Customers' Purchasing Behavior Analysis

Using Google BigQuery SQL

Yasser A. Rahman
01/10/2023

Introduction

DVD rental dataset is one of the classical exemplary datasets for learning SQL.

I have chosen here to use the payments table in order to prepare a summary table of all customers with their first payment, total revenue, and customer value for 30, and 60 days.

The correlated subqueries technique has been used to generate the 30 and 60 days customer value for each customer as shown below.

Payment Table:

Storage info

Number of rows	14,596
Total logical bytes	684.19 KB
Active logical bytes	0 B
Long term logical bytes	684.19 KB
Total physical bytes	177.47 KB
Active physical bytes	0 B
Long term physical bytes	177.47 KB
Time travel physical	0 B

Schema:

Field name	Type	Mode
payment_id	INTEGER	NULLABLE
customer_id	INTEGER	NULLABLE
staff_id	INTEGER	NULLABLE
rental_id	INTEGER	NULLABLE
amount	FLOAT	NULLABLE
payment_date	TIMESTAMP	NULLABLE

Customers' summary table with CLV values for 30, and 60 days:

```
  -- Customers' Summary TABLE
WITH
  customers_first_order AS(
  SELECT
    *
  FROM (
    SELECT
      customer_id,
      amount,
      payment_date,
      ROW_NUMBER() OVER (PARTITION BY customer_id ORDER BY payment_date) AS order_nth
    FROM
      `jrjames83-1171.sampledata.payments`
    ORDER BY
      1)
  WHERE
    order_nth = 1 ),
  summary_so_far AS (
  SELECT
    p.customer_id,
    c.amount AS first_order_amount,
    MIN(p.payment_date) AS first_order_date,
    ROUND(SUM(p.amount), 2) AS total_revenue,
    ROUND(c.amount / SUM(p.amount) * 100, 2) || '%' AS first_order_as_pct_total_revenue
  FROM
    `jrjames83-1171.sampledata.payments` AS p
  JOIN
    customers_first_order AS c
  ON
    c.customer_id = p.customer_id
  GROUP BY
    1,
```

```
    2
  ORDER BY
    4 DESC)
SELECT
  sf.*,
  (
  SELECT
    SUM(p2.amount)
  FROM
    `jrjames83-1171.sampledata.payments` AS p2
  WHERE
    sf.customer_id = p2.customer_id
    AND DATE(p2.payment_date) BETWEEN DATE(sf.first_order_date)
    AND DATE_ADD(DATE(sf.first_order_date), INTERVAL 30 DAY) ) AS
first_30_days_customer_value,
  (
  SELECT
    SUM(p2.amount)
  FROM
    `jrjames83-1171.sampledata.payments` AS p2
  WHERE
    sf.customer_id = p2.customer_id
    AND DATE(p2.payment_date) BETWEEN DATE(sf.first_order_date)
    AND DATE_ADD(DATE(sf.first_order_date), INTERVAL 60 DAY) ) AS
first_60_days_customer_value
FROM
  summary_so_far AS sf
```

Summary Output:

customer_	first_order_amount	first_order_date	total_revenue	first_order_as_pct_total_re	first_30_days_customer_value	first_60_days_customer_
576	4.99	2/15/2007 12:49:27	135.68	3.68%	43.91	90.79
155	1.99	2/15/2007 22:23:54	106.76	1.86%	38.93	76.83
349	2.99	2/15/2007 0:11:13	133.72	2.24%	42.92	103.8
467	8.99	2/16/2007 14:28:11	134.72	6.67%	55.9	90.8
510	5.99	2/15/2007 17:00:57	102.77	5.83%	46.9	70.83
517	4.99	2/16/2007 8:03:12	100.76	4.95%	37.91	91.78
232	0.99	2/16/2007 5:42:40	86.77	1.14%	17.96	37.89
274	1.99	3/1/2007 20:07:04	116.73	1.7%	69.84	115.74
508	4.99	2/16/2007 8:41:24	121.77	4.1%	23.96	68.88
532	4.99	2/16/2007 11:08:50	149.69	3.33%	45.91	131.74
93	4.99	2/18/2007 3:50:23	76.8	6.5%	36.92	72.82
106	4.99	2/18/2007 6:24:45	95.79	5.21%	33.93	60.86
180	2.99	2/19/2007 12:00:19	95.78	3.12%	41.91	73.84
186	4.99	2/14/2007 23:47:06	111.71	4.47%	26.93	86.79