

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?

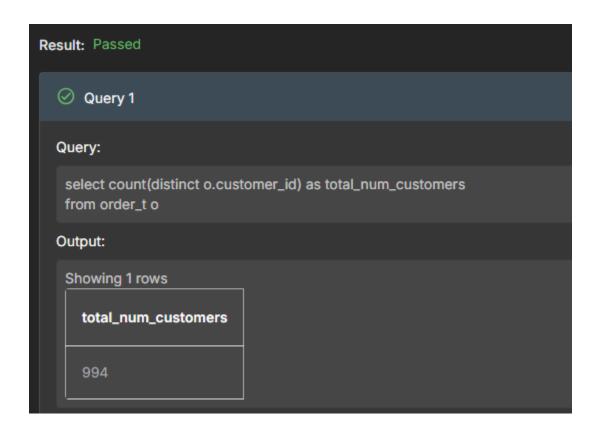
Solution Query:

1.a) Total number of customers placed orders:

select count(distinct o.customer_id) as total_num_customers

from order_t o:

output:



- The query counts the total number of distinct customers from the order_t table.
- The output indicates that there are 994 unique customers in the dataset who have placed orders.



1.b) The Distribution of the customers across states

select c.customer_id,o.order_id,c.state,

count(c.customer_id) as Number_of_customer

from customer_t c

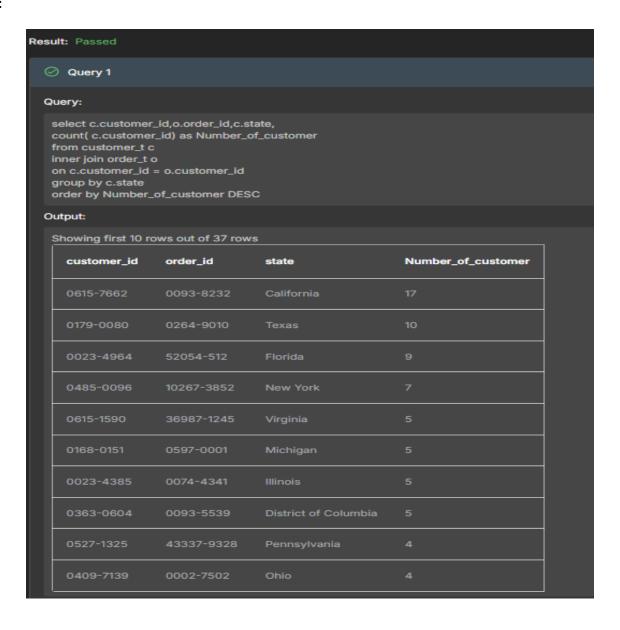
inner join order_t o

on c.customer_id = o.customer_id

group by c.state

order by Number_of_customer DESC;

Output:





- The query groups the data by state and counts the number of customers (Number_of_customer) in each state based on orders.
- The data shows a clear variation in customer distribution among states.
- California has the highest number of customers (17), indicating a strong market presence or larger customer base in the region.
- Other states like Texas (10), Florida (9), and New York (7) also show significant customer activity, though they trail behind California.
- States like Virginia, Michigan, Illinois, and others have a smaller number of customers (5 or fewer), suggesting lesser engagement or market penetration in those areas.

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

Solution Query:

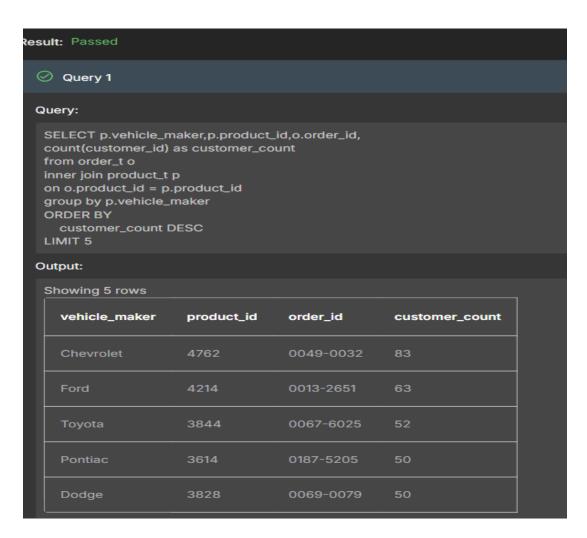
SELECT p.vehicle_maker,p.product_id,o.order_id,
count(customer_id) as customer_count
from order_t o
inner join product_t p
on o.product_id = p.product_id
group by p.vehicle_maker

ORDER BY
customer_count DESC

LIMIT 5;



Output:



- Chevrolet, Ford, Toyota, Pontiac, Dodge are the top 5 vehicle makers preferred by the customers.
- Chevrolet ranks first, with 83 customers.
- Ford follows with 63 customers, and Toyota ranks third with 52 customers.
- Both Pontiac and Dodge have 50 customers each.

Great Learning

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

state?

```
Solution Query:
SELECT *
FROM
  SELECT
    state,
    vehicle_maker,
    COUNT(c.customer_id) AS total_customers,
  RANK() OVER (PARTITION BY state ORDER BY COUNT(c.customer_id) DESC) AS ranking
  FROM product_t p
  JOIN order_t o
  on p.product_id = o.product_id
  JOIN customer_t c
  ON o.customer_id = c.customer_id
  GROUP BY 1, 2
) AS preferred_vehicle
WHERE ranking = 1
ORDER BY 3 DESC;
```





```
Result: Passed
  Query 1
  Query:
   SELECT *
   FROM
       SELECT
            state,
            vehicle_maker,
            COUNT(c.customer_id) AS total_customers,
     RANK() OVER (PARTITION BY state ORDER BY COUNT(c.customer_id) DESC) AS ranking
     FROM product_t p
     JOIN order_t o
     on p.product_id = o.product_id
     JOIN customer_t c
     ON o.customer_id = c.customer_id
       GROUP BY 1, 2
   ) AS preferred_vehicle
   WHERE ranking = 1
   ORDER BY 3 DESC
  Output:
   Showing first 10 rows out of 101 rows
     state
                   vehicle_maker
                                     total_customers
                                                        ranking
```



- California: There are four vehicle makers (Pontiac, Nissan, Ford, and Chevrolet) that are tied as the most preferred in California, each with 2 customers.
- This indicates a high level of competition and potentially a wide range of vehicle preferences among customers in California.
- Florida: Similarly, Volvo and Ford both have 2 customers as the most preferred vehicle makers. Like California, this suggests that Florida has multiple options for customers that they prefer, possibly due to availability or brand popularity.
- Texas: In Texas, Nissan leads with 2 customers, showing that there is a preference for Nissan in this state, but there is no mention of other vehicle makers leading. This could indicate lower competition or a strong brand presence for Nissan.
- Indiana: Mazda is the leading vehicle maker with only 2 customers, indicating that Indiana has a niche preference, and Mazda is highly preferred among its relatively small customer base.
- Alabama: Both Lincoln and Lexus lead in Alabama, with just 1 customer each. This reflects a limited preference and potentially lower overall customer volume in the state, suggesting that either fewer customers place orders or the market for vehicle makers is more specialized in this state.
- The rankings indicate a tied preference for many vehicle makers across states, which may reflect factors such as brand reputation, regional availability, or customer loyalty.
- Pontiac, Nissan, Ford, Chevrolet, and Volvo appear multiple times, suggesting they may be dominant vehicle makers in different states.

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:

4.a) overall average rating:

select

avg(CASE

WHEN o.customer_feedback = 'Very Bad' THEN 1

WHEN o.customer_feedback = 'Bad' THEN 2

WHEN o.customer_feedback = 'Okay' THEN 3

```
WHEN o.customer_feedback = 'Good' THEN 4
```



WHEN o.customer_feedback = 'Very Good' THEN 5

ELSE NULL

END

) AS overall_avg_rating

from order_t o;

OUTPUT:

```
Result: Passed
  Query 1
  Query:
   select
   avg(CASE
         WHEN o.customer_feedback = 'Very Bad' THEN 1
         WHEN o.customer_feedback = 'Bad' THEN 2
         WHEN o.customer_feedback = 'Okay' THEN 3
         WHEN o.customer_feedback = 'Good' THEN 4
         WHEN o.customer_feedback = 'Very Good' THEN 5
         ELSE NULL
       END
     ) AS overall_avg_rating
     from order_t o
  Output:
   Showing 1 rows
     overall_avg_rating
     3.135
```

Observation:

The overall average rating given by the customers is 3.135.

4.b) Average rating in each quarter:

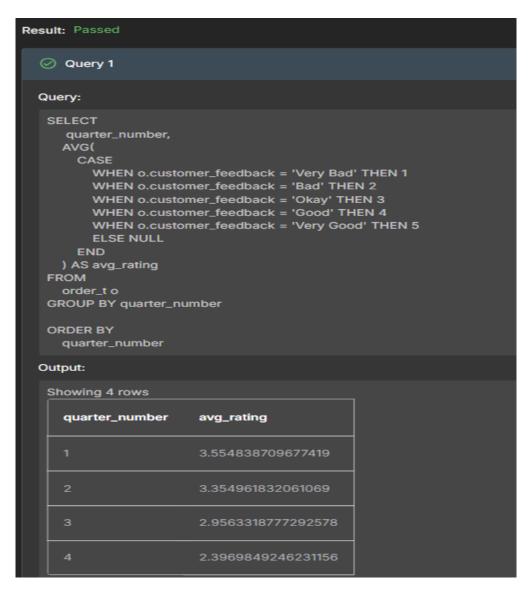
quarter_number;



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







- This output shows the **Decreasing Trend in Average Rating**.
- Q1 (3.55): The highest average rating.
- Q2 (3.35): Slight decline compared to Q1.
- Q3 (2.96): Significant drop from Q2.
- Q4 (2.40): The lowest average rating, indicating worsening customer feedback.
- The consistent decline in average ratings across quarters may indicate a decline in customer satisfaction.



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

Solution Query:

```
SELECT
  quarter_number,
  COUNT(o.customer_feedback) AS total_feedback,
  ROUND(
    100.0 * SUM(CASE WHEN o.customer_feedback = 'Very Bad' THEN 1 ELSE 0 END) /
COUNT(o.customer_feedback), 2
  ) AS percent_very_bad,
  ROUND(
    100.0 * SUM(CASE WHEN o.customer_feedback = 'Bad' THEN 1 ELSE 0 END) /
COUNT(o.customer_feedback), 2
 ) AS percent_bad,
  ROUND(
    100.0 * SUM(CASE WHEN o.customer_feedback = 'Okay' THEN 1 ELSE 0 END) /
COUNT(o.customer_feedback), 2
 ) AS percent_okay,
  ROUND(
    100.0 * SUM(CASE WHEN o.customer_feedback = 'Good' THEN 1 ELSE 0 END) /
COUNT(o.customer_feedback), 2
  ) AS percent_good,
  ROUND(
    100.0 * SUM(CASE WHEN o.customer_feedback = 'Very Good' THEN 1 ELSE 0 END) /
COUNT(o.customer_feedback), 2
 ) AS percent_very_good
FROM
  order_t o
```

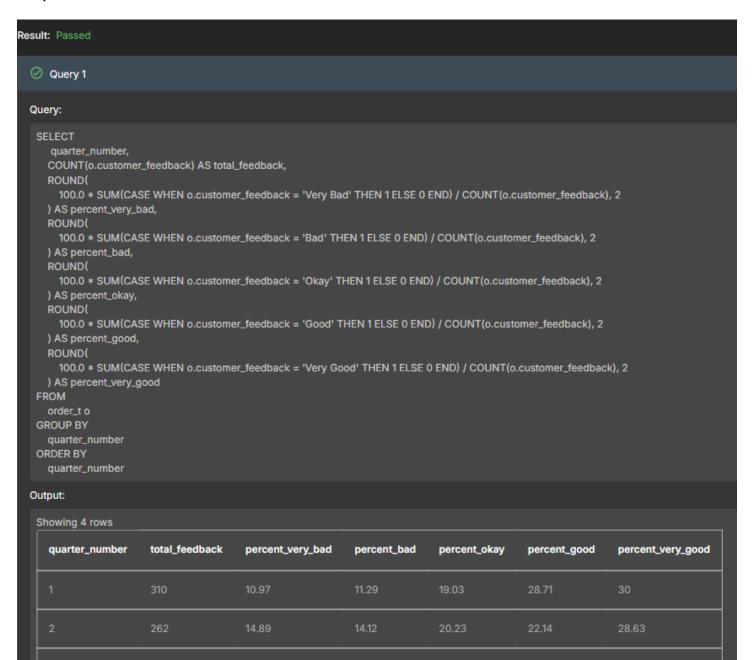


quarter_number

ORDER BY

quarter_number;

Output:





- The query calculates the total feedback count and the percentage distribution of customer feedback across different ratings: Very Bad, Bad, Okay, Good, and Very Good, grouped by quarter_number.
- As we move from Quarter 1 to Quarter 4:
- Very Bad Feedback percentage increases significantly, from 10.97% in Q1 to 30.65% in Q4.
- Bad Feedback also sees a steady increase from 11.29% in Q1 to 29.15% in Q4.
- Okay Feedback shows a declining trend, starting at 19.03% in Q1 and dropping to 20.1% in Q4.
- Good Feedback declines significantly, from 28.71% in Q1 to 10.05% in Q4.
- Very Good Feedback also decreases notably, from 30% in Q1 to 10.05% in Q4.
- The total feedback count decreases progressively, from 310 in Q1 to 199 in Q4.
- The steep increase in **negative feedback** (Very Bad and Bad) and the drop in **positive feedback** (Good and Very Good) indicate a significant decline in customer satisfaction over the quarters.

Yes, customers are getting more dissatisfied over time.

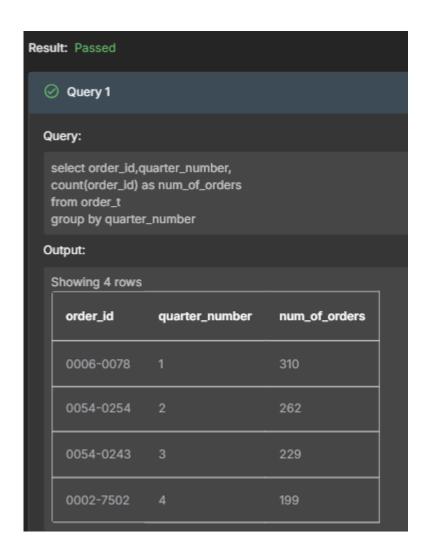
Question 6: What is the trend of the number of orders by quarter?

Solution Query:

select order_id,quarter_number,
count(order_id) as num_of_orders
from order_t
group by quarter_number;







- The number of orders steadily declines across the quarters:
- **Q1**: 310 orders
- **Q2:** 262 orders
- **Q3**: 229 orders
- **Q4**: 199 orders
- The rate of decline is consistent from one quarter to next quarter.
- The gradual reduction in the number of orders suggests a potential decline in customer interest or engagement over the quarters.
- This could be related to a decrease in customer satisfaction as observed in the feedback analysis.



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

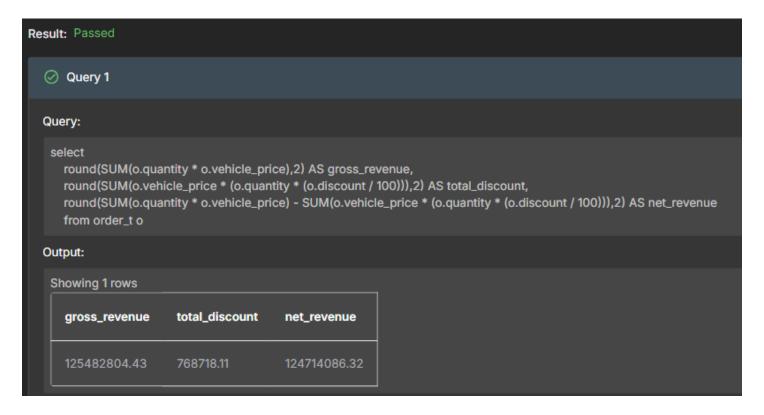
Solution Query:

7.a) The net revenue generated by the company:

select

```
round(SUM(o.quantity * o.vehicle_price),2) AS gross_revenue,
round(SUM(o.vehicle_price * (o.quantity * (o.discount / 100))),2) AS total_discount,
round(SUM(o.quantity * o.vehicle_price) - SUM(o.vehicle_price * (o.quantity * (o.discount / 100))),2) AS
net_revenue
from order_t o;
```

Output:



Observations and Insights:

• A net revenue of **124.71 million** after a minimal discount amounting to **768.71k** indicates the company retains a significant portion of its potential revenue.





```
SELECT
  o. quarter_number,
  SUM(o.quantity * o.vehicle_price) AS gross_revenue,
  SUM(o.vehicle_price * (o.quantity * (o.discount / 100))) AS total_discount,
  SUM(o.quantity * o.vehicle_price) - SUM(o.vehicle_price * (o.quantity * (o.discount / 100))) AS net_revenue,
  ROUND(
     100.0 * (
       SUM(o.quantity * o.vehicle_price) - SUM(o.vehicle_price * (o.quantity * (o.discount / 100)))
       - LAG(SUM(o.quantity * o.vehicle_price) - SUM(o.vehicle_price * (o.quantity * (o.discount / 100))))
        OVER (ORDER BY o. quarter number)
    ) / LAG(SUM(o.quantity * o.vehicle_price) - SUM(o.vehicle_price * (o.quantity * (o.discount / 100))))
      OVER (ORDER BY o. quarter number), 2
  ) AS qoq_change
FROM
  order_t o
GROUP BY
  o.quarter_number
ORDER BY
  o.quarter_number;
```



Result: Passed



Query 1

Query:

SELECT

o. quarter_number,

SUM(o.quantity * o.vehicle_price) AS gross_revenue,

SUM(o.vehicle_price * (o.quantity * (o.discount / 100))) AS total_discount,

SUM(o.quantity * o.vehicle_price) - SUM(o.vehicle_price * (o.quantity * (o.discount / 100))) AS net_revenue, ROUND(

100.0 * (

SUM(o.quantity * o.vehicle_price) - SUM(o.vehicle_price * (o.quantity * (o.discount / 100)))

- LAG(SUM(o.quantity * o.vehicle_price) - SUM(o.vehicle_price * (o.quantity * (o.discount / 100)))) OVER (ORDER BY o. quarter_number)

) / LAG(SUM(o.quantity * o.vehicle_price) - SUM(o.vehicle_price * (o.quantity * (o.discount / 100))))

OVER (ORDER BY o. quarter_number), 2

) AS qoq_change

FROM

order_t o

GROUP BY

o.quarter_number

ORDER BY

o.quarter_number

Output:

Showing 4 rows

quarter_number	gross_revenue	total_discount	net_revenue	qoq_change
1	39637630.97	216050.8107039999	39421580.159296	
2	32913737.76	197907.42003800007	32715830.339962002	-17.01
3	29435427.479999986	205531.28635099987	29229896.193648987	-10.66
4	23496008.219999995	149228.5893940001	23346779.630605996	-20.13

Insight and observation:



Quarter-over-Quarter (QoQ) % Change in Net Revenue:

- Q2 vs Q1: The net revenue decreased by 17.01%.
- Q3 vs Q2: The net revenue decreased by 10.66%.
- Q4 vs Q3: The net revenue decreased by 20.13%.
- Net revenue has consistently decreased across all quarters.
- The most significant drop occurred in Q4, with a decline of 20.13% compared to the previous quarter.
- There is a consistent downward trend in net revenue. This indicates potential issues such as declining sales, increasing competition, or inefficiencies in business operations.

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

Solution Query:

select quarter_number,

SUM(quantity * vehicle_price) AS gross_revenue,

SUM(vehicle_price * (quantity * (discount / 100))) AS total_discount,

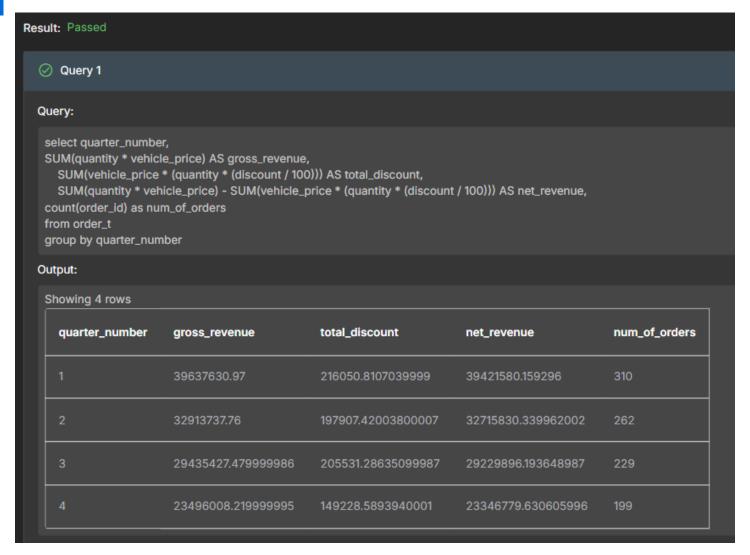
SUM(quantity * vehicle_price) - SUM(vehicle_price * (quantity * (discount / 100))) AS net_revenue,

count(order_id) as num_of_orders

from order t

group by quarter_number;





- Net revenue shows a consistent decline each quarter, with the steepest drop in Q4.
- The number of orders also exhibits a consistent decline across all quarters, with a relatively smaller but steady reduction.
- The decline in **net revenue** aligns closely with the drop in the **number of orders**, indicating fewer transactions as the primary driver of revenue decline.
- Both net revenue and the number of orders experienced their steepest declines in Q4, suggesting potential seasonal factors or business performance challenges during this period.

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



Solution Query:

order by 4 desc;

select c.customer_id,o.order_id,c.credit_card_type,
round(AVG(o.discount), 2) as Average_discount
from customer_t c
inner join order_t o
on c.customer_id = o.customer_id
group by c.credit_card_type

0555-0140

43353-914



Result: Passed Query 1 Query: select c.customer_id,o.order_id,c.credit_card_type, round(AVG(o.discount), 2) as Average_discount from customer_t c inner join order_t o on c.customer_id = o.customer_id group by c.credit_card_type order by 4 desc Output: Showing first 10 rows out of 16 rows customer_id credit_card_type Average_discount order_id 0172-3762 0363-0563 instapayment 0.77 0338-0520 49884-310 0.7 0363-0604 0093-5539 americanexpress 0179-0080 0264-9010 diners-club-enroute 0065-0643 0462-0395 0378-3005 0270-1314 diners-club-carte-blan... 0378-5504 0093-0248 0.64 visa-electron 0006-0740 16590-774 0023-4385 0074-4341

china-unionpay



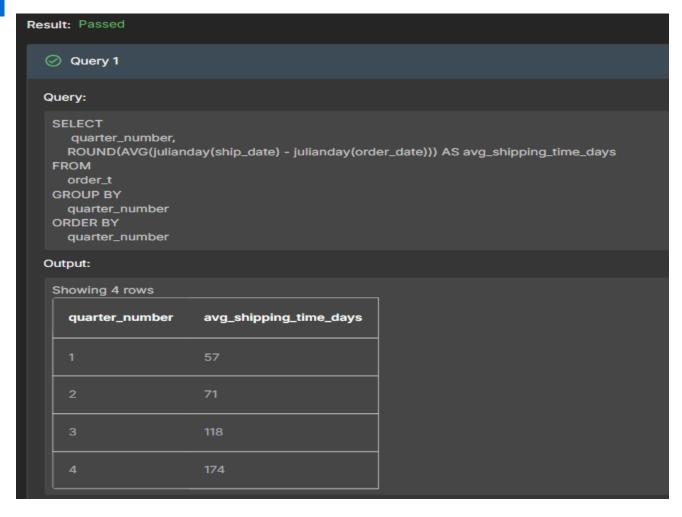
- **Instapayment** leads with the highest average discount (0.77), possibly indicating targeted promotions or rewards for this card type.
- Cards like Solo, American Express, and Diners Club Enroute receive significant discounts (above 0.67), suggesting a focus on incentivizing transactions from high-value customer segments.
- Cards such as **Mastercard**, **Visa Electron**, and **Maestro** have a consistent average discount around 0.64-0.65, indicating similar treatment for mainstream payment methods.
- Cards like Laser and China UnionPay receive the lowest discounts (0.62), possibly reflecting a lower priority for promotions targeting these card types.

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))) AS avg_shipping_time_days FROM order_t GROUP BY quarter_number ORDER BY quarter_number;







Insights and observation:

- Q1 had the least average shipping time of 57 days comparatively
- The average shipping time increased significantly each quarter, with the highest jump occurring in **Q3**. This suggests a growing delay in the fulfillment process.
- Q4 had the highest average shipping time of 174 days, nearly 3x the time in Q1. This could indicate inefficiencies or capacity issues during the later months.
- As shipping delays increase, they may negatively impact customer satisfaction and loyalty, contributing to declining orders and revenue observed previously.





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
23346779.630605996	199	98	21.5

Business Recommendations

- Address increasing delays by improving logistics, automating fulfillment, and offering expedited shipping options.
- Implement loyalty programs, personalized offers, and targeted campaigns to re-engage existing customers and drive repeat purchases.
- Offer discounts or promotions in states with fewer preferred brands (e.g., Alabama, Indiana).
- Promote brands in states like **California** and **Florida**, where brands like **Pontiac** and **Nissan** are popular.
- Focus on high-performing payment methods, align discounts with revenue impact, and introduce tiered incentives to maximize profitability.
- Ensure adequate inventory, launch targeted Q4 promotions, and proactively manage supply chain challenges.
- Monitor key performance metrics, segment customers effectively, and use predictive analytics to forecast demand and refine strategies.
- Investigate and resolve product issues impacting satisfaction.
- Establish continuous feedback loops from customers to product teams.
- Increase staffing and resources in Q3 and Q4 to handle demand surges.
- Use predictive analytics to forecast and prepare for high-demand periods.
- Be transparent with customers about service improvements.
- Engage dissatisfied customers with personalized recovery actions.
- Conduct quarterly reviews to track customer feedback trends.
- Implement real-time dashboards for immediate performance insights.
- Stay updated on customer preferences and change strategies accordingly.

