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

Solution Query:

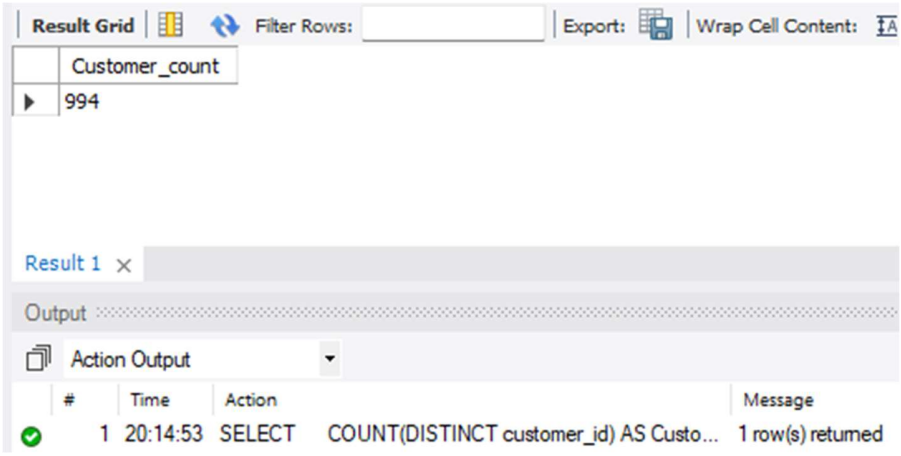
- **Query to calculate the total number of customers who have placed orders**

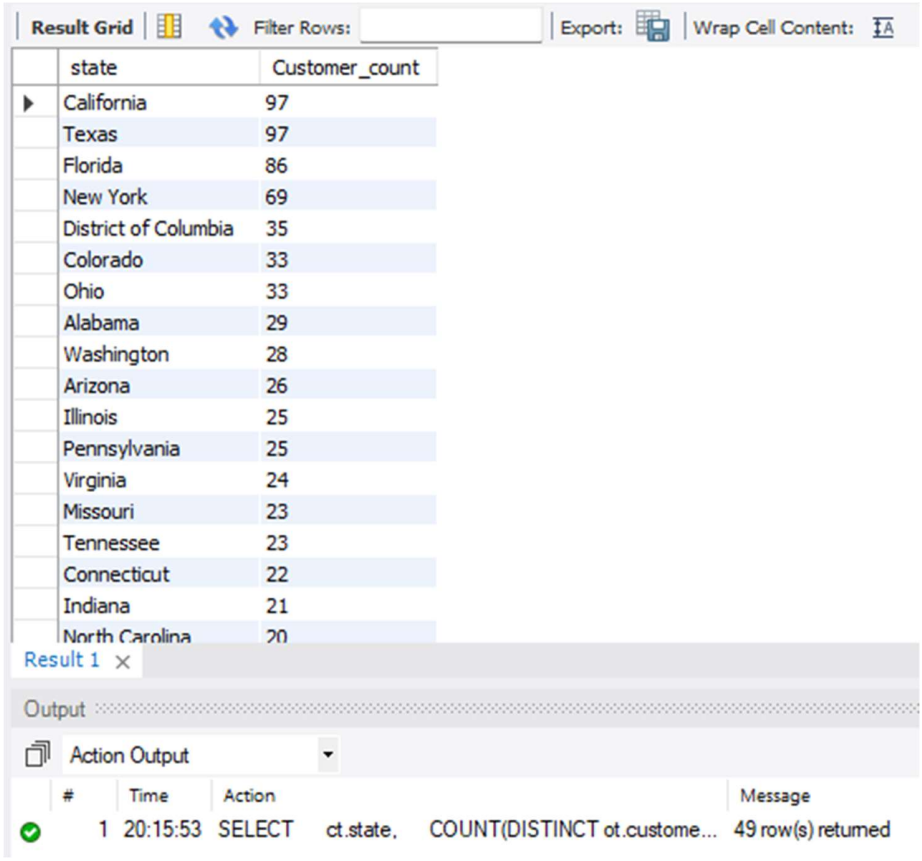
```
SELECT
    COUNT(DISTINCT customer_id) AS Customer_count
FROM
    order_t;
```

- **Query to calculate the distribution of the customers across states**

```
SELECT
    ct.state,
    COUNT(DISTINCT ot.customer_id) AS Customer_count
FROM
    customer_t AS ct
INNER JOIN
    order_t AS ot ON ct.customer_id = ot.customer_id
GROUP BY
    ct.state
ORDER BY
    Customer_count DESC;
```

Output:





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

state	Customer_count
California	97
Texas	97
Florida	86
New York	69
District of Columbia	35
Colorado	33
Ohio	33
Alabama	29
Washington	28
Arizona	26
Illinois	25
Pennsylvania	25
Virginia	24
Missouri	23
Tennessee	23
Connecticut	22
Indiana	21
North Carolina	20

Result 1

Output

Action Output

#	Time	Action			Message
1	20:15:53	SELECT	ct.state,	COUNT(DISTINCT ot.custome...	49 row(s) returned

Observations and Insights:

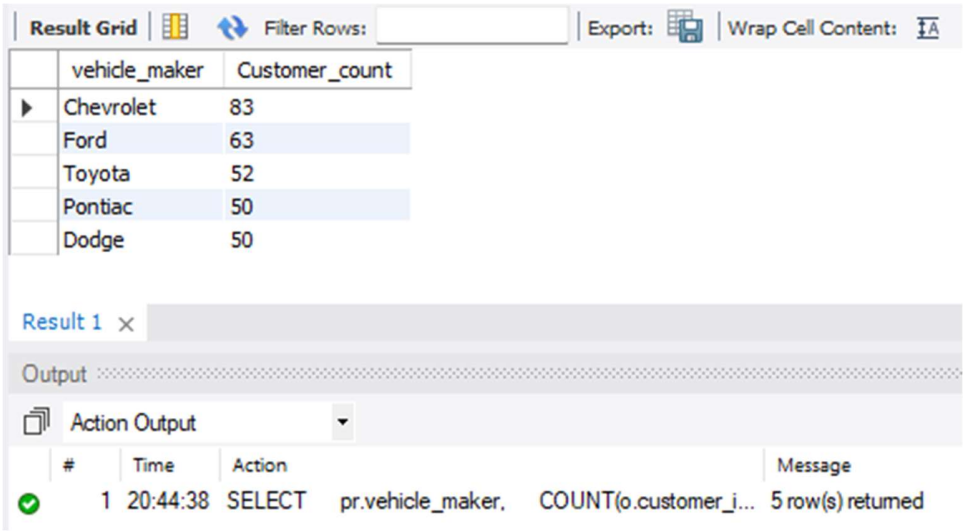
- Total number of customers who have placed orders: 994
- California and Texas have the highest count of customers of 97, followed by Florida and New York.
- Maine, Vermont and Wyoming have the lowest customer count.

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

Solution Query:

```
SELECT
    pr.vehicle_maker,
    COUNT(o.customer_id) AS Customer_count
FROM
    product_t AS pr
INNER JOIN
    order_t AS o
    ON o.product_id = pr.product_id
GROUP BY
    pr.vehicle_maker
ORDER BY
    Customer_count DESC
LIMIT 5;
```

Output:



Observations and Insights:

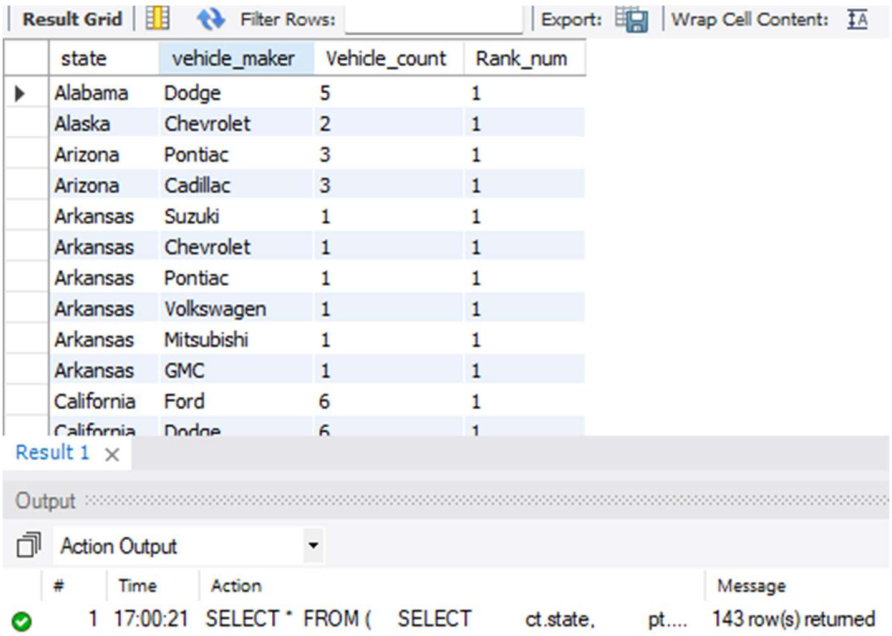
- Chevrolet was the most preferred vehicle maker with the highest customer count of 83.
- It is followed by Ford (63) and Toyota (52).
- Top 5 vehicle makers: Chevrolet, Ford, Toyota, Pontiac, Dodge.

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

Solution Query:

```
SELECT *
FROM (
    SELECT
        ct.state,
        pt.vehicle_maker,
        COUNT(ot.customer_id) AS Vehicle_count,
        RANK() OVER (PARTITION BY ct.state ORDER BY COUNT(ot.customer_id)
         DESC) AS Rank_num
    FROM
        product_t AS pt
    INNER JOIN
        order_t AS ot ON ot.product_id = pt.product_id
    INNER JOIN
        customer_t AS ct ON ct.customer_id = ot.customer_id
    GROUP BY
        ct.state, pt.vehicle_maker
) AS t
WHERE Rank_num = 1;
```

Output:



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

state	vehicle_maker	Vehicle_count	Rank_num
Alabama	Dodge	5	1
Alaska	Chevrolet	2	1
Arizona	Pontiac	3	1
Arizona	Cadillac	3	1
Arkansas	Suzuki	1	1
Arkansas	Chevrolet	1	1
Arkansas	Pontiac	1	1
Arkansas	Volkswagen	1	1
Arkansas	Mitsubishi	1	1
Arkansas	GMC	1	1
California	Ford	6	1
California	Dodge	6	1

Result 1

Output

Action Output

#	Time	Action	Message
1	17:00:21	SELECT * FROM (SELECT ct.state, pt....	143 row(s) returned

Observations and Insights:

- Chevrolet consistently maintains as the most preferred vehicle maker across many states.
- In majority of the states, a variety of vehicle makers tops the list.

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:

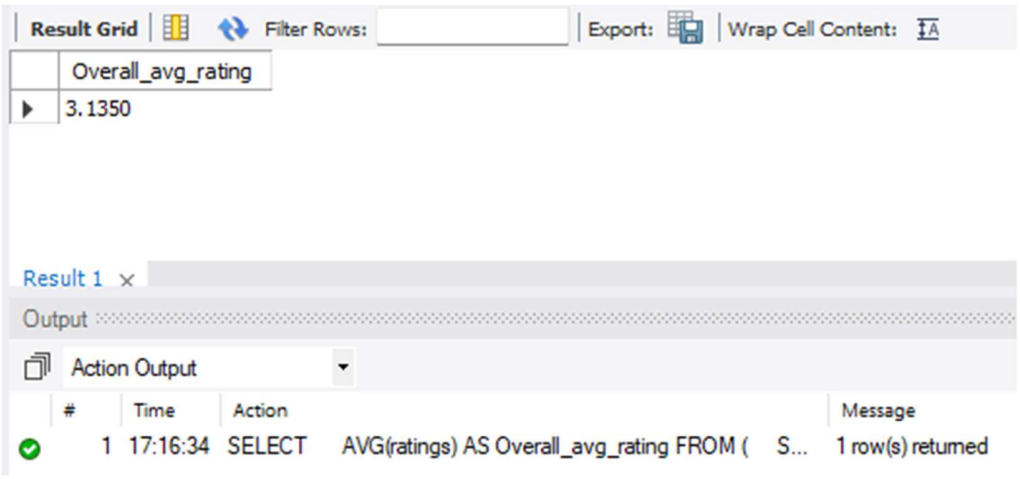
- **Query to calculate overall average rating**

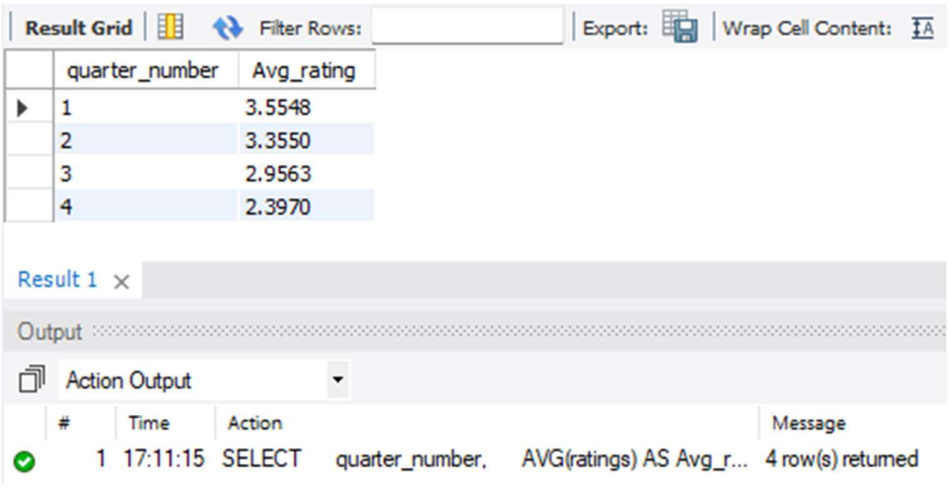
```
SELECT
    AVG(ratings) AS Overall_avg_rating
FROM (
    SELECT
        quarter_number,
        CASE
            WHEN customer_feedback = 'Very Bad' THEN 1
            WHEN customer_feedback = 'Bad' THEN 2
            WHEN customer_feedback = 'Okay' THEN 3
            WHEN customer_feedback = 'Good'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

- **Query to calculate average rating in each quarter**

```
SELECT
    quarter_number,
    AVG(ratings) AS Avg_rating
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 ratings
    FROM
        order_t
) AS subquery
GROUP BY
    quarter_number
ORDER BY
    quarter_number;
```

Output:





Observations and Insights:

- Overall Average rating given by the customers: 3.135
- The 1st quarter has the highest avg rating of 3.55 compared to other quarters.
- The Average ratings of customers clearly indicate a declining trend across the quarters.
- The 4th quarter has the lowest average rating of 2.39.

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

Solution Query:

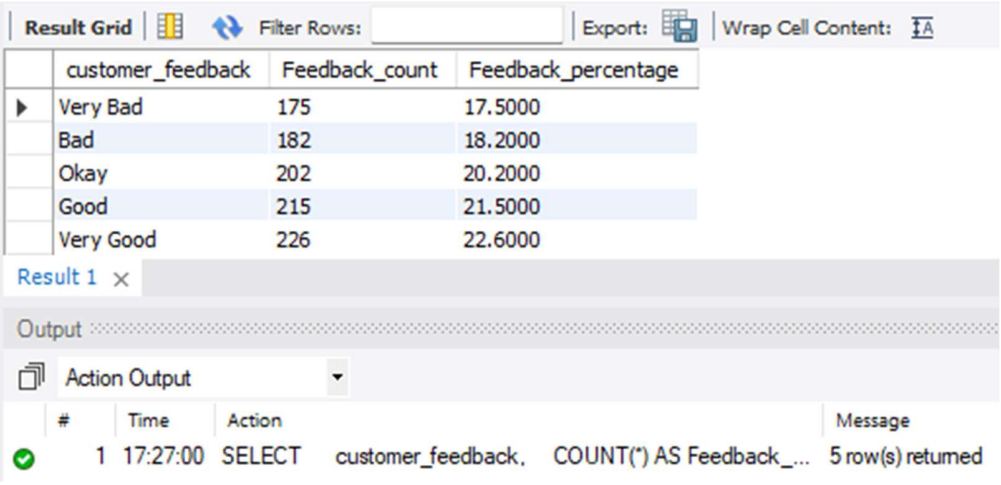
- **Query to calculate overall percentage distribution of customer feedback**

```
SELECT
    customer_feedback,
    COUNT(*) AS Feedback_count,
    COUNT(*) * 100 / SUM(COUNT(*)) OVER () AS Feedback_percentage
FROM
    order_t
GROUP BY
    customer_feedback;
```

- **Query to calculate percentage distribution of customers feedback in each quarter**

```
SELECT
    quarter_number,
    (SUM(CASE WHEN customer_feedback = 'Very Bad' THEN 1 ELSE 0 END) /
     COUNT(customer_feedback)) * 100.0 AS Very_bad_percentage,
    (SUM(CASE WHEN customer_feedback = 'Good' THEN 1 ELSE 0 END) /
     COUNT(customer_feedback)) * 100.0 AS Good_percentage,
    (SUM(CASE WHEN customer_feedback = 'Okay' THEN 1 ELSE 0 END) /
     COUNT(customer_feedback)) * 100.0 AS Okay_percentage,
    (SUM(CASE WHEN customer_feedback = 'Bad' THEN 1 ELSE 0 END) /
     COUNT(customer_feedback)) * 100.0 AS Bad_percentage,
    (SUM(CASE WHEN customer_feedback = 'Very Good' THEN 1 ELSE 0 END) /
     COUNT(customer_feedback)) * 100.0 AS Very_good_percentage
FROM
    order_t
GROUP BY
    quarter_number
ORDER BY
    quarter_number;
```

Output:



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

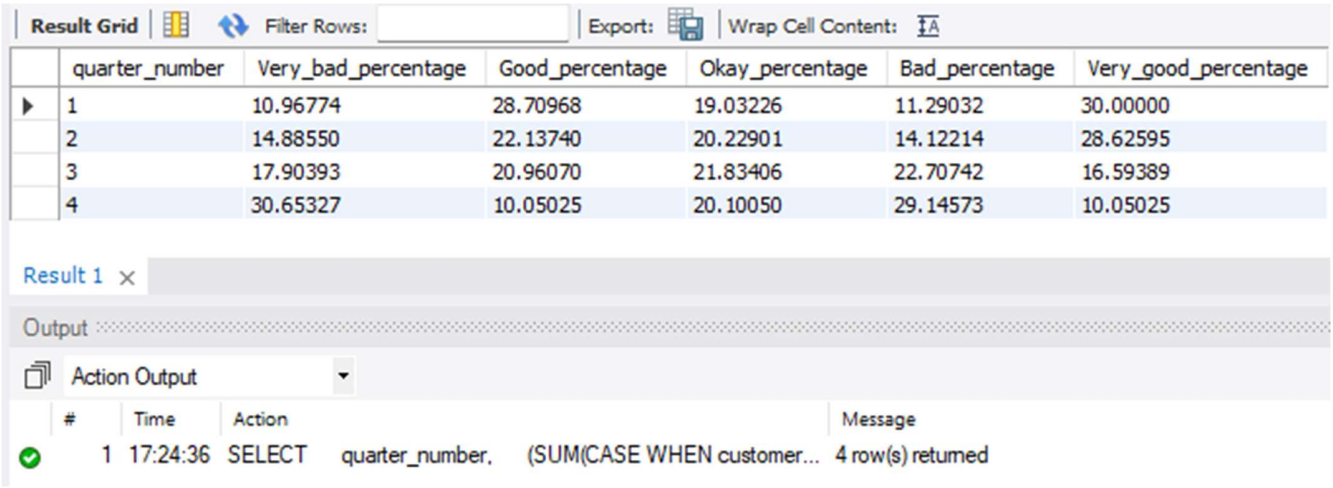
customer_feedback	Feedback_count	Feedback_percentage
Very Bad	175	17.5000
Bad	182	18.2000
Okay	202	20.2000
Good	215	21.5000
Very Good	226	22.6000

Result 1

Output

Action Output

#	Time	Action	Message
1	17:27:00	SELECT customer_feedback, COUNT(*) AS Feedback_...	5 row(s) returned



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

quarter_number	Very_bad_percentage	Good_percentage	Okay_percentage	Bad_percentage	Very_good_percentage
1	10.96774	28.70968	19.03226	11.29032	30.00000
2	14.88550	22.13740	20.22901	14.12214	28.62595
3	17.90393	20.96070	21.83406	22.70742	16.59389
4	30.65327	10.05025	20.10050	29.14573	10.05025

Result 1

Output

Action Output

#	Time	Action	Message
1	17:24:36	SELECT quarter_number, (SUM(CASE WHEN customer...	4 row(s) returned

Observations and Insights:

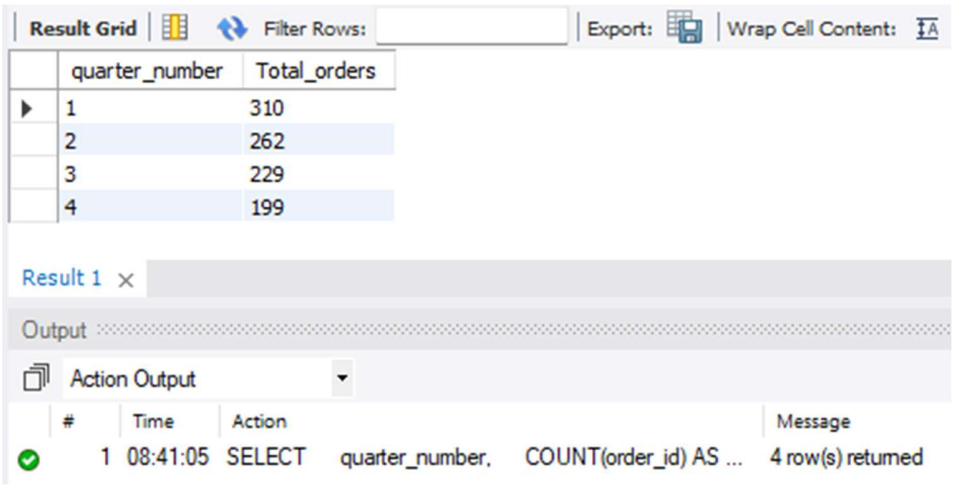
- Majority of the customers gave “Very Good” feedback (226).
- 44.1% of the customers gave good feedback (“Very Good” and “Good”)
- Feedback was drastically reduced over the quarters and customers are getting dissatisfied over the time.
- “Very Bad” feedback was increased from 10% to 30%.
- 1st quarter was the best performing quarter and 4th quarter was the worst performing quarter.

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

Solution Query:

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

Output:



Observations and Insights:

- Order count was significantly reduced from 1st quarter to 4th quarter which indicates a declining trend.
- 1st quarter have the best overall feedback and highest order count.
- 4th quarter have the worst overall feedback and lowest order count.

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

Solution Query:

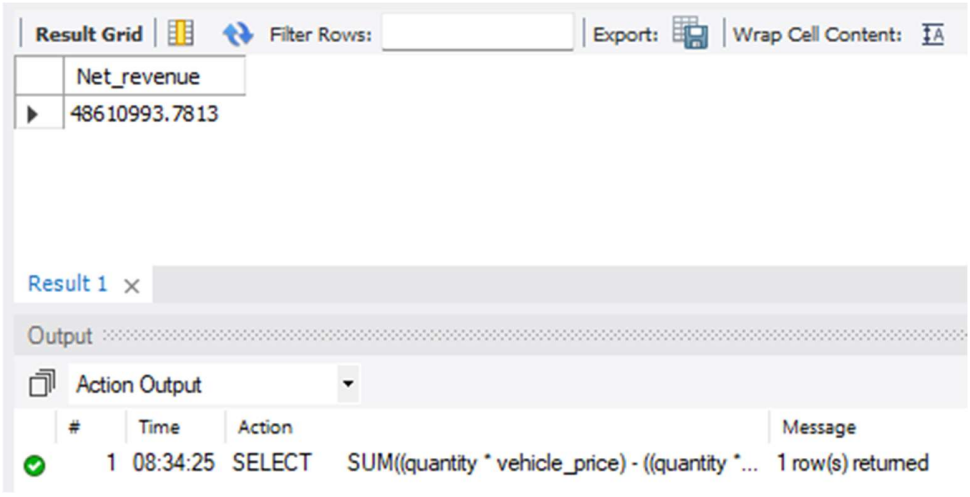
- **Query to calculate total net revenue generated by the company**

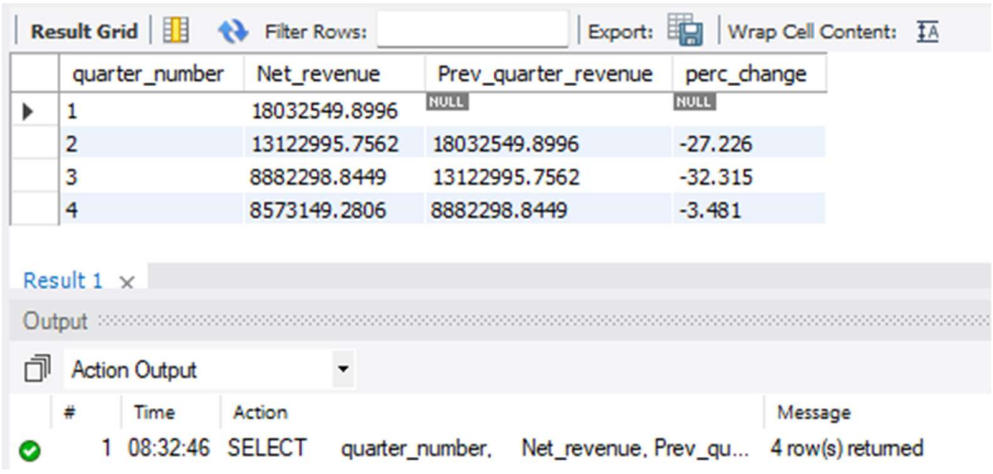
```
SELECT
    SUM((quantity * vehicle_price) - ((quantity * vehicle_price) *
    discount)) AS Net_revenue
FROM
    order_t;
```

- **Query to calculate quarter-over-quarter % change in net revenue**

```
SELECT
    quarter_number,
    Net_revenue,
    Prev_quarter_revenue,
    ROUND(((Net_revenue - prev_quarter_revenue) / prev_quarter_revenue) *
    100,3) AS perc_change
FROM (
    SELECT
        quarter_number,
        SUM((quantity * vehicle_price) - ((quantity * vehicle_price) *
        discount)) AS Net_revenue,
        LAG(SUM((quantity * vehicle_price) - ((quantity * vehicle_price) *
        discount)))
            OVER (ORDER BY quarter_number) AS Prev_quarter_revenue
    FROM
        order_t
    GROUP BY
        quarter_number
    ORDER BY
        quarter_number
) AS sub_query;
```

Output:





Observations and Insights:

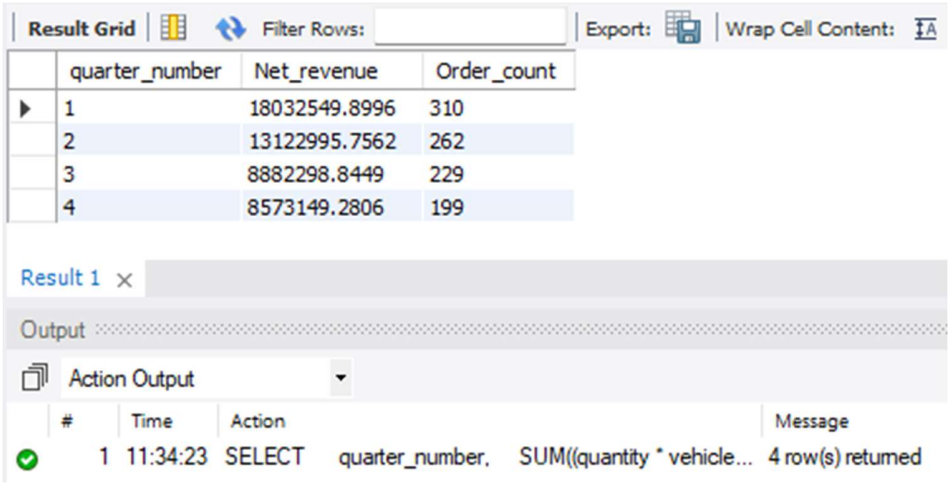
- Net-Revenue generated by the company: 48.61 million
- Net revenue was significantly reduced over the quarters.
- There was a sharp decline in net revenue from the 2nd quarter to 3rd quarter by -32.3%
- The decline in net revenue is slowed in 4th quarter with only -3.48%.

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

Solution Query:

```
SELECT
    quarter_number,
    SUM((quantity * vehicle_price)
    - ((quantity * vehicle_price) * discount)) AS Net_revenue,
    COUNT(order_id) AS Order_count
FROM
    order_t
GROUP BY
    quarter_number
ORDER BY
    quarter_number;
```

Output:



Observations and Insights:

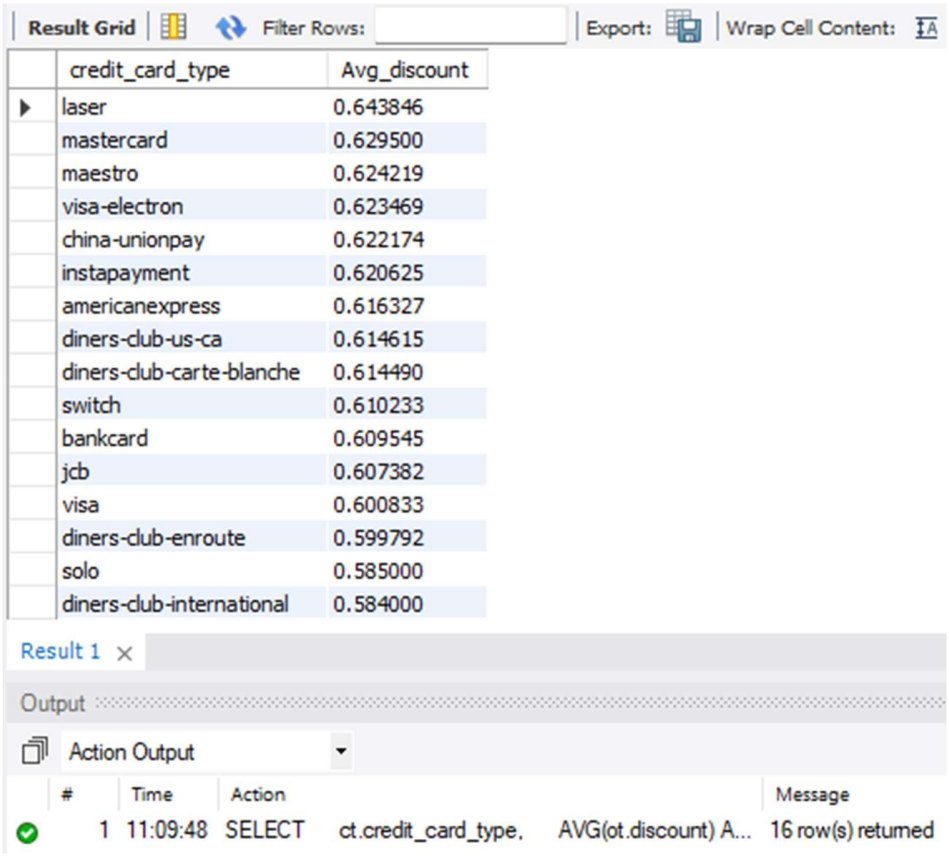
- Both the Net revenue and Order counts indicate a declining trend.
- The Net revenue is significantly reduced from 18 million to 8.5 million.
- The Order count is significantly reduced from 310 to 199.

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

Solution Query:

```
SELECT
    ct.credit_card_type,
    AVG(ot.discount) AS Avg_discount
FROM
    order_t AS ot
INNER JOIN
    customer_t AS ct
    ON ct.customer_id = ot.customer_id
GROUP BY
    ct.credit_card_type
ORDER BY
    avg_discount DESC;
```

Output:



Observations and Insights:

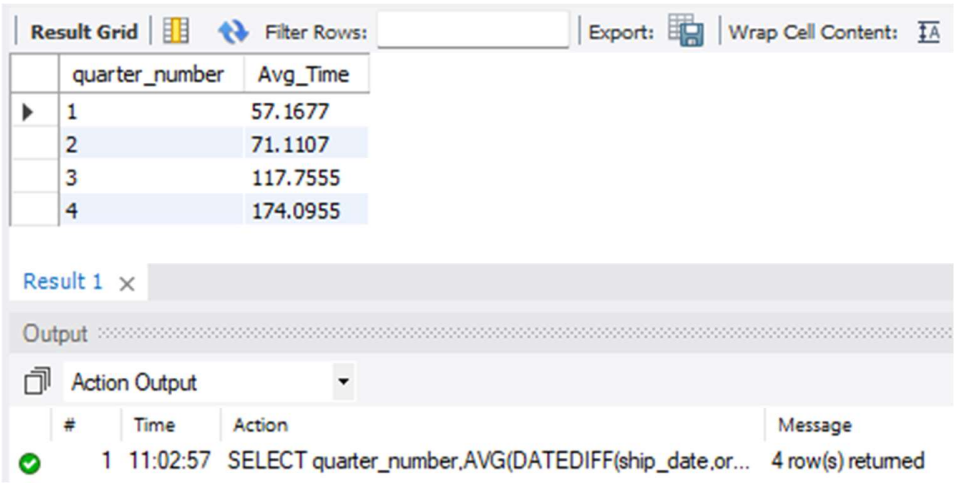
- There are 16 various types of credit cards in the dataset.
- "Laser" cards offer the highest average discount of 0.64, which is the best option for customers.
- "Diners-club-international" offer the lowest discount of 0.58

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

Solution Query:

```
SELECT
    quarter_number,
    AVG(DATEDIFF(ship_date, order_date)) AS Avg_Time
FROM
    order_t
GROUP BY
    quarter_number
ORDER BY
    quarter_number;
```

Output:



Observations and Insights:

- Average time taken to ship the placed orders increases over each quarter.
- The average duration for shipping orders increased from 57 days to 174 days over the time.
- 1st quarter has the lowest shipping time while the 4th quarter has the highest shipping time.

Business Metrics Overview

Total Revenue	Total Orders	Total Customers	Average Rating
48610993.7813	1000	994	3.1350
Last Quarter Revenue	**Last quarter Orders**	**Average Days to Ship**	**% Good Feedback**
8573149.2806	199	Approx. 98 days	21.5% (Good) 22.6% (Very Good)

Business Recommendations

- As we can clearly observe, average feedback was significantly reduced over the time. Surveys should be conducted and negative feedbacks should be addressed.
- The fourth quarter was the worst performing quarter in terms of feedback and sales. The root-cause of revenue decline should be analyzed.
- The Average shipping time should be reduced by partnering with quicker logistics services.
- By introducing loyalty services and offer programs for repeating customers.
- Marketing campaigns and strategies should be implemented to boost the revenue and orders in underperforming states like Maine, Vermont and Wyoming.
- Offering incentives and proper support for customers who reported negative feedback and understand the cause behind the reviews.
- Pricing strategies can be adjusted and discounts can be provided in the 4^{th} quarter to boost the net revenue and order counts.
- Promoting the business to a wider audience in last 2 quarters can increase the revenue.
- Partnering up with more credit card companies can improve the customer base.