AMAZON SALES DATA ANALYSIS USING SQL

“KEY INSIGHTS AND TRENDS FROM SALES DATA”

HARIPRIYA K

PROJECT AIM AND DATA OVERVIEW

Project Aim:
To analyze Amazon sales data and identify factors influencing sales across three branches.

About the Data:

- Sales transactions are from three branches: A, B, and C, located in the cities of Mandalay, Yangon, and Naypyitaw.
- The dataset includes 17 columns and 1000 rows detailing transactions.
- Customer types include Member and Normal, with data on both male and female customers.
- Key features include product quantity, VAT (Tax), revenue, cost of goods, and gross income.

APPROACH USED

This project follows a three-step approach: Data Wrangling to clean and prepare the data, Feature Engineering to enhance data with new tables, and Exploratory Data Analysis (EDA) to uncover trends and answer key business questions.

DATA WRANGLING

FEATURE ENGINEERING

EXPLORATORY DATA ANALYSIS (EDA)

OVERVIEW OF PROCESS

- **Data Wrangling:**

Ensure data quality by detecting and handling NULL or missing values using constraints like NOT NULL during table creation.

- **Feature Engineering:**

Create new columns like timeofday (sales by time of day), dayname (weekly trends), and monthname (monthly performance).

- **Exploratory Data Analysis (EDA):**

Analyze the data to answer key questions and extract actionable insights.

ANALYSIS LIST

Analyze the amazon dataset to derive meaningful insights.

- Product Analysis
- Sales Analysis
- Customer Analysis

1.Product Analysis

•Understand the performance of different product lines.

•Identify the best-performing product lines.

Highlight product lines that need improvement.

ANALYSIS LIST

2.Sales Analysis

- Analyze sales trends across various products.
- Measure the effectiveness of current sales strategies.
- Recommend modifications to boost sales.

3.Customer Analysis

- Uncover the different customer segments.
- Analyze purchase trends across segments
- Evaluate the profitability of each customer segment.

INSIGHTS AND ANALYSIS

1.Count of distinct city

Select count(distinct City) as Distinct_City_Count
from amazontable;

Distinct_City_Count
3

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

Select Branch,City from amazontable
group by Branch,City;

Branch	City
A	Yangon
C	Naypyitaw
B	Mandalay

INSIGHTS AND ANALYSIS

3.count of distinct product lines in the dataset?

```
select count(distinct Product_line) as
distinct_productline from amazontable;
```

distinct_productline
6

4.Which payment method occurs most frequently?

```
SELECT payment as payment_method, COUNT(*) AS
times_of_payment
FROM amazontable
GROUP BY payment_method
ORDER BY times_of_payment DESC limit 1;
```

payment_method	times_of_payment
Ewallet	345

INSIGHTS AND ANALYSIS

5.Which product line has the highest sales?

```
select product_line, sum(Quantity) as total_sales
from amazontable
group by product_line
order by total_sales  desc limit 1;
```

product_line	total_sales
Electronic accessories	971

6.How much revenue is generated each month?

```
select monthname as Month, sum(Total) as
Total_revenue
from amazontable
group by Month
order by FIELD(Month, 'Jan', 'Feb', 'Mar', 'Apr', 'May');
```

Month	Total_revenue
Jan	116291.86800000005
Feb	97219.37399999997
Mar	109455.50700000004

INSIGHTS AND ANALYSIS

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

```
select monthname as Month, sum(cogs) as
Cost_of_goods
from amazontable
group by Month
order by Cost_of_goods Desc limit 1;
```

Month	Cost_of_goods
Jan	110754.16000000002

8.Which product line generated the highest revenue?

```
select Product_line, sum(Total) as tot_revenue
from amazontable
group by Product_line
order by tot_revenue desc
limit 1;
```

Product_line	tot_revenue
Food and beverages	56144.844000000005

INSIGHTS AND ANALYSIS

9.In which city was the highest revenue recorded?

```
Select City, SUM(Total) as total_revenue
from amazontable
group by City
Order by total_revenue desc limit 1;
```

City	total_revenue
Naypyitaw	110568.70649999994

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

```
Select Product_line, SUM(Tax_5percent) as tax
from amazontable
group by Product_line
ORDER BY tax desc ;
```

Product_line	tax
Food and beverages	2673.5639999999994
Sports and travel	2624.8964999999994
Electronic accessories	2587.5015000000017
Fashion accessories	2585.995
Home and lifestyle	2564.853000000002
Health and beauty	2342.5589999999993

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 amazontable GROUP BY Product_line) AS overall_sum)
        THEN 'Good' ELSE 'Bad'
    END AS Sales_performance
FROM amazontable
GROUP BY Product_line
order by total_sales desc;
```

Product_line	total_sales	Sales_performance
Food and beverages	56144.844000000005	Good
Sports and travel	55122.826499999996	Good
Electronic accessories	54337.531500000005	Good
Fashion accessories	54305.895	Good
Home and lifestyle	53861.91300000001	Good
Health and beauty	49193.739000000016	Bad

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

Select distinct Branch , SUM(Quantity) as total_quantity
from amazontable group by Branch
Having SUM(Quantity) > (SELECT AVG(Quantity) from amazontable);

Branch	total_quantity
A	1859
C	1831
B	1820

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

SELECT Product_line, Gender, Count(*) as frequency from amazontable
group by Product_line, Gender
ORDER BY Gender, frequency DESC;

Product_line	Gender	frequen
Fashion accessories	Female	96
Food and beverages	Female	90
Sports and travel	Female	88
Electronic accessories	Female	84
Home and lifestyle	Female	79
Health and beauty	Female	64
Health and beauty	Male	88
Electronic accessories	Male	86
Food and beverages	Male	84
Fashion accessories	Male	82

14.Calculate the average rating for each product line.

```
SELECT Product_line, Avg(Rating) as average_rating
from amazontable
group by Product_line;
```

Product_line	average_rating
Health and beauty	7.003289473684212
Electronic accessories	6.92470588235294
Home and lifestyle	6.8375
Sports and travel	6.916265060240964
Food and beverages	7.113218390804598
Fashion accessories	7.029213483146067

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

```
Select dayname, timeofday, count(*) occurrence
from amazontable
where dayname in('Mon','Tue','Wed','Thu','Fri')
group by dayname, timeofday
order by occurrence asc ;
```

dayname	timeofday	occurrence
Mon	Morning	21
Wed	Morning	22
Fri	Morning	29
Thu	Morning	33
Tue	Morning	36
Mon	Afternoon	48
Thu	Afternoon	49
Fri	Evening	52
Tue	Afternoon	53
Mon	Evening	56

To find which time of day the sales is at the peak

```
WITH sales_time as (Select dayname, timeofday ,count(*) as
occurrence
from amazontable
where dayname in('Mon','Tue','Wed','Thu','Fri')
group by dayname, timeofday
order by timeofday,occurrence desc )

SELECT  timeofday , sum(occurrence) as tot_occurrence from
sales_time
group by timeofday
order by tot_occurrence desc;
```

timeofday	tot_occurrence
Evening	293
Afternoon	269
Morning	141

16.Identify the customer type contributing the highest revenue.

```
select `Customer type` , sum(Total) as Total_revenue
from amazontable
group  by `Customer type`
order by Total_revenue desc;
```

Customer type	Total_revenue
Member	164223.44400000002
Normal	158743.30500000005

17.Determine the city with the highest VAT percentage.

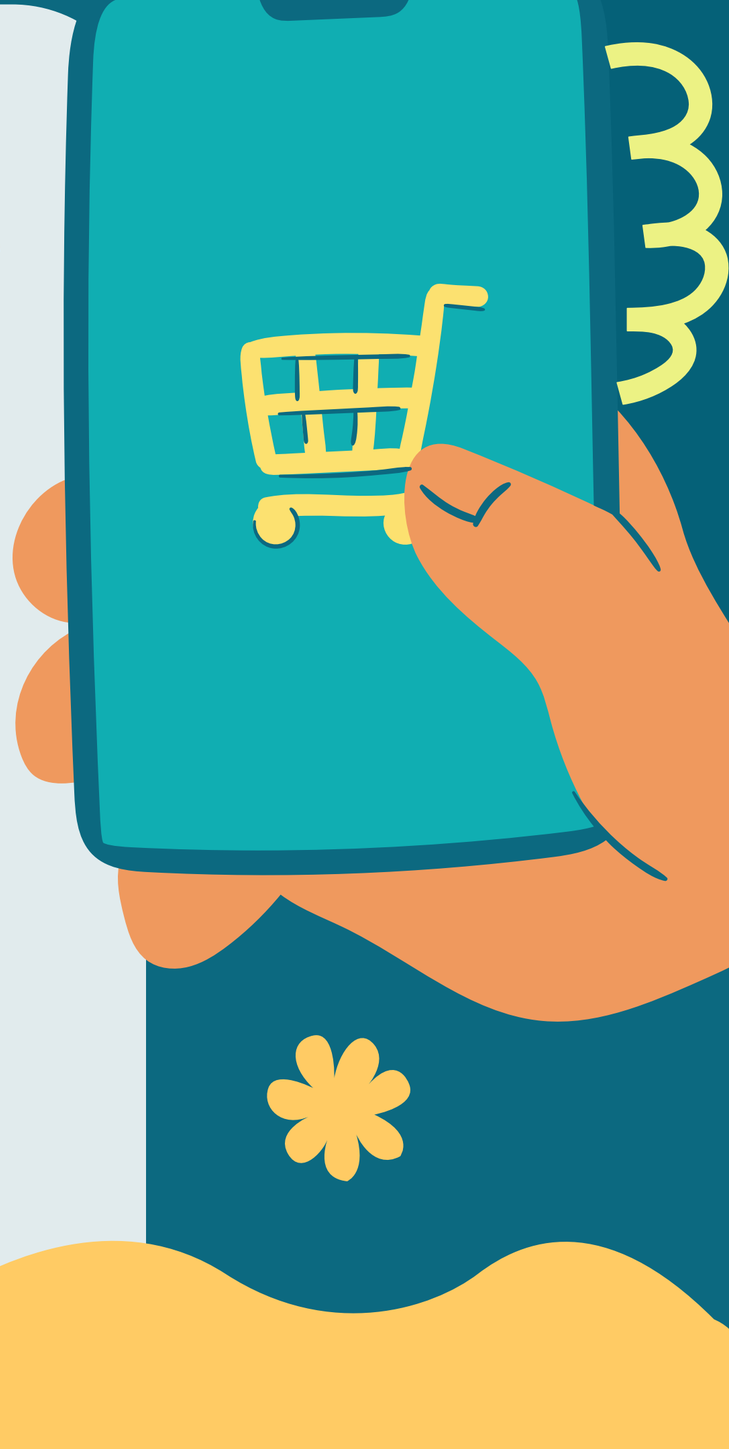
```
select City, MAX(Tax_5percent) as VAT
from amazontable
group by City
order by VAT desc;
```

City	VAT
Naypyitaw	49.65
Yangon	49.49
Mandalay	48.69

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

```
SELECT `Customer type` , SUM(Tax_5percent) as VAT
from amazontable
group by `Customer type`
order by VAT desc;
```

Customer type	VAT
Member	7820.164000000002
Normal	7559.205000000003



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

```
select count(distinct `Customer type`) as distinct_customer_types
from amazontable;
```

distinct_customer_types
2

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

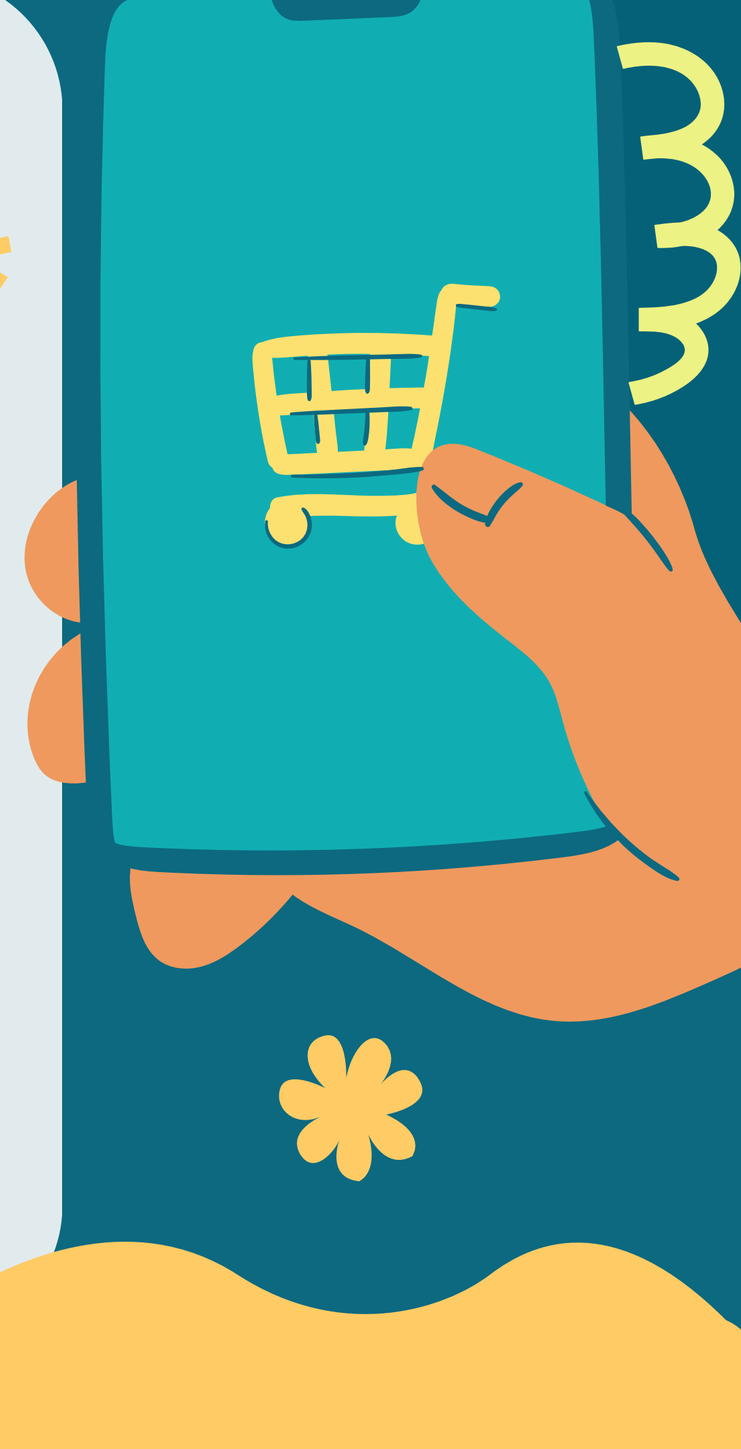
```
select count(distinct `payment`) as distinct_payment_methods
from amazontable;
```

distinct_payment_methods
3

21.Which customer type occurs most frequently?

```
Select `Customer type` , count(*) as Frequency
from amazontable
group by `Customer type`
order by frequency desc;
```

Customer type	Frequency
Member	501
Normal	499



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

```
Select `Customer type` , SUM(Quantity) as Highest_Purchase_frequency
from amazontable
group by `Customer type`
order by Highest_Purchase_frequency desc limit 1;
```

Customer type	Highest_Purchase_frequency
Member	2785

23.Determine the predominant gender among customers.

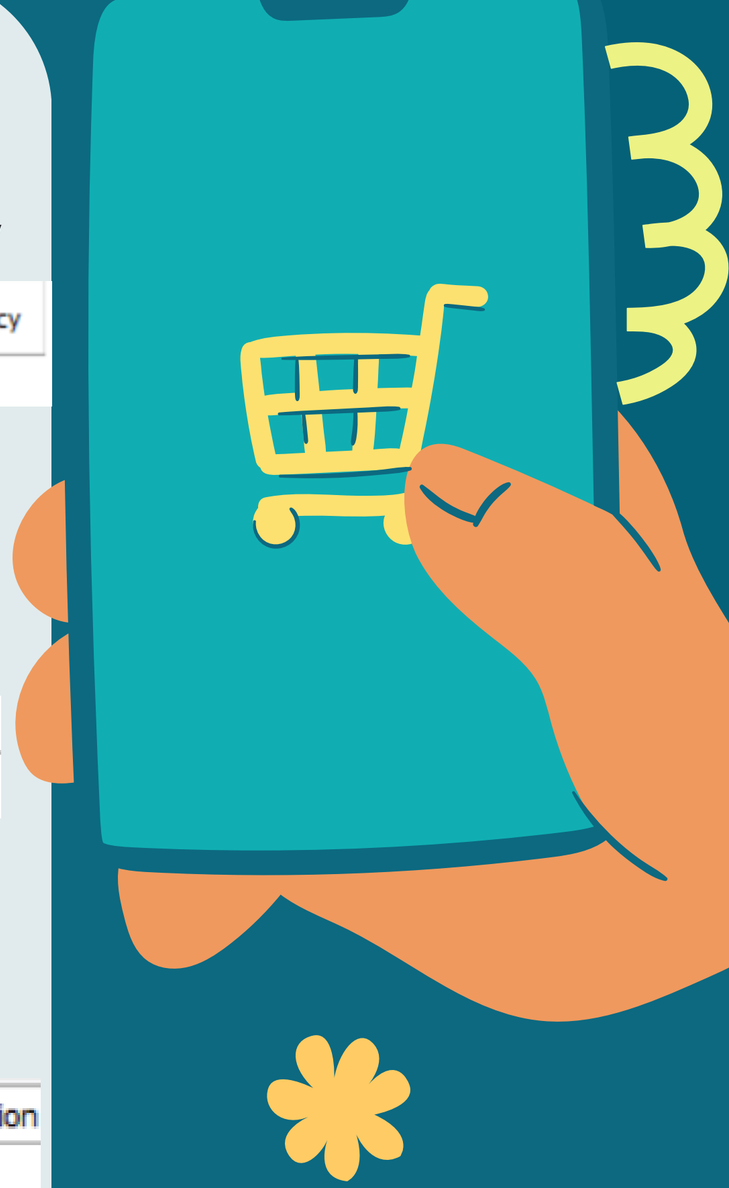
```
Select Gender, count(*) as Predominant_Gender
from amazontable
group by Gender
order by Predominant_Gender desc limit 1;
```

Gender	Predominant_Gender
Female	501

24.Examine the distribution of genders within each branch.

```
Select Gender, Branch, count(*) as Gender_distribution
from amazontable
group by Gender , Branch
order by Gender, Branch asc;
```

Gender	Branch	Gender_distribution
Female	A	161
Female	B	162
Female	C	178
Male	A	179
Male	B	170
Male	C	150



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

```
Select timeofday, COUNT(Rating) AS Most_rating
from amazontable
group by timeofday
order by Most_Rating desc limit 1;
```

timeofday	Most_rating
Evening	432

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

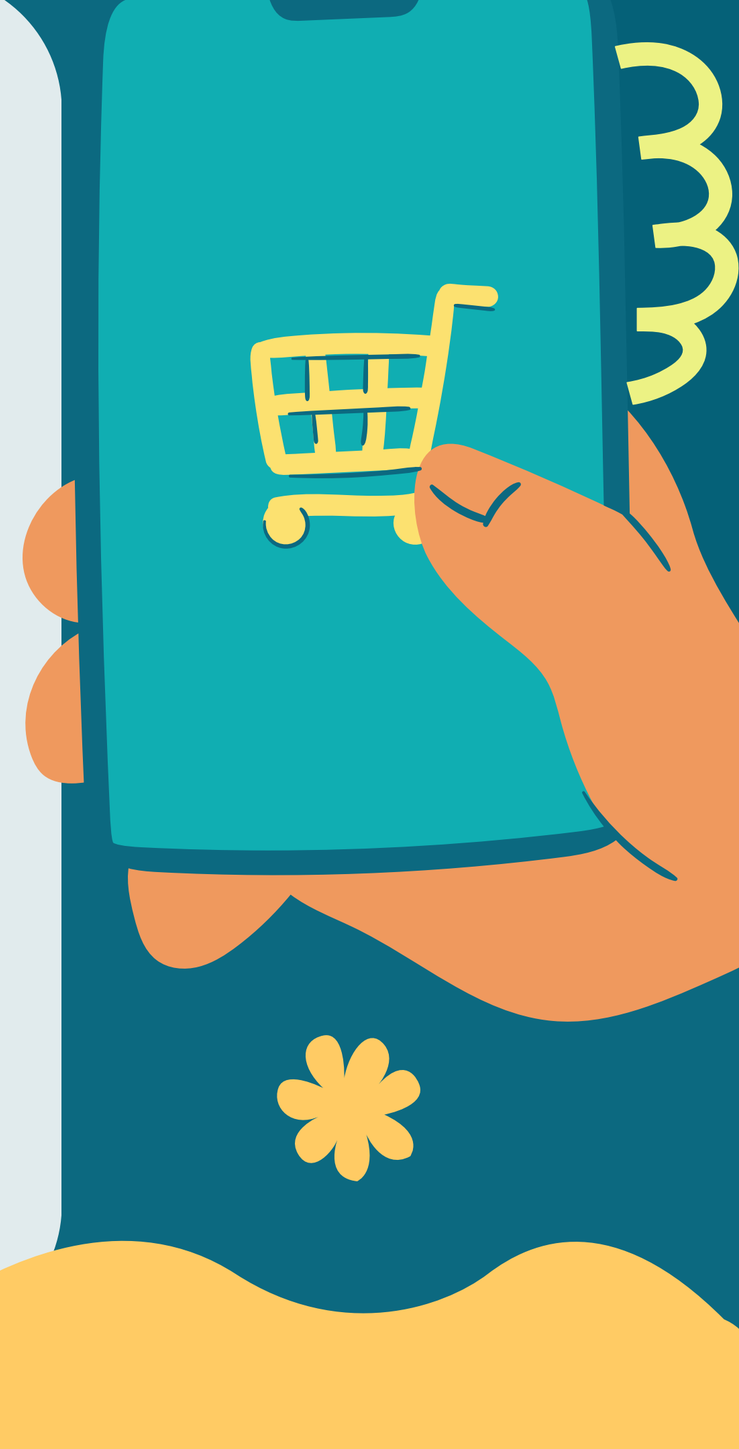
```
SELECT Branch, TimeOfDay, MAX(Rating)
 AS highest_rating FROM amazontable
GROUP BY Branch, TimeOfDay
ORDER BY Branch, highest_rating DESC;
```

Branch	TimeOfDay	highest_rating
A	Evening	10
A	Afternoon	9.9
A	Morning	9.9
B	Afternoon	10
B	Evening	10
B	Morning	10
C	Evening	10
C	Morning	9.9
C	Afternoon	9.9

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

```
Select dayname, AVG(Rating) AS highest_avg_rating
from amazontable
group by dayname
order by highest_avg_Rating desc limit 1;
```

dayname	highest_avg_rating
Mon	7.153599999999999



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

```sql
WITH AvgBranchratings AS(
SELECT Branch, dayname, AVG(Rating) as Avg_rating
from amazontable
group by Branch, dayname),

MaxBranchRatings AS (
    SELECT
        Branch,
        MAX(Avg_rating) AS max_avg_rating
    FROM AvgBranchratings
    GROUP BY Branch
)
SELECT a.Branch, a.dayname, a.avg_rating
FROM AvgBranchratings a
INNER JOIN MaxBranchRatings b
ON a.Branch = b.Branch  AND a.Avg_rating = b.max_avg_rating
ORDER BY a.Branch;
```

Branch	dayname	avg_rating
A	Fri	7.3119999999999985
B	Mon	7.335897435897434
C	Fri	7.278947368421051

PRODUCT LINE ANALYSIS

- There are 6 distinct product lines.
- Electronic Accessories recorded the highest sales with a quantity of 941 sold.
- Food and Beverages generated the highest revenue of 56,144.
- Food and Beverages also recorded the highest Value Added Tax (VAT) of 2,673.
- Food and beverages receiving an highest average rating of 7.11, Home and lifestyle receieved an lowest average rating of 6.8.
- For each product line, a "Good" or "Bad" performance indicator was added based on sales:
- Health and Beauty was categorized as "Bad" due to below-average sales performance.
- All other product lines were categorized as "Good" based on their sales being above average.

SALES ANALYSIS

- Electronic Accessories had the highest sales with 971 units sold
- Health and Beauty had the lowest with 854 units sold.
- Food and Beverages generated the highest revenue at 56144
- Health and Beauty recorded the lowest revenue.
- Naypyitaw achieved the highest revenue of 110568, while Mandalay had the least.
- All product lines, except Health and Beauty, showed a "Good" performance, with above-average sales.
- Branch A led with the highest sales, outperforming other branches.
- Sales peaked in the Evening, while Morning sales were the lowest.

CUSTOMER ANALYSIS

- eWallet was the most frequent payment method, while Credit Card was the least used.
- Females purchased Fashion Accessories more frequently, while Males preferred Health and Beauty.
- The average rating was highest for Food and Beverages (7.11), and lowest for Home and Lifestyle (6.8).
- Customers predominantly made purchases in the Evening, with Monday having the highest number of ratings.
- Member customers contributed higher revenue and more frequent purchases compared to Normal customers.
- Female customers were the majority.
- Branch A recorded the highest sales, with Health and Beauty having the lowest sales.
- Evening time showed the best sales performance, while Morning had the least sales.

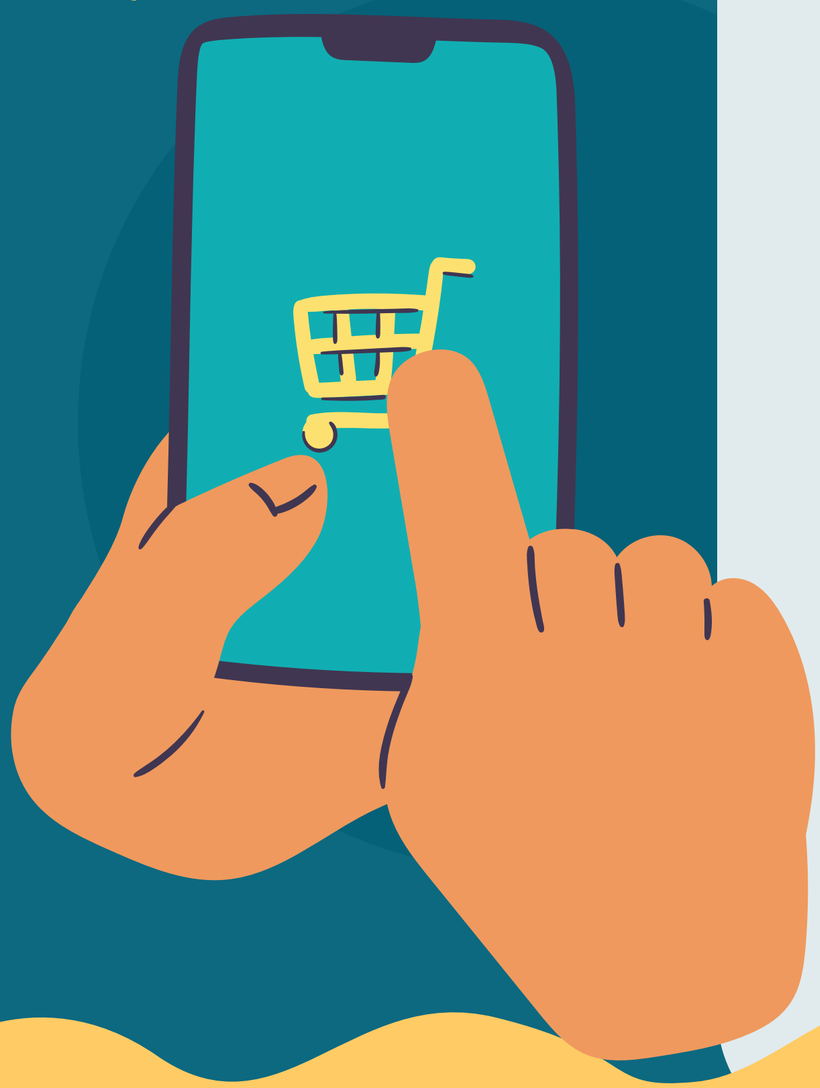
RECOMMENDATION

1. Since most sales occur in the Evening, launch targeted promotions or discounts during this time to increase sales volume.

2. Implement Different programs and exclusive offers to encourage Normal customers to become Members, who generate higher revenue and purchase more frequently.

3. Since eWallet is the most popular payment method, consider offering incentives or discounts for customers who use eWallet, encouraging even higher usage

4. Since the rating is very least in Home and lifestyle line,Enhance the customer experience by introducing personalized recommendations, or targeted marketing campaigns to improve customer satisfaction.



STRATEGIES TO ENHANCE SALES AND CUSTOMER ENGAGEMENT

- PERSONALIZED OFFERS
- HIGHLIGHT BEST-SELLING
- IMPROVED CUSTOMER SUPPORT
- SEASONAL CAMPAIGNS
- LEVERAGING DIGITAL PROMOTIONS

THANK YOU