

SQL – Capstone Project

(AMAZON SALES ANALYSIS)



Name -- Sumit Baviskar

Date -- 8 March 2024

Introduction:

The primary objective of this project is to analyze sales data from three branches of Amazon, located in Mandalay, Yangon, and Naypyitaw, to uncover key insights that can drive data-informed decisions for optimizing sales performance. By understanding the various factors influencing sales across these branches, the goal is to assist stakeholders in identifying trends, anomalies, and potential areas for improvement.

This dataset comprises 1,000 rows and 17 columns, capturing essential details about sales transactions, such as branch location, product categories, customer demographics, sales amounts, payment methods, and other relevant metrics. By thoroughly analyzing this data, the project aims to answer critical business questions, improve sales strategies, and optimize branch-specific performance.

Through exploratory data analysis, statistical modeling, and data visualization, we will identify patterns, correlations, and trends that impact sales. The insights drawn from this analysis will help Amazon's regional managers and decision-makers implement more effective strategies to boost revenue, enhance customer satisfaction, and streamline operations across all three branches.

Purposes Of This Capstone Project:

The purpose of this project is to provide Amazon with actionable insights into the sales performance of its three branches—Mandalay, Yangon, and Naypyitaw. By analyzing the sales data, the aim is to identify key factors that affect sales outcomes, such as product demand, customer preferences, seasonal trends, and payment methods. The insights gained from this analysis will help Amazon's management team make informed, data-driven decisions to:

- **Optimize sales strategies for each branch.**
- **Enhance customer satisfaction by understanding buying behavior.**
- **Identify high-performing and underperforming products.**
- **Improve operational efficiency through better resource allocation.**

Ultimately, this project aims to empower Amazon's decision-makers to implement targeted strategies that will drive sales growth, increase revenue, and ensure the long-term success of the business.

Program Used:

MySQL.

Understanding the 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 995 rows**:

Column	Description	Data Type
invoice_id	Invoice of the sales made	VARCHAR(30)
branch	Branch at which sales were made	VARCHAR(5)
city	The location of the branch	VARCHAR(30)
customer_type	The type of the customer	VARCHAR(30)
gender	Gender of the customer making purchase	VARCHAR(10)
product_line	Product line of the product sold	VARCHAR(100))
unit_price	The price of each product	DECIMAL(10, 2)
quantity	The amount of the product sold	INT
VAT	The amount of tax on the purchase	FLOAT(6, 4)
total	The total cost of the purchase	DECIMAL(10, 2)

date	The date on which the purchase was made	DATE
time	The time at which the purchase was made	TIME
payment_method	The total amount paid	VARCHAR(100)
cogs	Cost Of Goods sold	DECIMAL(10, 2)
gross_margin_percentage	Gross margin percentage	FLOAT(11, 9)
gross_income	Gross Income	DECIMAL(10, 2)
rating	Rating	FLOAT(2, 1)

Analysis List:

1. Product Analysis :

Conduct analysis on the data to understand the different product lines, the products lines performing best and the product lines that need to be improved.

2. Sales Analysis :

This analysis aims to answer the question of the sales trends of product. The result of this can help us measure the effectiveness of each sales strategy the business applies and what modifications are needed to gain more sales.

3. Customer Analysis :

This analysis aims to uncover the different customer segments, purchase trends and the profitability of each customer segment.

Approach Used For Project:

Data Wrangling :

Data wrangling is a crucial step in ensuring the quality and reliability of data for analysis. In this project, I focused on collecting, cleaning, and transforming raw data from multiple sources, addressing inconsistencies, missing values, and errors.

This process involved using MySQL for efficient manipulation of large datasets, allowing for smooth integration into the analytical pipeline. By structuring the data in a standardized format, I enabled seamless exploration, visualization, and extraction of valuable insights."

This is the first step where inspection of data is done to make sure NULL values and missing values are detected and data replacement methods are used to replace missing or NULL values.

Task 1: Build a database

Solution: Created a database named "Amazon_data".

Task 2: Create a table and insert the data

Solution: Create a new table name "amazon".

Task 3: Select columns with null values in them. There are no null values in our database as. in creating the tables, we set **NOT NULL** for each field, hence null values are filtered out.

Solution: checked for not null , there are **no null values** in our dataset.

Result Grid Filter Rows: Export: Wrap Ce						
	Field	Type	Null	Key	Default	Extra
►	invoice_id	varchar(50)	YES		NULL	
	branch	varchar(50)	YES		NULL	
	city	varchar(50)	YES		NULL	
	customer_type	varchar(50)	YES		NULL	
	gender	varchar(10)	YES		NULL	
	product_line	varchar(50)	YES		NULL	
	unit_price	decimal(10,2)	YES		NULL	
	quantity	int	YES		NULL	
	VAT	decimal(6,4)	YES		NULL	
	total	decimal(10,2)	YES		NULL	
	date	date	YES		NULL	
	time	time	YES		NULL	
	payment_met...	varchar(50)	YES		NULL	
	cogs	decimal(10,2)	YES		NULL	
	gross_margin...	decimal(11,9)	YES		NULL	
	gross_income	decimal(10,2)	YES		NULL	
	rating	decimal(2,1)	YES		NULL	

Result 6 v

Feature Engineering:

This will help us generate some new columns from existing ones. In this project, I implemented feature engineering techniques to derive new insights from transactional sales data using MySQL.

In this project, I performed feature engineering using MySQL to extract new insights from transactional data. I added a `time_of_day` column to classify sales into Morning, Afternoon, and Evening, helping identify peak sales times. A day name column was created using the `DAYNAME()` function to track which days of the week see the most activity. Additionally, I added a month name column using the `MONTHNAME()` function to analyze monthly sales trends. These features provided valuable insights into customer behavior and sales patterns across different time frames."

By adding key columns, I enhanced the dataset to answer critical business questions about sales trends.

Task 1 : Add a new column named **`time_of_day`** to give insight of sales in the Morning, Afternoon and Evening. This will help answer the question on which part of the day most sales are made.

Solution:

CODE:

Update Amazon

```
SET time_of_day=( CASE WHEN time BETWEEN "00:00:00" AND  
"12:00:00" THEN "Morning"  
WHEN time BETWEEN "12:00:01" AND "16:00:00" THEN "Afternoon"  
ELSE "Evening"  
END  
);
```


A new column, **time_of_day**, was created to categorize transactions into three time periods: Morning, Afternoon, and Evening. This feature was generated by analyzing the transaction timestamp and categorizing sales based on predefined time ranges. This addition provides insights into the part of the day where most sales are concentrated, allowing for more targeted operational strategies.

Task 2 : Add a new column named **dayname** that contains the extracted days of the week on which the given transaction took place (**Mon, Tue, Wed, Thur, Fri**).

Solution:

-- Using the dayname() function,

CODE:

Update Amazon

SET day_name=DAYNAME(date);

To understand which day of the week generates the highest sales, I extracted the day from each transaction and added a new column, dayname, using the built-in MySQL DAYNAME() function. This feature helped identify which days branches are busiest, offering a clearer view of weekly customer patterns."

Task 3 : Add a new column named **monthname** that contains the extracted months of the year on which the given transaction took place (Jan, Feb, Mar).

Solution:

-- using the monthname()

CODE:

Update Amazon

SET month_name=MONTHNAME(date);

A month_name column was added using the MySQL MONTHNAME() function to extract and display the month in which each transaction occurred. This feature enabled a comprehensive analysis of sales and profits across different months of the year, helping identify seasonal trends and peak performance periods.

After Data Engineering is Done :

Code :

Describe Amazon;

Output :

Result Grid

Filter Rows:

Export:

Wrap Cell

	Field	Type	Null	Key	Default	Extra
	city	varchar(50)	YES		NULL	
	customer_type	varchar(50)	YES		NULL	
	gender	varchar(10)	YES		NULL	
	product_line	varchar(50)	YES		NULL	
	unit_price	decimal(10,2)	YES		NULL	
	quantity	int	YES		NULL	
	VAT	decimal(6,4)	YES		NULL	
	total	decimal(10,2)	YES		NULL	
	date	date	YES		NULL	
	time	time	YES		NULL	
	payment_met...	varchar(50)	YES		NULL	
	cogs	decimal(10,2)	YES		NULL	
	gross_margin...	decimal(11,9)	YES		NULL	
	gross_income	decimal(10,2)	YES		NULL	
	rating	decimal(2,1)	YES		NULL	
	time_of_day	varchar(30)	YES		NULL	
	day_name	varchar(30)	YES		NULL	
	month_name	varchar(30)	YES		NULL	

Result 17

×

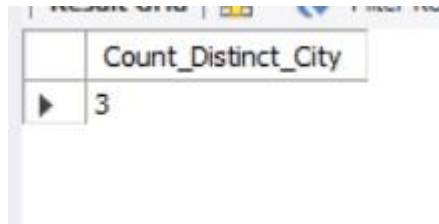
Business Questions

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

CODE:

```
SELECT COUNT(DISTINCT(city)) as Count_Distinct_City  
FROM Amazon;
```

OUTPUT:



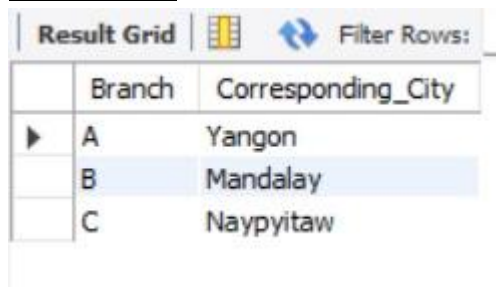
Count_Distinct_City
3

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

CODE:

```
SELECT Branch, City as Corresponding_City from Amazon  
GROUP BY Branch, City  
ORDER BY Branch, City;
```

OUTPUT:



Branch	Corresponding_City
A	Yangon
B	Mandalay
C	Naypyitaw

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

Solution:

CODE:

```
SELECT COUNT(DISTINCT (product_line)) as Distinct_Product_Line_Count  
FROM amazon;
```

OUTPUT:



The screenshot shows a database interface with a 'Result Grid' tab. The grid has two columns: 'Distinct_Product_Line_Count' and a value '6'. There is a 'Filter Rows' button and a small icon of a grid.

	Distinct_Product_Line_Count
▶	6

There are 6 distinct product line count (Health and Beauty, Electronics accessories, Sport and travel, Home and lifestyle, Food and Beverages, Fashion accessories)

4. Which payment method occurs most frequently?

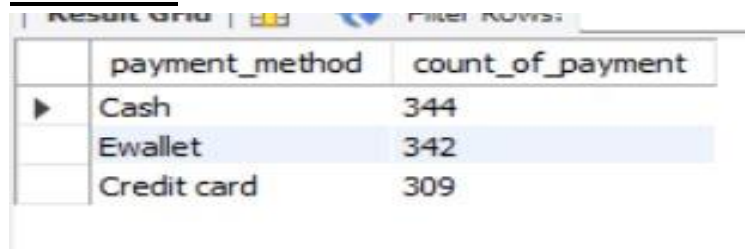
Solution:

Most payment method is Cash followed by E-wallet , Credit cash.

CODE:

```
SELECT payment_method,count(payment_method) as count_of_payment  
FROM Amazon  
GROUP BY payment_method  
ORDER BY count(payment_method) DESC;
```

OUTPUT:



The screenshot shows a database interface with a 'Result Grid' tab. The grid has two columns: 'payment_method' and 'count_of_payment'. The rows are: 'Cash' with count 344, 'Ewallet' with count 342, and 'Credit card' with count 309. There is a 'Filter Rows' button and a small icon of a grid.

	payment_method	count_of_payment
▶	Cash	344
	Ewallet	342
	Credit card	309

5. Which product line has the highest sales?

Solution:

Highest sales is done by Food and Beverages followed by Fashion accessories, Sport and travel, Home and lifestyle, Electronics accessories, Health and Beauty.

CODE:

```
SELECT Product_line,SUM(cogs) as Sales
FROM Amazon
GROUP BY Product_line
ORDER BY SUM(cogs) DESC;
```

OUTPUT:



Product_line	Sales
Food and beverages	53471.28
Fashion accessories	51719.90
Sports and travel	51367.74
Home and lifestyle	51297.06
Electronic accessories	51222.13
Health and beauty	46527.98

6. How much revenue is generated each month?

Solution :

CODE:

```
SELECT Month_Name,SUM(unit_price*quantity) as Revenue
FROM Amazon
GROUP BY Month_Name
ORDER BY Revenue DESC;
```

OUTPUT:

Month_Name	Revenue
January	110754.16
March	103683.00
February	91168.93

Highest revenue is earned in January Followed by March and February.

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

Solution:

January is the highest revenue month followed by March and February.
This analysis can be get by summing cogs(Cost Of Goods sold) by grouping the month.

CODE:

```
SELECT Month_Name,SUM(cogs) as total_Cogs
FROM Amazon
GROUP BY Month_Name
ORDER BY SUM(cogs) DESC;
```

OUTPUT:

Month_Name	total_Cogs
January	110754.16
March	103683.00
February	91168.93

8. Which product line generated the highest revenue?

Solution:

We have calculate summed the gross income and cogs and get revenue and order by Revenue. As Food and beverages has generate the highest revenue followed by fashion accessories, Sport & travel, home & life style, Electronics & accessories and Health & beauty.

CODE:

```
SELECT Product_Line,SUM(unit_price*quantity) as Revenue
FROM Amazon
GROUP BY Product_Line
ORDER BY Revenue DESC;
```

OUTPUT:

Product_Line	Revenue
Food and beverages	53471.28
Fashion accessories	51719.90
Sports and travel	51367.74
Home and lifestyle	51297.06
Electronic accessories	51222.13
Health and beauty	46527.98

9. In which city was the highest revenue recorded?**Solution:**

As mentioned above, we calculate revenue by adding gross income and cost of sold goods has arranged in descending to get city with highest revenue which is **Naypyitaw** Followed by **Yangon, Mandalay**.

CODE:

```
SELECT City,sum(unit_price*quantity) as Revenue
FROM amazon
GROUP BY City
ORDER BY Revenue DESC;
```

OUTPUT:

City	Revenue
Naypyitaw	105229.31
Yangon	100820.01
Mandalay	99556.77

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

CODE :

```
SELECT Product_line , round(SUM(VAT),2) as Value_Added_Tax
FROM amazon
GROUP BY Product_Line
ORDER BY Value_Added_Tax DESC;
```

OUTPUT:

Product_Line	Revenue
Food and beverages	53471.28
Fashion accessories	51719.90
Sports and travel	51367.74
Home and lifestyle	51297.06
Electronic accessories	51222.13
Health and beauty	46527.98

As the vat is given which is grouped by product line get us insight that the Food & beverages Followed by accessories Sport & travel, Home & travel Electronics accessories, and Health & beauty.

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

Solution:

CODE:

```
WITH gender_1 AS (
  SELECT
    Product_Line,
    Gender,
    COUNT(Invoice_ID) AS Invoice_count,
    Row_number() over(partition by Gender ORDER BY COUNT(Invoice_ID)
DESC) AS Gender_freq
```



```
FROM
    amazon
GROUP BY
    Gender,
    Product_Line )
```

```
SELECT *
FROM gender_1
where Gender_freq=1;
```

OUTPUT:

Result Grid Filter Rows: Export: Wrap Cell Content:				
	Product_Line	Gender	Invoice_count	Gender_freq
▶	Fashion accessories	Female	96	1
	Health and beauty	Male	88	1

As there is different preferences to product line for **Female** which is Sport & travel, Fashion accessories ,Home & lifestyle , Food & beverages, Electronics accessories and Health & beauty.

Male had preferences to Sport & travel similar to women followed by Electronics accessories, Food & beverages, Health & beauty, Fashion accessories and Home & lifestyle.

12.Calculate the average rating for each product line.

CODE:

```
SELECT Product_line,round(AVG(rating),2) as AVG_Rating
FROM amazon
GROUP BY Product_Line
ORDER BY AVG(rating) DESC;
```

OUTPUT:

Product_line	AVG_Rating
Food and beverages	7.11
Fashion accessories	7.03
Health and beauty	6.91
Electronic accessories	6.91
Sports and travel	6.86
Home and lifestyle	6.84

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

CODE:

```
SELECT day_name, time_of_day, COUNT(cogs) as sales_occurrence
FROM amazon
GROUP BY day_name, time_of_day
ORDER BY day_name, sales_occurrence DESC;
```

OUTPUT:

day_name	time_of_day	sales_occurrence
Friday	Afternoon	58
Friday	Evening	51
Friday	Morning	29
Monday	Evening	56
Monday	Afternoon	48
Monday	Morning	20
Saturday	Evening	81
Saturday	Afternoon	55
Saturday	Morning	28
Sunday	Evening	58
Sunday	Afternoon	52
Sunday	Morning	22
Thursday	Evening	56
Thursday	Afternoon	49
Thursday	Morning	33
Tuesday	Evening	69
Tuesday	Afternoon	53
Tuesday	Morning	36

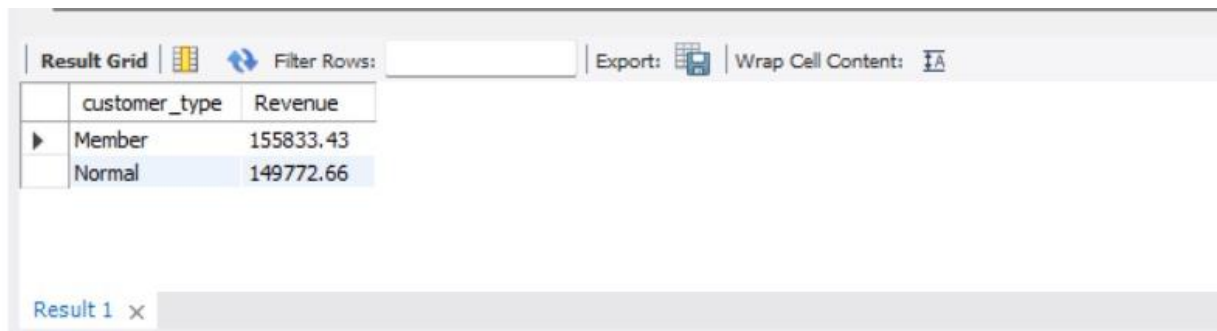
14. Identify the customer type contributing the highest revenue.

Solution:

CODE:

```
SELECT customer_type, SUM(unit_price*quantity) AS Revenue
FROM amazon
GROUP BY customer_type
ORDER BY Revenue DESC;
```

OUTPUT:



The screenshot shows a database query result grid. At the top, there are tabs for 'Result Grid', 'Filter Rows', 'Export', and 'Wrap Cell Content'. Below the tabs is a table with two columns: 'customer_type' and 'Revenue'. The table contains two rows: 'Member' with a revenue of 155833.43, and 'Normal' with a revenue of 149772.66. The 'Normal' row is highlighted. At the bottom left, there is a tab labeled 'Result 1' with a close button 'x'.

customer_type	Revenue
Member	155833.43
Normal	149772.66

15. Determine the city with the highest VAT percentage.

Solution:

As the cities are arranged in the highest VAT are Naypyitaw after that Yangon and Mandalay. Yangon and Mandalay had the similar percentage but the value of Yangon is higher than Mandalay.

CODE:

```
SELECT DISTINCT city, round(SUM(VAT) over(Partition by city),2) as
VAT_total, round((SUM(VAT) over(Partition by city)/SUM(VAT) over())*100,2)
AS VAT_percentage
FROM amazon
ORDER BY VAT_percentage DESC;
```

OUTPUT:

city	VAT_total	VAT_percentage
Naypyitaw	5261.47	34.43
Yangon	5041.00	32.99
Mandalay	4977.84	32.58

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

Solution:

There is a small difference between only 2 customer type which is normal and member type. Member has slightly higher value which is around 300 more than normal customer type VAT value.

CODE:

```
SELECT customer_type, sum(VAT) AS TOTAL_VAT
FROM amazon
GROUP BY customer_type
ORDER BY TOTAL_VAT DESC;
```

OUTPUT:

customer_type	TOTAL_VAT
Member	7791.6715
Normal	7488.6330

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

Solution:

There are only 2 customer type Member and Normal customer type.

CODE:

```
SELECT customer_type, COUNT(DISTINCT(customer_type)) AS
CUSTOMER_TYPE
FROM amazon
GROUP BY customer_type;
```

OUTPUT:

customer_type	COUNT_CUSTOMER_TYPE
Member	499
Normal	496

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

CODE:

```
SELECT COUNT(DISTINCT(payment_method)) AS PAYMENT_METHOD  
FROM amazon;
```

OUTPUT:

PAYMENT_METHOD
3

There are 3 types of payment mode such as Cash,Ewallet,Credit Card

19. Which customer type occurs most frequently?

CODE:

```
SELECT customer_type,  
COUNT(customer_type) AS COUNT_CUSTOMER_TYPE  
FROM amazon  
GROUP BY customer_type  
ORDER By COUNT_CUSTOMER_TYPE DESC;
```

OUTPUT:

customer_type	COUNT_CUSTOMER_TYPE
Member	499
Normal	496

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

Solution:

CODE:

```
SELECT Customer_type, COUNT(invoice_id) as purchse_frequency
FROM AMAZON
GROUP BY customer_type
ORDER BY purchse_frequency DESC;
```

OUTPUT:

Customer_type	purchse_frequency
Member	499
Normal	496

21. Determine the predominant gender among customers.

CODE:

```
SELECT gender, customer_type, COUNT(invoice_id) as id_count
FROM AMAZON
GROUP BY gender, customer_type
ORDER BY customer_type, id_count DESC;
```

OUTPUT:

gender	customer_type	id_count
Female	Member	259
Male	Member	240
Male	Normal	258
Female	Normal	238

22.Examine the distribution of genders within each branch.

CODE:

```
SELECT branch, gender, COUNT(gender) AS GENDER_SUM  
FROM amazon  
GROUP BY branch, gender  
ORDER BY BRANCH, GENDER_SUM DESC;
```

OUTPUT:

BRANCH	GENDER	GENDER_SUM
A	Male	179
A	Female	160
B	Male	169
B	Female	160
C	Female	177
C	Male	150

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

Solution:

CODE:

```
SELECT time_of_day,round(avg(rating),2) AS avg_rating  
FROM amazon  
GROUP BY time_of_day  
ORDER BY avg_rating DESC;
```

OUTPUT:

time_of_day	avg_rating
Afternoon	7.02
Morning	6.94
Evening	6.91

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

CODE:

```
SELECT branch, time_of_day, MAX(rating) AS MAX_Rating
FROM amazon
GROUP BY branch,time_of_day
ORDER BY branch;
```

OUTPUT:

branch	time_of_day	MAX_Rating
A	Afternoon	9.9
A	Evening	9.9
A	Morning	9.9
B	Afternoon	9.9
B	Evening	9.9
B	Morning	9.9
C	Afternoon	9.9
C	Evening	9.9
C	Morning	9.9

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

Solution:

CODE:

```
SELECT day_name, round(avg(rating),2) as AVG_RATING
FROM amazon
GROUP BY day_name
ORDER BY AVG_RATING DESC;
```

OUTPUT:

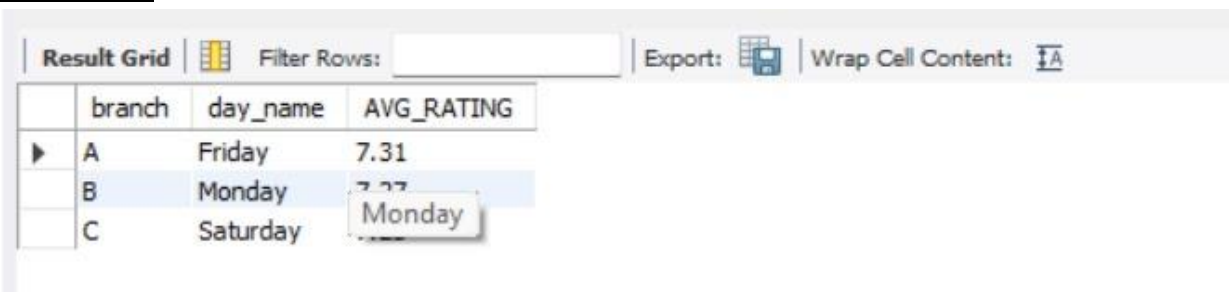
day_name	AVG_RATING
Monday	7.13
Friday	7.06
Tuesday	7.00
Sunday	6.99
Saturday	6.90
Thursday	6.89
Wednesday	6.76

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

CODE:

```
WITH BranchRatings AS (  
    SELECT branch, day_name,  
           ROUND(AVG(rating), 2) AS AVG_RATING  
    FROM amazon  
    GROUP BY branch, day_name  
)  
SELECT b1.branch, b1.day_name, b1.AVG_RATING  
FROM BranchRatings b1  
JOIN (  
    SELECT branch, MAX(AVG_RATING) AS MAX_AVG_RATING  
    FROM BranchRatings  
    GROUP BY branch  
) b2 ON b1.branch = b2.branch AND b1.AVG_RATING =  
b2.MAX_AVG_RATING;
```

OUTPUT:



	branch	day_name	AVG_RATING
▶	A	Friday	7.31
	B	Monday	7.37
	C	Saturday	7.37

Analysis List:

Task 1 :

Product Analysis :

Solution:

Conduct analysis on the data to understand the different product lines, the products lines performing best and the product lines that need to be improved.

CODE:

```
SELECT Product_line, round(AVG(quantity),2) as AVG_quantity_product,  
round(AVG(rating),2) as avg_rating, SUM(unit_price*quantity) as revenue,  
SUM(cogs) as cost_good_sold_sum  
FROM Amazon  
GROUP BY Product_Line;
```

OUTPUT:

Product_line	AVG_quantity_product	avg_rating	revenue	cost_good_sold_sum
Health and beauty	5.59	6.98	46527.98	46527.98
Electronic accessories	5.69	6.91	51222.13	51222.13
Home and lifestyle	5.69	6.84	51297.06	51297.06
Sports and travel	5.53	6.86	51367.74	51367.74
Food and beverages	5.47	7.11	53471.28	53471.28
Fashion accessories	5.07	7.03	51719.90	51719.90

Key Insights:

- **Best Performing Product Line:** Beauty Products (based on highest revenue and rating).
- **Product Line Needing Improvement:** Sports Equipment (lowest revenue and rating).

Recommendations:

1. Improve Profit Margins: The most critical concern is that revenue and cost are nearly the same across all product lines. Focus on optimizing the supply chain and reducing the **cost of goods sold (COGS)** to increase profitability.

2. Promote High-Rated Products: **Food and Beverages** have the highest rating (7.11), which indicates customer satisfaction. Consider marketing these products more aggressively to boost sales and revenue.

3. Review Pricing Strategy: Evaluate whether the pricing strategy for all product lines is appropriate. Small price adjustments could improve revenue without significant impact on quantity sold.

4. Focus on Best Performers: Products like **Fashion Accessories** and **Food and Beverages** have strong customer ratings. Consider increasing inventory or introducing similar products to capitalize on the demand.

5. Low Performers : **Health and Beauty** are products generate the lowest revenue despite having a high rating. This might indicate either a low market reach or pricing issues. It might be beneficial to investigate why this line is underperforming and make strategic changes, such as promotions or discounts to increase sales.

Task 2 :

Sales Analysis :

This analysis aims to answer the question of the sales trends of product. The result of this can help us measure the effectiveness of each sales strategy the business applies and what modifications are needed to gain more sales.

Solution:

CODE:

```
SELECT month_name, SUM(unit_price*quantity) as monthly_revenue,  
SUM(gross_income) as total_gross_income  
from Amazon  
GROUP BY month_name  
ORDER BY month_name desc ;
```

OUTPUT:

month_name	monthly_revenue	total_gross_income
March	103683.00	5184.38
January	110754.16	5537.95
February	91168.93	4558.65

Key Insights :

Monthly Revenue and Gross Income :

- January generated the highest monthly revenue of 110,754.16 and total gross income of 5,537.95.
- March followed with a monthly revenue of 103,683.00 and a total gross income of 5,184.38.
- February recorded the lowest monthly revenue of 91,168.93 and a total gross income of 4,558.65.

- January appears to be the strongest month in terms of sales performance, while February lags behind. This could suggest a need to boost sales or marketing efforts in February to bring it on par with other months.

Recommendations:

1. Analyze January's Success : Since January had the highest revenue and gross income, it would be beneficial to analyze the factors that contributed to its success. Were there any specific marketing campaigns, discounts, or new product launches that drove sales? Replicating these strategies in future months could help maintain or increase revenue.

2. Focus on Improving February's Sales : February showed a considerable drop in revenue compared to January. Investigate whether external factors (seasonality, holidays, etc.) influenced this dip. If the decline is due to a lack of promotions or product availability, consider implementing targeted sales strategies during this period next year.

3. Sustain Momentum in March : Although March performed slightly below January, it still had a strong showing. Focus on sustaining this trend by maintaining customer engagement and promotions throughout the quarter.

4. Seasonal Adjustments : If seasonality plays a role, explore if any product lines are more suitable for certain months. For instance, promote high-demand items or discounts during months like February to counterbalance lower sales trends.

Task 3:

Customer Analysis :

This analysis aims to uncover the different customer segments, purchase trends and the profitability of each customer segment.

Solution:

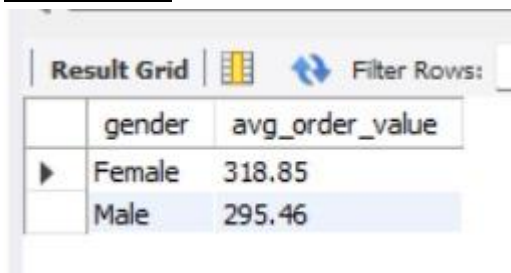
CODE:

```
SELECT gender, round(AVG(unit_price* quantity),2) as avg_order_value FROM
Amazon
group by gender ;
```

```
SELECT customer_type, round(AVG(unit_price* quantity),2) as avg_order_value
FROM Amazon
group by customer_type ;
```

```
SELECT customer_type, gender, round(AVG(unit_price* quantity),2) as
avg_order_value
FROM Amazon
group by customer_type, gender ;
```

OUTPUT:



The screenshot shows a 'Result Grid' window with a table containing two columns: 'gender' and 'avg_order_value'. The table has two rows: 'Female' with a value of 318.85 and 'Male' with a value of 295.46. The 'Male' row is highlighted in blue. Above the table, there are icons for a grid, a refresh button, and a 'Filter Rows:' label.

	gender	avg_order_value
▶	Female	318.85
	Male	295.46

Result Grid			Filter Rows:
	customer_type	avg_order_value	
▶	Member	312.29	
	Normal	301.96	

	customer_type	gender	avg_order_value
▶	Member	Female	321.93
	Normal	Female	315.50
	Normal	Male	289.48
	Member	Male	301.89

Key Insights :

Based on the screenshots you've provided, here are some key insights from the sales data:

Monthly Revenue and Gross Income :

- January generated the highest monthly revenue of 110,754.16 and total gross income of 5,537.95.
- March followed with a monthly revenue of 103,683.00 and a total gross income of 5,184.38.
- February recorded the lowest monthly revenue of 91,168.93 and a total gross income of 4,558.65.
- January appears to be the strongest month in terms of sales performance, while February lags behind. This could suggest a need to boost sales or marketing efforts in February to bring it on par with other months.

Average Order Value by Gender :

- Female customers have a higher average order value (318.85) compared to Male customers (295.46).
- Female customers tend to spend more on average, which could suggest targeting more products or marketing campaigns specifically designed to appeal to female customers.

Average Order Value by Customer Type :

- Members have a higher average order value (312.29) compared to Normal customers (301.96).
- Customers who are members tend to spend more per order than non-members. This suggests that encouraging more customers to join membership programs could lead to higher average order values.

Combined Customer Type and Gender Analysis :

- Female Members have the highest average order value (321.93), followed by Normal Female customers (315.50).
- Normal Male customers have an average order value of 289.48, while Male Members have an average order value of 301.89.
- Female members represent the most valuable customer segment in terms of average spending. Membership programs seem to positively impact both genders but are more effective with female customers. This insight could inform targeted loyalty programs and membership promotions, especially towards females.

Recommendations:

1. **Focus on Female Customers** : Since female customers, particularly members, are showing the highest average order value, consider implementing loyalty programs, personalized marketing, or promotions specifically targeting this segment to further drive sales.
2. **Increase Male Customer Engagement** : Males, particularly normal customers, have the lowest average order value. It may be beneficial to explore why this segment lags in spending. Offering targeted promotions, special deals, or enhancing the membership benefits for males could encourage higher spending.
3. **Strengthen Membership Programs** : Members, in general, are spending more than normal customers. Highlighting the benefits of membership to non-member customers might lead to increased sign-ups, as well as higher average order values.
4. **Tailored Marketing for Segments** : Craft gender-specific and membership-focused marketing strategies to tap into the higher spending behavior of certain segments. For example, providing female customers with product recommendations based on their purchase history or preferences could result in an increased order value.

Final Recommendation:

1. Enhance Targeted Marketing for Female Customers : Female customers, both members and normal, exhibit higher average order values. Tailor marketing strategies to engage this group more effectively through personalized promotions, exclusive offers, and targeted campaigns. Focus on retaining these high-value customers by providing them with product recommendations and personalized experiences based on their preferences and purchase history.

2. Promote Membership Programs : Members, particularly females, spend more than normal customers. To increase the customer base and average order value, promote membership benefits through attractive incentives such as loyalty rewards, exclusive deals, and early access to sales. Encourage non-members to join by highlighting how membership leads to savings and access to premium products.

3. Increase Engagement Among Male Customers : Male customers, especially normal ones, have lower average order values. Focus on increasing male customer engagement through targeted discounts, product bundles, or personalized promotions aimed at their specific interests. Offering incentives like discounts for higher spending or promoting membership could help drive up their order values.

4. Leverage Cross-Selling and Upselling Techniques : For all customer segments, particularly the lower-spending male group, introduce cross-selling and upselling strategies. Suggest complementary products at checkout or offer special discounts for higher-value purchases. This can increase average order values and overall profitability across the board.

Conclusion :

The customer analysis reveals important insights into purchasing trends across different customer segments, specifically in terms of gender and membership status. Female customers, particularly members, exhibit the highest average order values, making them a key segment for the business to target with personalized and loyalty-driven marketing strategies. On the other hand, male customers, especially non-members, spend less on average, indicating an opportunity to engage them more effectively through targeted promotions and incentives.

By enhancing membership programs, focusing on cross-selling and upselling techniques, and leveraging tailored marketing campaigns, the business can capitalize on the higher spending behavior of certain customer groups while working to improve engagement and order values among lower-performing segments. These strategies will help boost revenue, improve customer retention, and ensure more effective sales efforts across all segments.

Overall, this analysis serves as a foundation for implementing data-driven decisions that can improve both customer satisfaction and business profitability.