CLEANING AND TRANSFORMING THE DATASET USING POWER QUERY

Introduction

Source File: malltransactions_mockdataset.csv

Description of the file

- A mocked data of well-known malls in the capital region of the Philippines
- Created from mockaroo.com

Dimensions

- transaction_id
- date
- mall
- city
- gender
- age
- product category
- price
- quantity
- discount
- payment

Business Requirements

- All Ages less than 21 or 60 and above will have 20% discount
- For undeclared ages, replaced it with average age per gender and per mall.
- For undeclared gender, replace it with “Undeclared”

Business Notes

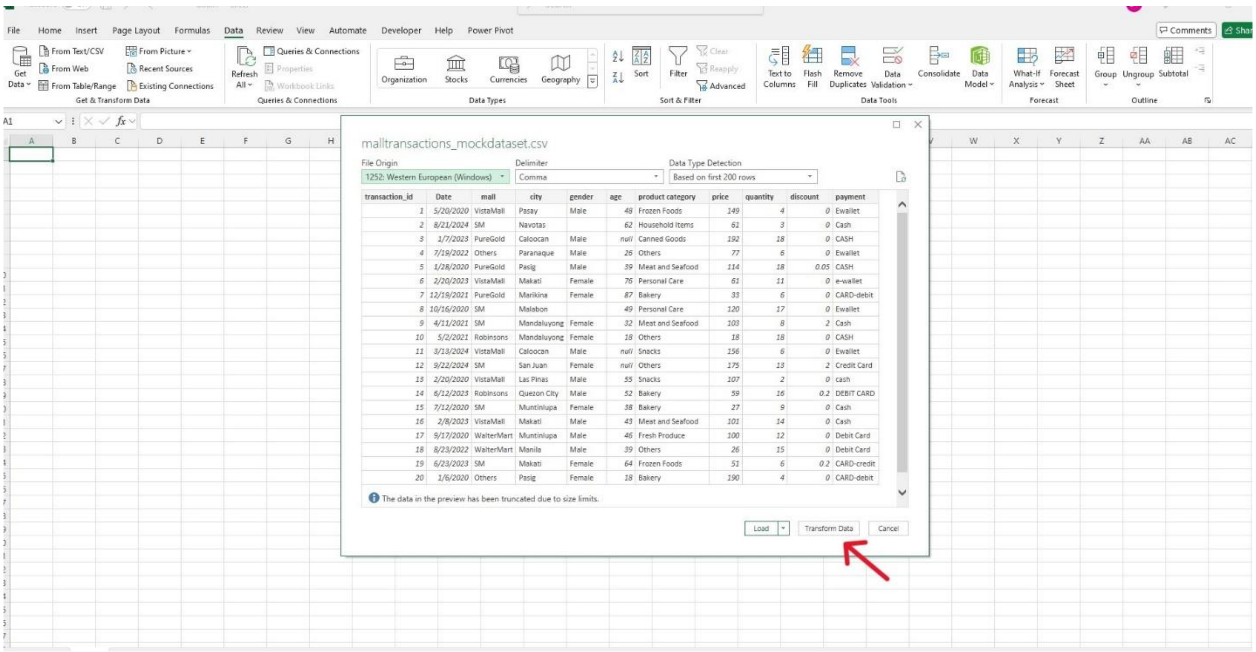
- There are data entry mistakes on discounts and payment

Steps In Cleaning and Transforming the Dataset

1. Open file in Power Query
2. Do an Exploratory Data Analysis of the file
    a. Check for error/null values on each column
    b. Check for misspelled / case inconsistency in the values
    c. Check for Outliers
3. Clean all necessary columns
    a. Correct misspelled / case inconsistency values
    b. Filter or Replace Outliers with standard values
    c. Replace or Remove error/null values
    d. Check if all columns have the right data type and format
4. Save and Close file to be used on Reports/Creating Dashboard

Step 1: Open file in Power Query

- I use Excel to use the built-in Power Query tool.
- I clicked the Transform Data for the next Step.

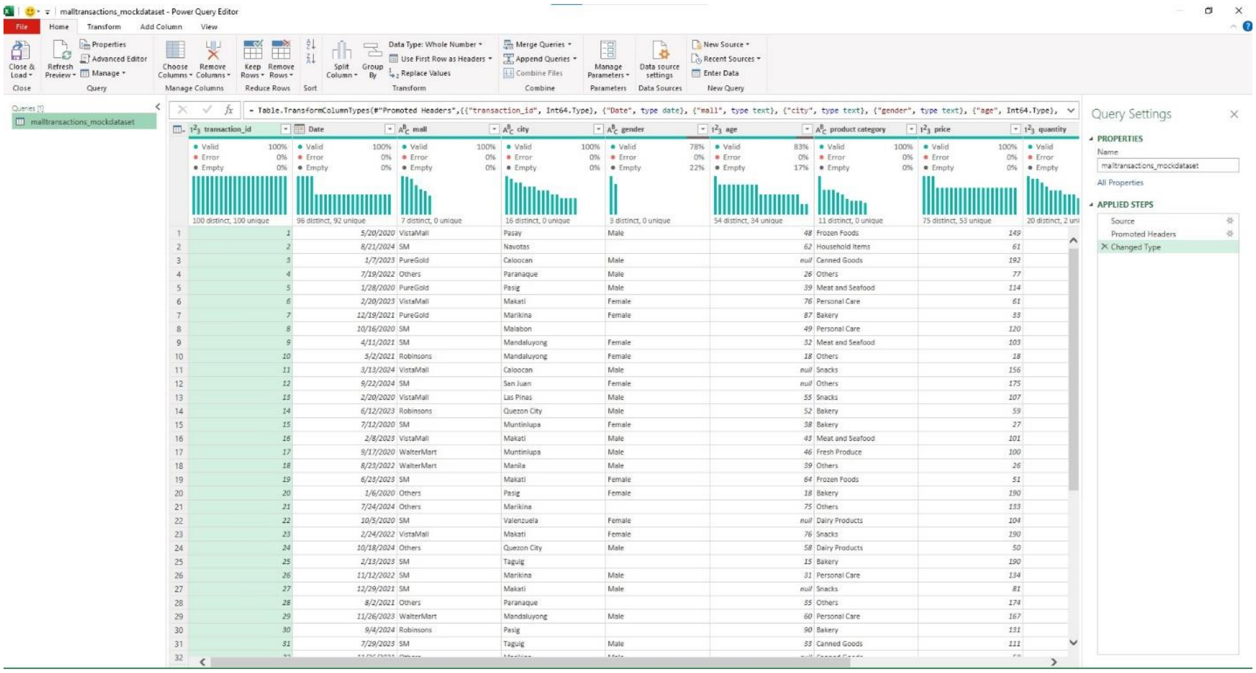


malltransactions_mockdataset.csv

File Origin: 1252: Western European (Windows)
Delimiter: Comma
Data Type Detection: Based on first 200 rows

transaction_id	Date	mall	city	gender	age	product category	price	quantity	discount	payment
1	5/20/2020	VistaMall	Pasay	Male	48	Frozen Foods	149	4	0	Ewallet
2	8/21/2024	SM	Navotas		62	Household Items	61	3	0	Cash
3	1/7/2023	PureGold	Caloocan	Male	null	Canned Goods	192	18	0	CASH
4	7/19/2022	Others	Paranaque	Male	26	Others	77	6	0	Ewallet
5	1/28/2020	PureGold	Pasig	Male	39	Meat and Seafood	114	18	0.05	CASH
6	2/20/2023	VistaMall	Makati	Female	76	Personal Care	61	11	0	e-wallet
7	12/19/2021	PureGold	Marikina	Female	87	Bakery	33	6	0	CARD-debit
8	10/16/2020	SM	Malabon		49	Personal Care	120	17	0	Ewallet
9	4/11/2021	SM	Mandaluyong	Female	32	Meat and Seafood	103	8	2	Cash
10	5/2/2021	Robinsons	Mandaluyong	Female	18	Others	18	18	0	CASH
11	3/13/2024	VistaMall	Caloocan	Male	null	Snacks	156	6	0	Ewallet
12	9/22/2024	SM	San Juan	Female	null	Others	175	13	2	Credit Card
13	2/20/2020	VistaMall	Las Pinas	Male	55	Snacks	107	2	0	cash
14	6/12/2023	Robinsons	Quezon City	Male	52	Bakery	59	16	0.2	DEBIT CARD
15	7/12/2020	SM	Muntinlupa	Female	38	Bakery	27	9	0	Cash
16	2/8/2023	VistaMall	Makati	Male	43	Meat and Seafood	101	14	0	Cash
17	9/17/2020	WalterMart	Muntinlupa	Male	46	Fresh Produce	100	12	0	Debit Card
18	8/23/2022	WalterMart	Manila	Male	39	Others	26	15	0	Debit Card
19	6/23/2023	SM	Makati	Female	64	Frozen Foods	51	6	0.2	CARD-credit
20	1/6/2020	Others	Pasig	Female	18	Bakery	190	4	0	CARD-debit

The data in the preview has been truncated due to size limits.



```
= Table.TransformColumnTypes(#"Promoted Headers",{{"transaction_id", Int64.Type}, {"Date", type date}, {"mall", type text}, {"city", type text}, {"gender", type text}, {"age", Int64.Type},
```

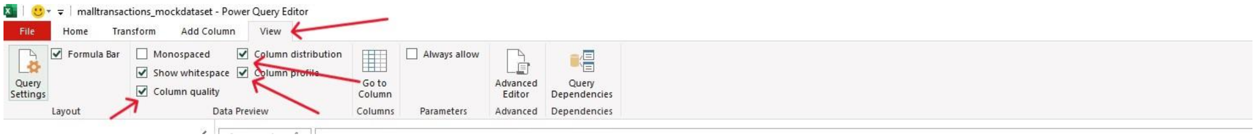
	transaction_id	Date	mall	city	gender	age	product category	price	quantity
	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 78% Error 0% Empty 22%	Valid 83% Error 0% Empty 17%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid Error Empty
	100 distinct, 100 unique	96 distinct, 92 unique	7 distinct, 0 unique	16 distinct, 0 unique	3 distinct, 0 unique	54 distinct, 34 unique	11 distinct, 0 unique	75 distinct, 53 unique	20 distinct, 2 uni
1	1	5/20/2020	VistaMall	Pasay	Male	48	Frozen Foods	149	
2	2	8/21/2024	SM	Navotas		62	Household Items	61	
3	3	1/7/2023	PureGold	Caloocan	Male	null	Canned Goods	192	
4	4	7/19/2022	Others	Paranaque	Male	26	Others	77	
5	5	1/28/2020	PureGold	Pasig	Male	39	Meat and Seafood	114	
6	6	2/20/2023	VistaMall	Makati	Female	76	Personal Care	61	
7	7	12/19/2021	PureGold	Marikina	Female	87	Bakery	33	
8	8	10/16/2020	SM	Malabon		49	Personal Care	120	
9	9	4/11/2021	SM	Mandaluyong	Female	32	Meat and Seafood	103	
10	10	5/2/2021	Robinsons	Mandaluyong	Female	18	Others	18	
11	11	3/13/2024	VistaMall	Caloocan	Male	null	Snacks	156	
12	12	9/22/2024	SM	San Juan	Female	null	Others	175	
13	13	2/20/2020	VistaMall	Las Pinas	Male	55	Snacks	107	
14	14	6/12/2023	Robinsons	Quezon City	Male	52	Bakery	59	
15	15	7/12/2020	SM	Muntinlupa	Female	38	Bakery	27	
16	16	2/8/2023	VistaMall	Makati	Male	43	Meat and Seafood	101	
17	17	9/17/2020	WalterMart	Muntinlupa	Male	46	Fresh Produce	100	
18	18	8/23/2022	WalterMart	Manila	Male	39	Others	26	
19	19	6/23/2023	SM	Makati	Female	64	Frozen Foods	51	
20	20	1/6/2020	Others	Pasig	Female	18	Bakery	190	
21	21	7/24/2024	Others	Marikina		75	Others	133	
22	22	10/5/2020	SM	Valenzuela	Female	null	Dairy Products	104	
23	23	2/24/2022	VistaMall	Makati	Female	76	Snacks	190	
24	24	10/18/2024	Others	Quezon City	Male	58	Dairy Products	50	
25	25	2/13/2023	SM	Taguig		15	Bakery	190	
26	26	11/12/2022	SM	Marikina	Male	31	Personal Care	134	
27	27	12/29/2021	SM	Makati	Male	null	Snacks	81	
28	28	8/2/2021	Others	Paranaque		35	Others	174	
29	29	11/26/2023	WalterMart	Mandaluyong	Male	60	Personal Care	167	
30	30	9/4/2024	Robinsons	Pasig		90	Bakery	131	
31	31	7/29/2023	SM	Taguig	Male	33	Canned Goods	111	

Step 2: Do an Exploratory Data Analysis (EDA)

***On View Ribbon, check the following*:**

- Column Profile

- Column Quality
- Column Distribution

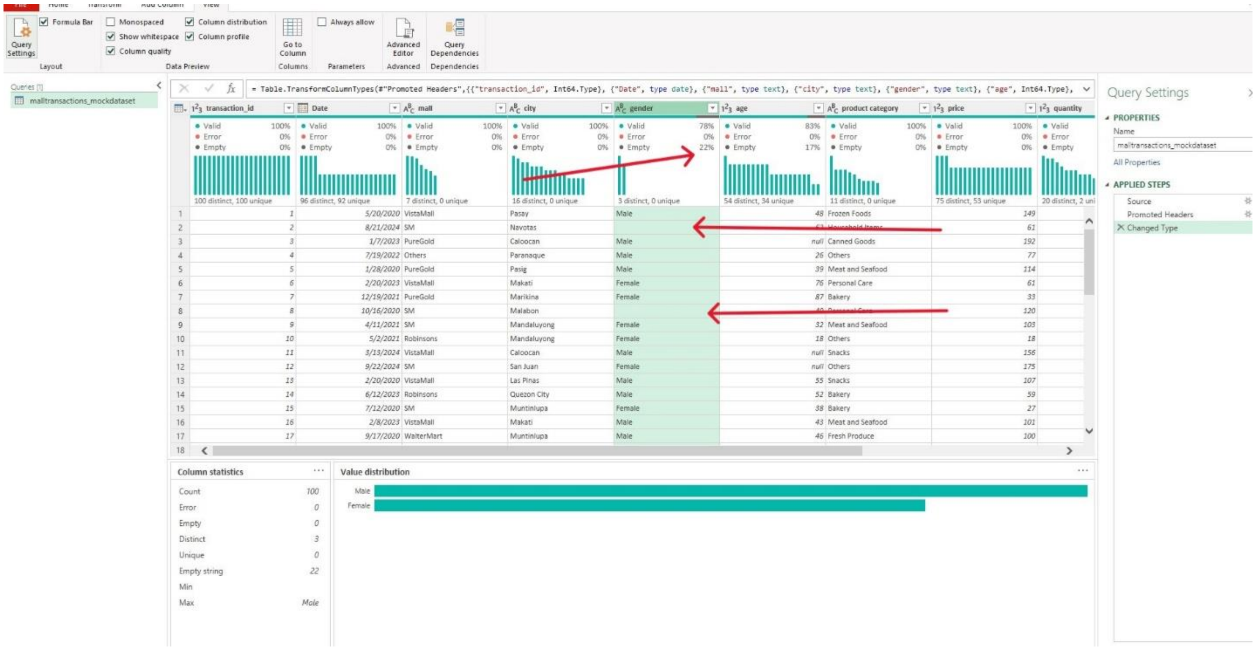


This is used for EDA, checking all the needed things for cleaning and transforming.

Check all column profiles, column quality, column distribution of each column

Gender column

- Gender column have EMPTY cells which is 22%.



```
= Table.TransformColumnTypes(#"Promoted Headers",{{"transaction_id", Int64.Type}, {"Date", type date}, {"mall", type text}, {"city", type text}, {"gender", type text}, {"age", Int64.Type},
```

	transaction_id	Date	mall	city	gender	age	product category	price	quantity
	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 78% Error 0% Empty 22%	Valid 83% Error 0% Empty 17%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid Error Empty
	100 distinct, 100 unique	96 distinct, 92 unique	7 distinct, 0 unique	16 distinct, 0 unique	3 distinct, 0 unique	54 distinct, 34 unique	11 distinct, 0 unique	75 distinct, 53 unique	20 distinct, 2 uni
1	1	5/20/2020	VistaMall	Pasay	Male	48	Frozen Foods	149	
2	2	8/21/2024	SM	Navotas		63	Household Items	61	
3	3	1/7/2023	PureGold	Caloocan	Male	null	Canned Goods	192	
4	4	7/19/2022	Others	Paranaque	Male	26	Others	77	
5	5	1/28/2020	PureGold	Pasig	Male	39	Meat and Seafood	114	
6	6	2/20/2023	VistaMall	Makati	Female	76	Personal Care	61	
7	7	12/19/2021	PureGold	Marikina	Female	87	Bakery	33	
8	8	10/16/2020	SM	Malabon		40	Personal Care	120	
9	9	4/11/2021	SM	Mandaluyong	Female	32	Meat and Seafood	103	
10	10	5/2/2021	Robinsons	Mandaluyong	Female	18	Others	18	
11	11	3/13/2024	VistaMall	Caloocan	Male	null	Snacks	156	
12	12	9/22/2024	SM	San Juan	Female	null	Others	175	
13	13	2/20/2020	VistaMall	Las Pinas	Male	55	Snacks	107	
14	14	6/12/2023	Robinsons	Quezon City	Male	52	Bakery	59	
15	15	7/12/2020	SM	Muntinlupa	Female	38	Bakery	27	
16	16	2/8/2023	VistaMall	Makati	Male	43	Meat and Seafood	101	
17	17	9/17/2020	WalterMart	Muntinlupa	Male	46	Fresh Produce	100	
18									

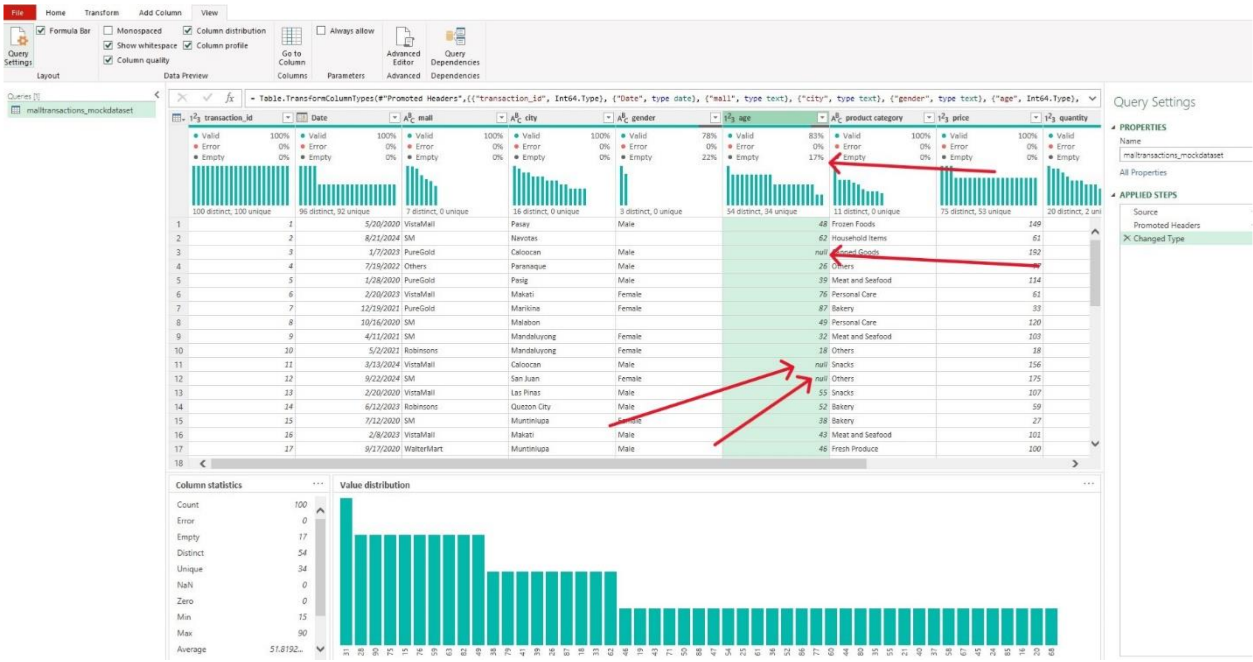
Column statistics

Count	100
Error	0
Empty	0
Distinct	3
Unique	0
Empty string	22
Min	
Max	Male

Value distribution

Age column

- Age column has null values and is about 17%.



```
= Table.TransformColumnTypes(#"Promoted Headers",{{"transaction_id", Int64.Type}, {"Date", type date}, {"mall", type text}, {"city", type text}, {"gender", type text}, {"age", Int64.Type},
```

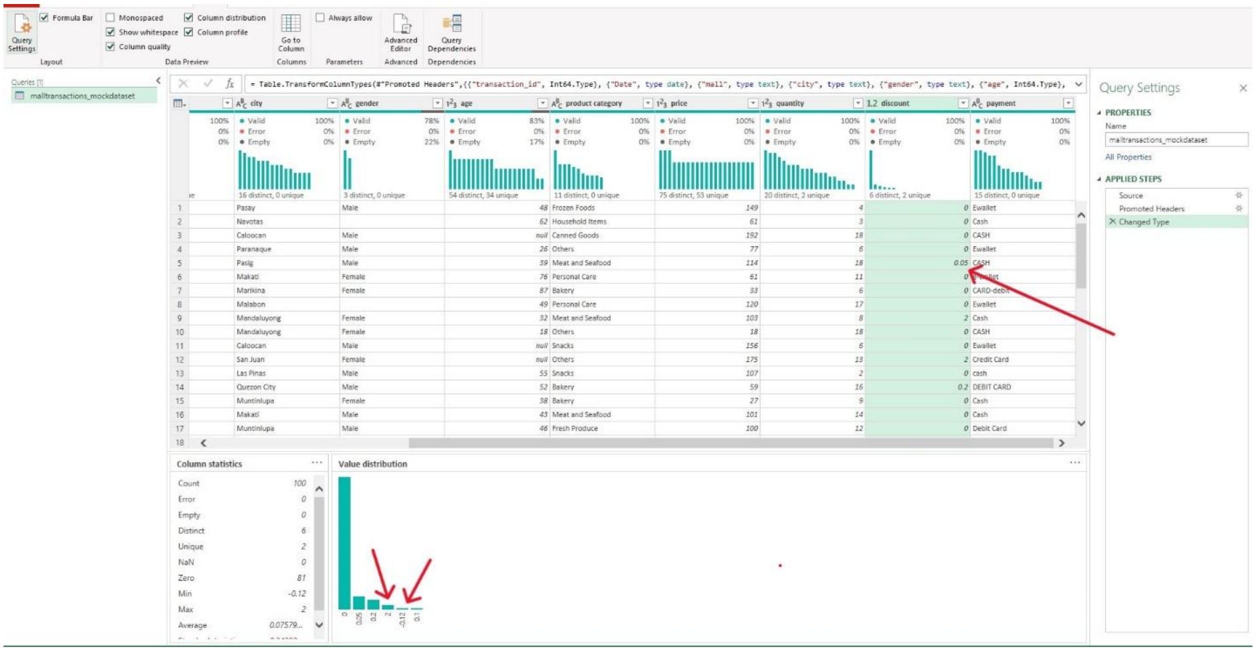
	transaction_id	Date	mall	city	gender	age	product category	price	quantity
Valid	100%	100%	100%	100%	78%	83%	100%	100%	
Error	0%	0%	0%	0%	0%	0%	0%	0%	
Empty	0%	0%	0%	0%	22%	17%	0%	0%	
	100 distinct, 100 unique	96 distinct, 92 unique	7 distinct, 0 unique	16 distinct, 0 unique	3 distinct, 0 unique	54 distinct, 34 unique	11 distinct, 0 unique	75 distinct, 53 unique	20 distinct, 2 uni
1	1	5/20/2020	VistaMall	Pasay	Male	48	Frozen Foods	149	
2	2	8/21/2024	SM	Navotas		62	Household Items	61	
3	3	1/7/2023	PureGold	Caloocan	Male	null	Canned Goods	192	
4	4	7/19/2022	Others	Paranaque	Male	26	Others	77	
5	5	1/28/2020	PureGold	Pasig	Male	39	Meat and Seafood	114	
6	6	2/20/2023	VistaMall	Makati	Female	76	Personal Care	61	
7	7	12/19/2021	PureGold	Marikina	Female	87	Bakery	33	
8	8	10/16/2020	SM	Malabon		49	Personal Care	120	
9	9	4/11/2021	SM	Mandaluyong	Female	32	Meat and Seafood	103	
10	10	5/2/2021	Robinsons	Mandaluyong	Female	18	Others	18	
11	11	3/13/2024	VistaMall	Caloocan	Male	null	Snacks	156	
12	12	9/22/2024	SM	San Juan	Female	null	Others	175	
13	13	2/20/2020	VistaMall	Las Pinas	Male	55	Snacks	107	
14	14	6/12/2023	Robinsons	Quezon City	Male	52	Bakery	59	
15	15	7/12/2020	SM	Muntinlupa	Female	38	Bakery	27	
16	16	2/8/2023	VistaMall	Makati	Male	43	Meat and Seafood	101	
17	17	9/17/2020	WalterMart	Muntinlupa	Male	46	Fresh Produce	100	
18									

Column statistics

Count	100
Error	0
Empty	17
Distinct	54
Unique	34
NaN	0
Zero	0
Min	15
Max	90
Average	51.8192...

Discount column

- Discount column has negative values and an Outlier. Data should be in percentage format.



```
= Table.TransformColumnTypes(#"Promoted Headers",{{"transaction_id", Int64.Type}, {"Date", type date}, {"mall", type text}, {"city", type text}, {"gender", type text}, {"age", Int64.Type},
```

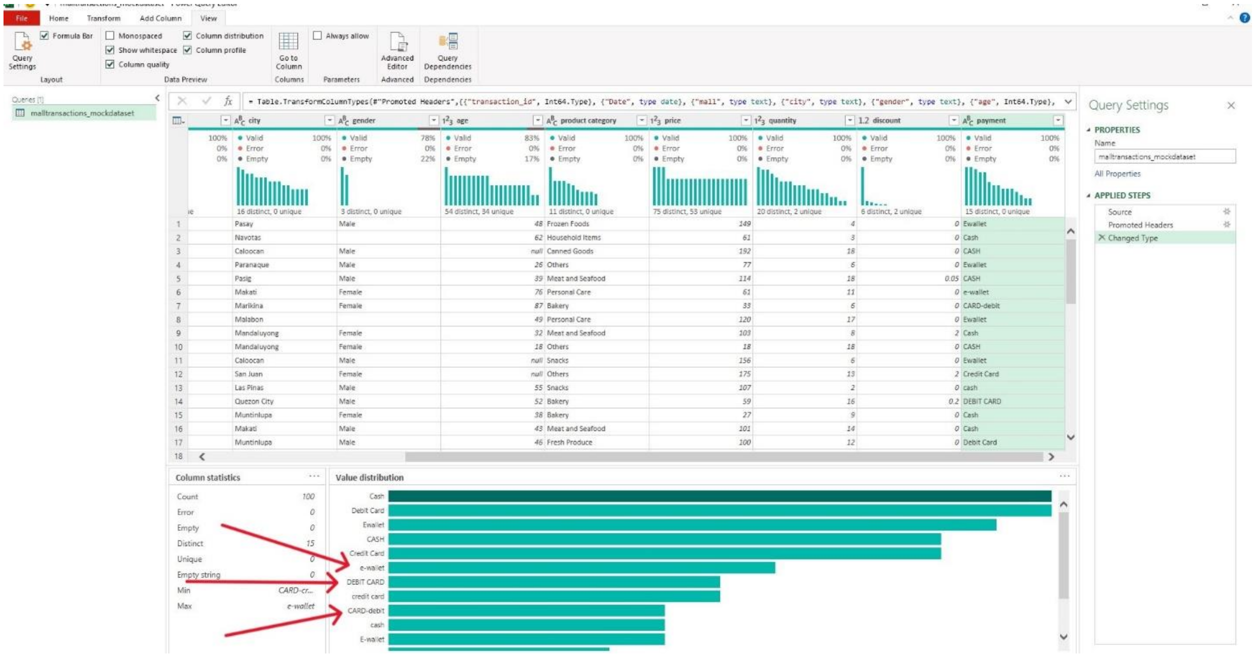
	city	gender	age	product category	price	quantity	discount	payment
Valid	100%	78%	83%	100%	100%	100%	100%	100%
Error	0%	0%	0%	0%	0%	0%	0%	0%
Empty	0%	22%	17%	0%	0%	0%	0%	0%
	16 distinct, 0 unique	3 distinct, 0 unique	54 distinct, 34 unique	11 distinct, 0 unique	75 distinct, 53 unique	20 distinct, 2 unique	6 distinct, 2 unique	15 distinct, 0 unique
1	Pasay	Male	48	Frozen Foods	149	4	0	Ewallet
2	Navotas		62	Household Items	61	3	0	Cash
3	Caloocan	Male	null	Canned Goods	192	18	0	CASH
4	Paranaque	Male	26	Others	77	6	0	Ewallet
5	Pasig	Male	39	Meat and Seafood	114	18	0.05	CASH
6	Makati	Female	76	Personal Care	61	11	0	Ewallet
7	Marikina	Female	87	Bakery	33	6	0	CARD-debit
8	Malabon		49	Personal Care	120	17	0	Ewallet
9	Mandaluyong	Female	32	Meat and Seafood	103	8	2	Cash
10	Mandaluyong	Female	18	Others	18	18	0	CASH
11	Caloocan	Male	null	Snacks	156	6	0	Ewallet
12	San Juan	Female	null	Others	175	13	2	Credit Card
13	Las Pinas	Male	55	Snacks	107	2	0	cash
14	Quezon City	Male	52	Bakery	59	16	0.2	DEBIT CARD
15	Muntinlupa	Female	38	Bakery	27	9	0	Cash
16	Makati	Male	43	Meat and Seafood	101	14	0	Cash
17	Muntinlupa	Male	46	Fresh Produce	100	12	0	Debit Card
18								

Column statistics

Count	100
Error	0
Empty	0
Distinct	6
Unique	2
NaN	0
Zero	81
Min	-0.12
Max	2
Average	0.07579...

Payment column

- Payment column has misspelled / case inconsistency in its values



```
= Table.TransformColumnTypes(#"Promoted Headers",{{"transaction_id", Int64.Type}, {"Date", type date}, {"mall", type text}, {"city", type text}, {"gender", type text}, {"age", Int64.Type},
```

	city	gender	age	product category	price	quantity	discount	payment
	Valid 100% Error 0% Empty 0%	Valid 78% Error 0% Empty 22%	Valid 83% Error 0% Empty 17%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%
	16 distinct, 0 unique	3 distinct, 0 unique	54 distinct, 34 unique	11 distinct, 0 unique	75 distinct, 53 unique	20 distinct, 2 unique	6 distinct, 2 unique	15 distinct, 0 unique
1	Pasay	Male	48	Frozen Foods	149	4	0	Ewallet
2	Navotas		62	Household Items	61	3	0	Cash
3	Caloocan	Male	null	Canned Goods	192	18	0	CASH
4	Paranaque	Male	26	Others	77	6	0	Ewallet
5	Pasig	Male	39	Meat and Seafood	114	18	0.05	CASH
6	Makati	Female	76	Personal Care	61	11	0	e-wallet
7	Marikina	Female	87	Bakery	33	6	0	CARD-debit
8	Malabon		49	Personal Care	120	17	0	Ewallet
9	Mandaluyong	Female	32	Meat and Seafood	103	8	2	Cash
10	Mandaluyong	Female	18	Others	18	18	0	CASH
11	Caloocan	Male	null	Snacks	156	6	0	Ewallet
12	San Juan	Female	null	Others	175	13	2	Credit Card
13	Las Pinas	Male	55	Snacks	107	2	0	cash
14	Quezon City	Male	52	Bakery	59	16	0.2	DEBIT CARD
15	Muntinlupa	Female	38	Bakery	27	9	0	Cash
16	Makati	Male	43	Meat and Seafood	101	14	0	Cash
17	Muntinlupa	Male	46	Fresh Produce	100	12	0	Debit Card
18								

Column statistics

Count	100
Error	0
Empty	0
Distinct	15
Unique	0
Empty string	0
Min	CARD-cr...
Max	e-wallet

Value distribution

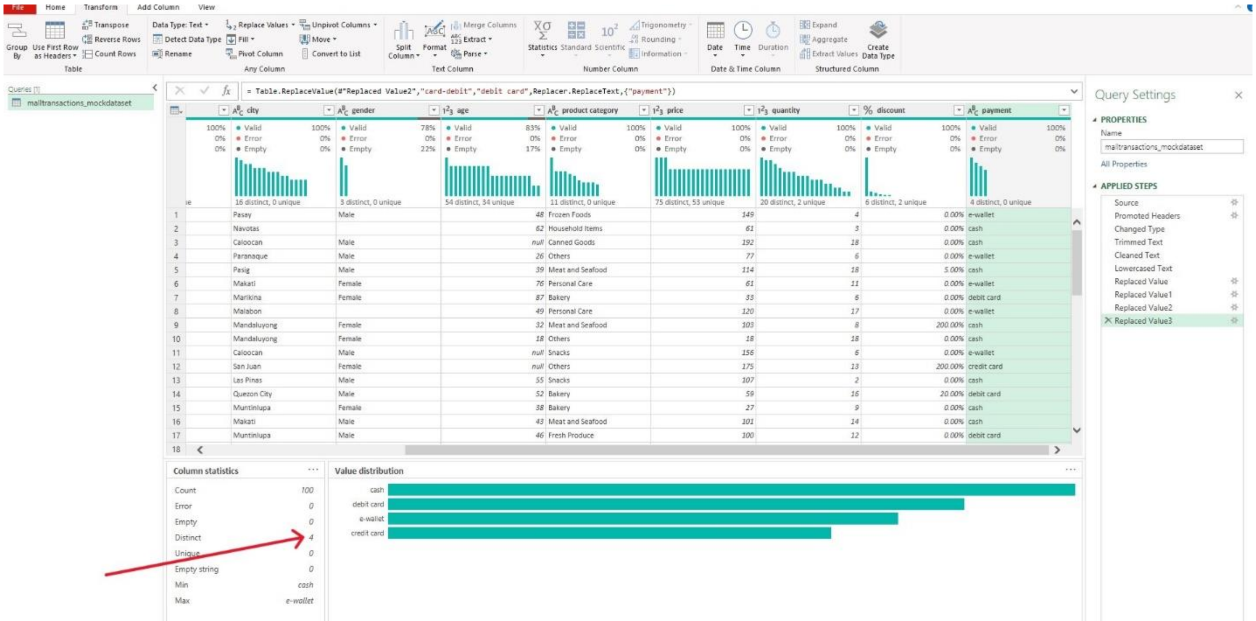
Name: malltransactions_mockdataset

APPLIED STEPS
- Source
- Promoted Headers
- Changed Type

Step 3: Clean all necessary columns

Payment Column

- For Payment Column, I trim, clean, and lower case the values of the column
- Replace ewallet to e-wallet, card-debit to debit card, coins to cash, card-credit to credit card for standardization
- Only 4 distinct values remain



```
= Table.ReplaceValue(#"Replaced Value2","card-debit","debit card",Replacer.ReplaceText,{"payment"})
```

	city	gender	age	product category	price	quantity	discount	payment
	Valid 100% Error 0% Empty 0%	Valid 78% Error 0% Empty 22%	Valid 83% Error 0% Empty 17%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%
	16 distinct, 0 unique	3 distinct, 0 unique	54 distinct, 34 unique	11 distinct, 0 unique	75 distinct, 53 unique	20 distinct, 2 unique	6 distinct, 2 unique	4 distinct, 0 unique
1	Pasay	Male	48	Frozen Foods	149	4	0.00%	e-wallet
2	Navotas		62	Household Items	61	3	0.00%	cash
3	Caloocan	Male	null	Canned Goods	192	18	0.00%	cash
4	Paranaque	Male	26	Others	77	6	0.00%	e-wallet
5	Pasig	Male	39	Meat and Seafood	114	18	5.00%	cash
6	Makati	Female	76	Personal Care	61	11	0.00%	e-wallet
7	Marikina	Female	87	Bakery	33	6	0.00%	debit card
8	Malabon		49	Personal Care	120	17	0.00%	e-wallet
9	Mandaluyong	Female	32	Meat and Seafood	103	8	200.00%	cash
10	Mandaluyong	Female	18	Others	18	18	0.00%	cash
11	Caloocan	Male	null	Snacks	156	6	0.00%	e-wallet
12	San Juan	Female	null	Others	175	13	200.00%	credit card
13	Las Pinas	Male	55	Snacks	107	2	0.00%	cash
14	Quezon City	Male	52	Bakery	59	16	20.00%	debit card
15	Muntinlupa	Female	38	Bakery	27	9	0.00%	cash
16	Makati	Male	43	Meat and Seafood	101	14	0.00%	cash
17	Muntinlupa	Male	46	Fresh Produce	100	12	0.00%	debit card
18								

Column statistics

Count	100
Error	0
Empty	0
Distinct	4
Unique	0
Empty string	0
Min	cash
Max	e-wallet

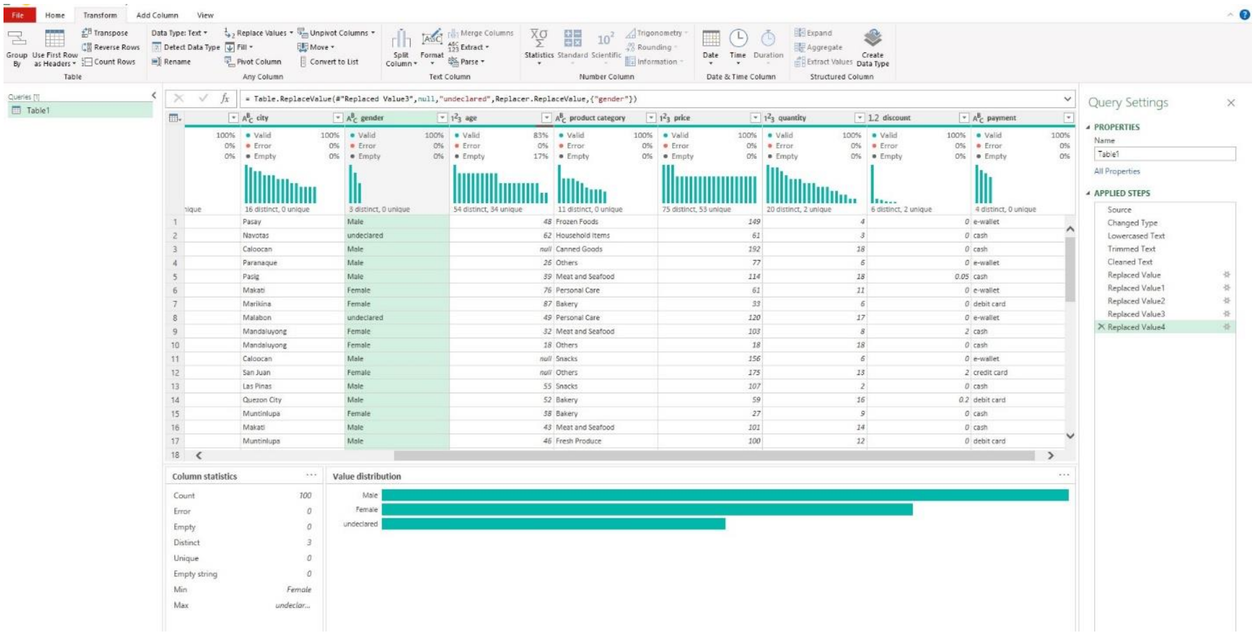
Value distribution

Name: malltransactions_mockdataset

APPLIED STEPS
- Source
- Promoted Headers
- Changed Type
- Trimmed Text
- Cleaned Text
- Lowercased Text
- Replaced Value
- Replaced Value1
- Replaced Value2
- Replaced Value3

Gender Column

- For Gender Column, I replaced blank values to undeclared values



```
= Table.ReplaceValue(#"Replaced Value3",null,"undeclared",Replacer.ReplaceValue,{"gender"})
```

	city	gender	age	product category	price	quantity	discount	payment
	16 distinct, 0 unique	3 distinct, 0 unique	54 distinct, 34 unique	11 distinct, 0 unique	75 distinct, 53 unique	20 distinct, 2 unique	6 distinct, 2 unique	4 distinct, 0 unique
1	Pasay	Male	48	Frozen Foods	149	4	0	e-wallet
2	Navotas	undeclared	62	Household Items	61	3	0	cash
3	Caloocan	Male	null	Canned Goods	192	18	0	cash
4	Paranaque	Male	26	Others	77	6	0	e-wallet
5	Pasig	Male	39	Meat and Seafood	114	18	0.05	cash
6	Makati	Female	76	Personal Care	61	11	0	e-wallet
7	Marikina	Female	87	Bakery	33	6	0	debit card
8	Malabon	undeclared	49	Personal Care	120	17	0	e-wallet
9	Mandaluyong	Female	32	Meat and Seafood	103	8	2	cash
10	Mandaluyong	Female	18	Others	18	18	0	cash
11	Caloocan	Male	null	Snacks	156	6	0	e-wallet
12	San Juan	Female	null	Others	175	13	2	credit card
13	Las Pinas	Male	55	Snacks	107	2	0	cash
14	Quezon City	Male	52	Bakery	59	16	0.2	debit card
15	Muntinlupa	Female	38	Bakery	27	9	0	cash
16	Makati	Male	43	Meat and Seafood	101	14	0	cash
17	Muntinlupa	Male	46	Fresh Produce	100	12	0	debit card

Column statistics

Count	100
Error	0
Empty	0
Distinct	3
Unique	0
Empty string	0
Min	Female
Max	undeclar...

Query Settings

PROPERTIES

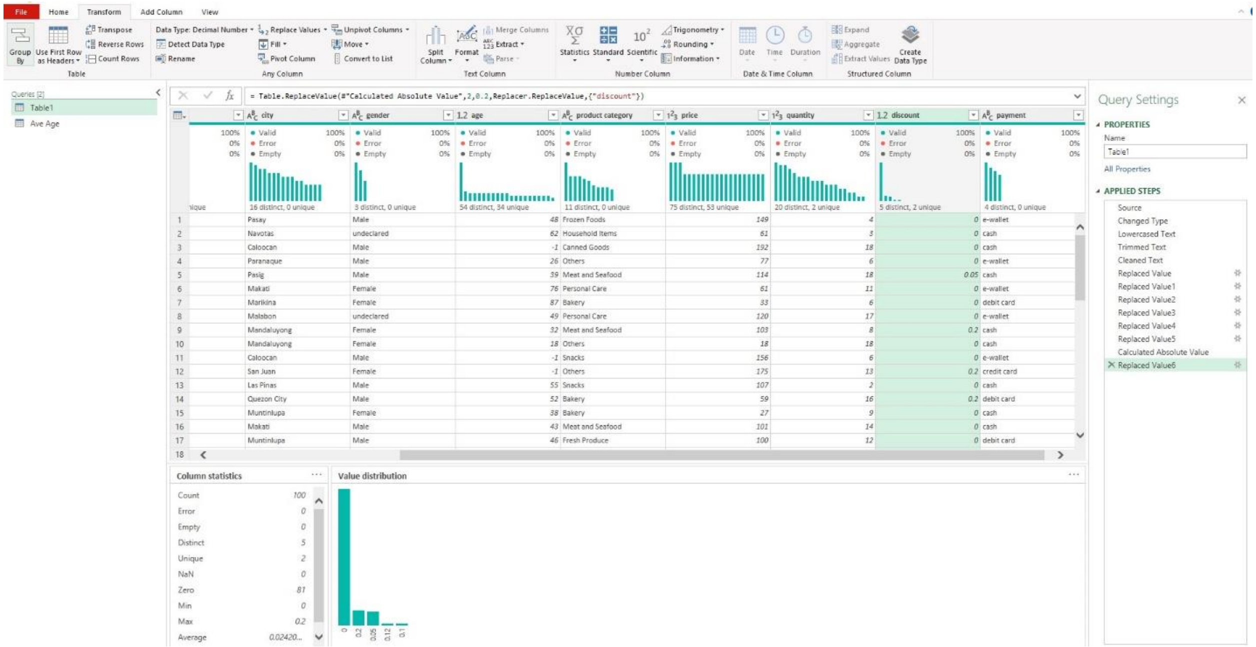
Name: Table1

APPLIED STEPS

- Source
- Changed Type
- Lowercased Text
- Trimmed Text
- Cleaned Text
- Replaced Value
- Replaced Value1
- Replaced Value2
- Replaced Value3
- Replaced Value4

Discount Column

- For Discount Column, I remove all negative values by using the absolute value feature
- Replace the Outlier 2 to .2



```
= Table.ReplaceValue(#"Calculated Absolute Value",2,0.2,Replacer.ReplaceValue,{"discount"})
```

	city	gender	age	product category	price	quantity	discount	payment
	16 distinct, 0 unique	3 distinct, 0 unique	54 distinct, 34 unique	11 distinct, 0 unique	75 distinct, 53 unique	20 distinct, 2 unique	5 distinct, 2 unique	4 distinct, 0 unique
1	Pasay	Male	48	Frozen Foods	149	4	0	e-wallet
2	Navotas	undeclared	62	Household Items	61	3	0	cash
3	Caloocan	Male	-1	Canned Goods	192	18	0	cash
4	Paranaque	Male	26	Others	77	6	0	e-wallet
5	Pasig	Male	39	Meat and Seafood	114	18	0.05	cash
6	Makati	Female	76	Personal Care	61	11	0	e-wallet
7	Marikina	Female	87	Bakery	33	6	0	debit card
8	Malabon	undeclared	49	Personal Care	120	17	0	e-wallet
9	Mandaluyong	Female	32	Meat and Seafood	103	8	0.2	cash
10	Mandaluyong	Female	18	Others	18	18	0	cash
11	Caloocan	Male	-1	Snacks	156	6	0	e-wallet
12	San Juan	Female	-1	Others	175	13	0.2	credit card
13	Las Pinas	Male	55	Snacks	107	2	0	cash
14	Quezon City	Male	52	Bakery	59	16	0.2	debit card
15	Muntinlupa	Female	38	Bakery	27	9	0	cash
16	Makati	Male	43	Meat and Seafood	101	14	0	cash
17	Muntinlupa	Male	46	Fresh Produce	100	12	0	debit card

Column statistics

Count	100
Error	0
Empty	0
Distinct	5
Unique	2
NaN	0
Zero	81
Min	0
Max	0.2
Average	0.02420...

Query Settings

PROPERTIES

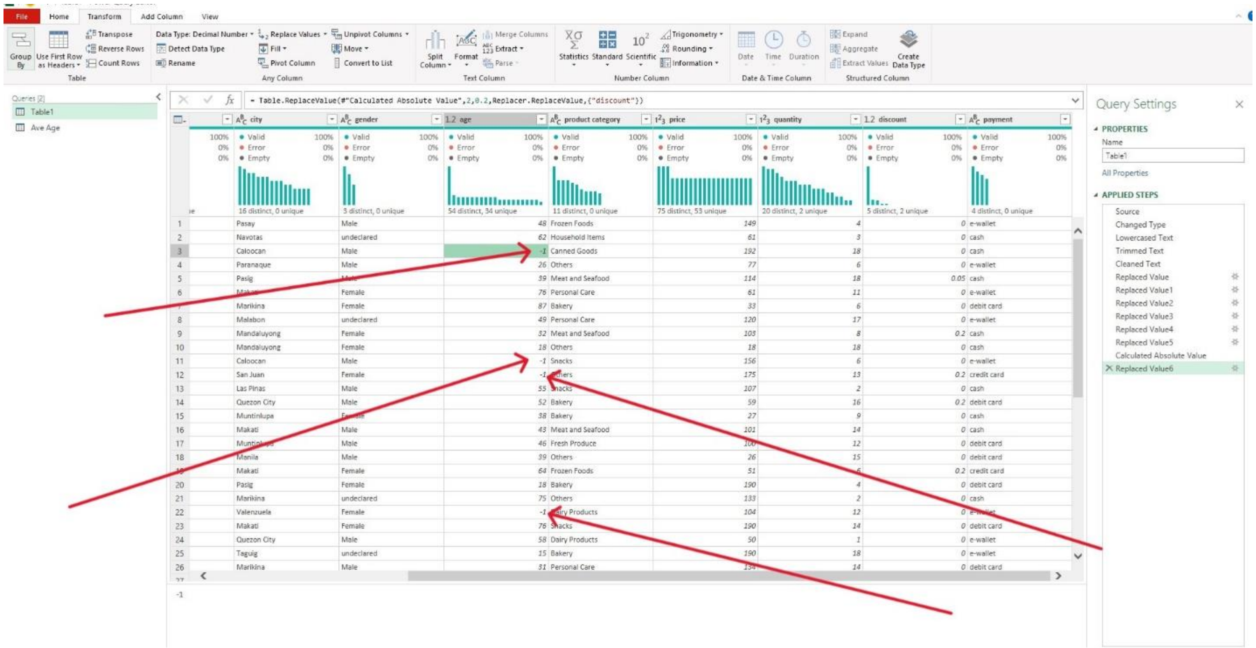
Name: Table1

APPLIED STEPS

- Source
- Changed Type
- Lowercased Text
- Trimmed Text
- Cleaned Text
- Replaced Value
- Replaced Value1
- Replaced Value2
- Replaced Value3
- Replaced Value4
- Replaced Value5
- Calculated Absolute Value
- Replaced Value6

- Created a new column “discount_new” that will change values based on the age (All Ages less than 21 or 60 and above will have 20% discount)

- Temporarily, I replace all null values with -1 on the Age column to avoid Errors after manipulation



```
= Table.ReplaceValue(#"Calculated Absolute Value",2,0.2,Replacer.ReplaceValue,{"discount"})
```

	city	gender	age	product category	price	quantity	discount	payment
	16 distinct, 0 unique	3 distinct, 0 unique	54 distinct, 34 unique	11 distinct, 0 unique	75 distinct, 53 unique	20 distinct, 2 unique	5 distinct, 2 unique	4 distinct, 0 unique
1	Pasay	Male	48	Frozen Foods	149	4	0	e-wallet
2	Navotas	undeclared	62	Household Items	61	3	0	cash
3	Caloocan	Male	-1	Canned Goods	192	18	0	cash
4	Paranaque	Male	26	Others	77	6	0	e-wallet
5	Pasig	Male	39	Meat and Seafood	114	18	0.05	cash
6	Makati	Female	76	Personal Care	61	11	0	e-wallet
7	Marikina	Female	87	Bakery	33	6	0	debit card
8	Malabon	undeclared	49	Personal Care	120	17	0	e-wallet
9	Mandaluyong	Female	32	Meat and Seafood	103	8	0.2	cash
10	Mandaluyong	Female	18	Others	18	18	0	cash
11	Caloocan	Male	-1	Snacks	156	6	0	e-wallet
12	San Juan	Female	-1	Others	175	13	0.2	credit card
13	Las Pinas	Male	55	Snacks	107	2	0	cash
14	Quezon City	Male	52	Bakery	59	16	0.2	debit card
15	Muntinlupa	Female	38	Bakery	27	9	0	cash
16	Makati	Male	43	Meat and Seafood	101	14	0	cash
17	Muntinlupa	Male	46	Fresh Produce	100	12	0	debit card
18	Manila	Male	39	Others	26	15	0	debit card
19	Makati	Female	64	Frozen Foods	51	6	0.2	credit card
20	Pasig	Female	18	Bakery	190	4	0	debit card
21	Marikina	undeclared	75	Others	133	2	0	cash
22	Valenzuela	Female	-1	Dairy Products	104	12	0	e-wallet
23	Makati	Female	76	Snacks	190	14	0	debit card
24	Quezon City	Male	58	Dairy Products	50	1	0	e-wallet
25	Taguig	undeclared	15	Bakery	190	18	0	e-wallet
26	Marikina	Male	31	Personal Care	134	14	0	debit card

-1

Query Settings

PROPERTIES

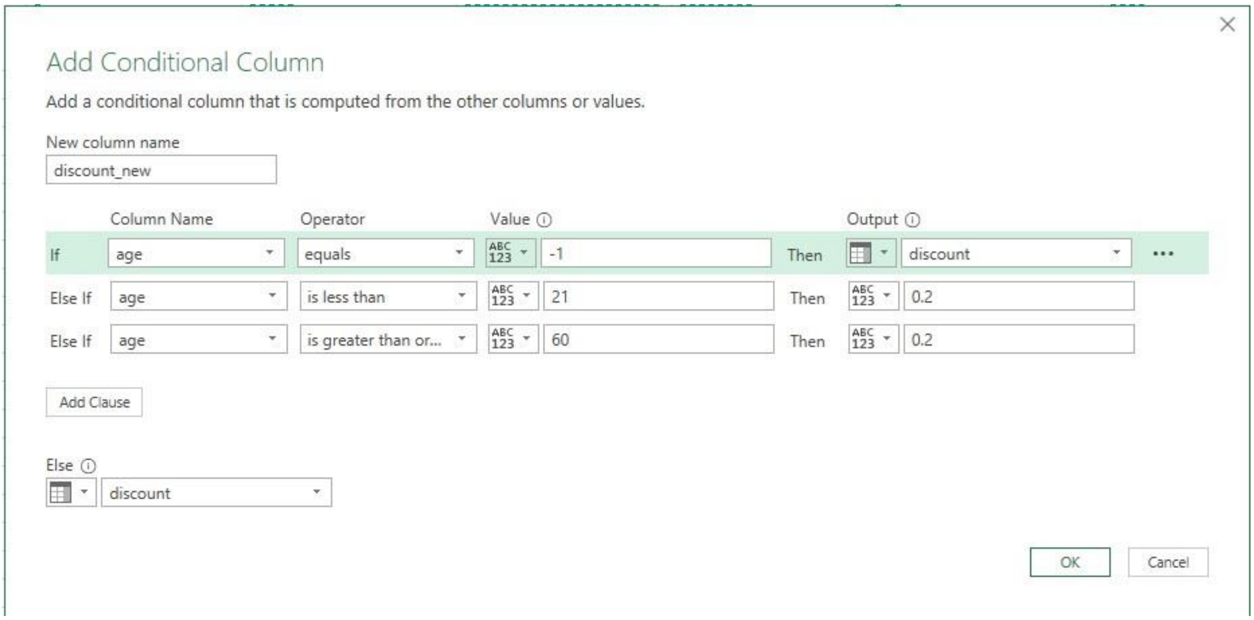
Name: Table1

All Properties

APPLIED STEPS

- Source
- Changed Type
- Lowercased Text
- Trimmed Text
- Cleaned Text
- Replaced Value
- Replaced Value1
- Replaced Value2
- Replaced Value3
- Replaced Value4
- Replaced Value5
- Calculated Absolute Value
- Replaced Value6

- I use conditional column for manipulation for the discount_new column



Add Conditional Column

Add a conditional column that is computed from the other columns or values.

New column name: discount_new

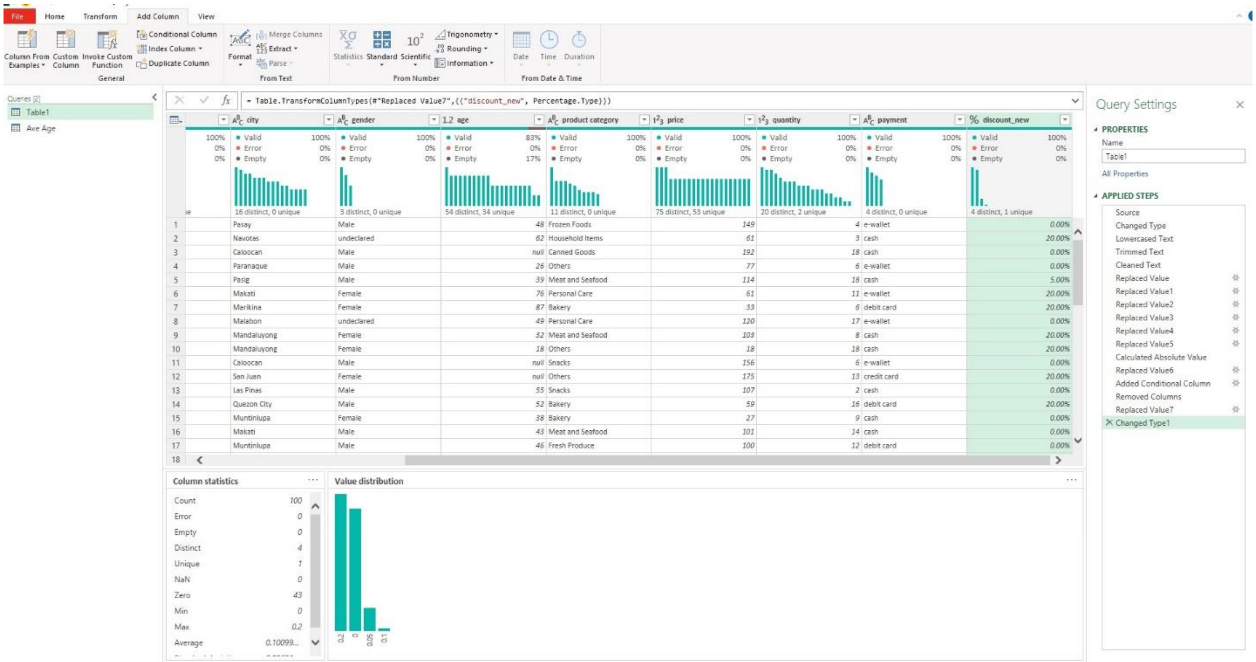
	Column Name	Operator	Value		Output
If	age	equals	ABC 123 -1	Then	discount
Else If	age	is less than	ABC 123 21	Then	ABC 123 0.2
Else If	age	is greater than or...	ABC 123 60	Then	ABC 123 0.2

Add Clause

Else: discount

OK Cancel

- Replace the -1 to Null again and delete the old column.
- Replace the format to percentage



```
= Table.TransformColumnTypes(#"Replaced Value7",{{"discount_new", Percentage.Type}})
```

		city	gender	age	product category	price	quantity	payment	discount_new
	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 83% Error 0% Empty 17%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%
	ie	16 distinct, 0 unique	3 distinct, 0 unique	54 distinct, 34 unique	11 distinct, 0 unique	75 distinct, 53 unique	20 distinct, 2 unique	4 distinct, 0 unique	4 distinct, 1 unique
1		Pasay	Male	48	Frozen Foods	149	4	e-wallet	0.00%
2		Navotas	undeclared	62	Household Items	61	3	cash	20.00%
3		Caloocan	Male	null	Canned Goods	192	18	cash	0.00%
4		Paranaque	Male	26	Others	77	6	e-wallet	0.00%
5		Pasig	Male	39	Meat and Seafood	114	18	cash	5.00%
6		Makati	Female	76	Personal Care	61	11	e-wallet	20.00%
7		Marikina	Female	87	Bakery	33	6	debit card	20.00%
8		Malabon	undeclared	49	Personal Care	120	17	e-wallet	0.00%
9		Mandaluyong	Female	32	Meat and Seafood	103	8	cash	20.00%
10		Mandaluyong	Female	18	Others	18	18	cash	20.00%
11		Caloocan	Male	null	Snacks	156	6	e-wallet	0.00%
12		San Juan	Female	null	Others	175	13	credit card	20.00%
13		Las Pinas	Male	55	Snacks	107	2	cash	0.00%
14		Quezon City	Male	52	Bakery	59	16	debit card	20.00%
15		Muntinlupa	Female	38	Bakery	27	9	cash	0.00%
16		Makati	Male	43	Meat and Seafood	101	14	cash	0.00%
17		Muntinlupa	Male	46	Fresh Produce	100	12	debit card	0.00%
18									

Column statistics

Count	100
Error	0
Empty	0
Distinct	4
Unique	1
NaN	0
Zero	43
Min	0
Max	0.2
Average	0.10099...

Query Settings

PROPERTIES

Name: Table1

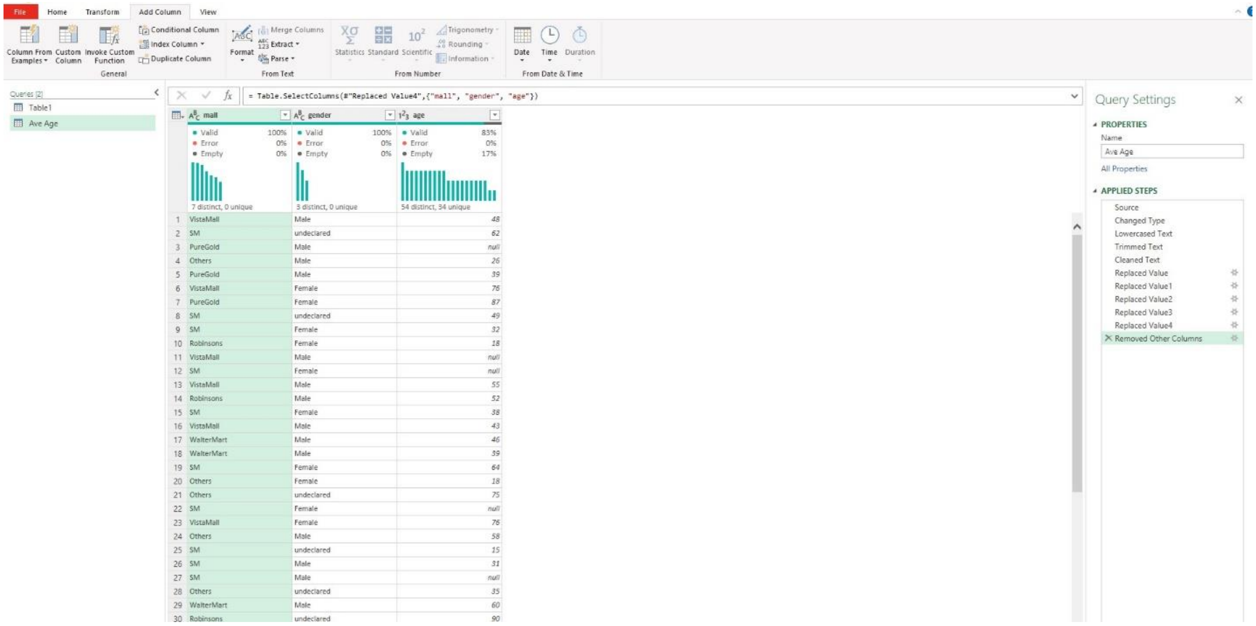
All Properties

APPLIED STEPS

- Source
- Changed Type
- Lowercased Text
- Trimmed Text
- Cleaned Text
- Replaced Value
- Replaced Value1
- Replaced Value2
- Replaced Value3
- Replaced Value4
- Replaced Value5
- Calculated Absolute Value
- Replaced Value6
- Added Conditional Column
- Removed Columns
- Replaced Value7
- Changed Type1

Age Column

- For Age Column, I need to get the average age per gender per mall.
  - I duplicate the query, remove other columns except Age, Mall, and Gender



```
= Table.SelectColumns(#"Replaced Value4",{"mall", "gender", "age"})
```

	mall	gender	age
	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 83% Error 0% Empty 17%
	7 distinct, 0 unique	3 distinct, 0 unique	54 distinct, 34 unique
1	VistaMall	Male	48
2	SM	undeclared	62
3	PureGold	Male	null
4	Others	Male	26
5	PureGold	Male	39
6	VistaMall	Female	76
7	PureGold	Female	87
8	SM	undeclared	49
9	SM	Female	32
10	Robinsons	Female	18
11	VistaMall	Male	null
12	SM	Female	null
13	VistaMall	Male	55
14	Robinsons	Male	52
15	SM	Female	38
16	VistaMall	Male	43
17	WalterMart	Male	46
18	WalterMart	Male	39
19	SM	Female	64
20	Others	Female	18
21	Others	undeclared	75
22	SM	Female	null
23	VistaMall	Female	76
24	Others	Male	58
25	SM	undeclared	15
26	SM	Male	31
27	SM	Male	null
28	Others	undeclared	35
29	WalterMart	Male	60
30	Robinsons	undeclared	90

Query Settings

PROPERTIES

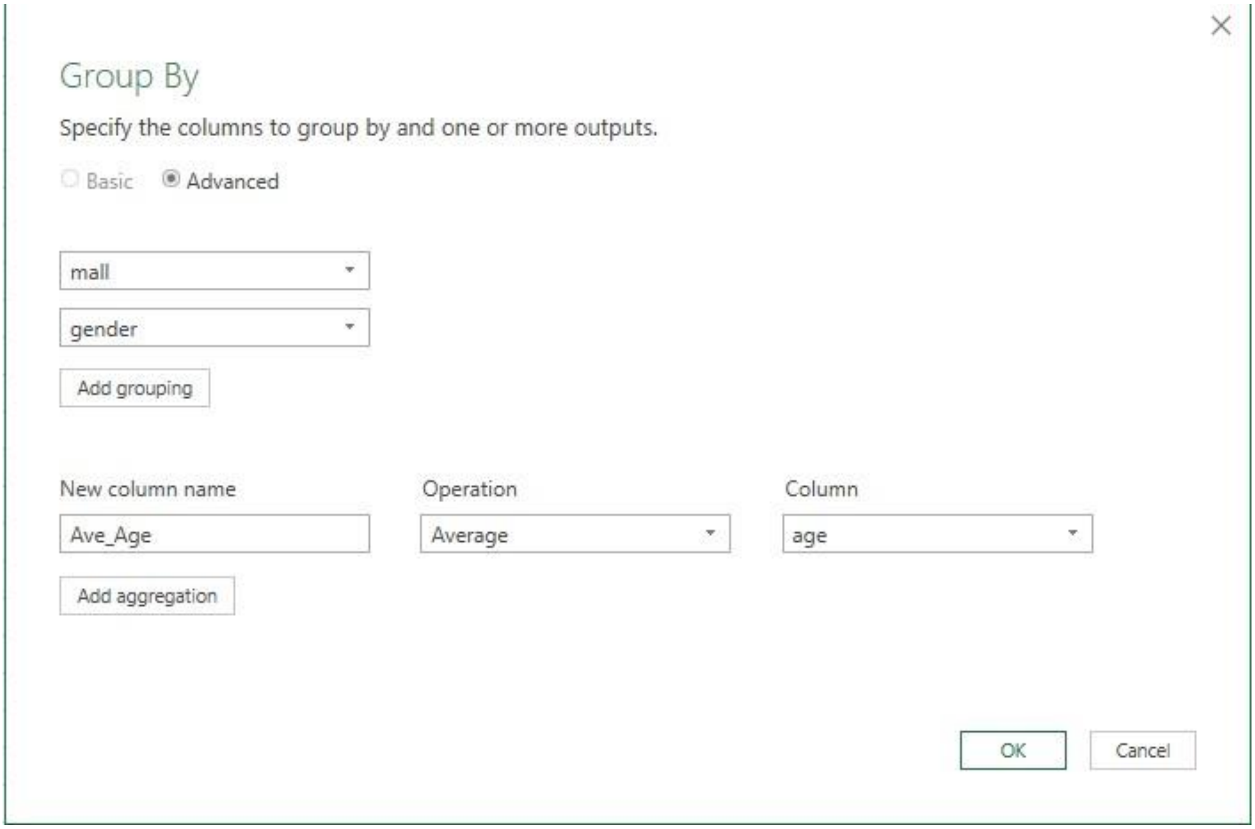
Name: Ave Age

All Properties

APPLIED STEPS

- Source
- Changed Type
- Lowercased Text
- Trimmed Text
- Cleaned Text
- Replaced Value
- Replaced Value1
- Replaced Value2
- Replaced Value3
- Replaced Value4
- Removed Other Columns

  - I use the Group BY feature to get the average age



Group By

Specify the columns to group by and one or more outputs.

○ Basic ● Advanced

mall

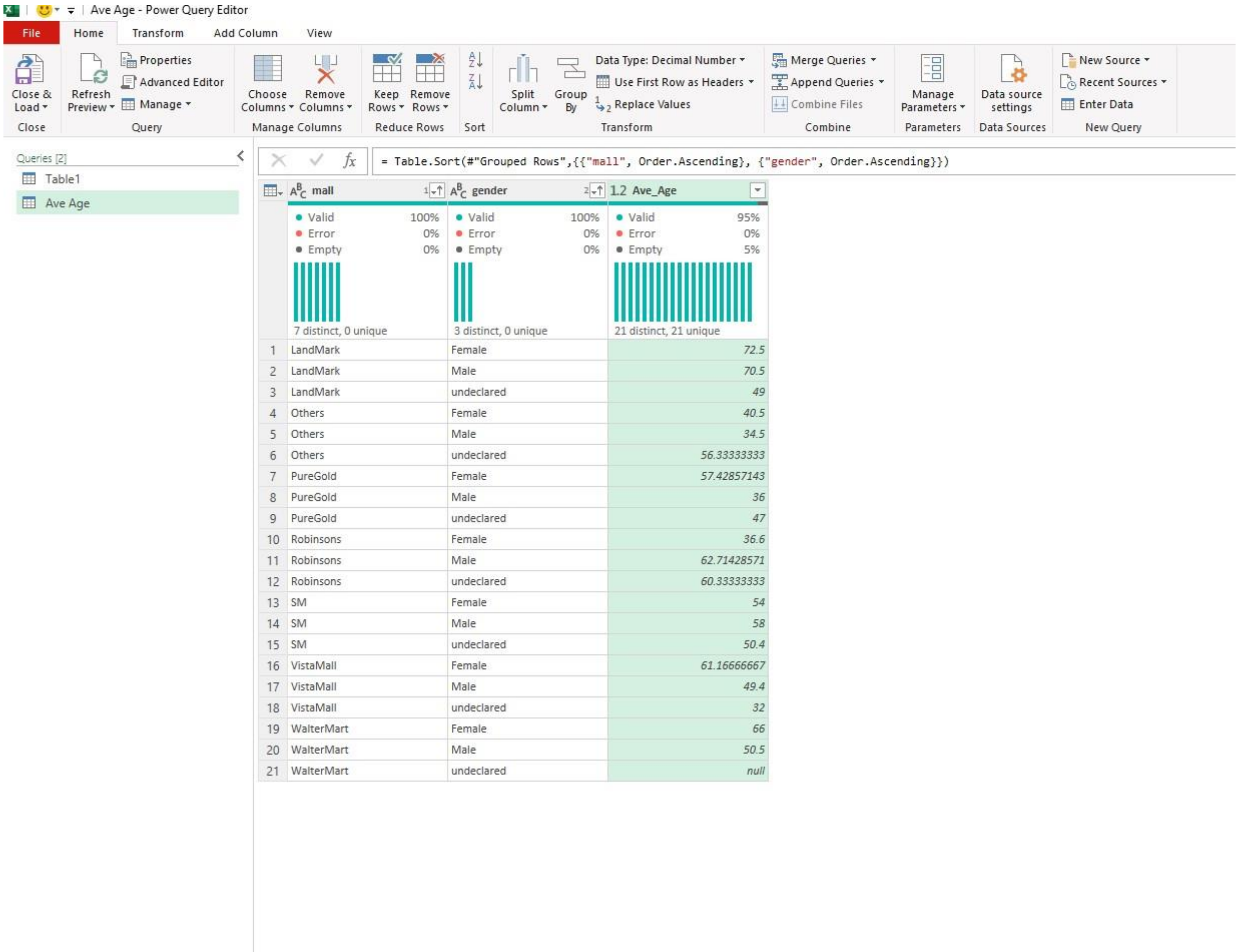
gender

New column name: Ave_Age

Operation: Average

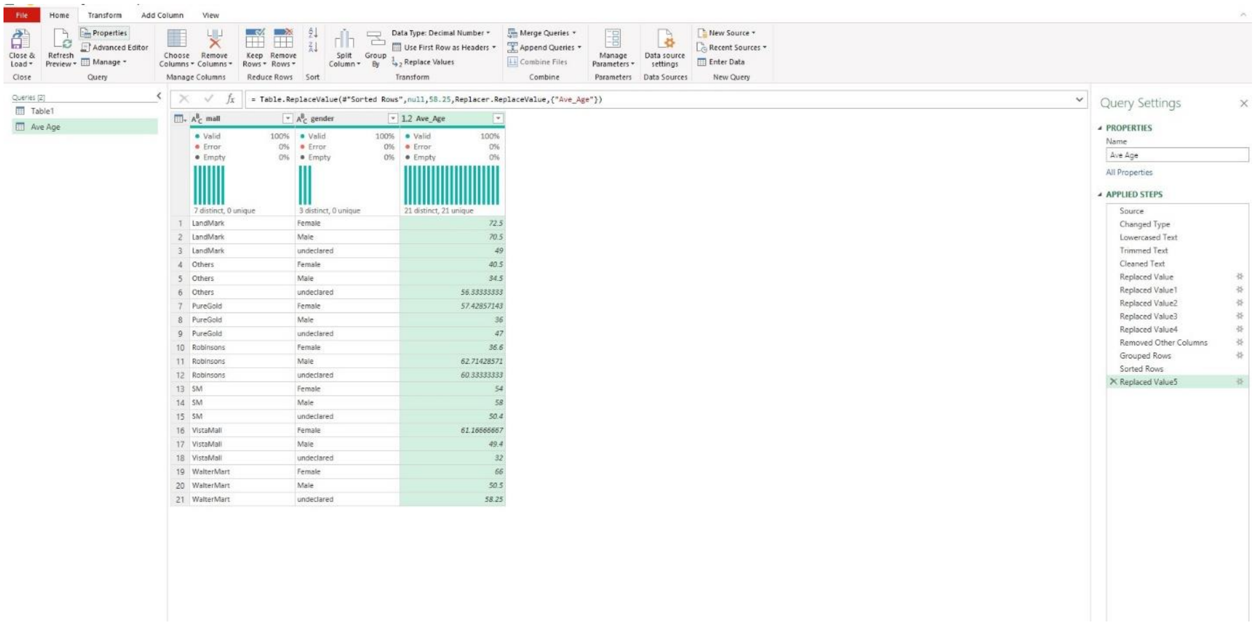
Column: age

- Replace the *null* to the average age between Female and Male of WalterMart



```
= Table.Sort(#"Grouped Rows",{{"mall", Order.Ascending}, {"gender", Order.Ascending}})
```

	mall	gender	Ave_Age
	Valid 100% Error 0% Empty 0% 7 distinct, 0 unique	Valid 100% Error 0% Empty 0% 3 distinct, 0 unique	Valid 95% Error 0% Empty 5% 21 distinct, 21 unique
1	LandMark	Female	72.5
2	LandMark	Male	70.5
3	LandMark	undeclared	49
4	Others	Female	40.5
5	Others	Male	34.5
6	Others	undeclared	56.33333333
7	PureGold	Female	57.42857143
8	PureGold	Male	36
9	PureGold	undeclared	47
10	Robinsons	Female	36.6
11	Robinsons	Male	62.71428571
12	Robinsons	undeclared	60.33333333
13	SM	Female	54
14	SM	Male	58
15	SM	undeclared	50.4
16	VistaMall	Female	61.16666667
17	VistaMall	Male	49.4
18	VistaMall	undeclared	32
19	WalterMart	Female	66
20	WalterMart	Male	50.5
21	WalterMart	undeclared	null



```
= Table.ReplaceValue(#"Sorted Rows",null,58.25,Replacer.ReplaceValue,{"Ave_Age"})
```

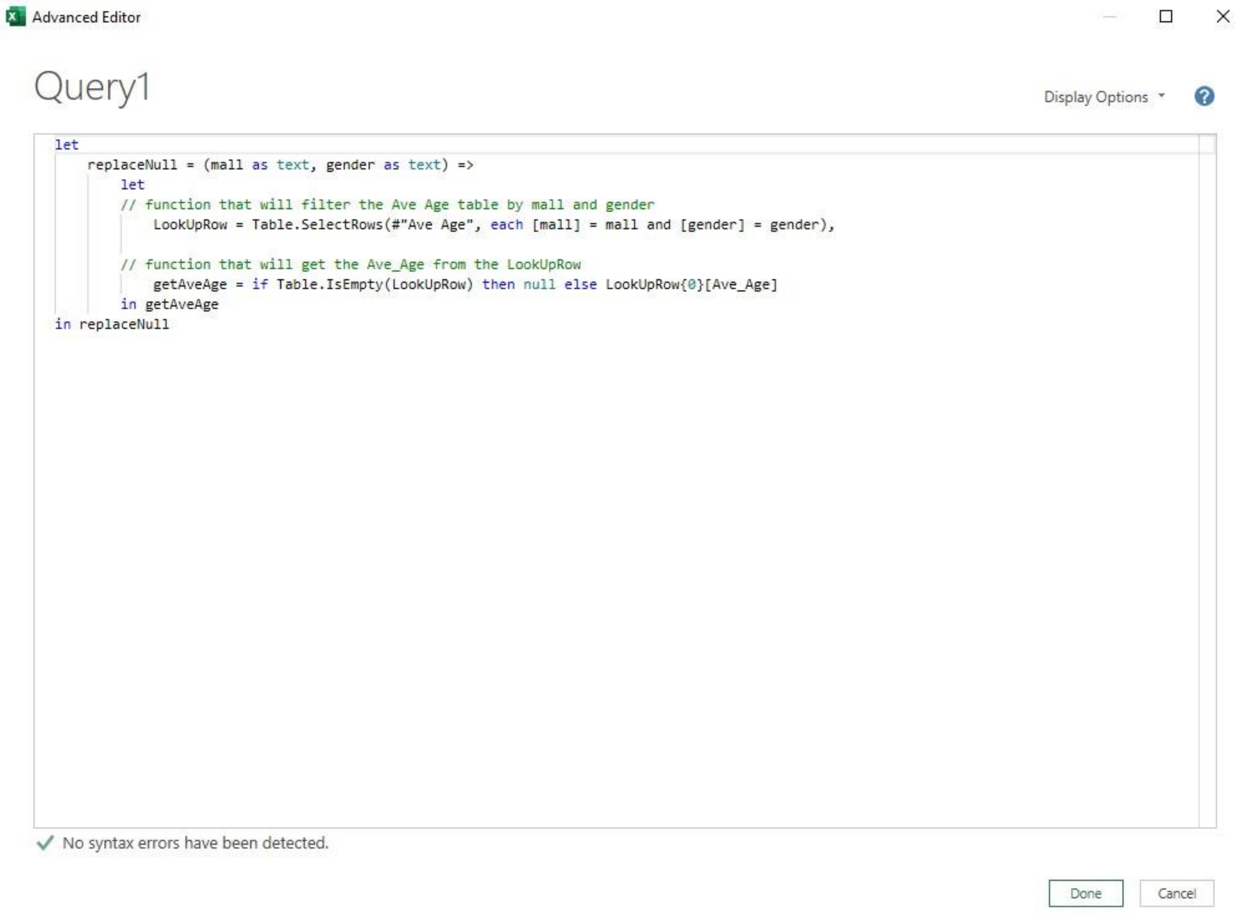
	mall	gender	Ave_Age
	Valid 100% Error 0% Empty 0% 7 distinct, 0 unique	Valid 100% Error 0% Empty 0% 3 distinct, 0 unique	Valid 100% Error 0% Empty 0% 21 distinct, 21 unique
1	LandMark	Female	72.5
2	LandMark	Male	70.5
3	LandMark	undeclared	49
4	Others	Female	40.5
5	Others	Male	34.5
6	Others	undeclared	56.33333333
7	PureGold	Female	57.42857143
8	PureGold	Male	36
9	PureGold	undeclared	47
10	Robinsons	Female	36.6
11	Robinsons	Male	62.71428571
12	Robinsons	undeclared	60.33333333
13	SM	Female	54
14	SM	Male	58
15	SM	undeclared	50.4
16	VistaMall	Female	61.16666667
17	VistaMall	Male	49.4
18	VistaMall	undeclared	32
19	WalterMart	Female	66
20	WalterMart	Male	50.5
21	WalterMart	undeclared	58.25

Name: Ave Age

Applied Steps:
Source
Changed Type
Lowercased Text
Trimmed Text
Cleaned Text
Replaced Value
Replaced Value1
Replaced Value2
Replaced Value3
Replaced Value4
Removed Other Columns
Grouped Rows
Sorted Rows
Replaced Value5

- This will be the reference table for replacing the null values.

- I’ll create a new query and create a custom function that will replace the null values based on the mall and the gender
    - On Advanced Editor, Put this code

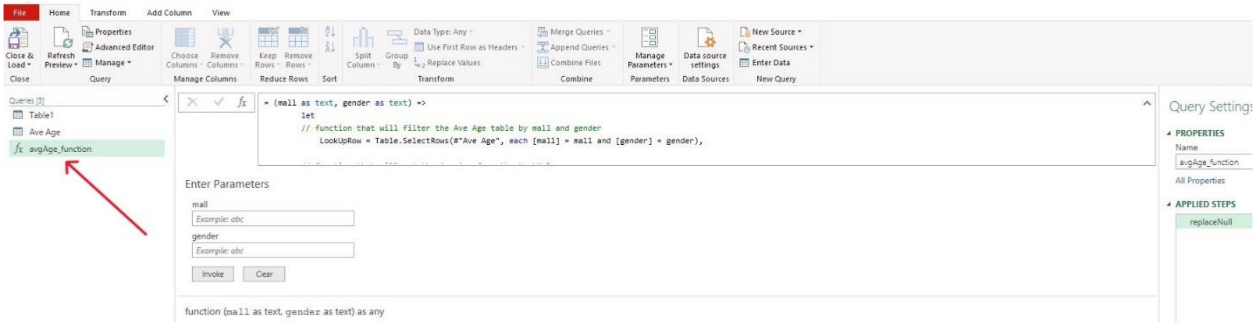


```
let
    replaceNull = (mall as text, gender as text) =>
        let
        // function that will filter the Ave Age table by mall and gender
            LookUpRow = Table.SelectRows(#"Ave Age", each [mall] = mall and [gender] = gender),

        // function that will get the Ave_Age from the LookUpRow
            getAveAge = if Table.IsEmpty(LookUpRow) then null else LookUpRow{0}[Ave_Age]
        in getAveAge
in replaceNull
```

No syntax errors have been detected.

- Rename the function avgAge_function



```
= (mall as text, gender as text) =>
        let
        // function that will filter the Ave Age table by mall and gender
            LookUpRow = Table.SelectRows(#"Ave Age", each [mall] = mall and [gender] = gender),
```

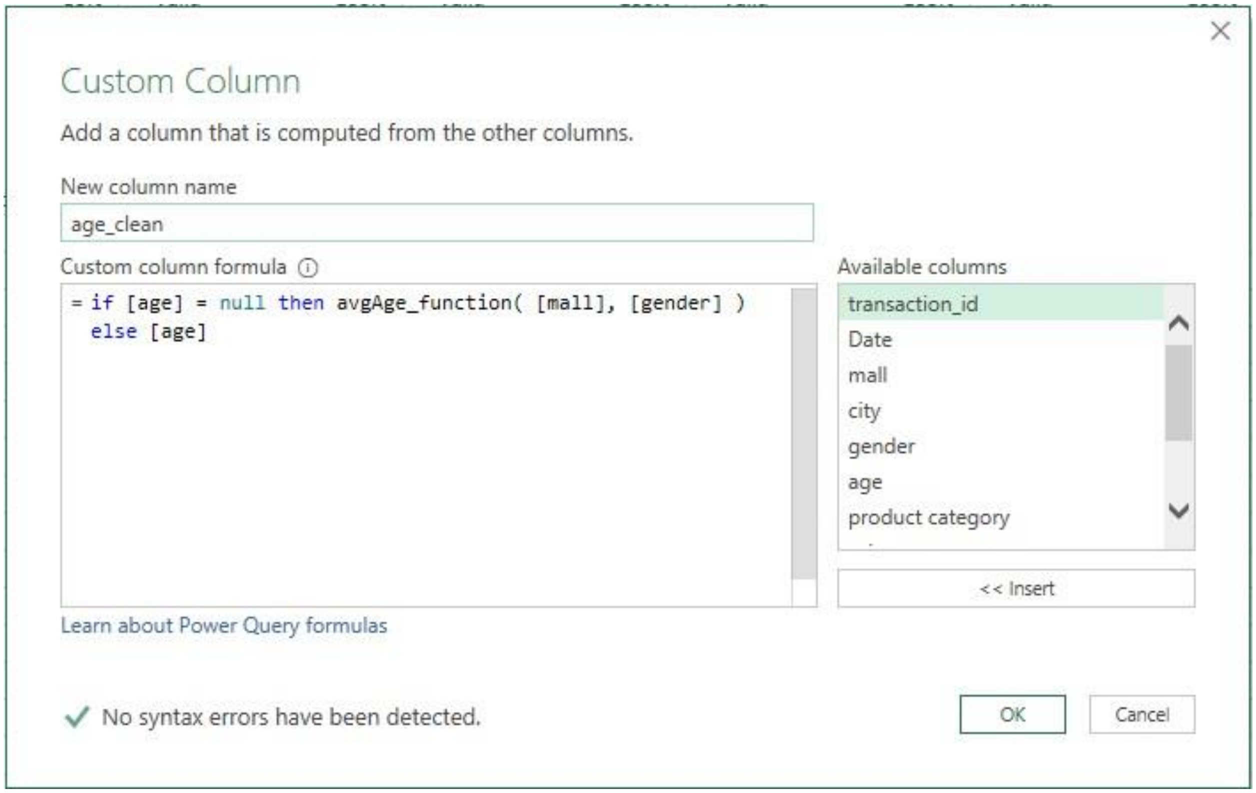
Enter Parameters

mall: Example: abc

gender: Example: abc

function (mall as text, gender as text) as any

- On the main table, add column by custom column and put this code:



Custom Column

Add a column that is computed from the other columns.

New column name

age_clean

Custom column formula

```
= if [age] = null then avgAge_function( [mall], [gender] )
  else [age]
```

No syntax errors have been detected.

- Remove the old age column and change format to whole Number, rename the column to age

Step 4: Save and Close file to be used on Reports/Creating Dashboard

- Close and Load the file. Save the file for Reports/ Creating Dashboard