

## Task 4 – SQL for Data Analysis

**Name:** Neha Jaiswal

**Project:** eCommerce Dataset Analysis using SQL

**Tool:** MySQL Workbench

Query 1:

```
1  CREATE DATABASE ecommerce;
2  USE ecommerce;
3
4  CREATE TABLE orders (
5      order_id VARCHAR(20),
6      customer_id VARCHAR(20),
7      product_id VARCHAR(20),
8      product_name VARCHAR(255),
9      category VARCHAR(100),
10     quantity INT,
11     price DECIMAL(10,2),
12     order_date DATE,
13     region VARCHAR(100),
14     payment_method VARCHAR(50),
15     total_amount DECIMAL(10,2)
16 );
17
18 -- Creating table products
19 CREATE TABLE products AS
20 SELECT DISTINCT product_id, product_name, category, price
21 FROM orders;
22
23 -- creating table customers
24 CREATE TABLE customers AS
25 SELECT DISTINCT customer_id
26 FROM orders;
27
28 show tables;
```

Output:

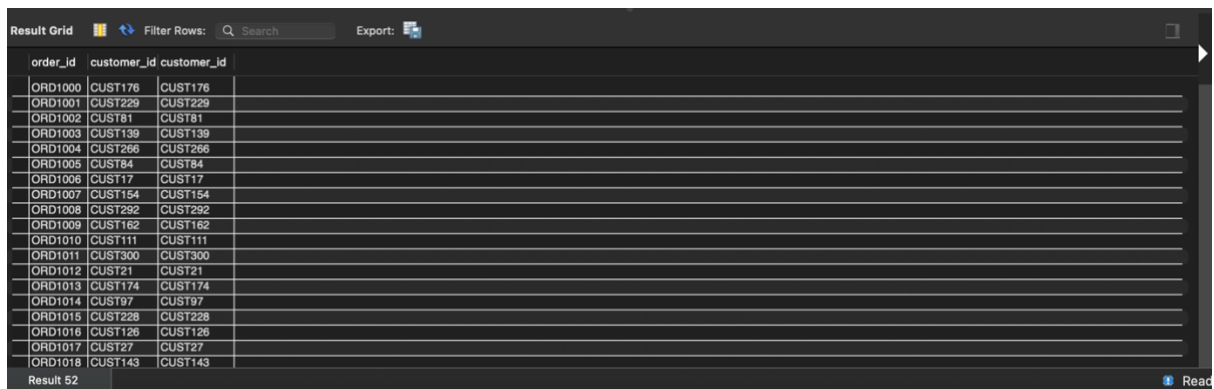


Tables_in_ecommerce
category_summary
customers
orders
products

Query 2:

```
29
30 -- INNER JOIN: Orders with customer details
31 SELECT o.order_id, o.customer_id, c.customer_id
32 FROM orders o
33 INNER JOIN customers c ON o.customer_id = c.customer_id;
34
```

Output:



order_id	customer_id	customer_id
ORD1000	CUST176	CUST176
ORD1001	CUST229	CUST229
ORD1002	CUST81	CUST81
ORD1003	CUST139	CUST139
ORD1004	CUST266	CUST266
ORD1005	CUST84	CUST84
ORD1006	CUST117	CUST117
ORD1007	CUST154	CUST154
ORD1008	CUST292	CUST292
ORD1009	CUST162	CUST162
ORD1010	CUST111	CUST111
ORD1011	CUST300	CUST300
ORD1012	CUST21	CUST21
ORD1013	CUST174	CUST174
ORD1014	CUST97	CUST97
ORD1015	CUST228	CUST228
ORD1016	CUST126	CUST126
ORD1017	CUST27	CUST27
ORD1018	CUST143	CUST143

### Query 3:

```
35 -- INNER JOIN orders with products (with sample output)
36 SELECT o.order_id, o.product_id, p.product_name, p.category, p.price
37 FROM orders o
38 INNER JOIN products p ON o.product_id = p.product_id
39 LIMIT 10;
```

### Output:

order_id	product_id	product_name	category	price
ORD1928	PROD94	Build	Home	313.75
ORD1858	PROD94	Build	Home	313.75
ORD1850	PROD94	Build	Home	313.75
ORD1813	PROD94	Build	Home	313.75
ORD1730	PROD94	Build	Home	313.75
ORD1676	PROD94	Build	Home	313.75
ORD1200	PROD94	Build	Home	313.75
ORD1094	PROD94	Build	Home	313.75
ORD1000	PROD94	Build	Home	313.75
ORD1976	PROD78	According	Clothing	99.55

### Query 4:

```
41
42 -- LEFT JOIN: Orders with product info
43 SELECT o.order_id, o.product_id, p.product_name, p.price
44 FROM orders o
45 LEFT JOIN products p ON o.product_id = p.product_id;
```

### Output:

order_id	product_id	product_name	price
ORD1000	PROD94	Within	212.97
ORD1000	PROD94	Table	118.88
ORD1000	PROD94	Field	339.67
ORD1000	PROD94	Increase	436.41
ORD1000	PROD94	They	327.16
ORD1000	PROD94	Field	105.84
ORD1000	PROD94	Build	313.75
ORD1001	PROD78	Congress	397.75
ORD1001	PROD78	Along	192.59
ORD1001	PROD78	Effect	127.29
ORD1001	PROD78	Either	178.41
ORD1001	PROD78	Hand	96.71
ORD1001	PROD78	Power	103.90
ORD1001	PROD78	A	297.23
ORD1001	PROD78	Indeed	160.50
ORD1001	PROD78	Nearly	404.19
ORD1001	PROD78	He	499.00
ORD1001	PROD78	Charge	195.24
ORD1001	PROD78	According	99.55

### Query 5:

```
46
47 -- RIGHT JOIN (if needed, note: MySQL may emulate it)
48 SELECT o.order_id, o.product_id, p.product_name
49 FROM orders o
50 RIGHT JOIN products p ON o.product_id = p.product_id;
```

### Output:

order_id	product_id	product_name
ORD1928	PROD94	Build
ORD1858	PROD94	Build
ORD1850	PROD94	Build
ORD1813	PROD94	Build
ORD1730	PROD94	Build
ORD1676	PROD94	Build
ORD1200	PROD94	Build
ORD1094	PROD94	Build
ORD1000	PROD94	Build
ORD1976	PROD78	According
ORD1844	PROD78	According
ORD1840	PROD78	According
ORD1668	PROD78	According
ORD1666	PROD78	According

Query 6:

```
52 -- View all data from orders table
53 SELECT * FROM orders;
54
```

Output:

order_id	customer_id	product_name	product_name	category	quantity	price	order_date	region	payment_method	total_amount
ORD1002	CUST81	PROD14	Rich	Electronics	2	375.53	2023-09-30	South	Credit Card	751.06
ORD1003	CUST139	PROD59	Almost	Books	3	438.23	2025-01-15	North	Cash	1314.69
ORD1004	CUST266	PROD15	Mention	Clothing	2	118.43	2024-09-13	North	UPI	236.86
ORD1005	CUST84	PROD2	Method	Electronics	1	118.15	2023-11-30	East	Credit Card	118.15
ORD1006	CUST117	PROD63	Gave	Electronics	4	204.71	2025-05-18	North	PayPal	818.84
ORD1007	CUST154	PROD79	Allow	Toys	3	224.19	2025-03-21	South	Credit Card	672.57
ORD1008	CUST292	PROD36	Pay	Toys	4	164.02	2024-06-30	North	Credit Card	656.08
ORD1009	CUST162	PROD7	Action	Books	3	206.23	2024-06-02	North	Credit Card	615.69
ORD1010	CUST111	PROD63	Actually	Toys	3	253.35	2023-08-24	East	Credit Card	760.05
ORD1011	CUST300	PROD27	Lose	Toys	1	353.48	2024-03-06	South	UPI	353.48
ORD1012	CUST21	PROD41	Decision	Home	2	441.25	2024-10-22	North	Credit Card	882.50
ORD1013	CUST174	PROD84	Leg	Toys	1	148.04	2025-04-13	East	PayPal	148.04
ORD1014	CUST97	PROD16	Project	Electronics	3	452.16	2024-09-27	West	Cash	1356.48
ORD1015	CUST228	PROD13	Former	Toys	2	342.77	2024-05-30	West	Cash	685.54
ORD1016	CUST126	PROD80	Job	Toys	5	377.79	2023-12-18	West	Cash	1888.95
ORD1017	CUST27	PROD37	Player	Clothing	2	286.58	2023-10-19	South	PayPal	573.16
ORD1018	CUST143	PROD67	Election	Clothing	1	244.76	2024-07-06	North	UPI	244.76

Query 7:

```
55 -- Get top 5 orders by highest total_amount
56 SELECT order_id, customer_id, total_amount
57 FROM orders
58 ORDER BY total_amount DESC
59 LIMIT 5;
60
```

Output:

order_id	customer_id	total_amount
ORD1253	CUST278	2498.45
ORD1118	CUST170	2494.20
ORD1626	CUST265	2492.35
ORD1753	CUST71	2477.70
ORD1484	CUST133	2472.70

Query 8:

```
61 -- Total quantity sold per product category
62 SELECT category, SUM(quantity) AS total_quantity
63 FROM orders
64 GROUP BY category;
65
```

Output:

category	total_quantity
Home	506
Clothing	634
Electronics	600
Books	651
Toys	576

Query 9:

```
66 -- Average price per region
67 SELECT region, AVG(price) AS avg_price
68 FROM orders
69 GROUP BY region;
70
```

Output:

region	avg_price
West	238.885850
South	251.530163
North	245.198112
East	261.663657

## Query 10:

```
71 -- Count the number of orders placed in each region
72 • SELECT region, COUNT(*) AS number_of_orders
73 FROM orders
74 GROUP BY region;
75
```

## Output:

region	number_of_ord...
West	253
South	246
North	233
East	268

## Query 11:

```
76 -- Subquery: Show orders where total_amount is greater than the average
77 • SELECT order_id, total_amount
78 FROM orders
79 WHERE total_amount > (
80     SELECT AVG(total_amount) FROM orders
81 );
```

## Output:

order_id	total_amou...
ORD1000	941.25
ORD1002	751.06
ORD1003	1314.69
ORD1006	818.84
ORD1010	760.05
ORD1012	882.50
ORD1014	1356.48
ORD1016	1888.95
ORD1021	1422.40
ORD1022	1816.65
ORD1024	1981.84
ORD1028	1231.98
ORD1027	1132.89

## Query 12:

```
83 -- View orders only from 'South' region
84 • SELECT * FROM orders
85 WHERE region = 'South';
86
```

## Output:

order_id	customer_id	product...	product_na...	category	quantity	price	order_date	region	payment_meth...	total_amou...
ORD1001	CUST229	PROD78	According	Clothing	5	99.55	2023-12-15	South	UPI	497.75
ORD1002	CUST81	PROD14	Rich	Electronics	2	375.53	2023-09-30	South	Credit Card	751.06
ORD1007	CUST154	PROD79	Allow	Toys	3	224.19	2025-03-21	South	Credit Card	672.57
ORD1011	CUST300	PROD27	Lose	Toys	1	353.48	2024-03-06	South	UPI	353.48
ORD1017	CUST27	PROD37	Player	Clothing	2	286.58	2023-10-19	South	PayPal	573.16
ORD1023	CUST253	PROD53	Field	Clothing	5	123.45	2024-09-08	South	Cash	617.25
ORD1030	CUST157	PROD53	Certainly	Electronics	1	76.02	2024-10-11	South	Cash	76.02
ORD1037	CUST32	PROD6	Wall	Toys	3	138.30	2024-06-02	South	Credit Card	408.90
ORD1039	CUST32	PROD58	Thank	Clothing	5	341.72	2024-02-16	South	PayPal	1708.60
ORD1042	CUST158	PROD25	No	Electronics	3	47.00	2024-12-08	South	PayPal	141.00
ORD1044	CUST178	PROD85	Federal	Electronics	4	92.61	2024-10-19	South	Credit Card	370.44
ORD1048	CUST243	PROD29	Young	Books	2	413.43	2023-12-22	South	Cash	826.86
ORD1054	CUST42	PROD38	Finish	Electronics	2	162.41	2023-08-17	South	Cash	304.82
ORD1055	CUST150	PROD17	Language	Electronics	1	159.00	2024-01-09	South	UPI	159.00
ORD1056	CUST61	PROD78	Charge	Home	1	195.24	2024-04-22	South	UPI	195.24

## Query 13:

```
92
93 -- Select from the view
94 • SELECT * FROM category_summary;
95
```



## Output:

category	total_sales
Home	127295.07
Clothing	159644.73
Electronics	153920.37
Books	156607.35
Toys	141638.97

## Query 14:

```
96 -- Create an index on order_date column to improve performance
97 ■ SHOW COLUMNS FROM orders;
98 ■ SHOW INDEX FROM orders;
99
```

## Output:

Result Grid  Filter Rows: <input type="text"/> Search <input type="text"/> Export: 														
Table	Non_unique	Key_name	Seq_in_index	Column_name	Collation	Cardinality	Sub_part	Packed	Null	Index_type	Comment	Index_comment	Visible	Expression
orders	1	idx_order_date	1	order_date	A	548			YES	BTREE			YES	