Café Sales

Assumption:

All ERRORs are system errors, UNKNOWNs and empty rows are unknown (not specified) not errors.

All items sold in the café have only one price per unit (Tea – 1.5, Cookie – 1.0, Coffee – 2.0, Salad – 5.0, Smoothie & Sandwich – 4.0, Cake & Juice – 3.0).

Null transactions were done around the same time as preceding ones.

Steps Taken:

- Reason: *To provide context to each column*

  Used the first row as headers

- Reason: *To avoid errors when changing data type*

  Replaced all "ERROR" and "UNKNOWN" values in "Quantity", "Price Per Unit", "Total Spent", and "Transaction Date" with null

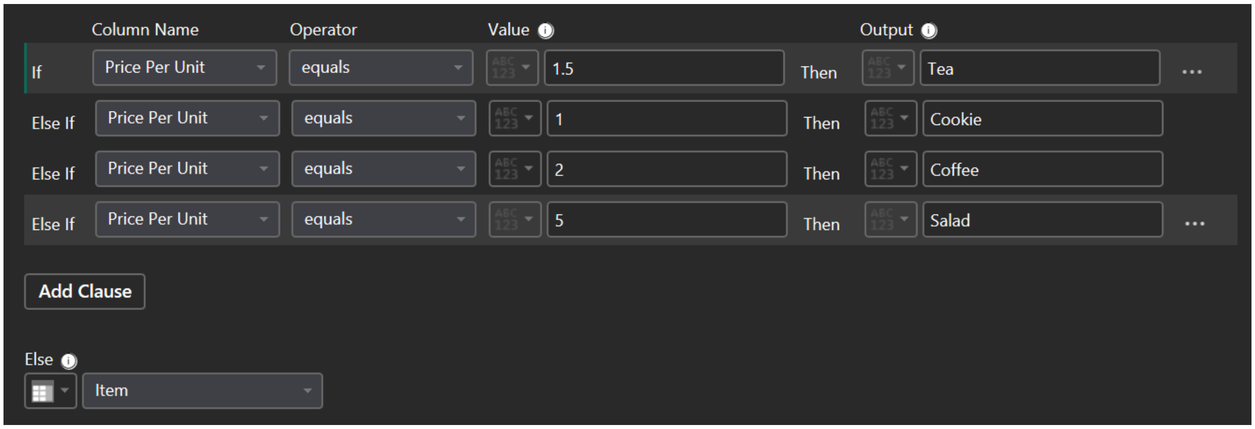
- Reason: *All columns need to be assigned proper data types according to context, to provide better analysis*

  Changed the data types of columns to appropriate types, then filled down the "Transaction Date" column

UPDATE ITEMS

- Reason: *To ensure items sold are properly filled according to their price, reducing the number of empty, "ERROR" and "UNKNOWN"*

  Created a conditional column where when Price equals a certain value, an Item is entered, otherwise, fill up with "Item" column, could only be used for non-intersecting prices (Tea – 1.5, Cookie – 1.0, Coffee – 2.0, Salad – 5.0).



	Column Name	Operator	Value		Output
If	Price Per Unit	equals	1.5	Then	Tea
Else If	Price Per Unit	equals	1	Then	Cookie
Else If	Price Per Unit	equals	2	Then	Coffee
Else If	Price Per Unit	equals	5	Then	Salad

Else: Item

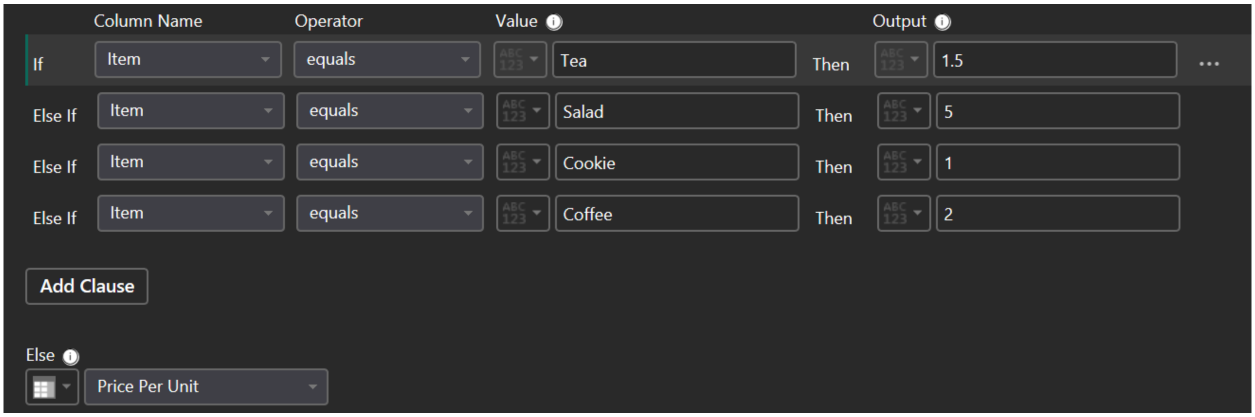
- Reason: *To avoid duplication*

  Replaced the "Item" column with the new conditional column

UPDATE PRICE PER UNIT

- Reason: *To ensure prices are properly filled according to their item, reducing the number of empty, "ERROR" and "UNKNOWN"*

  Created a conditional column where when items are a certain value, a price is entered, otherwise, fill up with "Price Per Unit" column, could only be used for non-intersecting prices (Tea – 1.5, Cookie – 1.0, Coffee – 2.0, Salad – 5.0).



	Column Name	Operator	Value		Output
If	Item	equals	Tea	Then	1.5
Else If	Item	equals	Salad	Then	5
Else If	Item	equals	Cookie	Then	1
Else If	Item	equals	Coffee	Then	2

Else: Price Per Unit

- Reason: *To avoid duplication*

  Replaced the "Price Per Item" column with the new conditional column

UPDATE QUANTITY

- Reason: *To fill quantity with appropriate values*

  Inserted a division column, where "Total Spent" was divided by "Price Per Unit"

- Reason: *To fill quantity with appropriate values, and reducing empty rows*

Merged the “Division” column with the “Quantity” column, then split the merged columns by position

- Reason: *To avoid duplication*

  Replaced the “Quantity" column with the “Merged.1” column that had fewer empty rows

UPDATE TOTAL SPENT

- Reason: *To fill amount spent with appropriate values*

  Inserted a multiplication column, where updated “Quantity” was multiplied by “Price Per Unit”

- Reason: *To fill amount spent with appropriate values, and reducing empty rows*

  Merged the “Multiplication” column with the “Total Spent” column by a comma, then split the merged columns by delimiter(comma).

- Reason: *To avoid duplication*

  Replaced the “Total Spent" column with the “Merged.1” column that had fewer empty rows

UPDATE PRICE PER UNIT

- Reason: *To fill price with appropriate values*

  Inserted a division column, where updated “Total Spent” was divided by updated “Quantity”

- Reason: *To fill price with appropriate values, and reducing empty rows*

  Merged the “Division” column with the “Price Per Unit” column by a comma, then split the merged columns by delimiter(comma)

- Reason: *To avoid duplication*

  Replaced the “Price Per Unit" column with the “Merged.1” column that had fewer empty rows

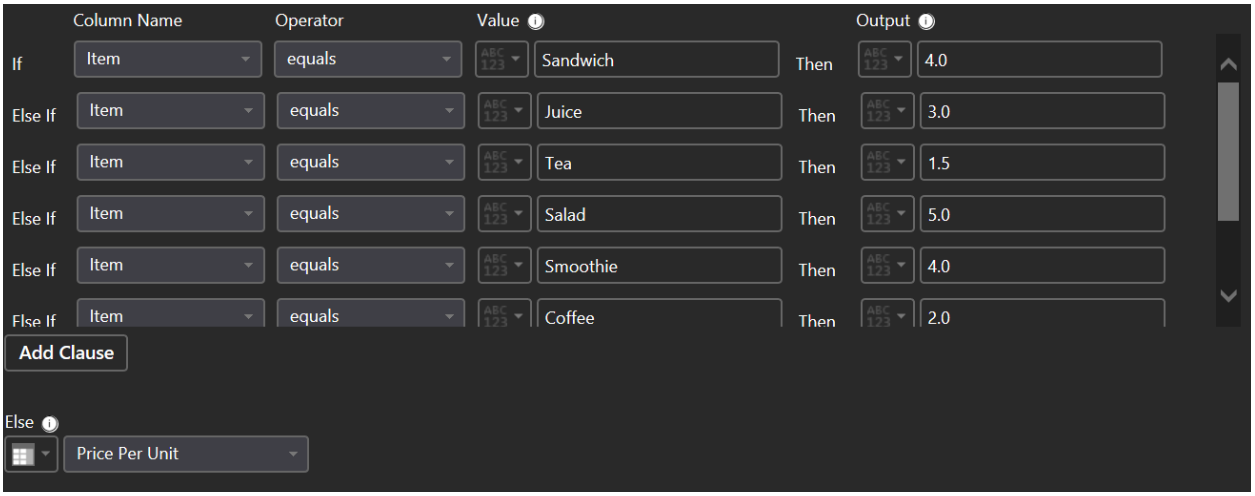
- Reason: *To remove empty rows*

  Replaced all null values in “Location” and “Payment Method” with “UNKNOWN”

UPDATE PRICE PER UNIT – Noticed there were some rows that has items, but no price

- Reason: *To ensure prices are properly filled according to their item, reducing the number of empty, “ERROR” and “UNKNOWN”*

Created a conditional column where when items are a certain value, a price is entered, otherwise, fill up with “Price Per Unit” column, for all prices (Tea – 1.5, Cookie – 1.0, Coffee – 2.0, Salad – 5.0, Smoothie & Sandwich – 4.0, Cake & Juice – 3.0).



	Column Name	Operator	Value		Output
If	Item	equals	Sandwich	Then	4.0
Else If	Item	equals	Juice	Then	3.0
Else If	Item	equals	Tea	Then	1.5
Else If	Item	equals	Salad	Then	5.0
Else If	Item	equals	Smoothie	Then	4.0
Else If	Item	equals	Coffee	Then	2.0

Else
Price Per Unit

- Reason: *To avoid duplication*

  Replaced the “Price Per Item” column with the new conditional column

UPDATE TOTAL SPENT – From the updated prices and non-null quantity values

- Reason: *To fill amount spent with appropriate values, and reduce null values*

  Inserted a multiplication column, where updated “Price Per Unit” was multiplied by “Quantity”

- Reason: *To avoid duplication*

  Replaced the “Total Spent" column with the “Multiplied” column that had fewer null rows

Nike Sales

Assumption:

The “Nike_Sales.csv” contain data on Nike shoes only.

Sizes 6 – 7, 8 – 9, 10 – 12, correspond to M, L, XL respectively.

MRP (Manufacturer’s Retail Price) refers to individual prices of items.

Where there is no discount specified, no discount was given/applied.

The data set was gotten from cities in India.

Steps Taken:

i. Reason: *The “Region” column contains names of Indian cities; errors need to be fixed.*

   Replaced entry mistake for “Region”, correcting each error to the intended Indian city.

ii. Reason: *All columns need to be assigned proper data types according to context, to provide better analysis*

   Changed the data types of the columns to the best represented data types, however I encountered some errors while changing that of “Order_Date”, hence, errors were removed, and the column was filled down

iii. Assumption: *There can’t be a negative number of units sold*

   The “Units_Sold” column was made absolute

iv. Assumption: *If no discount, means there was 0 discount given*

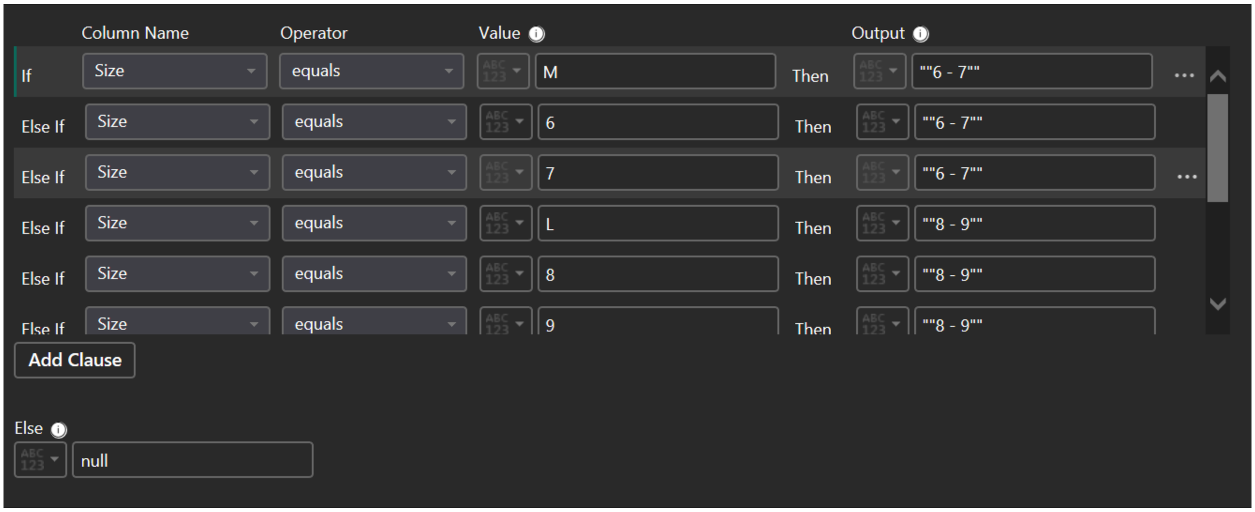
   For the “Discount_Applied” column, the null entries were replaced with 0

v. Reason*: To give a standard size applicable to each group*

   Created a conditional column, where if “Size” is “M” or “6” or “7”, then replace with “6 – 7”, the grouping was done by:

   “L” or “8” or “9” = “8 – 9”

   “XL” or “10” or “11” or “12” = “10 – 12”



	Column Name	Operator	Value		Output
If	Size	equals	M	Then	""6 - 7""
Else If	Size	equals	6	Then	""6 - 7""
Else If	Size	equals	7	Then	""6 - 7""
Else If	Size	equals	L	Then	""8 - 9""
Else If	Size	equals	8	Then	""8 - 9""
Else If	Size	equals	9	Then	""8 - 9""

Else: null

vi. Reason: *To provide clarity*

Renamed "MRP" to "Retail Price"

vii. Reason: *To find the selling price of products*

Created a multiplied column of "Unit_Sold" and "Retail Price", "Selling Price"

viii. Reason: *Calculating discount amount given*

Created a multiplied column of "Selling Price" and "Disccount_Applied", "Discounted_Amount"

ix. Reason: *To get the actual final revenue*

Created a subtracted column where "Discount" was deducted from "Selling Price"