amazon

SALES ANALYSIS

by Saswati Sadhu

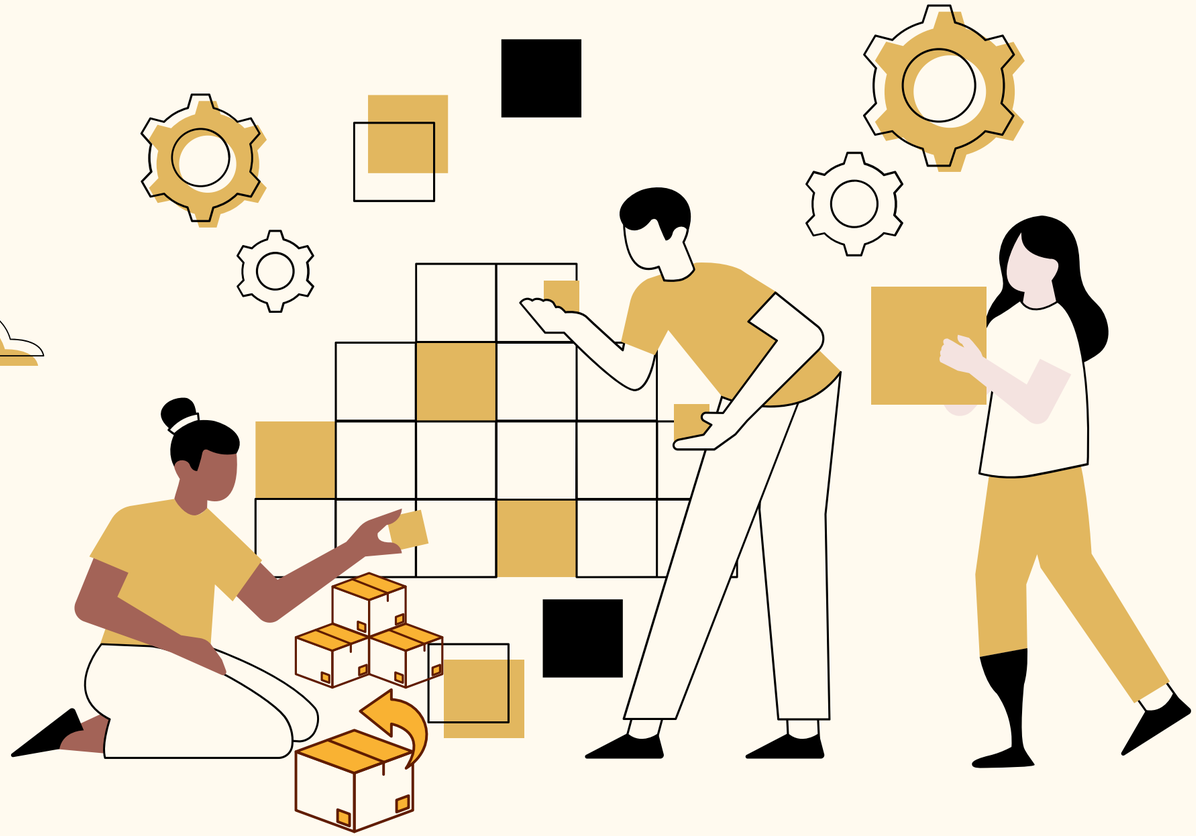


Table of content

This is the material point that will be delivered in the presentation.

01 Project Overview

02 Database Overview

03 Project workflow

04 Feature Engineering

05 EDA

06 Analysis

07 Insight Summary

08 Action & Recommendation

Objective

- **Analyze sales trends across branches and time periods**
- **Understand customer segments and behaviors**
- **Evaluate product performance and profitability**
- **Identify top revenue-generating factors**
- **Create insights using SQL and feature engineering**
- **Solve real-world business questions with data**

Database Overview

amazon_sales

- Invoice_ID VARCHAR(50)
- Branch ENUM('A', 'B', 'C')
- City VARCHAR(50)
- Customer_type VARCHAR(50)
- Gender VARCHAR(50)
- Product_line VARCHAR(200)
- Unit_price DECIMAL(10,0)
- Quantity INT
- Tax_5 DECIMAL(10,0)
- Total DECIMAL(10,0)
- Date DATE
- Time TIME
- Payment VARCHAR(50)
- cogs DECIMAL(10,0)
- gross_margin_percentage DECIMAL(10,...
- gross_income DECIMAL(10,0)
- Rating DECIMAL(10,0)
- timeofday VARCHAR(30)
- dayname VARCHAR(5)
- monthname VARCHAR(5)

Contains 1000 Rows

Contains 17 Columns

PROJECT WORKFLOW

- Data Wrangling
- Feature Engineering
- Exploratory Data Analysis

Data Wrangling

- **Database Setup:** Created database and inserted clean sales data
- **Schema Design:** Applied NOT NULL constraints to prevent missing values
- **Data Validation:** Verified no nulls exist due to enforced constraints
- **Data Ready:** Ensured structured, clean data for analysis

```
create database amazon_sales;
create table amazon_sales(
Invoice_ID  varchar (50) primary key not null,
Branch  enum('A','B','C') not null,
City    varchar(50) not null,
Customer_type   varchar(50) not null,
Gender  varchar(50) not null,
Product_line    varchar(200) not null,
Unit_price  decimal not null,
Quantity    int not null,
Tax_5       decimal not null,
Total     decimal not null,
Date    date,
Time    time,
Payment   varchar(50) not null,
cogs    decimal not null,
gross_margin_percentage    decimal not null,
gross_income    decimal not null,
Rating     decimal not null
);
```

FEATURE ENGINEERING

timeofday : Categorized sales into Morning, Afternoon, and Evening to analyze peak sales hours

dayname : Extracted weekday names to identify the busiest days for each branch

monthname : Extracted month names to evaluate monthly sales and profit trends

```sql
-- Add a new column named timeofday
alter table amazon_sales
add column timeofday varchar(30);
Update amazon_sales
set timeofday = case
   when hour(time)>=0 and hour(time) <12 then 'Morning'
   when hour(time)>=12 and hour(time) <17 then 'Afternoon'
   else 'Evening'
end ;
--  Add a new column named dayname
alter table amazon_sales
add column dayname varchar (5);
update amazon_sales
set dayname = date_format(date,'%a');
-- Add a new column named monthname
alter table amazon_sales
add column monthname varchar(5);
update amazon_sales
set monthname = date_format(date,'%b');
```

EXPLORATORAY DATA ANALYSIS

Let’s start the Analysis

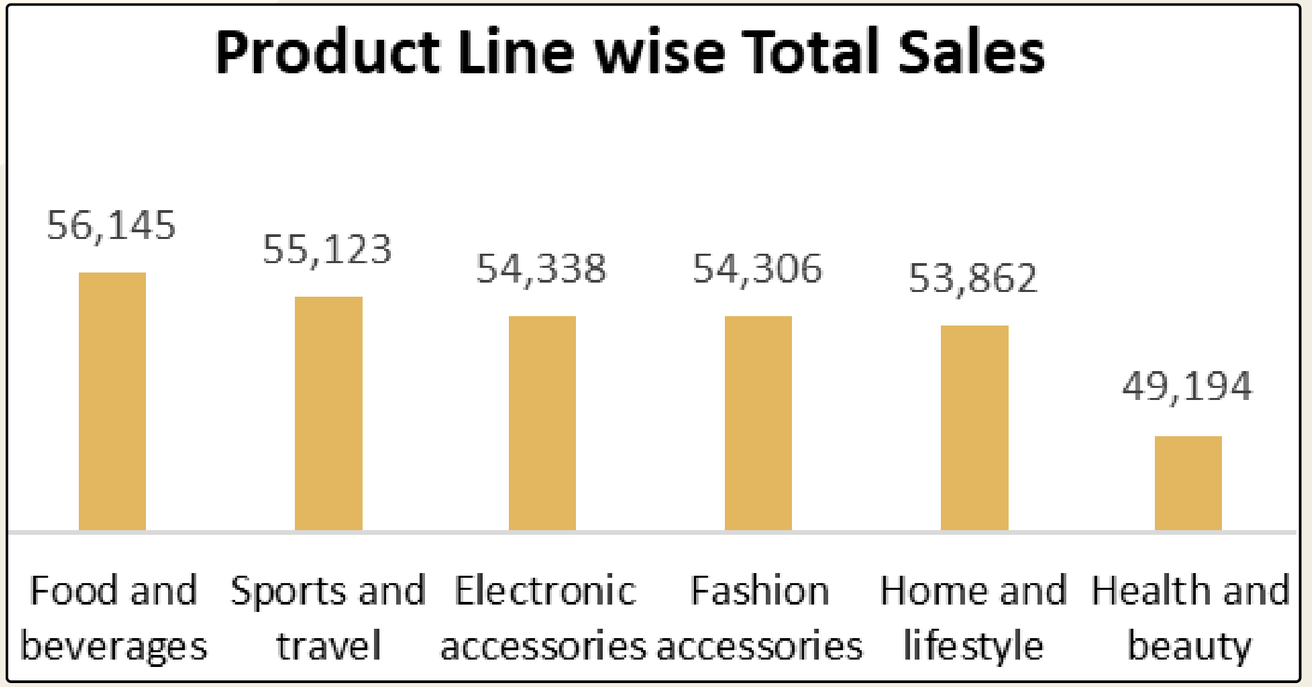
Next Slide

Product Analysis

Which product line has the highest sales?

Food and Beverages leads with the highest sales while **Health and Beauty** has the lowest

```sql
select product_line, count(*) as prodctline_sold_num
from amazon_sales
group by product_line
order by count(*) desc
limit 1;
```

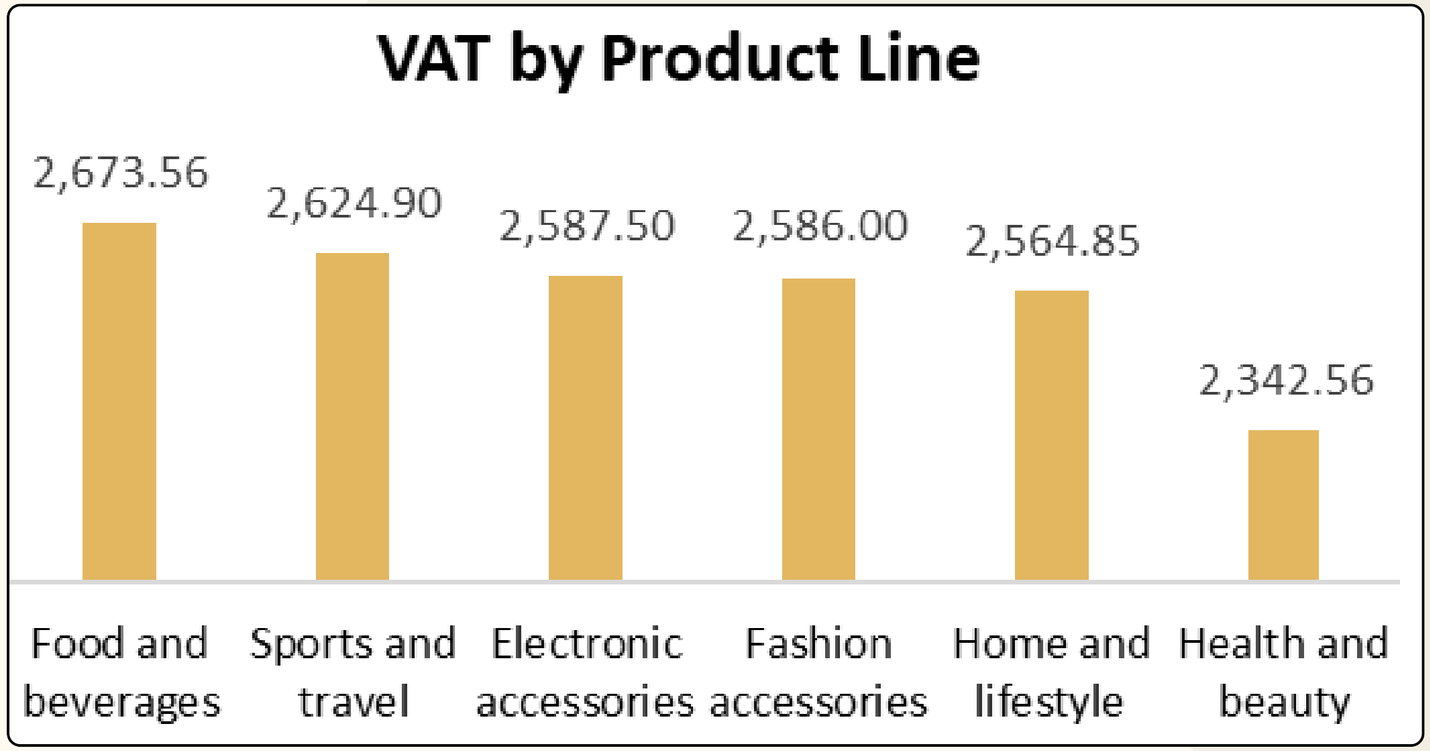


Product Analysis

Which product line incurred the highest Value Added Tax?

Food and Beverages has incurred the highest VAT while **Health and Beauty** has the lowest

```sql
select product_line, sum(Tax_5) as VAT
from amazon_sales
group by product_line
order by sum(Tax_5) desc
limit 1;
```

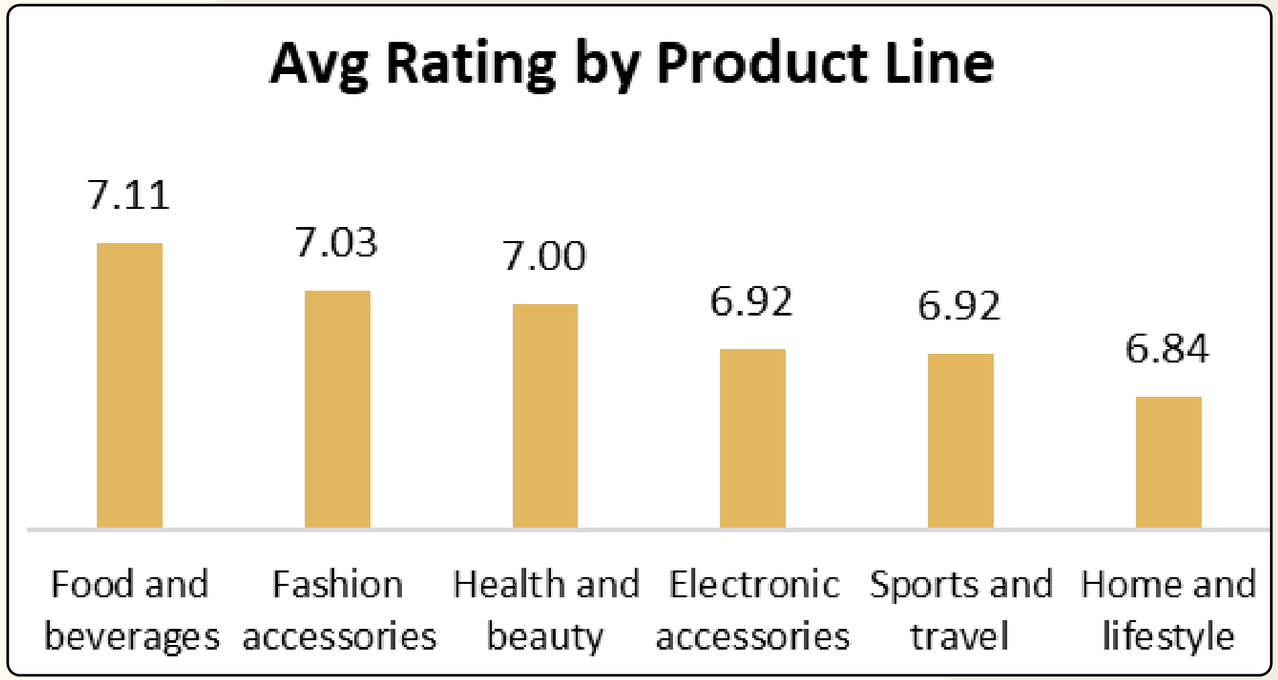


Product Analysis

Calculate the average rating for each product line

Food and Beverages leads with the highest sales while **Health and Beauty** has the lowest

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



Product Analysis

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

Only **Health and Beauty** has total sales below average sales rest all product lines falling under ‘Good’ category

```sql
with cte as (
select product_line, sum(Total) as total
from amazon_sales
group by Product_line),
avg_sale as (
select avg(total) as avg_total
from cte)
select cte.product_line , cte.total,avg_sale.avg_total,
case
   when avg_sale.avg_total< cte.total  then 'Good'
   else 'Bad'
   end as category
from cte,avg_sale ;
```

product_line	total	avg_total	category
Food and beverages	56153	53828.3333	Good
Health and beauty	49190	53828.3333	Bad
Sports and travel	55125	53828.3333	Good
Fashion accessories	54298	53828.3333	Good
Home and lifestyle	53866	53828.3333	Good
Electronic accessories	54338	53828.3333	Good

Product Analysis

Which product line is most frequently associated with each gender?

Female customers are frequently associated with **Fashion accessories,** on the other hand **Male customers** are associated with **Health and Beauty**

```sql
WITH ranked_product_lines AS (
    SELECT
        gender,
        product_line,
        COUNT(*) AS product_count,
        ROW_NUMBER() OVER (PARTITION BY gender ORDER BY COUNT(*) DESC) AS rn
    FROM amazon_sales
    GROUP BY gender, product_line
)
SELECT gender, product_line, product_count
FROM ranked_product_lines
WHERE rn = 1;
```

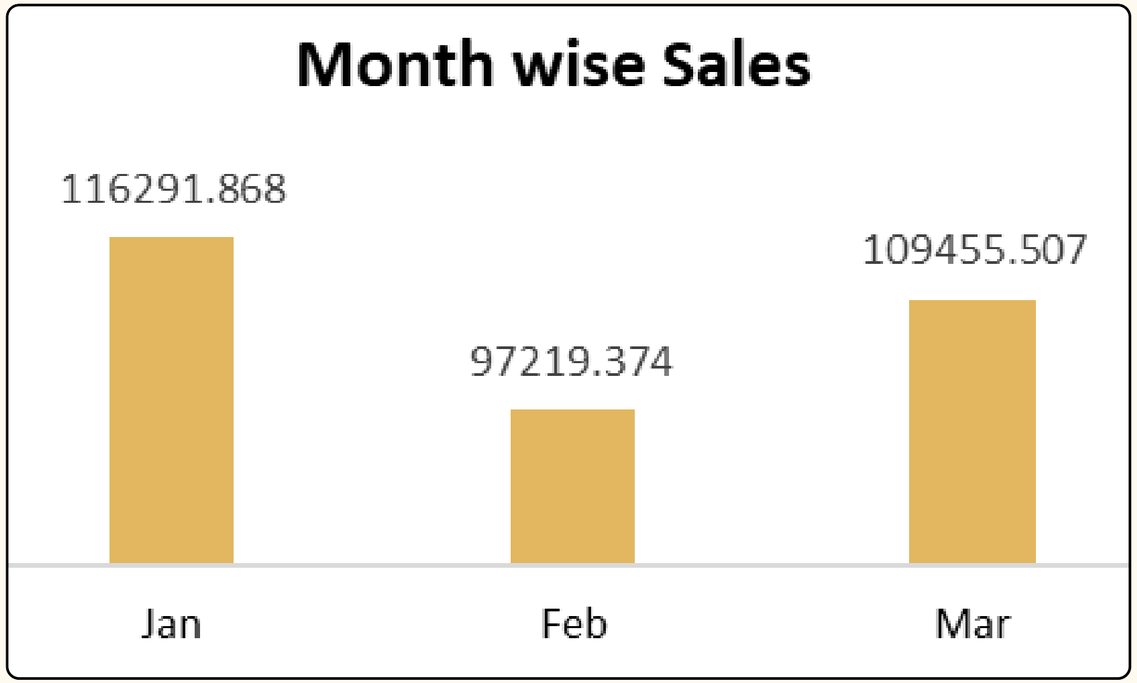
gender	product_line	product_count
Female	Fashion accessories	96
Male	Health and beauty	88

Sales Analysis

How much revenue is generated each month?

January has the maximum sales followed by **March** , major sales drop in **February**

```
select monthname , sum(total) as revenue
from amazon_sales
group by monthname
order by  sum(total) desc;
```

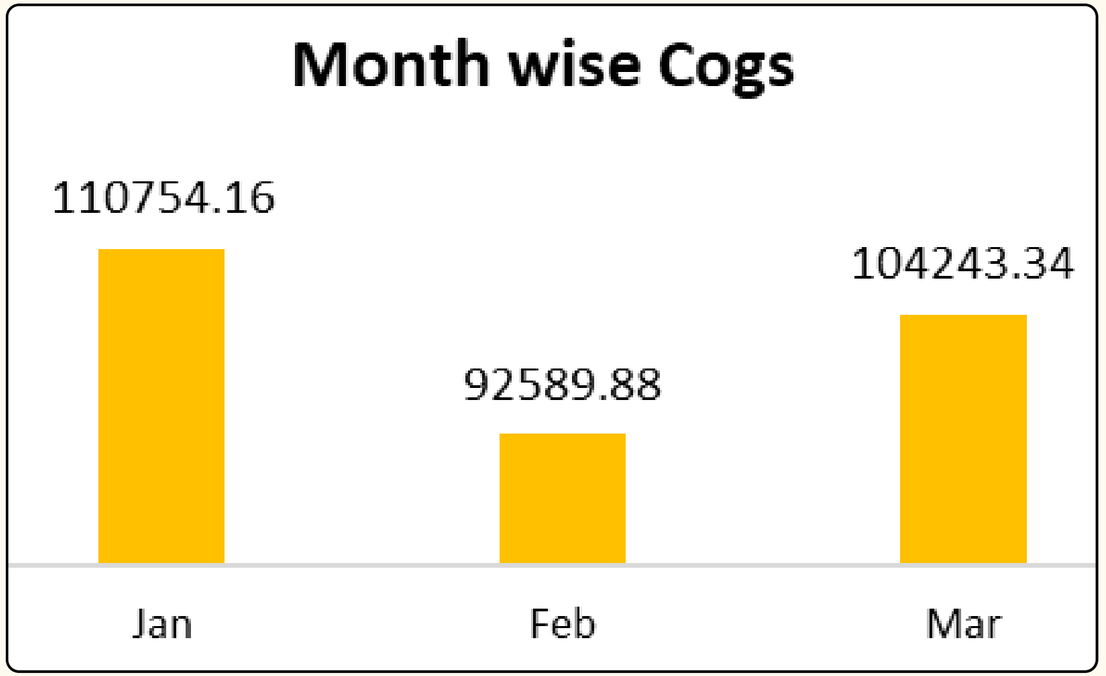


Sales Analysis

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

January has the maximum Cost of Goods Sold followed by **March ,** major Cost of Goods Sold drop in **February**

```sql
select product_line, count(*) as prodctline_sold_num
from amazon_sales
group by product_line
order by count(*) desc
limit 1;
```

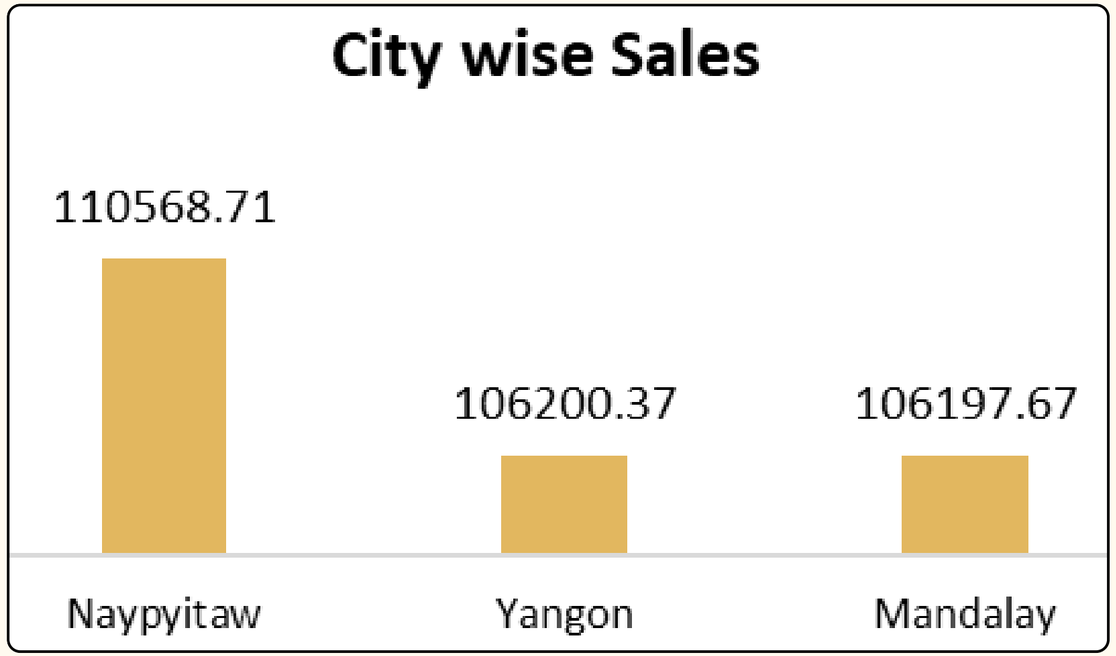


Sales Analysis

In which city was the highest revenue recorded?

Naypyitaw leads with the highest sales, followed by **Yangon** while **Mandalay** has the lowest sales

```
select city , sum(Total) as revenue
from amazon_sales
group by city
order by sum(Total) desc;
```

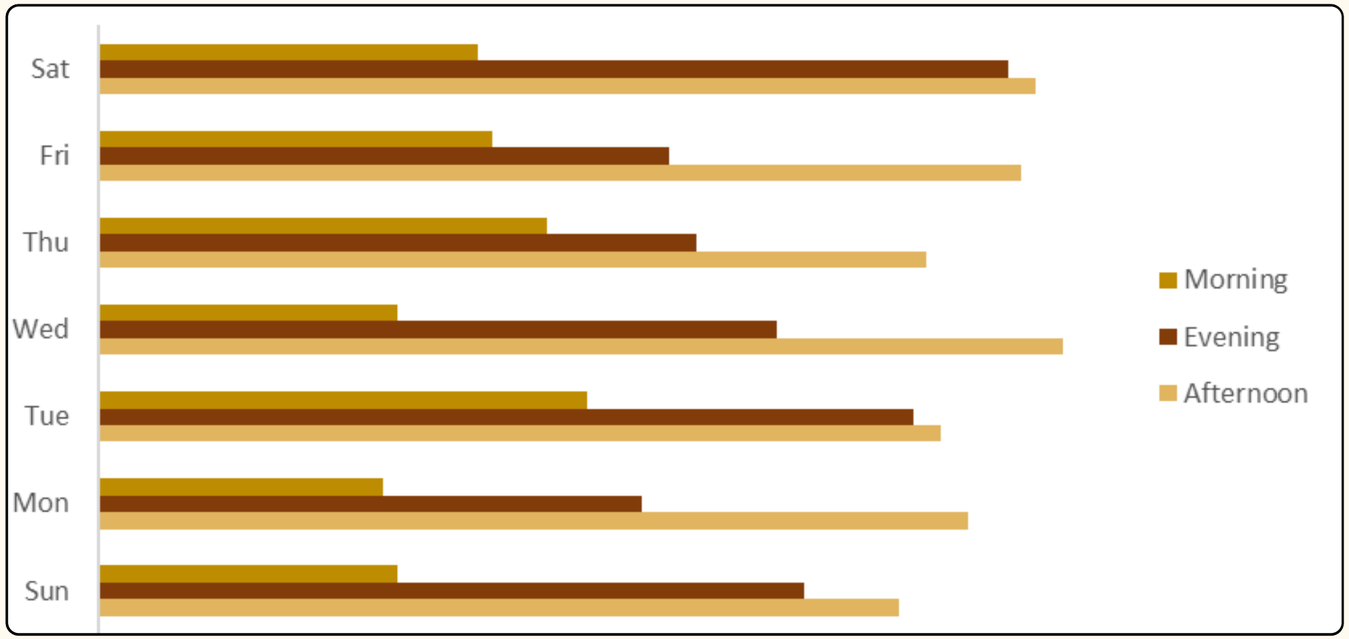


Sales Analysis

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

Afternoon consistently sees the **highest sales** activity across all weekdays, while **Morning** has the **lowest,** indicating customer preference peaks later in the day

```
select dayname,
timeofday,count(*) as sales_count
from amazon_sales
group by dayname, timeofday
order by dayname,timeofday;
```



Sales Analysis

Identify the branch that exceeded the average number of products sold

Branch A has exceeded average number of products sold

```
with cte as (
select branch , sum(Quantity) as total_quantity
from amazon_sales
group by branch),
avg_quantity as (
select avg(total_quantity) as avg_quantity
from cte)
select cte.branch , cte.total_quantity
from cte,avg_quantity
where avg_quantity.avg_quantity< cte.total_quantity;
```

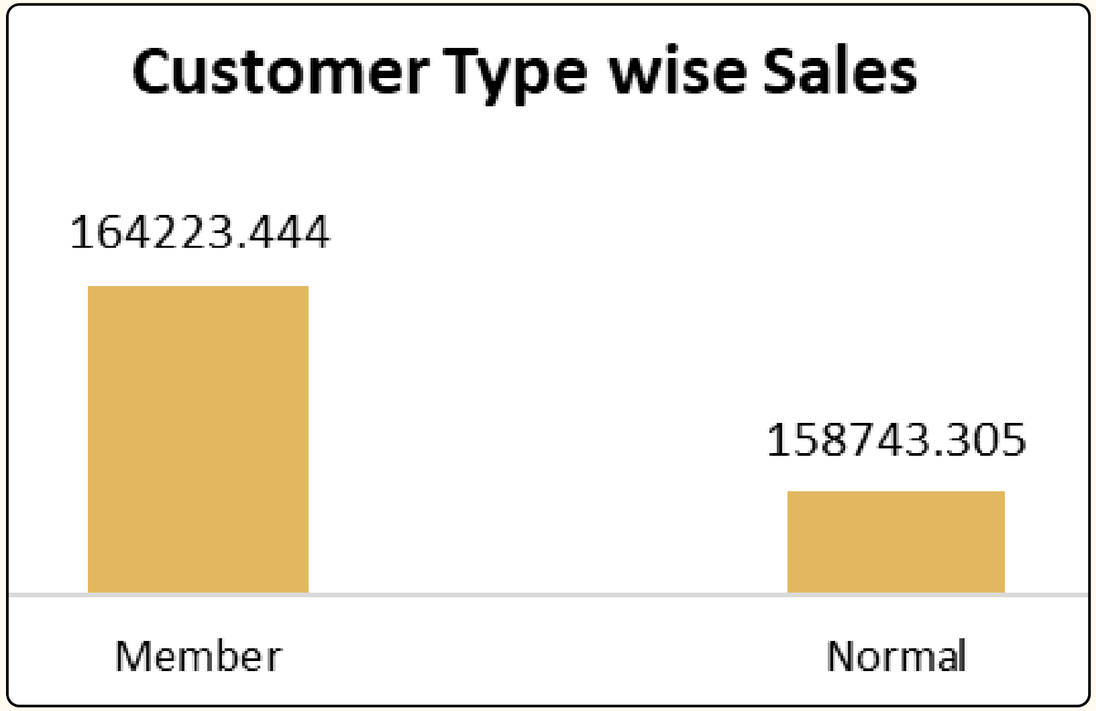
branch	total_quantity
A	1859

Customer Analysis

Identify the customer type contributing the highest revenue

```
select customer_type,
sum(total) as revenue
from amazon_sales
group by Customer_type;
```

Member Customer type leads with the highest sales while **Normal customer type** has the lowest

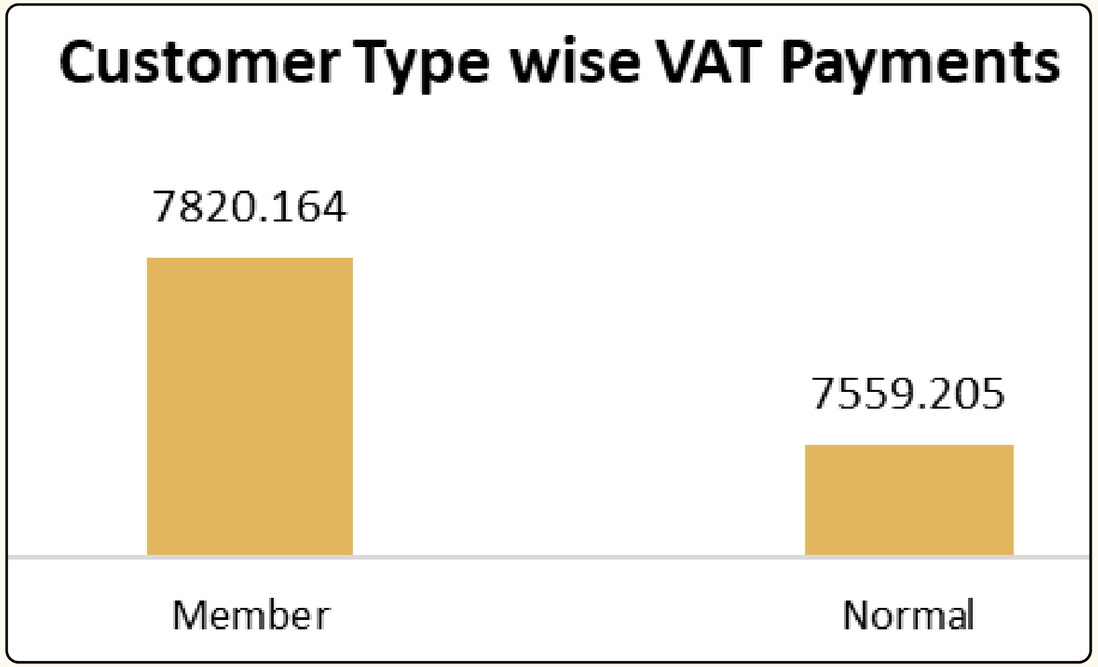


Customer Analysis

Identify the customer type with the highest VAT payments

Member Customer type leads with the highest VAT Payments while **Normal customer type** has paid lowest VAT

```
select customer_type , sum(tax_5) as vat
from amazon_sales
group by Customer_type
order by sum(tax_5) desc;
```

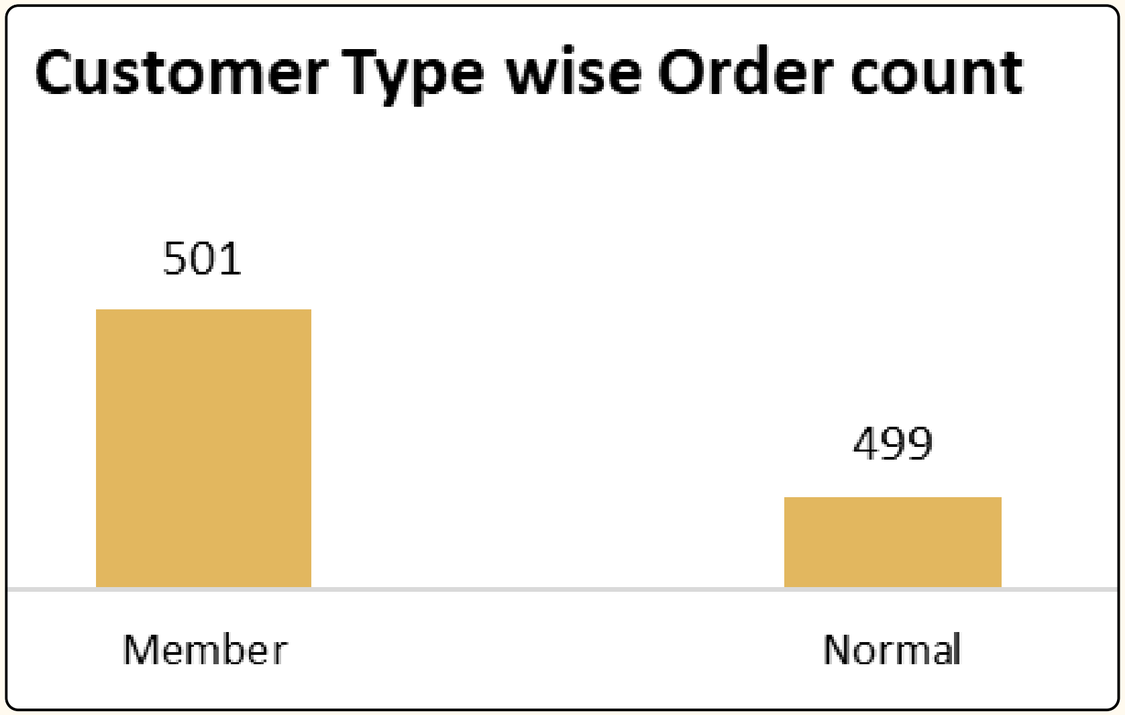


Customer Analysis

Which customer type occurs most frequently?

Member Customer type occurs most frequently while **Normal customer type** has lowest order count

```sql
select customer_type, count(*) as order_count
from amazon_sales
group by customer_type
order by count(*) desc
limit 1;
```

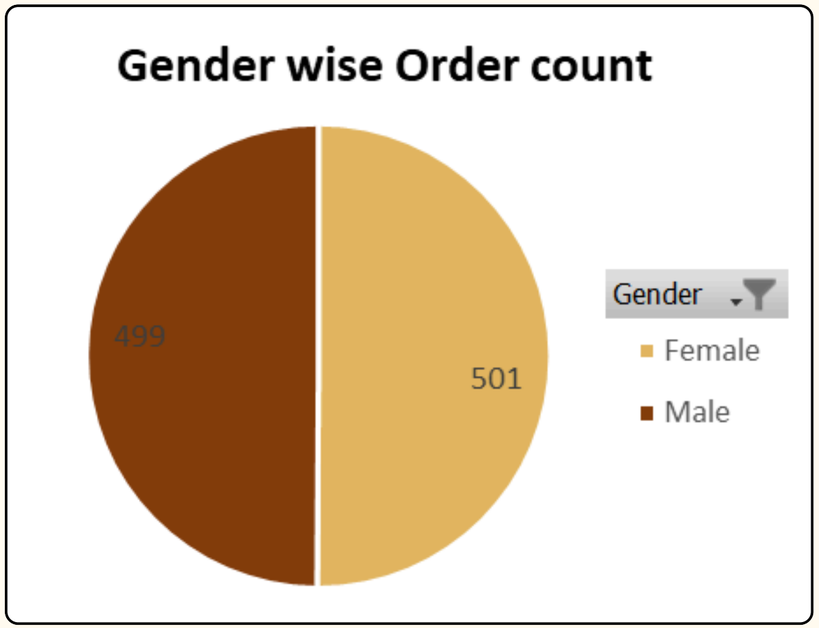


Customer Analysis

Determine the predominant gender among customers

Female customers are dominating with highest order count over **Male customers**

```sql
select gender, count(*) as order_count
from amazon_sales
group by Gender
order by count(*) desc
limit 1;
```

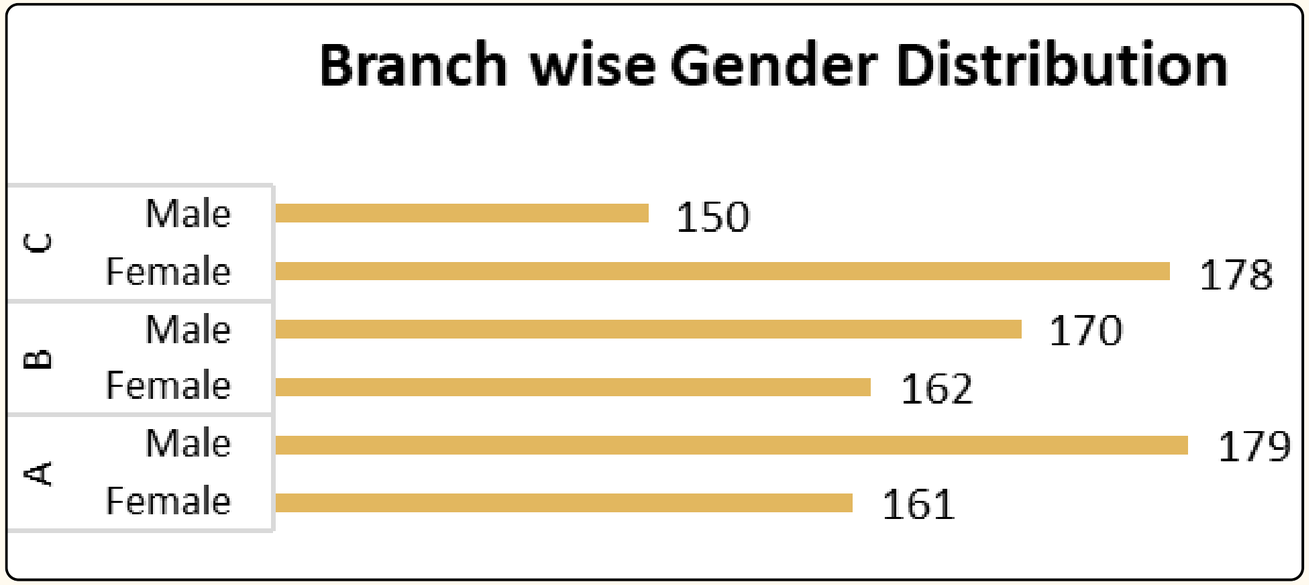


Customer Analysis

Examine the distribution of genders within each branch

```sql
select branch, Gender,
count(*) as gender_distribution
from amazon_sales
group by branch, Gender
order by branch, Gender;
```

Branch-wise gender distribution shows more **female customers** than male in all branches**,** with **Branch A** having the highest **male count** and **Branch C** the lowest **male count**



Business Insight

Top Seller

Food & Beverages has the highest sales and VAT

Needs Improvement

Health & Beauty underperforms in sales and VAT

Peak Sales Month

January, followed by March; drop in February

City Performance

Naypyitaw leads in revenue; Mandalay is lowest

Time Trend

Afternoon has the highest sales; Morning the lowest

Customer Type

Members drive most sales, VAT, and order frequency

Gender Insights

Females order more; males prefer Health & Beauty

Branch Performance

Branch A exceeds average product sales

Recommendation

- Boost Food & Beverages with stock, ads, and bundles.
- Revamp Health & Beauty via targeted offers and promos.
- Focus Sales in Jan & Mar, plan campaigns to balance Feb dip.
- Use Gender-Based Targeting for product marketing.
- Run Promotions in Afternoon to match sales peak time.
- Promote Memberships to convert Normal to Member customers.

Thank You

Tools used : MySQL & MS Excel