Ecommerce Data Analysis Report

Overview:

* The ecommerce project involves the analysis of various datasets related to an online retail business taken from Kaggle.com (Brazilian E-Commerce Public Dataset by Olist)
* The datasets typically include information about orders, customers, products, payments, sellers, and geolocation.

**Introduction:**

* The ecommerce industry has witnessed significant growth in recent years, with more consumers opting for online shopping due to convenience and a wide range of options.
* My analysis focuses on a specific ecommerce business, which operates in various regions and offers a diverse range of products.
* The datasets which i analyze comprises several tables containing information about orders, customers, products, payments, sellers, and geolocation.

**Objectives:**

* The primary objectives of the analysis are to gain actionable insights into various aspects of the ecommerce business, including customer behaviour, sales performance, product trends, geographic distribution, seller performance, and payment trends.
* By analyzing these key areas, we aim to provide valuable insights that can inform strategic decision-making, optimize operational processes, enhance customer experience, and drive business growth.

Data Overview:

**Datasets Description:**

**Dataset** -[**Link**](https://drive.google.com/drive/folders/1EMxuEKeoisz7LVy83Amdm4rykO7rDVKM?usp=drive_link)

* 1. **Order Items:** This dataset contains information about individual items within orders, including details such as order ID, product ID, seller ID, price, and shipping details.
  2. **Customers:** This dataset provides information about the customers, including unique identifiers, zip code, city, and state.
  3. **Orders:** The Orders dataset includes information about orders, such as order ID, customer ID, order status, and timestamps for various stages of the order process (purchase, approval, delivery, etc.).
  4. **Sellers:** This dataset contains information about sellers, including seller ID, zip code, city, and state.
  5. **Product Categories:** Provides information about product categories, including category names and their English translations.
  6. **Geolocation:** This dataset contains geographic information related to zip codes, including latitude, longitude, city, and state.
  7. **Payments:** Dataset containing payment details, including order ID, payment sequential, payment type, installments, and payment value.
  8. **Products:** Information about products, including product ID, category name, name length, description length, number of photos, weight, length, height, and width.

**Key Fields/Columns:**

**Order Items:**

* + - order\_id: Unique identifier for each order.
    - product\_id: Unique identifier for each product.
    - seller\_id: Unique identifier for each seller.
    - price: Price of the item.
    - shipping\_limit\_date: Deadline for shipping the item.

**Customers:**

* customer\_id: Unique identifier for each customer.
* customer\_unique\_id: Unique identifier for each customer (potentially a hashed value for privacy).
* customer\_zip\_code\_prefix: First part of the customer's zip code.
* customer\_city: City where the customer is located.
* customer\_state: State where the customer is located.

**Orders:**

* + - order\_id: Unique identifier for each order.
    - customer\_id: Unique identifier for each customer.
    - order\_status: Status of the order (e.g., pending, delivered, canceled).
    - order\_purchase\_timestamp: Timestamp when the order was placed.
    - order\_approved\_at: Timestamp when the order was approved.
    - order\_delivered\_carrier\_date: Timestamp when the order was handed over to the carrier.
    - order\_delivered\_customer\_date: Timestamp when the order was delivered to the customer.
    - order\_estimated\_delivery\_date: Estimated delivery date of the order.

**Sellers:**

* seller\_id: Unique identifier for each seller.
* seller\_zip\_code\_prefix: First part of the seller's zip code.
* seller\_city: City where the seller is located.
* seller\_state: State where the seller is located.

**Product Categories:**

* product\_category\_name: Name of the product category.
* product\_category\_name\_english: English translation of the product category name.

**Payments:**

* order\_id: Unique identifier for each order.
* payment\_sequential: Sequential number of the payment.
* payment\_type: Type of payment method used.
* payment\_installments: Number of installments for the payment.
* payment\_value: Value of the payment.

**Products:**

* product\_id: Unique identifier for each product.
* product\_category\_name: Name of the product category.
* product\_name\_length: Length of the product name.
* product\_description\_length: Length of the product description.
* product\_photos\_qty: Quantity of product photos.
* product\_weight\_g: Weight of the product in grams.
* product\_length\_cm: Length of the product in centimeters.
* product\_height\_cm: Height of the product in centimeters.
* product\_width\_cm: Width of the product in centimeters.

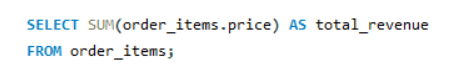
**Geolocation:**

* geolocation\_zip\_code\_prefix: First part of the zip code.
* geolocation\_lat: Latitude coordinate of the location.
* geolocation\_lng: Longitude coordinate of the location.
* geolocation\_city: City name.
* geolocation\_state: State abbreviation.

Analysis Findings:

**1.Total Revenue Analysis:**

This query calculates the total revenue generated from all orders.

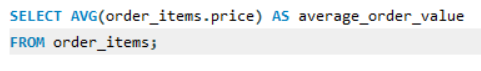


**OUTPUT:**

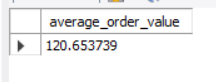


**2.Average Order Value Analysis:**

This query computes the average order value, which indicates the average amount spent per order.

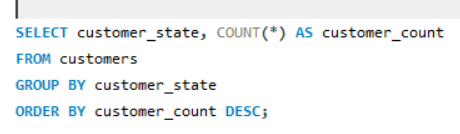


**OUTPUT:**

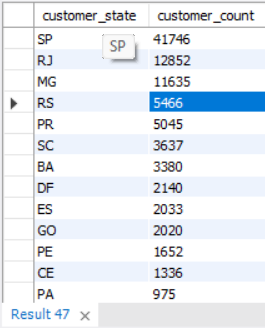


**3.Customer Distribution by State Analysis:**

This query provides a breakdown of the number of customers by state.

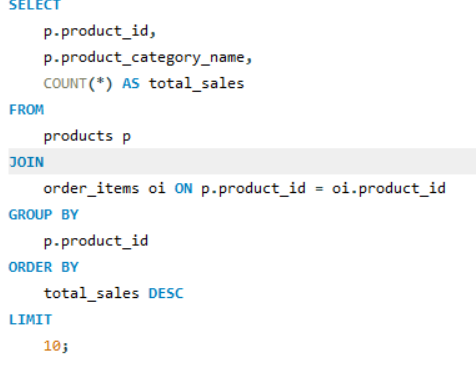


**OUTPUT:**



**4.Top Selling Products Analysis:**

This query identifies the top 10 best-selling products.

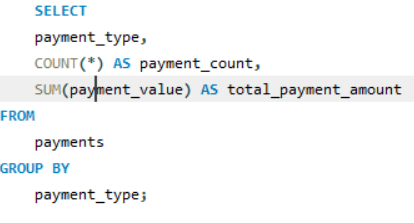


**OUTPUT:**

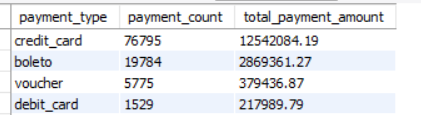


**5.Payment Method Analysis:**

This query provides insights into payment methods used by customers and their corresponding payment counts and total amounts.

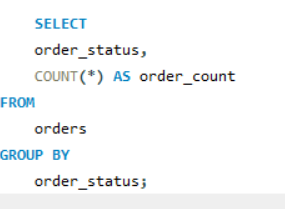


**OUTPUT:**

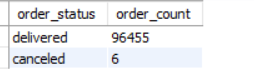


**6.Order Status Analysis:**

This query analyzes the distribution of orders by status, helping to understand order fulfillment efficiency.

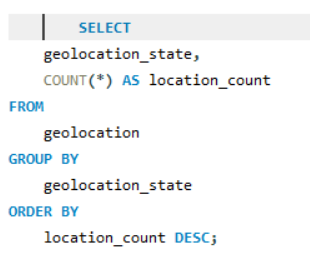


**OUTPUT:**

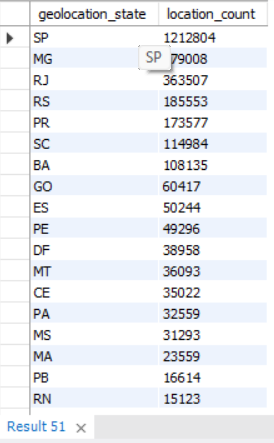


**7.Geolocation Analysis:**

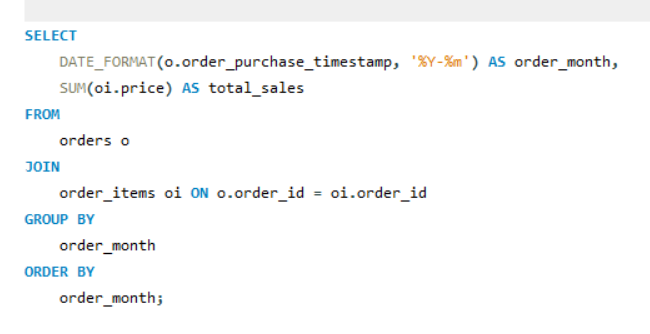
This query explores the geographic distribution of customers or sellers by state.



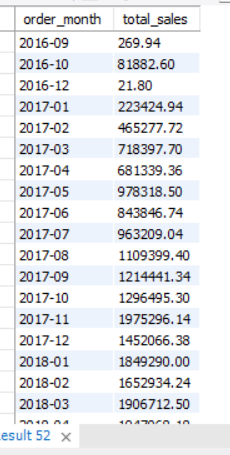
**OUTPUT:**



**8.Sales Performance Analysis:**

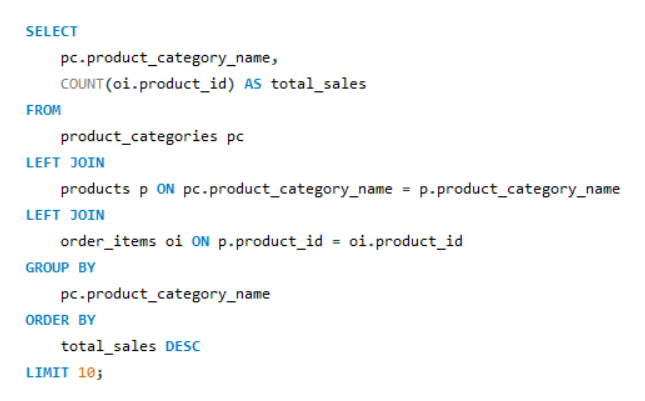
This query retrieves the total sales for each month by joining the orders and order\_items tables. 

**OUTPUT:**



**9.Top Selling Categories Analysis:**

This query identifies the top-selling product categories by joining the product\_categories, products, and order\_items tables.

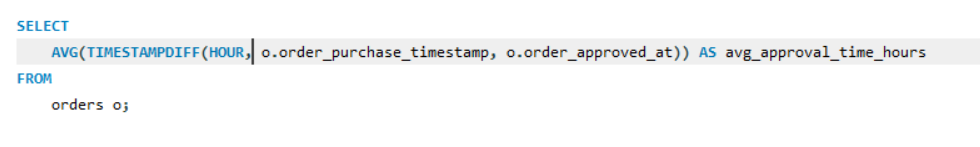


**OUTPUT:**

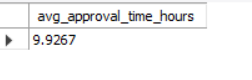


* 1. **Order Approval Time Analysis:**

This query calculates the average time taken for order approval by subtracting the order\_purchase\_timestamp from the order\_approved\_at, using TIMESTAMPDIFF function.

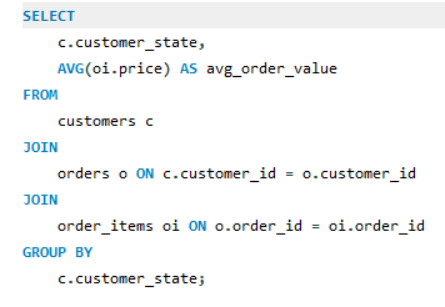


**OUTPUT:**

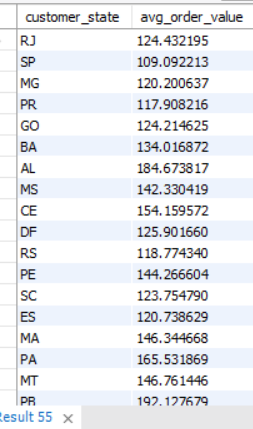


**11.Average Order Value by State Analysis:**

This query calculates the average order value for each state by joining the customers, orders, and order\_items tables.

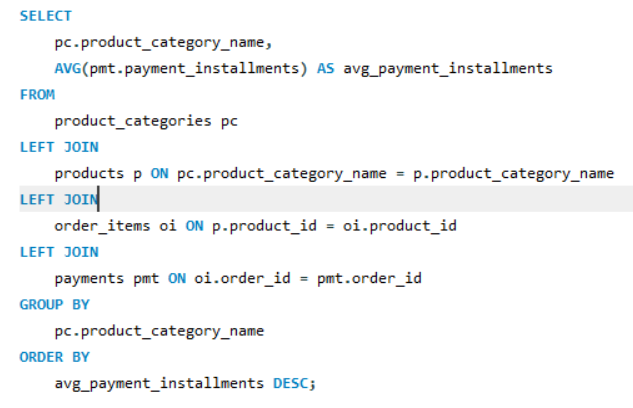


**OUTPUT:**



**12. Average Payment Installments by Product Category:**

This query calculates the average number of payment installments for each product category, providing insights into customer payment behaviour.

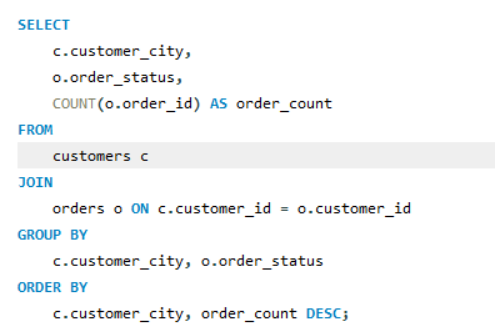


**OUTPUT:**



**13.Order Status Analysis by Customer City:**

This query analyzes the distribution of order statuses (e.g., delivered, processing) by customer city, helping identify areas with potential delivery or processing issues.

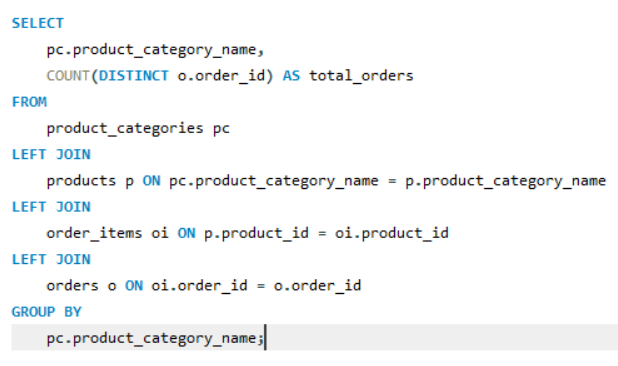


**OUTPUT:**

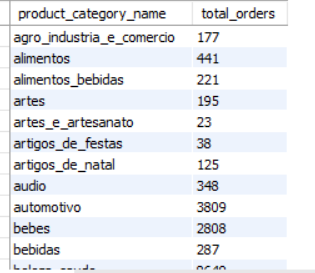


**14. Product Return Rate Analysis:**

This query calculates product category to the total number of orders.

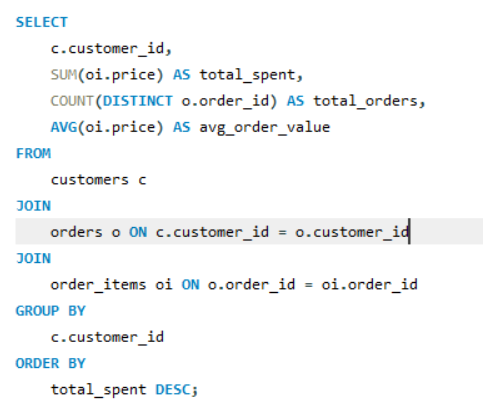


**OUTPUT:**

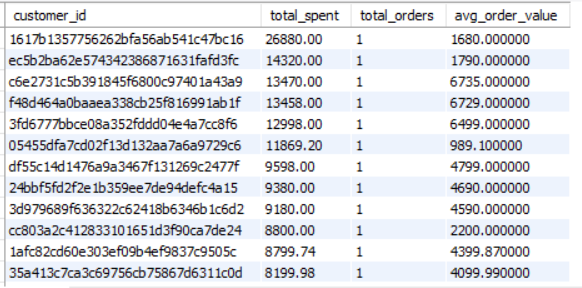


**15. Customer Lifetime Value Analysis:**

This query calculates the total amount spent, total number of orders, and average order value for each customer, allowing for the analysis of customer lifetime value.

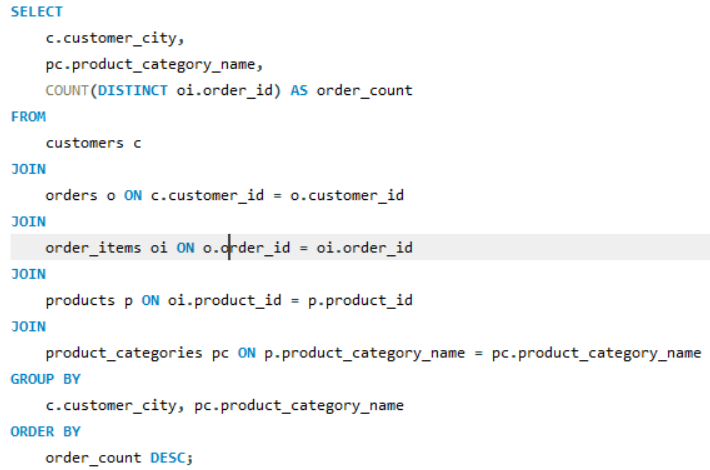


**OUTPUT:**

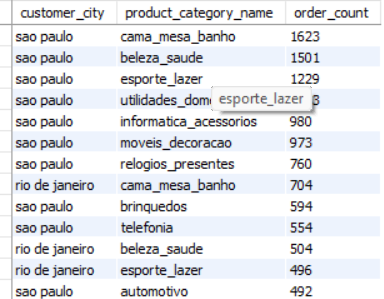


**16. Popular Product Categories by Customer City:**

This query identifies the most popular product categories for each customer city based on the number of orders, helping to understand regional preferences.

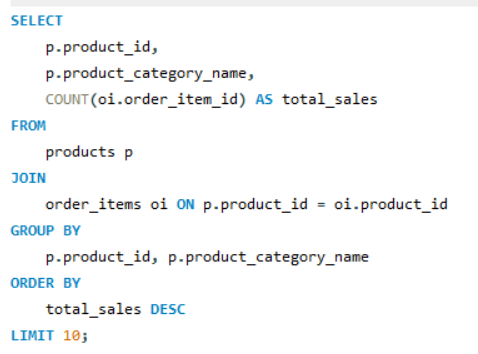


**OUTPUT:**

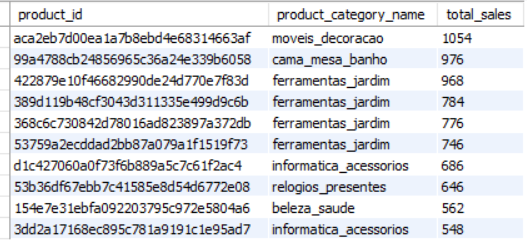


**17.Top Selling Products:**

This query identifies the top-selling products by counting the number of order items associated with each product, providing insights into product popularity.

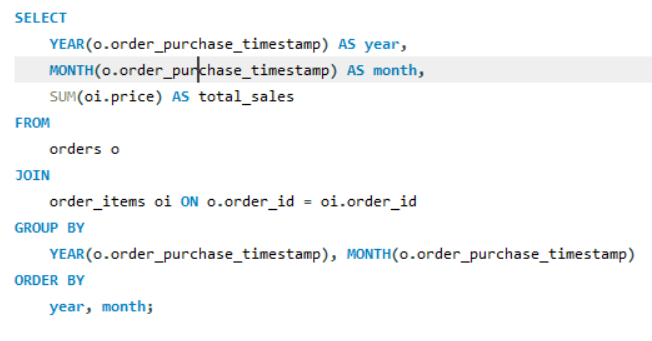


**OUTPUT:**



**18.Sales Distribution by Month:**

This query calculates the total sales for each month, allowing for the analysis of sales trends over time.

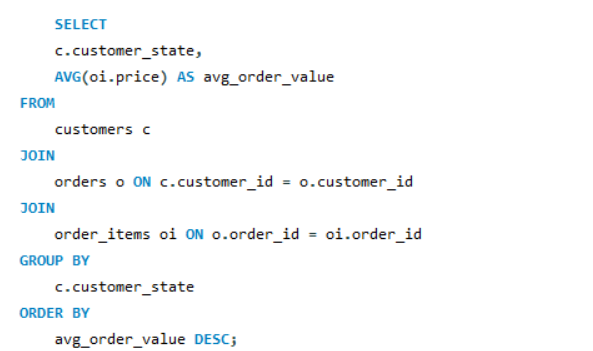


**OUTPUT:**

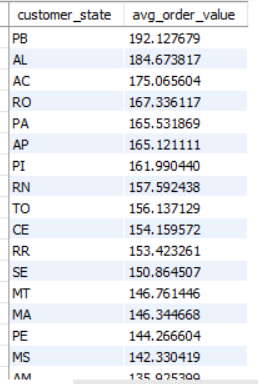


**19.** **Average Order Value by Customer State:**

This query calculates the average order value for each customer state, providing insights into regional spending habits.

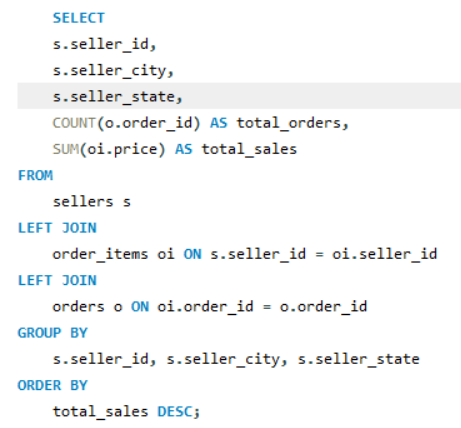


**OUTPUT:**



**20.Seller Performance Analysis:**

This query provides insights into seller performance by calculating the total number of orders and total sales generated by each seller.

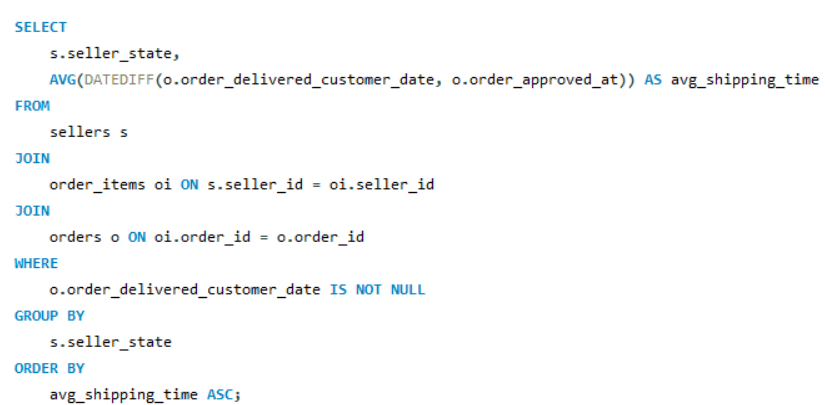


**OUTPUT:**

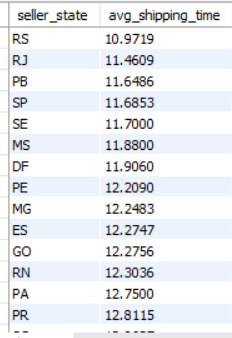


**21.Average Shipping Time by Seller State:**

This query calculates the average shipping time for orders fulfilled by sellers in each state, helping to identify regions with faster or slower delivery times.

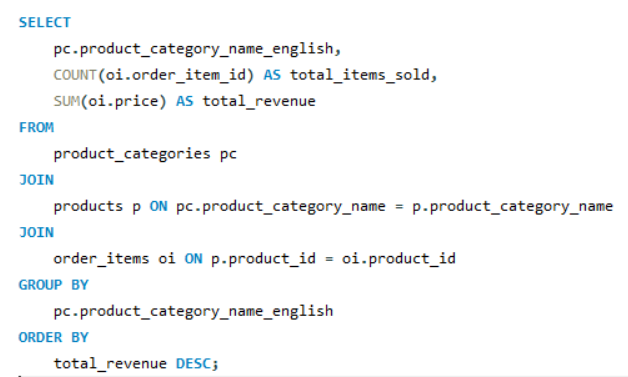


**OUTPUT:**

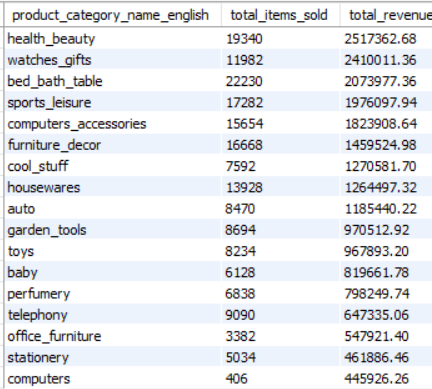


**22. Product Category Sales Analysis:**

This query provides insights into sales performance across different product categories, including the total number of items sold and total revenue generated for each category.

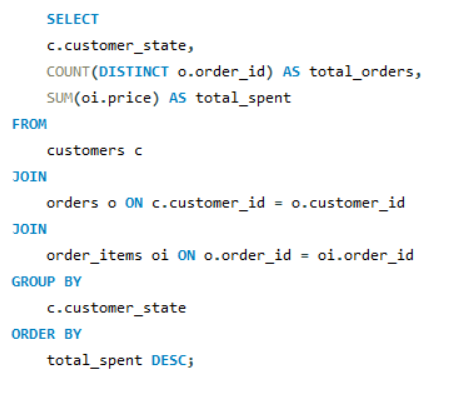


**OUTPUT:**

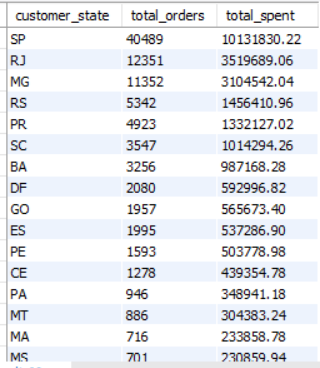


**23.Customer State Analysis:**

This query analyzes customer spending and order frequency by state, helping to identify regions with high-value customers and potential opportunities.

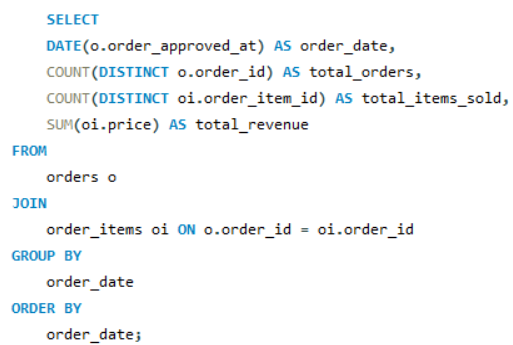


**OUTPUT:**

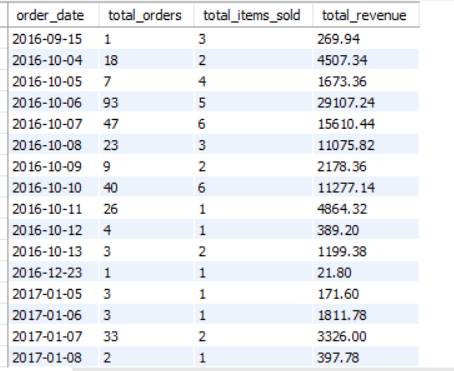


**24.Order Fulfillment Efficiency:**

This query tracks order fulfillment efficiency over time by analyzing the total number of orders, items sold, and revenue generated on each day.



**OUTPUT:**



Recommendations**:**

**Improving Sales Performance:**

* **Targeted Marketing Campaigns:** Utilize insights from the product category sales analysis to design targeted marketing campaigns for top-performing categories. For example, allocate a higher marketing budget to promote products in these categories through social media ads or email marketing.
* **Product Bundling:** Identify complementary products within top-selling categories and create bundled offerings to encourage higher average order value (AOV). For instance, offer discounts for purchasing related items together, such as pairing a camera with memory cards and tripods.
* **Seasonal Promotions:** Leverage seasonal trends identified in the sales data to launch time-sensitive promotions and capitalize on peak demand periods. For instance, offer discounts on winter clothing during the colder months to stimulate sales.

**Enhancing Customer Engagement:**

* **Personalized Recommendations:** Utilize customer state analysis insights to tailor product recommendations based on regional preferences and behaviors. Implement recommendation algorithms that suggest products similar to those previously purchased by customers in the same state.
* **Reward Programs:** Implement a loyalty program to incentivize repeat purchases and foster customer loyalty. Offer rewards such as discounts, exclusive access to new products, or free shipping for members who reach certain spending thresholds.
* **Interactive Content:** Create interactive content such as quizzes or polls on social media platforms to engage customers and gather feedback about their preferences. Use insights from customer interactions to refine product offerings and marketing strategies.

**Optimizing Operational Efficiency:**

* **Real-Time Inventory Management:** Implement an inventory management system that provides real-time updates on stock levels and automates replenishment processes. Use insights from order fulfillment efficiency analysis to forecast demand accurately and maintain optimal inventory levels to prevent stockouts or overstocking.
* **Streamlined Order Processing:** Utilize automation tools to streamline order processing and reduce manual errors. Implement order tracking systems that provide customers with real-time updates on the status of their orders, enhancing transparency and improving the overall shopping experience.
* **Supplier Collaboration:** Collaborate closely with suppliers to optimize logistics and minimize shipping delays. Negotiate favorable terms with shipping partners and explore alternative delivery options, such as local fulfillment centers, to reduce shipping times and costs.

Conclusion**:**

**Summary of Key Insights and Recommendations:**

* The analysis of the ecommerce dataset provided valuable insights into various aspects of the business, from sales performance across product categories to customer behaviour and order fulfillment efficiency.

Appendix**:**

Detailed SQL queries used for the analysis - [Link](https://drive.google.com/drive/folders/1Ewebpv3FCeo8QRSGPpa0zvUyKrJ9AyxW?usp=drive_link)