1. Pivoting and Transforming Measure Values

To combine multiple rows into a single row for each unique OrderNo, Customer, and ArticleCode combination by pivoting the dataset to create 5 new columns (Qty, Retail Price (SGD), Net Price (SGD), Item Discount (SGD), and Gst (SGD)) from the MeasureNames column and populating their respective values from the MeasureValues column. Sample below:

Order No	Customer	Dob	Customer Gender	Country	Remarks 1	Brand Description	Sales Transaction Type	Document Date	Store Id	Sales Employee Id	Article Code	Description	Category Description	Sales Channel	Measure Names	Measure Values
1234	abcd	13/11/1989	F	SG	ES00	CDEF	S	29/07/2023	11		CDEF40258	LOREM IPSUM EDT 65ML	Fragrance	3rd Party Marketplaces	Qty	1
1234	abcd	13/11/1989	F	SG	ES00	CDEF	S	29/07/2023	11		CDEF40258	LOREM IPSUM EDT 65ML	Fragrance	3rd Party Marketplaces	Retail Price	174
1234	abcd	13/11/1989	F	SG	ES00	CDEF	S	29/07/2023	11		CDEF40258	LOREM IPSUM EDT 65ML	Fragrance	3rd Party Marketplaces	Retail Price (SGD)	174
1234	abcd	13/11/1989	F	SG	ES00	CDEF	S	29/07/2023	11		CDEF40258	LOREM IPSUM EDT 65ML	Fragrance	3rd Party Marketplaces	Net Price (SGD)	161.11
1234	abcd	13/11/1989	F	SG	ES00	CDEF	S	29/07/2023	11		CDEF40258	LOREM IPSUM EDT 65ML	Fragrance	3rd Party Marketplaces	Net Price	161.11
1234	abcd	13/11/1989	F	SG	ES00	CDEF	S	29/07/2023	11		CDEF40258	LOREM IPSUM EDT 65ML	Fragrance	3rd Party Marketplaces	Item Discount	0
1234	abcd	13/11/1989	F	SG	ES00	CDEF	S	29/07/2023	11		CDEF40258	LOREM IPSUM EDT 65ML	Fragrance	3rd Party Marketplaces	Item Discount (SGD)	0
1234	abcd	13/11/1989	F	SG	ES00	CDEF	S	29/07/2023	11		CDEF40258	LOREM IPSUM EDT 65ML	Fragrance	3rd Party Marketplaces	Gst	12.89
1234	abcd	13/11/1989	F	SG	ES00	CDEF	S	29/07/2023	11		CDEF40258	LOREM IPSUM EDT 65ML	Fragrance	3rd Party Marketplaces	Gst (SGD)	12.89

The dataset initially contained 953,361 rows and 17 columns. After pivoting, the dataset was reduced to 105,929 rows and 24 columns, reflecting the consolidation of redundant rows into a single row per unique combination, while adding 5 new measure-related columns.

2. Category Description Standardization

Identified inconsistencies in the Category Description column where values had varying cases or typos (e.g., Home collection, SKIN CARE, Skin care). Consolidated these variations into a consistent format (Home Collection, Skin Care) to ensure uniformity in category representation.

3. Document Date Formatting

The Document Date column contained unformatted date values in ~10% of the data. Converted this column to a standard datetime format to align all entries and enable accurate date-based analysis.

4. Removal of Bulk Transactions

A total of 69 rows where Net Price (SGD) exceeded 20,000 were identified. These transactions were reviewed and confirmed by the business team to be bulk purchases, not relevant to the analysis of individual buyers. Therefore, they were excluded from the dataset to maintain consistency and focus on individual transactions.