**CASE 5: RFM ANALYSIS**

[E-Commerce Data](https://www.kaggle.com/datasets/carrie1/ecommerce-data)

Download the dataset from the link above and perform RFM analysis. When calculating Recency, use the date of the most recent order as the reference.

**Query**:

WITH rfm\_analyze AS (WITH scores AS (WITH frequency1 AS (WITH monetary1 AS (WITH recency1 AS

(SELECT DISTINCT customer\_id,

MAX (invoicedate)::date AS max\_inv\_date

FROM rfm

GROUP BY 1)

SELECT DISTINCT (rfm.customer\_id),

max\_inv\_date,

'2011-12-09'-max\_inv\_date::date AS recency,

round(quantity\*unitprice) AS payment

FROM recency1

JOIN rfm ON rfm.customer\_id=recency1.customer\_id

ORDER BY 3 DESC)

SELECT rfm.customer\_id,

recency,

SUM (payment)AS monetary

FROM monetary1

JOIN rfm ON rfm.customer\_id=monetary1.customer\_id

WHERE payment>0

GROUP BY 1,2

ORDER BY 3 DESC)

SELECT rfm.customer\_id,

recency,

monetary,

COUNT (DISTINCT invoiceno) AS frequency

FROM frequency1

JOIN rfm ON rfm.customer\_id=frequency1.customer\_id

GROUP BY 1,2,3

ORDER BY 4 DESC)

SELECT rfm.customer\_id,

recency,

NTILE(5) OVER (ORDER BY recency DESC) AS recency\_score,

frequency,

CASE WHEN frequency BETWEEN 1 AND 4 THEN frequency

ELSE 5

END AS frequency\_score,

monetary,

NTILE(5) OVER (ORDER BY monetary) AS monetary\_score

FROM scores

JOIN rfm ON rfm.customer\_id=scores.customer\_id

GROUP BY 1,2,4,6

ORDER BY 3 DESC,5 DESC,7 DESC)

SELECT customer\_id,

recency\_score::varchar ||'-'|| frequency\_score::varchar ||'-'|| monetary\_score::varchar AS RFM\_scores

FROM rfm\_analyze;

|  |
| --- |
| Customer Id RFM Scores |
| 13949 5-5-5 |
| 15804 5-5-5 |
| 13113 5-5-5 |
| 16665 1-1-1 |
| 14957 1-1-1 |
| 12498 4-3-2 |
| 13058 4-3-1 |

**Comment:** Examples of some RFM scores from the query are listed in the table. The RFM score for each customer can be examined in the following CSV file.

