Task 4: SQL for Data Analysis

Amazon sales data 2025

```
1    -- Q1: View orders
2 ●  SELECT * FROM amazon;
```

Order ID	Date	Product	Category	Price	Quantity	Total_Sales	Customer_Name	Customer_Location	Payment_Method	Status
ORD0001	14-03-2025	Running Shoes	Footwear	60	3	180	Emma Clark	New York	Debit Card	Cancelle
ORD0002	20-03-2025	Headphones	Electronics	100	4	400	Emily Johnson	San Francisco	Debit Card	Pending
ORD0003	15-02-2025	Running Shoes	Footwear	60	2	120	John Doe	Denver	Amazon Pay	Cancelle
ORD0004	19-02-2025	Running Shoes	Footwear	60	3	180	Olivia Wilson	Dallas	Credit Card	Pending
ORD0005	10-03-2025	Smartwatch	Electronics	150	3	450	Emma Clark	New York	Debit Card	Pending
ORD0006	14-03-2025	T-Shirt	Clothing	20	1	20	John Doe	Dallas	Credit Card	Pending
ORD0007	18-03-2025	Smartwatch	Electronics	150	4	600	Emma Clark	Houston	PayPal	Comple
ORD0008	02-03-2025	Smartphone	Electronics	500	1	500	Sophia Miller	Miami	PayPal	Comple
ORD0009	08-03-2025	T-Shirt	Clothing	20	3	60	Sophia Miller	Boston	PayPal	Comple

```
1    -- Q1: View orders
2 ●  SELECT * FROM amazon;
3
4    -- Q2: Count total orders
5 ●  SELECT COUNT(*) AS total_orders FROM amazon;
```

total_orders
250

```
7      -- Q3: Total revenue (sum of total_sale)
8  •   SELECT ROUND(SUM(total_sales),2) AS total_revenue FROM amazon;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: $\overline{I}A$

total_revenue
243845

amazon 10   Result 11   Result 12 ×

```
10     -- Q4: Total revenue for completed orders
11 •   SELECT ROUND(SUM(total_sales),2) AS revenue_completed
12     FROM amazon
13     WHERE status = 'Completed';
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: $\overline{I}A$

revenue_completed
88530

```
15     -- Q5: Total revenue by status
16 •   SELECT status, ROUND(SUM(total_sales),2) AS revenue
17     FROM amazon
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: $\overline{I}A$

status	revenue
Pending	90285
Completed	88530
Cancelled	65030

```sql
21    -- Q6: Total revenue by category
22  ● SELECT category, ROUND(SUM(total_sales),2) AS revenue
23    FROM amazon
24    GROUP BY category
25    ORDER BY revenue DESC;
```

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

category	revenue
Electronics	129950
Home Appliances	105000
Footwear	4320
Clothing	3540
Books	1035

```sql
27    -- Q7: Total quantity sold per product
28  ● SELECT product, SUM(quantity) AS total_quantity
29    FROM amazon
30    GROUP BY product
31    ORDER BY total_quantity DESC;
```

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

product	total_quantity
Smartwatch	105
Smartphone	97
Headphones	73
Laptop	73
Running Shoes	72
Book	69
Refrigerator	65
Jeans	62
T-Shirt	53
Washing Machine	45

```sql
32
33    -- Q8: Average order value (AOV)
34  ● SELECT ROUND(AVG(total_sales),2) AS avg_order_value FROM amazon;
```

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

avg_order_value
975.38

```sql
36    -- Q9: Highest single order by value
37  • SELECT order_id, date, customer_name, total_sales
38    FROM amazon
39    ORDER BY total_sales DESC
40    LIMIT 1;
```

order_id	date	customer_name	total_sales
ORD0126	04-02-2025	Olivia Wilson	6000

```sql
42    -- Q10: Lowest non-zero order
43  • SELECT order_id, date, total_sales
44    FROM amazon
45    WHERE total_sales > 0
46    ORDER BY total_sales ASC
47    LIMIT 1;
```

order_id	date	total_sales
ORD0070	10-03-2025	15

```sql
49    -- Q11: Orders placed in the last 30 days
50  • SELECT * FROM amazon
51    WHERE date >= DATE_SUB(CURDATE(), INTERVAL 30 DAY)
52    ORDER BY date DESC;
53
54
55
```

order_id	Date	Product	Category	Price	Quantity	Total_Sales	Customer_Name	Customer_Location	Payment_Method	Status

```sql
54    -- Q12: Revenue by customer (who bought the most in money)
55  • SELECT customer_name, ROUND(SUM(total_sales),2) AS total_spent
56    FROM amazon
57    GROUP BY customer_name
58    ORDER BY total_spent DESC
59    LIMIT 10;
60
```

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

customer_name	total_spent
Olivia Wilson	36170
Jane Smith	31185
Emma Clark	29700
John Doe	26870
Emily Johnson	23475
David Lee	22665
Michael Brown	22655
Daniel Harris	18945
Chris White	18885
Sophia Miller	13295

```sql
61    -- Q13: Count orders per payment method
62  • SELECT payment_method, COUNT(*) AS orders_count
63    FROM amazon
64    GROUP BY payment_method
65    ORDER BY orders_count DESC;
```

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

payment_method	orders_count
PayPal	60
Credit Card	54
Debit Card	53
Gift Card	42
Amazon Pay	41

```sql
67    -- Q14: Top 5 cities by revenue
68  • SELECT customer_location, ROUND(SUM(total_sales),2) AS revenue
69    FROM amazon
70    GROUP BY customer_location
71    ORDER BY revenue DESC
72    LIMIT 5;
```

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

customer_location	revenue
Miami	31700
Denver	29785
Houston	28390
Dallas	27145
Seattle	26890

```sql
74     -- Q15: List of pending orders (for follow-up)
75 •   SELECT order_id, date, customer_name, product, total_sales
76     FROM amazon
77     WHERE status = 'Pending'
78     ORDER BY date;
--
```

order_id	date	customer_name	product	total_sales
ORD0013	01-03-2025	Daniel Harris	Laptop	1600
ORD0034	02-04-2025	Jane Smith	T-Shirt	100
ORD0091	03-02-2025	Daniel Harris	Laptop	3200
ORD0126	04-02-2025	Olivia Wilson	Refrigerator	6000
ORD0196	04-03-2025	Daniel Harris	Headphones	500
ORD0155	05-02-2025	Sophia Miller	Refrigerator	4800
ORD0147	05-02-2025	Jane Smith	Running Shoes	300
ORD0088	06-02-2025	Chris White	Refrigerator	2400
ORD0027	07-02-2025	Daniel Harris	T-Shirt	20
ORD0134	07-02-2025	David Lee	Smartphone	2500

```sql
80     -- Q16: Data validation - show rows where total_sale differs from price * quantity
81 •   SELECT order_id, price, quantity, total_sales, ROUND(price * quantity, 2) AS calc_total,
82         ROUND(ABS(total_sales - price * quantity),2) AS diff
83     FROM amazon
84     WHERE total_sales IS NULL OR ROUND(ABS(total_sales - price * quantity),2) > 0.01;
85
86
--
```

order_id	price	quantity	total_sales	calc_total	diff

```sql
86     -- Q17: Create a simple view of completed order summary
87 •   DROP VIEW IF EXISTS vw_completed_orders;
88 •   CREATE VIEW vw_completed_orders AS
89     SELECT order_id,date, product, category, quantity, total_sales, customer_name, customer_location, payment_metho
90     FROM amazon
91     WHERE status = 'Completed';
92
93     -- Q18: Use the view
94 •   SELECT * FROM vw_completed_orders ORDER BY date DESC LIMIT 50;
```

order_id	date	product	category	quantity	total_sales	customer_name	customer_location	payment_method
ORD0111	31-03-2025	Laptop	Electronics	4	3200	Emma Clark	Los Angeles	Credit Card
ORD0096	30-03-2025	Smartwatch	Electronics	3	450	Jane Smith	New York	Amazon Pay
ORD0063	30-03-2025	Smartphone	Electronics	5	2500	Emma Clark	Miami	Gift Card
ORD0209	29-03-2025	Jeans	Clothing	5	200	Daniel Harris	Dallas	PayPal
ORD0171	28-02-2025	Smartwatch	Electronics	2	300	Daniel Harris	Houston	PayPal
ORD0173	28-02-2025	Refrigerator	Home Appliances	1	1200	Olivia Wilson	Seattle	PayPal
ORD0152	28-02-2025	Smartwatch	Electronics	1	150	John Doe	Seattle	Gift Card
ORD0176	27-03-2025	Book	Books	1	15	Michael Brown	Boston	Amazon Pay
ORD0117	27-02-2025	T-Shirt	Clothing	1	20	Daniel Harris	New York	Credit Card
ORD0205	27-02-2025	Headphones	Electronics	2	200	David Lee	Houston	PayPal

amazon 331   amazon 332   amazon 333   Result 334   Result 335   Result 336   amazon 337   Result 338   vw_completed_orders 3 ⓘ

```
96      -- Q19: Count cancelled orders
97 •    SELECT COUNT(*) AS cancelled_count
98      FROM amazon
```

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

cancelled_count
77

```
101     -- Q20: Distinct products sold
102 •   SELECT DISTINCT product FROM amazon ORDER BY product;
103
104
```

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

product
Book
Headphones
Jeans
Laptop
Refrigerator
Running Shoes
Smartphone
Smartwatch
T-Shirt
Washing Machine

```
104     -- Q21: First and last order date in dataset
105 •   SELECT MIN(date) AS first_order, MAX(date) AS last_order FROM amazon;
106
```

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

first_order	last_order
01-03-2025	31-03-2025

```
107     -- Q22: Top 3 products by revenue
108  •  SELECT product, ROUND(SUM(total_sales),2) AS revenue
109     FROM amazon
110     GROUP BY product
111     ORDER BY revenue DESC
112     LIMIT 3;
113
```

product	revenue
Refrigerator	78000
Laptop	58400
Smartphone	48500

```
114     -- Q23: Simple EXPLAIN for performance check (shows index usage)
115  •  EXPLAIN SELECT * FROM amazon WHERE status='Completed' AND category='Electronic';
116
```

id	select_type	table	partitions	type	possible_keys	key	key_len	ref	rows	filtered	Extra
1	SIMPLE	amazon	NULL	ALL	NULL	NULL	NULL	NULL	250	1.00	Using where