Growth Inventory and Cohort Analysis Using SQL BigQuery

-- To support the business, find monthly growth of inventory in percentage breakdown by product categories, ordered by time descendingly. After analyzing the monthly growth, is there any interesting insight that we can

-- pertumbuhan inventory perbulan dalam bentuk persen yang di breakdown by product kategori dan diurutkan berdasarkan waktu terrecent or desc

WITH monthly\_inventory AS(

SELECT

product\_category

,FORMAT\_DATE('%Y-%m', created\_at) AS year\_month

,COUNT (id) AS stock

FROM fsda-sql-01.TheLook\_Ecommerce.inventory\_items AS inventory

GROUP BY 1,2

),

previous\_monthly\_inventory AS (

SELECT

product\_category

,year\_month

,stock

,LAG(stock) OVER (PARTITION BY product\_category ORDER BY year\_month) AS prev\_month\_stock

FROM monthly\_inventory

)

SELECT

year\_month

,product\_category

,stock

,prev\_month\_stock

,ROUND((((stock - prev\_month\_stock)/ prev\_month\_stock)\*100),2) AS percentage\_growth\_inventory

FROM previous\_monthly\_inventory

ORDER BY 2 ASC, 1 DESC;

-- Create monthly retention cohorts (the groups, or cohorts, can be defined based upon the date that a user completely purchased a product) and then how many of them (%) coming back for the following months in 2022. After analyzing the retention cohort, is there any interesting insight that we can get?

-- order\_items: id, created\_at, orders: user\_id, delivered\_at, status

-- create cohort\_items

WITH cohort\_items AS(

SELECT DISTINCT

orders.user\_id AS user\_id

,MIN(DATE\_TRUNC(DATE(orders.created\_at),MONTH)) OVER (PARTITION BY orders.user\_id) AS first\_order\_date

,DATE\_TRUNC(DATE(orders.created\_at),MONTH) AS running\_order\_date

FROM fsda-sql-01.TheLook\_Ecommerce.order\_items AS orders

WHERE status = 'Complete'

ORDER BY 1

),

user\_period AS(

SELECT

\*

,DATE\_DIFF(running\_order\_date,first\_order\_date, MONTH) AS diff\_month

,COUNT(DISTINCT user\_id) OVER (PARTITION BY first\_order\_date) AS cohort\_size

FROM cohort\_items

),

total\_user\_retention AS (

SELECT

first\_order\_date

,diff\_month AS month

,cohort\_size

,COUNT (DISTINCT user\_id ) AS total\_user

FROM user\_period

WHERE first\_order\_date BETWEEN '2022-01-01' AND '2022-12-01'

AND running\_order\_date <='2022-12-01'

GROUP BY 1,2,3

ORDER BY 1,2

),

retention AS(

SELECT

\*

,ROUND(((total\_user/cohort\_size)\*100),2) AS retention\_rate

FROM total\_user\_retention

ORDER BY first\_order\_date, total\_user DESC

)

SELECT

first\_order\_date

,MAX (total\_user) AS total\_users

,MAX(CASE WHEN month = 0 THEN retention\_rate END) AS month\_0

,MAX(CASE WHEN month = 1 THEN retention\_rate END) AS month\_1

,MAX(CASE WHEN month = 2 THEN retention\_rate END) AS month\_2

,MAX(CASE WHEN month = 3 THEN retention\_rate END) AS month\_3

,MAX(CASE WHEN month = 4 THEN retention\_rate END) AS month\_4

,MAX(CASE WHEN month = 5 THEN retention\_rate END) AS month\_5

,MAX(CASE WHEN month = 6 THEN retention\_rate END) AS month\_6

,MAX(CASE WHEN month = 7 THEN retention\_rate END) AS month\_7

,MAX(CASE WHEN month = 8 THEN retention\_rate END) AS month\_8

,MAX(CASE WHEN month = 9 THEN retention\_rate END) AS month\_9

,MAX(CASE WHEN month = 10 THEN retention\_rate END) AS month\_10

,MAX(CASE WHEN month = 11 THEN retention\_rate END) AS month\_11

,MAX(CASE WHEN month = 12 THEN retention\_rate END) AS month\_12

FROM retention

GROUP BY 1

ORDER BY 1;