

Data Driven Analysis Customer Shopping Trends

1. Problem Statement

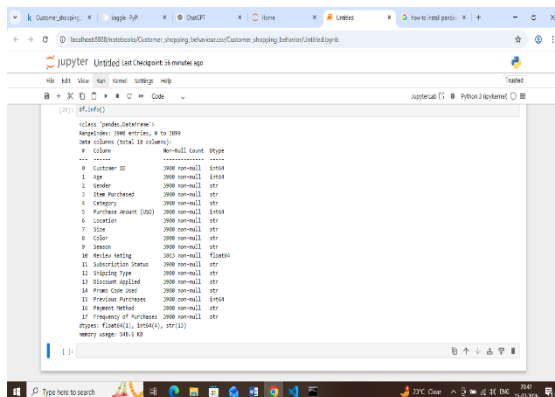
A leading retail company wants to better understand its customers' shopping behavior to improve sales, customer satisfaction, and long-term loyalty. The objective is to analyze purchasing patterns across demographics, product categories, and sales channels to identify key drivers such as discounts, reviews, 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()`



```
Out[1]:
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3900 entries, 0 to 3899
Data columns (total 18 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   Customer ID           3900 non-null   int64  
 1   Age                   3900 non-null   int64  
 2   Gender                3900 non-null   object  
 3   Item Purchased         3900 non-null   object  
 4   Category              3900 non-null   object  
 5   Purchase amount (USD)  3900 non-null   float64 
 6   Location              3900 non-null   object  
 7   Size                  3900 non-null   object  
 8   Color                 3900 non-null   object  
 9   Season                3900 non-null   object  
10  Review Rating         3663 non-null   float64 
11  Subscription Status   3900 non-null   object  
12  Shipping Type         3900 non-null   object  
13  Discount Applied      3900 non-null   object  
14  Promo Code Used       3900 non-null   object  
15  Previous Purchases     3900 non-null   int64  
16  Payment Method        3900 non-null   object  
17  Frequency of Purchases 3900 non-null   int64  
dtypes: float64(2), int64(4), object(12)
memory usage: 588.0 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

SQLQuery1.sql - DESKTOP-DGRAIVC\SQLEXPRESS\Customer_Shopping_Behavior (DESKTOP-DGRAIVC\Admin (84)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

Customer_Shopping_Beha... Execute

Object Explorer

- DESKTOP-DGRAIVC\SQLEXPRESS\SOL
- Databases
 - System Databases
 - Database Snapshots
 - AdventureWorks2019_02
 - AdventureWorks2019_03
 - AdventureWorks2019_04
 - Assignment10
 - Assignment11
 - Assignment4
 - Assignment5
 - Assignment6
 - Assignment7
 - Assignment8
 - Assignment9
 - Customer_Shopping_Behavior
 - Data analytics Course
 - DBName1
 - Security
 - Server Objects
 - Replication
 - Management
 - Event Profile

SQLQuery1.sql - DESKTOP-DGRAIVC\Admin (84) *
What is the total revenue generated by male vs. female customers?
SELECT Gender, SUM(Purchase_Amount_USD) AS Revenue
FROM Customer
GROUP BY Gender;

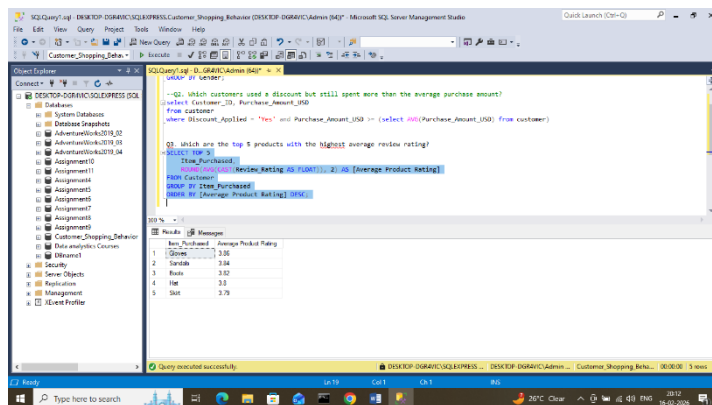
Results Messages

	Gender	Revenue
1	Male	157890
2	Female	79191

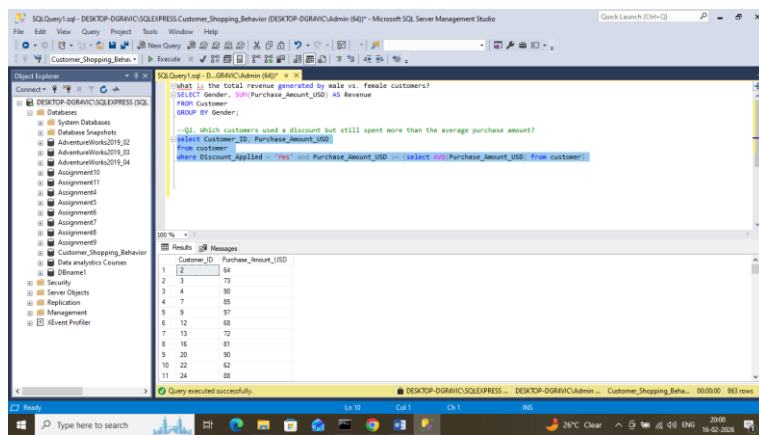
Query executed successfully. DESKTOP-DGRAIVC\SQLEXPRESS... DESKTOP-DGRAIVC\Admin... Customer_Shopping_Beha... 00:00:00 2 rows

Ready Type here to search 25°C Clear 18:54 16-02-2025

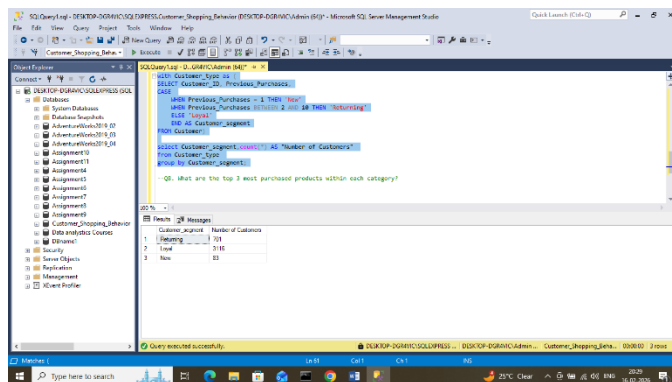
- Top 3 products per category



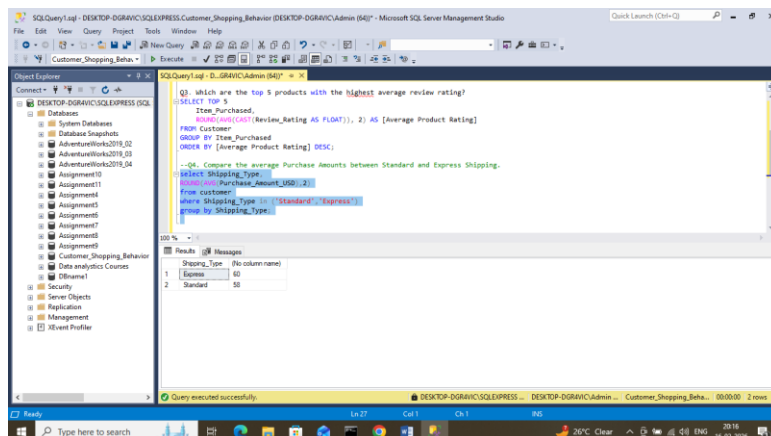
- High-spending discount users



- Customer segmentation insights



- Sales trends by age group and shipping type



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.

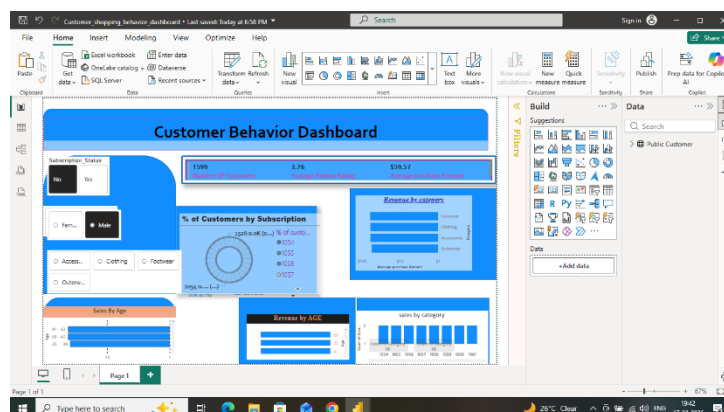
8. DAX Measures Used

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