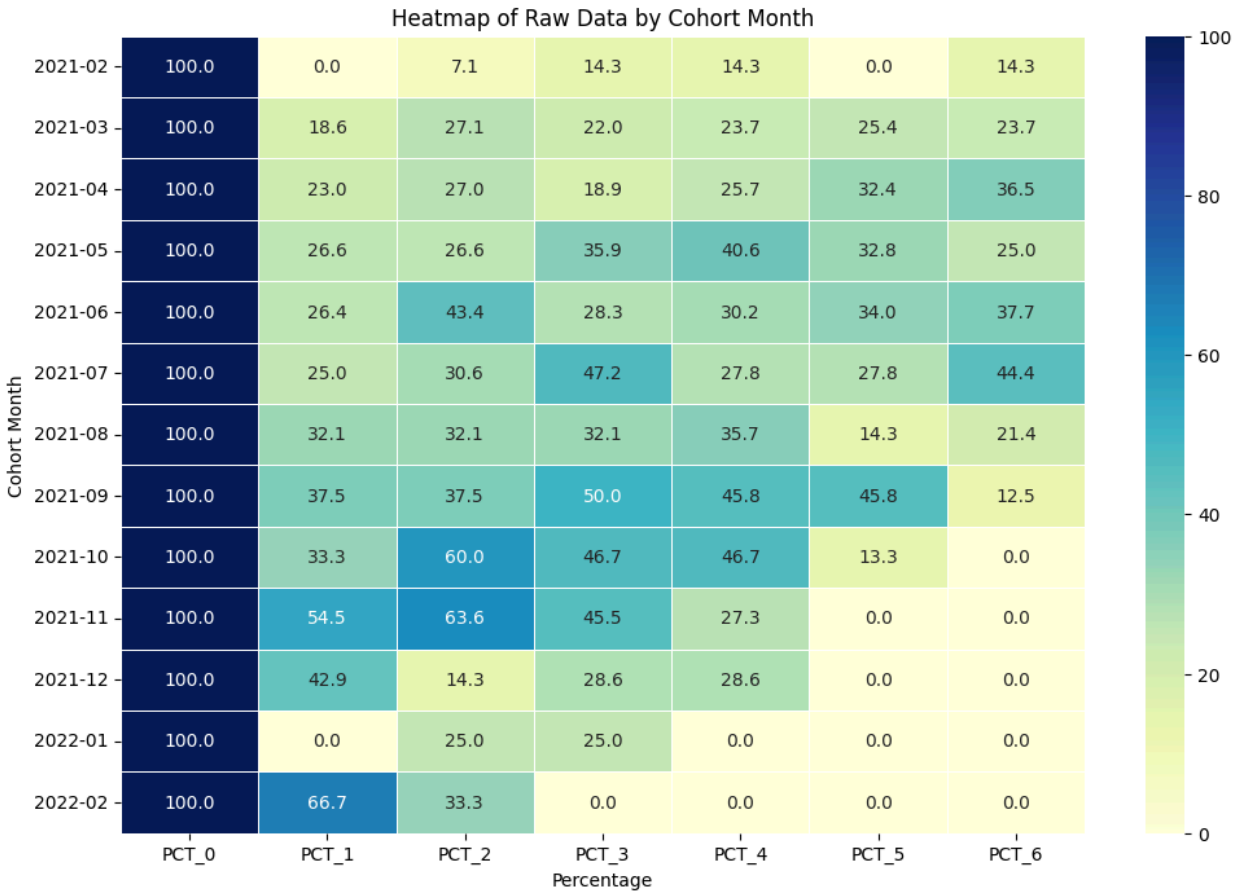
Cohort analysis

```
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."