Cohort analysis

WITH completed\_orders AS (

SELECT CUSTOMER\_ID, ORDER\_TMS, TRUNC(ORDER\_TMS, 'MONTH') AS ORDER\_MONTH

FROM CO.ORDERS

WHERE ORDER\_STATUS = 'COMPLETE'

),

first\_purchase AS (

-- cohort month per customer (date)

SELECT CUSTOMER\_ID, TRUNC(MIN(ORDER\_TMS), 'MONTH') AS COHORT\_MONTH

FROM completed\_orders

GROUP BY CUSTOMER\_ID

)

,cust\_months AS (

-- distinct customer x month rows (one row per customer per month)

SELECT DISTINCT co.CUSTOMER\_ID,

co.ORDER\_MONTH,

fp.COHORT\_MONTH

FROM completed\_orders co

JOIN first\_purchase fp ON co.CUSTOMER\_ID = fp.CUSTOMER\_ID

)

,customer\_cohort\_index AS (

-- compute months since cohort (0-based)

SELECT CUSTOMER\_ID,

ORDER\_MONTH,

COHORT\_MONTH,

(EXTRACT(YEAR FROM ORDER\_MONTH) \* 12 + EXTRACT(MONTH FROM ORDER\_MONTH))

- (EXTRACT(YEAR FROM COHORT\_MONTH) \* 12 + EXTRACT(MONTH FROM COHORT\_MONTH)) AS COHORT\_INDEX

FROM cust\_months

)

,cohort\_counts AS (

-- number of unique customers active per cohort\_month x cohort\_index

SELECT COHORT\_MONTH,

COHORT\_INDEX,

COUNT(\*) AS customers

FROM customer\_cohort\_index

WHERE COHORT\_INDEX >= 0

GROUP BY COHORT\_MONTH, COHORT\_INDEX

)

,cohort\_size AS (

-- cohort size = customers in index 0

SELECT COHORT\_MONTH, customers AS cohort\_size

FROM cohort\_counts

WHERE COHORT\_INDEX = 0

)

,retention AS (

SELECT cc.COHORT\_MONTH,

cc.COHORT\_INDEX,

cc.customers,

ROUND(100 \* cc.customers / cs.cohort\_size, 2) AS retention\_pct

FROM cohort\_counts cc

JOIN cohort\_size cs ON cc.cohort\_month = cs.cohort\_month

)

SELECT TO\_CHAR(ret.COHORT\_MONTH, 'YYYY-MM') AS cohort\_month,

COALESCE(MAX(CASE WHEN ret.COHORT\_INDEX = 0 THEN ret.retention\_pct END), 0) AS pct\_0,

COALESCE(MAX(CASE WHEN ret.COHORT\_INDEX = 1 THEN ret.retention\_pct END), 0) AS pct\_1,

COALESCE(MAX(CASE WHEN ret.COHORT\_INDEX = 2 THEN ret.retention\_pct END), 0) AS pct\_2,

COALESCE(MAX(CASE WHEN ret.COHORT\_INDEX = 3 THEN ret.retention\_pct END), 0) AS pct\_3,

COALESCE(MAX(CASE WHEN ret.COHORT\_INDEX = 4 THEN ret.retention\_pct END), 0) AS pct\_4,

COALESCE(MAX(CASE WHEN ret.COHORT\_INDEX = 5 THEN ret.retention\_pct END), 0) AS pct\_5,

COALESCE(MAX(CASE WHEN ret.COHORT\_INDEX = 6 THEN ret.retention\_pct END), 0) AS pct\_6

FROM retention ret

GROUP BY TO\_CHAR(ret.COHORT\_MONTH, 'YYYY-MM')

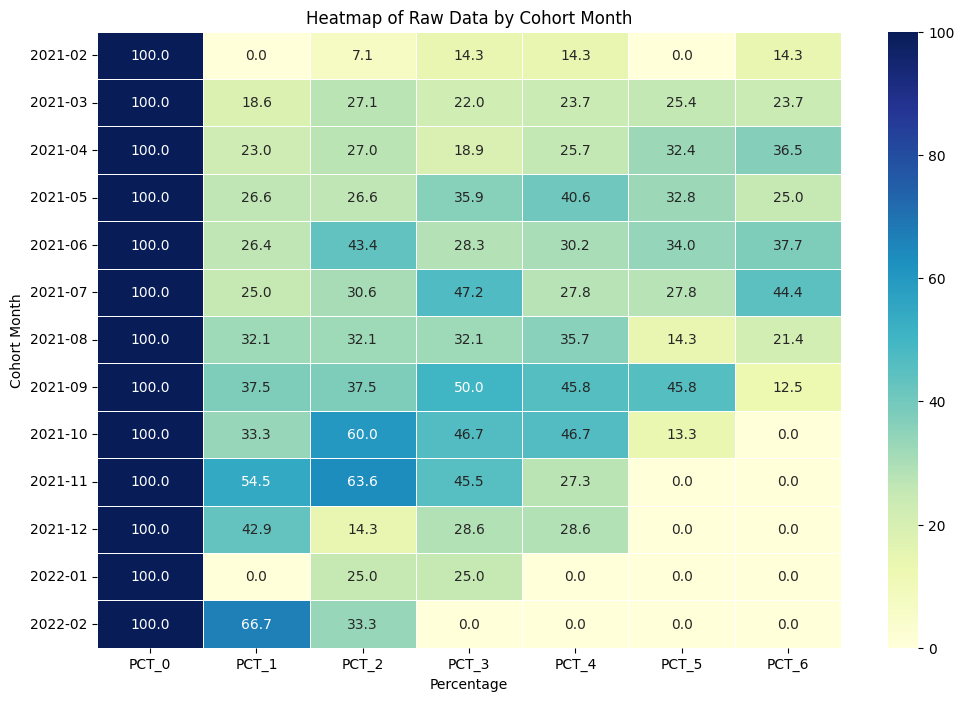
ORDER BY MIN(ret.COHORT\_MONTH);

“In my cohort analysis, I calculated the retention rate per cohort over time.

The columns pct\_0, pct\_1, pct\_2, etc., represent how many customers from each cohort were still active after 0, 1, 2, and so on months.

For example, if 100 customers joined in January and 85 made another purchase in February, the retention for month 1 (pct\_1) is 85%.

This helps businesses measure customer loyalty and understand if retention improves with new acquisition or marketing strategies.”



"This heatmap visualizes the percentage values across different stages (PCT\_0 to PCT\_6) for each cohort month from February 2021 to February 2022. The color intensity represents the magnitude of the percentages, with darker shades indicating higher values. We can observe that all cohorts start with a PCT\_0 of 100%. By examining the heatmap, we can analyze the trends within each cohort as they progress through the stages and compare the performance of different cohorts at similar stages. For example, the later cohorts in 2021 show a decline in percentages at later stages compared to earlier cohorts."