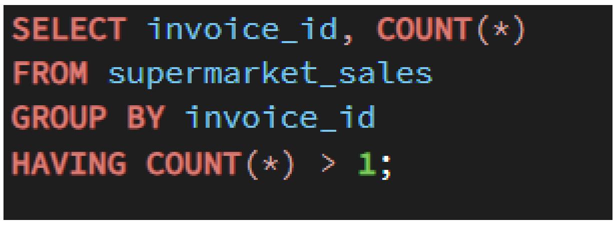
Task 4: - SQL Task Submission

Objective: Use SQL queries to extract and analyze data from a database.

Tools: - PostgreSQL

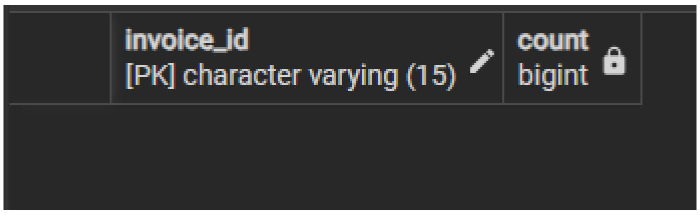
1. **Data Preprocessing: -**

   - Check for duplicates and ensure data consistency.
     Query: -



```
SELECT invoice_id, COUNT(*)
FROM supermarket_sales
GROUP BY invoice_id
HAVING COUNT(*) > 1;
```

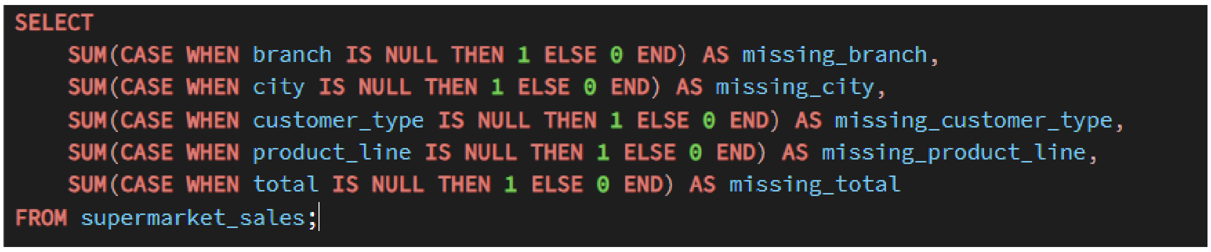
Output: -



	invoice_id [PK] character varying (15)	count bigint

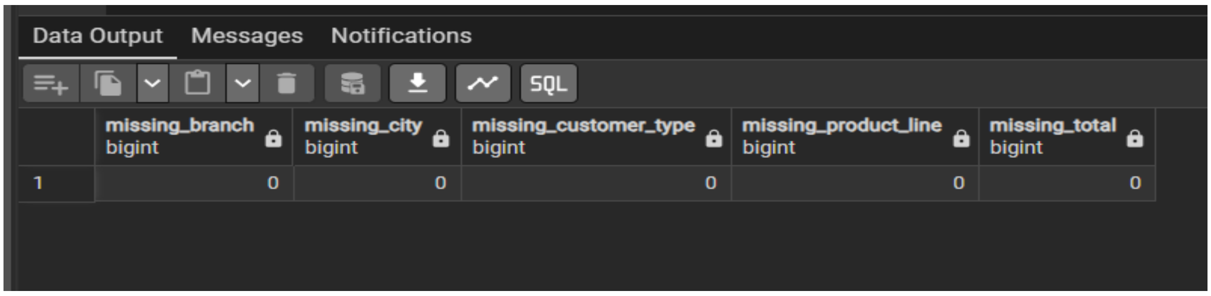
So, there is no duplicates values are present in the dataset.

   - Handling Missing value, if any.
     Query: -



```
SELECT
    SUM(CASE WHEN branch IS NULL THEN 1 ELSE 0 END) AS missing_branch,
    SUM(CASE WHEN city IS NULL THEN 1 ELSE 0 END) AS missing_city,
    SUM(CASE WHEN customer_type IS NULL THEN 1 ELSE 0 END) AS missing_customer_type,
    SUM(CASE WHEN product_line IS NULL THEN 1 ELSE 0 END) AS missing_product_line,
    SUM(CASE WHEN total IS NULL THEN 1 ELSE 0 END) AS missing_total
FROM supermarket_sales;
```

Output:



	missing_branch bigint	missing_city bigint	missing_customer_type bigint	missing_product_line bigint	missing_total bigint
1	0	0	0	0	0

- Convert data and time fields to appropriate SQL formats.

  Query: -

```sql
ALTER TABLE supermarket_sales
ALTER COLUMN sale_date TYPE DATE USING TO_DATE(sale_date, 'MM/DD/YYYY');

ALTER TABLE supermarket_sales
ALTER COLUMN sale_time TYPE TIME USING sale_time::TIME;
```

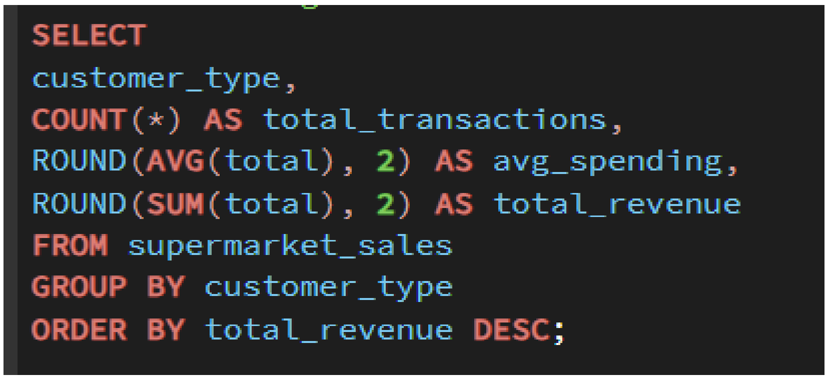
Output: -

I change date because when I try to import data it shows me wrong date format so firstly, I change it to text format. Then I import data. After that I change it to Proper date format.

2. Exploratory Data Analysis (EDA) using SQL Queries: -

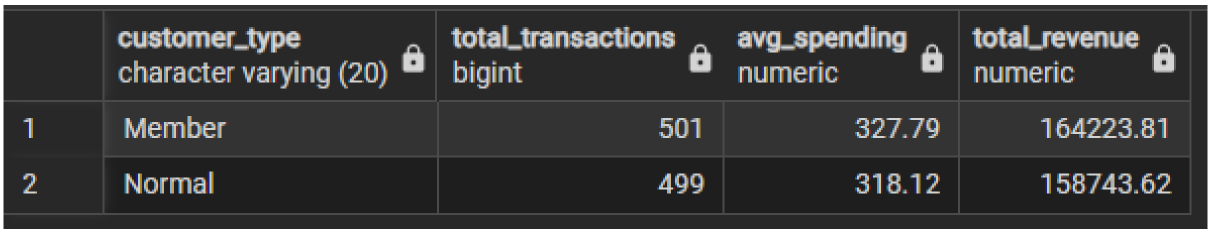
- **Customer Segmentation:** Count transactions by customer type, analyze average spending.

  Query: -



```
SELECT
customer_type,
COUNT(*) AS total_transactions,
ROUND(AVG(total), 2) AS avg_spending,
ROUND(SUM(total), 2) AS total_revenue
FROM supermarket_sales
GROUP BY customer_type
ORDER BY total_revenue DESC;
```

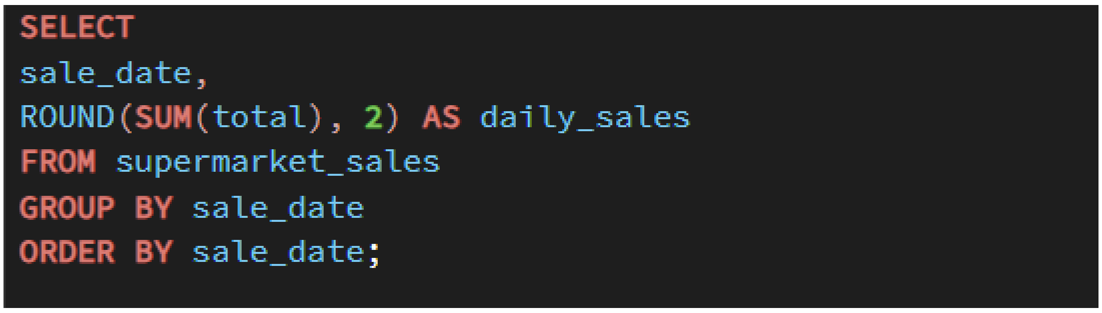
  Output: -



	customer_type character varying (20)	total_transactions bigint	avg_spending numeric	total_revenue numeric
1	Member	501	327.79	164223.81
2	Normal	499	318.12	158743.62

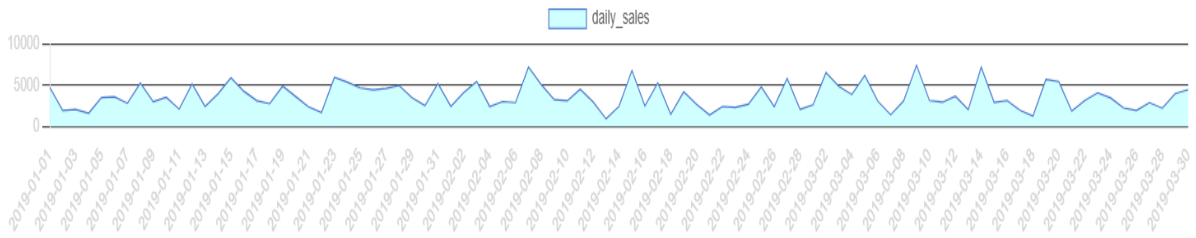
- **Sales Trend Analysis:** Identify sales performance over time, peak days, and time slots.

  Query: - Sales Over Time

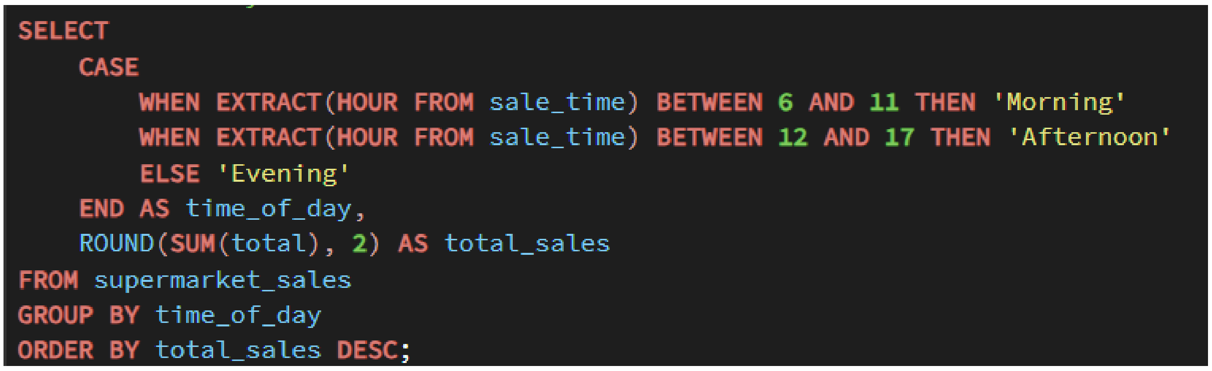


```
SELECT
sale_date,
ROUND(SUM(total), 2) AS daily_sales
FROM supermarket_sales
GROUP BY sale_date
ORDER BY sale_date;
```

Output: -

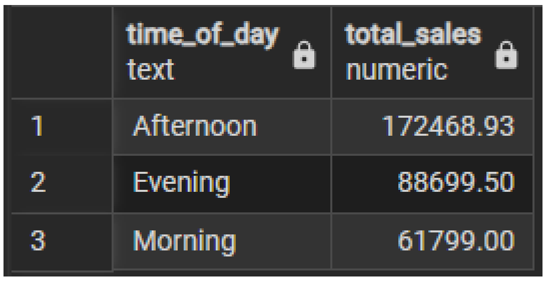


Query: - Peak Sales by Time Slot



```sql
SELECT
    CASE
        WHEN EXTRACT(HOUR FROM sale_time) BETWEEN 6 AND 11 THEN 'Morning'
        WHEN EXTRACT(HOUR FROM sale_time) BETWEEN 12 AND 17 THEN 'Afternoon'
        ELSE 'Evening'
    END AS time_of_day,
    ROUND(SUM(total), 2) AS total_sales
FROM supermarket_sales
GROUP BY time_of_day
ORDER BY total_sales DESC;
```

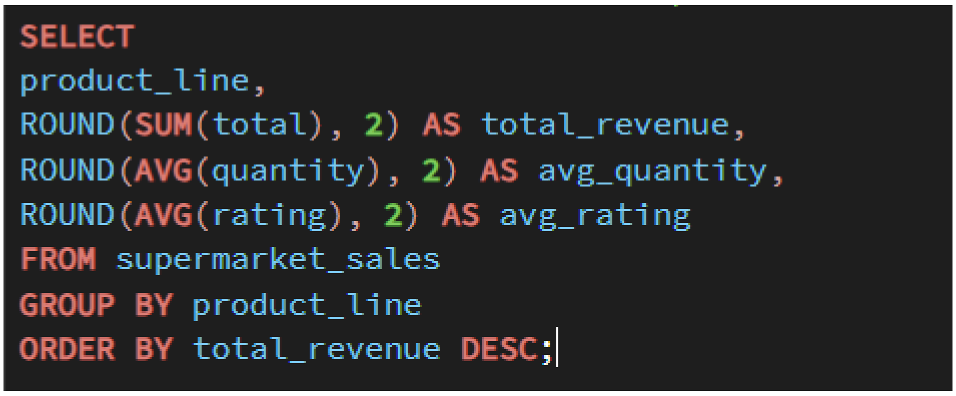
Output: -



	time_of_day text	total_sales numeric
1	Afternoon	172468.93
2	Evening	88699.50
3	Morning	61799.00

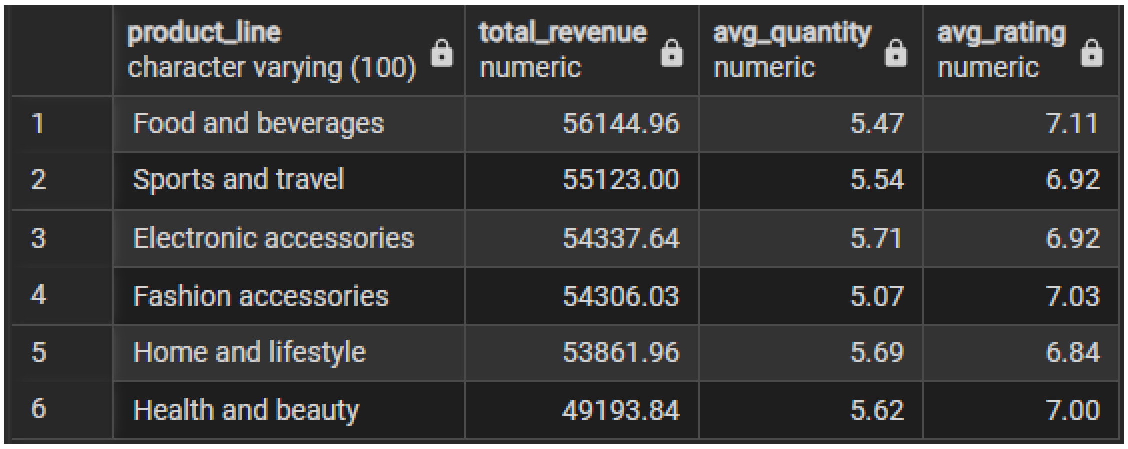
- **Product Line Performance:** Rank product lines by revenue, calculate average quantity sold.

Query: -



```sql
SELECT
product_line,
ROUND(SUM(total), 2) AS total_revenue,
ROUND(AVG(quantity), 2) AS avg_quantity,
ROUND(AVG(rating), 2) AS avg_rating
FROM supermarket_sales
GROUP BY product_line
ORDER BY total_revenue DESC;
```

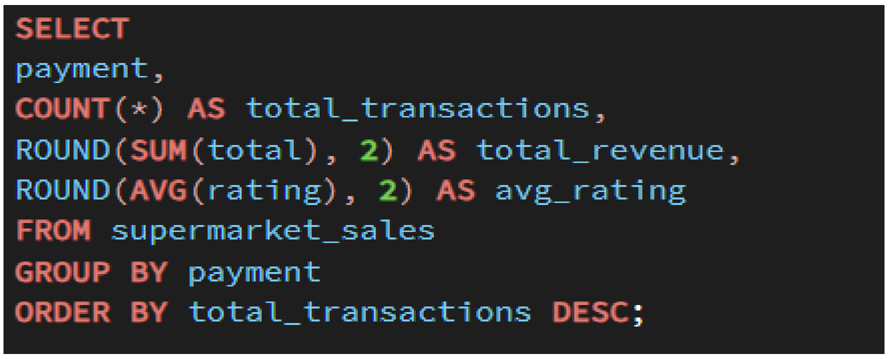
Output: -



	product_line character varying (100)	total_revenue numeric	avg_quantity numeric	avg_rating numeric
1	Food and beverages	56144.96	5.47	7.11
2	Sports and travel	55123.00	5.54	6.92
3	Electronic accessories	54337.64	5.71	6.92
4	Fashion accessories	54306.03	5.07	7.03
5	Home and lifestyle	53861.96	5.69	6.84
6	Health and beauty	49193.84	5.62	7.00

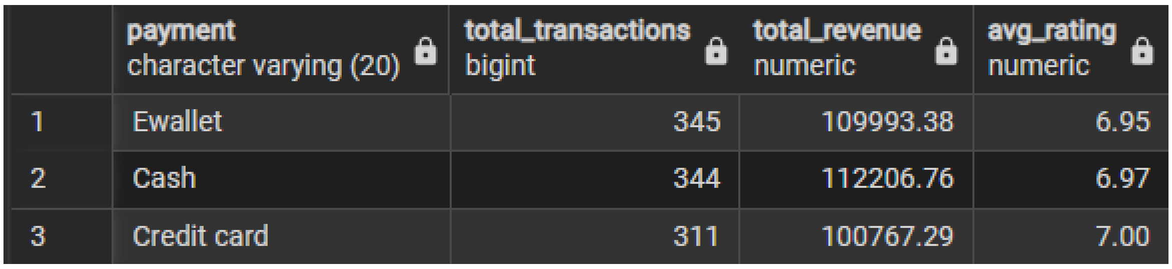
- **Payment Method Insights:** Identify preferred payment methods and correlation with satisfaction.

Query: -



```sql
SELECT
payment,
COUNT(*) AS total_transactions,
ROUND(SUM(total), 2) AS total_revenue,
ROUND(AVG(rating), 2) AS avg_rating
FROM supermarket_sales
GROUP BY payment
ORDER BY total_transactions DESC;
```

Output: -

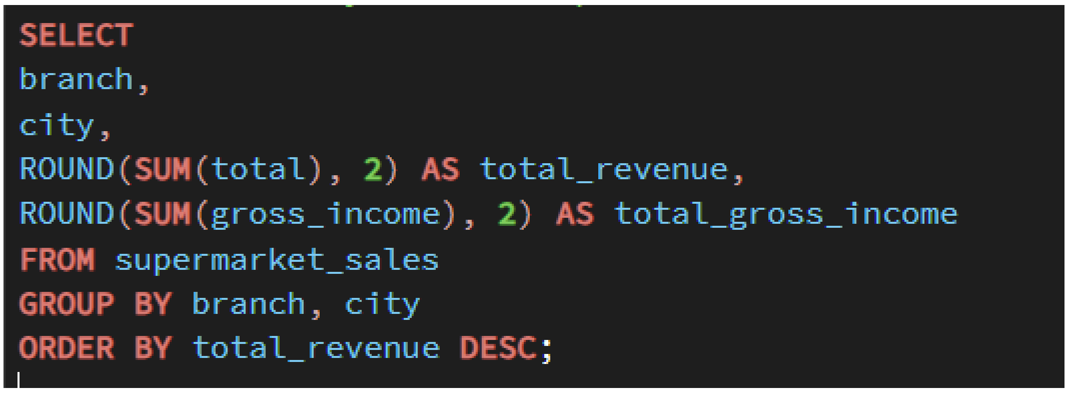


	payment character varying (20)	total_transactions bigint	total_revenue numeric	avg_rating numeric
1	Ewallet	345	109993.38	6.95
2	Cash	344	112206.76	6.97
3	Credit card	311	100767.29	7.00

3. Performance Analysis Using SQL:

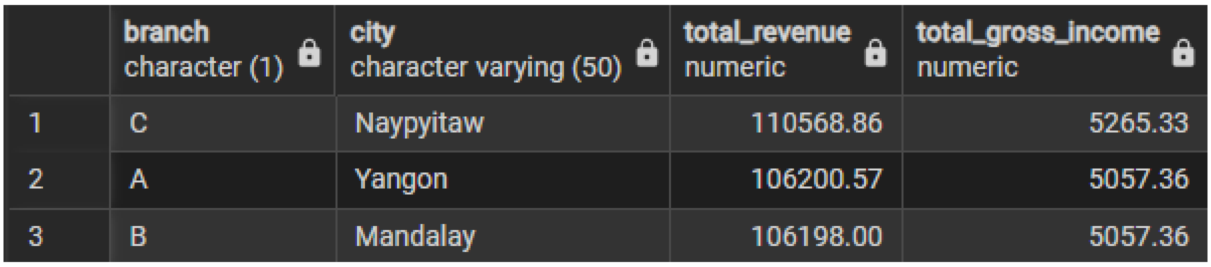
- **Branch and City Sales:** Compare revenue across branches and cities

  Query: -



```
SELECT
branch,
city,
ROUND(SUM(total), 2) AS total_revenue,
ROUND(SUM(gross_income), 2) AS total_gross_income
FROM supermarket_sales
GROUP BY branch, city
ORDER BY total_revenue DESC;
```

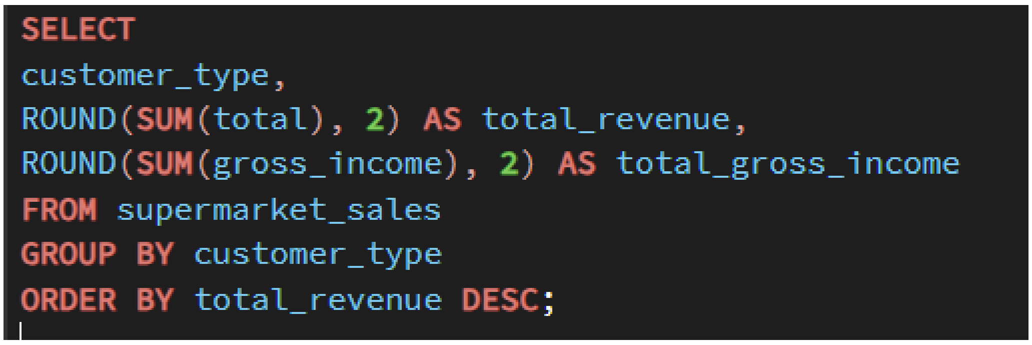
  Output: -



	branch character (1)	city character varying (50)	total_revenue numeric	total_gross_income numeric
1	C	Naypyitaw	110568.86	5265.33
2	A	Yangon	106200.57	5057.36
3	B	Mandalay	106198.00	5057.36

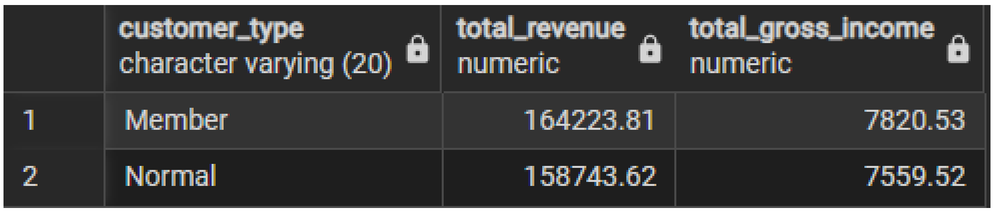
- **Customer Type Revenue Contribution:** Analyze revenue from members vs. normal customers.

  Query: -



```
SELECT
customer_type,
ROUND(SUM(total), 2) AS total_revenue,
ROUND(SUM(gross_income), 2) AS total_gross_income
FROM supermarket_sales
GROUP BY customer_type
ORDER BY total_revenue DESC;
```

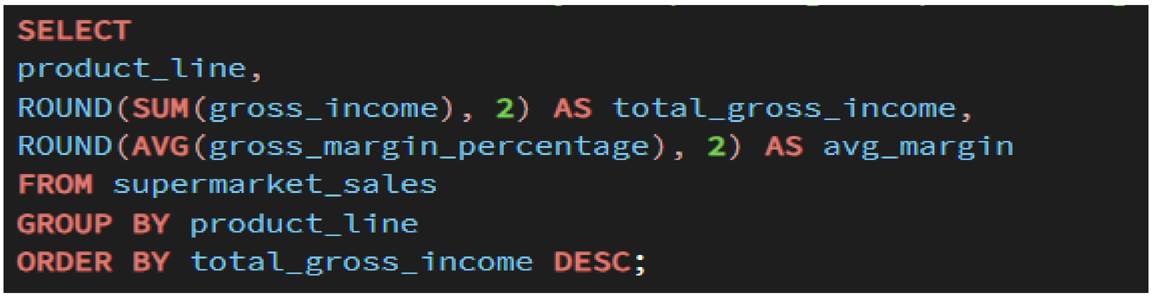
Output: -



	customer_type character varying (20)	total_revenue numeric	total_gross_income numeric
1	Member	164223.81	7820.53
2	Normal	158743.62	7559.52

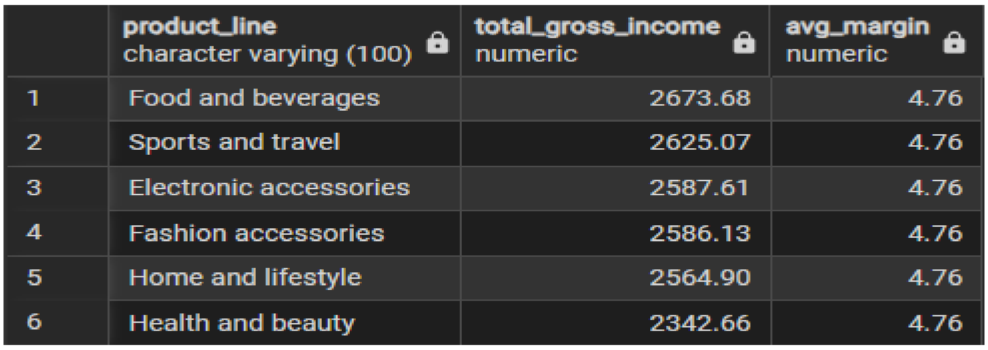
- **Product Line Profitability:** Compute highest profit margins by category.

Query: -



```
SELECT
product_line,
ROUND(SUM(gross_income), 2) AS total_gross_income,
ROUND(AVG(gross_margin_percentage), 2) AS avg_margin
FROM supermarket_sales
GROUP BY product_line
ORDER BY total_gross_income DESC;
```

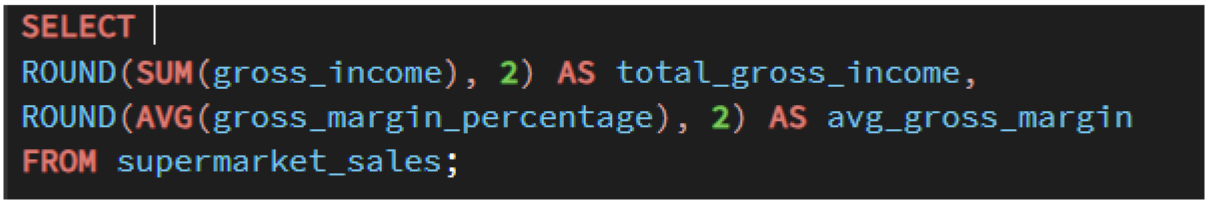
Output: -



	product_line character varying (100)	total_gross_income numeric	avg_margin numeric
1	Food and beverages	2673.68	4.76
2	Sports and travel	2625.07	4.76
3	Electronic accessories	2587.61	4.76
4	Fashion accessories	2586.13	4.76
5	Home and lifestyle	2564.90	4.76
6	Health and beauty	2342.66	4.76

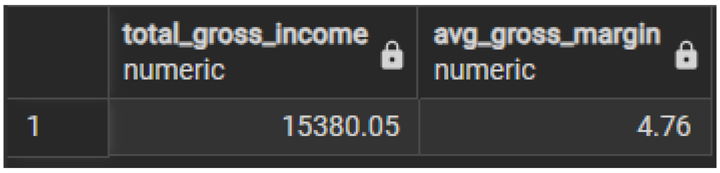
- **Gross Income & Margin Analysis:** Calculate total gross income and gross margin percentages.

  Query: -



```sql
SELECT
ROUND(SUM(gross_income), 2) AS total_gross_income,
ROUND(AVG(gross_margin_percentage), 2) AS avg_gross_margin
FROM supermarket_sales;
```

  Output: -

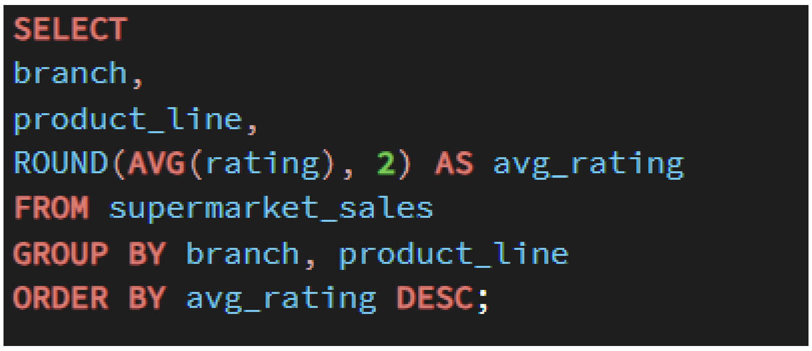


	total_gross_income numeric	avg_gross_margin numeric
1	15380.05	4.76

4. Customer Satisfaction Analysis:

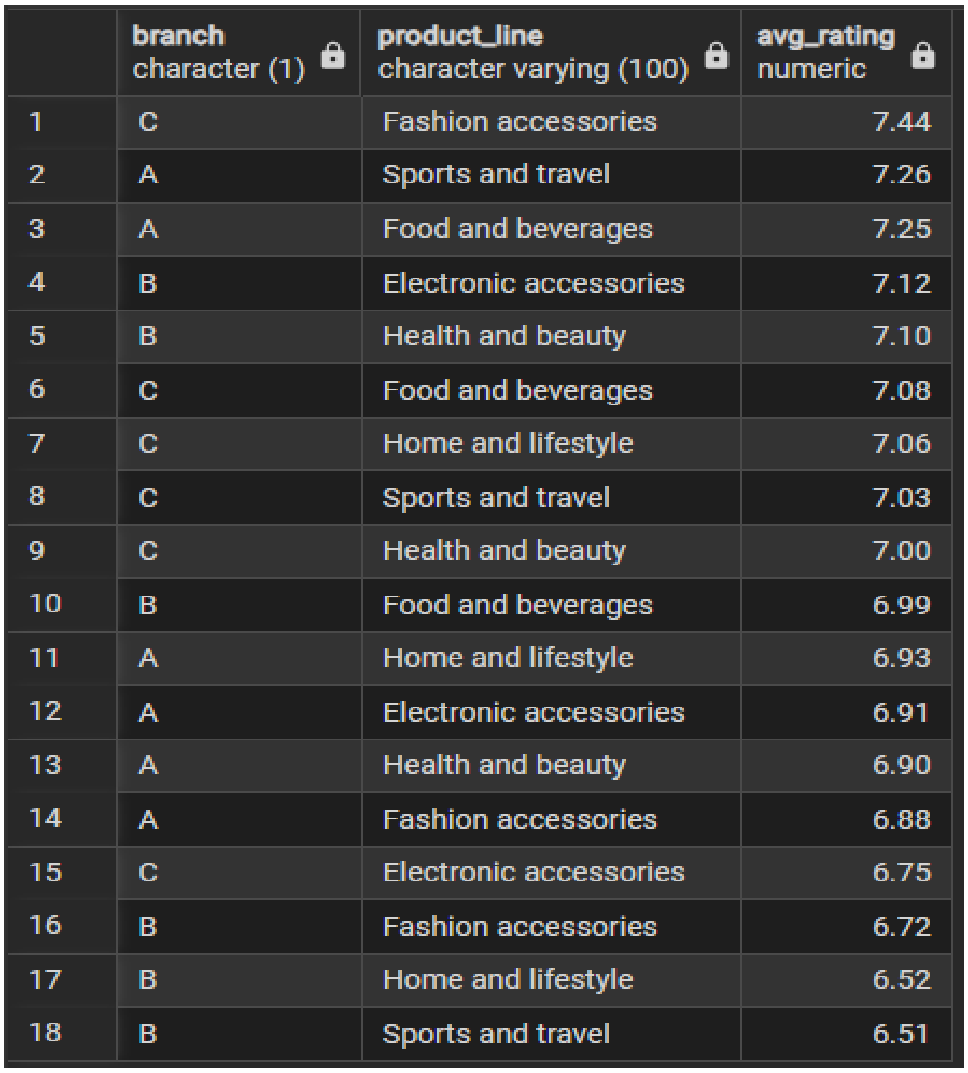
- Analyze ratings by product line and branch.

Query: -



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

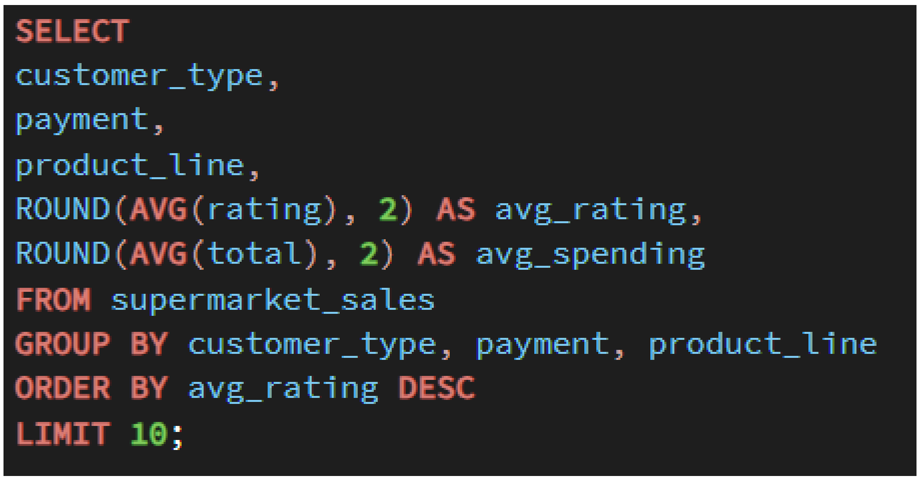
Output: -



	branch character (1)	product_line character varying (100)	avg_rating numeric
1	C	Fashion accessories	7.44
2	A	Sports and travel	7.26
3	A	Food and beverages	7.25
4	B	Electronic accessories	7.12
5	B	Health and beauty	7.10
6	C	Food and beverages	7.08
7	C	Home and lifestyle	7.06
8	C	Sports and travel	7.03
9	C	Health and beauty	7.00
10	B	Food and beverages	6.99
11	A	Home and lifestyle	6.93
12	A	Electronic accessories	6.91
13	A	Health and beauty	6.90
14	A	Fashion accessories	6.88
15	C	Electronic accessories	6.75
16	B	Fashion accessories	6.72
17	B	Home and lifestyle	6.52
18	B	Sports and travel	6.51

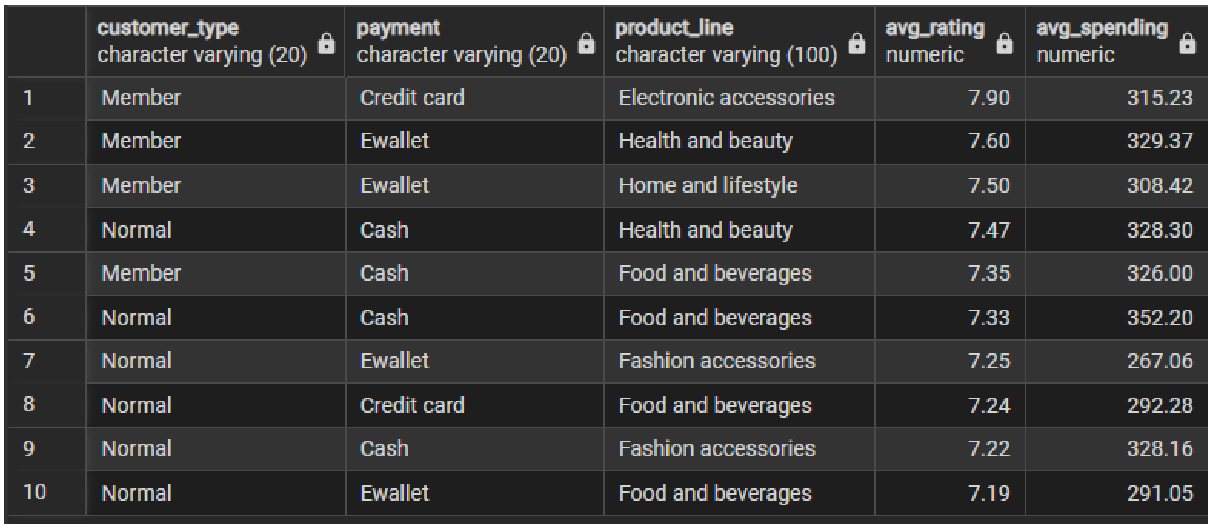
- Identify factors influencing higher satisfaction scores.

Query: -



```
SELECT
customer_type,
payment,
product_line,
ROUND(AVG(rating), 2) AS avg_rating,
ROUND(AVG(total), 2) AS avg_spending
FROM supermarket_sales
GROUP BY customer_type, payment, product_line
ORDER BY avg_rating DESC
LIMIT 10;
```

Output: -



	customer_type character varying (20)	payment character varying (20)	product_line character varying (100)	avg_rating numeric	avg_spending numeric
1	Member	Credit card	Electronic accessories	7.90	315.23
2	Member	Ewallet	Health and beauty	7.60	329.37
3	Member	Ewallet	Home and lifestyle	7.50	308.42
4	Normal	Cash	Health and beauty	7.47	328.30
5	Member	Cash	Food and beverages	7.35	326.00
6	Normal	Cash	Food and beverages	7.33	352.20
7	Normal	Ewallet	Fashion accessories	7.25	267.06
8	Normal	Credit card	Food and beverages	7.24	292.28
9	Normal	Cash	Fashion accessories	7.22	328.16
10	Normal	Ewallet	Food and beverages	7.19	291.05