

# Customer Retention and Sales Optimization in Retail

## Part 1 - Data Preparation, Excel Analytics, SQL Analytics & Excel Dashboarding

### 1. Data Preparation

The screenshot shows the Microsoft Excel interface with the 'Power Query' ribbon tab selected. A table named 'transactions' is open, displaying a list of transactions with columns: transaction\_id, customer\_id, product\_id, quantity, discount\_rate, and tots. The 'Data Analysis' ribbon tab is active. The 'Queries & Connections' pane on the right lists several loaded datasets: products (100 rows), campaigns (10 rows), customers (300 rows), and transactions (1,500 rows). Other queries listed include 'top 10 customers by revenue' (10 rows) and '1 vs repeat customer'.

The screenshot shows the MySQL Workbench interface. The 'Schemas' tree on the left shows a database named 'retail\_capstone' has been created. The 'SQL Editor' tab contains the following SQL code:

```
1 • CREATE DATABASE retail_capstone;
2 • USE retail_capstone;
3
4 • ○ CREATE TABLE customer_data (
5     customer_id VARCHAR(100) PRIMARY KEY,
6     name VARCHAR(100),
7     gender VARCHAR(10),
8     age INT,
9     region VARCHAR(10),
10    sign_up_date DATE,
11    loyalty_score INT
12 );
13
14 • ○ CREATE TABLE marketing_campaigns (
15     campaign_id VARCHAR(100) PRIMARY KEY,
16     channel VARCHAR(20),
17     start_date DATE,
18     end_date DATE,
```

The 'Output' tab shows the command 'CREATE DATABASE retail\_capstone;' was run successfully at 18:54:07. An error message indicates that the database already exists.

The data has been cleaned using power query in the excel and then imported into SQL. A database is created and containing four tables of main dataset.

## 2. Excel Analytics

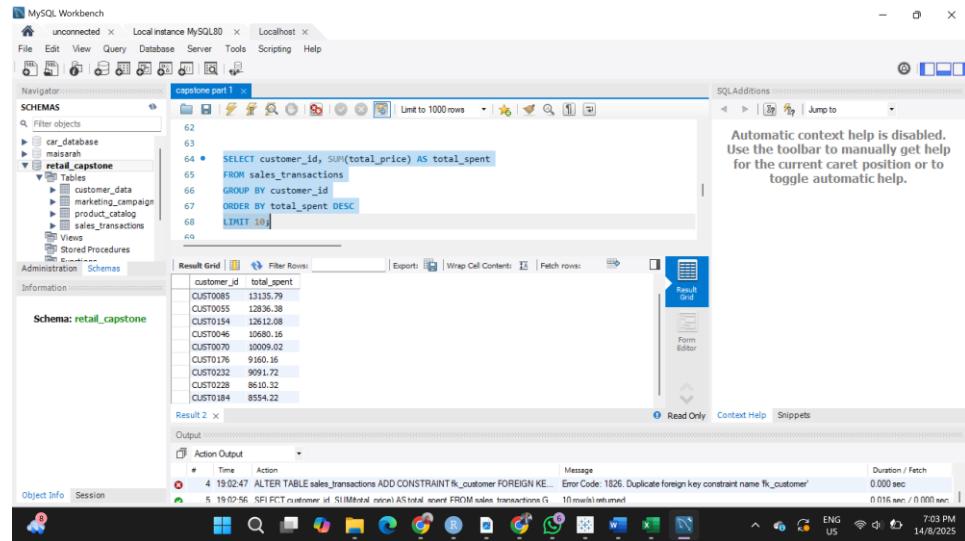
transactions		descriptive_stats_discount_rate		descriptive_stats_total_price	
descriptive_stats_quantity					
Mean	3.092667	Mean	0.0765	Mean	773.0681
Standard Error	0.037275	Standard Error	0.001442	Standard Error	15.62401
Median	3	Median	0.1	Median	638.26
Mode	5	Mode	0.15	Mode	1266.2
Standard Deviation	1.443654	Standard Deviation	0.055841	Standard Deviation	605.1155
Sample Variance	2.084136	Sample Variance	0.003118	Sample Variance	366164.7
Kurtosis	-1.34264	Kurtosis	-1.35598	Kurtosis	-0.31858
Skewness	-0.09296	Skewness	-0.03504	Skewness	0.771735
Range	4	Range	0.15	Range	2458.14
Minimum	1	Minimum	0	Minimum	8.81
Maximum	5	Maximum	0.15	Maximum	2466.95
Sum	4639	Sum	114.75	Sum	1159602
Count	1500	Count	1500	Count	1500

customers		products	
descriptive_stats_age		descriptive_stats_price	
Mean	43.32	Mean	265.8916
Standard Error	0.884887	Standard Error	15.52898
Median	43.5	Median	271.615
Mode	56	Mode	#N/A
Standard Deviation	15.32669	Standard Deviation	155.2898
Sample Variance	234.9073	Sample Variance	24114.92
Kurtosis	-1.17959	Kurtosis	-1.28082
Skewness	-0.03838	Skewness	-0.14831
Range	51	Range	486.16
Minimum	18	Minimum	10.36
Maximum	69	Maximum	496.52
Sum	12996	Sum	26589.16
Count	300	Count	100

## 3. SQL Analytics

### a) Top 10 customers by revenue



The screenshot shows the MySQL Workbench interface with the following details:

- Query Editor:** Contains the following SQL code:

```

SELECT customer_id, SUM(total_price) AS total_spent
FROM sales_transactions
GROUP BY customer_id
ORDER BY total_spent DESC
LIMIT 10;
    
```
- Result Grid:** Displays the results of the query, showing customer IDs and their total spent:

customer_id	total_spent
CUST0085	13135.79
CUST0055	12836.38
CUST0154	12612.08
CUST0046	10680.16
CUST0070	10009.02
CUST0176	9360.16
CUST0232	9091.72
CUST0228	8610.32
CUST0184	8594.22
- Output Tab:** Shows the following log entries:
  - Action: ALTER TABLE sales\_transactions ADD CONSTRAINT fk\_customer FOREIGN KEY (customer\_id) REFERENCES customer (customer\_id); Message: Error Code: 1826. Duplicate foreign key constraint name fk\_customer; Duration / Fetch: 0.000 sec / 0.000 sec
  - Action: SELECT customer\_id, SUM(total\_price) AS total\_spent FROM sales\_transactions; Message: 10 rows returned; Duration / Fetch: 0.016 sec / 0.000 sec

### b) Repeat vs one-time customers

MySQL Workbench

unconnected Local instance MySQL80 Localhost

File Edit View Query Database Server Tools Scripting Help

Navigator: SCHEMAS

- car\_database
- masharib
- retail\_capstone**
  - Tables: customer\_data, marketing\_campaign, product\_catalog, sales\_transactions
  - Views
  - Stored Procedures
  - Administration Schemas
  - Information

Schema: retail\_capstone

capstone part 1

```

72 COUNT(DISTINCT CASE WHEN purchase_count = 1 THEN customer_id END) AS one_time_customers
73
74 SELECT customer_id, COUNT(quantity) AS purchase_count
75 FROM sales_transactions
76 GROUP BY customer_id
77 ) AS sub1
78
79 ● SELECT customer_id

```

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

repeat_customers	one_time_customers
290	7

SQLAdditions

Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.

Action Output

- Time Action Message Duration / Fetch
- 5 19:02:56 SELECT customer\_id, SUM(total\_price) AS total\_spent FROM sales\_transactions GROUP BY customer\_id 10 rows(0) returned 0.016 sec / 0.000 sec
- 6 19:03:08 SPI PCTC COUNT(DISTINCT CASE WHEN purchase\_count > 1 THEN customer\_id END) AS repeat\_customers 1 rows(0) returned 0.016 sec / 0.000 sec

Object Info Session

Result 3 x

Output x

Read Only Context Help Snippets

7:04 PM ENG US 14/8/2025

### c) Monthly revenue growth

MySQL Workbench

unconnected Local instance MySQL80 Localhost

File Edit View Query Database Server Tools Scripting Help

Navigator: SCHEMAS

- car\_database
- masharib
- retail\_capstone**
  - Tables: customer\_data, marketing\_campaign, product\_catalog, sales\_transactions
  - Views
  - Stored Procedures
  - Administration Schemas
  - Information

Schema: retail\_capstone

capstone part 1

```

85 ● SELECT
86   DATE_FORMAT(transaction_date, '%Y-%m') AS month,
87   SUM(total_price * (1 - discount_rate)) AS monthly_revenue
88
89   FROM sales_transactions
90
91   WHERE transaction_date BETWEEN '2024-04-21' AND '2025-04-21'
92
93   GROUP BY DATE_FORMAT(transaction_date, '%Y-%m')
94
95   ORDER BY month

```

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

month	monthly_revenue
2024-04	38354.8000
2024-05	89485.2820
2024-06	11357.3420
2024-07	62001.7773
2024-08	67214.7215
2024-09	74934.8355
2024-10	95784.5720
2024-11	78804.7030
2024-12	93417.3335

SQLAdditions

Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.

Action Output

- Time Action Message Duration / Fetch
- 7 19:04:38 SELECT customer\_id FROM customer\_data WHERE customer\_id NOT IN (SELECT customer\_id FROM sales\_transactions WHERE transaction\_date < '2024-04-21') 3 rows(0) returned 0.015 sec / 0.000 sec
- 8 19:14:47 SPI PCTC DATE\_FORMAT(transaction\_date, '%Y-%m') AS month, SUM(total\_price) AS monthly\_revenue 13 rows(0) returned 0.016 sec / 0.000 sec

Object Info Session

Result 5 x

Output x

Read Only Context Help Snippets

7:05 PM ENG US 14/8/2025

### d) Lifetime value

MySQL Workbench

unconnected Local instance MySQL80 Localhost

File Edit View Query Database Server Tools Scripting Help

Navigator: SCHEMAS

- car\_database
- masharib
- retail\_capstone**
  - Tables: customer\_data, marketing\_campaign, product\_catalog, sales\_transactions
  - Views
  - Stored Procedures
  - Administration Schemas
  - Information

Schema: retail\_capstone

capstone part 1

```

91 ORDER BY month
92
93 ● SELECT
94   customer_id,
95   SUM(total_price * (1 - discount_rate)) AS lifetime_value
96
97   FROM sales_transactions
98
99   GROUP BY customer_id

```

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

customer_id	lifetime_value
CUST00001	6014.7850
CUST00002	3050.3990
CUST0003	3050.3960
CUST0004	3672.4655
CUST0005	1055.2790
CUST0006	4553.0095
CUST0007	5359.0115
CUST0008	1371.4525
CUST0009	1514.7570

SQLAdditions

Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.

Action Output

- Time Action Message Duration / Fetch
- 8 19:04:47 SELECT DATE\_FORMAT(transaction\_date, '%Y-%m') AS month, SUM(total\_price) AS monthly\_revenue 13 rows(0) returned 0.000 sec / 0.000 sec
- 9 19:05:24 SPI PCTC customer\_id, SUM(monthly\_revenue \* (1 - discount\_rate)) AS lifetime\_value FROM sales\_transactions WHERE transaction\_date < '2024-04-21' 297 rows(0) returned 0.016 sec / 0.000 sec

Object Info Session

Result 6 x

Output x

Read Only Context Help Snippets

7:05 PM ENG US 14/8/2025

## 4. Excel Dashboarding

### a) Pivot analysis

The screenshot shows a Microsoft Excel spreadsheet titled "Capstone Part 1 Maisarah assign...". The data is organized into two main pivot tables:

- Pivot Table 1 (Left):** Sum of net\_revenue by Region (East, North, South, West). The Grand Total is 1072811.593.
- Pivot Table 2 (Right):** Sum of net\_revenue by Month (2024, 2025) and Category (Books, Clothing, Electronics, Groceries, Home & Kitchen, Sports).

Below the pivot tables are two charts:

- A bar chart titled "Total" showing sales volume for each region.
- A pie chart titled "Total" showing the distribution of sales across categories.

The ribbon at the top includes tabs for File, Home, Insert, Page Layout, Formulas, Data, Review, View, Power Pivot, PivotTable Analyze, Design, and Data Analysis. A status bar at the bottom indicates the date as 14/8/2025 and the time as 7:08 PM.

### b) Dashboard

