**Retail Sales Analysis SQL Project**

# Objective:

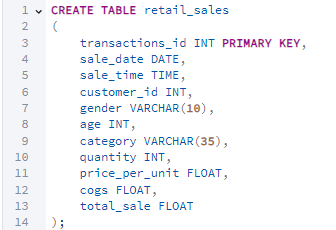
The objective of this project is to create, populate, and analyze a retail sales database to derive valuable insights that can inform business decisions. The project will involve the following steps:

1. Database Setup: Create a structured retail sales database to store sales transactions, products, customers, and related data.
2. Data Cleaning: Identify and remove records with missing or null values to ensure that the dataset is accurate and reliable for analysis.
3. Exploratory Data Analysis (EDA): Perform an initial exploration of the dataset to understand trends, patterns, and relationships among variables such as product sales, customer behavior, and time periods. This step will include summary statistics, data visualizations, and correlation analysis.
4. Business Analysis: Use SQL queries to answer key business questions such as sales performance, popular products, customer buying patterns, and overall profitability. The insights gained from this analysis will help in making data-driven decisions to optimize sales strategies and improve business operations.

# Key business problems:

1. How can we track daily sales performance to understand revenue trends on a specific date 11 November 2022?
2. How can we identify high-performing product categories, specifically for clothing, to optimize inventory for the month of November 2022?
3. What is the total sales contribution of each product category, and how can we use this to drive targeted marketing strategies?
4. Can we determine the average age of customers purchasing beauty products to create more personalized marketing campaigns?
5. How do we identify high-value transactions where sales exceed a specific threshold to analyze premium customer behavior?
6. What insights can we gain from the number of transactions segmented by gender and product category to enhance our customer segmentation efforts?
7. How can we analyze order volume across different times of the day (morning, afternoon, evening) to optimize staffing and inventory management?
8. Who are our top 5 customers based on total sales, and how can we leverage this data to increase customer retention and loyalty?
9. How many unique customers purchase from each category, and what does this indicate about category-specific customer engagement?
10. What is the average monthly sales performance, and which month stands out as the best-selling month each year? How can this data inform our seasonal sales strategies?

# Creating a table and data import:



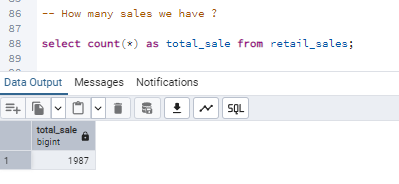
# Handling Missing Values:

# 

I found that there 13 rows that had missing values in them. I deleted all the 13 rows. The null value rows were deleted because it wasn’t significant to the total number of rows.

# Data Analysis & Findings:

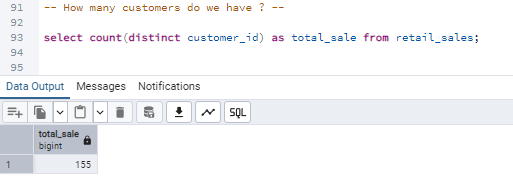
## **Exploratory Data Analysis:**



Q1. How many sales we have?

### **Findings:**

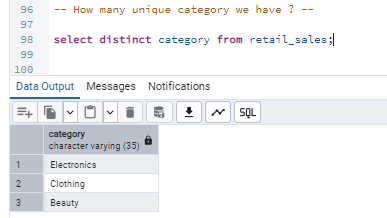
The total sales we have here is 1987.



Q2. How many customers do we have?

### **Findings:**

Total number of customers we have here is 155.



Q3. How many unique category we have ?

### **Findings:**

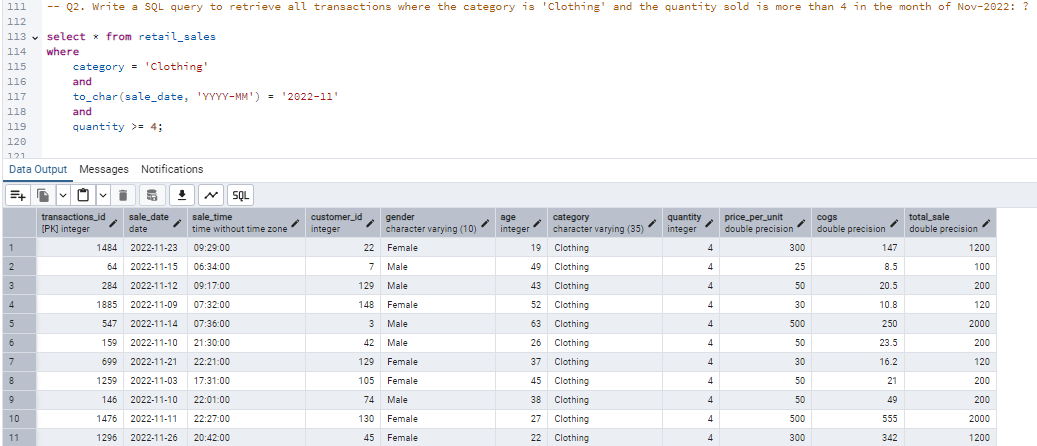
There are 3 unique categories Electronics, Clothing, Beauty.

# Key business problems and solution:

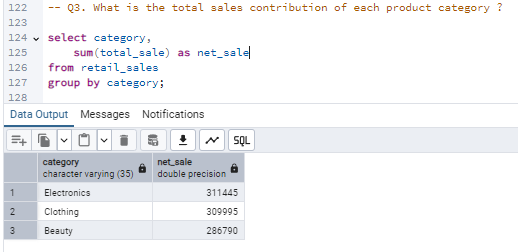
Q1. Write a SQL query to retrieve all columns for sales made on '2022-11-05?

# 

Q2. Write a SQL query to retrieve all transactions where the category is 'Clothing' and the quantity sold is more than 4 in the month of Nov-2022?



Q3. What is the total sales contribution of each product category?



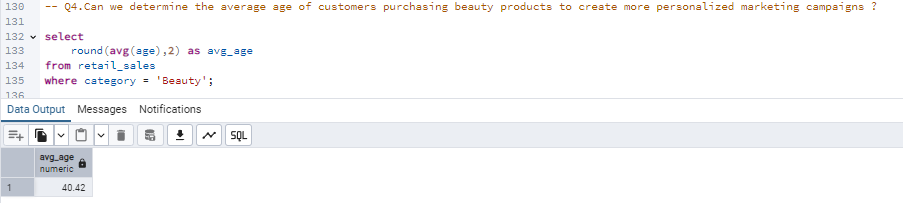
Findings:

Electronics: 311445

Clothing: 309995

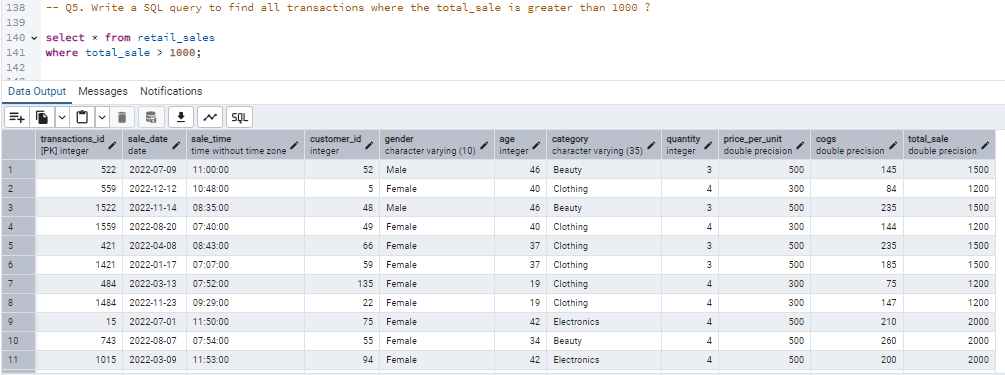
Beauty: 286790

Q4. Can we determine the average age of customers purchasing beauty products to create more personalized marketing campaigns?

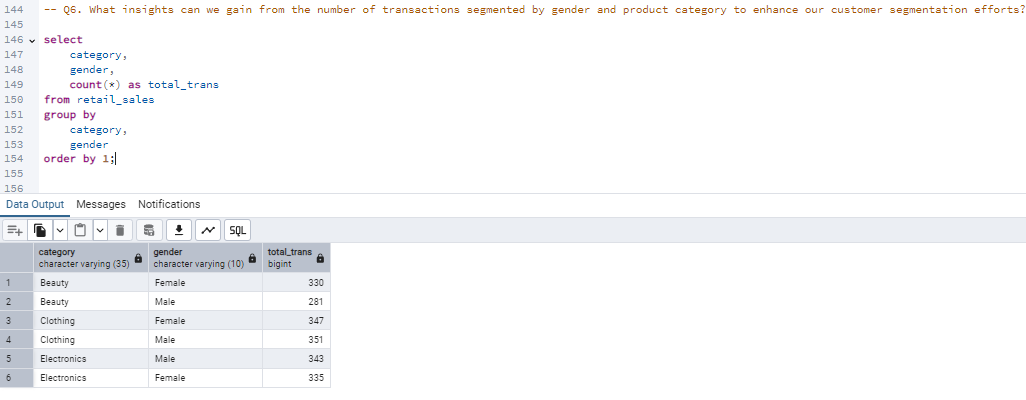


The average age here is around 40 years.

Q5. Write a SQL query to find all transactions where the total\_sale is greater than 1000?



Q6. What insights can we gain from the number of transactions segmented by gender and product category to enhance our customer segmentation efforts?



**Beauty Category:**

Females had 330 transactions, higher than males with 281.

Suggests stronger female engagement with this segment.

**Clothing Category:**

Females led with 347 transactions, slightly more than males at 351.

Indicates a more balanced interest between genders.

**Electronics Category:**

Males recorded 343 transactions, narrowly surpassing females at 335.

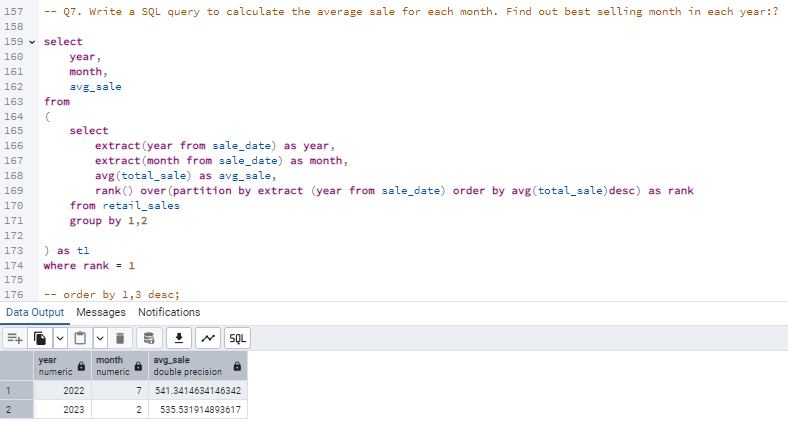
Suggests relatively equal interest, with a slight male preference.

Marketing Insights:

These patterns can be used to create targeted promotions and personalized recommendations.

Helps align marketing strategies with customer preferences across product categories.

Q7. Write a SQL query to calculate the average sale for each month. Find out bestselling month in each year?

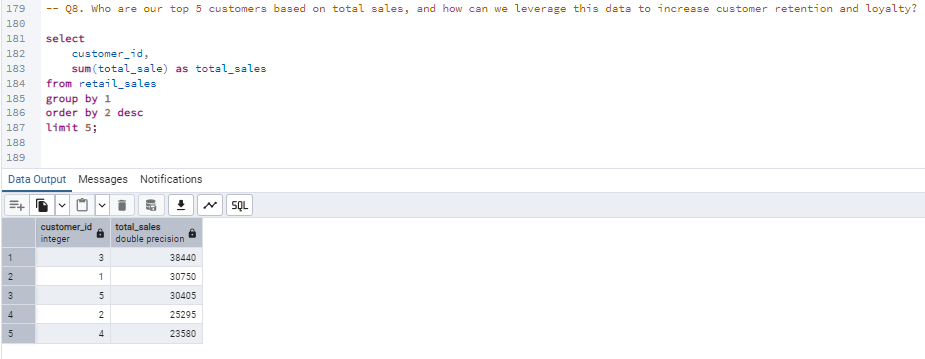


In 2022, the best-selling month was July, with an average sale of approximately 541.34.

In 2023, the best-selling month was February, with an average sale of approximately 535.53.

The results suggest that mid-year (July) and early-year (February) are strong sales periods for the respective years analyzed.

Q8. Who are our top 5 customers based on total sales, and how can we leverage this data to increase customer retention and loyalty?



**Findings:**

**Customer ID 3** is the top customer with total sales of 38,440.

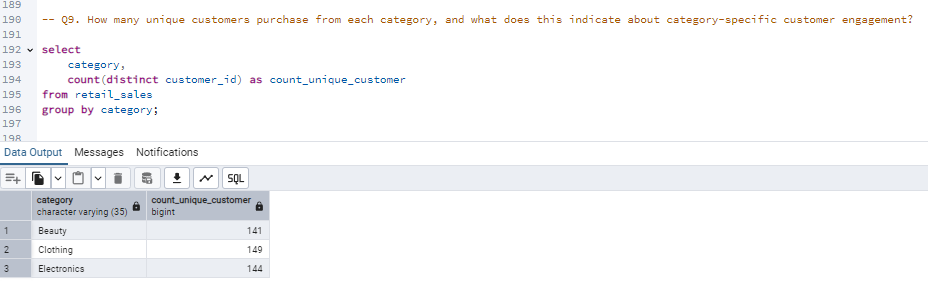
The subsequent top customers are ID 1 (30,750), ID 5 (30,405), ID 2 (25,295), and ID 4 (23,880).

These customers contribute significantly to the overall sales and are crucial for business revenue.

Leveraging this data, the company can focus on personalized marketing, loyalty programs, and exclusive offers to enhance customer retention and further increase their loyalty.

Understanding the purchasing patterns of these top customers can also help in tailoring products and services to meet their preferences, thereby fostering long-term relationships.

Q9. How many unique customers purchase from each category, and what does this indicate about category-specific customer engagement?



**Findings:**

The "Clothing" category has the highest number of unique customers at 149.

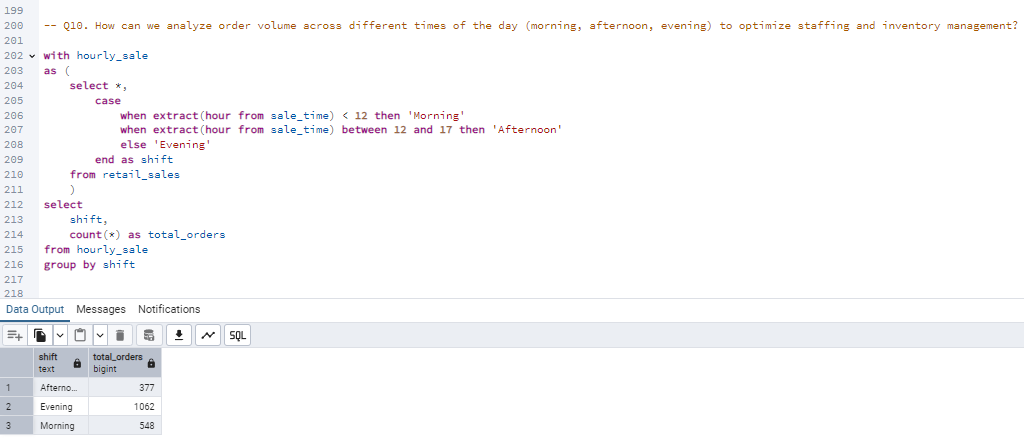
The "Electronics" category follows with 144 unique customers.

The "Beauty" category has 141 unique customers.

These numbers indicate that all categories have relatively similar levels of customer engagement.

The slight variation suggests that while "Clothing" is slightly more popular, "Beauty" and "Electronics" also attract a significant number of customers.

Q10. How can we analyze order volume across different times of the day (morning, afternoon, evening) to optimize staffing and inventory management?



**Morning Shift (before 12 PM):**

Total orders: 548.

Indicates steady activity, suggesting a good time for regular staffing and inventory checks.

**Afternoon Shift (12 PM - 5 PM):**

Total orders: 377.

Lowest order volume, implying a potential window for breaks, restocking, or lighter staffing.

Evening Shift (after 5 PM):

Total orders: 1062.

Peak order period, highlighting the need for maximum staffing and inventory availability.

Operational Insight:

Staffing and inventory management can be optimized by aligning resources with these demand patterns.

Consider promotions or faster service strategies to handle the evening rush efficiently.