```
/*
```

The provided document is a guide for the project. Follow the instructions and take the necessary steps to finish

the project in the SQL file

                                                Queries

--*/

```sql
USE vehdb;


-- A preliminary study indicates the presence of duplicates in a customer_id column of order_t table:
SELECT
    COUNT(DISTINCT shipper_id),
    COUNT(DISTINCT product_id),
    COUNT(DISTINCT order_id),
    COUNT(customer_id),
    COUNT(DISTINCT customer_id)
FROM
    order_t AS o
        JOIN
    shipper_t AS s USING (shipper_id)
        JOIN
    product_t AS p USING (product_id)
        JOIN
    customer_t AS c USING (customer_id);
```

-- Apparently, this means that some customers made several orders but with different vehicles, because the product_id and order_id columns don't have any duplicates.

-- So, we find these customers:

```sql
SELECT
    *
FROM
    order_t
GROUP BY customer_id
HAVING COUNT(customer_id) > 1;
```

-- --

```
/*-- QUESTIONS RELATED TO CUSTOMERS

    [Q1] What is the distribution of customers across states?

    Hint: For each state, count the number of customers.*/
```

```sql
SELECT
    state,
    COUNT(DISTINCT customer_id) AS number_of_customers_state
FROM
    customer_t
GROUP BY 1
ORDER BY 2 DESC;
```

```
/* From here it can be seen that the four states, where the company's services are most in demand, made about 70 or more orders each.

And this is twice or more than the number of orders from other states */
```

-- --

```
/* [Q2] What is the average rating in each quarter?

-- Very Bad is 1, Bad is 2, Okay is 3, Good is 4, Very Good is 5.
```

Hint: Use a common table expression and in that CTE, assign numbers to the different customer ratings.

Now average the feedback for each quarter.

Note: For reference, refer to question number 4. Week-2: mls_week-2_gl-beats_solution-1.sql.

You'll get an overview of how to use common table expressions from this question.*/

WITH assign_ratings AS (

SELECT order_id,

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

ELSE 5

END AS numeric_ratings

FROM order_t

)

SELECT quarter_number,

AVG(numeric_ratings) AS average_numeric_ratings

FROM assign_ratings

GROUP BY 1

ORDER BY 1;

/* From the latter, we can see that gradually, from quarter 1 to quarter 4, the company's average rating is decreasing.

This indicates a significant increase in negative customer feedbacks */

-- --

/* [Q3] Are customers getting more dissatisfied over time?

Hint: Need the percentage of different types of customer feedback in each quarter. Use a common table expression and

determine the number of customer feedback in each category as well as the total number of customer feedback in each quarter.

Now use that common table expression to find out the percentage of different types of customer feedback in each quarter.

Eg: (total number of very good feedback/total customer feedback)* 100 gives you the percentage of very good feedback.

Note: For reference, refer to question number 4. Week-2: mls_week-2_gl-beats_solution-1.sql.

You'll get an overview of how to use common table expressions from this question.*/


```sql
WITH assign_ratings AS (

SELECT order_id,

    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

      ELSE 5

    END AS numeric_ratings

FROM order_t

),


count_ratings AS (

SELECT quarter_number,

    numeric_ratings,

    COUNT(numeric_ratings) AS number_of_ratings

FROM assign_ratings

GROUP BY 1,2

ORDER BY 1

)
```

```sql
SELECT quarter_number,
    numeric_ratings,
        number_of_ratings,
    CASE
        WHEN quarter_number = 1 THEN (number_of_ratings / (SELECT SUM(number_of_ratings) FROM count_ratings WHERE quarter_number = 1))*100
        WHEN quarter_number = 2 THEN (number_of_ratings / (SELECT SUM(number_of_ratings) FROM count_ratings WHERE quarter_number = 2))*100
        WHEN quarter_number = 3 THEN (number_of_ratings / (SELECT SUM(number_of_ratings) FROM count_ratings WHERE quarter_number = 3))*100
        ELSE (number_of_ratings / (SELECT SUM(number_of_ratings) FROM count_ratings WHERE quarter_number = 4))*100
    END AS percentage_in_quarter
FROM count_ratings;


-- From the last output, we can see that over the course of the year, customers become more and more dissatisfied.
-- For example, if in the 1st quarter the number of "Very bad" feedbacks was about 11%, then in the 4th quarter the same value is already about 31%.

----------------------------------------------------------------------------------------------------------------


-- The next query is to determine the overall average annual percentage of "Good" and "Very Good" feedbacks for "Business Overview":
WITH assign_ratings AS (
SELECT order_id,
    customer_feedback,
    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
        ELSE 5
    END AS numeric_ratings
FROM order_t
```

```
)

SELECT numeric_ratings,

    customer_feedback,

    COUNT(numeric_ratings) AS number_of_ratings,

    ((COUNT(numeric_ratings) / (SELECT COUNT(numeric_ratings) FROM assign_ratings))*100) AS
percentage_of_ratings

FROM assign_ratings

GROUP BY 1

HAVING numeric_ratings IN (4, 5)

ORDER BY 1;


-- Based on the latter, the value of the total average annual percentage of "Good" and "Very Good" feedbacks is 44.1%

-- -----------------------------------------------------------------------------------------------------------------------------------


/*[Q4] Which are the top 5 vehicle makers preferred by the customer.


Hint: For each vehicle maker what is the count of the customers.*/


--  This statement is auxiliary, to understand the statistics of the data
SELECT
    COUNT(DISTINCT order_id) number_of_orders,
    MIN(quantity) min_quantity_vehicles_per_order,
    MAX(quantity) max_quantity_vehicles_per_order,
    (SELECT
        COUNT(DISTINCT order_id)
     FROM
        order_t
     WHERE
        quantity = 1) number_of_orders_1_vehicle,
    (SELECT
        COUNT(DISTINCT order_id)
```

```
      FROM

         order_t

      WHERE

         quantity = 2) number_of_orders_2_vehicles,

   AVG(quantity) average_quantity_vehicles_per_order,

   SUM(quantity) total_quantity_ordered_vehicles

FROM

   order_t;
```

/*Explanations for question 4. As we can see from the previous output, more than half of the individual orders (512) include 2 cars ordered, not just one.

Therefore, it will be objective to evaluate which vehicle maker customers prefer not only by the number of customers themselves and their orders,

but also by the number of ordered cars. */

-- The next two statements are the answers to the question. So, which are the top 5 vehicle makers preferred by the customer?

```
SELECT

   p.vehicle_maker,

   SUM(o.quantity) AS number_of_ordered_vehicles

FROM

   order_t AS o

      JOIN

   product_t AS p USING (product_id)

GROUP BY 1

ORDER BY 2 DESC

LIMIT 5;


SELECT

   p.vehicle_maker,

   COUNT(o.customer_id) AS number_of_customers

FROM

   order_t AS o
```

```
    JOIN
  product_t AS p USING (product_id)
GROUP BY 1
ORDER BY 2 DESC
LIMIT 5;
```

/* Based on the results of the last two queries, we can observe that the top 5 vehicle makers maintain their positions both

in terms of the number of customers and the number of ordered cars. */

-- ---

/*[Q5] What is the most preferred vehicle maker in each state?

Hint: Use the window function RANK() to rank based on the count of customers for each state and vehicle maker.

After ranking, take the vehicle maker whose rank is 1.*/

-- The next statement is for a deeper understanding of the data (ranking of each vehicle maker by customers across all states)

```
WITH count_customers AS (
SELECT p.vehicle_maker,
    COUNT(o.customer_id) AS number_of_customers
FROM
  order_t AS o JOIN product_t AS p USING (product_id)
GROUP BY 1
)

SELECT vehicle_maker,
    number_of_customers,
    RANK() OVER (ORDER BY number_of_customers DESC) AS customers_rank_overall
FROM count_customers;
```

```sql
/* The next statement is for a deeper understanding of the data too (calculation of number of customers and vehicle makers for each state,

and also ranking of each state by overall customers) */

WITH prepare_t AS (

SELECT p.vehicle_maker,

    c.customer_id,

    c.state

FROM order_t AS o

    JOIN product_t AS p USING (product_id)

    JOIN customer_t AS c USING (customer_id)

),


count_t AS (

SELECT state,

    COUNT(DISTINCT customer_id) AS number_of_customers_state,

    COUNT(DISTINCT vehicle_maker) AS number_of_makers_from_orders

FROM prepare_t

GROUP BY 1

)


SELECT state,

    number_of_customers_state,

    RANK() OVER (ORDER BY number_of_customers_state DESC) AS state_rank,

    number_of_makers_from_orders

FROM count_t;
-----------------------------------------------------------------------------
-- The next statement is the answer to the question. So, what is the most preferred vehicle maker in each state?

WITH prepare_t AS (

SELECT c.state,

        p.vehicle_maker,

    COUNT(c.customer_id) OVER (PARTITION BY vehicle_maker, state) AS number_customers_per_maker,

    c.customer_id
```

```
FROM order_t AS o

    JOIN product_t AS p USING (product_id)

    JOIN customer_t AS c USING (customer_id)

ORDER BY 1

),


rank_t AS (

SELECT state,

    vehicle_maker,

    number_customers_per_maker,

    RANK() OVER (PARTITION BY state ORDER BY number_customers_per_maker DESC) AS maker_rank_per_state

FROM prepare_t

ORDER BY 1

)


SELECT state,

        vehicle_maker,

    number_customers_per_maker,

    maker_rank_per_state

FROM rank_t

GROUP BY 1

HAVING maker_rank_per_state = 1

ORDER BY 1;
```

-- --
-

/*QUESTIONS RELATED TO REVENUE and ORDERS

-- [Q6] What is the trend of number of orders by quarters?

Hint: Count the number of orders for each quarter.*/

```sql
SELECT

    quarter_number,

    COUNT(DISTINCT order_id) AS number_of_orders

FROM

    order_t

GROUP BY 1

ORDER BY 1;
```

-- Based on the results of the last query, we can observe a constantly falling trend in the number of orders from quarter to quarter.

-- --

/* [Q7] What is the quarter over quarter % change in revenue?

Hint: Quarter over Quarter percentage change in revenue means what is the change in revenue from the subsequent quarter to the previous quarter in percentage.

    To calculate you need to use the common table expression to find out the sum of revenue for each quarter.

    Then use that CTE along with the LAG function to calculate the QoQ percentage change in revenue.

*/


```sql
--  This statement is auxiliary, to understand the statistics of the data

WITH calc_t AS (

SELECT vehicle_price,

        discount,

        quantity,

        ROUND((vehicle_price * discount / 100),2) AS discount_per_vehicle,

        (vehicle_price*quantity) AS revenue,

        ((vehicle_price - ROUND((vehicle_price * discount / 100),2))*quantity) AS revenue_discounted

FROM order_t

)


SELECT SUM(quantity) AS quantity_of_ordered_vehicles,

        SUM(discount_per_vehicle*quantity) AS discount_total,
```

```sql
        SUM(revenue) AS revenue_total,

        SUM(revenue_discounted) AS revenue_discounted_total
FROM calc_t;


-- In subsequent statements revenue is considered as revenue, taking into account the discount (previously,
revenue_discounted)

-- The next statement is the answer to the question. So, what is the quarter over quarter % change in revenue?

WITH calc_revenue AS (

SELECT quarter_number,

        SUM((vehicle_price - ROUND((vehicle_price * discount / 100),2))*quantity) AS revenue_quarter
FROM order_t

GROUP BY 1

ORDER BY 1

),


previous_quarter AS (

SELECT quarter_number,

        revenue_quarter,

                LAG(revenue_quarter, 1) OVER (ORDER BY quarter_number) AS previous_quarter_revenue
FROM calc_revenue

)


SELECT quarter_number,

        revenue_quarter,

        previous_quarter_revenue,

        (ROUND(((((revenue_quarter - previous_quarter_revenue) / previous_quarter_revenue)*100),3)) AS
revenue_change_percentage

FROM previous_quarter

ORDER BY quarter_number;


/* Based on the results of the last query, we can see that the revenue is constantly decreasing from quarter to quarter.

The drop in the 4th quarter, by 20%, is especially noticeable. */
```

-- ---

/* [Q8] What is the trend of revenue and orders by quarters?

Hint: Find out the sum of revenue and count the number of orders for each quarter.*/

SELECT

   quarter_number,

   COUNT(DISTINCT order_id) AS number_of_orders,

   SUM((vehicle_price - ROUND((vehicle_price * discount / 100), 2)) * quantity) AS revenue_quarter

FROM

   order_t

GROUP BY 1

ORDER BY 1;

/* From the results of the last query, it is clear that the number of orders and revenue are naturally correlated with each other,

and maintain a constantly falling trend throughout the year. */

-- --

/* QUESTIONS RELATED TO SHIPPING

   [Q9] What is the average discount offered for different types of credit cards?

Hint: Find out the average of discount for each credit card type.*/

SELECT

   c.credit_card_type,

   AVG(o.discount) AS average_discount

FROM

   order_t AS o

      JOIN

   customer_t AS c USING (customer_id)

GROUP BY 1

ORDER BY 1;

-- Based on the results of the last query, we can observe that regardless of the type of credit card, discounts are in the range of 0.58-0.65%.

-- --

/* [Q10] What is the average time taken to ship the placed orders for each quarters?

        Hint: Use the DATEDIFF() function to find the difference between the ship date and the order date.

*/

SELECT

    quarter_number,

    ROUND(AVG(DATEDIFF(ship_date, order_date)), 2) AS average_shipping_in_days

FROM

    order_t

GROUP BY 1

ORDER BY 1;

/* From here we can see that from quarter to quarter the average number of delivery days is constantly growing.

And this value is 3 times more in the 4th quarter than in the 1st.

Most likely, we can conclude that this is one of, if not the main, reason for the growing dissatisfaction of customers,

and, accordingly, the decrease in the number of orders and the company's revenue. */

-- ---Done--

-- --

-- Additional questions

-- [Q11] What is the trend of average vehicle prices by quarters?

SELECT

    quarter_number,

    ROUND(AVG(vehicle_price), 2) AS average_vehicle_price

FROM

    order_t

```
GROUP BY 1

ORDER BY 1;


/* Based on the results of the last query, we can see that the drop in sales is unlikely to be related to vehicle prices.

Moreover, in the 4th quarter, where the lowest revenue was recorded, the average vehicle price decreased significantly
compared to the 1st quarter. */

-- ---------------------------------------------------------------------------------------------------------------------------


-- [Q12] Is there a relationship between a decrease in overall customer feedback and an increase in the number of
shipping days?


WITH assign_ratings AS (

SELECT order_id,

        customer_feedback,

        quarter_number,

        ship_date,

        order_date,

        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

          ELSE 5

        END AS numeric_ratings

FROM order_t

)


SELECT quarter_number,

        AVG(numeric_ratings) AS average_numeric_ratings,

      ROUND(AVG(DATEDIFF(ship_date, order_date)), 2) AS average_shipping_in_days

FROM assign_ratings

GROUP BY 1

ORDER BY 1;
```

/* Observation: apparently we can observe a correlation between a decrease in the overall buyer rating and an increase in the number of days of delivery.

Therefore, from the point of view of our business task, it is advisable to dive deeper into the study of this area. */

-- --

-- [Q13] How does the average shipping time (in days) change across the year depending on the ship mode?

```sql
SELECT
    quarter_number,
    ship_mode,
    ROUND(AVG(DATEDIFF(ship_date, order_date)), 1) AS average_shipping_in_days
FROM
    order_t
GROUP BY 1 , 2
ORDER BY 1;
```

/* Observation: the average number of shipping days has increased by almost 3 or more times across the year for all ship modes approximately the same.

But perhaps we should pay closer attention to the "First Class" and "Same Day" mode. And we will see it later in the visualization. */

-- --

-- [Q14] How does the average shipping time (in days) change across the year depending on the shipping?

```sql
SELECT
    quarter_number,
    shipping,
    ROUND(AVG(DATEDIFF(ship_date, order_date)), 1) AS average_shipping_in_days
FROM
    order_t
GROUP BY 1 , 2
ORDER BY 1;
```

-- Observation: the average number of shipping days has increased more than 3 times across the year for all types of shipping approximately the same.

-- --

-- [Q15] Orders of which vehicle makers (top 10) are shipped the longest (average, in days) across the year?

```sql
WITH calc_shipping AS (

SELECT p.vehicle_maker,

    ROUND(AVG(DATEDIFF(ship_date, order_date)), 1) AS average_shipping_in_days

FROM

   order_t AS o JOIN product_t AS p USING (product_id)

GROUP BY 1

)


SELECT vehicle_maker,

    average_shipping_in_days,

    RANK() OVER (ORDER BY average_shipping_in_days DESC) AS shipping_rank_overall

FROM calc_shipping

LIMIT 10;
```

-- --

-- [Q16] Which shippers (top 10) make the longest deliveries (average, in days) across the year?

-- [Q16.1] The next statement is for a deeper understanding of the data (ranking of each shipper by customers overall):

```sql
WITH count_customers AS (

SELECT s.shipper_name,

    s.shipper_id,

    COUNT(o.customer_id) AS number_of_customers

FROM

   order_t AS o JOIN shipper_t AS s USING (shipper_id)
```

```sql
GROUP BY 1

)


SELECT shipper_name,

    shipper_id,

    number_of_customers,

    RANK() OVER (ORDER BY number_of_customers DESC) AS customers_rank_overall

FROM count_customers;
```

-- [Q16.2] The next query is the answer to the question. So, which shippers (top 10) make the longest deliveries (average, in days) across the year?

```sql
WITH calc_shipping AS (

SELECT s.shipper_name,

    s.shipper_id,

    ROUND(AVG(DATEDIFF(ship_date, order_date)), 1) AS average_shipping_in_days

FROM

  order_t AS o JOIN shipper_t AS s USING (shipper_id)

GROUP BY 1

)


SELECT shipper_name,

    shipper_id,

    average_shipping_in_days,

    RANK() OVER (ORDER BY average_shipping_in_days DESC) AS shipping_rank_overall

FROM calc_shipping

LIMIT 10;
```

-- --

-- [Q17] Now the opposite question to the previous one. Which shippers (top 10) make the fastest deliveries (average, in days) across the year?

```sql
WITH calc_shipping AS (

SELECT s.shipper_name,
```

```sql
    s.shipper_id,

    ROUND(AVG(DATEDIFF(ship_date, order_date)), 1) AS average_shipping_in_days
FROM

  order_t AS o JOIN shipper_t AS s USING (shipper_id)
GROUP BY 1
)


SELECT shipper_name,

    shipper_id,

    average_shipping_in_days,

    RANK() OVER (ORDER BY average_shipping_in_days) AS shipping_rank_overall
FROM calc_shipping
LIMIT 10;
```

-- ---Done---

-- --