

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:

- Find the total number of customers who have placed orders

Query -> select count(distinct customer_id) from order_t;

- What is the distribution of the customers across states?

Query->SELECT state,

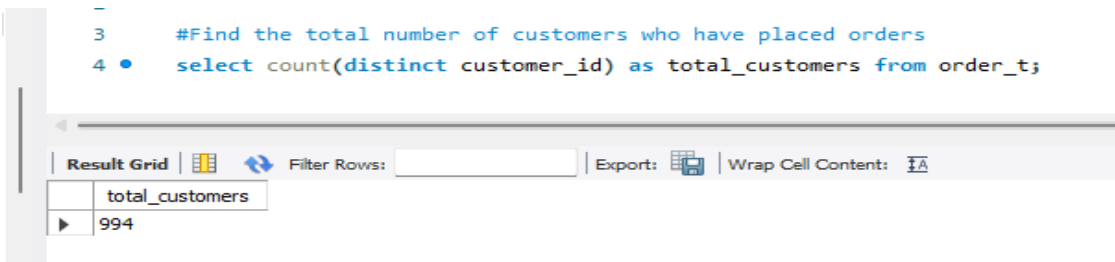
COUNT(customer_id) AS total_customers

FROM customer_t

GROUP BY 1

ORDER BY 2 DESC;

Output:



```
-  
3 #Find the total number of customers who have placed orders  
4 • select count(distinct customer_id) as total_customers from order_t;
```

total_customers
994

```

6  # What is the distribution of the customers across states?
7  • SELECT
8      state,
9      COUNT(customer_id) AS customer_cnt_per_state
10 FROM customer_t
11 GROUP BY 1
12 ORDER BY 2 DESC;

```

Result Grid			Filter Rows:	Export:	Wrap Cell Content:
	state	customer_cnt_per_state			
▶	Texas	97			
	California	97			
	Florida	86			
	New York	69			
	District of Columbia	35			
	Ohio	33			
	Colorado	33			
	Alabama	29			
	Washington	28			
	Arizona	26			
	Pennsylvania	25			
	Illinois	25			
	Virginia	24			
	Tennessee	23			
	Missouri	23			
	Connecticut	22			

Observations and Insights:

- Total Unique customers that placed order is 994.
- Texas and California are having most number of customer followed by Florida and New York.

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

Solution Query:

```
SELECT  
  
    vehicle_maker AS top_vehicle_makers,  
  
    COUNT(customer_id) AS total_customers  
  
FROM product_T JOIN customer_t  
  
GROUP BY 1  
  
ORDER BY 2 DESC  
  
LIMIT 5;
```

Output:

```
14  #Which are the top 5 vehicle makers preferred by the customers?  
15  •  SELECT  
16      vehicle_maker AS top_vehicle_makers,  
17      COUNT(customer_id) AS total_customers  
18  FROM product_T JOIN customer_t  
19  GROUP BY 1  
20  ORDER BY 2 DESC  
21  LIMIT 5;
```

top_vehicle_makers	total_customers
Chevrolet	82502
Ford	62622
Toyota	51688
Dodge	49700
Pontiac	49700

Observations and Insights:

- Chevrolet and Ford are top vehicle makers followed by Toyota.
- Dodge and Pontiac are also in top 5 vehicle makers

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

Solution Query:





```
SELECT *  
FROM  
(  
    SELECT  
        state,  
        vehicle_maker,  
        COUNT(customer_id) AS total_customers,  
        RANK() OVER (PARTITION BY state ORDER BY COUNT(customer_id) DESC) AS ranking  
    FROM product_t  
    JOIN order_t USING(product_id)  
    JOIN customer_t USING(customer_id)  
    GROUP BY 1, 2  
) AS preferred_vehicle  
WHERE ranking = 1  
ORDER BY 3 DESC;
```

Output:

```

22
23 # Which is the most preferred vehicle maker in each state?
24 • SELECT *
25 FROM
26 (
27     SELECT
28         state,
29         vehicle_maker,
30         COUNT(customer_id) AS total_customers,
31         RANK() OVER (PARTITION BY state ORDER BY COUNT(customer_id) DESC) AS ranking
32     FROM product_t
33     JOIN order_t USING(product_id)
34     JOIN customer_t USING(customer_id)
35     GROUP BY 1, 2
36 ) AS preferred_vehicle
37 WHERE ranking = 1
38 ORDER BY 3 DESC;
39

```

Result Grid   Filter Rows: Export:  Wrap Cell Content: 

	state	vehicle_maker	total_customers	ranking
▶	Texas	Chevrolet	9	1
	Florida	Toyota	7	1
	California	Nissan	6	1
	California	Ford	6	1
	California	Chevrolet	6	1
	California	Audi	6	1
	California	Dodge	6	1
	Ohio	Chevrolet	6	1
	New York	Pontiac	5	1
	Colorado	Chevrolet	5	1
	Virginia	Ford	5	1
	Maryland	Ford	5	1
	New York	Toyota	5	1
	Washing...	Chevrolet	5	1
	Alabama	Dodge	5	1
	Indiana	Mazda	4	1
	District o...	Chevrolet	4	1
	Missouri	Chevrolet	4	1
	Arizona	Pontiac	3	1
	Minnesota	GMC	3	1

Observations and Insights:

- In Texas Chevrolet is top choice of customers
- In Florida Toyota is top choice of customers
- In California there is equal domination of Nissan, Ford, Chevrolet, Audi and Dodge
- There is different vehicle maker are choice of customers in different state

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 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 avg_rating
FROM order_t;
```

- Average rating in each quarter

```
WITH rating AS
(
  SELECT
    customer_feedback,
    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 total_rating
  FROM order_t
)
SELECT
  quarter_number,
  ROUND(AVG(total_rating), 2) AS average_rating
FROM rating
GROUP BY 1
ORDER BY 1 ASC;
```

Output:

```

40  #Find the overall average rating given by the customers
41  • SELECT
42      AVG(
43          CASE
44              WHEN customer_feedback = 'very bad' THEN '1'
45              WHEN customer_feedback = 'bad' THEN '2'
46              WHEN customer_feedback = 'okay' THEN '3'
47              WHEN customer_feedback = 'good' THEN '4'
48              WHEN customer_feedback = 'very good' THEN '5'
49          END
50      )as avg_rating
51  FROM order_t;
52
53

```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: [IA](#)

avg_rating
3.135

```

53  #Find the overall average rating given by the customers. What is the average rating in each quarter?
54  • WITH rating AS
55      (
56          SELECT
57              customer_feedback,
58              quarter_number,
59          CASE
60              WHEN customer_feedback = 'very bad' THEN '1'
61              WHEN customer_feedback = 'bad' THEN '2'
62              WHEN customer_feedback = 'okay' THEN '3'
63              WHEN customer_feedback = 'good' THEN '4'
64              WHEN customer_feedback = 'very good' THEN '5'
65          END AS total_rating
66          FROM order_t
67      )
68      SELECT
69          quarter_number,
70          ROUND(AVG(total_rating), 2) AS average_rating
71      FROM rating
72      GROUP BY 1
73      ORDER BY 1 ASC;
74

```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: [IA](#)

quarter_number	average_rating
1	3.55
2	3.35
3	2.96
4	2.40

Observations and Insights:

- Average rating of customer is around 3.135
- Average rating is highest in 1st Quarter
- Average rating is going down each quarter with lowest of 2.4 in 4th quarter.

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

Solution Query:

WITH cust_feed AS

```
(  
    SELECT  
        quarter_number,  
        ROUND(SUM(CASE WHEN customer_feedback = 'very good' THEN 1 ELSE 0 END), 2) AS very_good,  
        ROUND(SUM(CASE WHEN customer_feedback = 'good' THEN 1 ELSE 0 END), 2) AS good,  
        ROUND(SUM(CASE WHEN customer_feedback = 'okay' THEN 1 ELSE 0 END), 2) AS okay,  
        ROUND(SUM(CASE WHEN customer_feedback = 'bad' THEN 1 ELSE 0 END), 2) AS bad,  
        ROUND(SUM(CASE WHEN customer_feedback = 'very bad' THEN 1 ELSE 0 END), 2) AS very_bad,  
        ROUND(COUNT(customer_feedback), 2) AS total_feedback  
    FROM order_t  
    GROUP BY 1  
    ORDER BY 1 ASC  
)
```

```
SELECT  
    quarter_number,  
    ROUND((very_good/total_feedback), 2) AS very_good,  
    ROUND((good/total_feedback), 2) AS good,  
    ROUND((okay/total_feedback), 2) AS okay,  
    ROUND((bad/total_feedback), 2) AS bad,  
    ROUND((very_bad/total_feedback), 2) AS very_bad  
FROM cust_feed  
GROUP BY 1  
ORDER BY 1 ASC;
```

Output:

```

74 #Find the percentage distribution of feedback from the customers. Are customers getting more dissatisfied over time?
75 • WITH cust_feed AS
76 (
77     SELECT
78         quarter_number,
79         ROUND(SUM(CASE WHEN customer_feedback = 'very good' THEN 1 ELSE 0 END), 2) AS very_good,
80         ROUND(SUM(CASE WHEN customer_feedback = 'good' THEN 1 ELSE 0 END), 2) AS good,
81         ROUND(SUM(CASE WHEN customer_feedback = 'okay' THEN 1 ELSE 0 END), 2) AS okay,
82         ROUND(SUM(CASE WHEN customer_feedback = 'bad' THEN 1 ELSE 0 END), 2) AS bad,
83         ROUND(SUM(CASE WHEN customer_feedback = 'very bad' THEN 1 ELSE 0 END), 2) AS very_bad,
84         ROUND(COUNT(customer_feedback), 2) AS total_feedback
85     FROM order_t
86     GROUP BY 1
87     ORDER BY 1 ASC
88 )
89 SELECT
90     quarter_number,
91     ROUND((very_good/total_feedback), 2) AS very_good,
92     ROUND((good/total_feedback), 2) AS good,
93     ROUND((okay/total_feedback), 2) AS okay,
94     ROUND((bad/total_feedback), 2) AS bad,
95     ROUND((very_bad/total_feedback), 2) AS very_bad
96 FROM cust_feed
97 GROUP BY 1
98 ORDER BY 1 ASC;

```

result Grid | Filter Rows: | Export: | Wrap Cell Content: [iA](#)

quarter_number	very_good	good	okay	bad	very_bad
1	0.30	0.29	0.19	0.11	0.11
2	0.29	0.22	0.20	0.14	0.15
3	0.17	0.21	0.22	0.23	0.18
4	0.10	0.10	0.20	0.29	0.31

Observations and Insights:

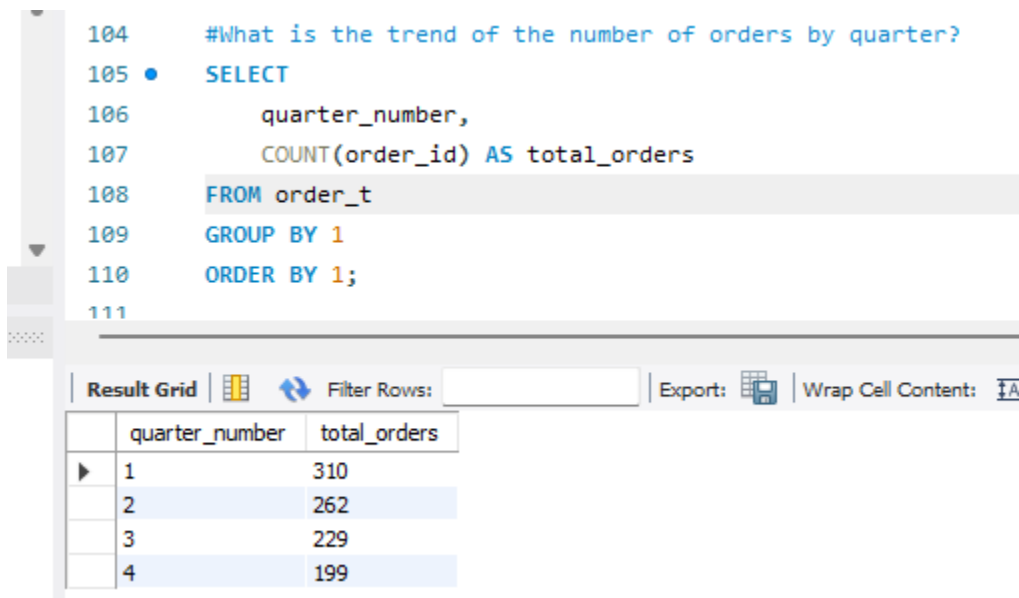
- As we can clearly see that very_good rating percentage is continuously decreasing with each quarter.
- Like that good rating is also decreasing each quarter,
- Okay rating is quite same in each quarter.
- While bad and very_bad rating continuously increasing by each quarter.
- With these observations we can clearly see that customer are getting more dissatisfied over time.

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 1
ORDER BY 1;
```

Output:



The screenshot shows a SQL query editor with the following query:

```
104 #What is the trend of the number of orders by quarter?
105 SELECT
106     quarter_number,
107     COUNT(order_id) AS total_orders
108 FROM order_t
109 GROUP BY 1
110 ORDER BY 1;
111
```

Below the query editor, the output is displayed in a table format. The table has two columns: 'quarter_number' and 'total_orders'. The data shows a decreasing trend in the number of orders over the four quarters.

quarter_number	total_orders
1	310
2	262
3	229
4	199

Observations and Insights:

- Total orders are highest in 1st quarter as 310.
- Least orders i.e. 199 received in last quarter
- We can see orders are continuously decreasing with each quarter.

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

Solution Query:

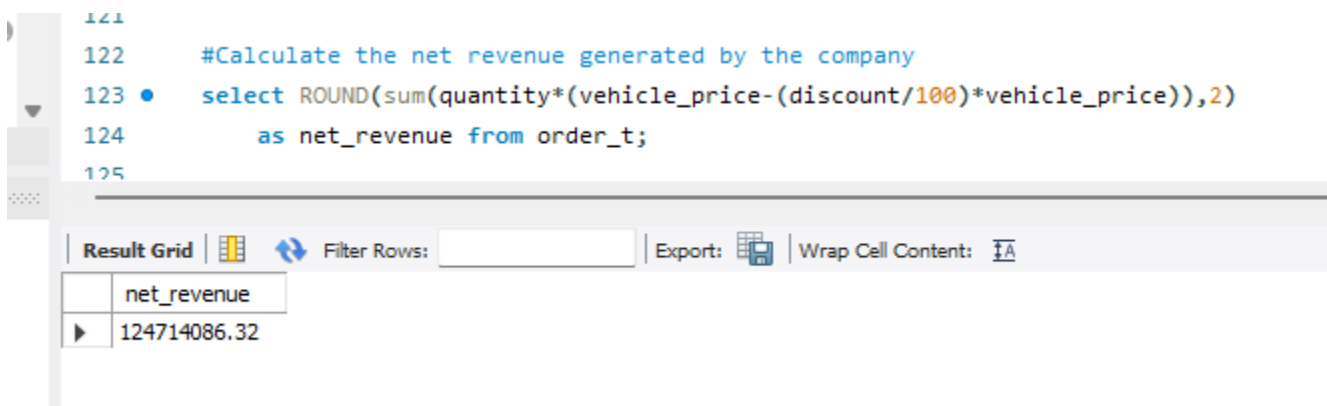
- Net revenue generated by the company

```
select ROUND(sum(quantity*(vehicle_price-(discount/100)*vehicle_price)),2)
as net_revenue
from order_t;
```

- quarter-over-quarter % change in net revenue.

```
WITH QoQ AS
(
    SELECT quarter_number,
    ROUND(SUM(quantity * (vehicle_price - ((discount/100)*vehicle_price))), 0) AS revenue
    FROM order_t
    GROUP BY quarter_number)
SELECT quarter_number, revenue,
ROUND(LAG(revenue) OVER(ORDER BY quarter_number), 2) AS previous_revenue,
ROUND((revenue - LAG(revenue) OVER(ORDER BY quarter_number))/LAG(revenue) OVER(ORDER BY
quarter_number), 2) AS qoq_perc_change
FROM QoQ;
```

Output:



The screenshot shows a SQL IDE interface. The query editor contains the following SQL code:

```
121
122 #Calculate the net revenue generated by the company
123 • select ROUND(sum(quantity*(vehicle_price-(discount/100)*vehicle_price)),2)
124 as net_revenue from order_t;
125
```


Below the query editor, there is a toolbar with options: Result Grid, Filter Rows, Export, and Wrap Cell Content. The Result Grid shows the output of the query:

net_revenue
124714086.32

```

122 #What is the quarter-over-quarter % change in net revenue?
123 • WITH QoQ AS
124 (
125     SELECT quarter_number,
126            ROUND(SUM(quantity * (vehicle_price - ((discount/100)*vehicle_price))), 0) AS revenue
127     FROM order_t
128     GROUP BY quarter_number)
129 SELECT quarter_number, revenue,
130        ROUND(LAG(revenue) OVER(ORDER BY quarter_number), 2) AS previous_revenue,
131        ROUND((revenue - LAG(revenue) OVER(ORDER BY quarter_number))/LAG(revenue) OVER(ORDER BY quarter_number), 2) AS qoq_perc_change
132 FROM QoQ;
133

```

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

	quarter_number	revenue	previous_revenue	qoq_perc_change
▶	1	39421580	NULL	NULL
	2	32715830	39421580.00	-0.17
	3	29229896	32715830.00	-0.11
	4	23346780	29229896.00	-0.20

Observations and Insights:

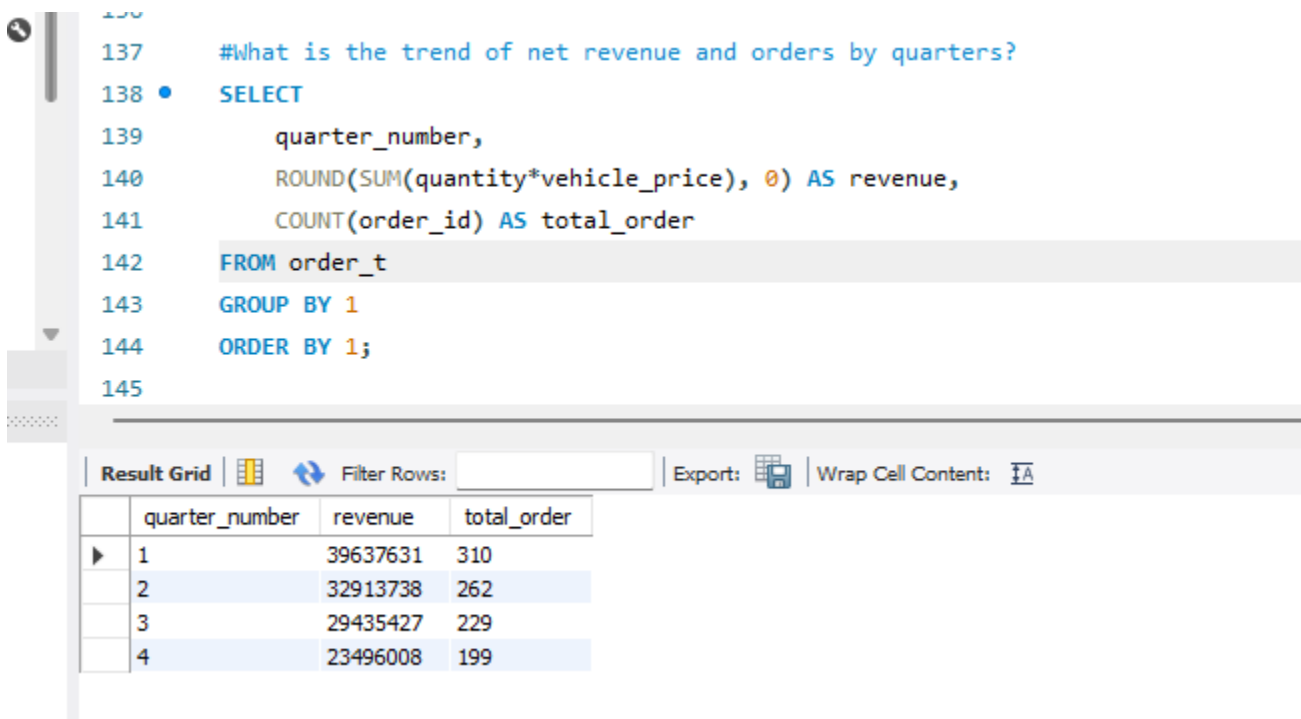
- Net revenue is around 124714086.32
- Revenue use declining in each quarter.

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

Solution Query:

```
SELECT
    quarter_number,
    ROUND(SUM(quantity*vehicle_price), 0) AS revenue,
    COUNT(order_id) AS total_order
FROM order_t
GROUP BY 1
ORDER BY 1;
```

Output:



The screenshot shows a SQL query editor with the following query:

```
137 #What is the trend of net revenue and orders by quarters?
138 SELECT
139     quarter_number,
140     ROUND(SUM(quantity*vehicle_price), 0) AS revenue,
141     COUNT(order_id) AS total_order
142 FROM order_t
143 GROUP BY 1
144 ORDER BY 1;
145
```

Below the query, the results are displayed in a table with the following columns: quarter_number, revenue, and total_order.

quarter_number	revenue	total_order
1	39637631	310
2	32913738	262
3	29435427	229
4	23496008	199

Observations and Insights:

- We can see orders and revenue is highest in first quarter
- Orders are decreasing in each quarter so it the revenue

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 t1
INNER JOIN customer_t t2
ON t1.customer_id = t2.customer_id
GROUP BY 1
ORDER BY 2 DESC;
```


Output:

```

145
146 #What is the average discount offered for different types of credit cards?
147 • SELECT
148     credit_card_type,
149     ROUND(AVG(discount), 2) AS average_discount
150 FROM order_t t1
151 INNER JOIN customer_t t2
152     ON t1.customer_id = t2.customer_id
153 GROUP BY 1
154 ORDER BY 2 DESC;
155

```

credit_card_type	average_discount
laser	0.64
mastercard	0.63
maestro	0.62
americanexpress	0.62
instapayment	0.62
visa-electron	0.62
china-unionpay	0.62
jcb	0.61
diners-club-carte-blanche	0.61
switch	0.61
bankcard	0.61
diners-club-us-ca	0.61
diners-club-enroute	0.60
visa	0.60
solo	0.59
diners-club-international	0.58



Observations and Insights:

- laser cards give the highest average discount followed by MasterCard.
- diners-club-international card gives lowest discount.

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

Solution Query:

```
SELECT
    quarter_number,
    ROUND(AVG(DATEDIFF(ship_date, order_date)), 0) AS average_shipping_time
FROM order_t
GROUP BY 1
ORDER BY 1;
```

Output:

```
156  #What is the average time taken to ship the placed orders for each quarter?
157  •  SELECT
158      quarter_number,
159      ROUND(AVG(DATEDIFF(ship_date, order_date)), 0) AS average_shipping_time
160  FROM order_t
161  GROUP BY 1
162  ORDER BY 1;
```

quarter_number	average_shipping_time
1	57
2	71
3	118
4	174

Observations and Insights:

- Average shipping time is getting increased in each quarter

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
23346780	199	97.9	44.1

Business Recommendations

- As we can see from rating trend that customers are getting dissatisfied bad rating are getting increased while good rating getting decreased in each quarter so New wheel need to understand what is causing issue and need to work on the their service to improve customer satisfaction
- From shipping time, we can see increase in shipping time in each quarter. New wheels need to work on this and should decrease shipping time.
- New wheel sale is getting decreased in each quarter for this they need to check with customer and do market survey to understand reason for the same.