

# SQL- Capstone Project

## Purposes Of The Capstone Project

The major aim of this project is to gain insight into the sales data of Amazon to understand the different factors that affect sales of the different branches.

About Data: This dataset contains sales transactions from three different branches of Amazon, respectively located in Mandalay, Yangon and Naypyitaw. The data contains 17 columns and 1000 rows:

### ANALYSIS LIST: 1. PRODUCT ANALYSIS

#### 2. SALES ANALYSIS

#### 3. CUSTOMER ANALYSIS

##### 1. *Product Analysis:*

- **Top-Performing Product Lines** The "Food and beverages" product line generates the highest revenue \$56144.84, contributing 174 to the overall sales. It also has the highest customer satisfaction rating of 7.1 This line should be prioritized for inventory restocking, marketing, and strategic expansion.
- **Underperforming Product Lines:** Low-performing product lines , such as "'Health and beauty'," require targeted marketing efforts, potential quality improvements, or discontinuation if they remain unprofitable despite intervention. Its contributing least, \$49193.73. contributing 152 to the overall sales
- **Customer Satisfaction:** High ratings indicate customer satisfaction. Focus on replicating the success of well-rated product lines across others.

##### 2. *Sales Analysis:*

- **Monthly Trends:** Peak revenue months , such as March and January, indicate high-demand periods (e.g., holidays). These months can benefit from additional stock, promotional offers, and strategic advertising. Conversely, low-revenue months like February require campaigns or discounts to stimulate sales.
- **Time of Day Trends:** Peak purchasing periods reveal when customers are most active, such as afternoons from 12 PM to 5 PM. Contributing Revenue \$172468.55 .Sales strategies should align with these times, offering incentives during high-traffic periods.
- **Branch Comparisons:** High-performing branches like "C" should serve as benchmarks. Its contributes Highest revenue \$110568.70 Investigate and address

location-specific issues in underperforming branches, "B" which experiences lower sales.

- **Cost Analysis:** High COGS during 'January', aligns with increased sales activity. Contributes COGS 110754. Monitoring these trends ensures profitability while maintaining operational efficiency.

### 3. Customer Analysis:

- **Customer Types:** "Members" contribute significantly to revenue, generating \$164223.44 in total sales. They should be incentivized through loyalty programs or exclusive offers. Non-members generating \$40,000 in revenue could be encouraged to join loyalty schemes.
- **Gender Insights:** Female customers dominate the customer base, making up 51% of total purchases. Focused marketing campaigns tailored to this demographic can increase engagement and sales. Efforts to attract male customers may expand the customer base.
- **Profitability Segments:** Member also contribute the Highest , 7820.16 TAX payments, indicating they purchase premium products. Special offers and programs for these segments can further drive sales and loyalty.

### 4. Recommendations

- **Boost Best-Sellers:** Prioritize resources, advertising, and promotions for top-performing product lines to maintain their market dominance.
- **Address Low Revenue Periods:** Use seasonal insights to plan targeted promotions and campaigns during slow months to maximize revenue potential.
- **Improve Branch Performance:** Conduct detailed evaluations of underperforming branches and implement tailored solutions to overcome challenges.
- **Enhance Customer Engagement:** Strengthen loyalty programs and develop strategies for underrepresented or less profitable customer segments to broaden the customer base and improve profitability.

## ➤ Business Questions To Answer:

### Question with Query codes:

#### 1. WHAT IS THE COUNT OF DISTINCT CITIES IN THE DATASET?

```
SELECT
  COUNT(DISTINCT CITY), CITY
FROM
  SALES_AMAZON
GROUP BY CITY;
# 3 DISTINCT CITIES '1', 'MANDALAY'1', 'NAYPYITAW'1', 'YANGON'
```

## 2.FOR EACH BRANCH, WHAT IS THE CORRESPONDING CITY?

```
SELECT
    CITY, BRANCH
FROM
    SALES_AMAZON;
# FOR EACH AS BRANCH 'YANGON', 'A', 'MANDALAY', 'B', 'NAYPYITAW', 'C'
```

## 3.WHAT IS THE COUNT OF DISTINCT PRODUCT LINES IN THE DATASET?

```
SELECT
    COUNT(DISTINCT PRODUCT_LINE), PRODUCT_LINE
FROM
    SALES_AMAZON
GROUP BY PRODUCT_LINE;
# THE COUNT OF PRODUCT LINE '1', 'ELECTRONIC ACCESSORIES'1, 'FASHION ACCESSORIES'1, 'FOOD
AND BEVERAGES'1,
# 'HEALTH AND BEAUTY'1, 'HOME AND LIFESTYLE'1, 'SPORTS AND TRAVEL'
```

## 4.WHICH PAYMENT METHOD OCCURS MOST FREQUENTLY?

```
SELECT
    PAYMENT, COUNT(*) AS FREQUENCY
FROM
    SALES_AMAZON
GROUP BY PAYMENT
ORDER BY FREQUENCY DESC
LIMIT 1;
#EWALLET PAYMENT OCCURS 345 TIMES
```

## 5.WHICH PRODUCT LINE HAS THE HIGHEST SALES?

```
SELECT
    PRODUCT_LINE, SUM(QUANTITY) AS HIGH_SALES
FROM
    SALES_AMAZON
GROUP BY PRODUCT_LINE
ORDER BY HIGH_SALES DESC
LIMIT 1;
#'ELECTRONIC ACCESSORIES', HIGHEST SALES '971'
```

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

```
select sum(total) as revenue , month_name
from sales_amazon
group by month_name
order by month_name;
#revenue generated by each month '97219.3740', 'February'116291.8680,
'January'109455.5070, 'March'
```

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

```
select month_name, sum(cogs) as peak
from sales_amazon
group by month_name
order by peak desc
limit 1;
#In 'January', is the peak cogs with '110754.1600'
```

8. Which product line generated the highest revenue?

```
select product_line, sum(total) as revenue
from sales_amazon
group by Product_line
order by revenue desc
limit 1;
# 'Food and beverages', highest revenue with '56144.8440'
```

9. In which city was the highest revenue recorded?

```
select city, sum(total) as highest_revenue
from sales_amazon
group by city
order by highest_revenue desc
limit 1;
# 'Naypyitaw', as highest revenue recorded with '110568.7065'
```

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

```
select product_line ,sum(tax) as high_tax
from sales_amazon
```

```
group by Product_line
order by high_tax desc
limit 1 ;
#'Food and beverages', as highest vat '2673.5640'
```

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

```
SELECT
    product_line,
    SUM(total) AS total_sales,
    CASE
        WHEN
            SUM(total) > (SELECT
                AVG(total_sales)
            FROM
                (SELECT
                    SUM(total) AS total_sales
                FROM
                    sales_amazon
                GROUP BY product_line) AS subquery)
        THEN
            'Good'
        ELSE 'Bad'
    END AS sales_performance
FROM
    sales_amazon
GROUP BY product_line
ORDER BY product_line DESC;
```

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

```
SELECT branch, SUM(quantity) AS Total_Quantity
FROM sales_amazon
GROUP BY branch
HAVING SUM(quantity) > (SELECT AVG(total_quantity)
    FROM (SELECT SUM(quantity) AS total_quantity
        FROM sales_amazon
        GROUP BY branch) AS subquery);
#'A' branch as the exceeded avg no of product sold '1859'
```

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

```
select gender, product_line, count(*) as frequency
from sales_amazon
group by product_line, gender
order by frequency desc
;
```

14.Calculate the average rating for each product line

```
select avg (rating) as avg_rating , product_line
from sales_amazon
group by product_line
order by avg_rating desc;
```

```
# the avg rating for each product_line '7.113218', 'Food and beverages'7.029213',
'Fashion accessories'7.003289',
#'Electronic accessories'6.916265', 'Sports and travel'6.837500', 'Home and lifestyle'
'Health and beauty'6.924706',
```

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

```
select day name(datee) as week_day, time_of_day as time_off_day, count(*) as total_
sales
from sales_amazon
group by week_day, time_of_day
order by field(week_day ,
'monday','tuesday','wednesday','thursday','friday','saturday','sunday'),timeofday;
```

16.Identify the customer type contributing the highest revenue.

```
select sum(total) as highest_sales, customer_type
from sales_amazon
group by Customer_type
order by highest_sales desc
limit 1;
#the highest revenue contributing by "member" with 164223.4440'
```

17. Determine the city with the highest VAT percentage

```
SELECT city,  
       round( AVG(tax/ total * 100),2) AS Average_ VAT_ Percentage  
FROM sales_ amazon  
GROUP BY city  
ORDER BY Average_ VAT_ Percentage DESC  
limit 1;  
# 'Yangon', as highest tax '4.76'
```

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

```
select customer_ type, sum(tax) highest_ tax  
from sales_ amazon  
group by Customer_ type  
order by highest_ tax desc ;  
# 'Member', as highest tax '7820.1640'  
# 'Normal', lowest tax '7559.2050'
```

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

```
select count(distinct customer_ type), customer_ type as distinct_ customers  
from sales_ amazon  
group by Customer_ type;  
# The distinct customer type is 2 'MEMBER','NORMAL'
```

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

```
select count(distinct payment) as payment_ type, Payment  
from sales_ amazon  
group by payment;  
# '1' - 'Cash', '2' - 'Credit_card', '3' - 'Ewallet'
```

21. Which customer type occurs most frequently?

```
select customer_ type, count(*) as frequency
```

```
from sales_ amazon
group by Customer_ type
order by frequency desc
limit 1;
#'Member' type occurs most frequently '501'
```

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

```
SELECT customer_ type, sum(total) as total,
       COUNT(*) AS Purchase_ Frequency
FROM sales_ amazon
GROUP BY customer_ type
ORDER BY Purchase_ Frequency DESC
LIMIT 1;
#'Member' type as highest revenue '164223.4440' with most frequently '501'
```

23. Determine the predominant gender among customers.

```
select gender, count(*) as frequency , customer_ type
from sales_ amazon
group by gender, Customer_ type
order by frequency desc ;
#'Female'-'261'-'Member','Male'-'259'-'Normal','Male'-'240'-'Member','Female'-'240'-'Normal'
```

24. Examine the distribution of genders within each branch.

```
select gender, branch, count(*) as gender_ count
from sales_ amazon
group by Gender, branch
order by gender_ count desc;
#'Male'-'A'-'179'
#'Female'-'C'-'178'
#'Male'-'B'-'170'
#'Female'-'B'-'162'
#'Female'-'A'-'161'
#'Male'-'C'-'150'
```



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

```
select timeofday, count(rating) as ratingss
from sales_amazon
group by timeofday
order by ratingss desc;
# 'Afternoon', '528' 'Evening', '281' 'Morning', '191'
# in afternoon we can see that most ratings happen with 528
```

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

```
select timeofday, branch, avg(rating) as avg_ratings
from sales_amazon
group by timeofday, branch
order by avg_ratings desc;
```

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

```
SELECT DAYNAME(datee) AS Weekday,
       AVG(rating) AS Average_Rating
FROM sales_amazon
GROUP BY Weekday
ORDER BY Average_Rating DESC
LIMIT 1;
# 'Monday' as highest avg rating with '7.153600'
```

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

```
select dayname(datee) as weekday, avg(rating) as avg_rating, branch
from sales_amazon
group by branch, weekday
order by branch, avg_rating desc;
```

# Thank You,

Presented By,

Ajay Kumar  
Student\_id-S9693