# Customer Shopping Behaviour Analysis Report

## 1. Project overview

This project analyzes customer shopping behavior using transactional data from **3,900 purchases** across various product categories. The objective is to identify spending patterns, customer segments, product preferences, and subscription trends to support data-driven business decisions.

## 2. Dataset summary

The dataset contains **3,900 records** and **18 features**, capturing customer demographics, purchase details, and shopping behavior. Customer attributes include age, gender, location, and subscription status, while transactional data covers items purchased, product categories, purchase amounts, and seasonal factors. Behavioral variables include discount usage, promo code application, purchase frequency, previous purchases, review ratings, and shipping type. The dataset contains **37 missing values in the review rating column**, which were addressed during data cleaning.

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Customer | Age | Gender | Item Purcl | Category | Purchase / | Location | Size | Color | Season | Review Rat | Subscripti | Shipping T | Discount A | Promo Cod | Previous P | Payment M | Frequency |
| 1 | 55 | Male | Blouse | Clothing | 53 | Kentucky | L | Gray | Winter | 3.1 | Yes | Express | Yes | Yes | 14 | Venmo | Fortnightly |
| 2 | 19 | Male | Sweater | Clothing | 64 | Maine | L | Maroon | Winter | 3.1 | Yes | Express | Yes | Yes | 2 | Cash | Fortnightly |
| 3 | 50 | Male | Jeans | Clothing | 73 | Massachu | S | Maroon | Spring | 3.1 | Yes | Free Shipp | Yes | Yes | 23 | Credit Car | Weekly |
| 4 | 21 | Male | Sandals | Footwear | 90 | Rhode Isla | M | Maroon | Spring | 3.5 | Yes | Next Day A | Yes | Yes | 49 | PayPal | Weekly |
| 5 | 45 | Male | Blouse | Clothing | 49 | Oregon | M | Turquoise | Spring | 2.7 | Yes | Free Shipp | Yes | Yes | 31 | PayPal | Annually |
| 6 | 46 | Male | Sneakers | Footwear | 20 | Wyoming | M | White | Summer | 2.9 | Yes | Standard | Yes | Yes | 14 | Venmo | Weekly |
| 7 | 63 | Male | Shirt | Clothing | 85 | Montana | M | Gray | Fall | 3.2 | Yes | Free Shipp | Yes | Yes | 49 | Cash | Quarterly |
| 8 | 27 | Male | Shorts | Clothing | 34 | Louisiana | L | Charcoal | Winter | 3.2 | Yes | Free Shipp | Yes | Yes | 19 | Credit Car | Weekly |
| 9 | 26 | Male | Coat | Outerwear | 97 | West Virgi | L | Silver | Summer | 2.6 | Yes | Express | Yes | Yes | 8 | Venmo | Annually |
| 10 | 57 | Male | Handbag | Accessori | 31 | Missouri | M | Pink | Spring | 4.8 | Yes | 2-Day Ship | Yes | Yes | 4 | Cash | Quarterly |
| 11 | 53 | Male | Shoes | Footwear | 34 | Arkansas | L | Purple | Fall | 4.1 | Yes | Store Picki | Yes | Yes | 26 | Bank Trans | Bi-Weekly |
| 12 | 30 | Male | Shorts | Clothing | 68 | Hawaii | S | Olive | Winter | 4.9 | Yes | Store Picki | Yes | Yes | 10 | Bank Trans | Fortnightly |
| 13 | 61 | Male | Coat | Outerwear | 72 | Delaware | M | Gold | Winter | 4.5 | Yes | Express | Yes | Yes | 37 | Venmo | Fortnightly |
| 14 | 65 | Male | Dress | Clothing | 51 | New Hamp | M | Violet | Spring | 4.7 | Yes | Express | Yes | Yes | 31 | PayPal | Weekly |
| 15 | 64 | Male | Coat | Outerwear | 53 | New York | L | Teal | Winter | 4.7 | Yes | Free Shipp | Yes | Yes | 34 | Debit Card | Weekly |

## 3. Exploratory data analysis using python

Exploratory data analysis was conducted in python to assess data structure, quality, and distributions. The dataset was loaded using pandas, with initial inspection performed using df.info() and df.describe(). Missing values in the review rating column were imputed using the median rating per product category. Columns were standardized to snake case,

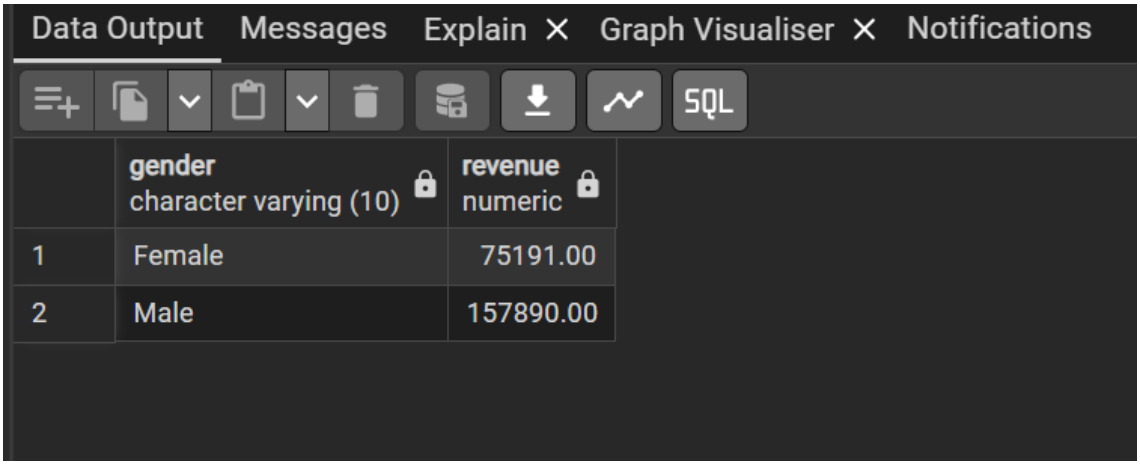
new features such as age groups and purchase frequency were engineered, redundant variables were removed, and the cleaned data was loaded into a postgresql database for sql analysis.

| | Customer ID | Age | Gender | Item Purchased | Category | Purchase Amount (USD) | Location | Size | Color | Season | Review Rating | Subscription Status | Shipping Type | Disco Appl |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 3900.000000 | 3900.000000 | 3900 | 3900 | 3900 | 3900.000000 | 3900 | 3900 | 3900 | 3900 | 3863.000000 | 3900 | 3900 | 3 |
| unique | NaN | NaN | 2 | 25 | 4 | NaN | 50 | 4 | 25 | 4 | NaN | 2 | 6 | |
| top | NaN | NaN | Male | Blouse | Clothing | NaN | Montana | M | Olive | Spring | NaN | No | Free Shipping | |
| freq | NaN | NaN | 2652 | 171 | 1737 | NaN | 96 | 1755 | 177 | 999 | NaN | 2847 | 675 | 2 |
| mean | 1950.500000 | 44.068462 | NaN | NaN | NaN | 59.764359 | NaN | NaN | NaN | NaN | 3.750065 | NaN | NaN | N |
| std | 1125.977353 | 15.207589 | NaN | NaN | NaN | 23.685392 | NaN | NaN | NaN | NaN | 0.716983 | NaN | NaN | N |
| min | 1.000000 | 18.000000 | NaN | NaN | NaN | 20.000000 | NaN | NaN | NaN | NaN | 2.500000 | NaN | NaN | N |
| 25% | 975.750000 | 31.000000 | NaN | NaN | NaN | 39.000000 | NaN | NaN | NaN | NaN | 3.100000 | NaN | NaN | N |
| 50% | 1950.500000 | 44.000000 | NaN | NaN | NaN | 60.000000 | NaN | NaN | NaN | NaN | 3.800000 | NaN | NaN | N |
| 75% | 2925.250000 | 57.000000 | NaN | NaN | NaN | 81.000000 | NaN | NaN | NaN | NaN | 4.400000 | NaN | NaN | N |
| max | 3900.000000 | 70.000000 | NaN | NaN | NaN | 100.000000 | NaN | NaN | NaN | NaN | 5.000000 | NaN | NaN | N |

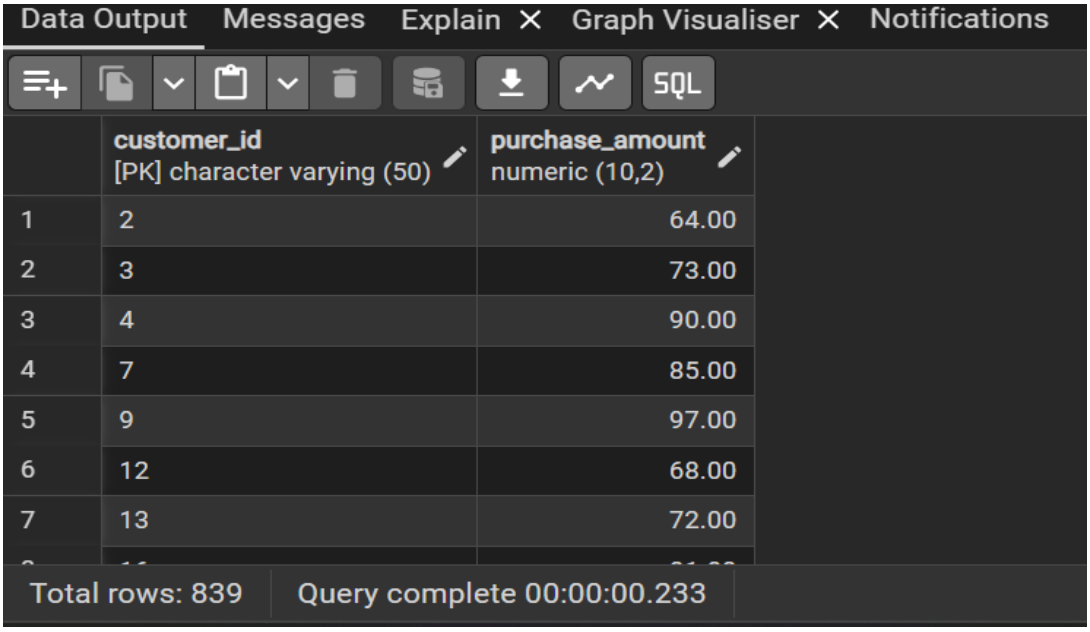
**4. Data analysis using sql (business transactions)**

We performed structured analysis in postgresql to answer key business questions:

1. Revenue by gender – compared total revenue generated by male vs. Female customers.




2. High-spending discount users – identified customers who used discounts but still spent above the average purchase amount.



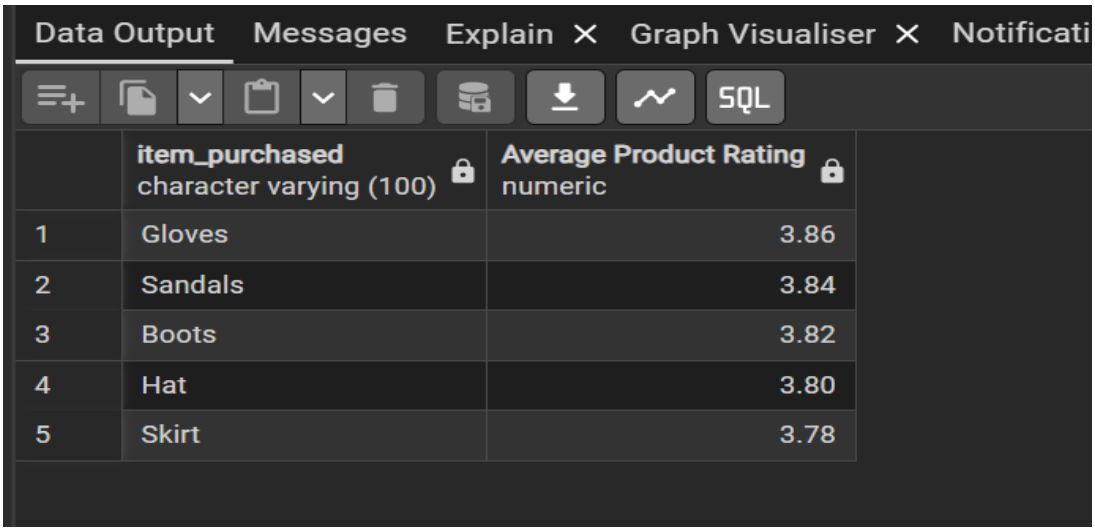
Data Output | Messages | Explain | Graph Visualiser | Notifications

SQL

| | customer_id [PK] character varying (50) | purchase_amount numeric (10,2) |
|---|---|---|
| 1 | 2 | 64.00 |
| 2 | 3 | 73.00 |
| 3 | 4 | 90.00 |
| 4 | 7 | 85.00 |
| 5 | 9 | 97.00 |
| 6 | 12 | 68.00 |
| 7 | 13 | 72.00 |

Total rows: 839 | Query complete 00:00:00.233

3. Top 5 products by rating – found products with the highest average review ratings.

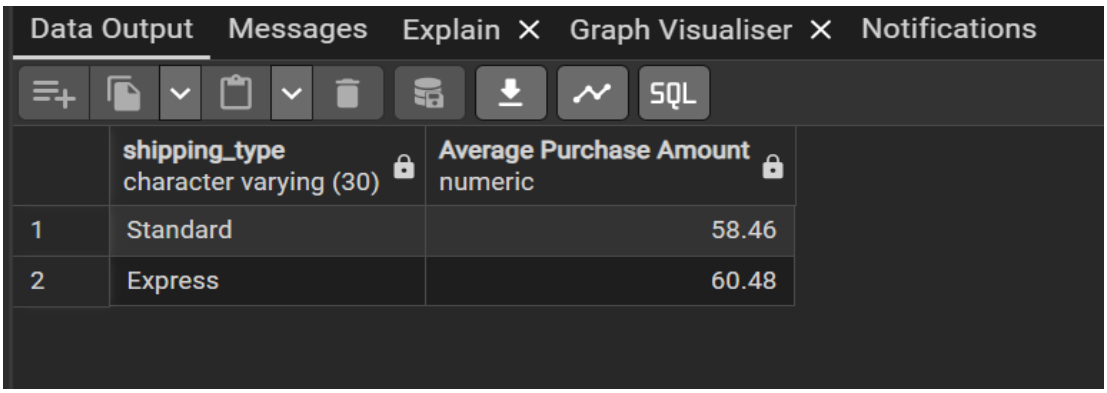


Data Output | Messages | Explain | Graph Visualiser | Notificati

SQL

| | item_purchased character varying (100) | Average Product Rating numeric |
|---|---|---|
| 1 | Gloves | 3.86 |
| 2 | Sandals | 3.84 |
| 3 | Boots | 3.82 |
| 4 | Hat | 3.80 |
| 5 | Skirt | 3.78 |

4. Shipping type comparison – compared average purchase amounts between standard and express shipping.

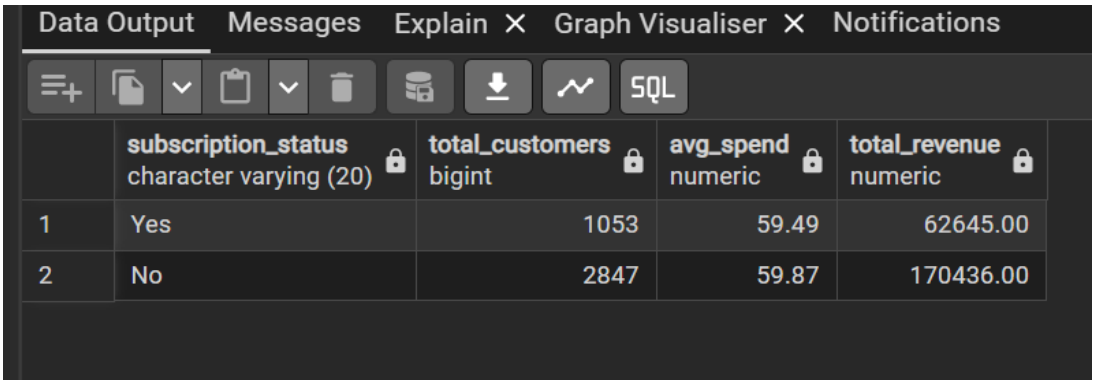


Data Output | Messages | Explain | Graph Visualiser | Notifications

SQL

| | shipping_type character varying (30) | Average Purchase Amount numeric |
|---|---|---|
| 1 | Standard | 58.46 |
| 2 | Express | 60.48 |

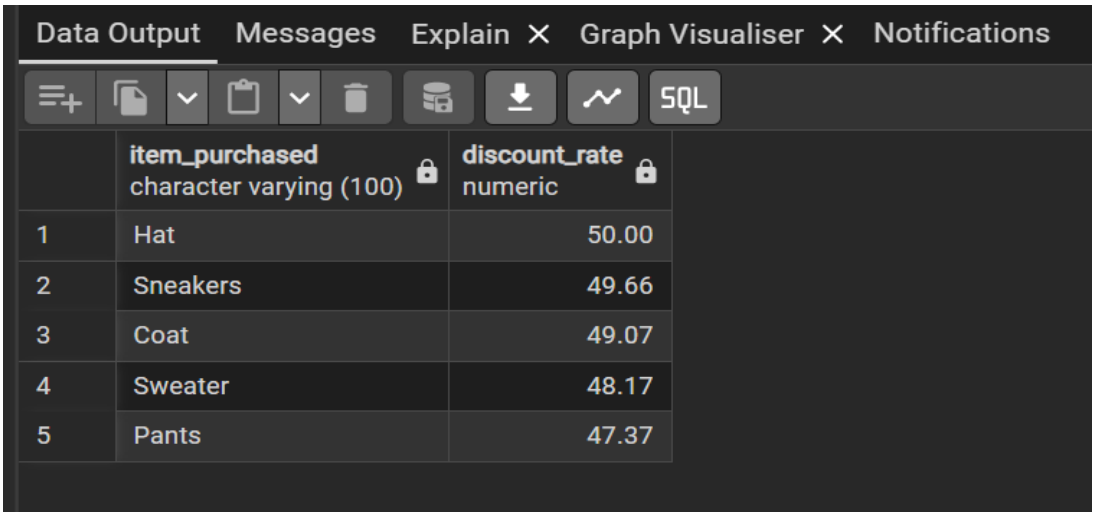
5. Subscribers vs. Non-subscribers – compared average spend and total revenue across subscription status.



Data Output | Messages | Explain | Graph Visualiser | Notifications

| | subscription_status character varying (20) | total_customers bigint | avg_spend numeric | total_revenue numeric |
|---|---|---|---|---|
| 1 | Yes | 1053 | 59.49 | 62645.00 |
| 2 | No | 2847 | 59.87 | 170436.00 |

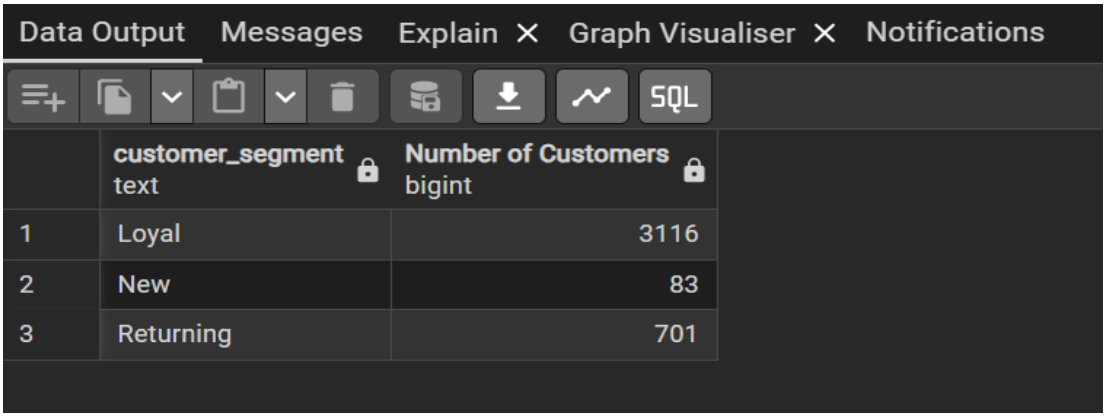
6. Discount-dependent products – identified five products with the highest percentage of discounted purchases.



Data Output | Messages | Explain | Graph Visualiser | Notifications

| | item_purchased character varying (100) | discount_rate numeric |
|---|---|---|
| 1 | Hat | 50.00 |
| 2 | Sneakers | 49.66 |
| 3 | Coat | 49.07 |
| 4 | Sweater | 48.17 |
| 5 | Pants | 47.37 |

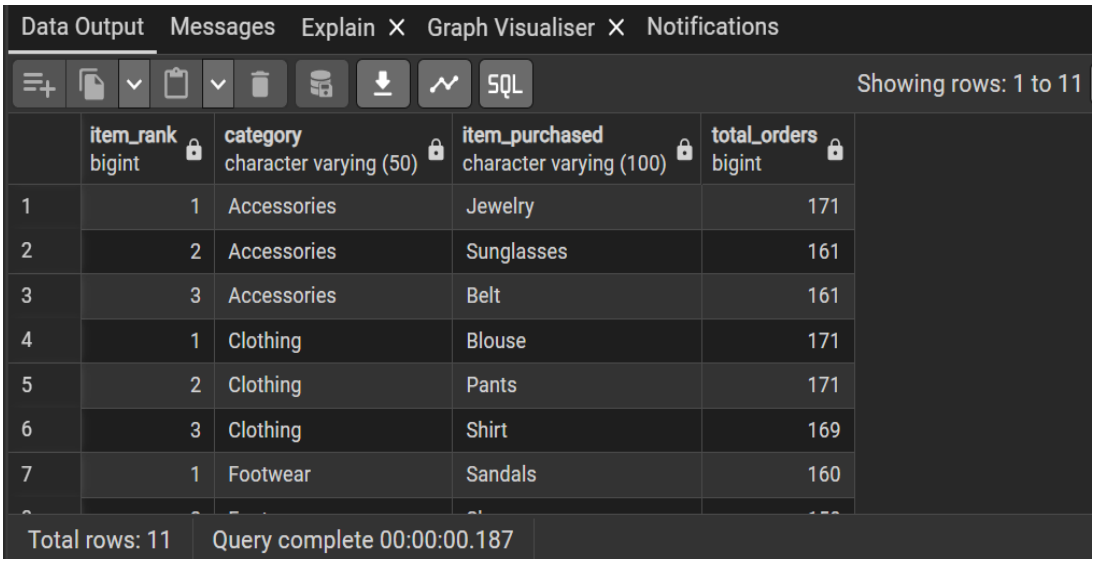
7. Customer segmentation – classified customers into new, returning, and loyal segments based on purchase history.



Data Output | Messages | Explain | Graph Visualiser | Notifications

| | customer_segment text | Number of Customers bigint |
|---|---|---|
| 1 | Loyal | 3116 |
| 2 | New | 83 |
| 3 | Returning | 701 |

8. Top 3 products per category – listed the most purchased products within each category.



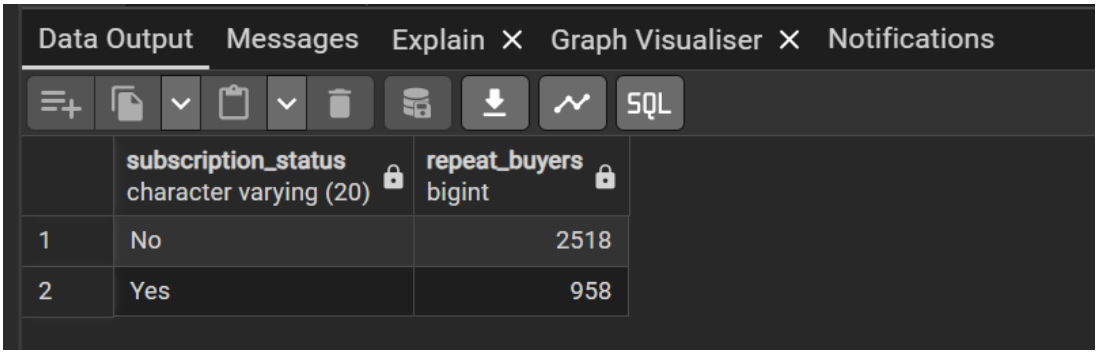
Data Output | Messages | Explain | Graph Visualiser | Notifications

Showing rows: 1 to 11

| | item_rank bigint | category character varying (50) | item_purchased character varying (100) | total_orders bigint |
|---|---|---|---|---|
| 1 | 1 | Accessories | Jewelry | 171 |
| 2 | 2 | Accessories | Sunglasses | 161 |
| 3 | 3 | Accessories | Belt | 161 |
| 4 | 1 | Clothing | Blouse | 171 |
| 5 | 2 | Clothing | Pants | 171 |
| 6 | 3 | Clothing | Shirt | 169 |
| 7 | 1 | Footwear | Sandals | 160 |

Total rows: 11 | Query complete 00:00:00.187

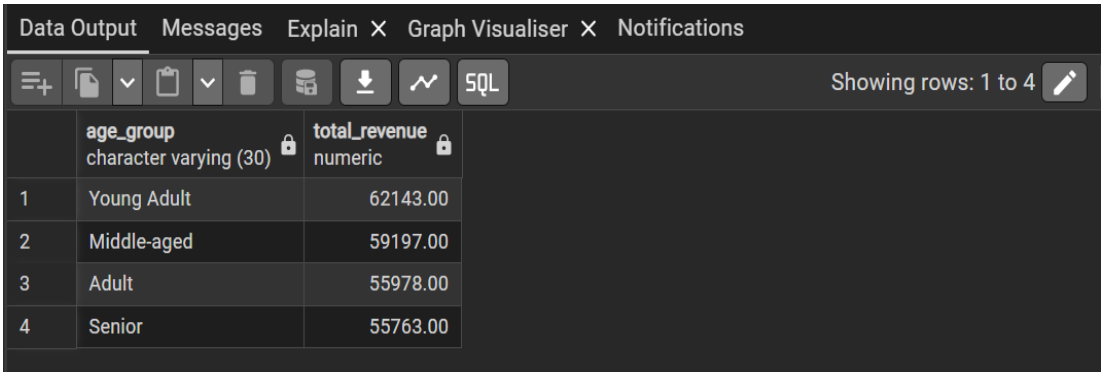
9. Repeat buyers & subscriptions – examined whether customers with more than five purchases were more likely to subscribe.



Data Output | Messages | Explain | Graph Visualiser | Notifications

| | subscription_status character varying (20) | repeat_buyers bigint |
|---|---|---|
| 1 | No | 2518 |
| 2 | Yes | 958 |

10. Revenue by age group – calculated total revenue contribution of each age group.



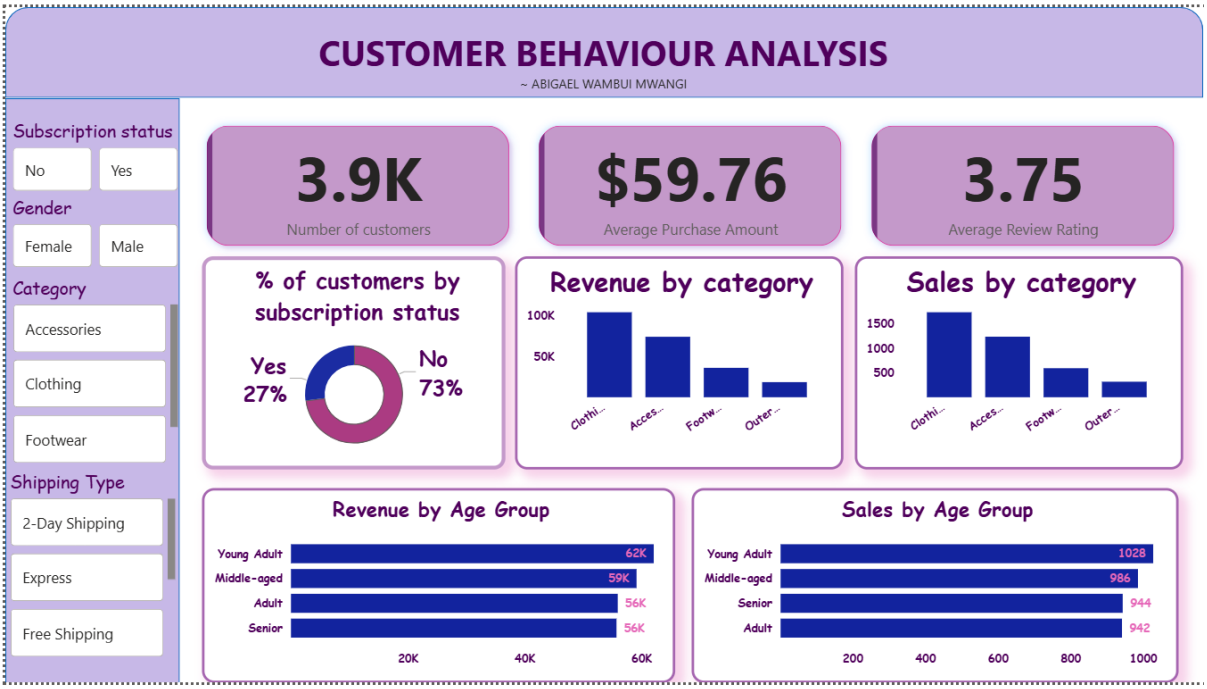
Data Output | Messages | Explain | Graph Visualiser | Notifications

Showing rows: 1 to 4

| | age_group character varying (30) | total_revenue numeric |
|---|---|---|
| 1 | Young Adult | 62143.00 |
| 2 | Middle-aged | 59197.00 |
| 3 | Adult | 55978.00 |
| 4 | Senior | 55763.00 |

## 5. Dashboard in Power Bi

An interactive power bi dashboard was developed to visually present insights related to customer behavior, revenue trends, product performance, and subscription patterns. It is as shown below;




## 6. Business recommendations

Key recommendations include;

1. Strengthening subscription offerings
2. Implementing customer loyalty programs
3. Reviewing discount strategies to protect margins
4. Highlighting top-rated and best-selling products
5. Applying targeted marketing to high-revenue age groups and express-shipping customers.