New Wheels Project

Introduction to SQL

Problem Statement

Business Context

A lot of people in the world share a common desire: to own a vehicle. A car or an automobile is seen as an object that gives the freedom of mobility. Many now prefer pre-owned vehicles because they come at an affordable cost, but at the same time, they are also concerned about whether the after-sales service provided by the resale vendors is as good as the care you may get from the actual manufacturers.

New-Wheels, a vehicle resale company, has launched an app with an end-to-end service from listing the vehicle on the platform to shipping it to the customer's location. This app also captures the overall after-sales feedback given by the customer.

Objective

New-Wheels sales have been dipping steadily in the past year, and due to the critical customer feedback and ratings online, there has been a drop in new customers every quarter, which is concerning to the business. The CEO of the company now wants a quarterly report with all the key metrics sent to him so he can assess the health of the business and make the necessary decisions.

As a data analyst, you see that there is an array of questions that are being asked at the leadership level that need to be answered using data. Import the dump file that contains various tables that are present in the database. Use the data to answer the questions posed and create a quarterly business report for the CEO.

Business Questions

Question 1: a) Find the total number of customers who have placed orders. b) What is the distribution of the customers across states?

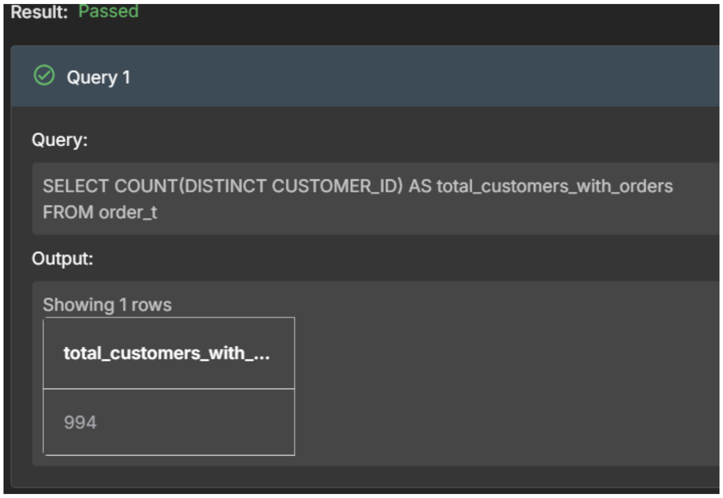
Solution Query:

a)
```
SELECT COUNT(DISTINCT CUSTOMER_ID) AS total_customers_with_orders
FROM order_t;
```

b)
```
SELECT c.STATE, COUNT(DISTINCT c.CUSTOMER_ID) AS num_customers
FROM customer_t c
JOIN order_t o
        ON c.CUSTOMER_ID = o.CUSTOMER_ID
GROUP BY c.STATE
ORDER BY num_customers DESC;
```

Output:

a)



Result: Passed

Query 1

Query:

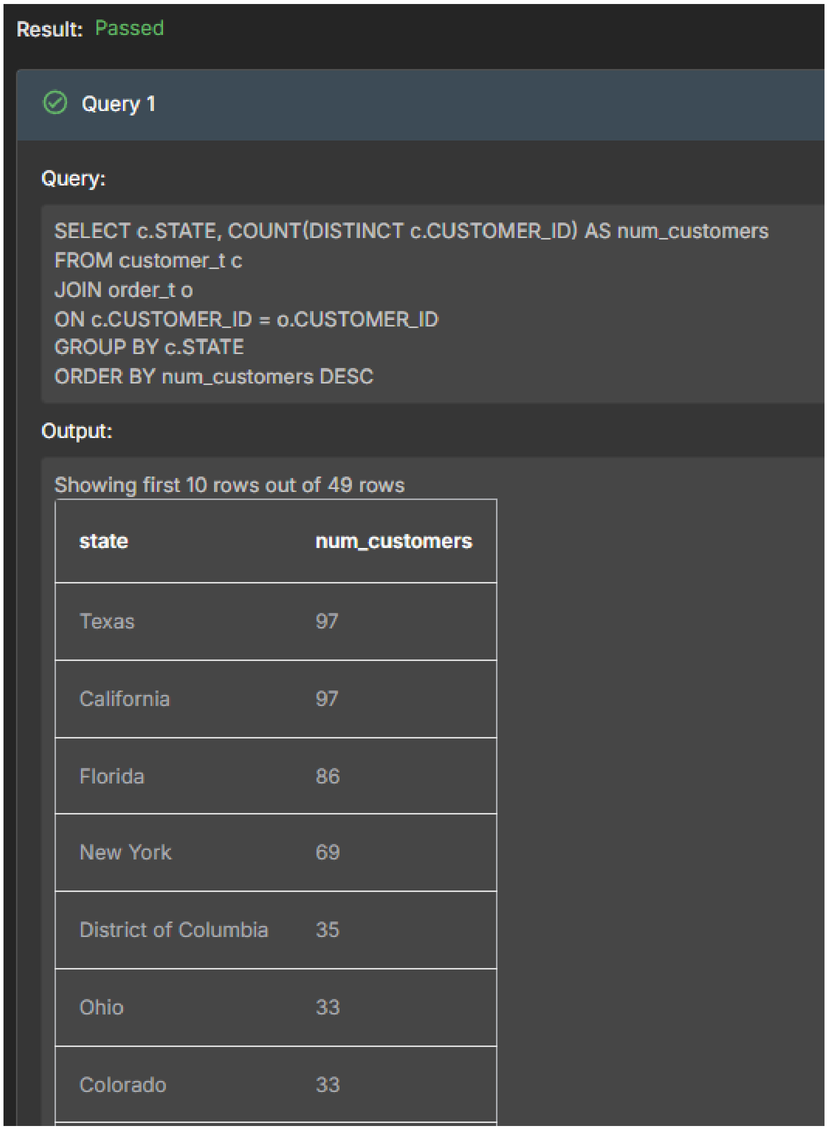
```
SELECT COUNT(DISTINCT CUSTOMER_ID) AS total_customers_with_orders
FROM order_t
```

Output:

Showing 1 rows

total_customers_with_...
994

b)



Result: Passed

Query 1

Query:

```
SELECT c.STATE, COUNT(DISTINCT c.CUSTOMER_ID) AS num_customers
FROM customer_t c
JOIN order_t o
ON c.CUSTOMER_ID = o.CUSTOMER_ID
GROUP BY c.STATE
ORDER BY num_customers DESC
```

Output:

Showing first 10 rows out of 49 rows

state	num_customers
Texas	97
California	97
Florida	86
New York	69
District of Columbia	35
Ohio	33
Colorado	33

Observations and Insights:

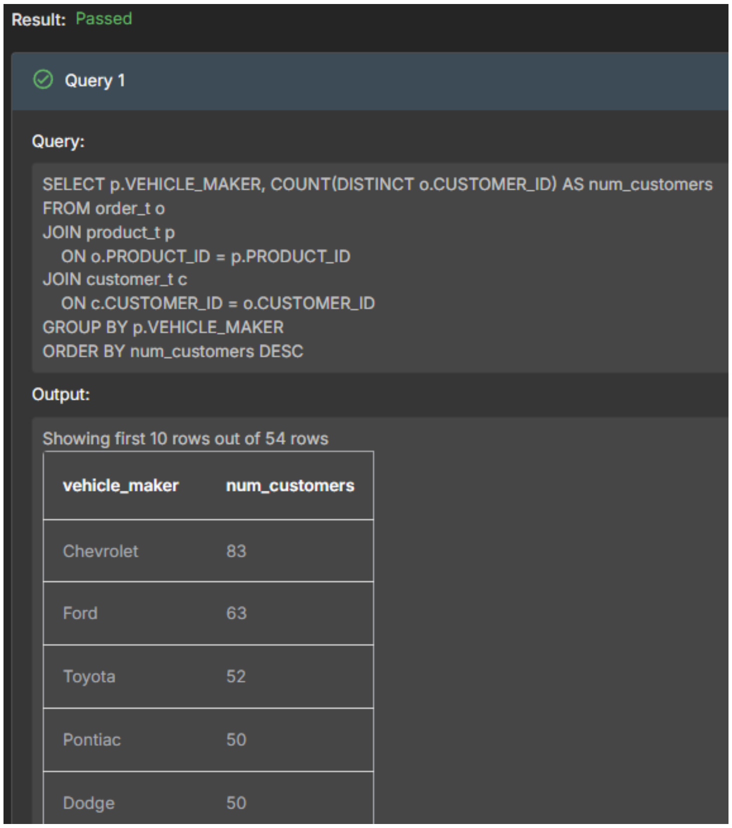
- A total of 994 customers placed orders.
- The states with highest populations have the most number or orders, e.g. Texas California, and Florida.
- The states with the lowest number of orders are Maine and Vermont, but it is not safe to assume that the region has a significant impact on the number of care orders.

Question 2: Which are the top 5 vehicle makers preferred by the customers?

Solution Query:

```
SELECT p.VEHICLE_MAKER, COUNT(DISTINCT o.CUSTOMER_ID) AS num_customers

FROM order_t o

JOIN product_t p

    ON o.PRODUCT_ID = p.PRODUCT_ID

JOIN customer_t c

    ON c.CUSTOMER_ID = o.CUSTOMER_ID

GROUP BY p.VEHICLE_MAKER

ORDER BY num_customers DESC;
```

Output:



Query:

```
SELECT p.VEHICLE_MAKER, COUNT(DISTINCT o.CUSTOMER_ID) AS num_customers
FROM order_t o
JOIN product_t p
   ON o.PRODUCT_ID = p.PRODUCT_ID
JOIN customer_t c
   ON c.CUSTOMER_ID = o.CUSTOMER_ID
GROUP BY p.VEHICLE_MAKER
ORDER BY num_customers DESC
```

Output:

Showing first 10 rows out of 54 rows

vehicle_maker	num_customers
Chevrolet	83
Ford	63
Toyota	52
Pontiac	50
Dodge	50

Observations and Insights:

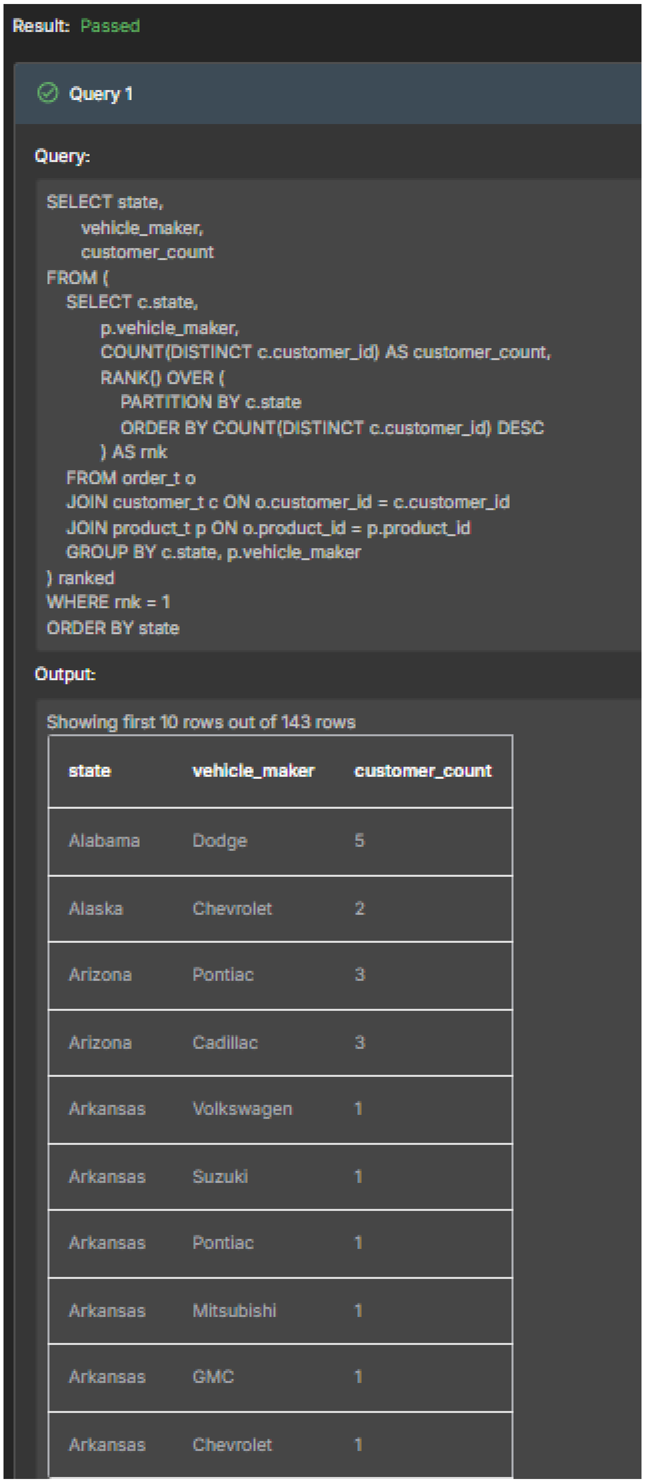
- There are 54 total car makers.
- The top 5 car makers preferred by customers are Chevrolet, Ford, Toyota, Pontiac, and Dodge.
- The 5 car makers least preferred by customers, as seen when we change DESC to ASC

Question 3: Which is the most preferred vehicle maker in each state?

Solution Query:

```
SELECT state,
    vehicle_maker,
    customer_count
FROM (
  SELECT c.state,
      p.vehicle_maker,
      COUNT(DISTINCT c.customer_id) AS customer_count,
      RANK() OVER (
        PARTITION BY c.state
        ORDER BY COUNT(DISTINCT c.customer_id) DESC
      ) AS rnk
  FROM order_t o
  JOIN customer_t c ON o.customer_id = c.customer_id
  JOIN product_t p ON o.product_id = p.product_id
  GROUP BY c.state, p.vehicle_maker
) ranked
WHERE rnk = 1
ORDER BY state;
```

Output:



Query:

```
SELECT state,
      vehicle_maker,
      customer_count
FROM (
   SELECT c.state,
         p.vehicle_maker,
         COUNT(DISTINCT c.customer_id) AS customer_count,
         RANK() OVER (
            PARTITION BY c.state
            ORDER BY COUNT(DISTINCT c.customer_id) DESC
         ) AS rnk
   FROM order_t o
   JOIN customer_t c ON o.customer_id = c.customer_id
   JOIN product_t p ON o.product_id = p.product_id
   GROUP BY c.state, p.vehicle_maker
) ranked
WHERE rnk = 1
ORDER BY state
```

Output:

Showing first 10 rows out of 143 rows

state	vehicle_maker	customer_count
Alabama	Dodge	5
Alaska	Chevrolet	2
Arizona	Pontiac	3
Arizona	Cadillac	3
Arkansas	Volkswagen	1
Arkansas	Suzuki	1
Arkansas	Pontiac	1
Arkansas	Mitsubishi	1
Arkansas	GMC	1
Arkansas	Chevrolet	1

Observations and Insights:

- Although orders came from only 49 states, the results table contains 143 rows. This indicates that many states had multiple vehicle makers tied for the top spot.

- When focusing only on selected states, California reveals a unique pattern—a five-way tie in which Nissan, Ford, Dodge, Chevrolet, and Audi each attracted six customers. This even distribution highlights California's diverse customer preferences.

- Across the four most populous states (California, Florida, New York, and Texas), there is no consistent pattern in the leading vehicle brand. The only similarity is that Florida and New York both rank Toyota as the most purchased brand.

Question 4: a)Find the overall average rating given by the customers.

b)What is the average rating in each quarter?

Consider the following mapping for ratings: “Very Bad”: 1, “Bad”: 2, “Okay”: 3, “Good”: 4, “Very Good”: 5

Solution Query:

a)

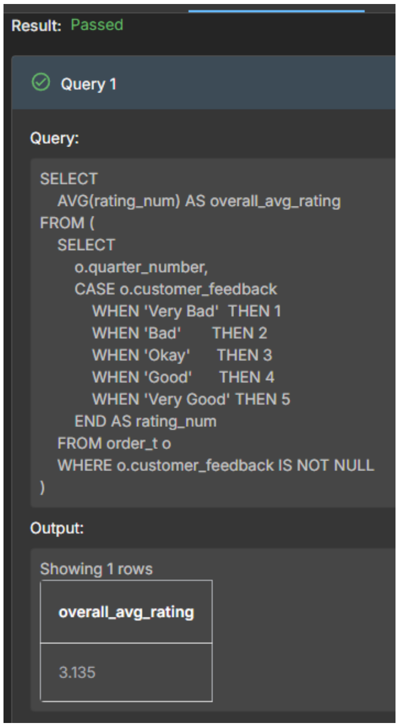
```
SELECT
    AVG(rating_num) AS overall_avg_rating
FROM (
    SELECT
        o.quarter_number,
        CASE o.customer_feedback
            WHEN 'Very Bad'  THEN 1
            WHEN 'Bad'       THEN 2
            WHEN 'Okay'      THEN 3
            WHEN 'Good'      THEN 4
            WHEN 'Very Good' THEN 5
        END AS rating_num
    FROM order_t o
    WHERE o.customer_feedback IS NOT NULL
) ;
```

b)

```
SELECT
    quarter_number,
    AVG(rating_num) AS avg_rating
FROM (
    SELECT
        o.quarter_number,
        CASE o.customer_feedback
            WHEN 'Very Bad'  THEN 1
            WHEN 'Bad'       THEN 2
            WHEN 'Okay'      THEN 3
            WHEN 'Good'      THEN 4
            WHEN 'Very Good' THEN 5
        END AS rating_num
    FROM order_t o
    WHERE o.customer_feedback IS NOT NULL
)
GROUP BY quarter_number
ORDER BY quarter_number;
```

Output:

a)



Result: Passed

Query 1

Query:

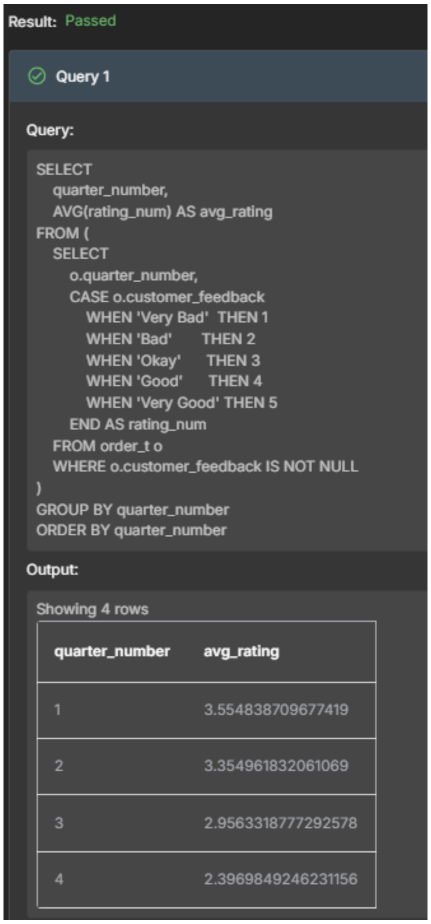
```
SELECT
   AVG(rating_num) AS overall_avg_rating
FROM (
   SELECT
      o.quarter_number,
      CASE o.customer_feedback
         WHEN 'Very Bad'  THEN 1
         WHEN 'Bad'       THEN 2
         WHEN 'Okay'      THEN 3
         WHEN 'Good'      THEN 4
         WHEN 'Very Good' THEN 5
      END AS rating_num
   FROM order_t o
   WHERE o.customer_feedback IS NOT NULL
)
```

Output:

Showing 1 rows

overall_avg_rating
3.135

b)



Result: Passed

Query 1

Query:

```
SELECT
   quarter_number,
   AVG(rating_num) AS avg_rating
FROM (
   SELECT
      o.quarter_number,
      CASE o.customer_feedback
         WHEN 'Very Bad'  THEN 1
         WHEN 'Bad'       THEN 2
         WHEN 'Okay'      THEN 3
         WHEN 'Good'      THEN 4
         WHEN 'Very Good' THEN 5
      END AS rating_num
   FROM order_t o
   WHERE o.customer_feedback IS NOT NULL
)
GROUP BY quarter_number
ORDER BY quarter_number
```

Output:

Showing 4 rows

quarter_number	avg_rating
1	3.554838709677419
2	3.354961832061069
3	2.9563318777292578
4	2.3969849246231156

Observations and Insights:

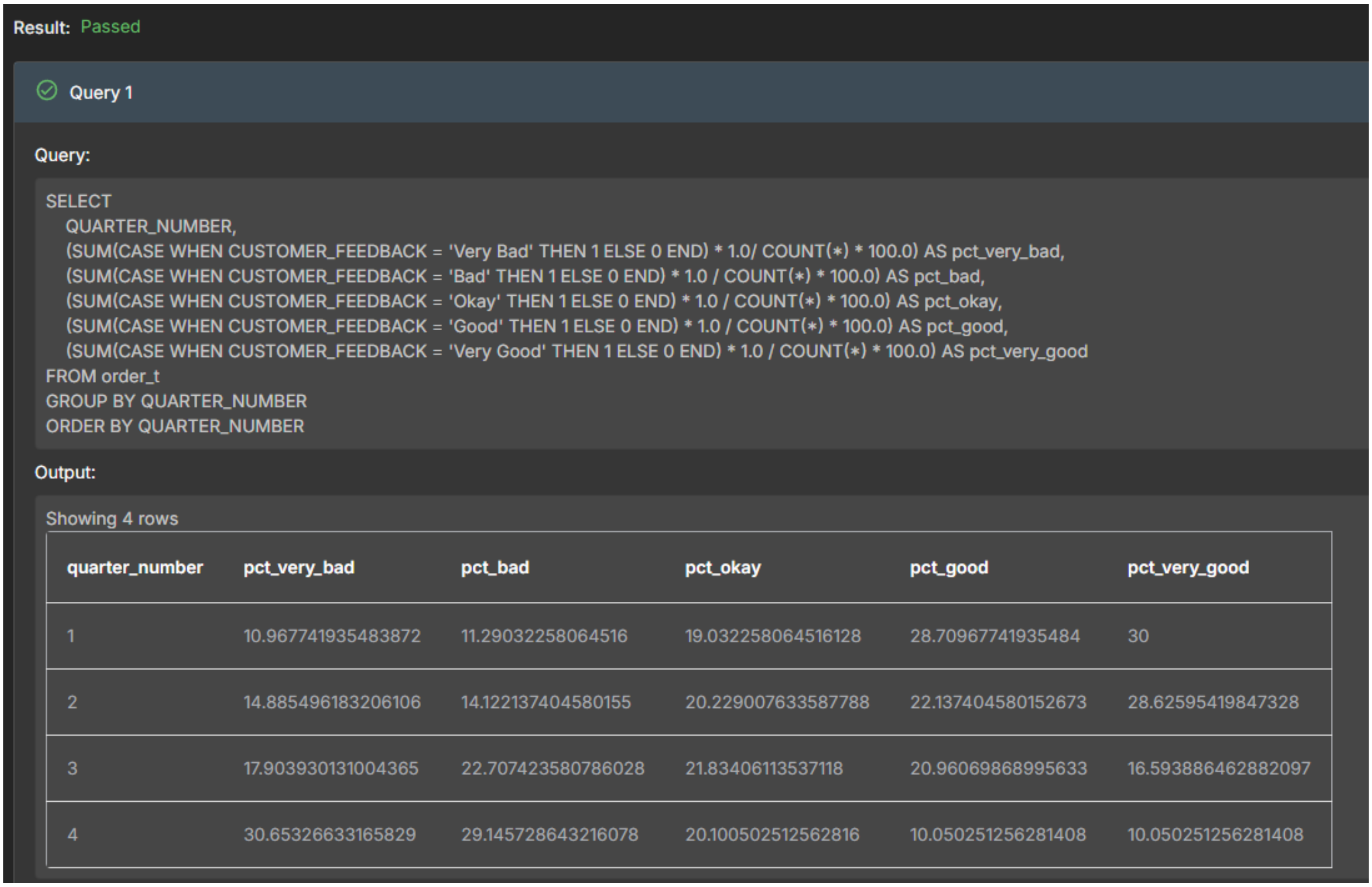
- Interestingly, the average customer rating decreased by one unit each quarter. In the first quarter, the average rating fell between ‘Okay’ and ‘Good,’ whereas by the last quarter, it had dropped closer to ‘Bad’.
- The largest decline in average ratings occurred between the third and fourth quarters, with approximately a 19% decrease.

Question 5: Find the percentage distribution of feedback from the customers. Are customers getting more dissatisfied over time?

Solution Query:

```
SELECT
    QUARTER_NUMBER,
    (SUM(CASE WHEN CUSTOMER_FEEDBACK = 'Very Bad' THEN 1 ELSE 0 END) * 1.0/ COUNT(*) * 100.0) AS pct_very_bad,
    (SUM(CASE WHEN CUSTOMER_FEEDBACK = 'Bad' THEN 1 ELSE 0 END) * 1.0 / COUNT(*) * 100.0) AS pct_bad,
    (SUM(CASE WHEN CUSTOMER_FEEDBACK = 'Okay' THEN 1 ELSE 0 END) * 1.0 / COUNT(*) * 100.0) AS pct_okay,
    (SUM(CASE WHEN CUSTOMER_FEEDBACK = 'Good' THEN 1 ELSE 0 END) * 1.0 / COUNT(*) * 100.0) AS pct_good,
    (SUM(CASE WHEN CUSTOMER_FEEDBACK = 'Very Good' THEN 1 ELSE 0 END) * 1.0 / COUNT(*) * 100.0) AS pct_very_good
FROM order_t
GROUP BY QUARTER_NUMBER
ORDER BY QUARTER_NUMBER;
```

Output:



Query:

```
SELECT
   QUARTER_NUMBER,
   (SUM(CASE WHEN CUSTOMER_FEEDBACK = 'Very Bad' THEN 1 ELSE 0 END) * 1.0/ COUNT(*) * 100.0) AS pct_very_bad,
   (SUM(CASE WHEN CUSTOMER_FEEDBACK = 'Bad' THEN 1 ELSE 0 END) * 1.0 / COUNT(*) * 100.0) AS pct_bad,
   (SUM(CASE WHEN CUSTOMER_FEEDBACK = 'Okay' THEN 1 ELSE 0 END) * 1.0 / COUNT(*) * 100.0) AS pct_okay,
   (SUM(CASE WHEN CUSTOMER_FEEDBACK = 'Good' THEN 1 ELSE 0 END) * 1.0 / COUNT(*) * 100.0) AS pct_good,
   (SUM(CASE WHEN CUSTOMER_FEEDBACK = 'Very Good' THEN 1 ELSE 0 END) * 1.0 / COUNT(*) * 100.0) AS pct_very_good
FROM order_t
GROUP BY QUARTER_NUMBER
ORDER BY QUARTER_NUMBER
```

Output:

Showing 4 rows

quarter_number	pct_very_bad	pct_bad	pct_okay	pct_good	pct_very_good
1	10.967741935483872	11.29032258064516	19.032258064516128	28.70967741935484	30
2	14.885496183206106	14.122137404580155	20.229007633587788	22.137404580152673	28.62595419847328
3	17.903930131004365	22.707423580786028	21.83406113537118	20.96069868995633	16.593886462882097
4	30.65326633165829	29.145728643216078	20.100502512562816	10.050251256281408	10.050251256281408

Observations and Insights:

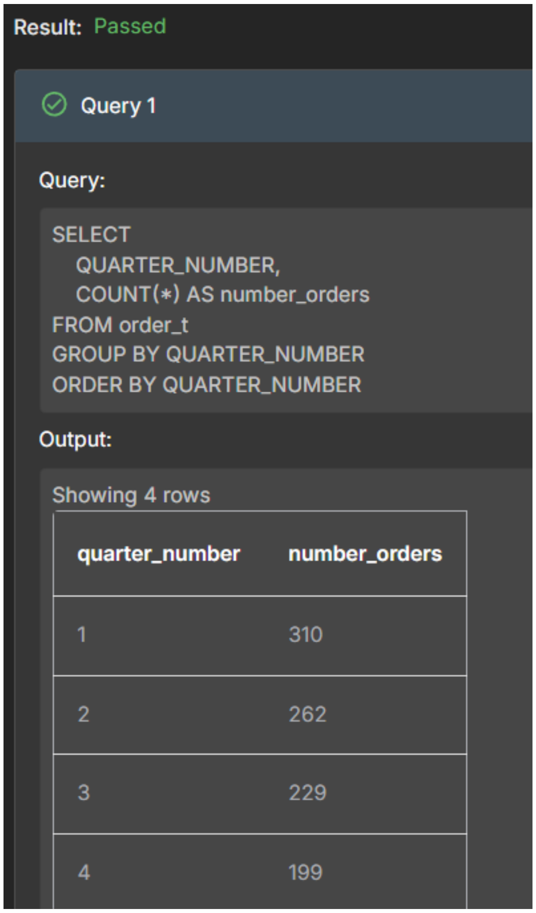
- Across all four quarters, the proportion of customers giving ‘Very Bad’ feedback increased each quarter.
- The percentage of customers that gave ‘Bad’ feedback increased over the 4 quarters, but less dramatically in comparison to the ‘Very Bad’ and ‘Very Good’ ratings.
- Customers that reported ‘Okay’ feed back remained relatively stable, with no drastic increases or decreases between any of the quarters.
- Over the same period, the share of customers giving ‘Very Good’ feedback decreased each quarter.
- The data suggests growing customer dissatisfaction over time, with the sharpest increase in ‘Very Bad’ feedback occurring between the 3rd and 4th quarters.

Question 6: What is the trend of the number of orders by quarter?

Solution Query:

```
SELECT

   QUARTER_NUMBER,

   COUNT(*) AS number_orders

FROM order_t

GROUP BY QUARTER_NUMBER

ORDER BY QUARTER_NUMBER;
```

Output:



Query:

```
SELECT
   QUARTER_NUMBER,
   COUNT(*) AS number_orders
FROM order_t
GROUP BY QUARTER_NUMBER
ORDER BY QUARTER_NUMBER
```

Output:

Showing 4 rows

quarter_number	number_orders
1	310
2	262
3	229
4	199

Observations and Insights:

- The number or orders decreased each quarter, with the steepest drop occurring between the 1st and 2nd quarters (48 fewer orders).
- The data indicates that rising customer dissatisfaction may have contributed to a decline in orders over time.

Question 7: a)Calculate the net revenue generated by the company.

b)What is the quarter-over-quarter % change in net revenue?

Solution Query:

a)

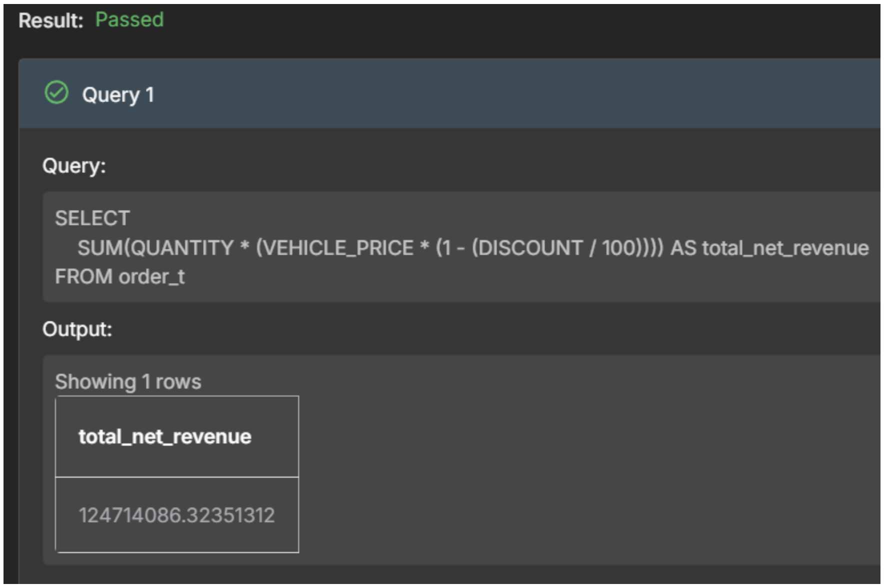
```
SELECT
    SUM(QUANTITY * (VEHICLE_PRICE * (1 - (DISCOUNT / 100)))) AS total_net_revenue
FROM order_t;
```

b)

```
SELECT
    QUARTER_NUMBER,
    SUM(QUANTITY * VEHICLE_PRICE * (1 - (DISCOUNT / 100))) AS net_revenue,
    LAG(SUM(QUANTITY * VEHICLE_PRICE * (1 - (DISCOUNT / 100))))
        OVER (ORDER BY QUARTER_NUMBER) AS prev_revenue,
    ROUND(
        (
            (SUM(QUANTITY * VEHICLE_PRICE * (1 - (DISCOUNT / 100)))
            - LAG(SUM(QUANTITY * VEHICLE_PRICE * (1 - (DISCOUNT / 100))))
                OVER (ORDER BY QUARTER_NUMBER))
            / LAG(SUM(QUANTITY * VEHICLE_PRICE * (1 - (DISCOUNT / 100))))
                OVER (ORDER BY QUARTER_NUMBER)
        ) * 100, 2
    ) AS qoq_pct_change
FROM order_t
GROUP BY QUARTER_NUMBER
ORDER BY QUARTER_NUMBER;
```

Output:

a)



Result: Passed

Query 1

Query:

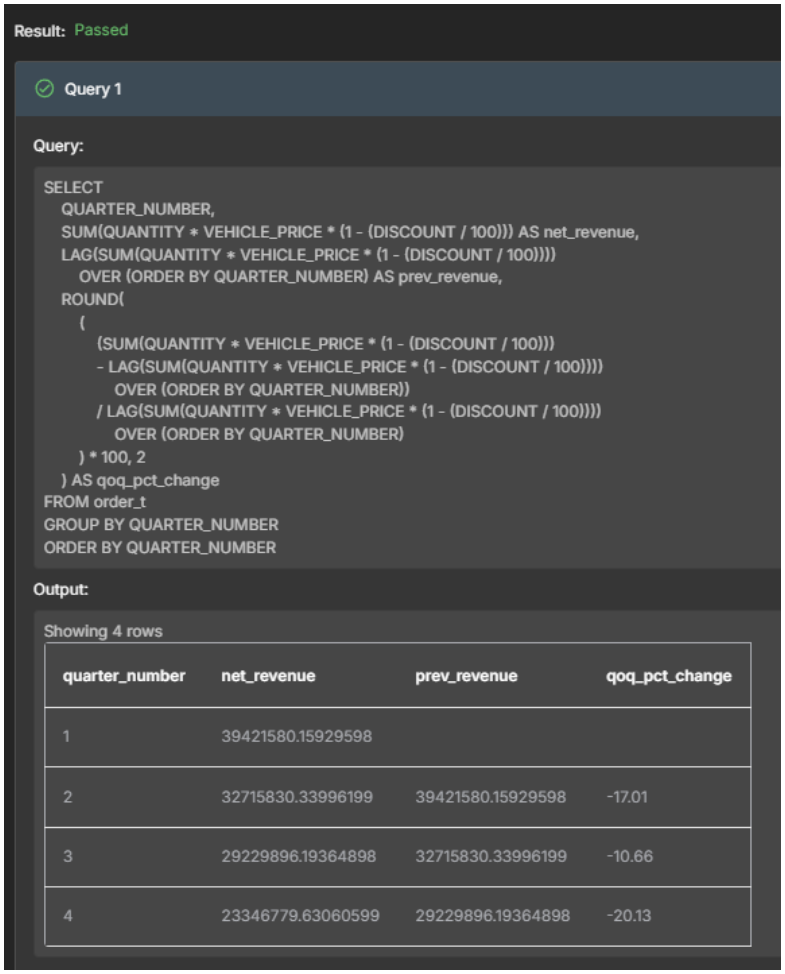
```
SELECT
  SUM(QUANTITY * (VEHICLE_PRICE * (1 - (DISCOUNT / 100)))) AS total_net_revenue
FROM order_t
```

Output:

Showing 1 rows

total_net_revenue
124714086.32351312

b)



Result: Passed

Query 1

Query:

```
SELECT
  QUARTER_NUMBER,
  SUM(QUANTITY * VEHICLE_PRICE * (1 - (DISCOUNT / 100))) AS net_revenue,
  LAG(SUM(QUANTITY * VEHICLE_PRICE * (1 - (DISCOUNT / 100))))
    OVER (ORDER BY QUARTER_NUMBER) AS prev_revenue,
  ROUND(
    (
      (SUM(QUANTITY * VEHICLE_PRICE * (1 - (DISCOUNT / 100)))
      - LAG(SUM(QUANTITY * VEHICLE_PRICE * (1 - (DISCOUNT / 100))))
        OVER (ORDER BY QUARTER_NUMBER))
      / LAG(SUM(QUANTITY * VEHICLE_PRICE * (1 - (DISCOUNT / 100))))
        OVER (ORDER BY QUARTER_NUMBER)
    ) * 100, 2
  ) AS qoq_pct_change
FROM order_t
GROUP BY QUARTER_NUMBER
ORDER BY QUARTER_NUMBER
```

Output:

Showing 4 rows

quarter_number	net_revenue	prev_revenue	qoq_pct_change
1	39421580.15929598		
2	32715830.33996199	39421580.15929598	-17.01
3	29229896.19364898	32715830.33996199	-10.66
4	23346779.63060599	29229896.19364898	-20.13

Observations and Insights:

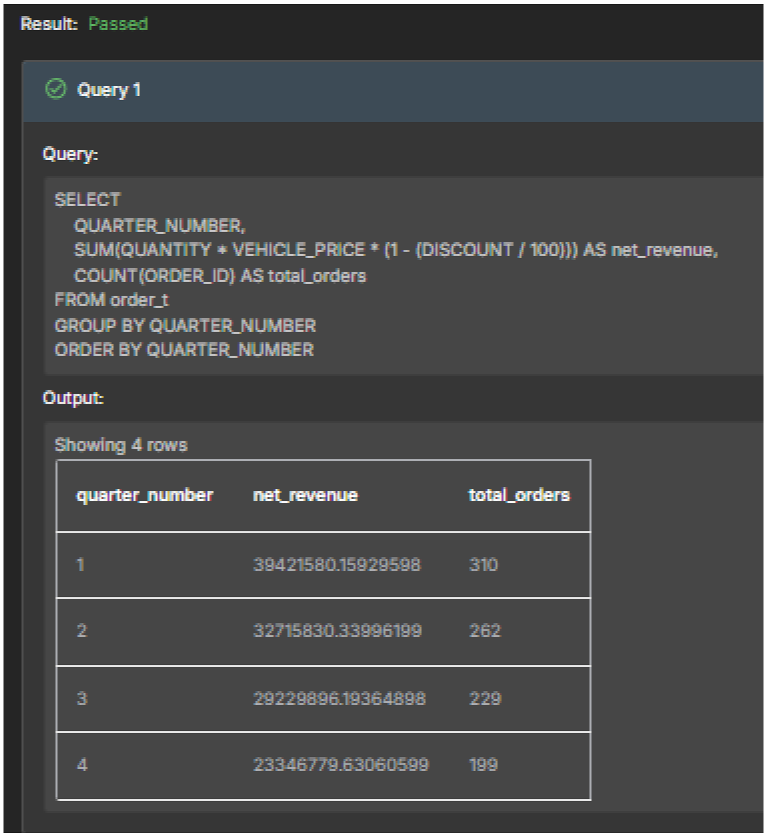
- The total revenue generated by the company is ~$124,714,086.32.
- Revenue peaked in the 1st quarter and declined in each subsequent quarter, with the largest drop (about 20%) occurring between the 3rd and 4th quarters.

Question 8: What is the trend of net revenue and orders by quarters?

Solution Query:

```
SELECT

   QUARTER_NUMBER,

   SUM(QUANTITY * VEHICLE_PRICE * (1 - (DISCOUNT / 100))) AS net_revenue,

   COUNT(ORDER_ID) AS total_orders

FROM order_t

GROUP BY QUARTER_NUMBER

ORDER BY QUARTER_NUMBER;
```

Output:



Result: Passed

Query 1

Query:

```
SELECT
   QUARTER_NUMBER,
   SUM(QUANTITY * VEHICLE_PRICE * (1 - (DISCOUNT / 100))) AS net_revenue,
   COUNT(ORDER_ID) AS total_orders
FROM order_t
GROUP BY QUARTER_NUMBER
ORDER BY QUARTER_NUMBER
```

Output:

Showing 4 rows

quarter_number	net_revenue	total_orders
1	39421580.15929598	310
2	32715830.33996199	262
3	29229896.19364898	229
4	23346779.63060599	199

Observations and Insights:

- The net revenue and number of orders show a clear correlation across all quarters, with both steadily declining each quarter.

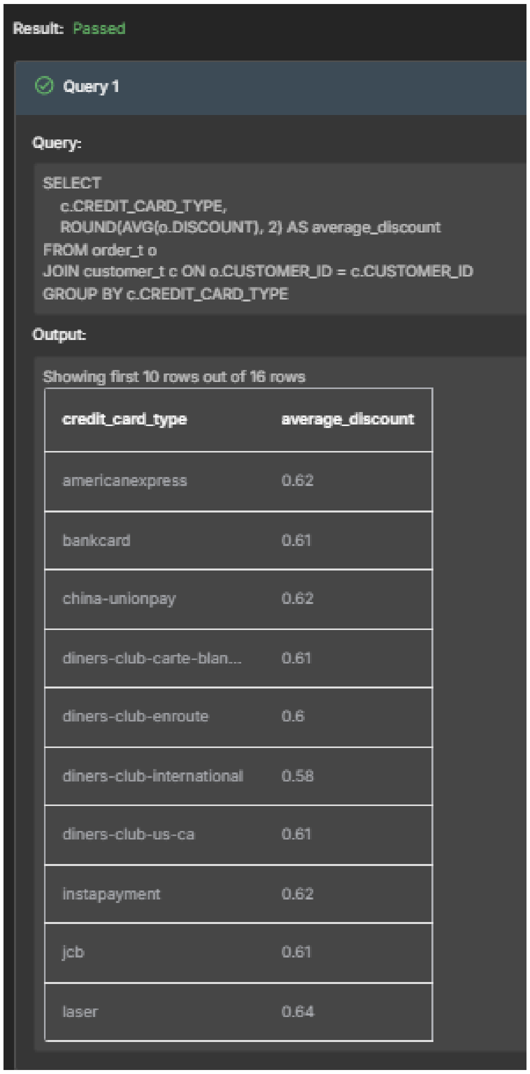
Question 9: What is the average discount offered for different types of credit cards?

Solution Query:

```
SELECT

   c.CREDIT_CARD_TYPE,

   ROUND(AVG(o.DISCOUNT), 2) AS average_discount

FROM order_t o

JOIN customer_t c ON o.CUSTOMER_ID = c.CUSTOMER_ID

GROUP BY c.CREDIT_CARD_TYPE;
```

Note: The ‘`discount`’ field in the ‘`order_t`’ table is stored as a percentage, i.e. `0.6` represents a discount of `0.6%`, not `60%`.

Output:



Result: Passed

Query 1

Query:

```
SELECT
   c.CREDIT_CARD_TYPE,
   ROUND(AVG(o.DISCOUNT), 2) AS average_discount
FROM order_t o
JOIN customer_t c ON o.CUSTOMER_ID = c.CUSTOMER_ID
GROUP BY c.CREDIT_CARD_TYPE
```

Output:

Showing first 10 rows out of 16 rows

credit_card_type	average_discount
americanexpress	0.62
bankcard	0.61
china-unionpay	0.62
diners-club-carte-blan...	0.61
diners-club-enroute	0.6
diners-club-international	0.58
diners-club-us-ca	0.61
instapayment	0.62
jcb	0.61
laser	0.64

Observations and Insights:

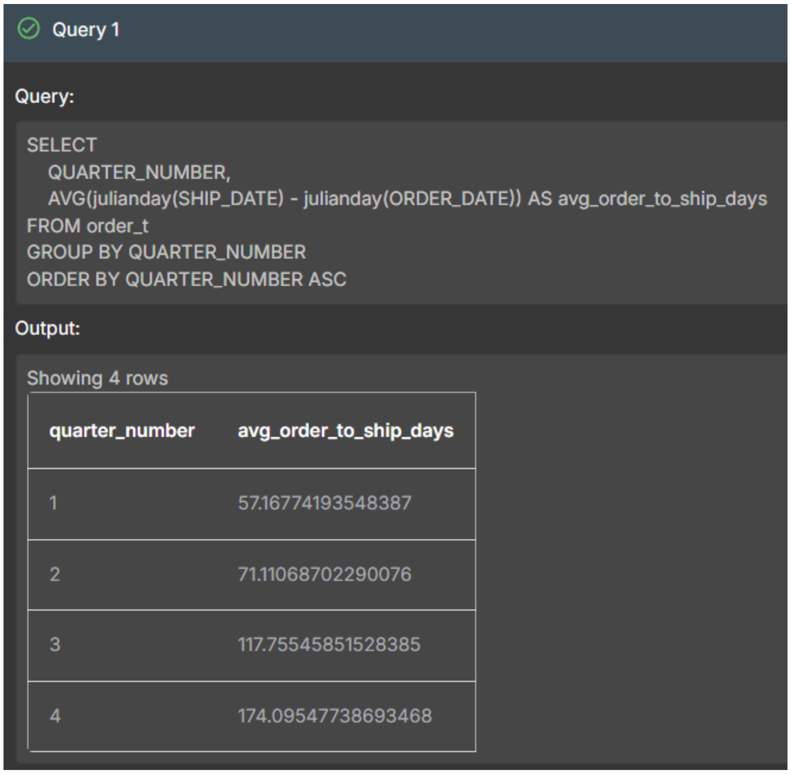
- For every type of credit card, the average discount is about 0.6%
- After ordering the credit cards in ascending and descending order, it was found that the credit card with the largest discount is ‘laser’ at 0.64%, and the credit card with the smallest discount is ‘diners_club_international’ at 0.58%.

Question 10: What is the average time taken to ship the placed orders for each quarter?

Solution Query:

```
SELECT

    QUARTER_NUMBER,

    AVG(julianday(SHIP_DATE) - julianday(ORDER_DATE)) AS avg_order_to_ship_days

FROM order_t

GROUP BY QUARTER_NUMBER

ORDER BY QUARTER_NUMBER ASC;
```

Output:



Query 1

Query:

```
SELECT
   QUARTER_NUMBER,
   AVG(julianday(SHIP_DATE) - julianday(ORDER_DATE)) AS avg_order_to_ship_days
FROM order_t
GROUP BY QUARTER_NUMBER
ORDER BY QUARTER_NUMBER ASC
```

Output:

Showing 4 rows

quarter_number	avg_order_to_ship_days
1	57.16774193548387
2	71.11068702290076
3	117.75545851528385
4	174.09547738693468

Observations and Insights:

- The average shipping duration increases each quarter, meaning customers are waiting longer for their orders as the year progresses.
- This delay may have contributed to the decline in customer ratings across quarters, since longer wait times often lead to lower satisfaction.

Business Metrics Overview

Total Revenue	Total Orders	Total Customers	Average Rating
124714086.32	1000	994	3.135
Last Quarter Revenue	**Last quarter Orders**	**Average Days to Ship**	**% Good Feedback**
23346779.63	199	97.96	21.5%

Note: These values must be derived using SQL queries. Some of them may have already been obtained while answering previous questions.

Business Recommendations

- Prioritize customer satisfaction in later quarters, as ratings consistently declined toward year-end.
- Address the decline in order volume and revenue by analyzing customer needs and implementing retention strategies.
- Reduce shipping delays by streamlining logistics, since the average time from order placement to delivery increased significantly in the last quarter.
- Introduce targeted year-end promotions or savings programs to boost order volume and enhance customer satisfaction during slower periods.