

Data Driven Analysis Customer Shopping Trends

1. Problem Statement

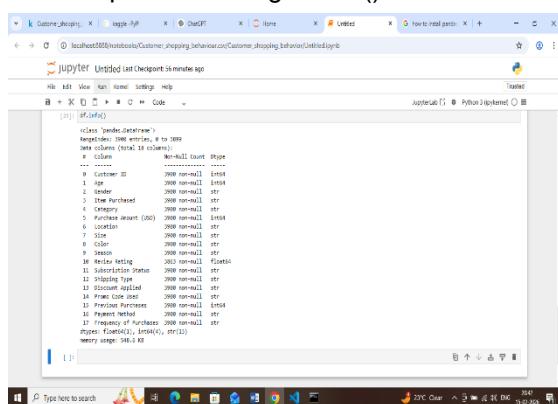
A leading retail company wants to better understand its customers' shopping behavior to improve sales, customer satisfaction, and long-term loyalty. The objective is to analyze purchasing patterns across demographics, product categories, and sales channels to identify key drivers such as discounts, reviews, seasons, and payment preferences that influence repeat purchases and engagement.

2. Dataset Summary

- Rows: 3,900
- Columns: 18
- Customer Demographics: Age, Gender, Location, Subscription Status
- Purchase Details: Item Purchased, Category, Purchase Amount, Season, Size, Color
- Shopping Behavior: Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type
- Missing Data: 37 values in Review Rating column

3. Data Preparation & Cleaning (Python)

- Data Loading using pandas
- Initial Exploration using df.info() and df.describe()

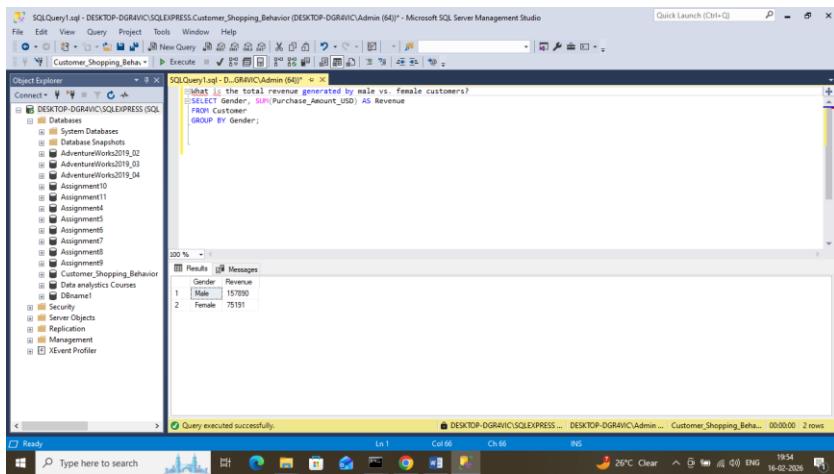


```
[1]: df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3900 entries, 0 to 3900
Data columns (total: 18 columns):
 #   column      : dtype : format:
 0   Customer ID : int64 : 0 to 3900
 1   Age          : int64 : 1 to 100
 2   Gender        : int64 : 0 or 1
 3   Item Purchased : int64 : str
 4   Category       : int64 : str
 5   Purchase amount (USD) : int64 : str
 6   Location        : int64 : str
 7   Season          : int64 : str
 8   Color           : int64 : str
 9   Size            : int64 : str
 10  Review Rating  : float64 : float64
 11  Subscription Status : int64 : str
 12  Previous purchases : int64 : str
 13  Discount applied : int64 : str
 14  Promo code used : int64 : str
 15  Payment method  : int64 : str
 16  Shipping type   : int64 : str
 17  Date            : datetime64[ns] : 2023-01-01 to 2023-12-31
memory usage: 545.0 kB
```

- Handled missing values by imputing median rating per category
- Standardized column names to snake_case
- Feature Engineering: Created age_group and purchase_frequency_days
- Data consistency check for discount_applied and promo_code_used
- Integrated cleaned data into SQL Server Management for SQL analysis

4. Data Analysis Using SQL

- Revenue analysis by gender



- Top 3 products per category

```
--Q1. Which are the top 5 products with the highest average review rating?  
SELECT TOP 5  
    Item_Purchased  
    ,AVG(Review_Rating) AS [Product Rating] , 2 AS [Average Product Rating]  
FROM Customer  
GROUP BY Item_Purchased  
ORDER BY [Average Product Rating] DESC  
  
--Q2. Which customers used a discount but still spent more than the average purchase amount?  
SELECT Customer_ID , Purchase_Amount_USD  
FROM Customer  
WHERE Discount_Applied = 'Yes' AND Purchase_Amount_USD > (SELECT AVG(Purchase_Amount_USD) FROM customer)
```

Item_Purchased	Average Product Rating
1	3.96
2	3.94
3	3.92
4	3.91
5	3.76

- High-spending discount users

Customer_Shopping_Behavior

Customer_ID Purchase_Amount_USD

Customer_ID	Purchase_Amount_USD
1	2
2	73
3	35
4	7
5	9
6	12
7	22
8	16
9	20
10	22
11	24

- #### • Customer segmentation insights

The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface. The title bar reads "SQL Query - DESKTOP-DIGITAL\ADMIN, Shopping_Behavior [master] - Microsoft SQL Server Management Studio". The left sidebar shows the database structure for "Customer_Shopping_Behavior". The main pane displays a T-SQL query:

```
SELECT Customer_type, COUNT(*) AS "Number of Customers"
FROM Customer_Segment
GROUP BY Customer_type
ORDER BY Customer_type
```

The results pane shows the output:

Customer_type	Number of Customers
1. Existing	311
2. First	316
3. New	33

At the bottom, a message says "Query executed successfully." The status bar at the bottom right shows "DESKTOP-DIGITAL\ADMIN\Customer_Shopping_Behavior [master] 00:00:02 7 rows".

- Sales trends by age group and shipping type

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left lists the database structure, including databases like 'AdventureworksLT', 'AdventureworksLTDW', and 'AdventureworksLTReport'. The 'Customer_Shopping_Behavior' database is selected. The Results pane on the right displays the output of two queries. The first query retrieves the top 5 products with the highest average review rating:

```
Q1. which are the top 5 products with the highest average review rating?  
-----  
SELECT Item_Purchased,  
       ROUND(AVG(CAST([Review_Rating] AS FLOAT)), 2) AS [Average Product Rating]  
FROM [AdventureworksLT].[dbo].[FactInternetSales]  
GROUP BY Item_Purchased  
ORDER BY [Average Product Rating] DESC;
```

The second query compares the average purchase amounts between Standard and Express shipping:

```
--Q2. Compare the average purchase Amounts between Standard and Express Shipping.  
-----  
select Shipping_Type,  
       round(Avg(TotalDue), 2)  
from Customer  
where Shipping_Type = 'Standard' or 'Express'  
group by Shipping_Type;
```

The Results pane shows the following data:

Shipping_Type	Avg(TotalDue)
Express	60
Standard	58

At the bottom, a status bar indicates "Query executed successfully." and the results are displayed across multiple tabs.

5. Tools & Technologies

- Microsoft Power BI
- Power Query (Data Cleaning & Transformation)
- DAX (Data Analysis Expressions)
- Data Modeling

6. Key Dashboard Metrics

- **Total Customers:** 3,900
- **Average Review Rating:** 3.76
- **Average Purchase Amount:** \$59.57
- Revenue by Category
- Sales by Age Group
- Subscription-Based Customer %

7. Key Insights

- Customers with subscriptions show higher engagement.
- Clothing and Accessories contribute significantly to revenue.
- Middle-aged customers form the largest purchasing segment.
- Average rating indicates moderate customer satisfaction.

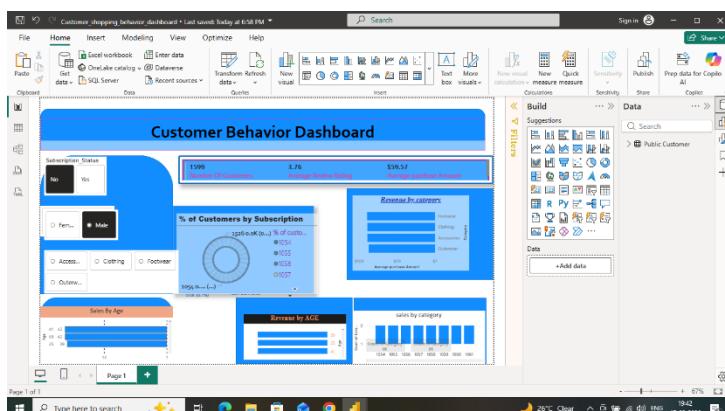
8. DAX Measures Used

Examples:

- Total Customers = COUNT(Customer[Customer ID])
- Average Purchase = AVERAGE(Customer[Purchase Amount])
- Average Rating = AVERAGE(Customer[Review Rating])

9. Dashboard Features

- Interactive slicers (Gender, Category, Subscription Status)
- KPI Cards
- Donut & Bar Charts
- Category-wise revenue analysis
- Age-based sales comparison



10. Business Recommendations

- Boost Subscriptions: Promote exclusive benefits for subscribers
- Customer Loyalty Programs: Reward repeat buyers to move them into the loyal segment
- Review Discount Policy: Balance sales boosts with margin control
- Product Positioning: Highlight top-rated and best-selling products in campaigns
- Targeted Marketing: Focus on high-revenue age groups and express-shipping users