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:

- **Finding the total number of customers who placed orders**

```
SELECT
    COUNT(DISTINCT c.customer_id) AS total_customers_with_orders
FROM customer_t c
INNER JOIN order_t o ON c.customer_id = o.customer_id;
```

- **Finding the distribution of customers across states who placed orders**

```
SELECT
    c.state,
    COUNT(DISTINCT c.customer_id) AS number_of_customers
FROM customer_t c
INNER JOIN order_t o ON c.customer_id = o.customer_id
GROUP BY c.state;
```

Output:

- **Output of the total number of customers who placed orders**

total_customers_with_orders
994

- **Output of the distribution of customers across states who placed orders**

state	number_of_customers
Alabama	29
Alaska	10
Arizona	26
Arkansas	6
California	97
Colorado	33
Connecticut	22
Delaware	6
District of Columbia	35
Florida	86
Georgia	18

Observations and Insights:

- There are 994 customers who have placed orders.
- California (97) and Florida (86) have the highest number of customers, while Arkansas (6) and Delaware (6) have the lowest.
- Some states (e.g., California, Florida, and Colorado) have a larger customer base, potentially due to high population density or demand for the company's products.
- Focusing markets on California and Florida for higher profitability, while targeting Arkansas and Delaware for expansion through brand awareness and incentives.

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

Solution Query:

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

Output:

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

Observations and Insights:

1. Chevrolet (83) and Ford (63) are the top two preferred vehicle makers, followed by Toyota (52), Pontiac (50), and Dodge (50).
2. American brands dominate customer preference, with Chevrolet, Ford, Pontiac, and Dodge making up 4 of the top 5 choices.
3. Stocking and promotional efforts should focus on these high-demand brands to maximize sales and customer satisfaction.

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

Solution Query:

```
WITH vehicle_rank AS (
   SELECT
      c.state,
      p.vehicle_maker,
      RANK() OVER (
         PARTITION BY c.state
         ORDER BY COUNT(DISTINCT o.customer_id) DESC
      ) AS rn
   FROM order_t o
   JOIN product_t p
      ON o.product_id = p.product_id
   JOIN customer_t c
      ON o.customer_id = c.customer_id
   GROUP BY c.state, p.vehicle_maker
)
SELECT
   state,
   vehicle_maker
FROM vehicle_rank
WHERE rn = 1;
```

Output:

	state	vehicle_maker
	Arkansas	Volkswagen
	California	Audi
	California	Chevrolet
	California	Dodge
	California	Ford
	California	Nissan
	Colorado	Chevrolet
	Connect...	Chevrolet
	Connect...	Maserati
	Connect...	Mercury
	Connect...	Volvo
	Delaware	Mitsubishi
	District ...	Chevrolet
	Florida	Toyota
	Georgia	Toyota

Observations and Insights:

- Chevrolet is the most preferred brand in multiple states, including California, Colorado, Connecticut, and the District of Columbia, indicating strong nationwide demand.
- California has diverse vehicle preferences, with multiple brands (Audi, Chevrolet, Dodge, Ford, Nissan) making it a key market for various automakers.
- Different states prefer different brands, suggesting that regional marketing strategies should focus on the most popular brands in each area to maximize sales.

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:

- **Finding the overall average rating given by the customers**

```
SELECT
  AVG(rating_value) AS overall_average_rating
FROM (
  SELECT
    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
      ELSE NULL
    END AS rating_value
  FROM order_t
) AS ratings;
```

- **Finding Avg rating given by the customers in each quarter**

```
SELECT
  quarter_number,
  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
        ELSE NULL
    END
  ) AS average_rating
FROM order_t
WHERE customer_feedback IS NOT NULL
GROUP BY quarter_number
ORDER BY quarter_number;
```

Output:

- **Output of the overall average rating given by the customers**

overall_average_rating
3.1350

- **Output of the Avg rating given by the customers in each quarter**

quarter_number	average_rating
1	3.5548
2	3.3550
3	2.9563
4	2.3970

Observations and Insights:

- The overall average rating is 3.135, which indicates a neutral to slightly positive feedback from customers.

- The ratings have been declining over time, starting from 3.55 in Q1 to 2.39 in Q4, showing a significant drop in customer satisfaction.
- The downward trend suggests that customers might be facing increasing issues with service or product quality, leading to lower ratings.

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

Solution Query:

```
SELECT
    quarter_number,
    COUNT(CASE WHEN customer_feedback = 'Very Bad' THEN 1 END) * 100.0 / COUNT(*) AS percent_very_bad,
    COUNT(CASE WHEN customer_feedback = 'Bad' THEN 1 END) * 100.0 / COUNT(*) AS percent_bad,
    COUNT(CASE WHEN customer_feedback = 'Okay' THEN 1 END) * 100.0 / COUNT(*) AS percent_okay,
    COUNT(CASE WHEN customer_feedback = 'Good' THEN 1 END) * 100.0 / COUNT(*) AS percent_good,
    COUNT(CASE WHEN customer_feedback = 'Very Good' THEN 1 END) * 100.0 / COUNT(*) AS percent_very_good
FROM order_t
GROUP BY quarter_number
ORDER BY quarter_number;
```

Output:

quarter_number	percent_very_bad	percent_bad	percent_okay	percent_good	percent_very_good
1	10.96774	11.29032	19.03226	28.70968	30.00000
2	14.88550	14.12214	20.22901	22.13740	28.62595
3	17.90393	22.70742	21.83406	20.96070	16.59389
4	30.65327	29.14573	20.10050	10.05025	10.05025

Observations and Insights:

- The percentage of "very bad" and "bad" ratings has increased every quarter, reaching 30.65% and 29.14% in Q4.
- The percentage of "good" and "very good" ratings has dropped significantly, from 30% (very good) in Q1 to just 10% in Q4.
- Overall, more customers are unhappy with the service/product as the year progresses. Immediate action is needed to understand and fix the issues.

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:

quarter_number	total_orders
1	310
2	262
3	229
4	199

Observations and Insights:

- Orders have been steadily decreasing each quarter, with a 35.8% total drop from Q1 (310 orders) to Q4 (199 orders).
- The decline suggests fewer customers are buying, possibly due to bad reviews, pricing issues, or competition.
- The company should investigate customer satisfaction, pricing strategy and marketing efforts.

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

Solution Query:

```
WITH revenue_calc AS (
   SELECT
      quarter_number,
      SUM(quantity * (vehicle_price - discount)) AS net_revenue
   FROM order_t
   GROUP BY quarter_number
),
qoq_change AS (
   SELECT
      quarter_number,
      net_revenue,
      LAG(net_revenue) OVER (ORDER BY quarter_number) AS prev_quarter_revenue,
      ROUND(
         ( (net_revenue - LAG(net_revenue) OVER (ORDER BY quarter_number)) /
          NULLIF(LAG(net_revenue) OVER (ORDER BY quarter_number), 0) ) * 100,
         2
      ) AS percent_change
   FROM revenue_calc
)
SELECT
   quarter_number,
   net_revenue,
   COALESCE(percent_change, 0) AS percent_change
```

```
FROM qoq_change

ORDER BY quarter_number;
```

Output:

quarter_number	net_revenue	percent_change
1	39637378.16	0.00
2	32913497.44	-16.96
3	29435188.49	-10.57
4	23495814.14	-20.18

Observations and Insights:

- Revenue is dropping every quarter, with a total decline of about 40.7% from Q1 to Q4
- The biggest drop happened in Q4 (-20.18%), showing a significant loss in earnings.
- The steady decline in revenue aligns with the drop in total orders, suggesting fewer sales are driving the loss.
- If this trend continues, the company may face serious financial challenges, so urgent action is needed to reverse the decline.

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

Solution Query:

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

Output:

quarter_number	net_revenue	total_orders
1	39637378.16	310
2	32913497.44	262
3	29435188.49	229
4	23495814.14	199

Observations and Insights:

- Both net revenue and total orders are decreasing each quarter, showing a consistent downward trend in sales.
- The decline in orders is directly affecting revenue, meaning fewer customers are buying vehicles.
- If this trend continues, the company's profitability will be at risk, requiring urgent intervention.
- The revenue per order seems to be fairly stable, suggesting that the drop in revenue is mainly due to fewer sales rather than lower vehicle prices.

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

Solution Query:

```
SELECT
    c.credit_card_type,
    (AVG(o.discount) * 100) AS avg_discount
FROM customer_t c
JOIN order_t o ON c.customer_id = o.customer_id
GROUP BY c.credit_card_type;
```

Output:

credit_card_type	avg_discount
laser	64.384615
china-unionpay	62.217391
diners-club-enroute	59.979167
americanexpress	61.632653
mastercard	62.950000
visa	60.083333
bankcard	60.954545
solo	58.500000
maestro	62.421875
diners-club-us-ca	61.461538
instapayment	62.062500
diners-club-international	58.400000

Observations and Insights:

- The highest average discount is given for Laser cards (64.38%)
- Popular cards like Visa, Mastercard, and American Express receive discounts between 60% and 63%, which is fairly competitive.
- Most credit card types receive discounts in the 58–63% range, meaning there isn't a huge variation in discounts.

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

Solution Query:

```
SELECT
   quarter_number,
   AVG(TIMESTAMPDIFF(DAY, order_date, ship_date)) AS avg_shipping_time
FROM order_t
WHERE ship_date IS NOT NULL
GROUP BY quarter_number
ORDER BY quarter_number;
```

Output:

quarter_number	avg_shipping_time
1	57.1677
2	71.1107
3	117.7555
4	174.0955

Observations and Insights:

- The average shipping time has increased every quarter, from 57.17 hours in Q1 to 174.10 hours in Q4.
- Q3 and Q4 saw the sharpest increases in shipping time, suggesting a worsening problem that needs urgent attention.
- The company should analyze regional delays to see if certain areas face more shipping issues than others.

Business Metrics Overview

Total Revenue	Total Orders	Total Customers	Average Rating
125481878.23	1000	994	3.1350
Last Quarter Revenue	**Last quarter Orders**	**Average Days to Ship**	**% Good Feedback**
23495814.14	199	97.96	44.10

Business Recommendations

- Customer ratings are dropping each quarter. So offering better customer services, resolving complaints quickly, and introducing loyalty to retain customers.
- Shipping delays are increasing, especially in later quarters which results in less repeat customers. So working with more delivery partners, optimize logistics, and opening regional warehouses to reduce wait times.
- Chevrolet, Ford, and Toyota have the highest demand. So Focusing more on increasing inventory for these type of brands to boost sales and meet customer preferences.
- Revenue has been slowing down over time. Boost sales by introducing trade-in programs, flexible financing options, seasonal discounts, and marketing campaigns to attract more buyers.
- Some vehicles sell faster than others, and demand changes every quarter. Use past sales data to predict demand and adjust stock levels, ensuring the right mix of vehicles is available at the right time.