# 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.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# Business Questions

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

**Solution Query:**

To find the total number of customers:

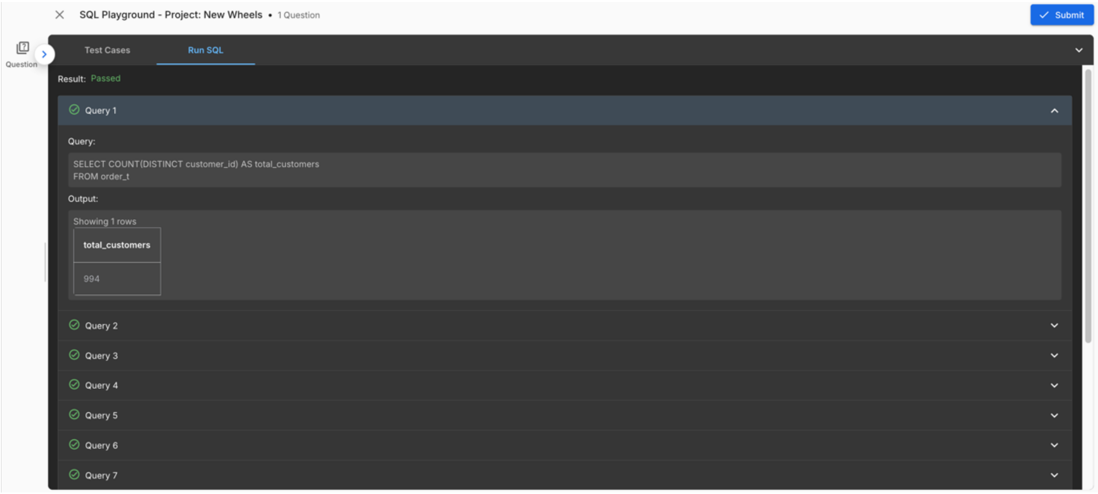
```
SELECT COUNT (DISTINCT customer_id) AS total_customers
FROM order_t;
```

To find the distribution of the customers across states:

```
SELECT state,COUNT(customer_id) AS TOTAL
FROM customer_t
GROUP BY state;
```

**Output:**

- To find the total number of customers:



Query:

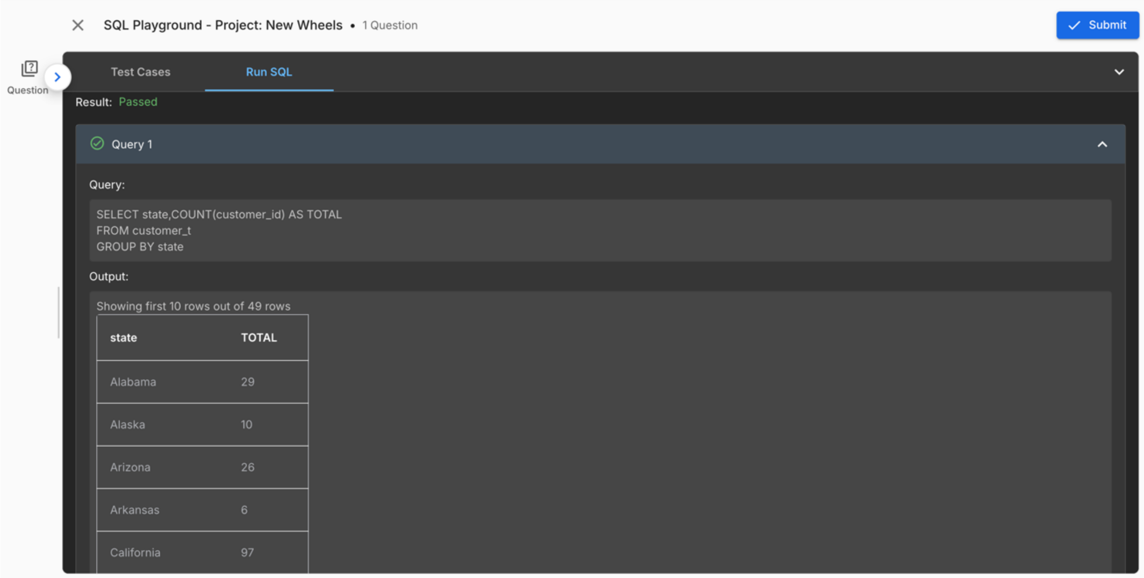
```
SELECT COUNT(DISTINCT customer_id) AS total_customers
FROM order_t
```

Output:

Showing 1 rows

| total_customers |
|---|
| 994 |

- To find the distribution of the customers across states



Query:

```
SELECT state,COUNT(customer_id) AS TOTAL
FROM customer_t
GROUP BY state
```

Output:

Showing first 10 rows out of 49 rows

| state | TOTAL |
|---|---|
| Alabama | 29 |
| Alaska | 10 |
| Arizona | 26 |
| Arkansas | 6 |
| California | 97 |

**Observations and Insights:**

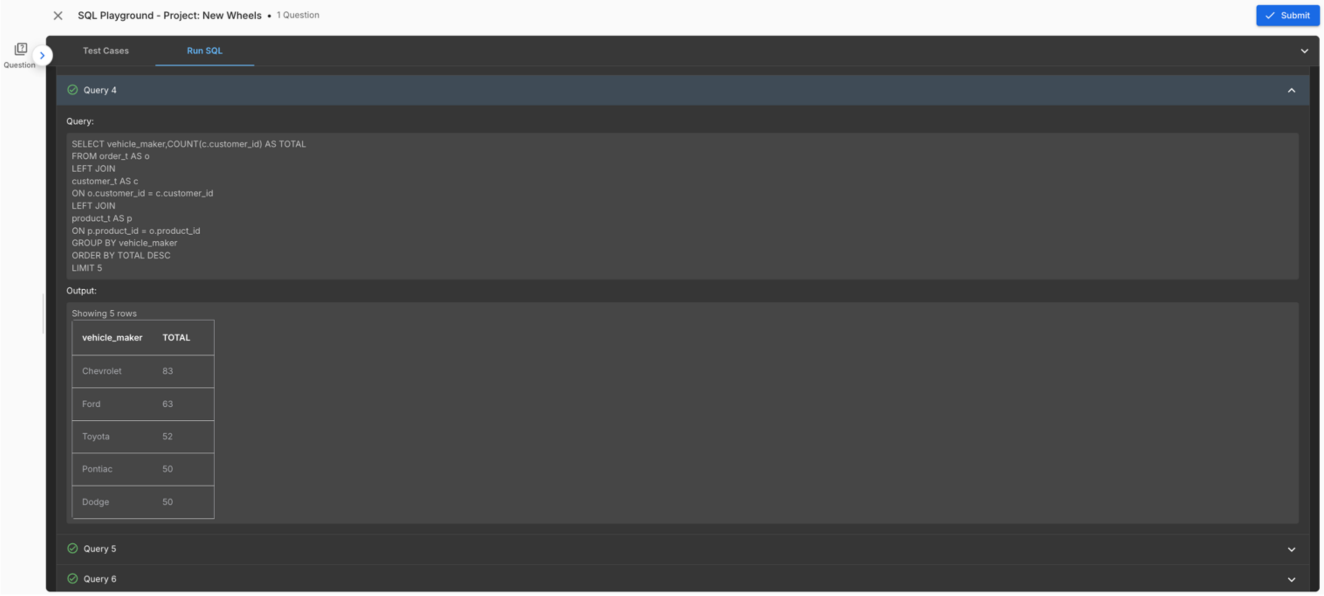
- There are 994 customers who ordered cars.
- There are 29 orders that come from the state of Alabama.
- There are 97 orders that come from the state of California.

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

**Solution Query:**

```
SELECT vehicle_maker,COUNT(c.customer_id) AS TOTAL
FROM order_t AS o
LEFT JOIN customer_t AS c
ON o.customer_id = c.customer_id
LEFT JOIN product_t AS p
ON p.product_id = o.product_id
GROUP BY vehicle_maker
ORDER BY TOTAL DESC
LIMIT 5;
```

**Output:**



Query:

```
SELECT vehicle_maker,COUNT(c.customer_id) AS TOTAL
FROM order_t AS o
LEFT JOIN
customer_t AS c
ON o.customer_id = c.customer_id
LEFT JOIN
product_t AS p
ON p.product_id = o.product_id
GROUP BY vehicle_maker
ORDER BY TOTAL DESC
LIMIT 5
```

Output:

Showing 5 rows

| vehicle_maker | TOTAL |
|---|---|
| Chevrolet | 83 |
| Ford | 63 |
| Toyota | 52 |
| Pontiac | 50 |
| Dodge | 50 |

**Observations and Insights:**

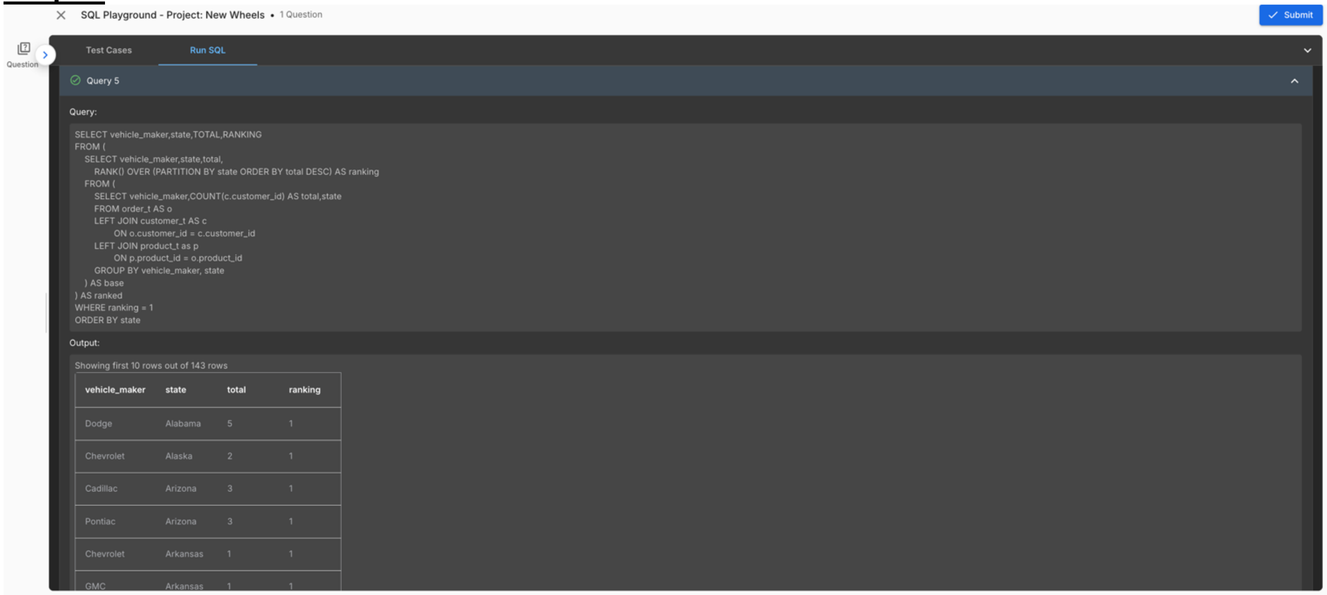
- Chevrolet is the first vehicle maker preferred by the customers.
- Ford is the second vehicle makers preferred by the customers.
- Pontiac and Dodge are tied for fourth vehicle makers preferred by the customers.

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

**Solution Query:**

```
SELECT vehicle_maker,state,TOTAL,RANKING
FROM (
        SELECT vehicle_maker,state,total,
                RANK() OVER (PARTITION BY state ORDER BY total DESC) AS ranking
        FROM (
                SELECT vehicle_maker,COUNT(c.customer_id) AS total,state
                FROM order_t AS o
                LEFT JOIN customer_t AS c
                        ON o.customer_id = c.customer_id
                LEFT JOIN product_t as p
                        ON p.product_id = o.product_id
                GROUP BY vehicle_maker, state
        ) AS base
) AS ranked
WHERE ranking = 1
ORDER BY state;
```

**Output:**



Query:

```
SELECT vehicle_maker,state,TOTAL,RANKING
FROM (
  SELECT vehicle_maker,state,total,
    RANK() OVER (PARTITION BY state ORDER BY total DESC) AS ranking
  FROM (
    SELECT vehicle_maker,COUNT(c.customer_id) AS total,state
    FROM order_t AS o
    LEFT JOIN customer_t AS c
        ON o.customer_id = c.customer_id
    LEFT JOIN product_t as p
        ON p.product_id = o.product_id
    GROUP BY vehicle_maker, state
  ) AS base
) AS ranked
WHERE ranking = 1
ORDER BY state
```

Output:

Showing first 10 rows out of 143 rows

| vehicle_maker | state | total | ranking |
|---|---|---|---|
| Dodge | Alabama | 5 | 1 |
| Chevrolet | Alaska | 2 | 1 |
| Cadillac | Arizona | 3 | 1 |
| Pontiac | Arizona | 3 | 1 |
| Chevrolet | Arkansas | 1 | 1 |
| GMC | Arkansas | 1 | 1 |

**Observations and Insights:**

- The preferred vehicle maker in Alabama is Dodge with a total of 5 people .
- The preferred vehicle maker in Alaska is Chevrolet with a total of 2 people.
- The preferred vehicle makers in Arizona are Pontiac and Cadillac with a tied total of 3 people.

## Question 4: Find the overall average rating given by the customers. 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:**

For finding the overall average rating:

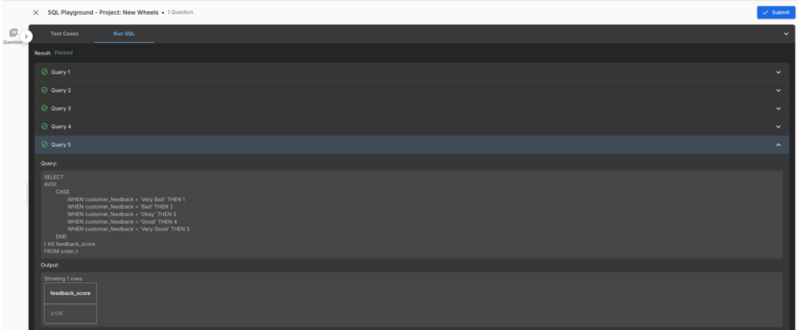
```
SELECT
AVG(
        CASE
                WHEN customer_feedback = 'Very Bad' THEN 1
                WHEN customer_feedback = 'Bad' THEN 2
                WHEN customer_feedback = 'Okay' THEN 3
                WHEN customer_feedback = 'Good' THEN 4
                WHEN customer_feedback = 'Very Good' THEN 5
        END
) AS feedback_score
FROM order_t;
```

For finding the average rating in each quarter:

```
SELECT quarter_number,
ROUND (AVG(
        CASE
        WHEN customer_feedback = 'Very Bad' THEN 1
        WHEN customer_feedback = 'Bad' THEN 2
        WHEN customer_feedback = 'Okay' THEN 3
        WHEN customer_feedback = 'Good' THEN 4
        WHEN customer_feedback = 'Very Good' THEN 5
        END
),2) AS feedback_score
FROM order_t
GROUP BY quarter_number
ORDER BY quarter_number;
```

**Output:**

- For finding the overall average rating

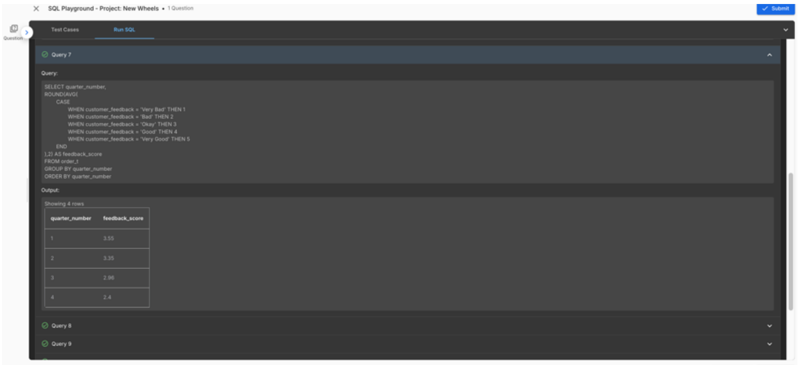


```
SELECT
AVG(
    CASE
        WHEN customer_feedback = 'Very Bad' THEN 1
        WHEN customer_feedback = 'Bad' THEN 2
        WHEN customer_feedback = 'Okay' THEN 3
        WHEN customer_feedback = 'Good' THEN 4
        WHEN customer_feedback = 'Very Good' THEN 5
    END
) AS feedback_score
FROM order_t
```

Showing 1 rows

| feedback_score |
|---|
| 3.135 |

- For finding the average rating in each quarter



```
SELECT quarter_number,
ROUND(AVG(
    CASE
        WHEN customer_feedback = 'Very Bad' THEN 1
        WHEN customer_feedback = 'Bad' THEN 2
        WHEN customer_feedback = 'Okay' THEN 3
        WHEN customer_feedback = 'Good' THEN 4
        WHEN customer_feedback = 'Very Good' THEN 5
    END
),2) AS feedback_score
FROM order_t
GROUP BY quarter_number
ORDER BY quarter_number
```

Showing 4 rows

| quarter_number | feedback_score |
|---|---|
| 1 | 3.55 |
| 2 | 3.35 |
| 3 | 2.96 |
| 4 | 2.4 |

**Observations and Insights:**

- The average rating given by customers was 3.135 which is about an “Okay” rating.
- When the Quarter goes by, the average rating went downward.
    - This makes sense given the problem statement expressed concerns of a downward trend.
- The range from Quarter 3 to Quarter 4 was when the rating went from plateaued to downward trend.

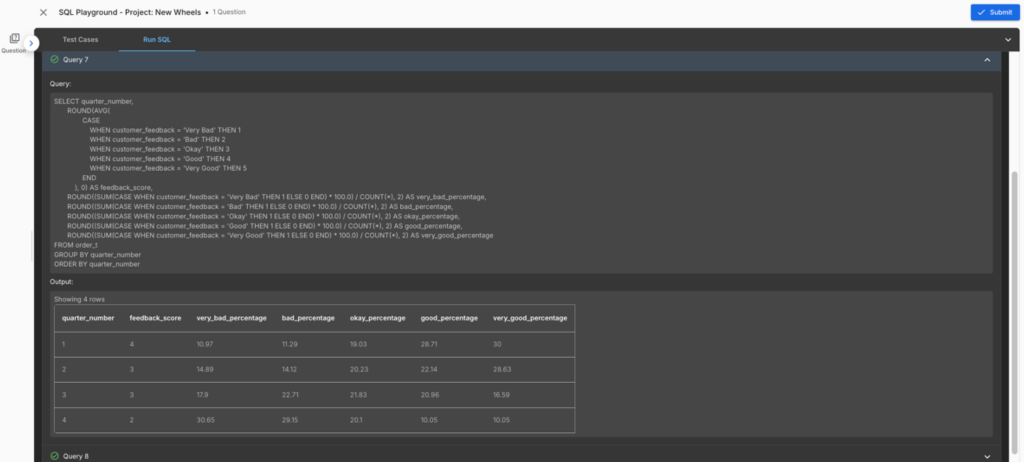
## Question 5: Find the percentage distribution of feedback from the customers. Are customers getting more dissatisfied over time?

**Solution Query:**

To find the percentage distribution of feedback from the customers and whether customers are getting dissatisfied over time:

```
SELECT quarter_number,
        ROUND(AVG(
                CASE
                        WHEN customer_feedback = 'Very Bad' THEN 1
                        WHEN customer_feedback = 'Bad' THEN 2
                        WHEN customer_feedback = 'Okay' THEN 3
                        WHEN customer_feedback = 'Good' THEN 4
                        WHEN customer_feedback = 'Very Good' THEN 5
                END
                ), 0) AS feedback_score,
        ROUND((SUM(CASE WHEN customer_feedback = 'Very Bad' THEN 1 ELSE 0 END) * 100.0) / COUNT(*),
        2) AS very_bad_percentage,
        ROUND((SUM(CASE WHEN customer_feedback = 'Bad' THEN 1 ELSE 0 END) * 100.0) / COUNT(*), 2) AS
        bad_percentage,
        ROUND((SUM(CASE WHEN customer_feedback = 'Okay' THEN 1 ELSE 0 END) * 100.0) / COUNT(*), 2) AS
        okay_percentage,
        ROUND((SUM(CASE WHEN customer_feedback = 'Good' THEN 1 ELSE 0 END) * 100.0) / COUNT(*), 2) AS
        good_percentage,
        ROUND((SUM(CASE WHEN customer_feedback = 'Very Good' THEN 1 ELSE 0 END) * 100.0) / COUNT(*),
        2) AS very_good_percentage
FROM order_t
GROUP BY quarter_number
ORDER BY quarter_number;
```

**Output:**



Query:

```
SELECT quarter_number,
    ROUND(AVG(
        CASE
          WHEN customer_feedback = 'Very Bad' THEN 1
          WHEN customer_feedback = 'Bad' THEN 2
          WHEN customer_feedback = 'Okay' THEN 3
          WHEN customer_feedback = 'Good' THEN 4
          WHEN customer_feedback = 'Very Good' THEN 5
        END
      ), 0) AS feedback_score,
    ROUND((SUM(CASE WHEN customer_feedback = 'Very Bad' THEN 1 ELSE 0 END) * 100.0) / COUNT(*), 2) AS very_bad_percentage,
    ROUND((SUM(CASE WHEN customer_feedback = 'Bad' THEN 1 ELSE 0 END) * 100.0) / COUNT(*), 2) AS bad_percentage,
    ROUND((SUM(CASE WHEN customer_feedback = 'Okay' THEN 1 ELSE 0 END) * 100.0) / COUNT(*), 2) AS okay_percentage,
    ROUND((SUM(CASE WHEN customer_feedback = 'Good' THEN 1 ELSE 0 END) * 100.0) / COUNT(*), 2) AS good_percentage,
    ROUND((SUM(CASE WHEN customer_feedback = 'Very Good' THEN 1 ELSE 0 END) * 100.0) / COUNT(*), 2) AS very_good_percentage
FROM order_t
GROUP BY quarter_number
ORDER BY quarter_number
```

Output:

Showing 4 rows

| quarter_number | feedback_score | very_bad_percentage | bad_percentage | okay_percentage | good_percentage | very_good_percentage |
|---|---|---|---|---|---|---|
| 1 | 4 | 10.97 | 11.29 | 19.03 | 28.71 | 30 |
| 2 | 3 | 14.89 | 14.12 | 20.23 | 22.14 | 28.63 |
| 3 | 3 | 17.9 | 22.71 | 21.83 | 20.96 | 16.59 |
| 4 | 2 | 30.65 | 29.15 | 20.1 | 10.05 | 10.05 |

**Observations and Insights:**

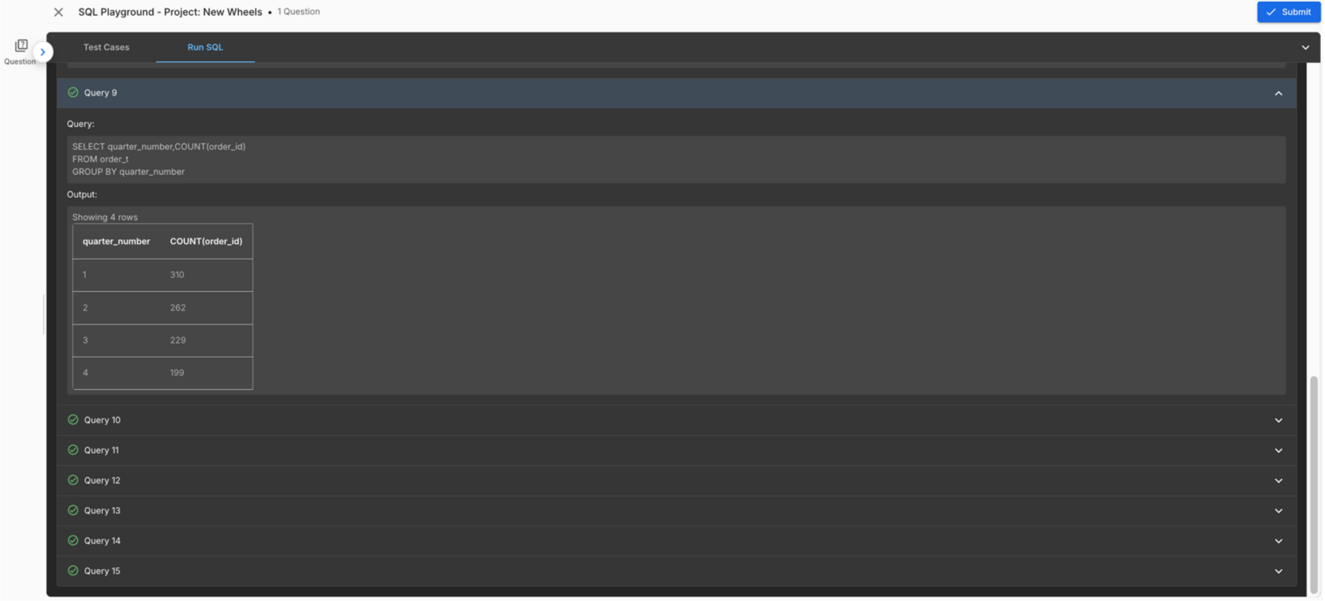
- As the quarter goes by, the “Very Bad” and “Bad” percentage had an increase as the quarter goes by.
  - This means that the customers are dissatisfied as the quarter goes by
- The “Good” and “Very Good” percentage had a decrease as the quarter goes by
  - This also means that the customers are dissatisfied as the quarter goes by.
- The “Okay” percentage had a steady increase until quarter 4 when it went down.

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

**Solution Query:**

```
SELECT quarter_number,COUNT(order_id)
FROM order_t
GROUP BY quarter_number;
```

**Output:**



Query:

```
SELECT quarter_number,COUNT(order_id)
FROM order_t
GROUP BY quarter_number
```

Output:

Showing 4 rows

| quarter_number | COUNT(order_id) |
|---|---|
| 1 | 310 |
| 2 | 262 |
| 3 | 229 |
| 4 | 199 |

**Observations and Insights:**

- As the quarter goes by, the quantity goes down.
- The range between Quarter 1 and Quarter 4 was 111.
- The first quarter to second quarter yielded the most drop of 48 less orders.

## Question 7: Calculate the net revenue generated by the company. What is the quarter-over-quarter % change in net revenue?

**Solution Query:**

To find the net revenue generated by the company:

```
SELECT ROUND(SUM(quantity*vehicle_price*(1 - discount)),2) AS total_net_revenue
FROM order_t;
```

To find the quarter over quarter percent change:

[find the quarter over quarter]

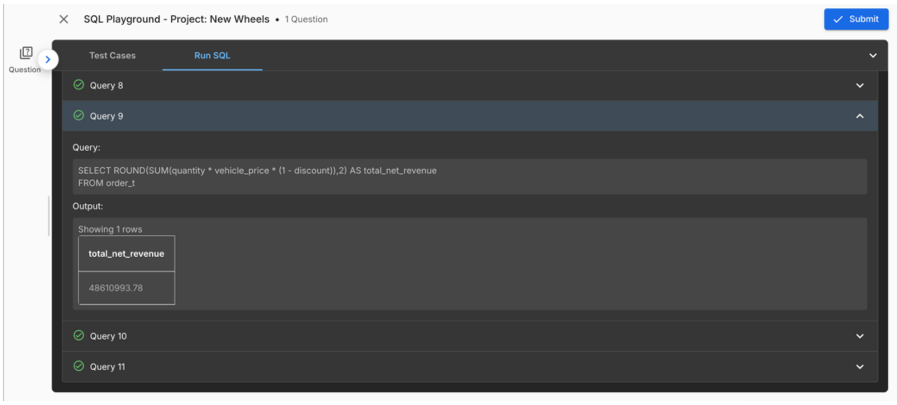
```
SELECT quarter_number,net_revenue,
  LAG(net_revenue) OVER (ORDER BY quarter_number) AS previous_net_revenue
FROM (
  SELECT quarter_number,ROUND(SUM(quantity * vehicle_price * (1 - discount)), 2) AS net_revenue
  FROM order_t
  GROUP BY quarter_number
) AS sub
ORDER BY quarter_number;
```

[find the quarter over quarter percent change]

```
SELECT quarter_number,net_revenue,
  LAG(net_revenue) OVER (ORDER BY quarter_number) AS previous_net_revenue,
  ROUND(100.0 * (net_revenue - LAG(net_revenue) OVER (ORDER BY quarter_number)) /
     NULLIF(LAG(net_revenue) OVER (ORDER BY quarter_number), 0),2) AS percent_change
  FROM (
     SELECT quarter_number, ROUND(SUM(quantity * vehicle_price * (1 - discount)), 2) AS net_revenue
     FROM order_t
     GROUP BY quarter_number
  ) AS sub
ORDER BY quarter_number;
```

**Output:**

- To find the net revenue generated by the company



Query:

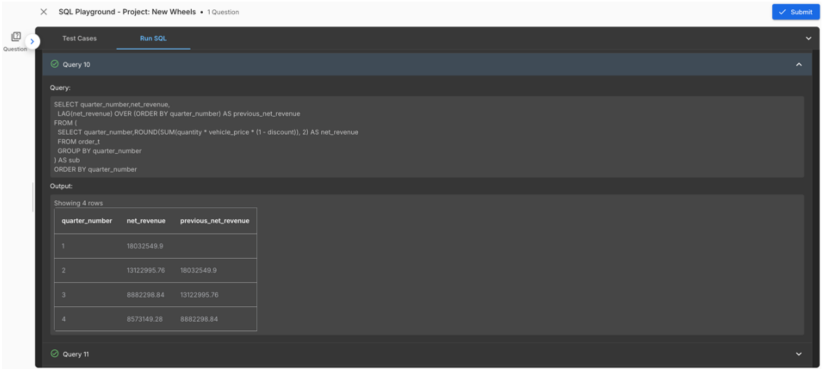
```
SELECT ROUND(SUM(quantity * vehicle_price * (1 - discount)),2) AS total_net_revenue
FROM order_t
```

Output:

Showing 1 rows

| total_net_revenue |
|---|
| 48610993.78 |

- To find the quarter over quarter percentage change
    - First find the quarter over quarter



Query:

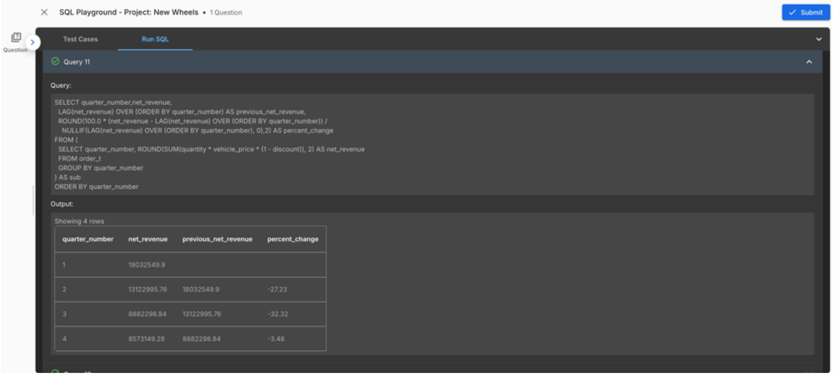
```
SELECT quarter_number,net_revenue,
 LAG(net_revenue) OVER (ORDER BY quarter_number) AS previous_net_revenue
FROM (
 SELECT quarter_number,ROUND(SUM(quantity * vehicle_price * (1 - discount)), 2) AS net_revenue
 FROM order_t
 GROUP BY quarter_number
) AS sub
ORDER BY quarter_number
```

Output:

Showing 4 rows

| quarter_number | net_revenue | previous_net_revenue |
|---|---|---|
| 1 | 18032549.9 | |
| 2 | 13122995.76 | 18032549.9 |
| 3 | 8882298.84 | 13122995.76 |
| 4 | 8573149.28 | 8882298.84 |

- Then calculate the quarter over quarter percentage change



Query:

```
SELECT quarter_number,net_revenue,
  LAG(net_revenue) OVER (ORDER BY quarter_number) AS previous_net_revenue,
  ROUND(100.0 * (net_revenue - LAG(net_revenue) OVER (ORDER BY quarter_number)) /
    NULLIF(LAG(net_revenue) OVER (ORDER BY quarter_number), 0),2) AS percent_change
FROM (
  SELECT quarter_number, ROUND(SUM(quantity * vehicle_price * (1 - discount)), 2) AS net_revenue
  FROM order_t
  GROUP BY quarter_number
) AS sub
ORDER BY quarter_number
```

Output:

Showing 4 rows

| quarter_number | net_revenue | previous_net_revenue | percent_change |
|---|---|---|---|
| 1 | 18032549.9 | | |
| 2 | 13122995.76 | 18032549.9 | -27.23 |
| 3 | 8882298.84 | 13122995.76 | -32.32 |
| 4 | 8573149.28 | 8882298.84 | -3.48 |

## Observations and Insights:

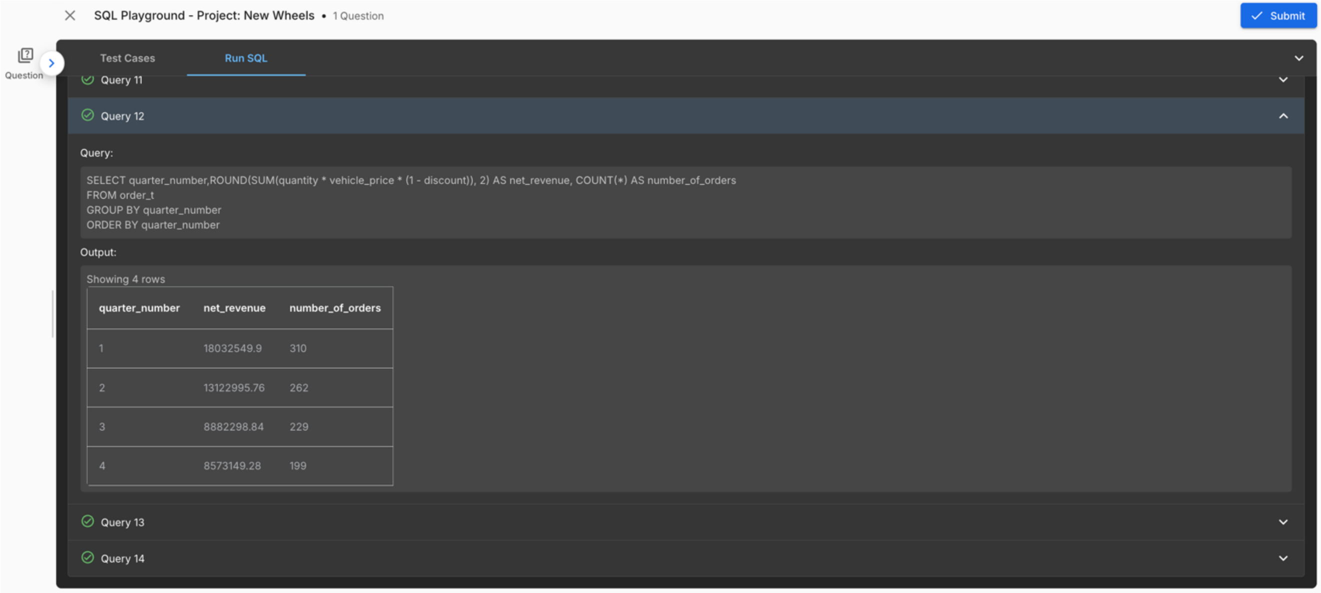
- The total net revenue is $48,610,993.78.
- As the quarter went by, the revenue went down.
- The biggest percentage change was from quarter 2 to quarter 3 with 32.32% decrease.

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

**Solution Query:**

```
SELECT quarter_number,ROUND(SUM(quantity * vehicle_price * (1 - discount)), 2) AS
net_revenue, COUNT(*) AS number_of_orders
FROM order_t
GROUP BY quarter_number
ORDER BY quarter_number;
```

**Output:**



Query:

```
SELECT quarter_number,ROUND(SUM(quantity * vehicle_price * (1 - discount)), 2) AS net_revenue, COUNT(*) AS number_of_orders
FROM order_t
GROUP BY quarter_number
ORDER BY quarter_number
```

Output:

Showing 4 rows

| quarter_number | net_revenue | number_of_orders |
|---|---|---|
| 1 | 18032549.9 | 310 |
| 2 | 13122995.76 | 262 |
| 3 | 8882298.84 | 229 |
| 4 | 8573149.28 | 199 |

**Observations and Insights:**

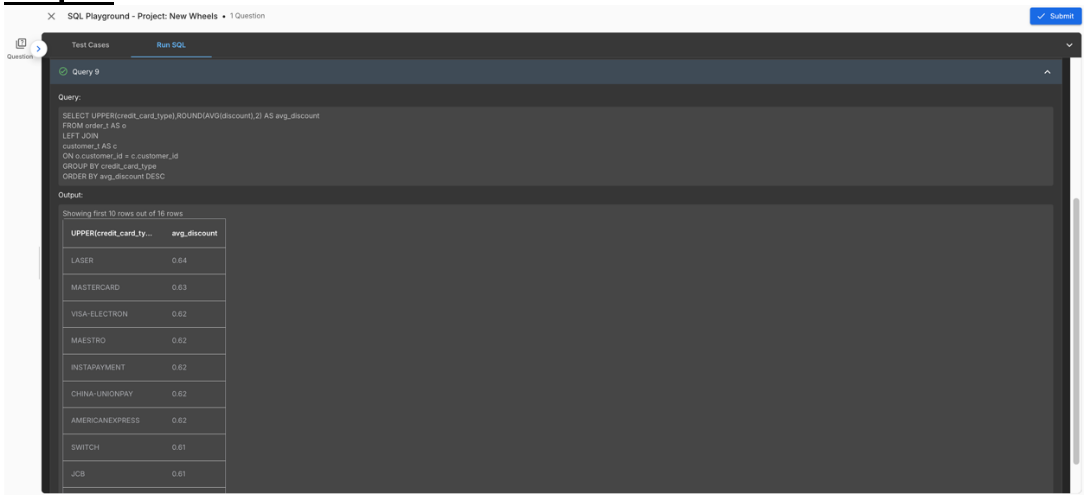
- As the quarter goes by, the net revenue and the number of orders decreases.
- Quarter 3 and Quarter 4 was when the net revenue dropped below $10,000,000.
- Quarter 4 had the smallest net revenue and the smallest number of orders.

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

**Solution Query:**

```
SELECT UPPER(credit_card_type),ROUND(AVG(discount),2) AS avg_discount
FROM order_t AS o
LEFT JOIN
customer_t AS c
ON o.customer_id = c.customer_id
GROUP BY credit_card_type
ORDER BY credit_card_type ASC;
```

**Output:**



Query:

```
SELECT UPPER(credit_card_type),ROUND(AVG(discount),2) AS avg_discount
FROM order_t AS o
LEFT JOIN
customer_t AS c
ON o.customer_id = c.customer_id
GROUP BY credit_card_type
ORDER BY avg_discount DESC
```

Output:

Showing first 10 rows out of 16 rows

| UPPER(credit_card_ty... | avg_discount |
|---|---|
| LASER | 0.64 |
| MASTERCARD | 0.63 |
| VISA-ELECTRON | 0.62 |
| MAESTRO | 0.62 |
| INSTAPAYMENT | 0.62 |
| CHINA-UNIONPAY | 0.62 |
| AMERICANEXPRESS | 0.62 |
| SWITCH | 0.61 |
| JCB | 0.61 |

**Observations and Insights:**

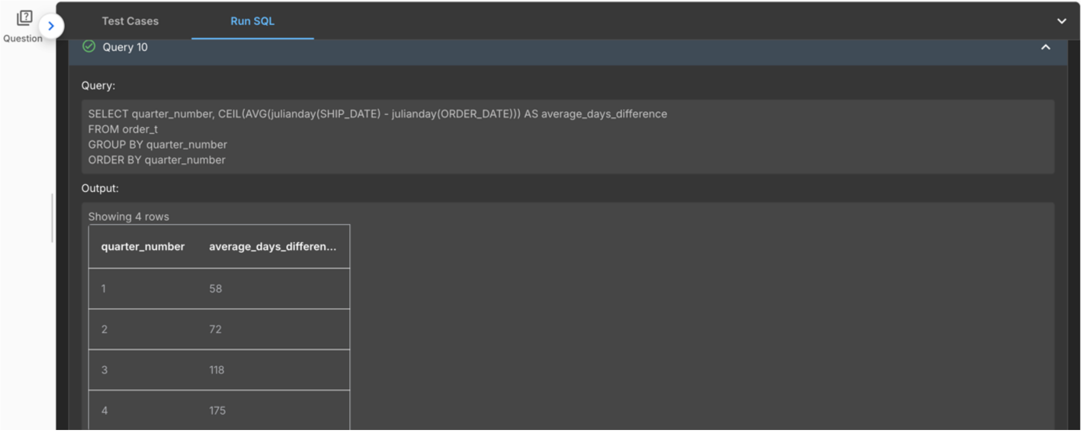
- Laser had the highest average discount with 64%.
- Mastercard had the second highest average discount with 63%.
- Diners Club International had the lowest average discount with 58%

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

**Solution Query:**

```
SELECT quarter_number, CEIL(AVG(julianday(SHIP_DATE) - julianday(ORDER_DATE))) AS
average_days_difference
FROM order_t
GROUP BY quarter_number
ORDER BY quarter_number;
```

**Output:**



Query:

```
SELECT quarter_number, CEIL(AVG(julianday(SHIP_DATE) - julianday(ORDER_DATE))) AS average_days_difference
FROM order_t
GROUP BY quarter_number
ORDER BY quarter_number
```

Output:

Showing 4 rows

| quarter_number | average_days_differen... |
|---|---|
| 1 | 58 |
| 2 | 72 |
| 3 | 118 |
| 4 | 175 |

**Observations and Insights:**

- As the quarter went by, the average days that it took to get the product increased
- Quarter 1 had the best outcome with an average of 58 days.
- The range from Quarter 3 to Quarter 4 was the highest at 57 average days it took for the product to ship after ordering.

# Business Metrics Overview

| Total Revenue | Total Orders | Total Customers | Average Rating |
|---|---|---|---|
| $48,610,993.78 | 1000 | 994 | 3.135 (Okay) |
| **Last Quarter Revenue** | **Last quarter Orders** | **Average Days to Ship** | **% Good Feedback** |
| $8,882,298.84 | 199 | 98 | 20.1% (combining Good and Very Good feedback) |

# Business Recommendations

- Work on getting shipments done in a timely matter.
  - The shipping time increased in days as the quarter went by so focus on more efficient ways to ship (like scout for local areas to create cars)
- Instead of checking per quarter, do a mid-quarter check in with the feedback.
  - This would help get more up to date feedback on what needs to be improved.
- Since Texas and California have the most orders, create some advertisements in the states like Wyoming, Maine, or Vermont where there was one purchase to increase the visibility of the company and offer options like credit card discount to incentive people from the respective states to look into the company for cars.