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.

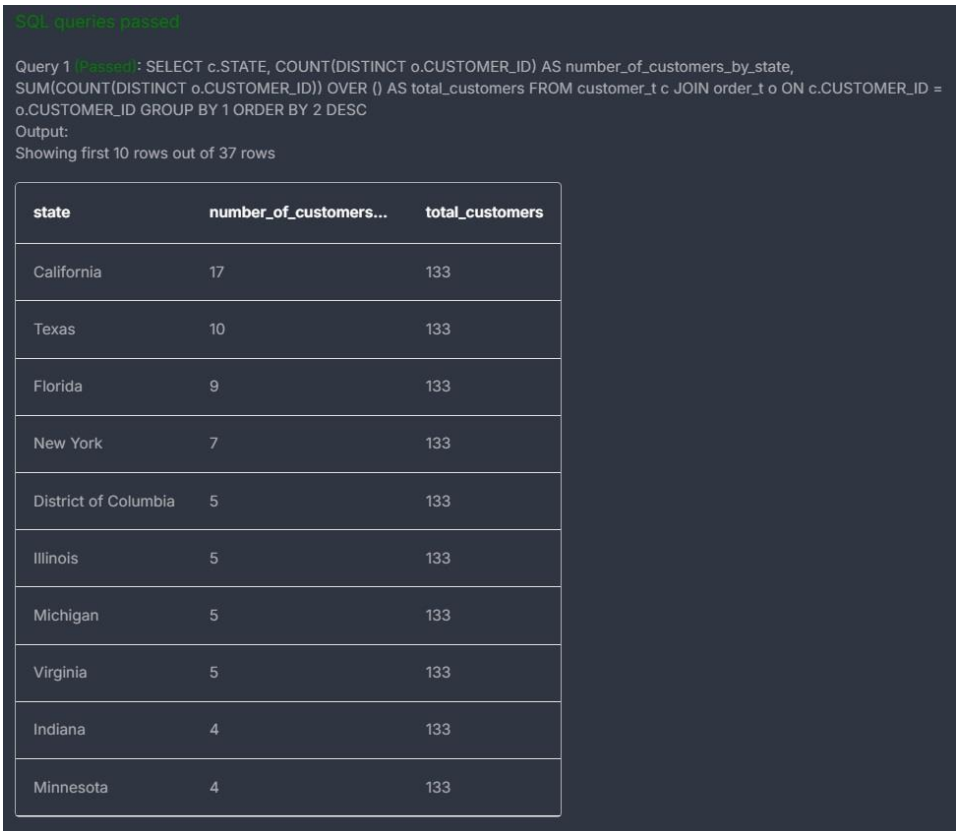
As a data analyst, you see that there is an array of questions that are being asked at the leadership level that need to be answered using data. Import the dump file that contains various tables that are present in the database. Use the data to answer the questions posed and create a quarterly business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

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

Solution Query:

```
SELECT
    c.STATE,
    COUNT(DISTINCT o.CUSTOMER_ID) AS number_of_customers_by_state,
    SUM(COUNT(DISTINCT o.CUSTOMER_ID)) OVER () AS total_customers
FROM
    customer_t c
JOIN
    order_t o
    ON c.CUSTOMER_ID = o.CUSTOMER_ID
GROUP BY
    1
ORDER BY
    2 DESC;
```

Output:



SQL queries passed

Query 1 (Passed): SELECT c.STATE, COUNT(DISTINCT o.CUSTOMER_ID) AS number_of_customers_by_state, SUM(COUNT(DISTINCT o.CUSTOMER_ID)) OVER () AS total_customers FROM customer_t c JOIN order_t o ON c.CUSTOMER_ID = o.CUSTOMER_ID GROUP BY 1 ORDER BY 2 DESC
Output:
Showing first 10 rows out of 37 rows

state	number_of_customers...	total_customers
California	17	133
Texas	10	133
Florida	9	133
New York	7	133
District of Columbia	5	133
Illinois	5	133
Michigan	5	133
Virginia	5	133
Indiana	4	133
Minnesota	4	133

Observations and Insights:

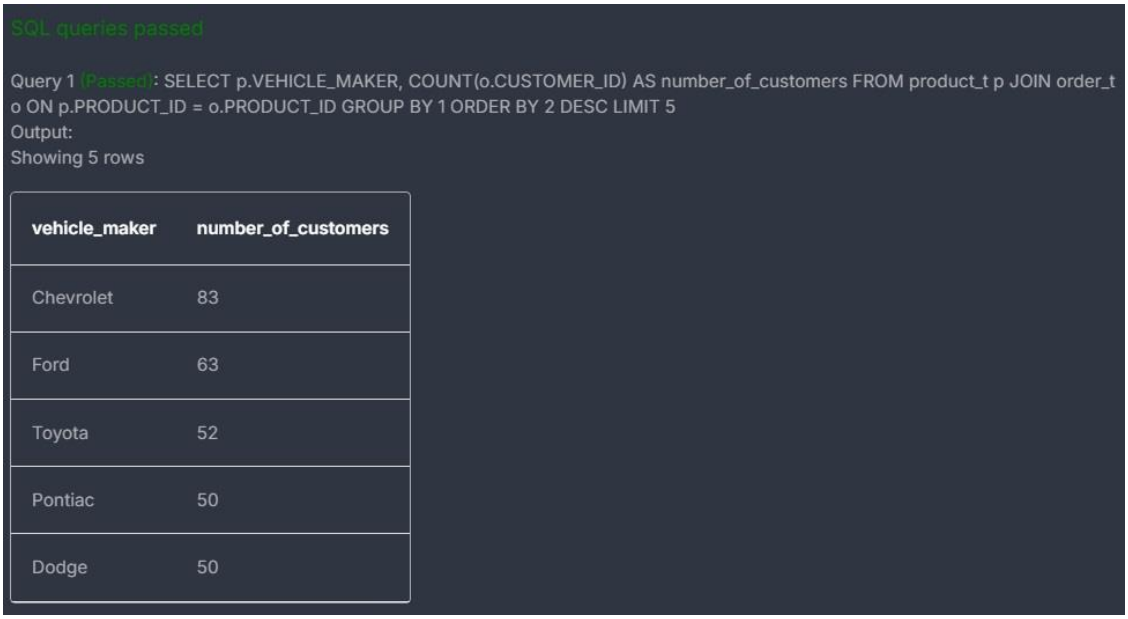
- California (17), Texas (10), and Florida (9) have the highest number of customers.
- States with lower customer count, such as Pennsylvania (4) and Ohio (4), may benefit from better marketing or sales efforts.

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

Solution Query:

```
SELECT
    p.VEHICLE_MAKER,
    COUNT(o.CUSTOMER_ID) AS number_of_customers
FROM
    product_t p
JOIN
    order_t o
    ON p.PRODUCT_ID = o.PRODUCT_ID
GROUP BY
    1
ORDER BY
    2 DESC
LIMIT
    5;
```

Output:


SQL queries passed

Query 1 (Passed): SELECT p.VEHICLE_MAKER, COUNT(o.CUSTOMER_ID) AS number_of_customers FROM product_t p JOIN order_t o ON p.PRODUCT_ID = o.PRODUCT_ID GROUP BY 1 ORDER BY 2 DESC LIMIT 5
Output:
Showing 5 rows

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

Observations and Insights:

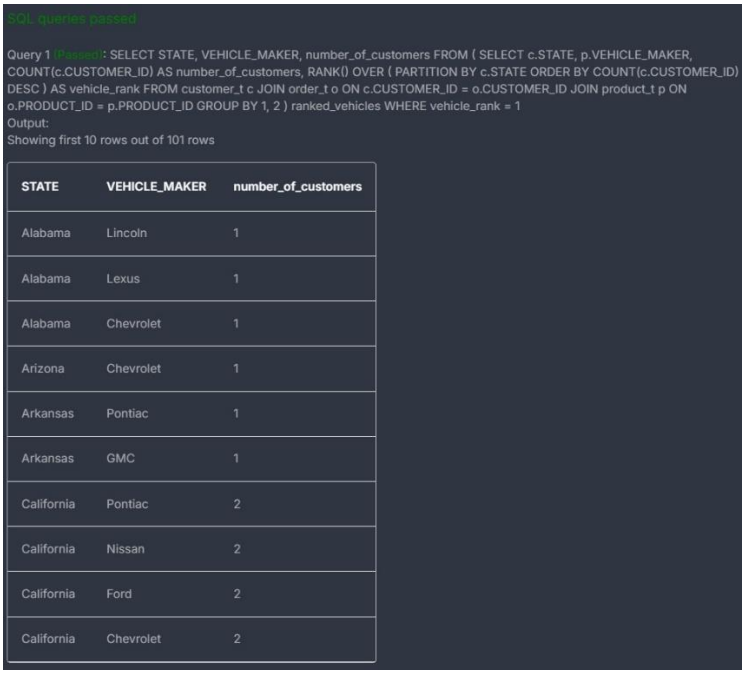
- Chevrolet is the most preferred vehicle maker with 83 customers.
- Pontiac and Dodge are tied in fourth place with 50 customers each.

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

Solution Query:

```
SELECT
    STATE,
    VEHICLE_MAKER,
    number_of_customers
FROM
    (
        SELECT
            c.STATE,
            p.VEHICLE_MAKER,
            COUNT(c.CUSTOMER_ID) AS number_of_customers,
            RANK() OVER (
                PARTITION BY c.STATE
                ORDER BY COUNT(c.CUSTOMER_ID) DESC
            ) AS vehicle_rank
        FROM
            customer_t c
        JOIN
            order_t o
            ON c.CUSTOMER_ID = o.CUSTOMER_ID
        JOIN
            product_t p
            ON o.PRODUCT_ID = p.PRODUCT_ID
        GROUP BY
            1, 2
    ) ranked_vehicles
WHERE
    vehicle_rank = 1;
```

Output:



SQL queries passed

Query 1 (Passed): SELECT STATE, VEHICLE_MAKER, number_of_customers FROM (SELECT c.STATE, p.VEHICLE_MAKER, COUNT(c.CUSTOMER_ID) AS number_of_customers, RANK() OVER (PARTITION BY c.STATE ORDER BY COUNT(c.CUSTOMER_ID) DESC) AS vehicle_rank FROM customer_t c JOIN order_t o ON c.CUSTOMER_ID = o.CUSTOMER_ID JOIN product_t p ON o.PRODUCT_ID = p.PRODUCT_ID GROUP BY 1, 2) ranked_vehicles WHERE vehicle_rank = 1

Output:

Showing first 10 rows out of 101 rows

STATE	VEHICLE_MAKER	number_of_customers
Alabama	Lincoln	1
Alabama	Lexus	1
Alabama	Chevrolet	1
Arizona	Chevrolet	1
Arkansas	Pontiac	1
Arkansas	GMC	1
California	Pontiac	2
California	Nissan	2
California	Ford	2
California	Chevrolet	2

Observations and Insights:

- Alabama has customers that prefer Lincoln, Lexus, and Chevrolet.
- Arkansas has customers that prefer Pontiac and GMC.
- Chevrolet appears to be the more consistent brand name among the vehicle makers that rank number 1.

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:

```
SELECT
    QUARTER_NUMBER,
    ROUND( AVG(rating) , 2) AS average_feedback_per_quarter,
    ROUND( AVG(rating) OVER(), 2) AS overall_average
FROM
    (
        SELECT
            QUARTER_NUMBER,
            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 rating
        FROM
            order_t
    ) feedback_table
GROUP BY
    1
ORDER BY
    1;
```

Output:

SQL queries passed

Query 1 (Passed): SELECT QUARTER_NUMBER, ROUND(AVG(rating) , 2) AS average_feedback_per_quarter, ROUND(AVG(rating) OVER(), 2) AS overall_average FROM (SELECT QUARTER_NUMBER, 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 rating FROM order_t) feedback_table GROUP BY 1 ORDER BY 1

Output:

Showing 4 rows

QUARTER_NUMBER	average_feedback_per...	overall_average
1	3.55	2.75
2	3.35	2.75
3	2.96	2.75
4	2.4	2.75

Observations and Insights:

- The overall average customer rating is 2.75.
- Quarter 1 has the highest feedback rating (3.55), compared to Quarter 4, which has the lowest feedback rating (2.4). This indicates a growing dissatisfaction with the company over time.

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

Solution Query:

```
SELECT
    QUARTER_NUMBER,
    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
    1
ORDER BY
    1;
```

Output:

SQL queries passed

Query 1 (Passed): SELECT QUARTER_NUMBER, 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 1 ORDER BY 1

Output:

Showing 4 rows

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

Observations and Insights:

- The percentage of “Very Bad” feedback increases from Quarter 1 (10.97%) to Quarter 4 (30.65%).
- This trend indicates problems with product quality, service delays, or unmet customer expectations.

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

Solution Query:

```
SELECT
    QUARTER_NUMBER,
    COUNT(ORDER_ID) AS number_of_orders
FROM
    order_t
GROUP BY
    1
ORDER BY
    1;
```

Output:

SQL queries passed

Query 1 (Passed): SELECT QUARTER_NUMBER, COUNT(ORDER_ID) AS number_of_orders FROM order_t GROUP BY 1 ORDER BY 1
Output:
Showing 4 rows

quarter_number	number_of_orders
1	310
2	262
3	229
4	199

Observations and Insights:

- The number of orders decreases each quarter, starting with Quarter 1 (310), and ending with Quarter 4 (199).
- Investigation into whether the downward trend has been brought on by internal or external factors is recommended.

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

Solution Query:

```
SELECT
    QUARTER_NUMBER,
    ROUND(net_revenue, 2) AS net_revenue,
    ROUND(LAG(net_revenue) OVER (ORDER BY QUARTER_NUMBER), 2)
        AS previous_quarter_revenue,
    CASE
        WHEN LAG(net_revenue) OVER (ORDER BY QUARTER_NUMBER) IS NOT NULL
        THEN ROUND(((net_revenue - LAG(net_revenue)
            OVER (ORDER BY QUARTER_NUMBER))
            / LAG(net_revenue) OVER (ORDER BY QUARTER_NUMBER)) * 100, 2)
        ELSE NULL
    END AS percentage_change
FROM
    (
        SELECT
            QUARTER_NUMBER,
            SUM(QUANTITY * (VEHICLE_PRICE * (1 - DISCOUNT)))
                AS net_revenue
        FROM
            order_t
        GROUP BY
            1
    ) revenue
ORDER BY
    1;
```

Output:

SQL queries passed

Query 1 (Passed): SELECT QUARTER_NUMBER, ROUND(net_revenue, 2) AS net_revenue, ROUND(LAG(net_revenue) OVER (ORDER BY QUARTER_NUMBER), 2) AS previous_quarter_revenue, CASE WHEN LAG(net_revenue) OVER (ORDER BY QUARTER_NUMBER) IS NOT NULL THEN ROUND(((net_revenue - LAG(net_revenue) OVER (ORDER BY QUARTER_NUMBER)) / LAG(net_revenue) OVER (ORDER BY QUARTER_NUMBER)) * 100, 2) ELSE NULL END AS percentage_change FROM (SELECT QUARTER_NUMBER, SUM(QUANTITY * (VEHICLE_PRICE * (1 - DISCOUNT))) AS net_revenue FROM order_t GROUP BY 1) revenue ORDER BY 1
Output:
Showing 4 rows

QUARTER_NUMBER	net_revenue	previous_quarter_reve...	percentage_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 biggest percentage drops occur in Quarter 2 and Quarter 3, -27.23% and -32.32% respectively.
- It seems to stabilize, somewhat, in Quarter 4 (-3.48%).
- This shows that efforts were made in Quarter 3 to counter the steep downward trend in revenue.

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

Solution Query:

```
SELECT
    QUARTER_NUMBER,
    COUNT(ORDER_ID) AS number_of_orders,
    ROUND(SUM(QUANTITY * (VEHICLE_PRICE * (1 - DISCOUNT))), 2) AS net_revenue
FROM
    order_t
GROUP BY
    1;
```

Output:

SQL queries passed

Query 1 (Passed): SELECT QUARTER_NUMBER, COUNT(ORDER_ID) AS number_of_orders, ROUND(SUM(QUANTITY * (VEHICLE_PRICE * (1 - DISCOUNT))), 2) AS net_revenue FROM order_t GROUP BY 1
Output:
Showing 4 rows

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

Observations and Insights:

- The number of orders and net revenue are both consistently decreasing quarter to quarter.
- This shows that a decrease in customer demand is the primary cause of the drop in revenue.

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

Solution Query:

```
SELECT
    c.CREDIT_CARD_TYPE,
    ROUND(AVG(DISCOUNT), 2) AS average_discount
FROM
    order_t o
JOIN
    customer_t c
    ON o.CUSTOMER_ID = c.CUSTOMER_ID
GROUP BY
    1
ORDER BY
    2 DESC;
```

Output:

SQL queries passed

Query 1 (Passed): SELECT c.CREDIT_CARD_TYPE, ROUND(AVG(DISCOUNT), 2) AS average_discount FROM order_t o JOIN customer_t c ON o.CUSTOMER_ID = c.CUSTOMER_ID GROUP BY 1 ORDER BY 2 DESC
Output:
Showing first 10 rows out of 16 rows

credit_card_type	average_discount
instapayment	0.77
solo	0.7
americanexpress	0.68
diners-club-enroute	0.67
mastercard	0.65
diners-club-carte-blan...	0.65
visa-electron	0.64
maestro	0.64
laser	0.62
china-unionpay	0.62

Observations and Insights:

- The highest average discount is offered to customers using Insta Payment (0.77%).
- Solo (0.7%), American Express (0.68%), and Diner's Club Enroute (0.67%), come in at second, third, and fourth place, respectively.

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

Solution Query:

```
SELECT
    QUARTER_NUMBER,
    ROUND(AVG(JULIANDAY(SHIP_DATE) - JULIANDAY(ORDER_DATE)), 0)
    AS average_shipping_days
FROM
    order_t
GROUP BY
    1;
```

Output:

SQL queries passed

Query 1 (Passed): SELECT QUARTER_NUMBER, ROUND(AVG(JULIANDAY(SHIP_DATE) - JULIANDAY(ORDER_DATE)), 0) AS average_shipping_days FROM order_t GROUP BY 1
Output:
Showing 4 rows

quarter_number	average_shipping_days
1	57
2	71
3	118
4	174

Observations and Insights:

- There is an increase in average shipping days from 57 days in Quarter 1 to 174 days in Quarter 4.
- The increased shipping time appears to be a contributing factor to the decrease in customer satisfaction.

Business Metrics Overview

Total Revenue	Total Orders	Total Customers	Average Rating
$48,611,933.78	1000	133	2.75
Last Quarter Revenue	**Last quarter Orders**	**Average Days to Ship**	**% Good Feedback**
$8,573,149.28	199	105	20.46%

Business Recommendations

- Work on getting the average shipping time back into the range of 50 to 60 days. With this improvement, net revenue should increase accordingly.
- Conduct customer surveys to gather additional feedback to understand the reasons behind the drop in satisfaction. Address key issues such as product quality, customer service response times, or communication gaps.
- Re-engage with customers who haven't placed an order recently. Offer limited-time discounts or promotions to encourage them to return.