

# Customer Data Analysis

## 1. Project Overview

The Customer Data Analysis project focuses on understanding customer purchasing behaviour based on gender, age, payment method, and product categories.

Using SQL for data extraction and Power BI for visualization, the aim is to derive meaningful business insights that help in better decision-making, customer segmentation, and strategic planning.

## 2. Dataset Summary

The dataset contains 1000 customer transactions recorded across multiple shopping malls.

It includes demographic details, purchase information, and transaction attributes such as category, quantity, revenue, and payment method.

Rows: 1000

Columns: 9

## 3. Project Objectives

The project is designed to answer the following business questions:

1. How is shopping distribution according to gender?
2. Which gender purchased more quantity?
3. Which gender generated more revenue?
4. What is the category distribution?
5. How is shopping distribution according to age?
6. Which age group purchased more quantity?
7. Which age group generated more revenue?
8. How category varies with gender/age?
9. How payment method is related to other columns?
10. What is the distribution of payment methods?

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

Below are the primary SQL queries used to answer the business questions.

### 1. Shopping distribution by gender

```
SELECT gender, COUNT(*) AS total_transactions
FROM customer
GROUP BY gender
```

gender	total_transactions
Female	59482
Male	39975

### 2. Quantity by gender

```
SELECT gender, SUM(quantity) AS total_quantity
FROM customer
GROUP BY gender;
```

gender	total_quantity
Female	178659
Male	120053

### 3. Revenue by gender

```
SELECT gender, SUM(quantity * price) AS total_revenue
FROM customer
GROUP BY gender;
```

gender	total_revenue
Female	150207136.02
Male	101298658.23

#### 4. Category distribution

```
1
2 • SELECT category , COUNT(*) AS total_transactions
3 FROM customer
4 GROUP BY category;
```

category	total_transactions
Books	4981
Clothing	34487
Cosmetics	15097
Food & Beverage	14776
Shoes	10034
Souvenir	4999
Technology	4996
Toys	10087

#### 5. Age group distribution

```
Query 1
1 • SELECT
2 CASE
3 WHEN age BETWEEN 18 AND 25 THEN '18-25'
4 WHEN age BETWEEN 26 AND 35 THEN '26-35'
5 WHEN age BETWEEN 36 AND 45 THEN '36-45'
6 WHEN age BETWEEN 46 AND 55 THEN '46-55'
7 ELSE '56+'
8 END AS age_group,
9 category,
10 COUNT(*) AS total_transactions
11 FROM customer
12 GROUP BY age_group, category
13 ORDER BY age_group, category;
```

age_group	category	total_transactions
18-25	Books	788
18-25	Clothing	5337
18-25	Cosmetics	2276
18-25	Food & Beverage	2306
18-25	Shoes	1550
18-25	Souvenir	768
18-25	Technology	751
18-25	Toys	1583
26-35	Books	920
26-35	Clothing	6614
26-35	Cosmetics	2895
26-35	Food & Beverage	2820

## 6. Quantity by age group

```
2 CASE
3   WHEN age BETWEEN 18 AND 25 THEN '18-25'
4   WHEN age BETWEEN 26 AND 35 THEN '26-35'
5   WHEN age BETWEEN 36 AND 45 THEN '36-45'
6   WHEN age BETWEEN 46 AND 55 THEN '46-55'
7   ELSE '56+'
8 END AS age_group,
9   sum(quantity) AS total_quantity
10 FROM customer
11 GROUP BY age_group;
```

The screenshot shows a data visualization interface. At the top, there is a 'Result Grid' tab, a grid icon, a refresh icon, and a 'Filter Rows:' input field. To the right is an 'Export:' button with a download icon. Below this is a table with two columns: 'age\_group' and 'total\_quantity'.

age_group	total_quantity
18-25	46028
26-35	57265
36-45	58483
46-55	57086
56+	79850

[sult 14](#) ×

## 7. Revenue by age group

```
Query 1 ×
SELECT
CASE
WHEN age BETWEEN 18 AND 25 THEN '18-25'
WHEN age BETWEEN 26 AND 35 THEN '26-35'
WHEN age BETWEEN 36 AND 45 THEN '36-45'
WHEN age BETWEEN 46 AND 55 THEN '46-55'
ELSE '56+'
END AS age_group,
sum(quantity*price) AS total_revenue
FROM customer
GROUP BY age_group;
```

Result Grid | Filter Rows:  | Export: | Wrap Cell Content:

age_group	total_revenue
18-25	38118271.35
26-35	47879659.58
36-45	50184235.79
46-55	48219742.64
56+	67103884.89

## 8. Category vs gender

```
1
2 • SELECT category , gender, COUNT(*) AS total_transactions
3 FROM customer
4 GROUP BY category, gender
5 ORDER BY category, gender;
```


category	gender	total_transactions
Books	Female	2906
Books	Male	2075
Clothing	Female	20652
Clothing	Male	13835
Cosmetics	Female	9070
Cosmetics	Male	6027
Food & Beverage	Female	8804
Food & Beverage	Male	5972
Shoes	Female	5967
Shoes	Male	4067
Souvenir	Female	3017
Souvenir	Male	1982
Technology	Female	2981
Technology	Male	2015
Toys	Female	6085
Toys	Male	4002

## 9. Payment method vs category

```
1 • SELECT category, payment_method,
2 COUNT(*) AS total_transactions
3 FROM customer
4 GROUP BY category, payment_method
5 ORDER BY category, payment_method;
```

category	payment_method	total_transactions
Books	Cash	2268
Books	Credit Card	1696
Books	Debit Card	1017
Clothing	Cash	15456
Clothing	Credit Card	12025
Clothing	Debit Card	7006
Cosmetics	Cash	6674
Cosmetics	Credit Card	5336
Cosmetics	Debit Card	3087
Food & Beverage	Cash	6587
Food & Beverage	Credit Card	5250
Food & Beverage	Debit Card	2939
Shoes	Cash	4474
Shoes	Credit Card	3553
Shoes	Debit Card	2007
Souvenir	Cash	2211
Souvenir	Credit Card	1770
Souvenir	Debit Card	1018
Technology	Cash	2235
Technology	Credit Card	1753
Technology	Debit Card	1008
Toys	Cash	4542
Toys	Credit Card	3548
Toys	Debit Card	1997

10. Payment method distribution



1 •

2



3

4

5

```
SELECT payment_method,  
COUNT(*) AS total_transactions  
FROM customer  
GROUP BY payment_method;
```

Result Grid



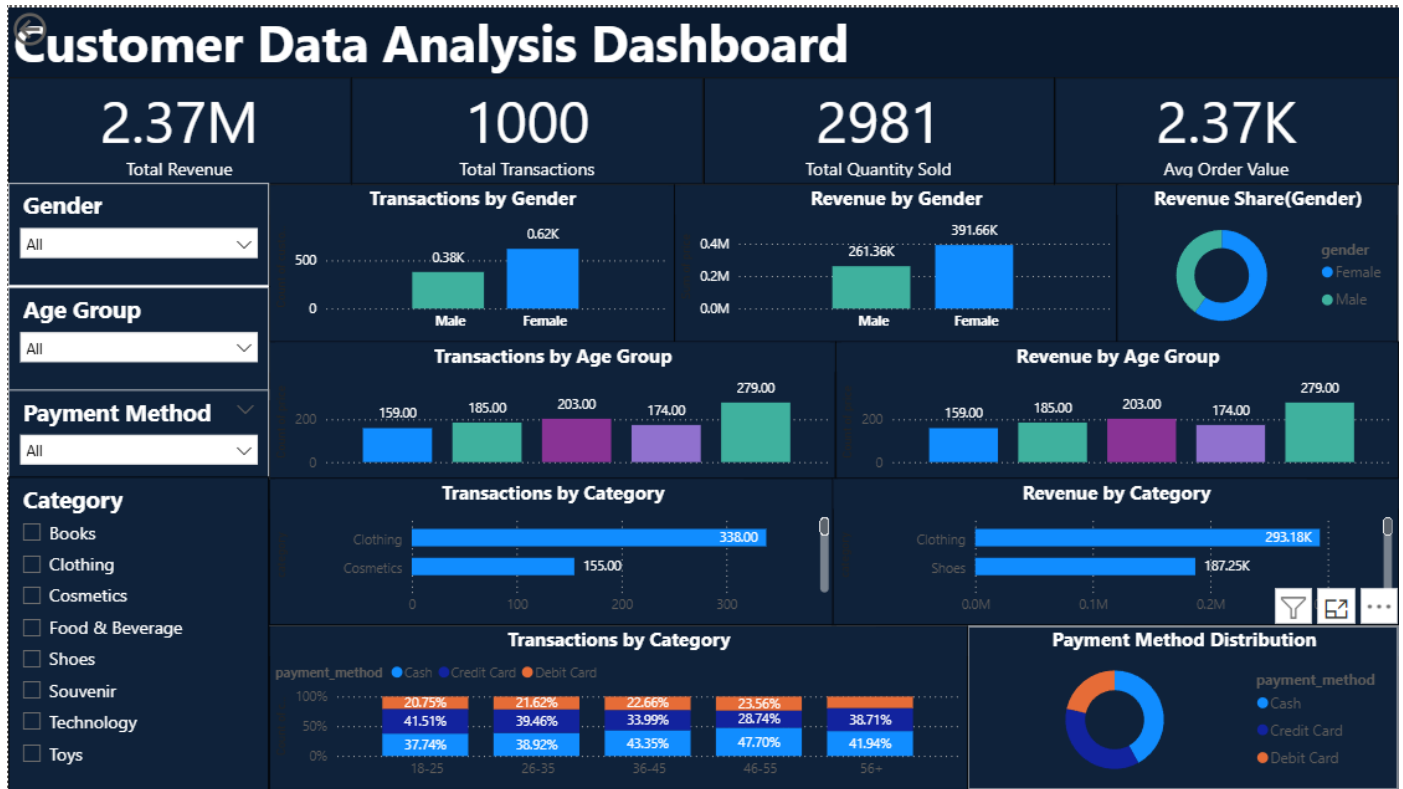
Filter Rows:

E

	payment_method	total_transactions
▶	Cash	44447
	Credit Card	34931
	Debit Card	20079

## 5. Dashboard in Power BI

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



## 6. Business Summary

1. **Female customers contribute higher total transactions compared to male customers**, making them the dominant shopping group in terms of volume.
2. **Male customers generate slightly higher revenue per transaction**, even though their total transactions are fewer, indicating higher-value purchases.
3. **The age group 46–55 contributes the highest total revenue**, followed by 36–45, making them the most valuable customer segments.
4. **Younger customers (18–25) show lower transaction counts and revenue**, suggesting limited purchasing power or fewer visits.
5. **Clothing, Shoes, and Accessories are the top-performing product categories**, consistently generating the highest revenue.
6. **Product categories like Souvenir, Books, and Cosmetics generate lower revenue**, indicating they are less preferred by customers.
7. **Credit Card is the most preferred payment method**, followed by Cash; Debit Card is the least used among all customer segments.
8. **Female customers purchase more across nearly all product categories**, showing strong engagement and shopping activity.

9. **Payment preferences vary slightly by gender**, but Credit Card remains the top choice for both male and female customers.
10. **Overall shopping trends show that middle-aged customers (36–55 years) are the most active and high-spending group**, making them the ideal target for marketing campaigns.

## 7. Business Recommendations

### 1. Strengthen Engagement with Female Customers

Female customers contribute the highest share of transactions and revenue. Implement tailored campaigns, exclusive collections, and targeted retention strategies to maximize their revenue potential.

### 2. Prioritize High-Performing Age Segments (36–55 Years)

Customers aged 36–55 are the most valuable demographic in terms of quantity purchased and total revenue. Introduce premium bundles, age-specific promotions, and loyalty benefits to deepen engagement within this segment.

### 3. Capitalize on Top-Selling Categories

Clothing, Shoes, and Accessories consistently drive the majority of sales. Feature these categories prominently across marketing channels, homepage placements, and recommendation engines to strengthen overall revenue.

### 4. Revitalize Underperforming Categories

Categories such as Books, Souvenir, and Cosmetics show lower contribution. Introduce visibility campaigns, bundle offers, or limited-time discounts to improve traction and category performance.

### 5. Optimize Payment Method Strategy

Credit card usage dominates customer payments. Partner with banks to offer cashback programs, EMI options, or reward-driven promotions to increase average order value and transaction frequency.

### 6. Leverage Customer Segmentation for Personalization

Utilize demographic insights (gender, age group, category preference) to deliver personalized product suggestions, email campaigns, and retargeting ads, improving conversion and retention rates.

### 7. Increase Average Order Value (AOV)

Implement product bundling, cross-sell recommendations, and loyalty incentives to encourage higher per-transaction spending across customer groups.

### 8. Strengthen Engagement Among Younger Customers (18–25)

This segment demonstrates lower revenue and quantity contribution. Drive engagement through trend-focused products, student discounts, influencer collaborations, and social-media-led campaigns.

