Coffee Sales Analysis

Power BI Data Modeling & Transformation

Danish Karim
Data Analyst

Objective

- The objective of this project is to transform and clean a given Coffee Sales dataset and develop a structured data model using Power BI. The process includes data preprocessing, normalization, and modeling to ensure data integrity and efficiency.

Tasks

- Data Transformation:

● Check Column names, data types, missing, and error values. (If any)

● Split the dataset into smaller tables to normalize the data:

○ Transactions: transaction_id, transaction_date, transaction_time, transaction_qty, store_id, product_id. ○ Stores: store_id, store_location.

○ Products: product_id, unit_price, product_category, product_type, product_detail.

Tasks

- Data Modeling
- Identify Fact and Dimension Tables.
- Create Relationships between tables.
- Identify the schema.

Tasks

- Power Query Analysis Perform the following tasks:

1. Create a column for Sales o Merge column ”Unit price” from “products” to the “Transaction” table.

o Create a custom column: Sales = unit price x transaction_qty.

2. Conditional column:

o Create a conditional column Is High Quantity: If transaction_qty > 4, return "Yes", otherwise "No".

3. Parameters:

o Calculate the given and store them as parameters: i. Calculate Total Sales: Sum of Total Sales. ii. Calculate Average Transaction Quantity: Average of transaction_qty.

Tasks

- 4. Filter based on parameters:

  ○ Create a duplicate of the Transactions Table.

  ○ Filter the transactions with a quantity greater than the parameter " Average transaction quantity".

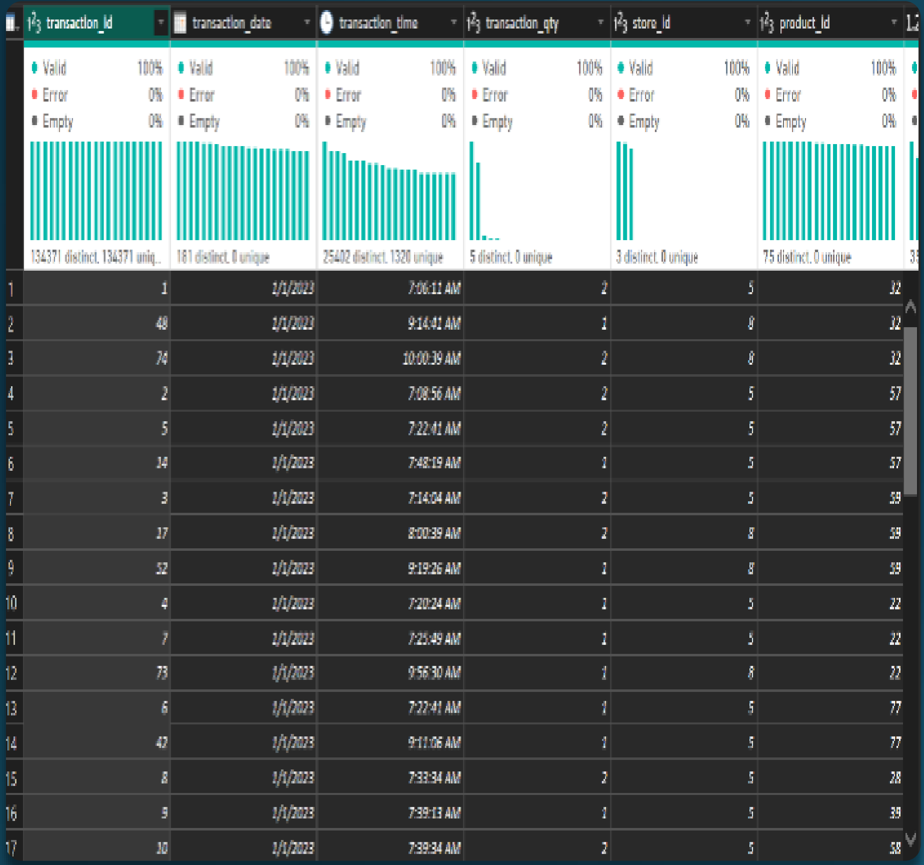
  5. Sales Based on Location:

  ○ Merge Sales from "Transaction Table" to "Store" and show the aggregated value "Sum of Sales".

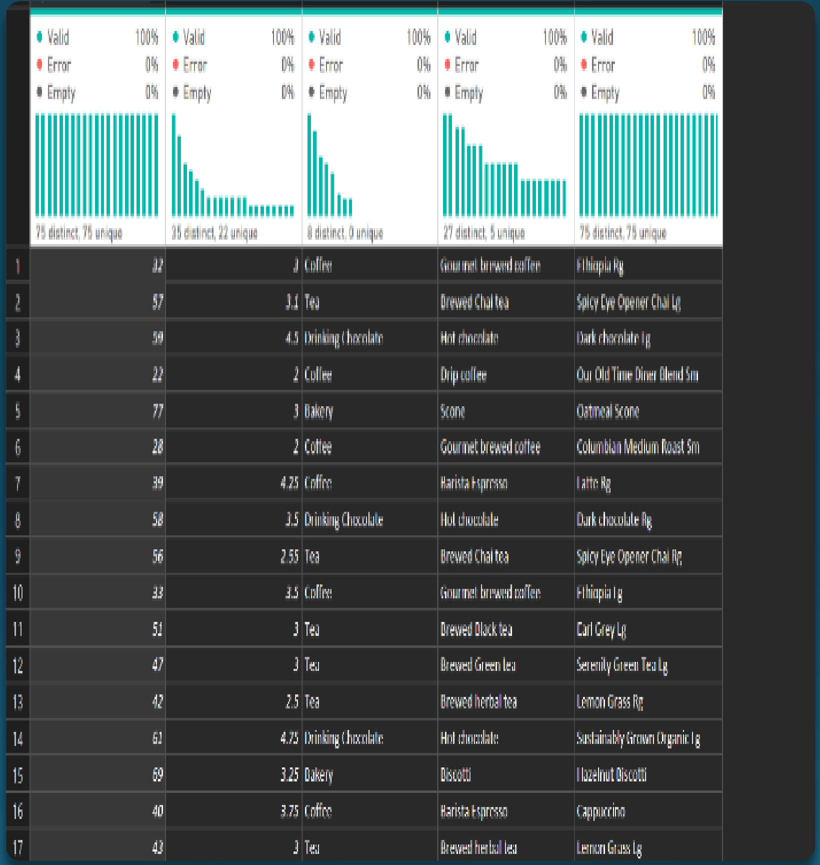
  6. Count of Products in each product Category:

  ○ Create a duplicate of "Products". Apply GroupBY to count products in each category. Rename this table as "Product summary"

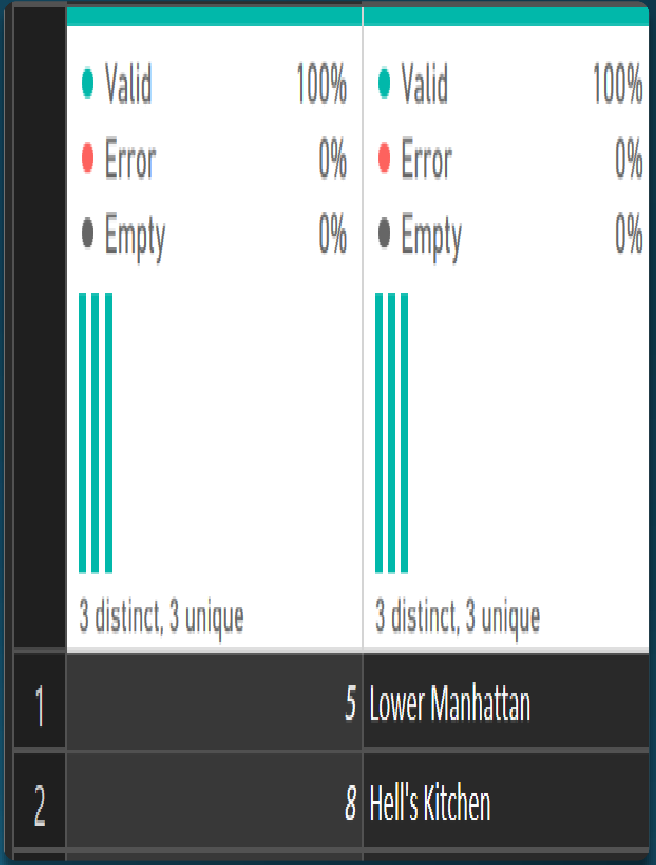
Data Transformation Solution



	transaction_id	transaction_date	transaction_time	transaction_qty	store_id	product_id
	Valid 100% Error 0% Empty 0% 134371 distinct, 134371 uniq...	Valid 100% Error 0% Empty 0% 181 distinct, 0 unique	Valid 100% Error 0% Empty 0% 25402 distinct, 1320 unique	Valid 100% Error 0% Empty 0% 5 distinct, 0 unique	Valid 100% Error 0% Empty 0% 3 distinct, 0 unique	Valid 100% Error 0% Empty 0% 75 distinct, 0 unique
1	1	1/1/2023	7:06:11 AM	2	5	32
2	48	1/1/2023	9:14:41 AM	1	8	32
3	74	1/1/2023	10:00:39 AM	2	8	32
4	2	1/1/2023	7:08:56 AM	2	5	57
5	5	1/1/2023	7:22:41 AM	2	5	57
6	14	1/1/2023	7:48:19 AM	1	5	57
7	3	1/1/2023	7:14:04 AM	2	5	59
8	17	1/1/2023	8:00:39 AM	2	8	59
9	52	1/1/2023	9:19:26 AM	1	8	59
10	4	1/1/2023	7:20:24 AM	1	5	22
11	7	1/1/2023	7:25:49 AM	1	5	22
12	73	1/1/2023	9:56:30 AM	1	8	22
13	6	1/1/2023	7:22:41 AM	1	5	77
14	47	1/1/2023	9:11:06 AM	1	5	77
15	8	1/1/2023	7:33:34 AM	2	5	28
16	9	1/1/2023	7:39:13 AM	1	5	39
17	10	1/1/2023	7:39:34 AM	2	5	58



	Valid 100% Error 0% Empty 0% 75 distinct, 75 unique	Valid 100% Error 0% Empty 0% 35 distinct, 22 unique	Valid 100% Error 0% Empty 0% 8 distinct, 0 unique	Valid 100% Error 0% Empty 0% 27 distinct, 5 unique	Valid 100% Error 0% Empty 0% 75 distinct, 75 unique
1	32	3	Coffee	Gourmet brewed coffee	Ethiopia Rg
2	57	3.1	Tea	Brewed Chai tea	Spicy Eye Opener Chai Lg
3	59	4.5	Drinking Chocolate	Hot chocolate	Dark chocolate Lg
4	22	2	Coffee	Drip coffee	Our Old Time Diner Blend Sm
5	77	3	Bakery	Scone	Oatmeal Scone
6	28	2	Coffee	Gourmet brewed coffee	Columbian Medium Roast Sm
7	39	4.25	Coffee	Barista Espresso	Latte Rg
8	58	3.5	Drinking Chocolate	Hot chocolate	Dark chocolate Rg
9	56	2.55	Tea	Brewed Chai tea	Spicy Eye Opener Chai Rg
10	33	3.5	Coffee	Gourmet brewed coffee	Ethiopia Lg
11	51	3	Tea	Brewed Black tea	Earl Grey Lg
12	47	3	Tea	Brewed Green tea	Serenity Green Tea Lg
13	42	2.5	Tea	Brewed herbal tea	Lemon Grass Rg
14	61	4.75	Drinking Chocolate	Hot chocolate	Sustainably Grown Organic Lg
15	69	3.25	Bakery	Biscotti	Hazelnut Biscotti
16	40	3.75	Coffee	Barista Espresso	Cappuccino
17	43	3	Tea	Brewed herbal tea	Lemon Grass Lg



	Valid 100% Error 0% Empty 0% 3 distinct, 3 unique	Valid 100% Error 0% Empty 0% 3 distinct, 3 unique
1	5	Lower Manhattan
2	8	Hell's Kitchen

Transactions Table

Products Table

Stores Table

Data Modeling Solution

Q1-Identify Fact and Dimension Tables.

Transaction Table is a Fact table because it contains:

- TransactionID = Unique identifier (Primary Key)
- Date & Time = When the transaction happened
- Quantity = Measurable metric (Fact)
- StoreID = Foreign Key linking to the Stores table
- ProductID = Foreign Key linking to the Product table

Stores Table is a Dimesnion table because it contains:

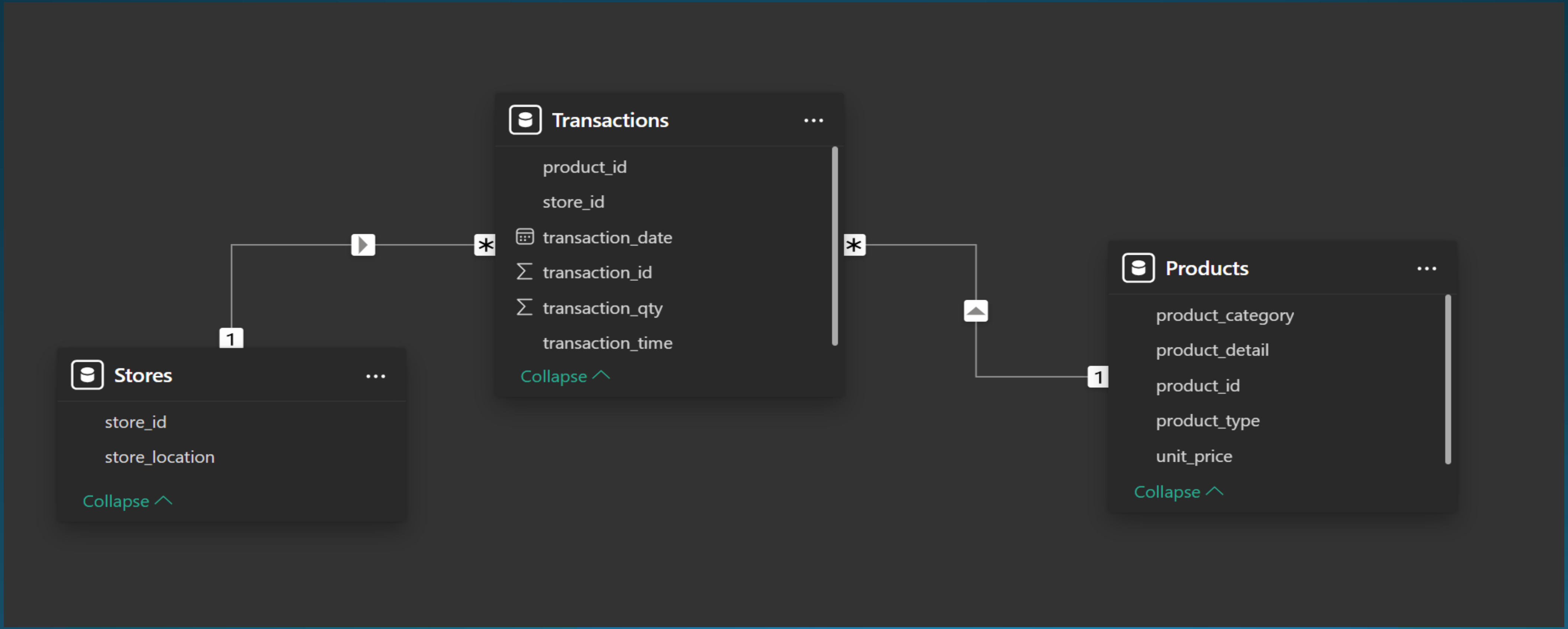
- StoreID = Unique identifier (Primary Key)
- Location = Descriptive attribute

Products Table is a Dimension table because it contains:

- ProductID (Foreign Key)
- Quantity Sold
- Total Sales Amount

Data Modeling Solution

- Q2-Create Relationships between tables.
- **Stores → Transactions → One-to-Many (1:*)**
- **Products → Transactions → One-to-Many (1:*)**



Transactions
- product_id
- store_id
- transaction_date
- transaction_id
- transaction_qty
- transaction_time

Stores
- store_id
- store_location

Products
- product_category
- product_detail
- product_id
- product_type
- unit_price

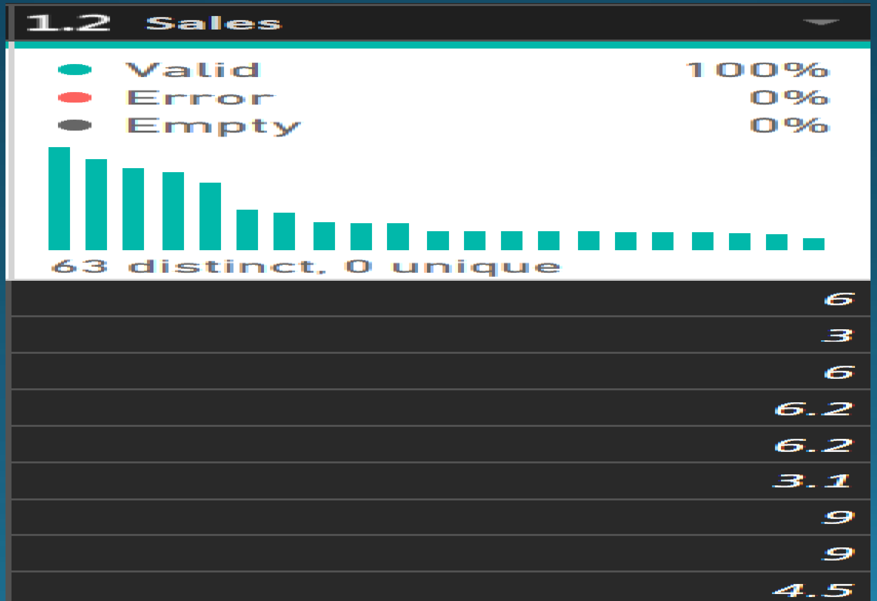
Data Modeling Solution

- Q3-Identify the schema.
- This a Star Schema.
- Single Fact Table (Transactions) at the center.
- Direct connections from the Fact Table to all Dimension Tables

Power Query Analysis Solution

Q-Create a column for Sales ○ Merge column "Unit price" from "products" to the "Transaction" table. ○ Create a custom column: Sales

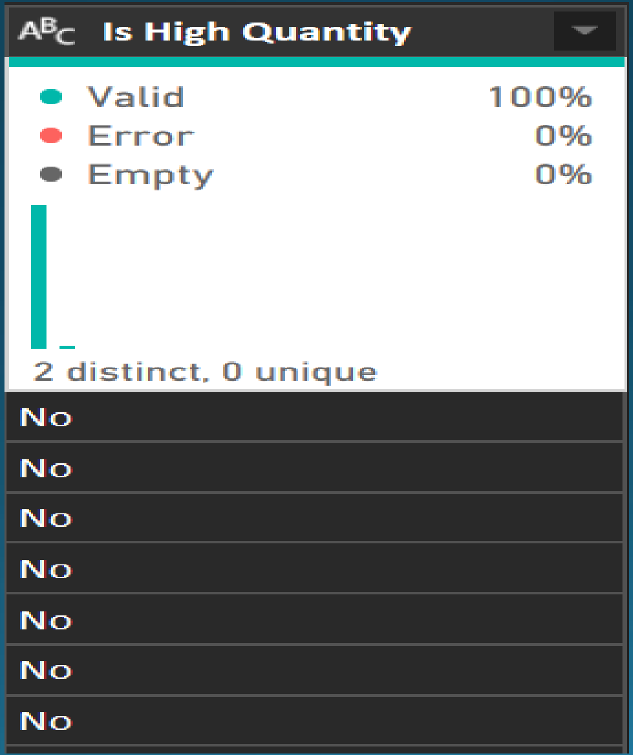
ANS-First, merge the **Unit Price** column from the **Products** table into the **Transactions** table. In Power Query, click **Merge Queries**, select **Transactions** as the primary table and **Products** as the secondary table, then match them using the **product_id** column. Use a **inner Join** to retain all transaction records and expand the merged table to include the **Unit Price** column. Next, add a custom column by navigating to **Add Column** → **Custom Column**, naming it **Sales**, and using the formula Sales = unit price x transaction_qty.



```
1.2 Sales
● Valid     100%
● Error       0%
● Empty       0%
63 distinct, 0 unique
  6
  3
  6
6.2
6.2
3.1
  9
  9
4.5
```

Power Query Analysis Solution

- Conditional column: ○ Create a conditional column Is High Quantity: If transaction_qty > 4, return "Yes", otherwise "No".



ABC Is High Quantity

● Valid	100%
● Error	0%
● Empty	0%

2 distinct, 0 unique

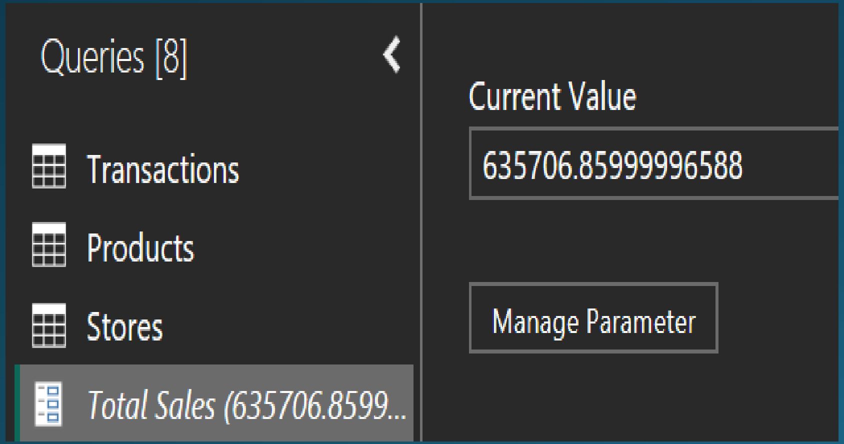
```
No
No
No
No
No
No
No
```

Power Query Analysis Solution

- Parameters: ○ Calculate the given and store them as parameters:

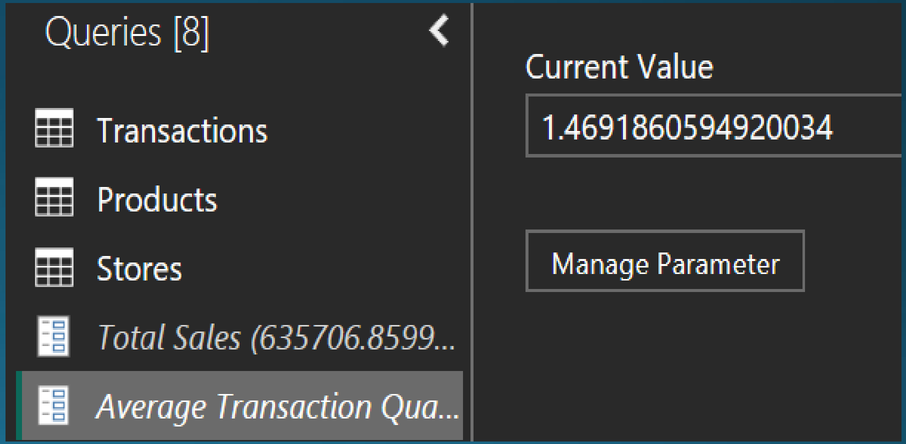
i. Calculate Total Sales: Sum of Total Sales.

ii. Calculate Average Transaction Quantity: Average of transaction_qty.



Current Value

635706.85999996588

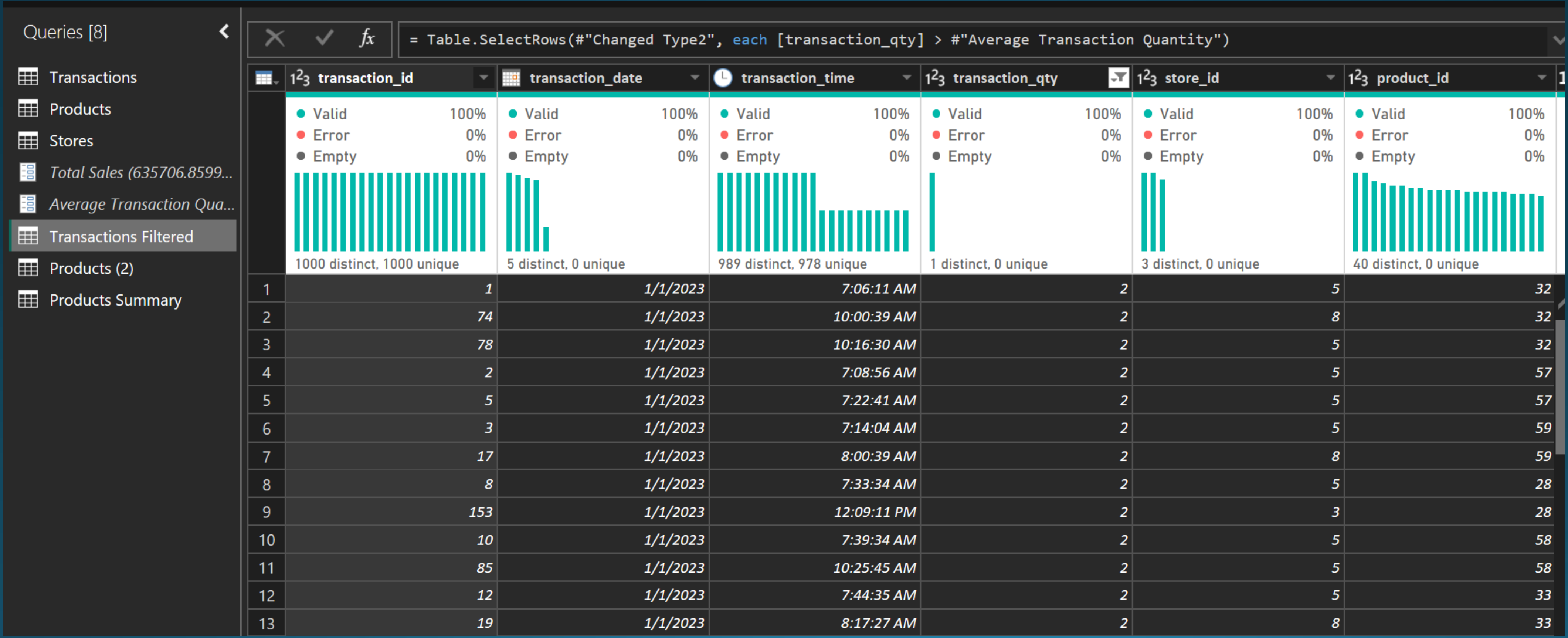


Current Value

1.4691860594920034

Power Query Analysis Solution

- Filter based on parameters: ○ Create a duplicate of the Transactions Table. ○ Filter the transactions with a quantity greater than the parameter " Average transaction quantity".



Queries [8]

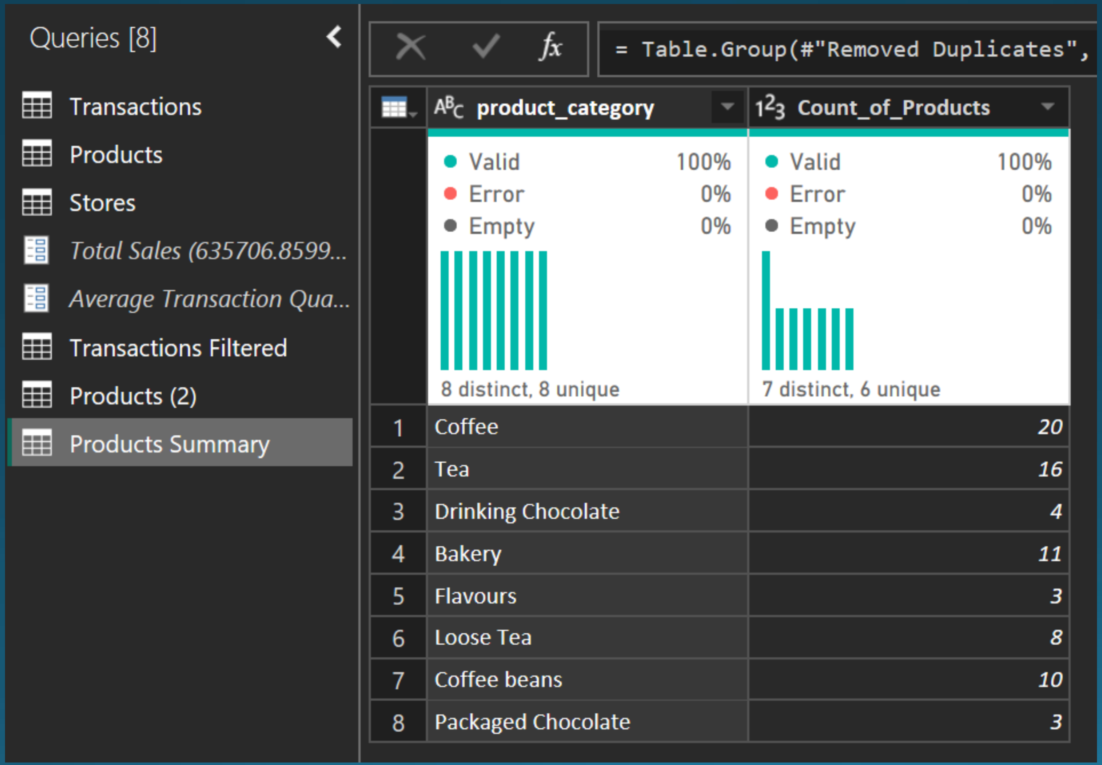
- Transactions
- Products
- Stores
- *Total Sales (635706.8599...*
- *Average Transaction Qua...*
- Transactions Filtered
- Products (2)
- Products Summary

```
= Table.SelectRows(#"Changed Type2", each [transaction_qty] > #"Average Transaction Quantity")
```

	transaction_id	transaction_date	transaction_time	transaction_qty	store_id	product_id
	Valid 100% Error 0% Empty 0% 1000 distinct, 1000 unique	Valid 100% Error 0% Empty 0% 5 distinct, 0 unique	Valid 100% Error 0% Empty 0% 989 distinct, 978 unique	Valid 100% Error 0% Empty 0% 1 distinct, 0 unique	Valid 100% Error 0% Empty 0% 3 distinct, 0 unique	Valid 100% Error 0% Empty 0% 40 distinct, 0 unique
1	1	1/1/2023	7:06:11 AM	2	5	32
2	74	1/1/2023	10:00:39 AM	2	8	32
3	78	1/1/2023	10:16:30 AM	2	5	32
4	2	1/1/2023	7:08:56 AM	2	5	57
5	5	1/1/2023	7:22:41 AM	2	5	57
6	3	1/1/2023	7:14:04 AM	2	5	59
7	17	1/1/2023	8:00:39 AM	2	8	59
8	8	1/1/2023	7:33:34 AM	2	5	28
9	153	1/1/2023	12:09:11 PM	2	3	28
10	10	1/1/2023	7:39:34 AM	2	5	58
11	85	1/1/2023	10:25:45 AM	2	5	58
12	12	1/1/2023	7:44:35 AM	2	5	33
13	19	1/1/2023	8:17:27 AM	2	8	33

Power Query Analysis Solution

- Sales Based on Location: ○ Merge Sales from “Transaction Table” to “Store” and show the aggregated value “Sum of Sales”.



Queries [8]

- Transactions
- Products
- Stores
- *Total Sales (635706.8599...*
- *Average Transaction Qua...*
- Transactions Filtered
- Products (2)
- Products Summary

```
= Table.Group(#"Removed Duplicates",
```

	product_category	Count_of_Products
1	Coffee	*20*
2	Tea	*16*
3	Drinking Chocolate	*4*
4	Bakery	*11*
5	Flavours	*3*
6	Loose Tea	*8*
7	Coffee beans	*10*
8	Packaged Chocolate	*3*