

# New Wheels Project Introduction to SQL

Partha Sarathy Mohapatra

#### **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:**

#### **Unique Customer orders**

select count(distinct customer\_id) as Unique\_Customer\_Orders from order\_t;

#### **Total Customer orders**

select count(order\_id) as Total\_Customer\_Orders from order\_t;

#### Distribution of the customers across states

select state, count (customer\_id) as No\_Of\_Customers from customer\_t group by state;

#### **Output:**

```
mysql> select count(distinct customer_id) as Unique_Customer_Orders from order_t;
+------+
| Unique_Customer_Orders |
+-----+
| 994 |
+-----+
1 row in set (0.00 sec)
```

```
mysql> select count(order_id) as Total_Customer_Orders from order_t;
+------+
| Total_Customer_Orders |
+------+
| 1000 |
+------+
1 row in set (0.02 sec)
```



| mysql> select state,cou | nt(customer_id) as | No_Of_Customers | from | customer_t | group | by | state; |
|-------------------------|--------------------|-----------------|------|------------|-------|----|--------|
| state                   | No_Of_Customers    |                 |      |            |       |    |        |
| New Jersey              | 9                  |                 |      |            |       |    |        |
| Indiana                 | 21                 |                 |      |            |       |    |        |
| Washington              | 28                 |                 |      |            |       |    |        |
| California              | 97                 |                 |      |            |       |    |        |
| District of Columbia    | 35                 |                 |      |            |       |    |        |
| Texas                   | 97                 |                 |      |            |       |    |        |
| Virginia                | 24                 |                 |      |            |       |    |        |
| Illinois                | 25                 |                 |      |            |       |    |        |
| Florida                 | 86                 |                 |      |            |       |    |        |
| Colorado                | 33                 |                 |      |            |       |    |        |
| Iowa                    | 11                 |                 |      |            |       |    |        |
| Alabama                 | 29                 |                 |      |            |       |    |        |

| mysql> select state,cou | nt(customer_id) as | No_Of_Customers | from customer_t | group by st | tate order by | / No_Of_Customer | s desc; |
|-------------------------|--------------------|-----------------|-----------------|-------------|---------------|------------------|---------|
| state                   | No_Of_Customers    |                 |                 |             |               |                  |         |
| <br>  California        | 97                 |                 |                 |             |               |                  |         |
| Texas                   | 97                 |                 |                 |             |               |                  |         |
| Florida                 | 86                 |                 |                 |             |               |                  |         |
| New York                | j 69 j             |                 |                 |             |               |                  |         |
| District of Columbia    | j 35 j             |                 |                 |             |               |                  |         |
| Colorado                | j 33 j             |                 |                 |             |               |                  |         |
| Objo                    | j 22 j             |                 |                 |             |               |                  |         |

- Total 1000 orders by customers.
- Out of which 994 unique customers have ordered.
- California, Texas and Florida are the top 3 states with highest customers

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



#### customers?

#### **Solution Query:**

select vehicle\_maker,count(vehicle\_maker) as count from product\_t where product\_id in (select product\_id from order\_t) group by vehicle\_maker order by count desc limit 10;

#### Output:



- Chevrolet, Ford, Toyota are the topmost preferred vehicle makers.
- Mercedes-Benz, GMC, Buick are also among the top 10 preferred vehicle makers.
- MG,Citroen, Austin and Ram are the least preferred vehicle makers.



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

#### state?

#### **Solution Query:**

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

#### **Output:**

| +          | +             | ·+           |
|------------|---------------|--------------|
| state      | vehicle_maker | total_orders |
| Alabama    | <br>  Dodge   | 5            |
| Alaska     | Chevrolet     | 2            |
| Arizona    | Pontiac       | 3            |
| Arizona    | Cadillac      | 3            |
| Arkansas   | GMC           | 1            |
| Arkansas   | Mitsubishi    | 1            |
| Arkansas   | Suzuki        | 1            |
| Arkansas   | Volkswagen    | 1            |
| Arkansas   | Chevrolet     | 1            |
| Arkansas   | Pontiac       | 1            |
| California | Chevrolet     | 6            |



- There are many states with multiple vehicle makers having same no of orders.
- Chevrolet, Dodge and Pontiac are the highest ranked vehicles in 17,12 and 11 states respectively.
- Jaguar, Maybach and Oldsmobile are highest ranked vehicles in one state only.



# Question 4: Find the overall average rating given by the customers.

### What is the average rating in each quarter?

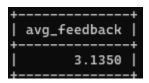
Consider the following mapping for ratings: "Very Bad": 1, "Bad": 2, "Okay": 3, "Good": 4, "Very Good": 5 Solution Query:

#### **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_feedback
      FROM
        order_t;
Avg rating per quarter
      SELECT
        quarter_number,
        AVG(
          CASE
            WHEN customer_feedback = 'Very Bad' THEN 1
```

```
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
END
) AS avg_feedback
FROM
order_t
```

#### **Output:**



**GROUP BY** 

quarter\_number;



```
mysql> 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
    ->
              END
            ) AS avg_feedback
    ->
    -> FROM order_t
    -> GROUP BY quarter_number;
  quarter_number | avg_feedback
                 4
                            2.3970
                 1
                            3.5548
                 3
                            2.9563
                            3.3550
4 rows in set (0.00 sec)
```

- Avg rating by all customers is 3.135
- Avg rating is highest in the 1<sup>st</sup> and 2<sup>nd</sup> quarter
- Avg rating is the lowest in the 4<sup>th</sup> quarter



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

#### **Solution Query:**

```
SELECT
quarter_number,
customer_feedback,
COUNT(customer_feedback) * 100.0 /
    (SELECT COUNT(*) FROM order_t WHERE quarter_number = o.quarter_number) AS
feedback_percentage
FROM order_t o
GROUP BY quarter_number, customer_feedback
ORDER BY quarter_number, feedback_percentage DESC;
```

#### Output:

| +              |                   |                     |
|----------------|-------------------|---------------------|
| quarter_number | customer_feedback | feedback_percentage |
| 1              | Very Good         | 30.00000            |
| 1              | Good              | 28.70968            |
| 1              | 0kay              | 19.03226            |
| 1              | Bad               | 11.29032            |
| 1              | Very Bad          | 10.96774            |
| j 2 i          | Very Good         | 28.62595            |
| j 2 i          | Good              | 22.13740            |
| 2              | 0kay              | 20.22901            |
| 2              | Very Bad          | 14.88550            |
| 2              | Bad               | 14.12214            |
| 3              | Bad               | 22.70742            |
| 3              | 0kay              | 21.83406            |
| 3              | Good              | 20.96070            |
| 3              | Very Bad          | 17.90393            |
| 3              | Very Good         | 16.59389            |
| 4              | Very Bad          | 30.65327            |
| 4              | Bad               | 29.14573            |
| 4              | 0kay              | 20.10050            |
| 4              | Good              | 10.05025            |
| 4              | Very Good         | 10.05025            |
| +              | 05 sec)           | ++                  |

20 rows in set (0.05 sec)

- Over every quarter the good feedback is going downhill.
- Negative feedback seems to be increasing

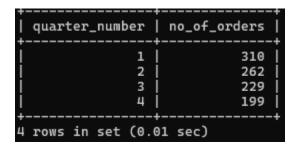




#### **Solution Query:**

select quarter\_number,count(order\_id) as no\_of\_orders from order\_t group by quarter\_number order by quarter\_number;

#### **Output:**



- There seems to be falling orders every quarter
- Q1 with 310 and Q4 with 199 orders.



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

#### **Solution Query:**

#### **Total Revenue**

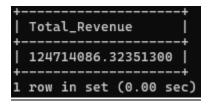
select sum((vehicle\_price\*(1-(discount/100)))\*quantity) as Total\_Revenue from order\_t;

#### QoQ % Change

```
SELECT
quarter_number,
net_revenue,
LAG(net_revenue) OVER (ORDER BY quarter_number) AS previous_quarter_revenue,
(net_revenue - LAG(net_revenue) OVER (ORDER BY quarter_number)) / LAG(net_revenue) OVER
(ORDER BY quarter_number) * 100 AS qoq_percentage_change
FROM (
SELECT
quarter_number,
SUM(quantity * (vehicle_price*(1-(discount/100)))) AS net_revenue
FROM order_t
GROUP BY quarter_number
) AS revenue_per_quarter;
```

#### Output:

#### **Total Revenue**



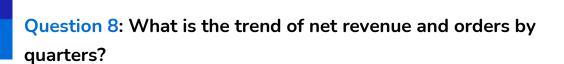
#### QoQ % Change

| +- | quarter_number   | net_revenue  | previous_quarter_revenue  | ++<br>  qoq_percentage_change  |  |  |
|----|------------------|--|---|--|--|--|
|    | 1<br>2<br>3<br>4 | 39421580.15929600<br>32715830.33996200<br>29229896.19364900<br>23346779.63060600 | NULL<br>39421580.15929600<br>32715830.33996200<br>29229896.19364900 | NULL  <br>  -17.010352685603  <br>  -10.655190805458  <br>  -20.127052535757 |  |  |
| 4  | +                |  |   |  |  |  |





- The revenue growth seems to be decreasing after every quarter
- Q4 seems to have the highest decline in revenue.





### **Solution Query:**

select quarter\_number, COUNT(order\_id) AS total\_orders, SUM(quantity\*vehicle\_price\*(1-(discount/100))) AS net\_revenue from order\_t group by quarter\_number order by quarter\_number;

#### Output:

| quarter_number           | total_orders             | net_revenue  |  |  |
|--------------------------|--------------------------|--|--|--|
| 1<br>2<br>3<br>4         | 310<br>262<br>229<br>199 | 39421580.15929600  <br>  32715830.33996200  <br>  29229896.19364900  <br>  23346779.63060600 |  |  |
| 4 rows in set (0.00 sec) |                          |  |  |  |

#### Observations and Insights:

• Revenue and Orders both are failing for every quarter.



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

#### **Solution Query:**

select credit\_card\_type, avg(o.discount) as avg\_discount from customer\_t c inner join order\_t o on o.customer\_id=c.customer\_id group by c.credit\_card\_type order by avg\_discount;

#### Output:

| credit_card_type   | avg_discount   |
|--|--|
| diners-club-international   solo   diners-club-enroute   visa   jcb   bankcard   switch   diners-club-carte-blanche   diners-club-us-ca   americanexpress   instapayment   china-unionpay   visa-electron   maestro   mastercard | 0.584000   0.584000   0.585000   0.599792   0.600833   0.607382   0.609545   0.610233   0.614490   0.614615   0.616327   0.620625   0.623469   0.624219   0.629500 |
| laser<br>+   | 0.643846   |
| 16 rows in set (0.01 sec)  |  |

- Diners-club-international gives the lowest discount followed by solo card.
- Laser card followed by mastercard provides the highest discount.

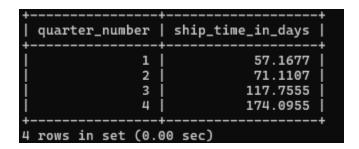


# Question 10: What is the average time taken to ship the placed orders for each quarter?

#### **Solution Query:**

select quarter\_number, avg(datediff(ship\_date,order\_date)) as ship\_time\_in\_days from order\_t group by quarter\_number order by quarter\_number;

#### **Output:**



#### Observations and Insights:

Avg Shipping days has increased from 2 months to 6 months from Q1 to Q4

# **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 |
| 23346779.63          | 199                 | 97.96                | 44.1            |

# **Business Recommendations**

- New wheels need to improve the shipping time of the orders.
- Company needs to focus more on understanding failing satisfaction level of customers and act on it.
- Company should do a market survey to understand the falling sales.

