Customer Shopping Behavior Analysis

1. Project Overview

This project analyzes customer shopping behavior using transactional data from 1,000 purchases across various product categories. The goal is to uncover insights into spending patterns, customer segments, product preferences, and subscription behavior to guide strategic business decisions.

2. Dataset Summary

- Rows: 1,000
- Columns: 11
- Key Features:
  - Customer demographics (Customer id, Age, Gender, Age Group)
  - Purchase details (Transaction id, Product Category, Total Amount, Date, Quarters, Quantity, Price per Unit)

3. Exploratory Data Analysis using Python

We began with data preparation and cleaning in Python:

- **Data Loading:** Imported the dataset using `pandas`.
- **Initial Exploration:** Used `df.info()` to check structure and `.describe()` for summary statistics.

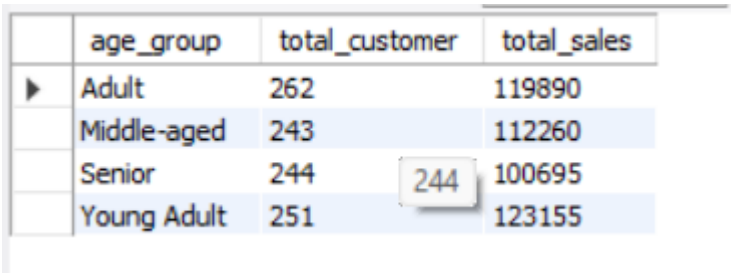
	Transaction ID	Date	Customer ID	Gender	Age	Product Category	Quantity	Price per Unit	Total Amount
count	1000.000000	1000	1000	1000	1000.00000	1000	1000.000000	1000.000000	1000.000000
unique	NaN	345	1000	2	NaN	3	NaN	NaN	NaN
top	NaN	2023-05-16	CUST001	Female	NaN	Clothing	NaN	NaN	NaN
freq	NaN	11	1	510	NaN	351	NaN	NaN	NaN
mean	500.500000	NaN	NaN	NaN	41.39200	NaN	2.514000	179.890000	456.000000
std	288.819436	NaN	NaN	NaN	13.68143	NaN	1.132734	189.681356	559.997632
min	1.000000	NaN	NaN	NaN	18.00000	NaN	1.000000	25.000000	25.000000
25%	250.750000	NaN	NaN	NaN	29.00000	NaN	1.000000	30.000000	60.000000
50%	500.500000	NaN	NaN	NaN	42.00000	NaN	3.000000	50.000000	135.000000
75%	750.250000	NaN	NaN	NaN	53.00000	NaN	4.000000	300.000000	900.000000
max	1000.000000	NaN	NaN	NaN	64.00000	NaN	4.000000	500.000000	2000.000000

- **Missing Data Handling:** Checked for null values and imputed missing values in the columns using the median rating of each product category.
- **Column Standardization:** Renamed columns to **snake case** for better readability and documentation.
- **Feature Engineering:**
    - Created **age_group** column by binning customer ages.
    - Created **quarters** column from date.
- **Database Integration:** Connected Python script to MYSQL and loaded the cleaned DataFrame into the database for SQL analysis.

4. Data Analysis using SQL (Business Transactions)

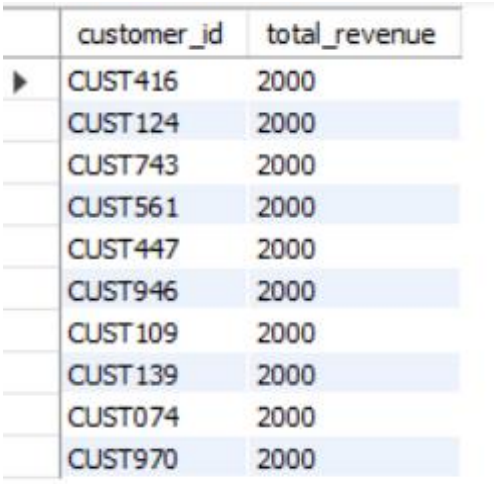
We performed structured analysis in PostgreSQL to answer key business questions:

1. How many unique customers belong to each age group, and what is the total sales generated by each group?



age_group	total_customer	total_sales
Adult	262	119890
Middle-aged	243	112260
Senior	244	100695
Young Adult	251	123155

2. Which are the top 10 customers based on total revenue generated?



customer_id	total_revenue
CUST416	2000
CUST124	2000
CUST743	2000
CUST561	2000
CUST447	2000
CUST946	2000
CUST109	2000
CUST139	2000
CUST074	2000
CUST970	2000

3. Which customers have made more than one purchase?

customer_id	purchase_count

4. What is the total sales amount for each month?

month	total_sales
2023-01	35450
2023-02	44060
2023-03	28990
2023-04	33870
2023-05	53150
2023-06	36715
2023-07	35465
2023-08	36960
2023-09	23620
2023-10	46580
2023-11	34920
2023-12	44690
2024-01	1530

5. which product category generates the highest total sales?

product_category	total_sales
Electronics	156905
Clothing	155580
Beauty	143515

6. What is the total lifetime spending of each customer?

customer_id	lifetime_amount
CUST001	150
CUST002	1000
CUST003	30
CUST004	500
CUST005	100

--it return all customer_id, lifetime_amount column records

7. Which customers have total spending higher than the overall average spending?

customer_id	total_spent
CUST002	1000
CUST004	500
CUST009	600
CUST013	1500
CUST015	2000
CUST016	1500
CUST020	900
CUST021	500
CUST026	1000
CUST028	500
CUST030	900
CUST031	1200
CUST035	900

8. What are the total and average sales for each gender?

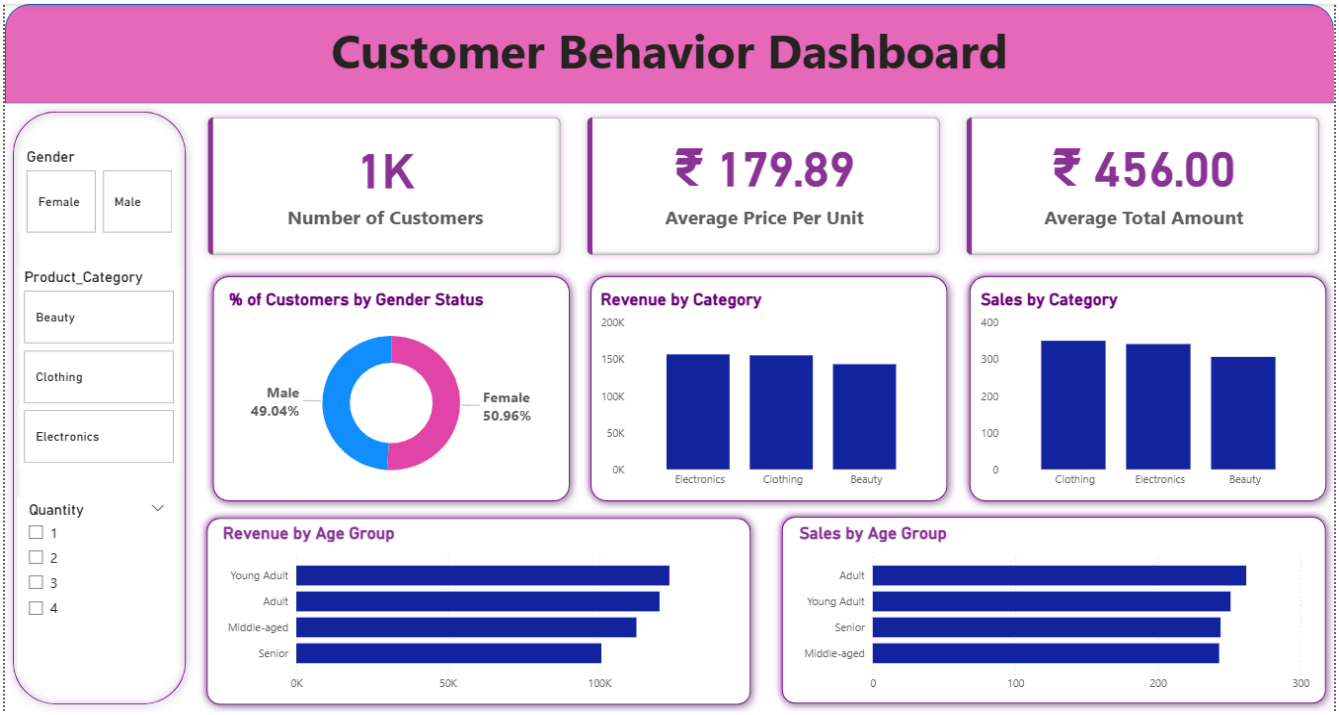
gender	total_sales	avg_sales
Male	223160	455.43
Female	232840	456.55

9. What is the total sales generated in each quarter?

quarters	total_sales
Q1	122870
Q2	108335
Q3	106300
Q4	118495

5. Dashboard in Power BI

Finally, we built an interactive dashboard in **Power BI** to present insights visually.



6. Business Recommendations

- **Product Positioning** – Highlight top-rated and best-selling products in campaigns.
- **Targeted Marketing** – Focus efforts on high-revenue age groups and express-shipping users.