

SQL

Q1: 2020 spam_indicator = 0 Ad Load

ad_load

1. (country) (dt)
2. spam_indicator = 0
3. 2020

SQL

```
WITH AdImpressions AS (  
    -- Step 1:   
    SELECT  
        ud.country,  
        pi.dt AS calendar_date,  
        CASE WHEN pp.advertiser_id IS NOT NULL THEN 1 ELSE 0 END AS is_ad  
    FROM pin_impressions pi  
    LEFT JOIN promoted_pins pp  
        ON pi.pin_id = pp.pin_id AND pi.dt = pp.dt  
    JOIN user_dimension ud  
        ON pi.user_id = ud.user_id  
    WHERE ud.spam_indicator = 0  
        AND TO_DATE(pi.dt) BETWEEN TO_DATE('2020-01-01') AND TO_DATE('2020-12-31')  
) ,  
AdLoadByCountryDay AS (  
    -- Step 2:   
    SELECT  
        country,  
        calendar_date,  
        SUM(is_ad) AS ad_count,           --   
        COUNT(*) AS total_count,         --   
        SUM(is_ad) / COUNT(*) AS ad_load --   
    FROM AdImpressions  
    GROUP BY country, calendar_date  
)  
-- Step 3:   
SELECT  
    country,  
    calendar_date,  
    ad_load  
FROM AdLoadByCountryDay  
ORDER BY country, calendar_date;
```

1. AdImpressions CTE:

- pin_impressions promoted_pins LEFT JOIN CASE WHEN advertiser_id
- user_dimension spam_indicator = 0

- 2020 年

2. AdLoadByCountryDay CTE:

- `SUM(is_ad)` 和 `COUNT(*)`
- `SUM(is_ad) / COUNT(*)`

3. 步骤:

-

Q2: 2020 年

2020 年

Q1 2020 年

SQL

```
WITH AdLoad AS (
  -- Step 1: 2020 年
  SELECT
    ud.country,
    pi.dt AS calendar_date,
    SUM(CASE WHEN pp.advertiser_id IS NOT NULL THEN 1 ELSE 0 END) / COUNT(*) AS
ad_load
  FROM pin_impressions pi
  LEFT JOIN promoted_pins pp
    ON pi.pin_id = pp.pin_id AND pi.dt = pp.dt
  JOIN user_dimension ud
    ON pi.user_id = ud.user_id
  WHERE ud.spam_indicator = 0
    AND TO_DATE(pi.dt) BETWEEN TO_DATE('2020-01-01') AND TO_DATE('2020-12-31')
  GROUP BY ud.country, pi.dt
),
TodayVsYesterday AS (
  -- Step 2: 2020 年
  SELECT
    t.country,
    t.calendar_date AS today_date,
    y.calendar_date AS yesterday_date,
    t.ad_load AS today_ad_load,
    y.ad_load AS yesterday_ad_load
  FROM AdLoad t
  JOIN AdLoad y
    ON t.country = y.country
    AND DATEDIFF(t.calendar_date, y.calendar_date) = 1
  WHERE t.ad_load > y.ad_load
)
-- Step 3: 2020 年
SELECT
  country,
  today_date,
  today_ad_load,
```

```
yesterday_ad_load
FROM TodayVsYesterday
WHERE today_date = CURRENT_DATE;
```

####

1. AdLoad CTE:

- #### Q1 #####

2. TodayVsYesterday CTE:

- ##### JOIN #####
- #### DATEDIFF(t.calendar_date, y.calendar_date) = 1 #####
- #####

3. ####:

- #####
- ##### CURRENT_DATE ####

Transaction #### Dense Rank

####

Transaction ##### dense_rank ##### MM/DD/YYYY

SQL####

```
WITH RankedTransactions AS (
  -- Step 1: #####
  SELECT
    CustomerID,
    Date,
    STR_TO_DATE(Date, '%m/%d/%Y') AS formatted_date, -- #####
    DENSE_RANK() OVER (PARTITION BY CustomerID ORDER BY STR_TO_DATE(Date,
'%m/%d/%Y')) AS rank
  FROM Transaction
)
-- Step 2: ####
SELECT
  CustomerID,
  Date,
  formatted_date,
  rank
FROM RankedTransactions
ORDER BY CustomerID, rank;
```

####

1. RankedTransactions CTE:

- #### STR_TO_DATE ##### MM/DD/YYYY #####
- #### DENSE_RANK() #####
- PARTITION BY CustomerID ##### ORDER BY STR_TO_DATE(Date, '%m/%d/%Y')
#####

2. 注意：

- 数据库表名不区分大小写
- 表名 ID 不区分大小写

注意

- CTE 即 Common Table Expressions
- CTE 即 Common Table Expressions
- CTE 即 Common Table Expressions

Markdown 在 GitHub 上