Data Load in Product Table.

Problem

While loading the data from Staging table into product table, I encountered an error. There were many repeated product_id in the csv file. Which makes Sense because Same Id Clould have multiple transaction from multiple location. However in RDBMS Product_id being the P.K we cannot have duplicates.

Solution.

I'm gonna use CTE & window function to group distinct product_id & rank each groups respectively. Then insert into the product table casting appropriate data types

Let's Look at the Code to understand Clearly.

```
WITH ranked_products AS (

        SELECT
              product_id,
              product_name,
              category,
              ROW_NUMBER() OVER (PARTITION BY product_id   ORDER BY product_name) AS rn
        FROM  staging_data_raw
        WHERE  product_id  IS NOT NULL
                  )
        INSERT INTO  product  (product_ID, product_name, category)
        SELECT
              CAST (product_id  AS  INT),
              CAST (product_name  AS  VARCHAR (50)
              CAST (Category  AS  VARCHAR (50)
        FROM  ranked_products
        WHERE  rn = 1
```

Common Table Expression [CTE] - usually Starts "WITH" to call CTE
Think of it like creating a temporary view or table that you can use right after.

SELECT - Choosing the columns

ROW_NUMBER() OVER(...) AS rn - This generates a row number for each row within the same product_id group.

PARTITION BY product_id - This tells SQL to re-start the row numbering for each product_id group

ORDER BY product_name decides the order in which the row numbers are assigned

AS rn - This gives a name to the generated row number Column

FROM Staging_data_raw - This is the Source table - where all raw CSV data was loaded.

WHERE product_id IS NOT NULL - This Skips rows that are missing a product ID, Since they can't be inserted into table that requires it.

```
INSERT INTO product
SELECT CAST( . . . . )
FROM ranked_products
WHERE rn =1
```