**CASE STUDY: “Sales Performance and Customer Segmentation (RFM) Analysis for Strategic Growth”**

**Introduction**

In today’s fast-moving retail world, understanding how customers behave is key to growing sales and boosting performance. This case study demonstrates how RFM-based customer segmentation, combined with sales performance metrics, reveal customer behavior and guide smarter business decisions.

Using a fictional retail dataset, I analyze transaction data to identify high-value customers, reveal spending habits, and assess their influence on overall business performance. The goal is to translate raw data into actionable strategies for improving customer targeting, retention, and marketing efficiency.

**Business Challenge**

A mid-sized retail business wants to boost revenue by understanding customer purchasing patterns and identifying its most valuable shoppers. The leadership team suspects that marketing efforts are too broad, and they’re missing opportunities to personalize outreach and improve retention.

The goal of this analysis is to uncover seasonal trends and segment customers based on Recency, Frequency, and Monetary (RFM) value. The aim is to evaluate how each customer segment contributes to overall sales performance, guiding smarter decisions in marketing, inventory management, and customer engagement strategies.

The analysis will follow the 6 phases of the Data Analysis process: Ask, Prepare, Process, Analyze, Share and Act (APPASA).

***ASK***

**Key Questions**

1. Sales Performance and Customer Demographics Analysis
2. Which month exhibit peak sales performance, and what key factors contribute to those peaks?
3. **Which months have higher or lower customer purchase activity?**
4. How does customer demographics such as age and gender influence spending and product preferences?
5. Which product contributes the most to revenue?
6. Customer Segmentation (RFM Analysis)
   1. What percentage of customers belong to the high-value segment? (e.g., Champions, Loyal Customers)?
   2. Which RFM segments contribute the most to overall revenue?
   3. How can we re-engage at-risk or low engagement customers?

***PREPARE***

This dataset gives a quick look at a fictional retail setting, showing key details about how the business runs and how customers behave.

**Data Overview**

* Data source: Kaggle.com
* Dataset: retail sales
* Dataset size: 1,000 rows × 9 columns
* Unique values: 1000 transaction ID

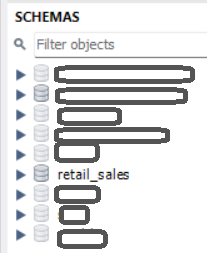
**Key fields:**

* Transaction ID – unique identifier for each purchase
* Date – transaction date
* Customer ID – unique identifier for each customer
* Gender – customer gender
* Age – customer age
* Product Category – category of purchased product
* Quantity – number of items purchased
* Price per Unit – cost of a single item
* Total Amount – total sales

**Tools used**

* MySQL: data cleaning, RFM scoring, JOINS, and view creation
* Excel: preliminary preprocessing and exploratory checks (pivot table)

Created a database named retail\_sales and imported the retail dataset to the database.



Created and used a staging table to keep the original data untouched while cleaning and analyzing it.

SQL Query:



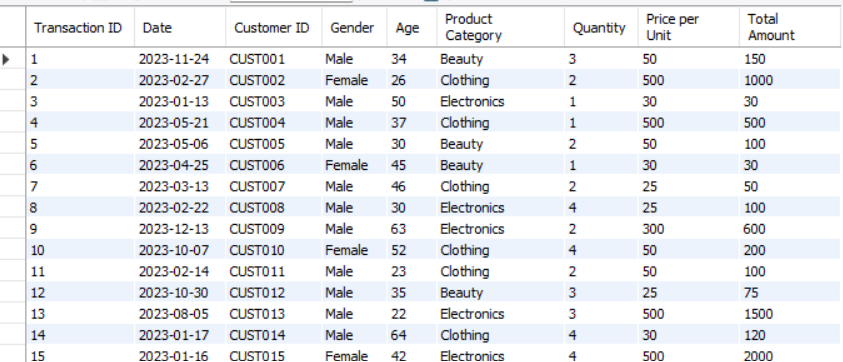
View staging table

SQL Query:

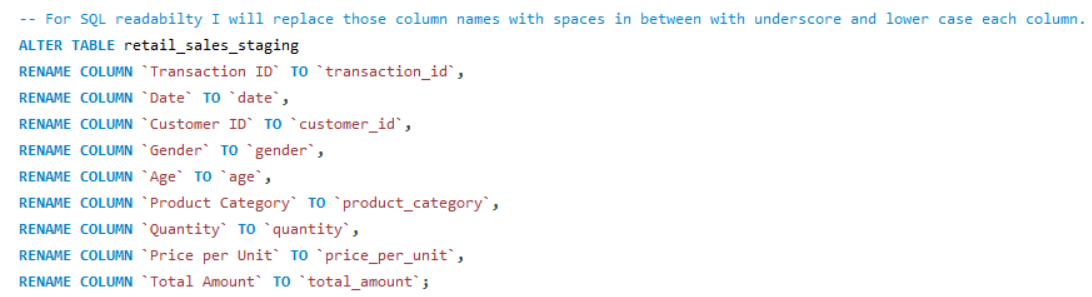


Output:

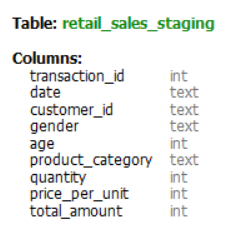
First 15 rows



Checking if the data is clean. If not, I will clean it first before exploring the dataset.



Output:

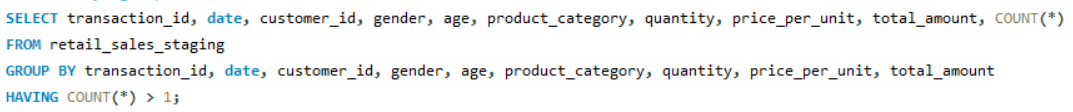


***PROCESS***

In data cleaning, I followed some steps:

1. **Removed duplicates**

SQL Query:

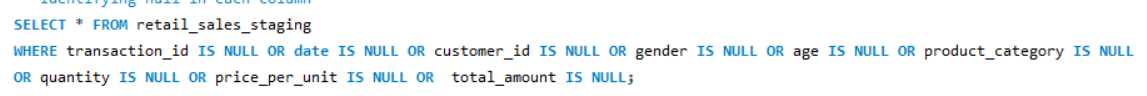


Output:

Number of duplicates: 0 row(s) returned

1. **Identified null in each column**

SQL Query:



Output:

Identifying Nulls: 0 row(s) returned; No nulls found

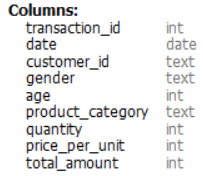
1. **Standardizing data types**

Changed data type from text to date

SQL Query:



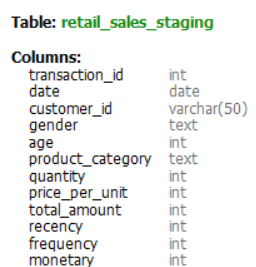
Output:



Changed data type from text to VARCHAR



Output:



After verifying that the data is clean by removing duplicates, identifying nulls and standardizing formats, I proceeded to the analysis.

***ANALYZE***

**Note:** The dataset includes January 2024, I focused only on the full year of 2023 to ensure complete and consistent analysis.

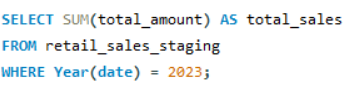
* + - * 1. **Sales performance and Customer Demographics Analysis (EDA)**

**KPIs**

Purpose: These KPIs provides instant clarity for decision-makers by highlighting key performance metrics by a glance, they can immediately understand how business is performing.

* 1. **What is the total sales amount for the year 2023?**

SQL Query:

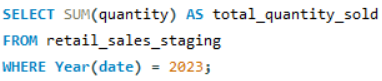


Output:



* 1. **How many units were sold overall?**

SQL Query:

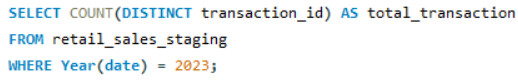


Output:



* 1. **How many transactions were made?**

SQL Query:



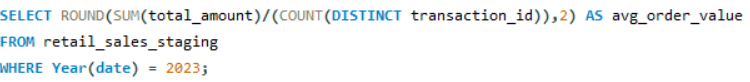
Output:



Observation: I noticed that there is only one transaction per customer.

* 1. **What is the average order value?**

SQL Query:



Output:



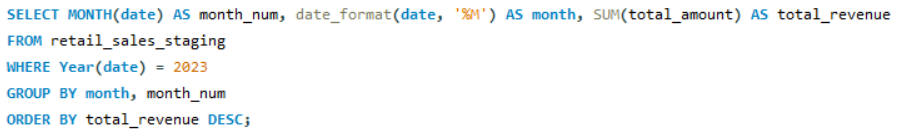
**Sales Performance Analysis**

**Monthly sales trends**

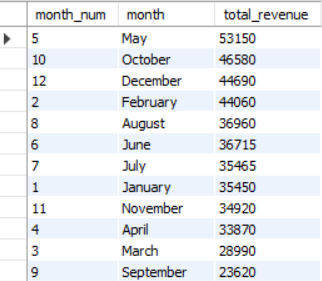
Purpose: The purpose of analyzing sales trends over a year is to identify patterns and seasonal variations in sales performance. This provides insights into:

* High-demand months where sales peak, such as holiday seasons or promotional periods.
* Low-demand months that may require strategic interventions to boost sales.
  1. **Which month exhibit peak sales performance?**

SQL Query:



Output:

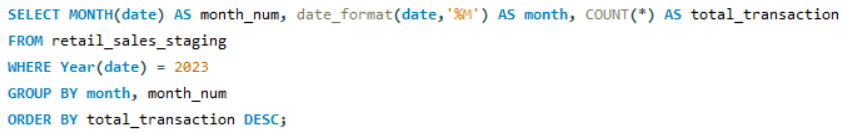


**Transaction Frequency Over Months**

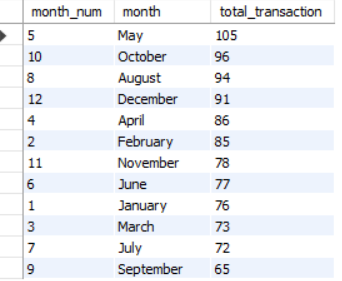
Purpose: The purpose of analyzing transaction frequency over a year is to understand how often transactions occur during specific months. By examining the distribution of transactions, businesses can:

* Identify high-activity months where transactions peak, indicating increased customer engagement.
* Detect low-activity months that may require targeted marketing or promotional efforts.
  1. **Which months have higher or lower customer purchase activity?**

SQL Query:



Output:

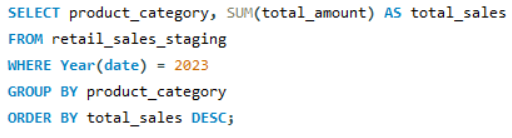


**Product Category by sales**

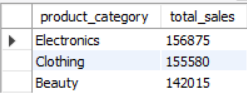
Purpose: The purpose of this is to analyze how much revenue each product category brings in.

* Identify which categories are driving the most sales.
* Helps to spot top-performing items and make decisions about inventory, marketing, or promotions.
  1. **Which product contributes the most to revenue?**

SQL Query:



Output:

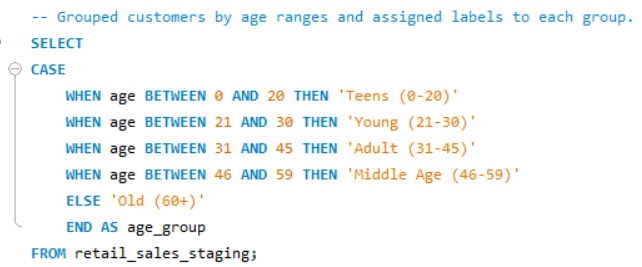


**Customer Demographics**

**Segmentation of Customers by Age**

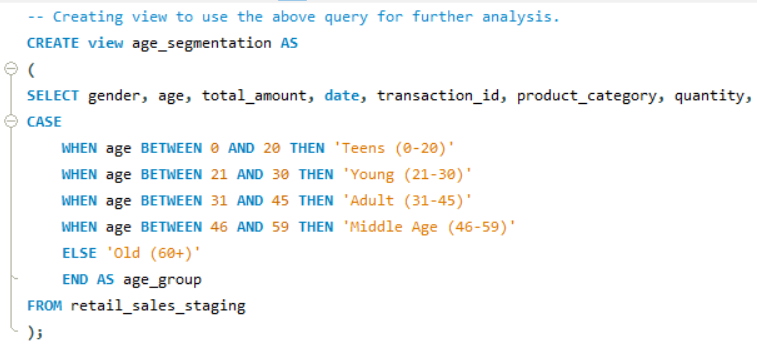
Grouped customers by age ranges and assigned labels to each group to better understand behavior and create focused insights.

SQL Query:

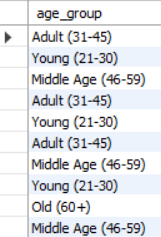


Created view to store selected columns for further analysis.

SQL Query:



Age Group Segmentation:

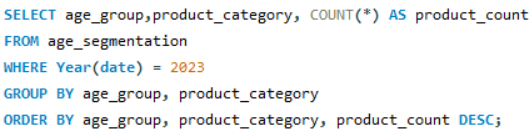


**Product categories purchased by age group**

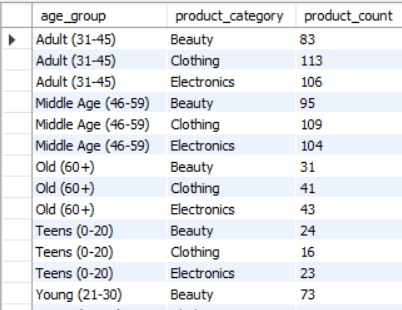
Purpose: To see what types of products each age group buys, so businesses can offer better choices, send the right promotions, and improve customer satisfaction.

* 1. **How does customer demographics such as age influence spending and product preferences?**

SQL Query:



Output:

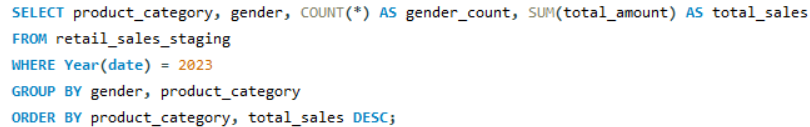


**Gender Distribution by Product category**

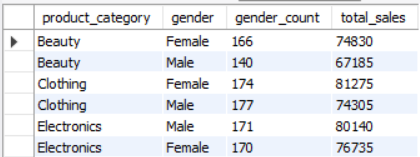
Purpose: The purpose of this is to analyze how different genders contribute to sales across various product categories.

* 1. **How does customer demographics such gender influence spending and product preferences?**

SQL Query:



Output:



1. **Customer Segmentation (RFM)**

Purpose: Helps businesses know which customers buy often, spend more, or might stop buying so they can improve marketing and service. By segmenting customers through RFM scores, business can:

* Tailor marketing strategies to match each group's preferences and needs
* Personalize product recommendations and promotions to increase engagement
* Identify high-value segments like loyal customers or big spenders
* Spot customers who are at risk of churning and create strategies to retain them

Brief description of Recency, Frequency and Monetary (RFM)

* Recency – Days since the last purchase (lower score is better)
* Frequency – Total number of purchases (higher score is better)
* Monetary – Total amount spent (higher score is better)

By combing RFM scores, I will classify customers into actionable segments like:

* **Champions** – Recent, frequent, and high-value buyers
* **Loyal Customers** – Consistent shoppers with strong history
* **Big Spenders** – Customers who spend a lot
* **At-Risk** – Used to buy, but activity is dropping
* **Low Engagement** – Rarely shop and spend little

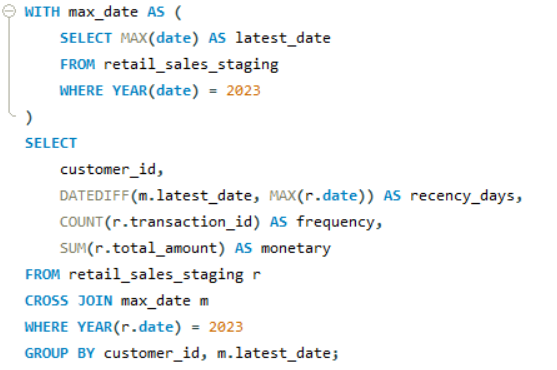
After calculating RFM scores, I created customer segments labeled ‘Champions’, ‘Loyal Customers’, ‘Big Spenders’, ‘At Risk’, and ‘Low Engagement’. Then, I analyzed how customers in each segment behaved and what they tended to buy.

**Below is the step-by-step process:**

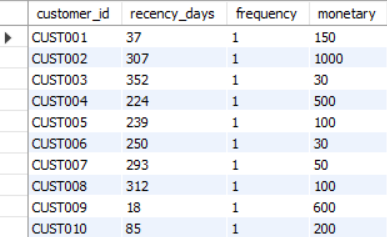
1. **RFM Calculation (recency, frequency, monetary)**

Used CTE and CROSS JOIN to calculate the recency

SQL Query:

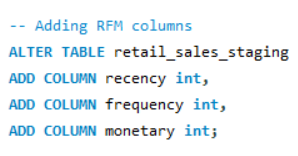


Output:

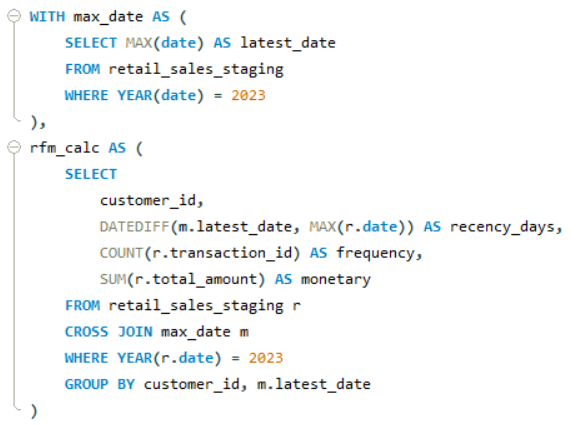


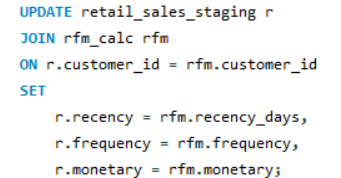
Created new columns to store the RFM values used for scoring and segmentation.

SQL Query:



Inserted the data to each column



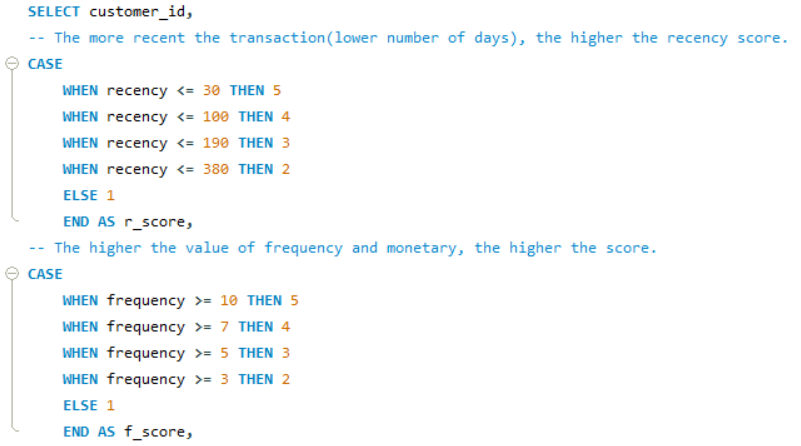


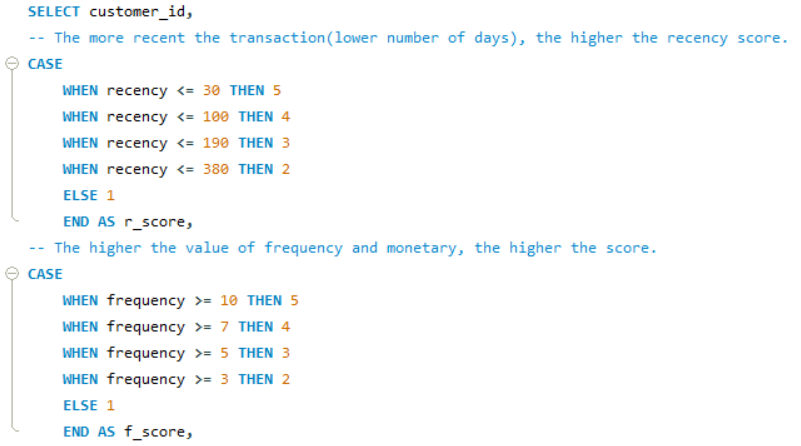
Successfully created and updated the calculated RFM individual columns.

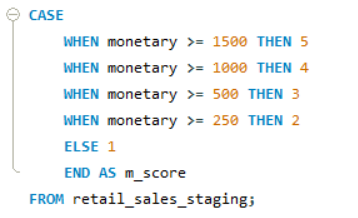
1. **RFM Score Creation**

Based on the computed RFM, I assigned RFM scores for each customer.

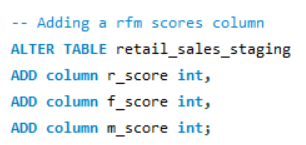
SQL Query:

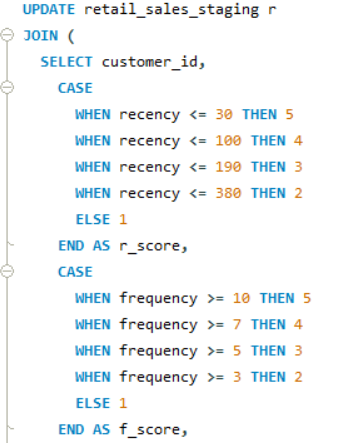


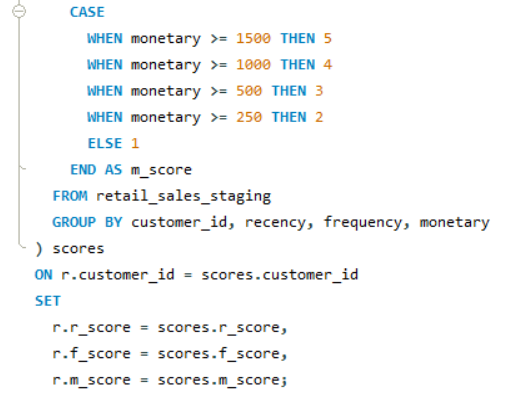




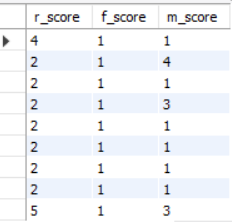
Saved each customer's RFM scores, combined them into a single RFM code, then use that code to assign a customer label.



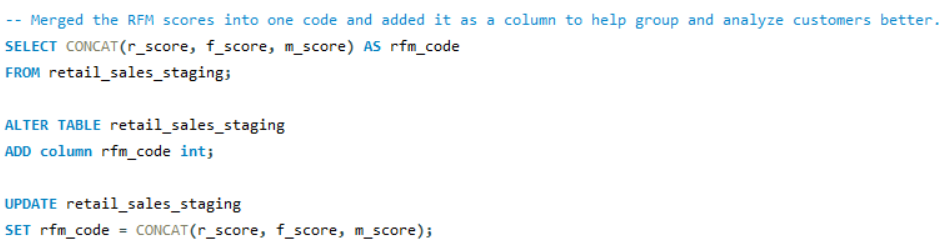




Output:

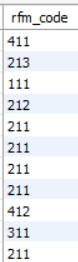


1. **RFM Code Creation**

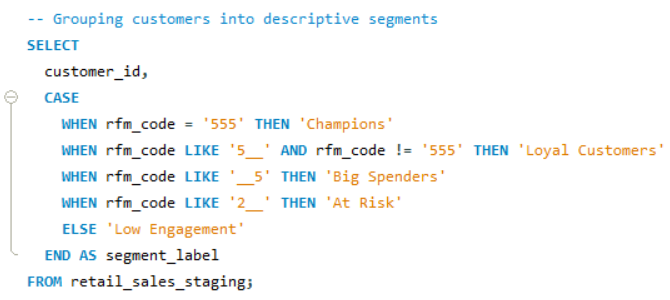


**Note**: At first, I stored the data type as int, I keep getting an error. After analyzing further, it should be a string (varchar or char) to use the rfm\_code for labeling the segments.

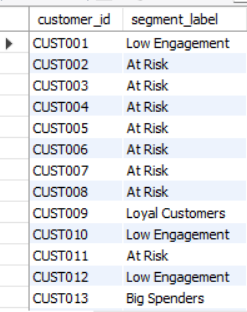
Output:

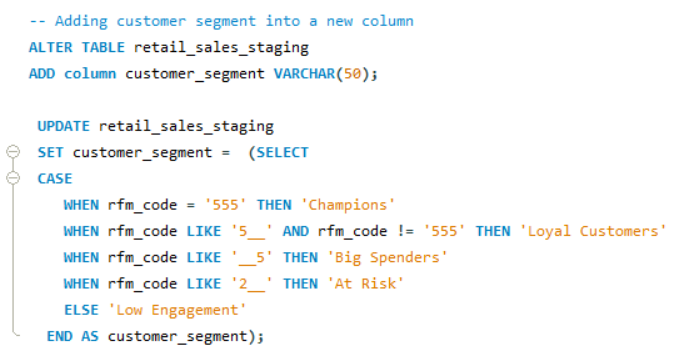


1. **Customer Segmentation**



Output:



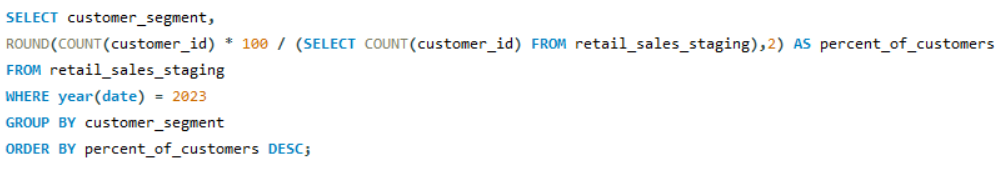


Successfully created and updated the customer segmentation based on RFM codes.

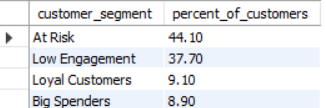
We can now analyze customer behavior and purchasing patterns across segments.

* + - * 1. **What percentage of customers belong to the high-value segment?**

SQL Query:

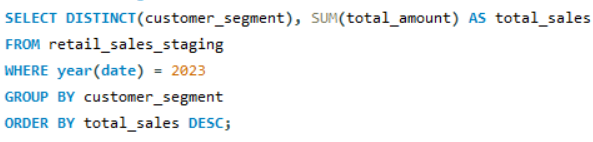


Output:

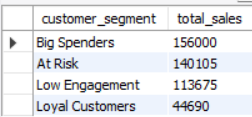


* + - * 1. **Which RFM segments contribute the most to overall revenue?**

SQL Query:

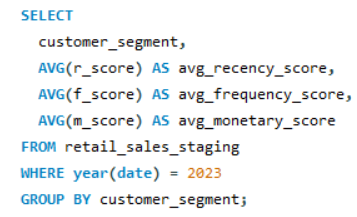


Output:

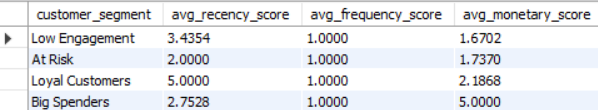


* + - * 1. **What is the average RFM scores by segment?**

SQL Query:

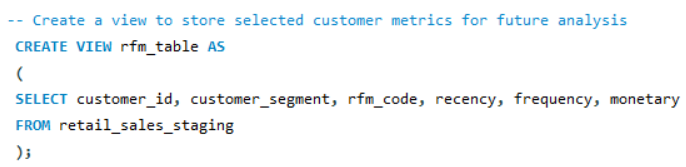


Output:

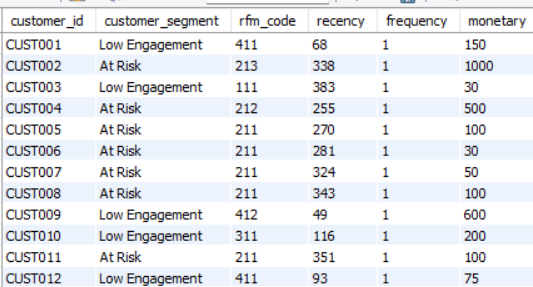


Created view to store selected metrics for further analysis.

SQL Query:



Output:



***SHARE***

**Data Visualization in Tableau**

After cleaning and structuring the retail dataset using SQL, I imported the refined data into Tableau to develop interactive dashboards.

**Purpose:** These visualizations were designed to clearly present customer segmentation and sales performance insights, enabling stakeholders to explore trends and make data-driven decisions with ease.

**Target Audience**

This dashboard is designed for sales managers, marketing analysts, and e-commerce strategists seeking to enhance customer retention and maximize revenue.

**Design and Interactivity**

1. **Dashboard Dynamics**

* This dashboard provides users with the ability to navigate between the sales and customers dashboards easily.
* The charts and graphs are interactive, enabling users to filter data using charts.
* I refined the **tooltips** to make the key results easier to understand at a glance. The updated tooltips display only the most relevant details, helping users quickly interpret data without clutter or distraction.

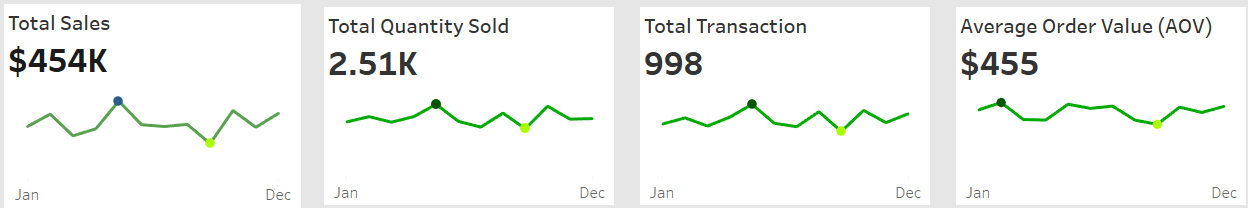
1. **Data Filters**

* Allow users to filter data by appropriate information.
* These filters are designed to support flexible analysis, enabling stakeholders to drill down into specific customer behaviors or time periods for deeper insights.

**Note:** The dataset includes January 2024, I filtered and focused only on the full year of 2023 to ensure complete and consistent analysis.

**1st Dashboard: KPI Overview**

The dashboard highlights key metrics Total Sales, Total Quantity Sold, Total Transactions, and Average Order Value (AOV) using bold number displays (BANs) for quick reference. Each KPI is paired with a sparkline to show its performance trend over time.



**Insights:**

**Total Revenue:**

* Total revenue is currently high, but the sparkline shows a recent decline. The current figure sits closer to the minimum ($24K) than the maximum ($53K) for the year, which may suggest a slowdown in customer activity and a need to boost engagement.

**Total Quantity Sold:**

* The total quantity sold has dropped from its peak during the mid-year period. Although sales are still active, the current volume is much closer to the yearly low, suggesting reduced customer demand or fewer large orders in recent months.

**Total Transactions:**

* Transaction volume peaked in May but has slowed down in September. It might be time to refresh promotions or explore why fewer orders are coming in.

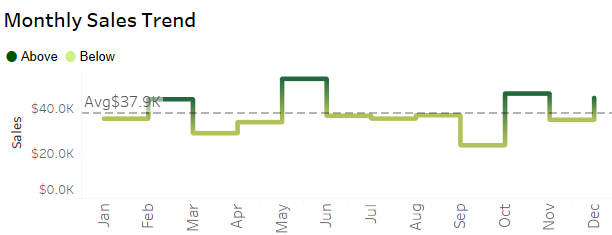
**Average Order Value(AOV)**

* AOV reached its highest point in February at $518, and its lowest in September at $363. The current value is closer to the lower end, which could mean customers are making smaller purchases recently. This might be a good time to explore upselling strategies or limited-time bundles.

**Sales Performance Analysis**

* 1. **Monthly Sales Trend**

To display monthly sales and total amount, I used a **time series line chart** in Tableau based on transaction dates. I then changed the line style to “step” to highlight how sales values change from one month to the next. I also added an average reference line to show the overall sales benchmark, using single color with varying shades—darker for above average and lighterfor below average.



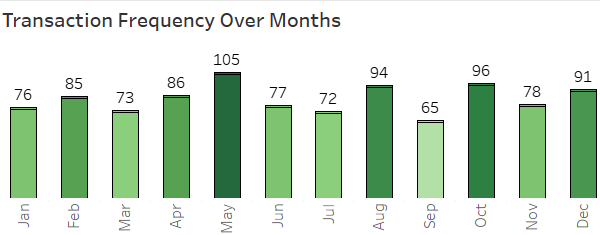
**Insights:**

* Sales peaked in May, experienced a dipped in September and sales bounced back in October.
* Spring, month of May, and the holiday season, month of December, are key sales periods.

These seasonal surges likely reflect consumer behaviors tied to events like back-to-school, summer breaks, or holiday gift-giving.

* 1. **Transaction Frequency Over Months**

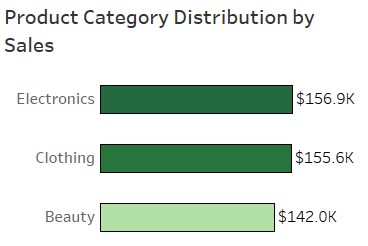
I used a **bar chart** to show the number of transactions each month, making it easy to compare activity over time. To highlight the exact monthly totals, I added a Gantt bar overlay that displays the numeric values just above each bar. This makes it quick for users to see both the trend and the specific figures without needing to hover or check tooltips.



**Insight:**

* Transaction frequency reached its peak in May, indicating high customer activity, and experienced a noticeable dip in September.
  1. **Top-Performing Product Categories**

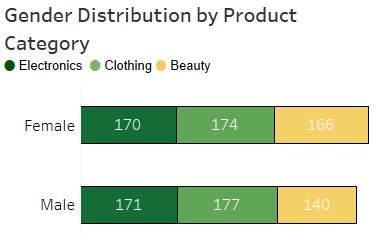
I used **a horizontal bar chart** sorted from highest to lowest sales. To show the difference in performance, I applied a single color with varying shades—darker for higher sales and lighter for lower ones. This keeps the chart easy to read while still highlighting category rankings.



**Insight:**

* Electronics contributes most to sales while beauty contributes least to sales.
  1. **Gender Distribution by Product Category**

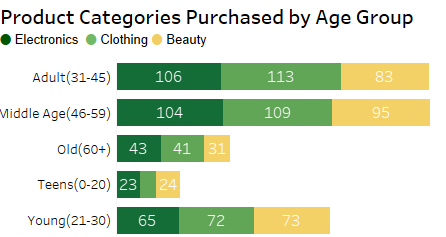
I used a **horizontal stacked bar chart** to display the count of each gender within every product category. To enhance clarity, I applied a green-gold color palette—visually distinguishing product categories while keeping the chart easy to read.



**Insight:**

* Female customers show a stronger preference for clothing. While male customers also draw strong interest to electronics.
  1. **Product Categories Purchased by Age Group**

I used a **stacked bar chart** to show which product categories are most popular across different age groups. To make the chart visually engaging and easy to interpret, I applied a green-gold color palette that clearly separates each category.



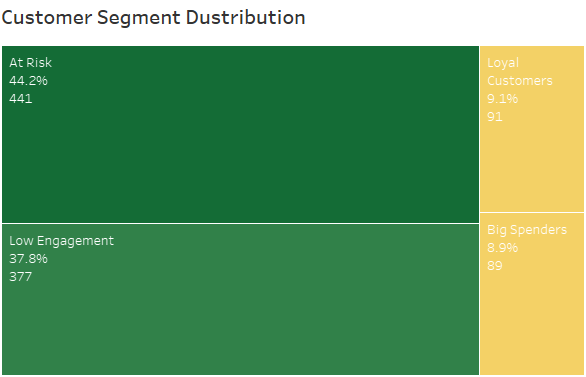
**Insight:**

* Electronics, clothing and beauty are most popular for middle-aged and adult customers. While young customers show interest to beauty products.

**2nd Dashboard: Customer Segmentation (RFM)**

1. **Customer Segment Distribution**

I used a heatmap to visualize customer segments, showing each segment’s name, its percentage of the total, and the customer count. The color intensity helps highlight which segments are most notable, making it easy to spot key patterns at a glance.

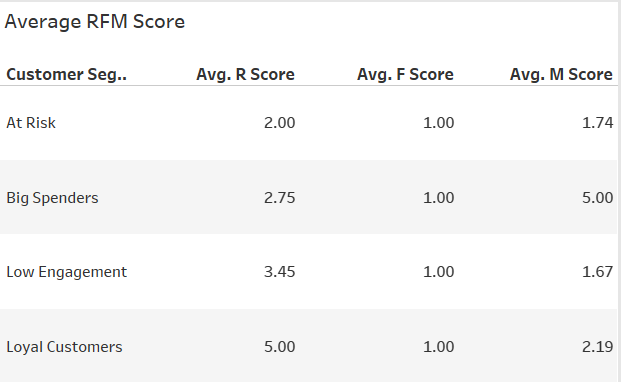


**Insight:**

* Only 18.0% of customers fall into high-value segments (Big Spenders and Loyal Customers), highlighting a small but important group that contributes significantly to sales. The majority of customers fall under the At Risk (44.2%) and Low Engagement (37.8%) segments, indicating a need for re-engagement strategies.

1. **Average RFM Score Table**

I created a table showing the average RFM scores for each segment, making it easy to spot which ones need immediate action.

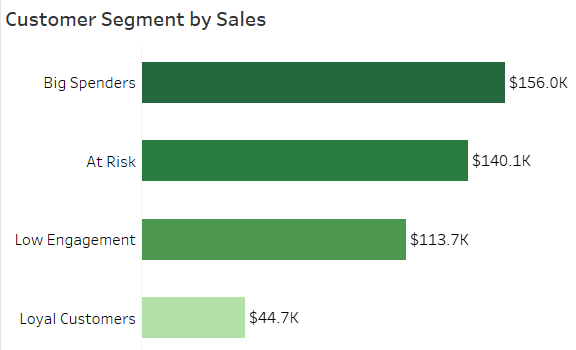


**Insight:**

* Loyal Customers buy most recently, with the highest Recency score (5.0), while Big Spenders spend the most, with the top Monetary score (5.0). In contrast, Low Engagement and At-Risk segments have lower scores, showing they buy less often and spend less.

1. **Customer Segment by Sales**

I created a horizontal bar chart to show the segment which contributes most to sales, I sorted the segments from highest to lowest sales and used a green color scale—darker green shows higher sales.

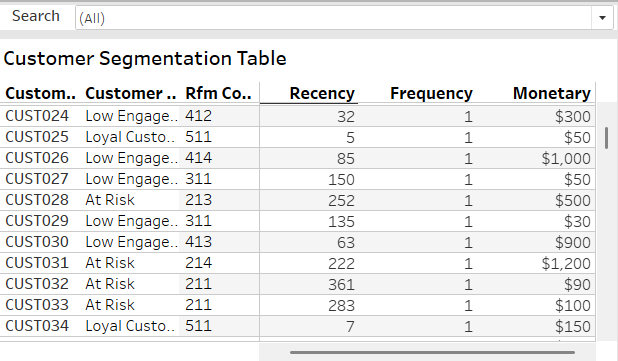


**Insight:**

* Analysis showed that the Big Spenders and At-Risk segments contributed the most to sales, while the Loyal Customers segment had a surprisingly minimal impact. This suggests that recent transactions may have come from less consistent buyers, highlighting an opportunity to convert these segments into more loyal ones through targeted re-engagement strategies.

1. **Look Up Table**

To easily Identify each customer’s RFM segment, I created a lookup table, allowing for faster analysis and targeted strategy decisions based on behavior patterns



***ACT***

**Conclusion and Recommended Actions**

**Sales and Customer Demographics**

Retail sales in 2023 showed strong performance overall. Sales and transaction volume peaked in May, dropped in September, and began to recover in October. May and December are strong sales months, likely due to holidays and seasonal events like school breaks. Female customers among adult and middle-aged group have a strong preference for clothing. While male customers among adult and middle-aged group have a strong interest for electronics product. Beauty products are most popular to young female customers. Overall, electronics contributed the most to total sales, making it the top-performing product category for the year.

**Recommended Action:**

* Introduce targeted promotions for September and set clear performance goals.
* Re-engage inactive customers during low-activity months through strategic campaigns.
* Ensure balanced inventory and tailor promotions across product categories.
* Align marketing by audience: beauty for females, electronics for males, and unisex campaigns for clothing.
* Focus electronics promotions toward middle-aged and older customers; promote beauty products to younger buyers.

**RFM Analysis**

Only 18.0% of customers fall into high-value segments such as Big Spenders and Loyal Customers, while most customers fall into the At-Risk and Low Engagement groups. Though Big Spenders customers aren’t purchase frequently, they still make up a big part of the sales, so it's worth focusing on keeping them. To grow value, we can re-engage inactive customers with discounts, rewards, or surveys to understand what’s stopping them from buying again.

**Recommended Actions:**

* **High-Value Retention:** Strengthen relationships with Loyal Customers and Big Spenders by offering exclusive perks such as early access to products, premium support, or surprise rewards. This reinforces their value and encourages continued loyalty.
* **Retention Focus:** Prioritize reactivation campaigns for At-Risk customers, especially those high monetary value from the past purchases.
* **Value Growth:** Introduce loyalty programs, exclusive discounts, or upsell strategies for low-engagement customers. The goal is to increase their frequency and spend, moving them toward more valuable segments.
* **Customer Experience:** Use feedback tools such as short surveys or exit polls to understand why customers stop purchasing and identify ways to enhance their journey.