

# Customer Shopping Analysis

## 1. Project Overview

This project examines customer shopping behaviour using transactional data from 3,900 purchases across multiple product categories.

The purpose of the analysis is to understand:

- How customers spend their money
- Which products and categories perform best
- How different customer groups behave (age, gender, location)
- How discounts, shipping methods, and subscription status affect purchasing decisions
- What factors contribute to customer satisfaction and repeat purchases

## 2. Dataset Summary

- Total Rows: 3,900
- Total Columns: 18

Key Feature Groups

### 2.1. Customer Demographics

*Age, Gender, Location (State), Subscription Status*

### 2.2. Purchase Details

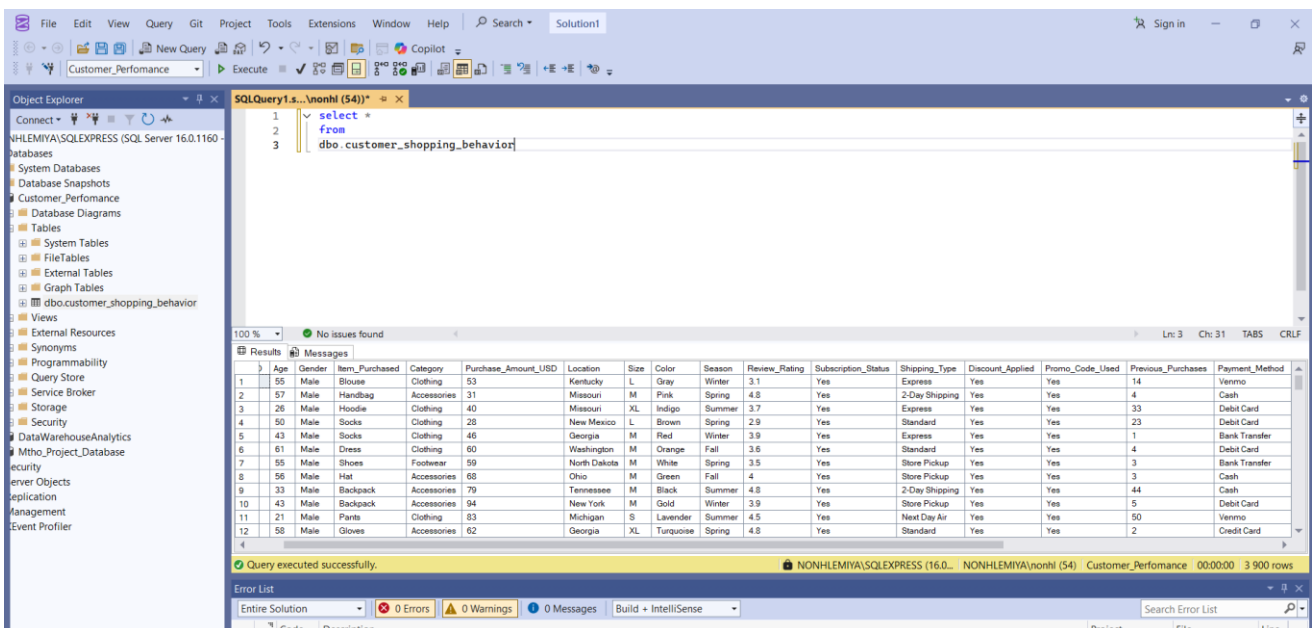
*Item Purchased, Category, Purchase Amount, Season, Size, Colour*

### 2.3. Shopping Behaviour

*Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type*

## 3. Exploratory Data Analysis (EDA)

I began with data preparation and cleaning in SQL and viewed data



The screenshot shows the SQL Server Enterprise Manager interface. The query window displays the following SQL query:

```
1 select *
2 from
3 dbo.customer_shopping_behavior
```

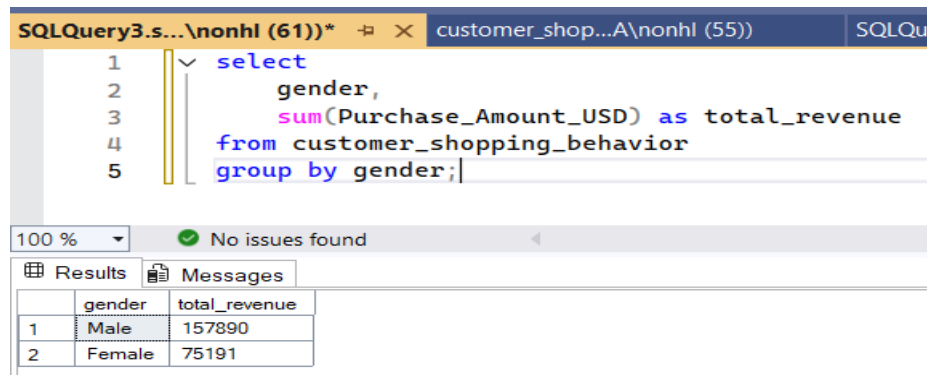
The results pane shows a table with 18 columns and 12 rows of data. The columns are: Age, Gender, Item\_Purchased, Category, Purchase\_Amount\_USD, Location, Size, Color, Season, Review\_Rating, Subscription\_Status, Shipping\_Type, Discount\_Applied, Promo\_Code\_Used, Previous\_Purchases, and Payment\_Method. The data is as follows:

	Age	Gender	Item_Purchased	Category	Purchase_Amount_USD	Location	Size	Color	Season	Review_Rating	Subscription_Status	Shipping_Type	Discount_Applied	Promo_Code_Used	Previous_Purchases	Payment_Method
1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Express	Yes	Yes	14	Venmo
2	57	Male	Handbag	Accessories	31	Missouri	M	Pink	Spring	4.8	Yes	2-Day Shipping	Yes	Yes	4	Cash
3	26	Male	Hoodie	Clothing	40	Missouri	XL	Indigo	Summer	3.7	Yes	Express	Yes	Yes	33	Debit Card
4	50	Male	Socks	Clothing	28	New Mexico	L	Brown	Spring	2.9	Yes	Standard	Yes	Yes	23	Debit Card
5	43	Male	Socks	Clothing	46	Georgia	M	Red	Winter	3.9	Yes	Express	Yes	Yes	1	Bank Transfer
6	61	Male	Dress	Clothing	60	Washington	M	Orange	Fall	3.6	Yes	Standard	Yes	Yes	4	Debit Card
7	55	Male	Shoes	Footwear	59	North Dakota	M	White	Spring	3.5	Yes	Store Pickup	Yes	Yes	3	Bank Transfer
8	56	Male	Hat	Accessories	68	Ohio	M	Green	Fall	4	Yes	Store Pickup	Yes	Yes	3	Cash
9	33	Male	Backpack	Accessories	79	Tennessee	M	Black	Summer	4.8	Yes	2-Day Shipping	Yes	Yes	44	Cash
10	43	Male	Backpack	Accessories	94	New York	M	Gold	Winter	3.9	Yes	Store Pickup	Yes	Yes	5	Debit Card
11	21	Male	Pants	Clothing	83	Michigan	S	Lavender	Summer	4.5	Yes	Next Day Air	Yes	Yes	50	Venmo
12	58	Male	Gloves	Accessories	62	Georgia	XL	Turquoise	Spring	4.8	Yes	Standard	Yes	Yes	2	Credit Card

#### 4. Data Analysis using SQL (Business Transactions)

I performed structured analysis in SQL Server to answer **key business questions**:

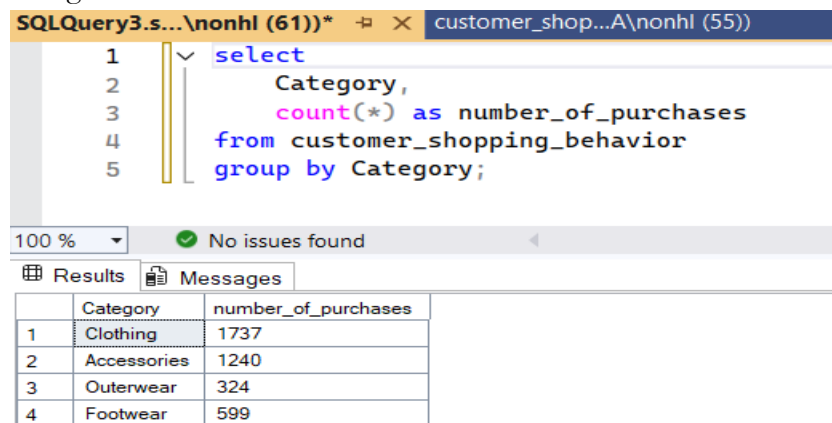
- 4.1. **Revenue by Gender** – Compared total revenue generated by male vs. female customers.



The screenshot shows a SQL query in SQL Server Enterprise Manager. The query is: `select gender, sum(Purchase_Amount_USD) as total_revenue from customer_shopping_behavior group by gender;`. The results pane shows a table with two rows: Male with a total revenue of 157890, and Female with a total revenue of 75191.

	gender	total_revenue
1	Male	157890
2	Female	75191

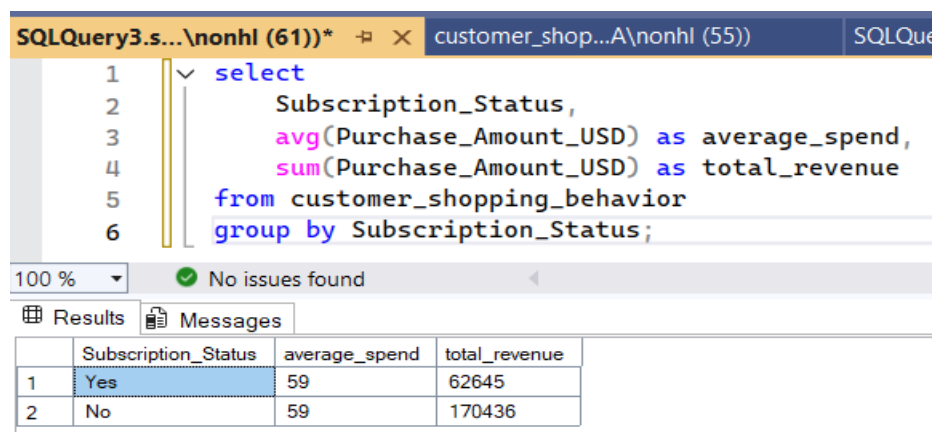
- 4.2. **Top 5 Products by Rating** – Found products with the highest average review ratings. **Dimensions**



The screenshot shows a SQL query in SQL Server Enterprise Manager. The query is: `select Category, count(*) as number_of_purchases from customer_shopping_behavior group by Category;`. The results pane shows a table with four rows: Clothing (1737), Accessories (1240), Outerwear (324), and Footwear (599).

	Category	number_of_purchases
1	Clothing	1737
2	Accessories	1240
3	Outerwear	324
4	Footwear	599

- 4.3. **Subscribers vs. Non-Subscribers** – Compared average spend and total revenue across subscription status. **Aggregates**



The screenshot shows a SQL query in SQL Server Enterprise Manager. The query is: `select Subscription_Status, avg(Purchase_Amount_USD) as average_spend, sum(Purchase_Amount_USD) as total_revenue from customer_shopping_behavior group by Subscription_Status;`. The results pane shows a table with two rows: Yes (average spend: 59, total revenue: 62645) and No (average spend: 59, total revenue: 170436).

	Subscription_Status	average_spend	total_revenue
1	Yes	59	62645
2	No	59	170436

- 4.5. **Customer Segmentation** – Classified customers into New, Returning, and Loyal segments based on purchase history. , **Common table Exploratory**

SQLQuery3.s... \nonhl (61))\* customer\_shop...A\nonhl (55)) SQLQuery1.sq...A\nonhl (54))\*

```

1  ;with cte as (
2      select
3          case
4              when Previous_Purchases = 1 then 'New'
5              when Previous_Purchases between 2 and 5 then 'Returning'
6              else 'Loyal'
7          end as Customer_Segment
8      from dbo.customer_shopping_behavior
9  )
10     select
11         Customer_Segment,
12         count(*) as number_of_customers
13     from cte
14     group by Customer_Segment
15     order by number_of_customers;

```

100 % No issues found

	Customer_Segment	number_of_customers
1	New	83
2	Returning	341
3	Loyal	3476

- 4.6. **Top 3 Products per Category** – Listed the most purchased products within each category. **Window function**

```

WITH Ranked_Products AS (
    SELECT
        Category,
        Item_Purchased,
        purchase_count,
        ROW_NUMBER() OVER (PARTITION BY Category ORDER BY purchase_count DESC) AS rank
    FROM (
        SELECT
            Category,
            Item_Purchased,
            COUNT(*) AS purchase_count
        FROM customer_shopping_behavior
        GROUP BY Category, Item_Purchased
    ) AS Sub
)
SELECT *
FROM Ranked_Products
WHERE rank <= 3;

```

100 % No issues found

	Category	Item_Purchased	purchase_count	rank
1	Accessories	Jewelry	171	1
2	Accessories	Sunglasses	161	2
3	Accessories	Belt	161	3
4	Clothing	Blouse	171	1
5	Clothing	Pants	171	2
6	Clothing	Shirt	169	3
7	Footwear	Sandals	160	1
8	Footwear	Shoes	150	2
9	Footwear	Sneakers	145	3
10	Outerwear	Jacket	163	1
11	Outerwear	Coat	161	2

- 4.7. **Revenue by Age Group** – Calculated total revenue contribution of each age group. **Window Function**

```

SELECT
    CASE
        WHEN Age BETWEEN 13 AND 19 THEN 'teenager'
        WHEN Age BETWEEN 20 AND 35 THEN 'young adult'
        WHEN Age BETWEEN 36 AND 55 THEN 'adult'
        ELSE 'senior'
    END AS Age_Group,
    SUM(Purchase_Amount_USD) AS total_revenue
FROM customer_shopping_behavior
GROUP BY
    CASE
        WHEN Age BETWEEN 13 AND 19 THEN 'teenager'
        WHEN Age BETWEEN 20 AND 35 THEN 'young adult'
        WHEN Age BETWEEN 36 AND 55 THEN 'adult'
        ELSE 'senior'
    END
ORDER BY total_revenue desc;

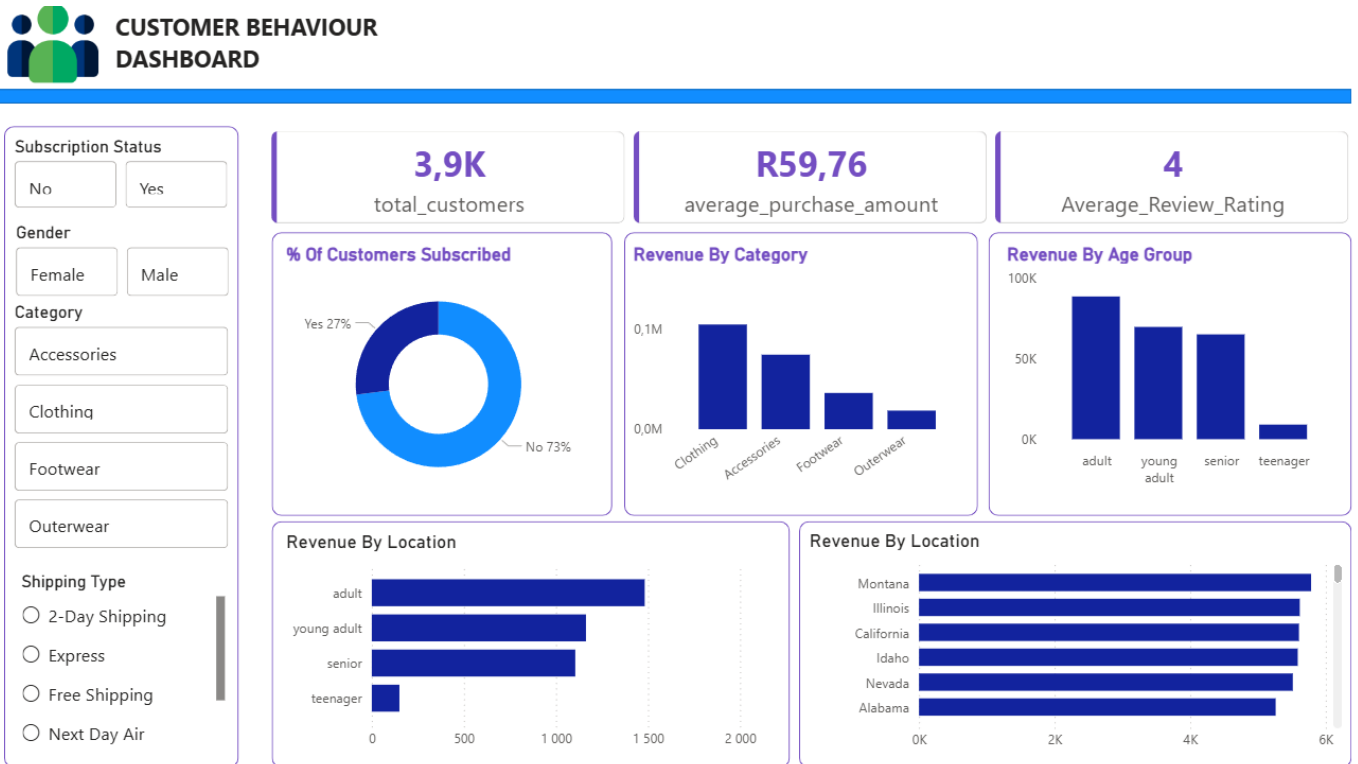
```

100 % No issues found

	Age_Group	total_revenue
1	adult	88853
2	young adult	69892
3	senior	65256
4	teenager	9080

## 5. Dashboard

Below is the I built an interactive dashboard in Power BI to present insights visually



## 6. Business Recommendation

Based on the analysis I have conducted above, following are the recommendations:

- **Boost Subscriptions** – Promote exclusive benefits for subscribers
- **Customer Loyalty Programs** – Reward repeats buyers to move them into the “Loyal” segment.