Question # 1

For each unique user in the dataset, find the latest date when their flags got reviewed. Then, find how many distinct videos were removed on that date.

Output the the first and last name of the user (in two columns), the date and the number of removed videos. Only include these users who had at least one of their flags reviewed. If no videos got removed on a certain date, output 0.

Tables: user_flags, flag_review

User_flags

User_firstname:varchar
User_lastname:varchar
Video_id:varchar
Flag_id:varchar

Flag_review

Flag_id:varchar
Reviewed_by_yt:bool
Reviewed_date:datetime
Reviewed_outcome:varchar

Explanation:

1. Filter the users first name, last name and latest date a video was reviewed by youtube. If it is not reviewed by youtube, the information will be filtered out

2. Aggregate the data by the date, and count how many unique videos were removed on each date

3. Join both tables on the date, and replace any null values with a 0 using the COALESCE function

4. Output: first name, last name, the date and # of removed videos

SQL:

```
WITH

latest_dates AS(

    SELECT user_firstname AS first_name,
        user_lastname AS last_name,
        MAX(reviewed_date) AS latest_date
    FROM user_flags AS f
    INNER JOIN flag_review AS r ON f.flag_id = r.flag_id
    WHERE r.reviewed_by_yt = 'TRUE'
    GROUP BY first_name, last_name

),

review_dates AS (

    SELECT fr.reviewed_date,
    COUNT(DISTINCT video_id) AS removed
    FROM flag_review AS fr
    INNER JOIN user_flags AS uf ON fr.flag_id = uf.flag_id
    WHERE reviewed_outcome = 'REMOVED'
    GROUP BY reviewed_date

)

    SELECT ld.*,
    COALESCE(r.removed, 0)
    FROM latest_dates AS ld
    LEFT JOIN review_dates AS r ON ld.latest_date = r.reviewed_date
```

first_name	last_name	latest_date	videos_removed
William	Kwan	2022-03-14	1
Daniel	Bell	2022-03-16	1
Gina	Korman	2022-03-17	2
Richard	Hasson	2022-03-18	0
Evelyn	Johnson	2022-03-17	2
Mark	May	2022-03-15	2
Mark	Johnson	2022-03-16	1
Pauline	Wilks	2022-03-17	2

Question # 2

Write a query to find the Market Share at the Product Brand level for each Territory, for Time Period Q4-2021. Market Share is the number of Products of a certain Product Brand brand sold in a territory, divided by the total number of Products sold in this Territory.

Output the ID of the Territory, name of the Product Brand and the corresponding Market Share in percentages. Only include these Product Brands that had at least one sale in a given territory.

Tables: fct_customer_sales, map_customer_territory, dim_product

Fct_customer_sales

Cust_id:varchar
Prod_sku_id:varchar
Order_date:datetime
Order_value:int
Order_id:varchar

Map_customer_territory

Cust_id:varchar
Territory_id:varchar

Dim_product

Prod_sku_id:varchar
Prod_sku_name:varchar
Prod_brand:varchar
market_name:varchar

SQL:

```
WITH

q4 AS (

    SELECT t.territory_id,
    p.prod_brand,
    COUNT(*) AS n_sales
    FROM fct_customer_sales AS s
    LEFT JOIN map_customer_territory AS t ON s.cust_id = t.cust_id
    LEFT JOIN dim_product AS p ON s.prod_sku_id = p.prod_sku_id
    WHERE EXTRACT(QUARTER FROM order_date) = 4
    AND EXTRACT(YEAR FROM order_date) = 2021
    GROUP BY t.territory_id, p.prod_brand, p.market_name

),

territories AS (


    SELECT territory_id,
    prod_brand,
    SUM(n_sales) AS sales
    FROM q4
    GROUP BY territory_id, prod_brand
    ORDER BY territory_id ASC


)

    SELECT territory_id,
    prod_brand,
    ROUND(sales / SUM(sales) OVER (PARTITION BY territory_id) * 100,3) AS market_share
    FROM territories
```

Explanation:

1. Join the map_customer_territory table to the customer sales table on the customer id, and the product table on the prod_sku_id. This creates a table with all of the information in one dataset called ‘q4’

2. Once the dataset is created, filter the data using the EXTRACT() function to get the 4th quarter in the year 2021 and get the territory, brand, and # of sales

3. In a second table, aggregate the data by territory and prod_brand, and add the total sales

4. To get the result, get the territory_id, and prod_brand from the territories table, and then divide the total sales by the brand / total sales in the territory * 100. Using the rank window function and partition by territory_id will give the total sales by territory

5. That will output the territory, brand, and market share

Output:

territory_id	prod_brand	market_share
T1	Apple	33.333
T1	JBL	16.667
T1	Samsung	50
T2	Apple	25
T2	Samsung	75
T3	Apple	37.5
T3	Canon	12.5
T3	Dell	12.5
T3	GoPro	25
T3	JBL	12.5
T4	Apple	41.667

Question # 3

Write a query to return Territory and corresponding Sales Growth. Compare growth between periods Q4-2021 vs Q3-2021.
If Territory (say T123) has Sales worth $100 in Q3-2021 and Sales worth $110 in Q4-2021, then the Sales Growth will be 10%

i.e. = ((110 - 100)/100) * 100

Output the ID of the Territory and the Sales Growth. Only output these territories that had any sales in both quarters.

Tables: fct_customer_sales, map_customer_territory

Fct_customer_sales

Cust_id:varchar
Prod_sku_id:varchar
Order_date:datetime
Order_value:int
order_id:varchar

Map_customer_territory

Cust_id:varchar
territory_id:varchar

Explanation:

1. Join the map_customer_territory table with the fct_customer_sales on the cust_id. Get the cust_id, territory, quarter using the EXTRACT() function, and sum the sales
2. Using a subquery, sum the sales in Q3, and Q4 using the CASE WHEN function and then perform the following: (Q4-Q3)/Q3 * 100 to calculate the quarter over quarter growth
3. Select the territory, growth from the subquery above

SQL:

```
WITH

d AS (

    SELECT DISTINCT s.cust_id,
    t.territory_id AS territory,
    EXTRACT(QUARTER FROM s.order_date) AS quarter,
    SUM(s.order_value) AS sales
    FROM fct_customer_sales AS s
    LEFT JOIN map_customer_territory AS t ON s.cust_id = t.cust_id
    WHERE EXTRACT(QUARTER FROM s.order_date) IN (3,4)
    AND EXTRACT(YEAR FROM s.order_date) = 2021
    GROUP BY s.cust_id, territory, quarter
    ORDER BY s.cust_id ASC, quarter ASC

)

    SELECT territory,
    growth
    FROM
    (SELECT territory,
    SUM(CASE WHEN quarter = 3 THEN sales END) AS q3,
    SUM(CASE WHEN quarter = 4 THEN sales END) AS q4,
    (SUM(CASE WHEN quarter = 4 THEN sales END) - SUM(CASE WHEN quarter = 3 THEN
sales END)) / SUM(CASE WHEN quarter = 3 THEN sales END) * 100 AS growth
    FROM d
    GROUP BY territory) AS a
    WHERE a.q3 IS NOT NULL
```

Output:

territory	growth
T3	67.12
T5	6.07
T1	4.44
T4	17.36