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:

***Total number of customers who have placed orders:**

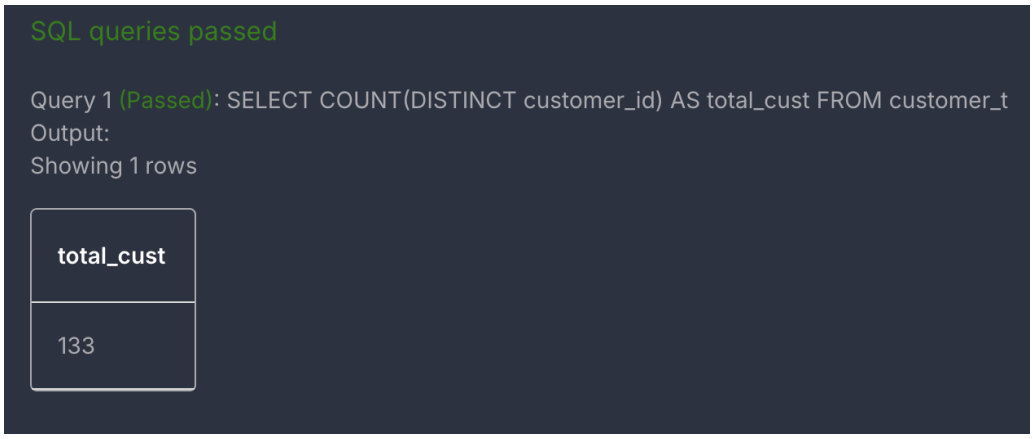
```
SELECT COUNT(DISTINCT customer_id) AS total_cust
FROM customer_t;
```

***Distribution of customers across states:**

```
SELECT COUNT(customer_id) AS total_cust, state
FROM customer_t
GROUP BY state
ORDER BY total_cust DESC;
```

Output:

***Total number of customers who have placed orders:**



***Distribution of customers across states:**

SQL queries passed

Query 1 (Passed): SELECT COUNT(customer_id) AS total_cust, state FROM customer_t GROUP BY state ORDER BY total_cust DESC
Output:
Showing first 10 rows out of 37 rows

total_cust	state
17	California
10	Texas
9	Florida
7	New York
5	Virginia
5	Michigan
5	Illinois
5	District of Columbia
4	Pennsylvania
4	Ohio

*The query below is the exact same query as the one above. The only difference is the total number of customers is in ascending order to show the states with the least amount of customers first.

SQL queries passed

Query 1 (Passed): SELECT COUNT(customer_id) AS total_cust, state FROM customer_t GROUP BY state ORDER BY total_cust ASC
Output:
Showing first 10 rows out of 37 rows

total_cust	state
1	Arizona
1	Idaho
1	Nebraska
1	New Hampshire
1	New Jersey
1	North Dakota
1	Oklahoma
1	West Virginia
2	Arkansas
2	Massachusetts

Observations and Insights:

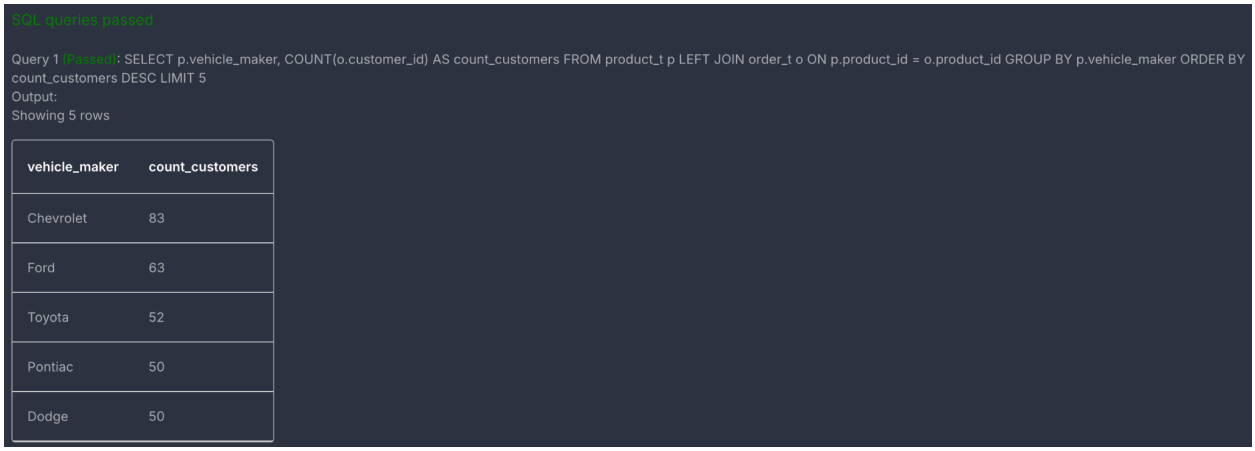
- There's a direct relationship between the number of customers and the population of the state.
- In this case, the more populated the state is, the more customers it has. For example, California, Texas, Florida and New York (the highest populated states in the USA) have the highest number of customers.
- When we order the query to reveal the states with the least amount of customers first, Arizona, Idaho, Nebraska, New Hampshire, New Jersey, North Dakota, Oklahoma and West Virginia come up. In general, these states are less populated and tend to have more rural spaces than the states with more customers.
- Generally, states containing higher population and more urban areas will generate a higher clientele.

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

Solution Query:

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

Output:


SQL queries passed

Query 1 (Passed): SELECT p.vehicle_maker, COUNT(o.customer_id) AS count_customers FROM product_t p LEFT JOIN order_t o ON p.product_id = o.product_id GROUP BY p.vehicle_maker ORDER BY count_customers DESC LIMIT 5
Output:
Showing 5 rows

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

Observations and Insights:

- The top five vehicle makers preferred by the customers are Chevrolet, Ford, Toyota, Pontiac and Dodge.
- All of these are esteemed vehicle makers, which can indicate that customers prefer to buy popular brands with a distinguished trajectory.

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

Solution Query:

```
SELECT *
FROM
  (
    SELECT c.state, p.vehicle_maker, RANK() OVER (PARTITION BY c.state ORDER
           BY COUNT(c.customer_id) DESC) AS rnk
    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 c.state, p.vehicle_maker
  )
  AS table1
WHERE rnk = 1;
```

Output:

SQL queries passed

Query 1 (Passed): SELECT * FROM(SELECT c.state, p.vehicle_maker, RANK() OVER (PARTITION BY c.state ORDER BY COUNT(c.customer_id) DESC) AS rnk 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 c.state, p.vehicle_maker) AS table1 WHERE rnk = 1
Output:
Showing first 10 rows out of 101 rows

state	vehicle_maker	rnk
Alabama	Lincoln	1
Alabama	Lexus	1
Alabama	Chevrolet	1
Arizona	Chevrolet	1
Arkansas	Pontiac	1
Arkansas	GMC	1
California	Pontiac	1
California	Nissan	1
California	Ford	1
California	Chevrolet	1

Observations and Insights:

- Chevrolet appears to be one of the preferred vehicle makers across states like Alabama, Arizona and California. These states vary in population size and provide a variety of rural or urban areas.
- However, Lincoln, Lexus, Pontiac, Nissan and Ford also rank #1 in Alabama and California.
- These results imply that no one vehicle maker brand presides in popularity.

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:

***Overall rating given by customers:**

```
SELECT
  ROUND(AVG(rating), 2) 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 rating
    FROM order_t
  )
  AS table1;
```

***Average rating in each quarter:**

```
SELECT
  quarter_number, ROUND(AVG(rating), 2) 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 rating
    FROM order_t
  )
  AS table1
GROUP BY quarter_number;
```

Output:

***Overall rating given by customers:**

SQL queries passed

Query 1 (Passed): SELECT ROUND(AVG(rating), 2) 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 rating FROM order_t) AS table1
Output:
Showing 1 rows

avg_rating
3.14

***Average rating in each quarter:**

SQL queries passed

Query 1 (Passed): SELECT quarter_number, ROUND(AVG(rating),2) 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 rating FROM order_t) AS table1 GROUP BY quarter_number
Output:
Showing 4 rows

quarter_number	avg_rating
1	3.55
2	3.35
3	2.96
4	2.4

Observations and Insights:

- The average rating in quarter 1 is 3.55 (Good/Okay), 3.35 (Okay) in quarter 2, 2.96 (Okay) in quarter 3 and 2.4 (Bad) in quarter 4.
- We can see that as time progresses the average rating declines. We started with an average rating of 3.55 in quarter 1 and by the end of quarter 4 the average rating ends at 2.4.
- This trend suggests that customer satisfaction seems to decline with the progression of time.
- New Wheels needs to investigate possible events (poor customer service, higher costs, low product quality, etc.) that transpired after quarter 1 that are making customers more dissatisfied 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(100.0 * SUM(CASE WHEN customer_feedback = 'Very Bad' THEN 1 ELSE 0
      END) / COUNT(customer_feedback), 2) AS very_bad_percentage,
      ROUND(100.0 * SUM(CASE WHEN customer_feedback = 'Bad' THEN 1 ELSE 0 END) /
      COUNT(customer_feedback), 2) AS bad_percentage,
      ROUND(100.0 * SUM(CASE WHEN customer_feedback = 'Okay' THEN 1 ELSE 0 END) /
      COUNT(customer_feedback), 2) AS okay_percentage,
      ROUND(100.0 * SUM(CASE WHEN customer_feedback = 'Good' THEN 1 ELSE 0 END) /
      COUNT(customer_feedback), 2) AS good_percentage,
      ROUND(100.0 * SUM(CASE WHEN customer_feedback = 'Very Good' THEN 1 ELSE 0
      END) / COUNT(customer_feedback), 2) AS very_good_percentage
FROM order_t
GROUP BY quarter_number
ORDER BY quarter_number;
```

Output:

SQL queries passed

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

- From the query results, we can see that customers are getting more dissatisfied over time.
- There is a decrease in positive feedback, such as in the “very_good_percentage” column, which starts at 30% in quarter 1 but ends at 10.05% in quarter 4. Simultaneously, the negative feedback increases as the year progresses, such as the “very_bad_percentage” column, which starts at 10.97% in quarter 1 and ends at 30.65% in quarter 4.
- Another insight to note is that the “okay_percentage” column remains pretty consistent around the 20% mark. In other words, customers are not entirely satisfied or dissatisfied.

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

Solution Query:

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

Output:

SQL queries passed

Query 1 (Passed): SELECT quarter_number, COUNT(order_id) AS number_orders FROM order_t GROUP BY quarter_number
Output:
Showing 4 rows

quarter_number	number_orders
1	310
2	262
3	229
4	199

Observations and Insights:

- From this query, it is clear that the number of orders declined as the year progressed (quarter number increased).
- From analyzing the percent distribution of feedback over time on the previous question, it is coherent why as the year progresses there are less number of orders: the customers are not satisfied, therefore, not buying more.
- Although, the decline in orders can be related to previously mentioned events (ex.: poor customer service, higher costs, low product quality, etc.), there also exists the possibility that the decline in orders is due to factors out of New Wheel's control (ex.: market crashes, people spending less money on a vehicle due to seasonal expenses, etc.).

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, total_net_revenue,
      ROUND(LAG(total_net_revenue) OVER (ORDER BY quarter_number), 2) AS
      prev_quarter_revenue,
      ROUND(((total_net_revenue - LAG(total_net_revenue) OVER (ORDER BY
      quarter_number)) / LAG(total_net_revenue) OVER (ORDER BY quarter_number))
      * 100, 2) AS qoq_percentage_change
FROM
  (
    SELECT quarter_number, ROUND(SUM(net_revenue), 2) AS total_net_revenue
    FROM
      (
        SELECT quarter_number,
           quantity * (vehicle_price - (discount / 100) * vehicle_price) AS
           net_revenue
        FROM order_t
      )
      AS table1
    GROUP BY quarter_number
    ORDER BY total_net_revenue DESC
  )
  AS table2;
```

Output:

SQL queries passed

Query 1 (Passed): SELECT quarter_number, total_net_revenue, ROUND(LAG(total_net_revenue) OVER (ORDER BY quarter_number), 2) AS prev_quarter_revenue, ROUND(((total_net_revenue - LAG(total_net_revenue) OVER (ORDER BY quarter_number)) / LAG(total_net_revenue) OVER (ORDER BY quarter_number)) * 100, 2) AS qoq_percentage_change FROM (SELECT quarter_number, ROUND(SUM(net_revenue), 2) AS total_net_revenue FROM (SELECT quarter_number, quantity * (vehicle_price - (discount / 100) * vehicle_price) AS net_revenue FROM order_t) AS table1 GROUP BY quarter_number ORDER BY total_net_revenue DESC) AS table2
Output:
Showing 4 rows

quarter_number	total_net_revenue	prev_quarter_revenue	qoq_percentage_change
1	39421580.16		
2	32715830.34	39421580.16	-17.01
3	29229896.19	32715830.34	-10.66
4	23346779.63	29229896.19	-20.13

Observations and Insights:

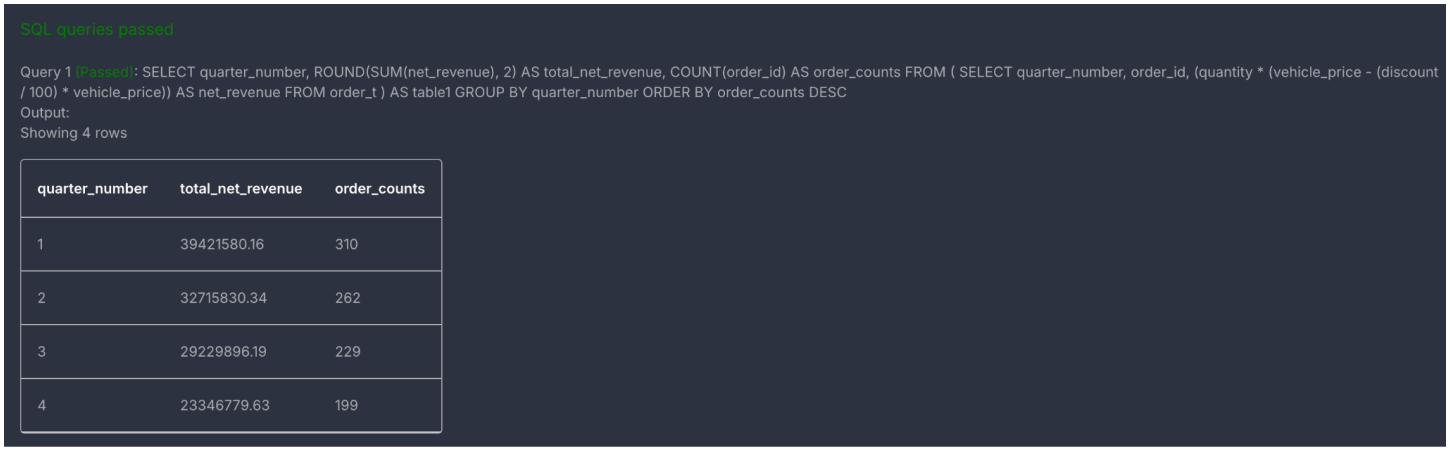
- There is a clear depreciation of revenue as the quarters progress. Quarter 1 starts at a net revenue of $39,421,580.16 and by quarter 4 the revenue is at a total of $23,346,779.63.
- This devaluation trend can be further broken down through the quarter-over-quarter percent change in the net revenue. From quarter 1 to quarter 2 there is a 17.01% decrease of net revenue. From quarter 2 to quarter 3 there is a 10.66% decrease of net revenue. Lastly, from quarter 3 to quarter 4 there is a 20.33% decrease of net revenue.

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

Solution Query:

```
SELECT quarter_number, ROUND(SUM(net_revenue), 2) AS total_net_revenue,
      COUNT(order_id) AS order_counts
FROM
  (
    SELECT quarter_number, order_id,
      (quantity * (vehicle_price - (discount / 100) * vehicle_price)) AS
      net_revenue
    FROM order_t
  )
  AS table1
GROUP BY quarter_number
ORDER BY order_counts DESC;
```

Output:



SQL queries passed

Query 1 (Passed): SELECT quarter_number, ROUND(SUM(net_revenue), 2) AS total_net_revenue, COUNT(order_id) AS order_counts FROM (SELECT quarter_number, order_id, (quantity * (vehicle_price - (discount / 100) * vehicle_price)) AS net_revenue FROM order_t) AS table1 GROUP BY quarter_number ORDER BY order_counts DESC
Output:
Showing 4 rows

quarter_number	total_net_revenue	order_counts
1	39421580.16	310
2	32715830.34	262
3	29229896.19	229
4	23346779.63	199

Observations and Insights:

- The trend between the net revenue and the order by quarters is that as the total net revenue decreases so does the number of orders.
- This confirms that when people buy less products then less revenue comes in.

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

Solution Query:

```
SELECT credit_card_type, ROUND(AVG(discount), 2) AS average_discount
FROM order_t JOIN customer_t USING (customer_id)
GROUP BY credit_card_type
ORDER BY average_discount DESC;
```

Output:

SQL queries passed

Query 1 (Passed): SELECT credit_card_type, ROUND(AVG(discount),2) AS average_discount FROM order_t JOIN customer_t USING (customer_id) GROUP BY credit_card_type ORDER BY average_discount 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:

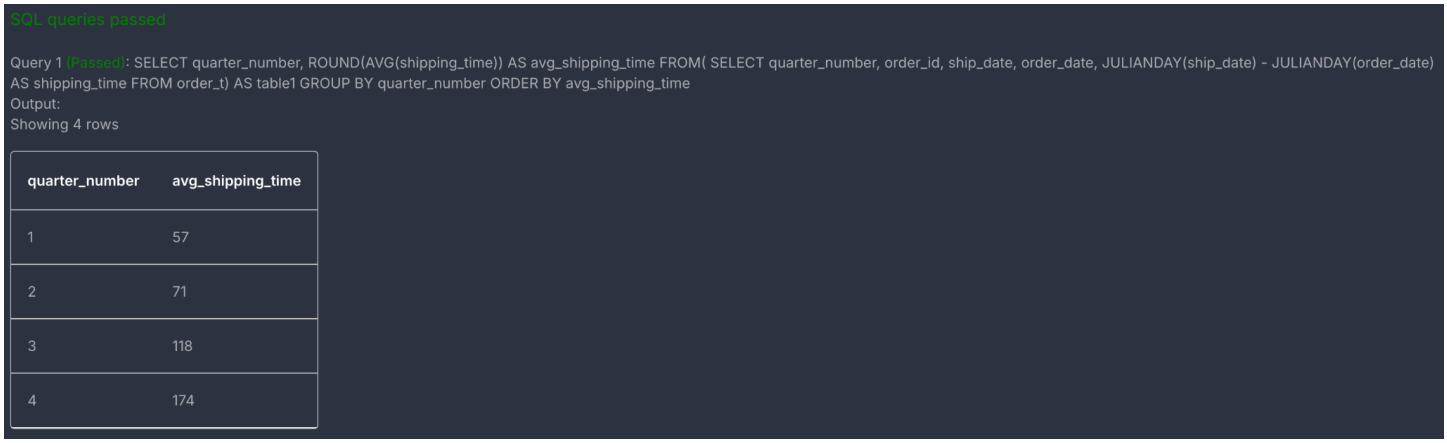
- The top two credit cards with the highest average discounts are Instapayment (0.77) and Solo (0.70).
- Surprisingly, well-known credit cards like Mastercard and Visa Electron sit in the middle range of averages at 0.65 and 0.64.
- In the bottom bracket, we have Laser and China Unionpay each with an average of 0.62.

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

Solution Query:

```
SELECT quarter_number, ROUND(AVG(shipping_time)) AS avg_shipping_time
FROM
  (
    SELECT quarter_number, order_id, ship_date, order_date,
      JULIANDAY(ship_date) - JULIANDAY(order_date) AS shipping_time
    FROM order_t
  )
  AS table1
GROUP BY quarter_number
ORDER BY avg_shipping_time;
```

Output:


SQL queries passed

Query 1 (Passed): SELECT quarter_number, ROUND(AVG(shipping_time)) AS avg_shipping_time FROM(SELECT quarter_number, order_id, ship_date, order_date, JULIANDAY(ship_date) - JULIANDAY(order_date) AS shipping_time FROM order_t) AS table1 GROUP BY quarter_number ORDER BY avg_shipping_time
Output:
Showing 4 rows

quarter_number	avg_shipping_time
1	57
2	71
3	118
4	174

Observations and Insights:

- The average shipping time in quarter 1 is 57 days, 71 days in quarter 2, 118 days in quarter 3 and 174 days in quarter 4.
- At the start of quarter 1 (which had the highest number of orders, highest total net revenue and highest “very good percent feedback” by customers), the shipping time was around 57 days – the quickest shipping time of the entire year. However, by quarter 4, the shipping time increases exponentially by 3 – totaling to 174 days. These results validate why in quarter 4 there were lower number of orders, lower total net revenue and higher “very bad percent feedback” by customers.

Business Metrics Overview

Total Revenue	Total Orders	Total Customers	Average Rating
$124,714,086.32	1,000	133	3.14
Last Quarter Revenue	**Last quarter Orders**	**Average Days to Ship**	**% Good Feedback**
$23,346,779.63	199	98	21.50%

Business Recommendations

- It is fundamental to focus marketing efforts in Arizona, Idaho, Nebraska, New Hampshire, New Jersey, North Dakota, Oklahoma and West Virginia (states with the lower total number of customers) to increase clientele in these regions.
- For California, Texas, Florida and New York (states with higher total number of customers), the demand for products is a lot higher; therefore, evaluate the inventory needed to meet customers' expectations in these regions.
- Chevrolet, Ford, Toyota, Pontiac and Dodge should be the focal brands in marketing campaigns across the nation. Nevertheless, it is important to note that while promoting these campaigns, the products presented should align with the demands and terrain of specific regions. For example, rural areas might prefer 4x4's and trucks with large cargo bed space while urban areas might prefer hybrid and/or smaller vehicles.
- Consider promotion campaigns that include Instapayment and Solo credit card users since these have the highest average discounts.
- The decline in demand, net revenue and customer satisfaction is highly a consequence of an exponential increase in shipping time towards the end of quarter 4. Possible causes for these shipping delays could be due to seasonal factors, congested production, lack of inventory and/or labor cuts. Focus on improving warehouse logistics, increasing inventory or staff to shorten the shipping times.
- In order to enhance customer satisfaction and maintain clientele, it is very important to identify the cause of these shipping delays, as soon as possible.