON
- GROUP BY City
- ORDER BY Total_Revenue DESC;

- -- NAYPYITAW IS THE CITY WHO GENERATE MOST REVENUE.

-- 10. Which product line incurred the highest Value Added Tax?

```
SELECT Product_line,
       MAX(`Tax_5%`) AS Highest_Tax
FROM AMAZON
GROUP BY Product_line
ORDER BY Highest_Tax DESC;
```

-- FASHION ACCESSORIES ARE THE PRODUCT WHO INCURRED HIGHEST TAX.

-- 11. For each product line, add a column indicating "Good" if its sales are above average, otherwise "Bad."

```
WITH SALES AS
 (
SELECT Product_line,
   COUNT(*) AS Total_Sales
FROM AMAZON
GROUP BY Product_line
    )

SELECT Product_line,
 CASE
    WHEN Total_Sales > (SELECT AVG(Total_Sales) FROM SALES) THEN 'Good'
 ELSE 'Bad'
END AS Above_AVG,
(SELECT ROUND(AVG(Total_Sales),2) FROM SALES) AS Average_Sales
FROM SALES;
-- ELECTRONIC ACCESSORIES, FOOD AND BEVERAGES, FASHION ACCESORIES ARE THE PRODUCT WHOSE SALES ARE MORE THAN AVERAGE SALES.
```

-- 12. Identify the branch that exceeded the average number of products sold.

```
WITH TOTAL_SALES AS
 (
SELECT Branch,
   COUNT(*) AS Total_Sales
FROM AMAZON
GROUP BY Branch
),
AVG_SALES AS
 (
SELECT ROUND(AVG(Total_Sales),2) AS AVERAGE_SALES
FROM TOTAL_SALES
)

SELECT TOTAL_SALES.Branch,
   AVG_SALES.AVERAGE_SALES
FROM TOTAL_SALES, AVG_SALES
WHERE TOTAL_SALES.Total_Sales > AVG_SALES.AVERAGE_SALES;
-- BRANCH A SALES IS GREATER THAN AVERAGE SALES.
```

-- 13. Which product line is most frequently associated with each gender?

```
WITH Gender_Product_Count AS
 (
SELECT Gender,
   Product_line,
   COUNT(Product_line) AS Count_Product
FROM AMAZON
GROUP BY Gender, Product_line
 ),

Ranked_Product_Line AS
 ( SELECT Gender,
        Product_line,
        Count_Product,
        ROW_NUMBER() OVER(PARTITION BY Gender ORDER BY Count_Product DESC) AS RN
   FROM Gender_Product_Count)

SELECT Gender,
   Product_line,
      Count_Product
FROM Ranked_Product_Line
WHERE RN = 1;
-- Fashion Accessories are the most liked product line by females, while Health and Beauty is the most liked product line by males.
```

-- 14. Calculate the average rating for each product line.

- SELECT Product_line,
- ROUND(AVG(Rating),2) AS Average_Rating
- FROM AMAZON
- GROUP BY Product_line
- ORDER BY Average_Rating DESC;
- -- These are the average ratings of the products, with the highest rating given by users for Food and Beverages

-- 15. Count the sales occurrences for each time of day on every weekday.

- SELECT TIME_OF_DAY,
- COUNT(*) AS Sales_Occurence
- FROM AMAZON
- GROUP BY TIME_OF_DAY;
- -- IN AFTERNOON THE SALE IS HIGHER THAN OTHER.

-- 16. Identify the customer type contributing the highest revenue.

- SELECT Customer_type,
- ROUND(SUM(Total),2) AS TotaL_Revenue
- FROM AMAZON
- GROUP BY Customer_type;
- -- CUSTOMER HAVE MEMBERSHIP WHO GENERATE MORE REVENUE THAN NORMAL CUSTOMER.

-- 17. Determine the city with the highest VAT percentage.

- SELECT City,
- MAX(`Tax_5%`) AS Tax_Percentage
- FROM AMAZON
- GROUP BY City
- ORDER BY Tax_Percentage DESC;
- -- NAYPYITAW IS THE CITY WHERE HIGHER VAT PERCENTAGE.

-- 18. Identify the customer type with the highest VAT payments.

- SELECT Customer_type,
- ROUND(SUM(`Tax_5%`),2) AS Total_VAT
- FROM AMAZON
- GROUP BY Customer_type
- ORDER BY Total_VAT DESC;
- -- CUSTOMER WHO HAVE MEMBERSHIP HAVE MORE TOTAL VAT THAN NORMAL.

-- 19. What is the count of distinct customer types in the dataset?

- SELECT Customer_type,
- COUNT(DISTINCT Customer_type) AS Unique_Customer
- FROM AMAZON
- GROUP BY Customer_type;
- -- IN THIS DATASET THERE ARE TWO CUSTOMER TYPE (MEMBER, NORMAL).

-- 20. What is the count of distinct payment methods in the dataset?

- SELECT Payment,
- COUNT(DISTINCT Payment) AS Payment_Type
- FROM AMAZON
- GROUP BY Payment;
- -- IN THIS DATASET THERE ARE THREE PAYMENT METHOD (CASH, CREDIT CARD, EWALLET).

-- 21. Which customer type occurs most frequently?

- SELECT Customer_type,
- COUNT(*) AS Count_Customer_Type
- FROM AMAZON
- GROUP BY Customer_type;
- -- THERE ARE 501 MEMBER CUSTOMER AND 499 NORMAL CUSTOMER.

-- 22. Identify the customer type with the highest purchase frequency.

- SELECT Customer_type,
- SUM(Quantity) AS Total_Quantity_Purchase
- FROM AMAZON
- GROUP BY Customer_type;
- -- CUSTOMER HAVE MEMBERSHIP WHO BOUGHT MORE PRODUCT.

-- 23. Determine the predominant gender among customers.

```
SELECT Gender,
       COUNT(*) AS Total_Order
FROM AMAZON
GROUP BY Gender;
-- FEMALE ARE PREDOMINANT GENDER IN THIS DATASET.
```

-- 24. Examine the distribution of genders within each branch.

```
SELECT Branch,
       Gender,
       COUNt(Gender) AS Count_Gender
FROM AMAZON
GROUP BY Branch, Gender
ORDER BY Branch;
```

-- 25. Identify the time of day when customers provide the most ratings.

- SELECT TIME_OF_DAY,
- COUNT(Rating) AS Rating_Count
- FROM AMAZON
- GROUP BY TIME_OF_DAY
- ORDER BY Rating_Count DESC;
- -- IN AFTERNOON THE CUSTOMER PROVIDE MORE RATING.

-- 26. Determine the time of day with the highest customer ratings for each branch.

- SELECT Branch,
- TIME_OF_DAY,
- Rating AS Highest_Rating
- FROM AMAZON
- WHERE (Branch, Rating) IN (SELECT Branch, MAX(Rating) FROM AMAZON GROUP BY Branch)
- ORDER BY Branch ASC;
- -- BRANCH A AND C HIGHEST RATING IN AFTERNOON ONLY BUT,
- -- BRANCH B HIGHEST RATING IN MORNING, AFTERNOON, EVENING.

-- 27. Identify the day of the week with the highest average ratings.

- SELECT DAY_NAME,
- ROUND(AVG(Rating),2) AS AVG_RATING
- FROM AMAZON
- GROUP BY DAY_NAME
- ORDER BY AVG_RATING DESC;
- -- MONDAY IS THE DAY WHERE HIGHEST AVERAGE RATING IS 7.15.

-- 28. Determine the day of the week with the highest average ratings for each branch.

- SELECT Branch,
- DAY_NAME,
- ROUND(AVG(Rating),2) AS AVG_RATING
- FROM AMAZON
- GROUP BY DAY_NAME, Branch
- ORDER BY AVG_RATING DESC;
- -- BRANCH B DAY MONDAY WHERE AVG RATING IS 7.34 WHICH IS HIGHEST THAN OTHER.