Customer Segmentation Analysis using MySQL

1.Project Overview

This project aims to analyse customer purchase behaviour and sales performance using transactional retail data. The analysis focuses on understanding how age groups, locations, seasons, discounts, and product categories influence overall revenue and customer value.

Using MySQL, structured queries were developed to calculate key performance indicators such as total sales, customer count, average spend, high-value customers, and underperforming product categories. SQL techniques like aggregations, subqueries, grouping, and window functions were applied to derive meaningful business insights.

2.Dataset Summary

The dataset contains transaction-level customer purchase data collected from a retail environment. Each record represents a single purchase made by a customer and includes demographic, transactional, product, and sales-related attributes.

- **Customer Information: Customer ID, age, age group, gender**
- **Transaction Details: Purchase date, quantity, payment method**
- **Product Information: Product category, unit price**
- **Sales Metrics: Discount amount, total purchase amount**
- **Time & Seasonality: Month, season (Spring, Summer, Fall, Winter)**
- **Geography: Customer location (city/state/region)**

3.Project Objectives

The project is designed to answer the following business questions

1. **Retrieve all customer purchase records from the dataset.**
2. **Find the total number of customers**
3. **Calculate the total sales amount**
4. **Find average order value**
5. **Count total transactions using payment method**
6. **Total sales per project category**
7. **Top 5 locations by total sales**
8. **Which age group generate highest revenue**
9. **Average spending by age group**

4.Data Analysis using MySQL

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

##-- Retrieve all customer purchase records from the dataset?

```
1
2    -- Retrieve all customer purchase records from the dataset.
3 •  SELECT * FROM `customer segmentation`.shopping_trends_dataset;
4
5
```

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

CustomerID	Age	Gender	ItemPurchased	Category	PurchaseAmountUSD	Location	Size	Color	Season	ReviewRating	SubscriptionStatus	ShippingType	DiscountApplied	PromoCodeUsed	Previou
1	56	Male	Bag	Accessories	374.86	New York	L	Red	Fall	3	No	Standard	Yes	No	7
2	46	Male	Shoes	Home Decor	33.28	Los Angeles	S	Black	Spring	2	No	Express	Yes	No	3
3	32	Female	Laptop	Home Decor	117.45	Houston	S	Black	Fall	4	No	Standard	Yes	Yes	13
4	60	Male	Bag	Home Decor	356.05	Chicago	M	Green	Summer	2	Yes	Express	No	No	14
5	25	Female	Laptop	Accessories	393.50	Houston	M	Black	Winter	2	Yes	Express	No	No	5
6	38	Female	Phone	Accessories	233.11	Los Angeles	M	Green	Winter	5	No	Standard	No	Yes	4
7	56	Male	Shoes	Accessories	352.06	Chicago	S	Green	Winter	2	Yes	Standard	Yes	No	12
8	36	Female	T-Shirt	Clothing	200.10	Chicago	XL	Green	Fall	2	Yes	Express	Yes	No	9
9	40	Female	Watch	Electronics	88.36	Chicago	M	Red	Fall	2	Yes	Standard	Yes	Yes	5
10	28	Male	Football	Home Decor	205.09	Chicago	XL	Black	Summer	1	No	Standard	Yes	No	3
11	28	Female	Football	Clothing	294.18	Houston	M	Black	Spring	3	Yes	Express	Yes	Yes	3
12	41	Female	Watch	Home Decor	474.55	Chicago	S	Green	Spring	3	Yes	Express	No	Yes	6
13	53	Female	Phone	Clothing	43.69	Houston	XL	Black	Winter	2	Yes	Express	Yes	Yes	0
14	57	Female	Shoes	Accessories	258.62	Chicago	XL	Black	Winter	5	Yes	Standard	No	Yes	8
15	41	Male	Shoes	Home Decor	171.17	Chicago	S	Green	Fall	5	Yes	Express	Yes	Yes	9
16	20	Female	Laptop	Accessories	259.07	Los Angeles	XL	White	Winter	5	Yes	Express	No	Yes	2
17	39	Female	T-Shirt	Home Decor	21.75	Houston	S	White	Summer	3	Yes	Standard	Yes	Yes	3
18	19	Female	Curtains	Sports	163.38	Houston	S	Green	Summer	4	No	Standard	Yes	Yes	14
19	41	Female	Bag	Clothing	411.74	Miami	L	Red	Winter	2	Yes	Standard	Yes	Yes	8
20	61	Male	Shoes	Sports	281.43	Los Angeles	M	Black	Winter	3	Yes	Express	Yes	Yes	4
21	47	Female	Watch	Clothing	230.79	Houston	L	Black	Winter	5	Yes	Standard	No	Yes	11
22	55	Male	Curtains	Electronics	149.91	Los Angeles	S	White	Winter	3	No	Express	Yes	Yes	3
23	19	Female	Football	Home Decor	156.04	Los Angeles	S	Green	Winter	3	Yes	Express	No	Yes	11
24	38	Male	Curtains	Electronics	197.72	Los Angeles	XL	Red	Summer	3	No	Express	No	No	5
25	50	Male	T-Shirt	Electronics	419.43	New York	S	Green	Spring	5	Yes	Express	No	Yes	0
26	29	Male	Watch	Clothing	483.93	Houston	S	White	Summer	4	No	Standard	Yes	Yes	9
27	39	Male	T-Shirt	Sports	224.94	Los Angeles	S	White	Summer	1	No	Standard	Yes	Yes	0
28	61	Male	Curtains	Accessories	202.74	New York	M	Black	Winter	2	No	Standard	No	No	4

1. Find the total number of customers?

```
21
22      -- 1. Find the total number of customers?
23 •    SELECT COUNT(DISTINCT `CustomerID`) AS total_customers
24      FROM shopping_trends_dataset;
25
```

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

total_customers
3900

2. Calculate the total sales amount?

```
26      -- 2. Caluclate the total sales amount?
27      SELECT
28          SUM(`PurchaseAmountUSD`) AS total_sales_amount
29      FROM shopping_trends_dataset;
```

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

total_sales_amount
996378.80

3. Find the average order value?

```
32      -- 3. Find the average order value?
33      SELECT
34          ROUND(AVG(`PurchaseAmountUSD`), 2) AS average_order_value
35      FROM shopping_trends_dataset;
```

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

average_order_value
255.48

4. Count total transaction made using each payment method?

```
37      -- 4. Count total transaction made using each payment method?
38      SELECT
39       PaymentMethod,
40      COUNT(*) AS total_transactions
41      FROM shopping_trends_dataset
42      GROUP BY PaymentMethod
43      ORDER BY total_transactions DESC;
44
```

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

PaymentMethod	total_transactions
Credit Card	802
UPI	792
PayPal	791
Debit Card	772
Cash	743

5. Show total sales per product category?

```sql
-- 5. Show total slaes per product category?
SELECT
    Category AS product_category,
    ROUND(SUM(`PurchaseAmountUSD`), 2) AS total_sales
FROM shopping_trends_dataset
GROUP BY Category
ORDER BY total_sales DESC;
```

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

product_category	total_sales
Clothing	208775.03
Accessories	208063.00
Home Decor	196999.83
Electronics	193682.52
Sports	188858.42

6. Identify top 5 locations by total sales?

```sql
-- 6. Identyfy top 5 locations by total sales?
SELECT
    Location,
    ROUND(SUM(`PurchaseAmountUSD`), 2) AS total_sales
FROM shopping_trends_dataset
GROUP BY Location
ORDER BY total_sales DESC
LIMIT 5;
```

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

Location	total_sales
Chicago	205405.71
New York	205032.64
Los Angeles	202583.83
Miami	193080.46
Houston	190276.16

7. Which age group generates the highest sales?

```sql
-- 7. Which age_group generates the highest sales?
SELECT
    age_group,
    ROUND(SUM(`PurchaseAmountUSD`), 2) AS total_sales
FROM (
    SELECT
        CASE
            WHEN Age < 25 THEN 'Under 25'
            WHEN Age BETWEEN 25 AND 34 THEN '25-34'
            WHEN Age BETWEEN 35 AND 44 THEN '35-44'
            WHEN Age BETWEEN 45 AND 54 THEN '45-54'
            ELSE '55+'
        END AS age_group,
        `PurchaseAmountUSD`
    FROM shopping_trends_dataset
) t
GROUP BY age_group
ORDER BY total_sales DESC
LIMIT 1;
```

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

age_group	total_sales
45-54	223678.23

8. Calculate the average spending by age group?

```sql
-- 8. Calculate the average spending by age group?
SELECT
    age_group,
    ROUND(AVG(`PurchaseAmountUSD`), 2) AS avg_spend
FROM (
    SELECT
        CASE
            WHEN Age < 25 THEN 'Under 25'
            WHEN Age BETWEEN 25 AND 34 THEN '25-34'
            WHEN Age BETWEEN 35 AND 44 THEN '35-44'
            WHEN Age BETWEEN 45 AND 54 THEN '45-54'
            ELSE '55+'
        END AS age_group,
        `PurchaseAmountUSD`
    FROM shopping_trends_dataset
) t
GROUP BY age_group
ORDER BY avg_spend DESC
LIMIT 1;
```

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

age_group	avg_spend
35-44	257.53

5.Business Summary

1. Analysed customer purchase data to understand sales performance and buying behavior.

2. Identified high-value customers and underperforming product categories using SQL analysis.

3. Evaluated the impact of discounts, seasons, and locations on overall revenue.

4. Built an interactive Power BI dashboard to visualize key business metrics and trends.

5. Enabled data-driven decisions for targeted marketing, pricing optimization, and inventory planning.

6.Conclusion

This project demonstrates strong SQL skills including joins, aggregations, filtering, sorting, and real-world data analysis. The analysis provides actionable insights into customer Buying behaviour.