

CAPSTONE PROJECT

SQL

Name: Saiyyam Sushil Jain

ID: S10159

LIST OF CONTENTS

03	TITLE OF THE PROJECT
04	BUSINESS PROBLEM AND OBJECTIVE
05	DATASET OVERVIEW
06	TOOLS & METHOD
07-33	QUERY ANALYSIS & RESULT
34	KEY INSIGHTS
35	RECOMMENDATIONS
36	CONCLUSION

introducing

“SALES DECODED”

amazon

The Amazon logo, featuring a thick orange curved arrow pointing from the letter 'a' to the letter 'z'.

BUSINESS PROBLEM:

Amazon operates multiple branches with diverse products, but identifying key factors driving sales success across branches, product lines, and customer segments remains challenging

OBJECTIVE:

- Product Analysis
- Sales Analysis
- Customer Analysis

DATASET OVERVIEW

- 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

APPROACH USED

- Data Wrangling
- Feature Engineering
- Exploratory Data Analysis

invoice_id
branch
city
customer_type
gender
product_line
unit_price
quantity
VAT
total
date
time
payment_method
cogs
gross_margin_percentage
gross_income
rating

TOOLS & METHOD

- Tools: MySQL Workbench, Excel (for preparation).
- Techniques: SQL queries (GROUP BY, JOIN, AGGREGATE FUNCTION, SUBQUERIES, CASE etc.).



QUERY ANALYSIS & RESULT

-- Q1 = What is the count of distinct cities in the dataset?

```
select count(distinct city) as distinct_city  
from data_amazon;
```

distinct_city
3

-- Q2 = For each branch, what is the corresponding city?

```
select branch, city  
from data_amazon  
group by branch;
```

branch	city
A	Yangon
C	Naypyitaw
B	Mandalay

-- Q3 = What is the count of distinct product lines in the dataset?

```
select count(distinct product_line) as distinct_product_line  
from data_amazon;
```

distinct_product_line
6

-- Q4 = Which payment method occurs most frequently?

```
select payment_method, count(*) as pay_method_count
from data_amazon
group by payment_method
order by pay_method_count desc;
```

payment_method	pay_method_count
Ewallet	345
Cash	344
Credit card	311

```
-- Q5 = Which product line has the highest sales?
```

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

product_line	total_sales
Food and beverages	56153



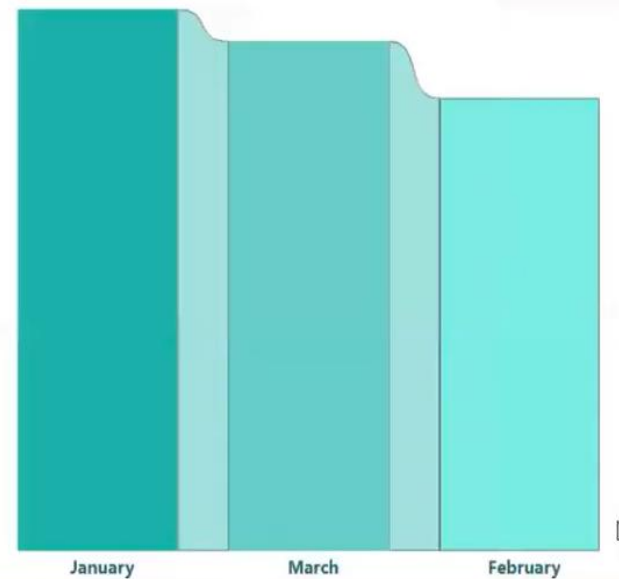
product_line

- Food and beverages
- Sports and travel
- Electronic accessories
- Fashion accessories
- Home and lifestyle
- Health and beauty

-- Q6 = How much revenue is generated each month?

```
select month_name, sum(total) as monthly_revenue  
from data_amazon  
group by month_name  
order by monthly_revenue desc;
```

month_name	monthly_revenue
January	116294
March	109463
February	97213



-- Q7 = In which month did the cost of goods sold reach its peak?

```
select month_name, sum(cogs) as monthly_cogs
from data_amazon
group by month_name
order by monthly_cogs desc
limit 1;
```

month_name	monthly_cogs
January	110759

month_name	monthly_cogs
January	110759
March	104255
February	92598

-- Q8 = Which product line generated the highest revenue?

```
select product_line, sum(total) as total_revenue
from data_amazon
group by product_line
order by total_revenue desc
limit 1;
```

product_line	total_revenue
Food and beverages	56153

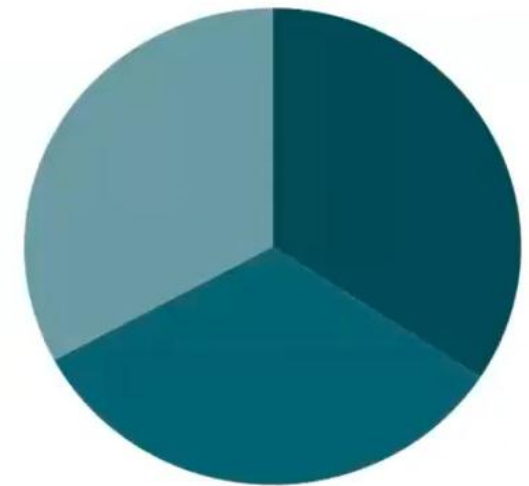
product_line	total_sales
Food and beverages	56153
Sports and travel	55125
Electronic accessories	54338
Fashion accessories	54298
Home and lifestyle	53866
Health and beauty	49190

-- Q9 = In which city was the highest revenue recorded?

```
select city, sum(total) as total_revenue
from data_amazon
group by city
order by total_revenue desc
limit 1;
```

city	total_revenue
Naypyitaw	110569

city	total_revenue
Naypyitaw	110569
Yangon	106202
Mandalay	106199



● Naypyitaw ● Yangon ● Mandalay

-- Q10= Which product line incurred the highest Value Added Tax?

```
select product_line, sum(VAT) as total_VAT
from data_amazon
group by product_line
order by total_VAT desc
limit 1;
```

product_line	total_VAT
Food and beverages	2673

product_line	total_VAT
Food and beverages	2673
Sports and travel	2621
Electronic accessories	2590
Fashion accessories	2590
Home and lifestyle	2565
Health and beauty	2343


```
-- Q11 = 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 data_amazon
                                group by product_line) as Sales)
         THEN 'Good'
         ELSE 'Bad'
       END AS PL_Category
from data_amazon
group by product_line;
```

product_line	total_sales	PL_Category
Health and beauty	49190	Bad
Electronic accessories	54338	Good
Home and lifestyle	53866	Good
Sports and travel	55125	Good
Food and beverages	56153	Good
Fashion accessories	54298	Good

```
-- Q12 = Identify the branch that exceeded  
--      the average number of products sold.
```

```
select branch, sum(quantity) as total_qty_sold  
from data_amazon  
group by branch  
having sum(quantity) > (select avg(branch_sales)  
                        from  
                        (select sum(quantity) as branch_sales  
                         from data_amazon  
                         group by branch) as Qty_Sold  
                        );
```

branch	total_qty_sold
A	1859

Avg Branch Sales 1836.6667
:

-- Q13 = Which product line is most frequently associated with each gender?

```
select gender, product_line, count(*) as PL_count  
from data_amazon  
group by gender;
```

gender	product_line	PL_count
Female	Health and beauty	501
Male	Home and lifestyle	499

```
-- Q14 = Calculate the average rating for each product line.
```

```
select product_line, avg(rating) as PL_Rating  
from data_amazon  
group by product_line  
order by PL_Rating desc;
```

product_line	PL_Rating
Food and beverages	7.1494
Fashion accessories	7.1348
Health and beauty	7.0395
Electronic accessories	7
Sports and travel	6.988
Home and lifestyle	6.8563

day_name	time_of_day	sales_occurrence
Friday	Afternoon	74
Friday	Evening	36
Friday	Morning	29
Monday	Afternoon	75
Monday	Evening	29
Monday	Morning	21
Saturday	Afternoon	81
Saturday	Evening	55
Saturday	Morning	28
Sunday	Afternoon	70
Sunday	Evening	41
Sunday	Morning	22
Thursday	Afternoon	76
Thursday	Evening	29
Thursday	Morning	33
Tuesday	Afternoon	71
Tuesday	Evening	51
Tuesday	Morning	36
Wednesday	Afternoon	81
Wednesday	Evening	40
Wednesday	Morning	22

```
-- Q15 = Count the sales occurrences for each
--       time of day on every weekday.
```

```
select day_name, time_of_day,
       count(*) as sales_occurrence
from data_amazon
group by day_name, time_of_day
order by day_name, time_of_day;
```

```
-- Q16 = Identify the customer type contributing the highest revenue.
```

```
select customer_type, sum(total) as highest_revenue  
from data_amazon  
group by customer_type  
order by highest_revenue desc  
limit 1;
```

customer_type	highest_revenue
Member	164230

-- Q17 = Determine the city with the highest VAT percentage.

```
select city, max(VAT) as highest_VAT
from data_amazon
group by city
order by highest_VAT desc
limit 1;
```

city	highest_VAT
Naypyitaw	50

-- Q18 = Identify the customer type with the highest VAT payments.

```
select customer_type, max(VAT) as highest_VAT
from data_amazon
group by customer_type
order by highest_VAT desc
limit 1;
```

customer_type	highest_VAT
Member	50

-- Q19 = What is the count of distinct customer types in the dataset?

```
select count(distinct customer_type) as cust_type  
from data_amazon;
```

cust_type
2

-- Q20 = What is the count of distinct payment methods in the dataset?

```
select count(distinct payment_method) as pay_method  
from data_amazon;
```

pay_method
3

-- Q21 = Which customer type occurs most frequently?

```
select customer_type, count(*) as frequency
from data_amazon
group by customer_type
order by frequency desc
limit 1;
```

customer_type	frequency
Member	501

-- Q22 = Identify the customer type with the highest purchase frequency.

```
select customer_type, count(invoice_id) as purchase_frequency
from data_amazon
group by customer_type
order by purchase_frequency desc
limit 1;
```

customer_type	purchase_frequency
Member	501

-- Q23 = Determine the predominant gender among customers.

```
select gender, count(*) as gender_count
from data_amazon
group by gender
order by gender_count desc;
```

gender	gender_count
Female	501
Male	499

-- Q24 = Examine the distribution of genders within each branch.

```
select branch, gender, count(*) as distribution
from data_amazon
group by branch, gender
order by distribution desc;
```

branch	gender	distribution
A	Male	179
C	Female	178
B	Male	170
B	Female	162
A	Female	161
C	Male	150



-- Q25 = Identify the time of day when customers provide the most ratings.

```
select time_of_day, count(*) as rating_count
from data_amazon
group by time_of_day
order by rating_count desc;
```

time_of_day	rating_count
Afternoon	528
Evening	281
Morning	191


```
-- Q26 = Determine the time of day with the  
--       highest customer ratings for each branch.
```

```
select branch, time_of_day, count(*) as rating_count  
from data_amazon  
group by branch  
order by rating_count desc;
```

branch	time_of_day	rating_count
A	Afternoon	340
B	Afternoon	332
C	Morning	328

-- Q27 = Identify the day of the week with the highest average ratings.

```
select day_name, avg(rating) as avg_rating
from data_amazon
group by day_name
order by avg_rating desc
limit 1;
```

day_name	avg_rating
Monday	7.232

day_name	avg_rating
Monday	7.232
Friday	7.1367
Tuesday	7.0823
Sunday	7.0451
Thursday	6.9565
Saturday	6.9329
Wednesday	6.8671


```
-- Q28 = Determine the day of the week with the  
--       highest average ratings for each branch.
```

```
select branch, day_name, avg(rating) as avg_rating  
from data_amazon  
group by branch  
order by branch;
```

branch	day_name	avg_rating
A	Saturday	7.0912
B	Wednesday	6.8855
C	Friday	7.1159

KEY INSIGHTS

- The product line “**Food and beverages**” generated the highest revenue, indicating strong customer demand.
- The sales data reveals that “**Mandalay**” branch consistently outperforms others in total sales.
- Customer type “**Member**” is the most frequent buyer, contributing significantly to overall sales.
- The month of “**January**” sees peak sales activity, aligning with seasonal trends or promotions.
- Branch “**A**” has recorded the highest number of products sold, showcasing its strong sales performance and customer demand.

RECOMMENDATIONS

- Focus marketing efforts on promoting the “**Home and Lifestyle**” product line to further capitalize on its popularity.
- Allocate more resources to “**C**” branch to maximize its sales potential
- Introduce time-specific offers during “**Afternoon**” to boost sales during peak hours.

CONCLUSION

- By addressing underperforming areas and leveraging high-performing segments through this analysis, Amazon can achieve greater profitability.
- Optimizing operations based on these insights will enhance efficiency and customer satisfaction.
- Understanding customer behavior and sales trends empowers Amazon to make data-driven decisions that drive growth.

**THANK YOU FOR
YOUR TIME!**