Sales Analysis Project : Understanding Amazon Sales Data

Name : Shraddhanath Kar

Objectives

- ✓ The main aim of the project is to gain insights into sales data of Amazon in order to understand the different factors that affect sales across various branches. The analysis focus on three key areas.

1. Product Analysis :
   - ✓ Identification of different product lines
   - ✓ Determination of the best performing product lines
   - ✓ Identify high performing product lines that contribute significantly to overall sales
   - ✓ Identification of product lines that require improvement

2. Sales Analysis :
   - ✓ analysis of sales trend over time to understand pattern and fluctuation
   - ✓ Evaluation of the effectiveness of sales strategies implement by business
   - ✓ Identification of areas for sales improvement and improving sales revenue
   - ✓ Identify factor influencing sales performance such as seasonality or promotional activities

3. Customer Analysis :
   - ✓ Identification of different customer segment
   - ✓ Analysis of purchase trends by customer segment to understand their performance and buying habits
   - ✓ Evaluation of the profitability of each customer segment by analysing their contribution to revenue and profit
   - ✓ Identify opportunities for targeting specific customer segment more effectively and improving cutomer satisfaction and loyalty

Importance of understanding sales data

- ✓ Identifying trends : Sales data provide available insights into trend overtime such as seasonal fluctuations, product popularity and changes in customere preference. By understanding these trends, business can adapt their strategies accordingly to capitalize on opportunity and reduce risk.
- ✓ Measuring performance : Sales data allows to track the performance of product. By analyzing revenue,profit margin business can access the effectiveness of their strategies and make adjustment as needed to improve performance.
- ✓ Target & Goal : Sales data help to set target & Goal for revenue growth,market share and customer asset. by analyzing sales data we can identify areas for improvement to drive future growth.

Feature Engineering

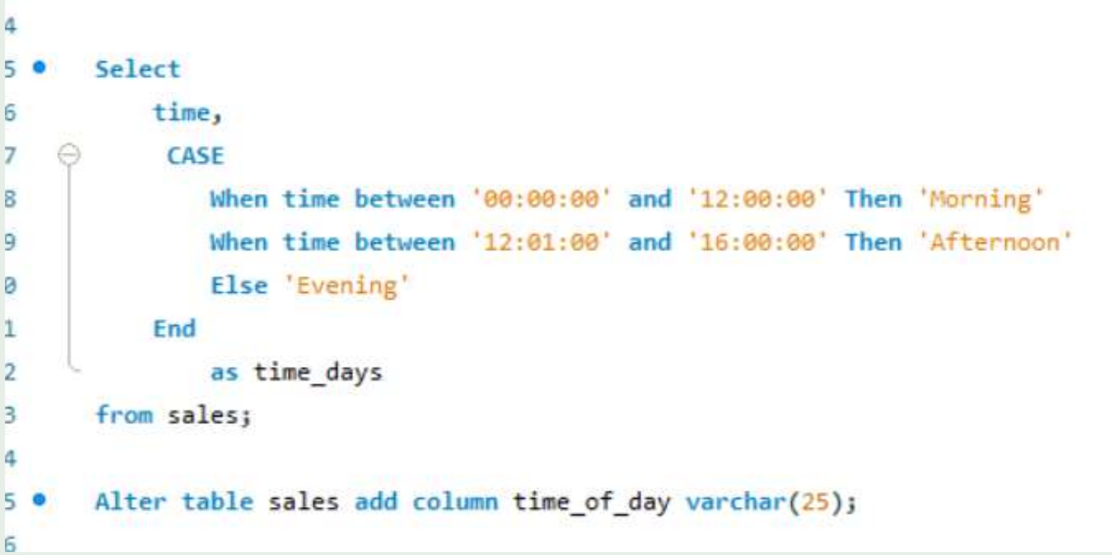
- ✓ Here we added 3 columns using a CASE statement.
  - ❑ Time of day : Which indicate the time category for each transaction
  - ❑ Day name : Which contain day name for each transaction.
  - ❑ Month name : which contain month name for each transaction

Handling missing value

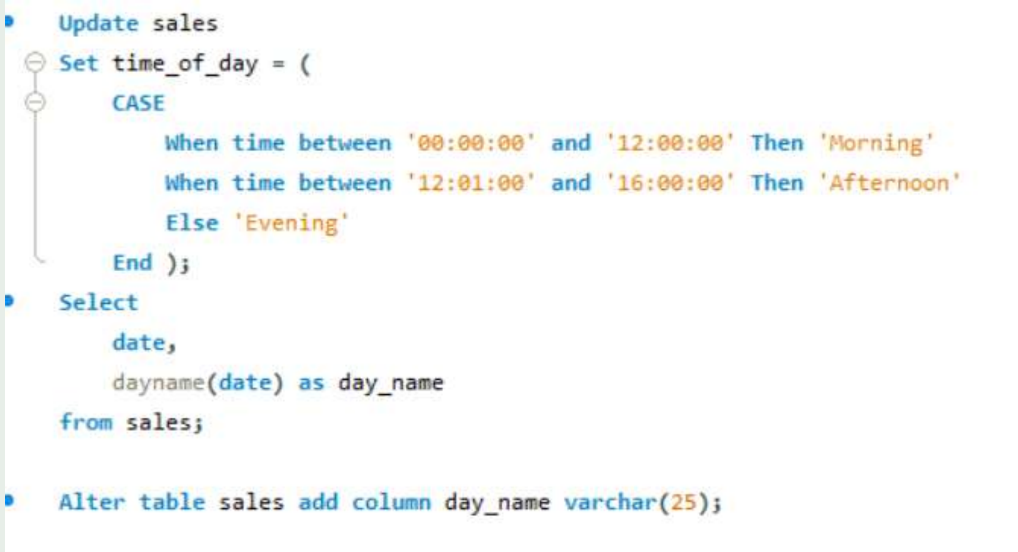
- ✓ In this dataset there is NOT NULL constraint for each field. It ensure that null values are not present. How ever it always a good practice to verify the dataset for missing value and handle them appropriately if they exist.

IDE TOOL

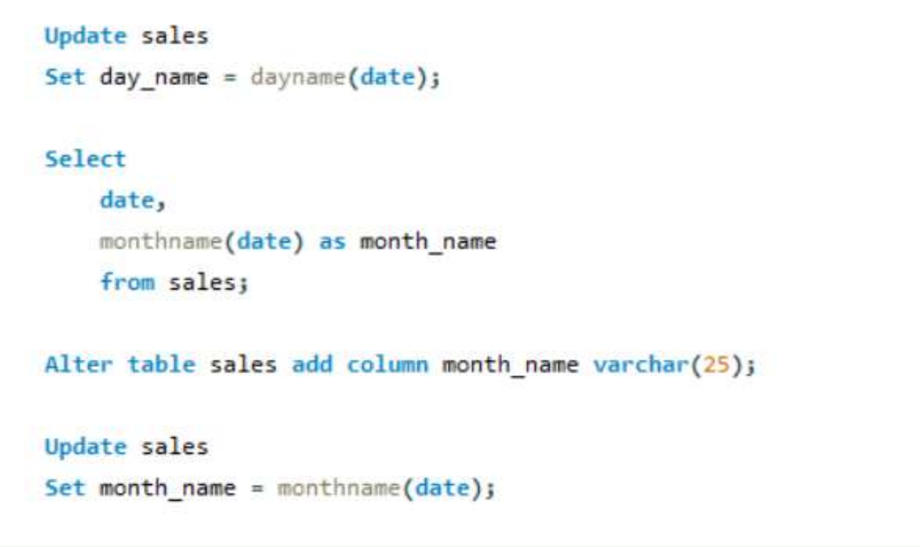
Feature Engineering



```sql
Select
    time,
     CASE
        When time between '00:00:00' and '12:00:00' Then 'Morning'
        When time between '12:01:00' and '16:00:00' Then 'Afternoon'
        Else 'Evening'
    End
        as time_days
from sales;

Alter table sales add column time_of_day varchar(25);
```



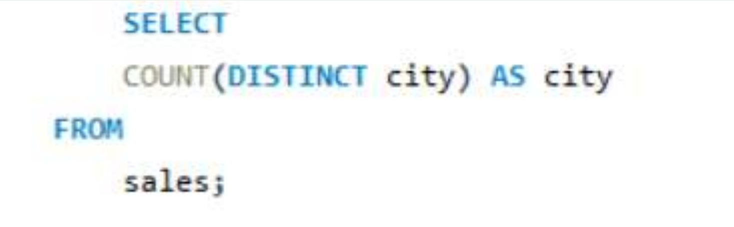
```sql
Update sales
Set time_of_day = (
    CASE
        When time between '00:00:00' and '12:00:00' Then 'Morning'
        When time between '12:01:00' and '16:00:00' Then 'Afternoon'
        Else 'Evening'
    End );
Select
    date,
    dayname(date) as day_name
from sales;

Alter table sales add column day_name varchar(25);
```



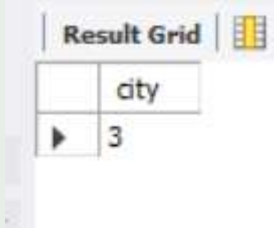
```sql
Update sales
Set day_name = dayname(date);

Select
    date,
    monthname(date) as month_name
    from sales;

Alter table sales add column month_name varchar(25);

Update sales
Set month_name = monthname(date);
```

EDA

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

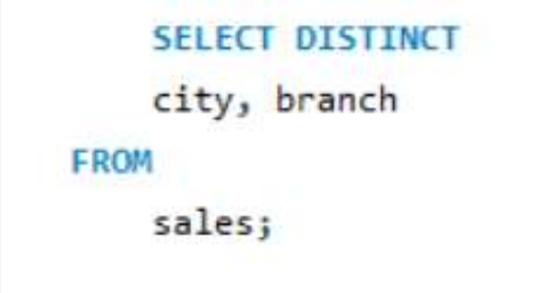


```
    SELECT
    COUNT(DISTINCT city) AS city
FROM
    sales;
```

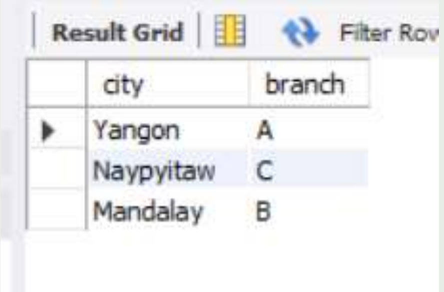


	city
▶	3

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

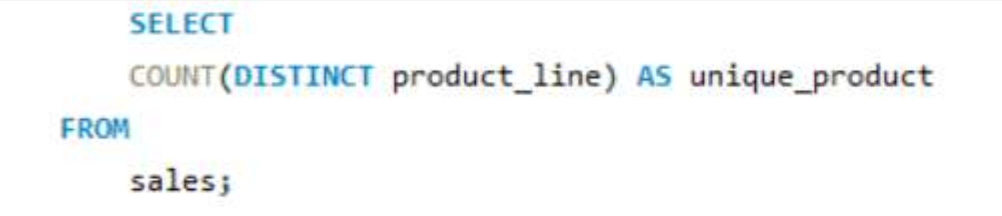


```
    SELECT DISTINCT
    city, branch
FROM
    sales;
```

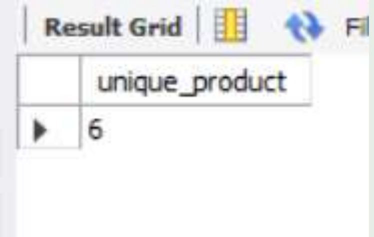


	city	branch
▶	Yangon	A
	Naypyitaw	C
	Mandalay	B

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



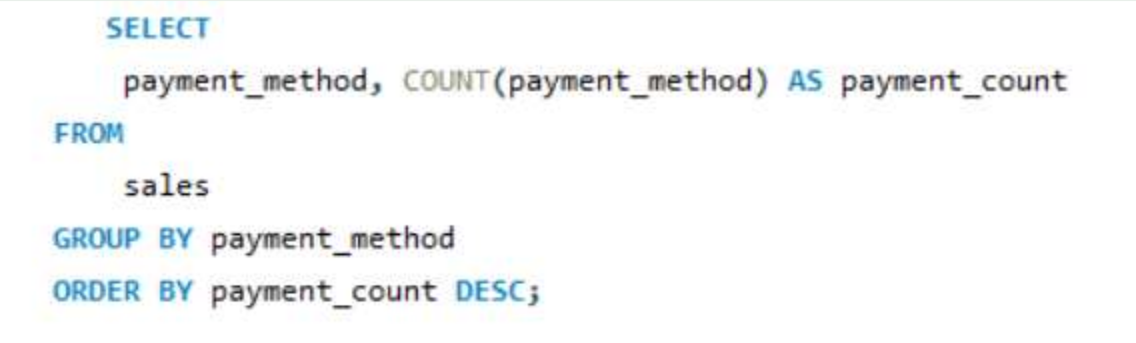
```
    SELECT
    COUNT(DISTINCT product_line) AS unique_product
FROM
    sales;
```



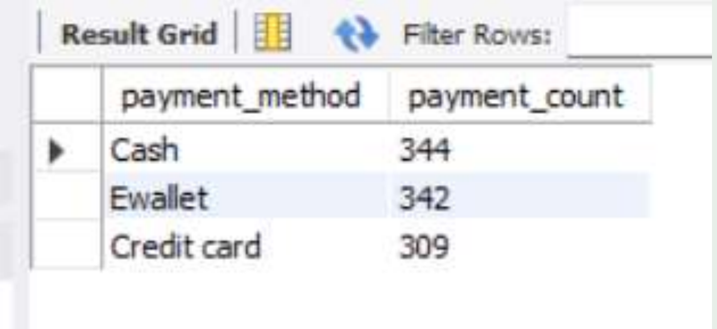
	unique_product
▶	6

EDA

4. Which payment method occurs most frequently?

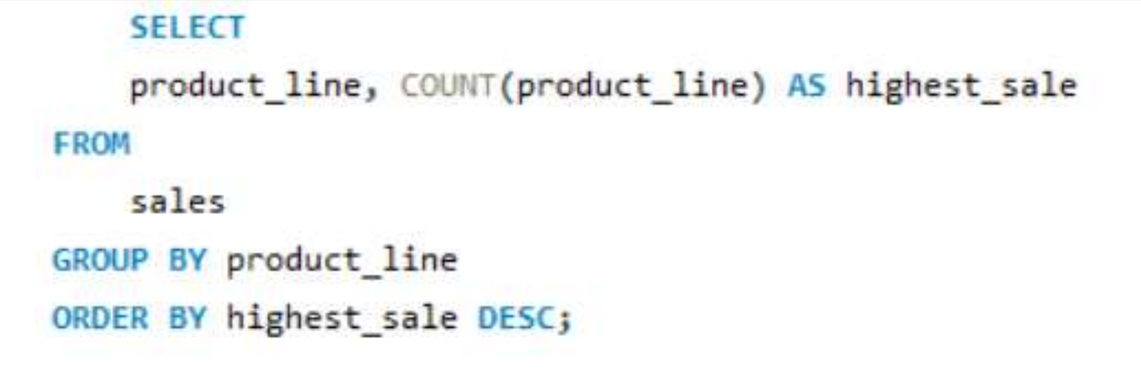


```
    SELECT
     payment_method, COUNT(payment_method) AS payment_count
FROM
     sales
GROUP BY payment_method
ORDER BY payment_count DESC;
```

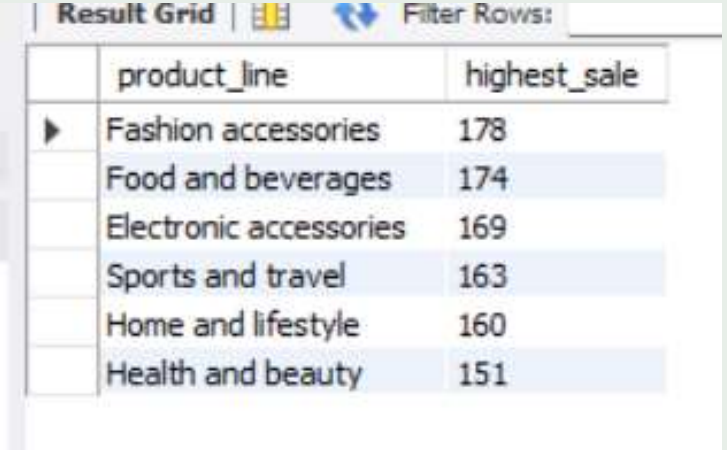


payment_method	payment_count
Cash	344
Ewallet	342
Credit card	309

5. Which product line has the highest sales?



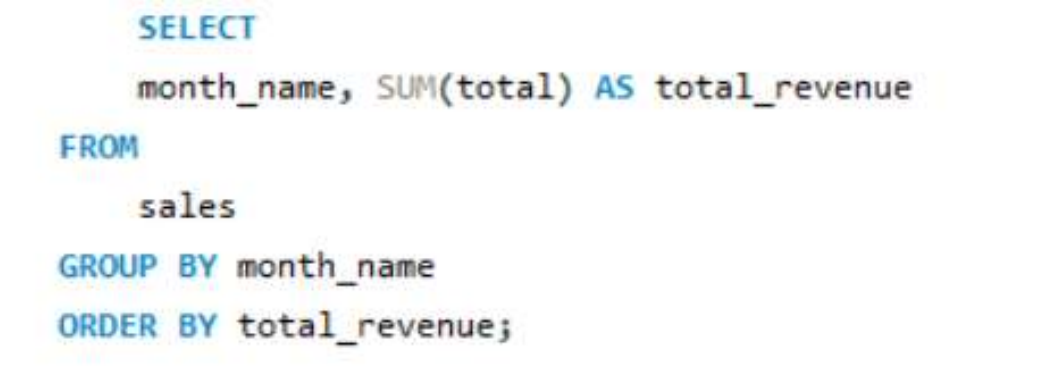
```
    SELECT
    product_line, COUNT(product_line) AS highest_sale
FROM
    sales
GROUP BY product_line
ORDER BY highest_sale DESC;
```



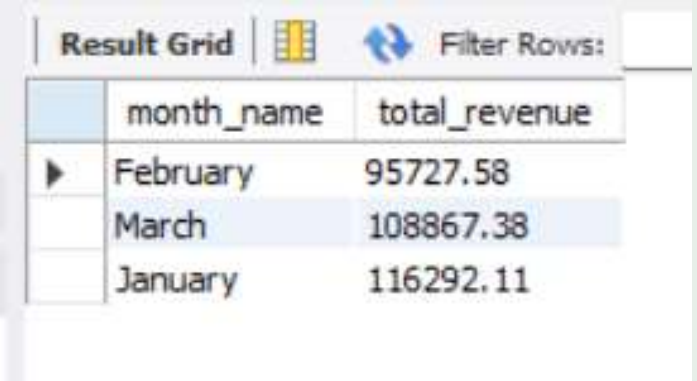
product_line	highest_sale
Fashion accessories	178
Food and beverages	174
Electronic accessories	169
Sports and travel	163
Home and lifestyle	160
Health and beauty	151

EDA

6. How much revenue is generated each month?

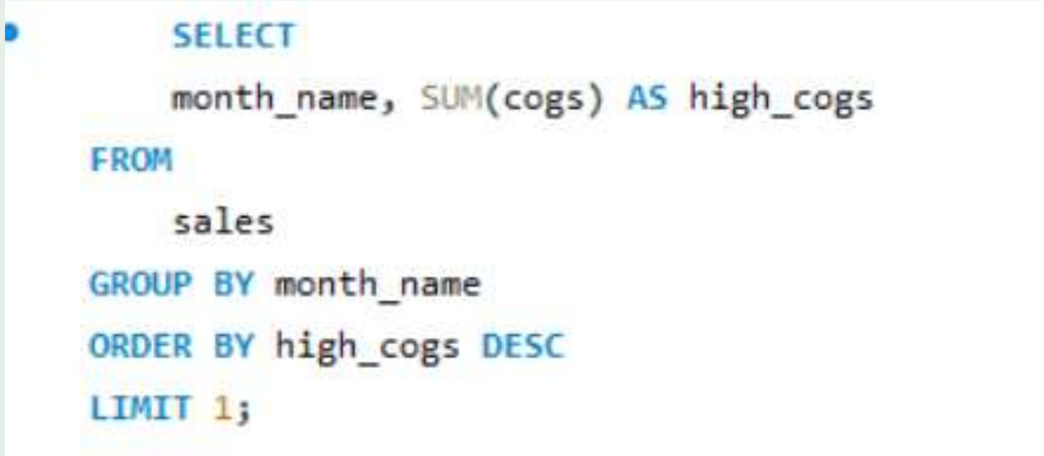


```
    SELECT
    month_name, SUM(total) AS total_revenue
FROM
    sales
GROUP BY month_name
ORDER BY total_revenue;
```

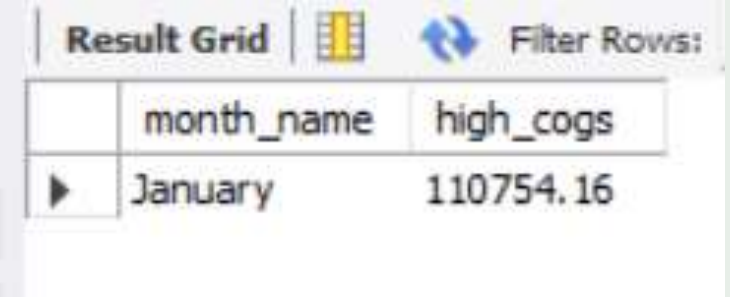


month_name	total_revenue
February	95727.58
March	108867.38
January	116292.11

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



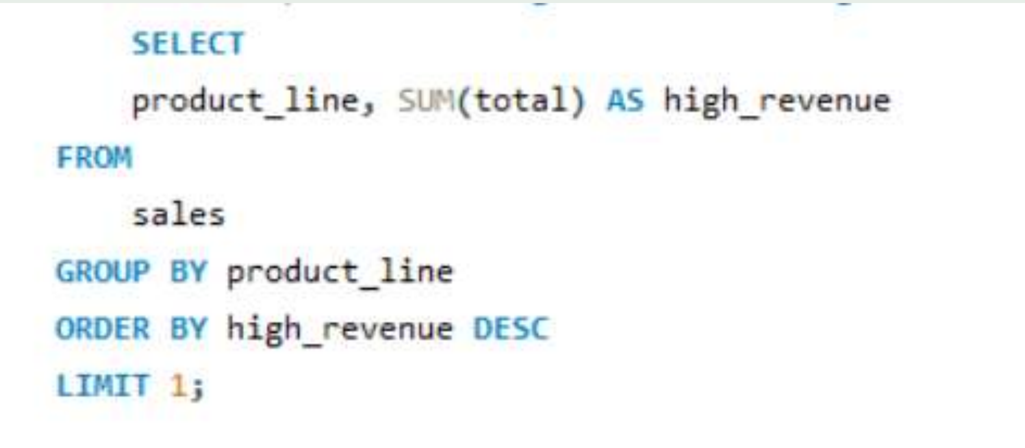
```
    SELECT
    month_name, SUM(cogs) AS high_cogs
FROM
    sales
GROUP BY month_name
ORDER BY high_cogs DESC
LIMIT 1;
```



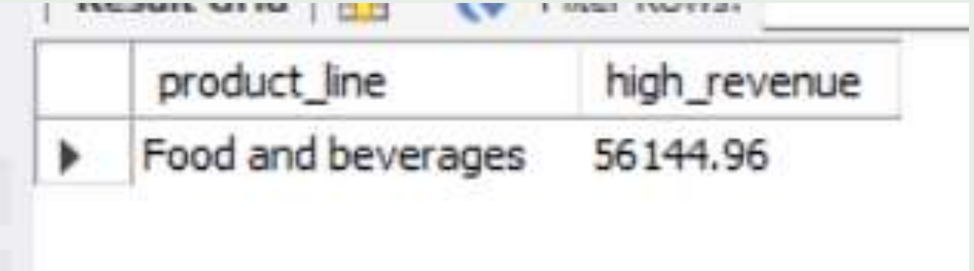
month_name	high_cogs
January	110754.16

EDA

8. Which product line generated the highest revenue?

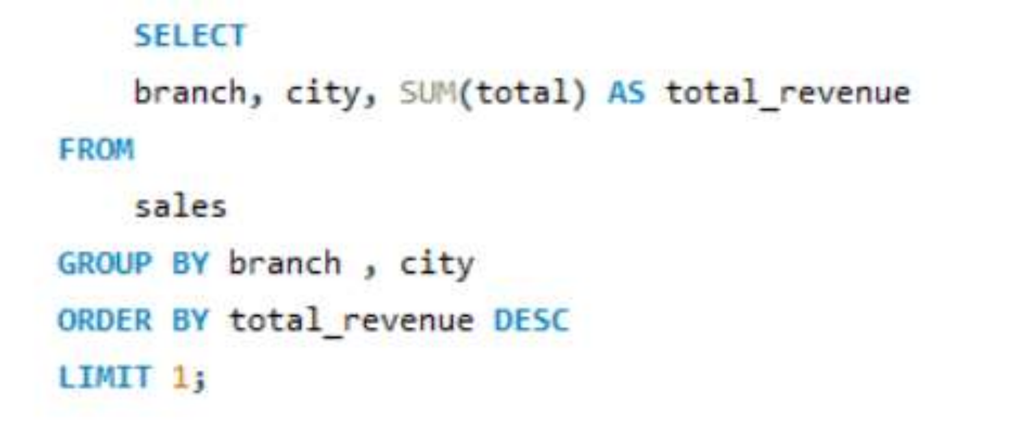


```
    SELECT
    product_line, SUM(total) AS high_revenue
FROM
    sales
GROUP BY product_line
ORDER BY high_revenue DESC
LIMIT 1;
```

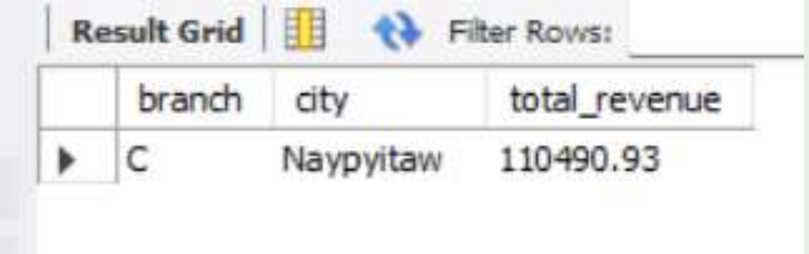


	product_line	high_revenue
▶	Food and beverages	56144.96

9. In which city was the highest revenue recorded?



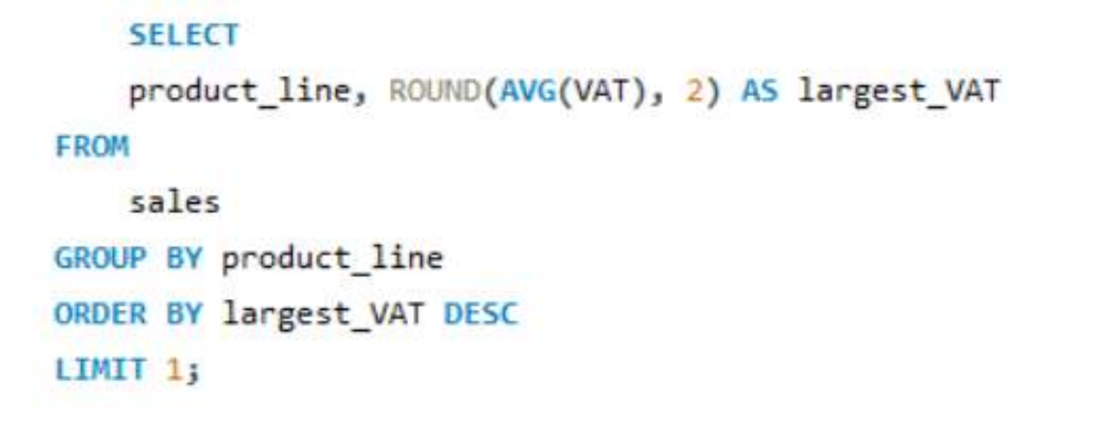
```
    SELECT
    branch, city, SUM(total) AS total_revenue
FROM
    sales
GROUP BY branch , city
ORDER BY total_revenue DESC
LIMIT 1;
```



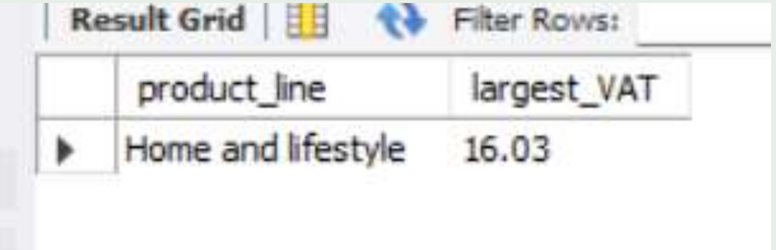
	branch	city	total_revenue
▶	C	Naypyitaw	110490.93

EDA

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

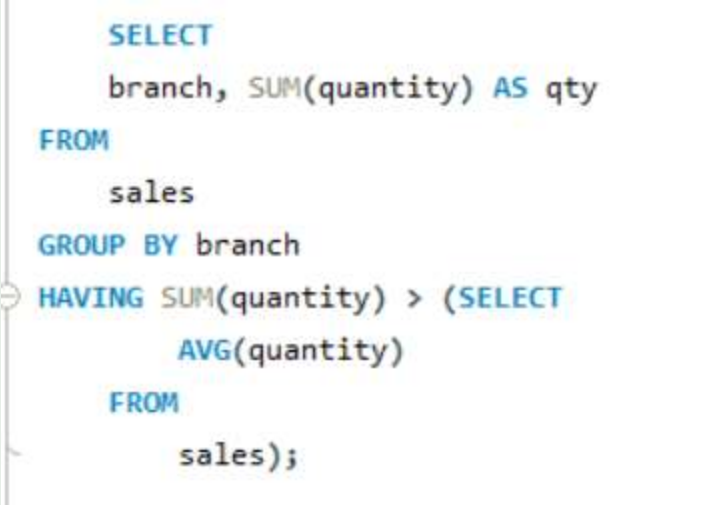


```
    SELECT
    product_line, ROUND(AVG(VAT), 2) AS largest_VAT
FROM
    sales
GROUP BY product_line
ORDER BY largest_VAT DESC
LIMIT 1;
```

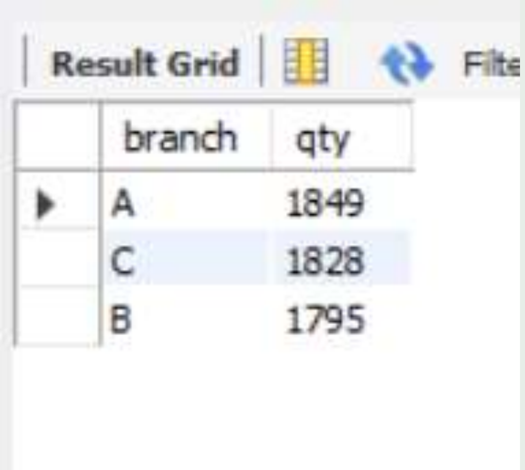


product_line	largest_VAT
Home and lifestyle	16.03

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



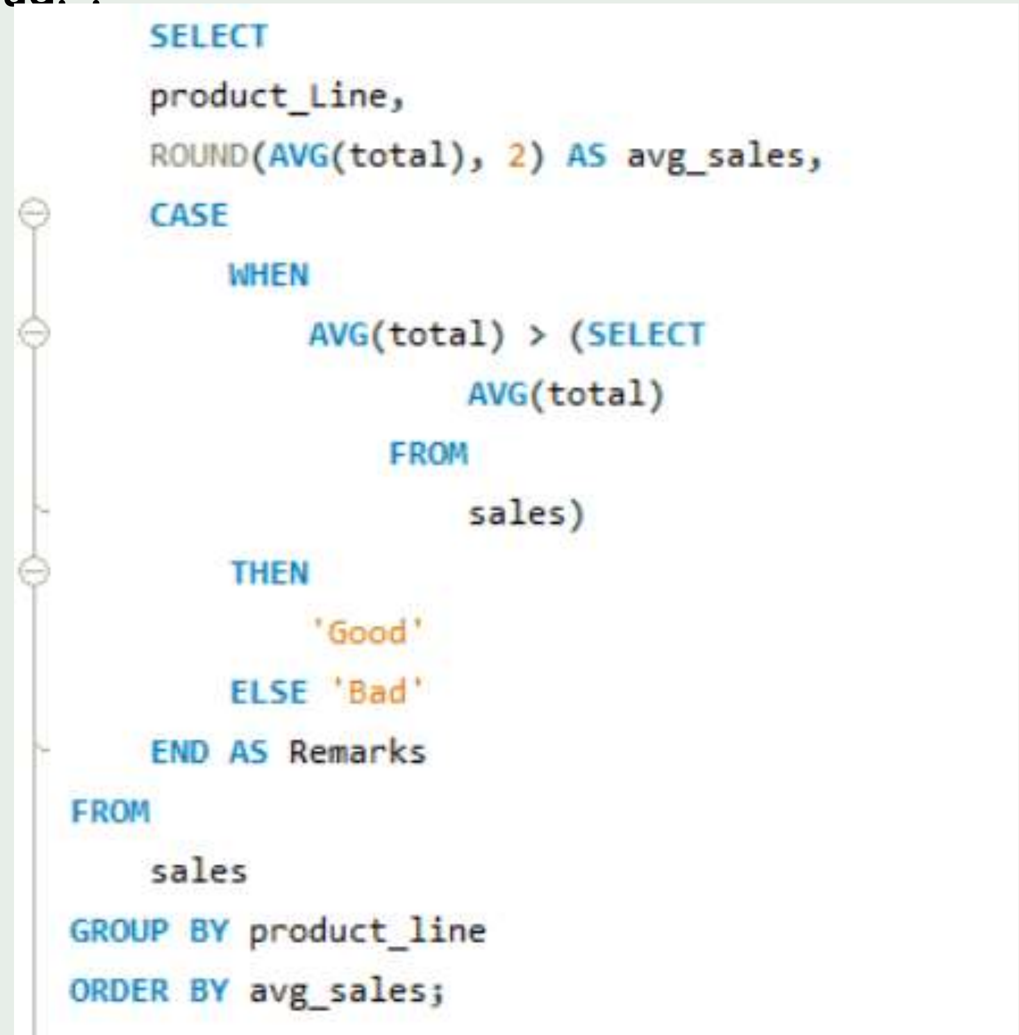
```
    SELECT
    branch, SUM(quantity) AS qty
FROM
    sales
GROUP BY branch
HAVING SUM(quantity) > (SELECT
        AVG(quantity)
    FROM
        sales);
```



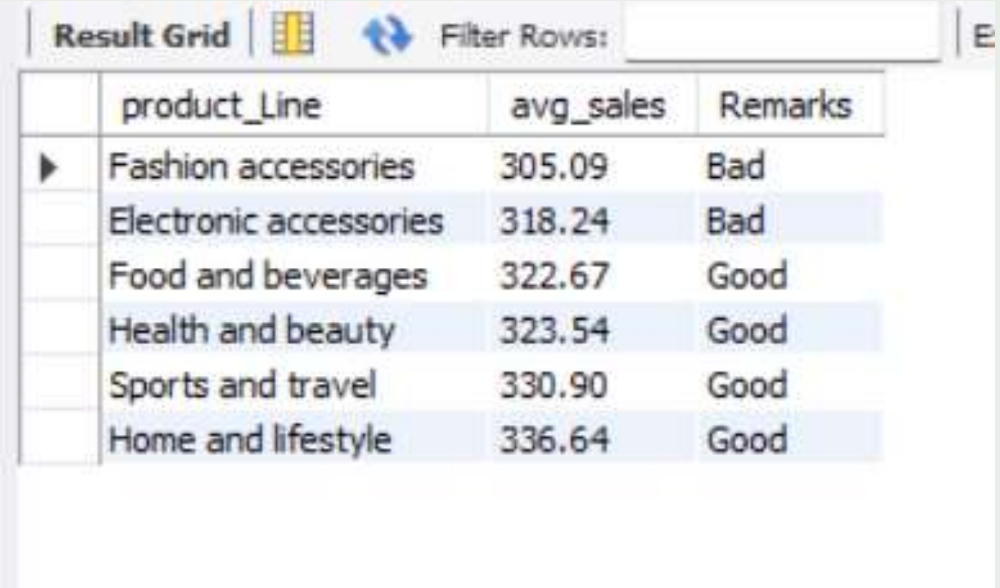
branch	qty
A	1849
C	1828
B	1795

EDA

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



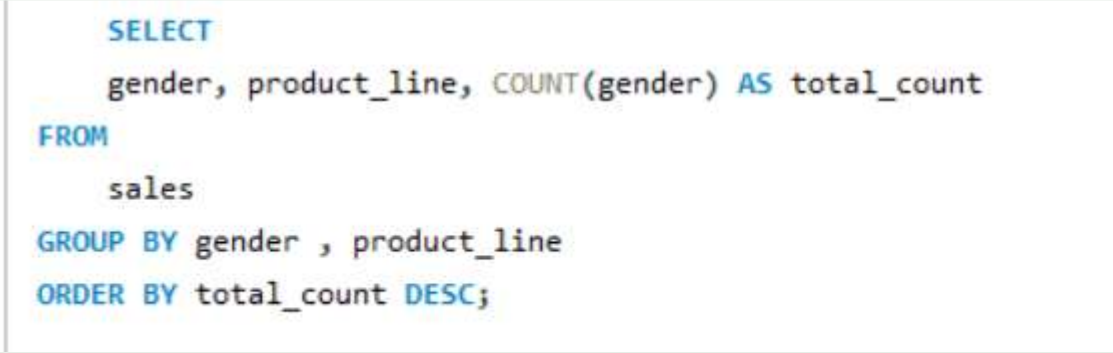
```
    SELECT
    product_Line,
    ROUND(AVG(total), 2) AS avg_sales,
    CASE
        WHEN
            AVG(total) > (SELECT
                    AVG(total)
                FROM
                    sales)
        THEN
            'Good'
        ELSE 'Bad'
    END AS Remarks
FROM
    sales
GROUP BY product_line
ORDER BY avg_sales;
```



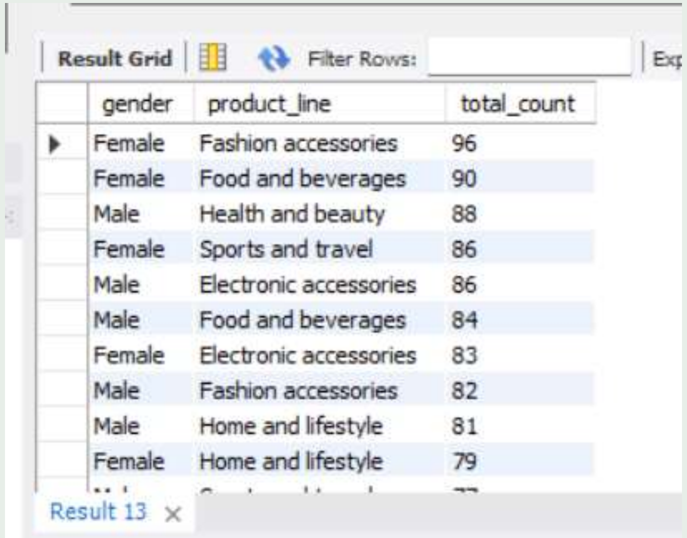
product_Line	avg_sales	Remarks
Fashion accessories	305.09	Bad
Electronic accessories	318.24	Bad
Food and beverages	322.67	Good
Health and beauty	323.54	Good
Sports and travel	330.90	Good
Home and lifestyle	336.64	Good

EDA

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

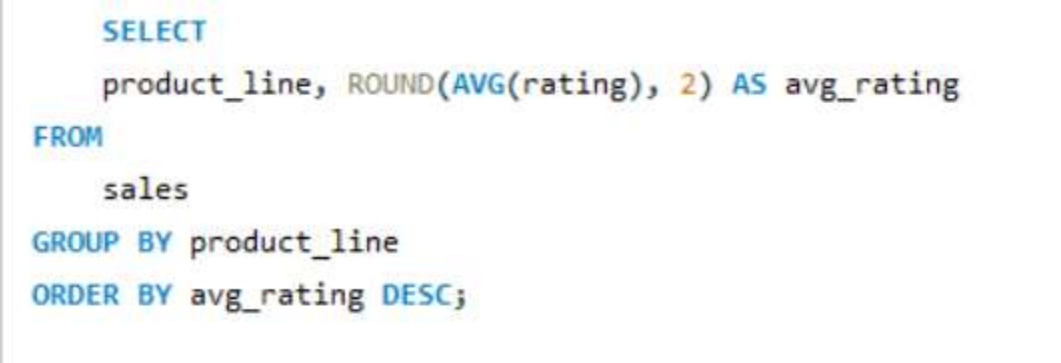


```
    SELECT
    gender, product_line, COUNT(gender) AS total_count
FROM
    sales
GROUP BY gender , product_line
ORDER BY total_count DESC;
```

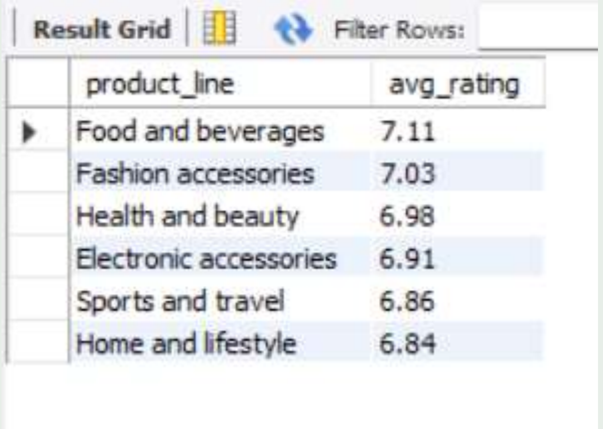


gender	product_line	total_count
Female	Fashion accessories	96
Female	Food and beverages	90
Male	Health and beauty	88
Female	Sports and travel	86
Male	Electronic accessories	86
Male	Food and beverages	84
Female	Electronic accessories	83
Male	Fashion accessories	82
Male	Home and lifestyle	81
Female	Home and lifestyle	79

14. Calculate the average rating for each product line.



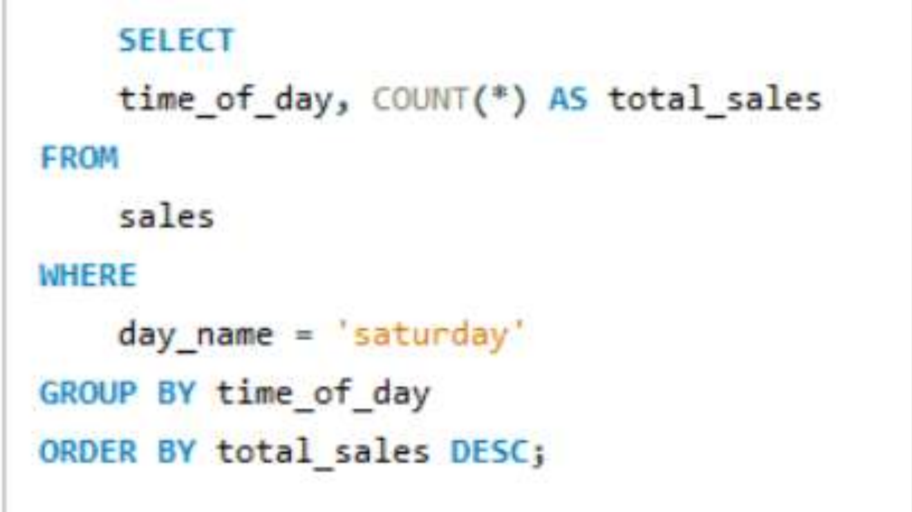
```
    SELECT
    product_line, ROUND(AVG(rating), 2) AS avg_rating
FROM
    sales
GROUP BY product_line
ORDER BY avg_rating DESC;
```



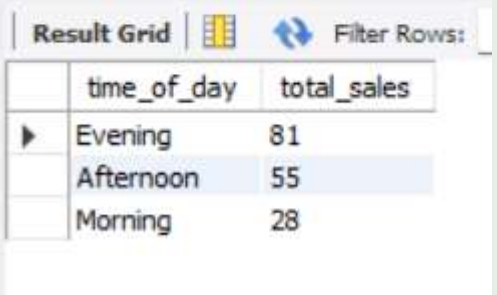
product_line	avg_rating
Food and beverages	7.11
Fashion accessories	7.03
Health and beauty	6.98
Electronic accessories	6.91
Sports and travel	6.86
Home and lifestyle	6.84

EDA

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

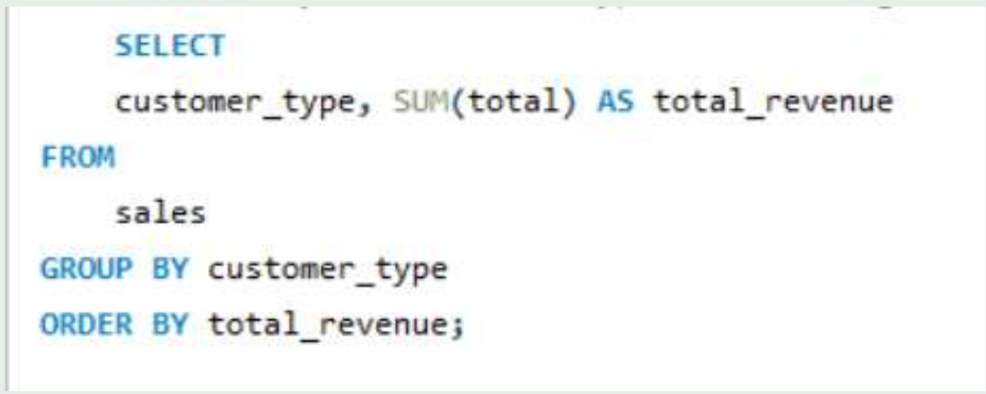


```
    SELECT
    time_of_day, COUNT(*) AS total_sales
FROM
    sales
WHERE
    day_name = 'saturday'
GROUP BY time_of_day
ORDER BY total_sales DESC;
```

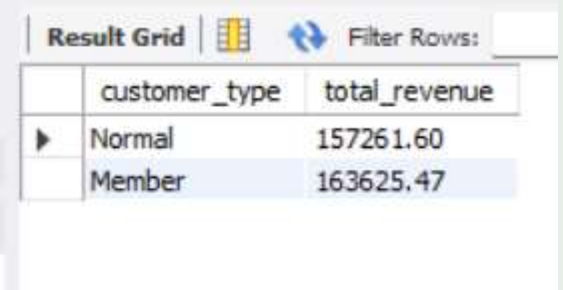


time_of_day	total_sales
Evening	81
Afternoon	55
Morning	28

16. Identify the customer type contributing the highest revenue.



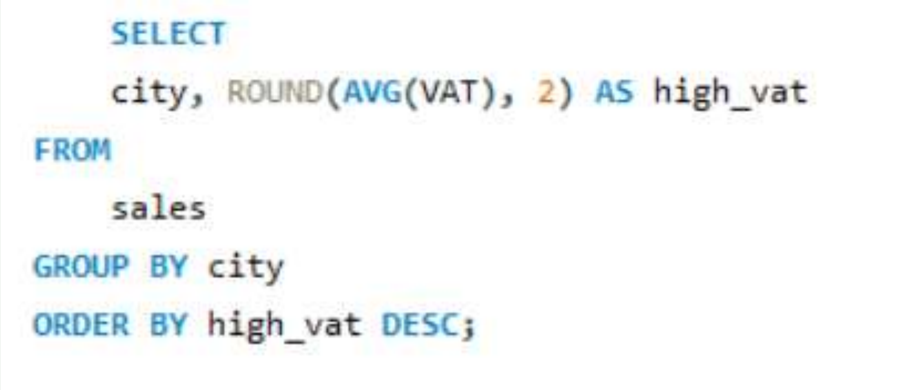
```
    SELECT
    customer_type, SUM(total) AS total_revenue
FROM
    sales
GROUP BY customer_type
ORDER BY total_revenue;
```



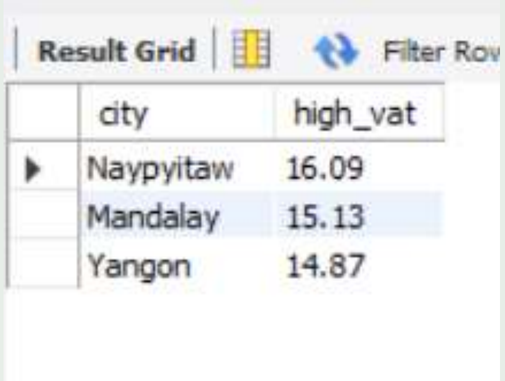
customer_type	total_revenue
Normal	157261.60
Member	163625.47

EDA

17. Determine the city with the highest VAT percentage.

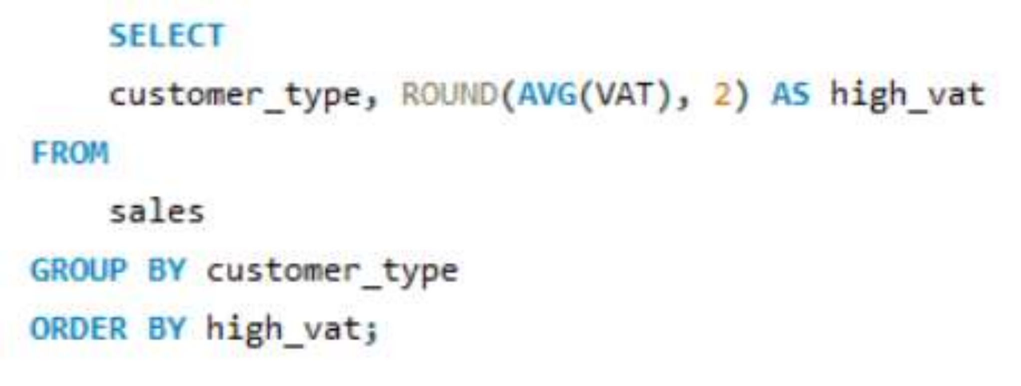


```
    SELECT
    city, ROUND(AVG(VAT), 2) AS high_vat
FROM
    sales
GROUP BY city
ORDER BY high_vat DESC;
```

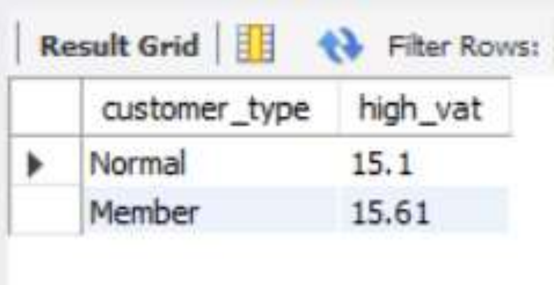


	city	high_vat
▶	Naypyitaw	16.09
	Mandalay	15.13
	Yangon	14.87

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



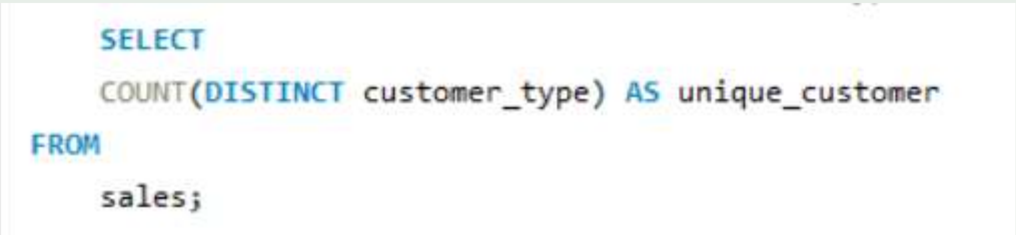
```
    SELECT
    customer_type, ROUND(AVG(VAT), 2) AS high_vat
FROM
    sales
GROUP BY customer_type
ORDER BY high_vat;
```



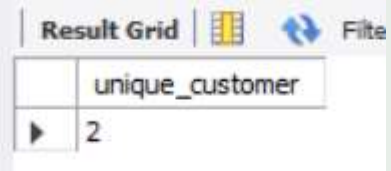
	customer_type	high_vat
▶	Normal	15.1
	Member	15.61

EDA

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

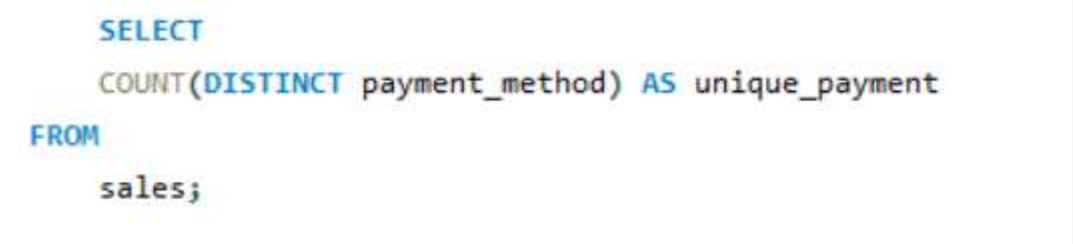


```
    SELECT
    COUNT(DISTINCT customer_type) AS unique_customer
FROM
    sales;
```

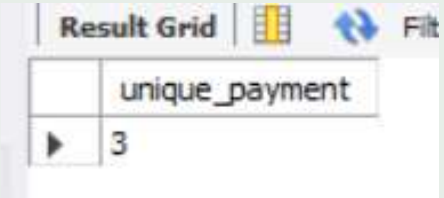


```
   unique_customer
▶  2
```

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

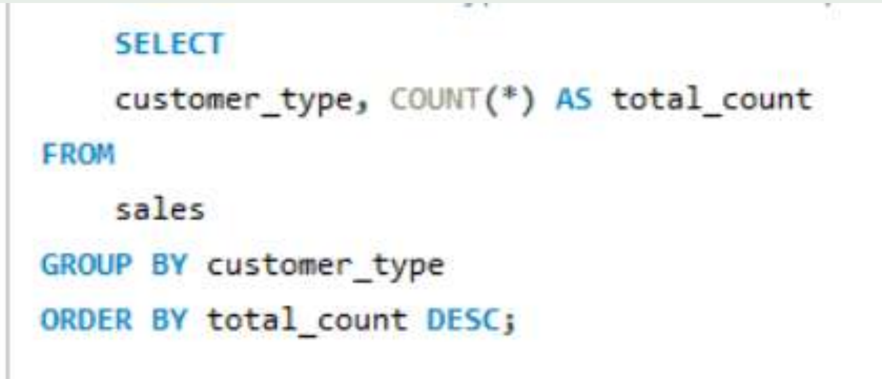


```
    SELECT
    COUNT(DISTINCT payment_method) AS unique_payment
FROM
    sales;
```

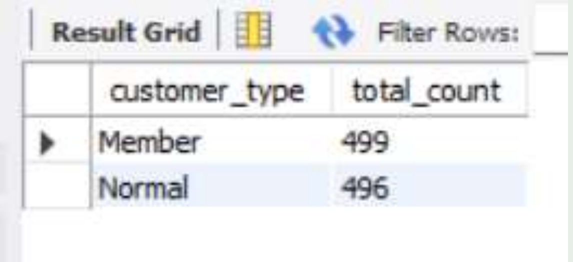


```
   unique_payment
▶  3
```

21. Which customer type occurs most frequently?



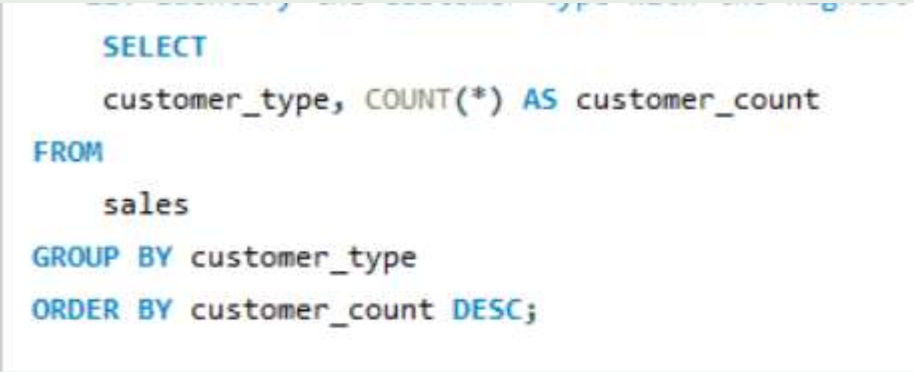
```
    SELECT
    customer_type, COUNT(*) AS total_count
FROM
    sales
GROUP BY customer_type
ORDER BY total_count DESC;
```



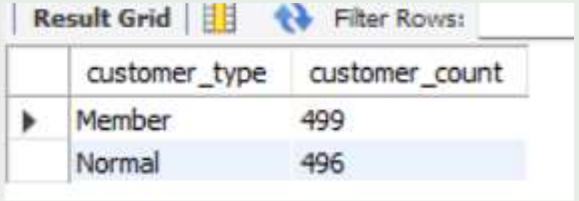
```
   customer_type   total_count
▶  Member          499
   Normal          496
```

EDA

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

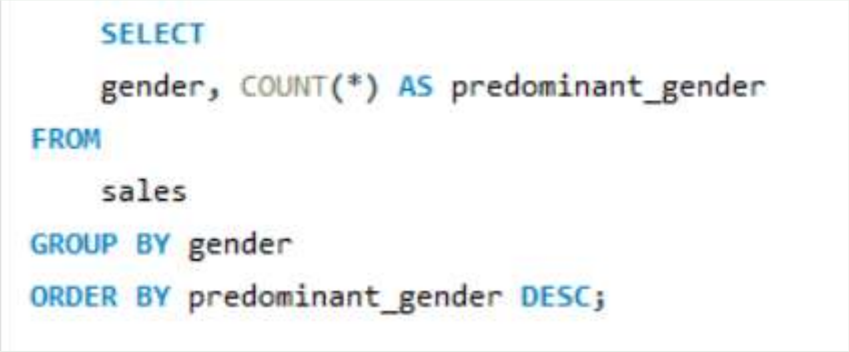


```
    SELECT
    customer_type, COUNT(*) AS customer_count
FROM
    sales
GROUP BY customer_type
ORDER BY customer_count DESC;
```

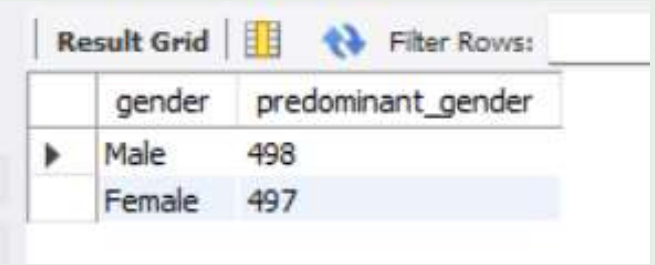


customer_type	customer_count
Member	499
Normal	496

23. Determine the predominant gender among customers.



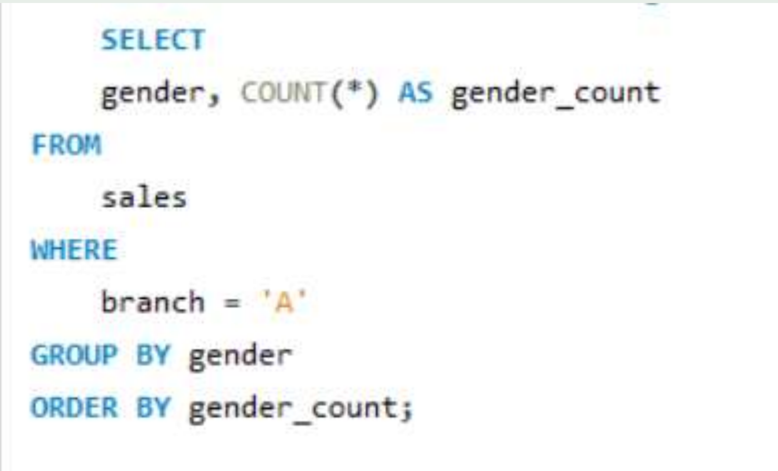
```
    SELECT
    gender, COUNT(*) AS predominant_gender
FROM
    sales
GROUP BY gender
ORDER BY predominant_gender DESC;
```



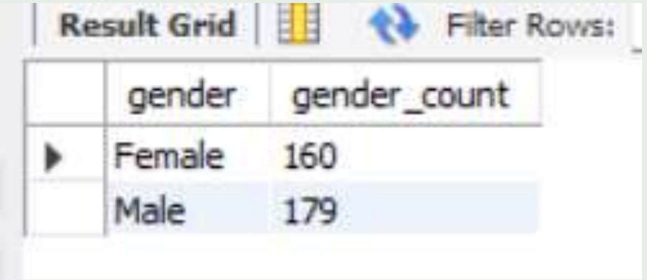
gender	predominant_gender
Male	498
Female	497

EDA

24. Examine the distribution of genders within each branch.

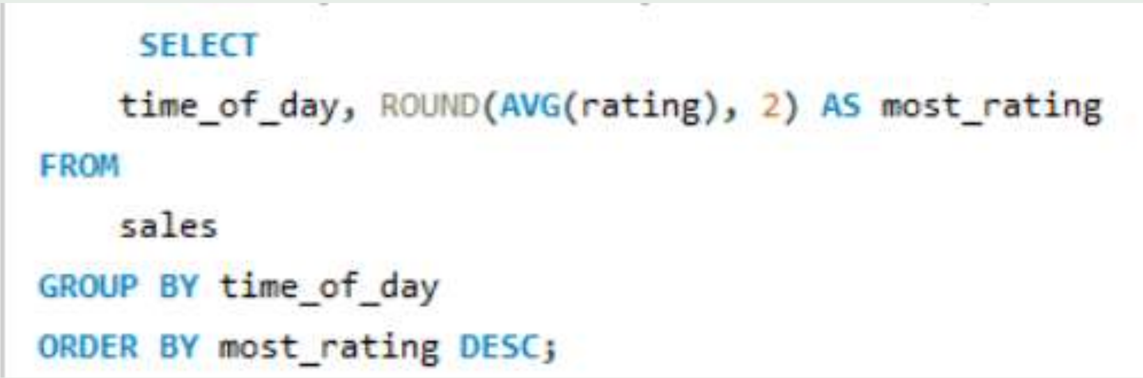


```
    SELECT
    gender, COUNT(*) AS gender_count
FROM
    sales
WHERE
    branch = 'A'
GROUP BY gender
ORDER BY gender_count;
```

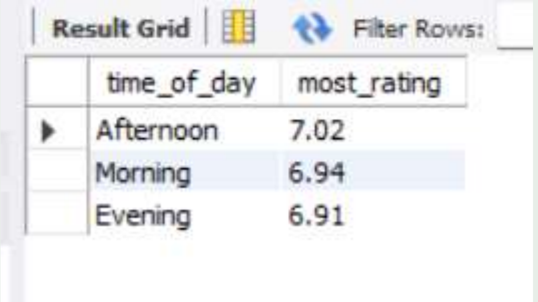


gender	gender_count
Female	160
Male	179

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



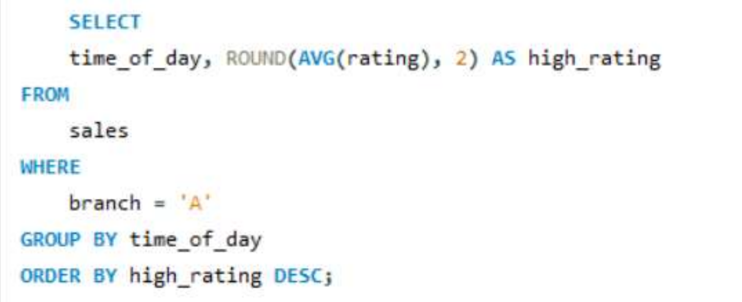
```
    SELECT
    time_of_day, ROUND(AVG(rating), 2) AS most_rating
FROM
    sales
GROUP BY time_of_day
ORDER BY most_rating DESC;
```



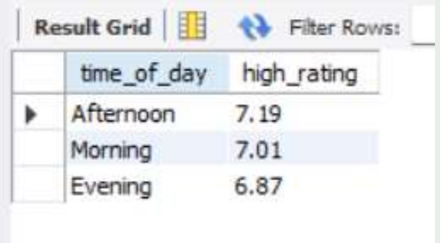
time_of_day	most_rating
Afternoon	7.02
Morning	6.94
Evening	6.91

EDA

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

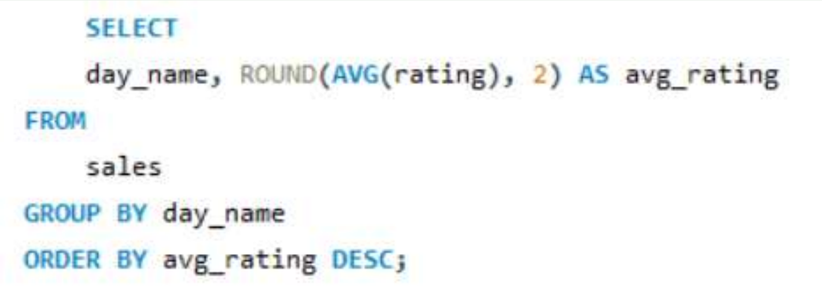


```
    SELECT
    time_of_day, ROUND(AVG(rating), 2) AS high_rating
FROM
    sales
WHERE
    branch = 'A'
GROUP BY time_of_day
ORDER BY high_rating DESC;
```

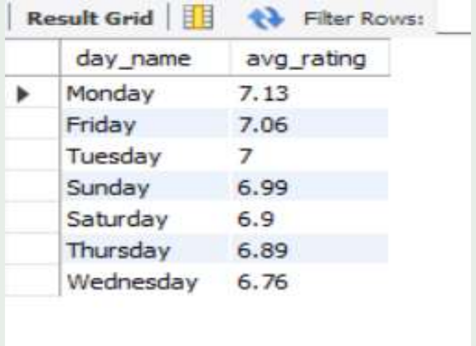


time_of_day	high_rating
Afternoon	7.19
Morning	7.01
Evening	6.87

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

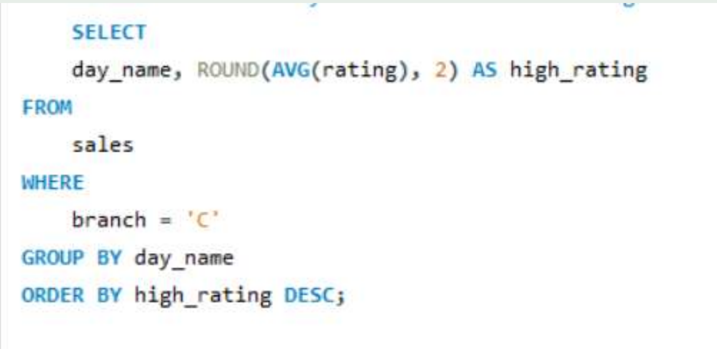


```
    SELECT
    day_name, ROUND(AVG(rating), 2) AS avg_rating
FROM
    sales
GROUP BY day_name
ORDER BY avg_rating DESC;
```

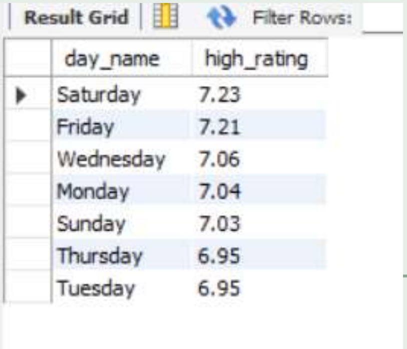


day_name	avg_rating
Monday	7.13
Friday	7.06
Tuesday	7
Sunday	6.99
Saturday	6.9
Thursday	6.89
Wednesday	6.76

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



```
    SELECT
    day_name, ROUND(AVG(rating), 2) AS high_rating
FROM
    sales
WHERE
    branch = 'C'
GROUP BY day_name
ORDER BY high_rating DESC;
```



day_name	high_rating
Saturday	7.23
Friday	7.21
Wednesday	7.06
Monday	7.04
Sunday	7.03
Thursday	6.95
Tuesday	6.95

Insights

Product Analysis

- ✓ Identification of product line : The analysis identified several product lines available in Amazon sales data including city and customer type.
- ✓ Best performing product line : Based on sales volume and revenue Fashion accessories has highest sales and Food and beverages has highest revenue.
- ✓ Product lines needs to improvement : Other product line requires improvement such as Health and beauty in sales and revenue. Health and beauty considered as low rating. So we nee to focus on this particular product line.

Sales Analysis

- ✓ Sales trend over time : Here we can observe in January month highest revenue generated. In weekdays Monday has highest rating according to product sold
- ✓ Areas for improvement Have to focus February month sales through target efforts.

Customer Analysis

- ✓ As we can see Female are more interested in fashion accessories and Male are interested in health and beauty.
- ✓ On weekend customer prefers to shop at evening
- ✓ Total revenue generated maximum for membership customers
- ✓ Customer prefer to rating their products mostly at afternoon time.

Thank you

I appreciate the opportunity to share this project with you