-- Tạo mô hình Clustering - Kmeans

-- 3 clusters (3 Annual Income (k$))

-- n = 200 mau --> log10(200)) =2.3 cho nên ta chọn k=2 hoặc k=3

--k=2

CREATE OR REPLACE MODEL `project-tuananh-2023.Tuananh\_ML.shopping\_data\_clustering\_2 `

  OPTIONS(model\_type='kmeans',num\_clusters=2)

  AS

    SELECT Age,Annual\_Income\_\_k\_\_

    FROM `project-tuananh-2023.Tuananh\_ML.shopping\_data`;

-- Evaluate model

SELECT \*

FROM ML.EVALUATE(MODEL `project-tuananh-2023.Tuananh\_ML.shopping\_data\_clustering\_2 `);

--k=3

CREATE OR REPLACE MODEL `project-tuananh-2023.Tuananh\_ML.shopping\_data\_clustering\_3 `

  OPTIONS(model\_type='kmeans',num\_clusters=3)

  AS

    SELECT Age,Annual\_Income\_\_k\_\_

    FROM `project-tuananh-2023.Tuananh\_ML.shopping\_data`;

-- Evaluate model

SELECT \*

FROM ML.EVALUATE(MODEL `project-tuananh-2023.Tuananh\_ML.shopping\_data\_clustering\_3 `);

SELECT \*

FROM ML.PREDICT(MODEL `project-tuananh-2023.Tuananh\_ML.shopping\_data\_clustering\_3 `,

    ( SELECT \*

    FROM `project-tuananh-2023.Tuananh\_ML.shopping\_data`));