

Transform the Data

Venkata Reddy Konasani

Contents

- Power Query
- Handling null values
- Rename
- M-Language
- Creating new columns
- Creating columns from examples
- Handling errors

Have a Look at the data

id	gender	age	ever_married	work_type	Residence_type
60182	Female	49	Yes	WT -Private	Urban
60491	Female	78	Yes	WT -Private	Urban
12175	Female	54	Yes	WT -Private	Urban
5317	Female	79	Yes	WT -Private	Urban
62602	Female	49	Yes	WT -Private	Urban
1845	Female	63	Yes	WT -Private	Urban
47472	Female	58	Yes	WT -Private	Urban
17004	Female	70	Yes	WT -Private	Urban
71673	Female	79	Yes	WT -Private	Urban
45805	Female	51	Yes	WT -Private	Urban
28291	Female	79	Yes	WT -Private	Urban
5563	Female	77	Yes	WT -Private	Urban
72918	Female	53	Yes	WT -Private	Urban
14164	Female	72	Yes	WT -Private	Urban
70943	Female	80	Yes	WT -Private	Urban
11762	Female	76	Yes	WT -Private	Urban
8045	Female	74	Yes	WT -Private	Urban
17308	Female	72	Yes	WT -Private	Urban

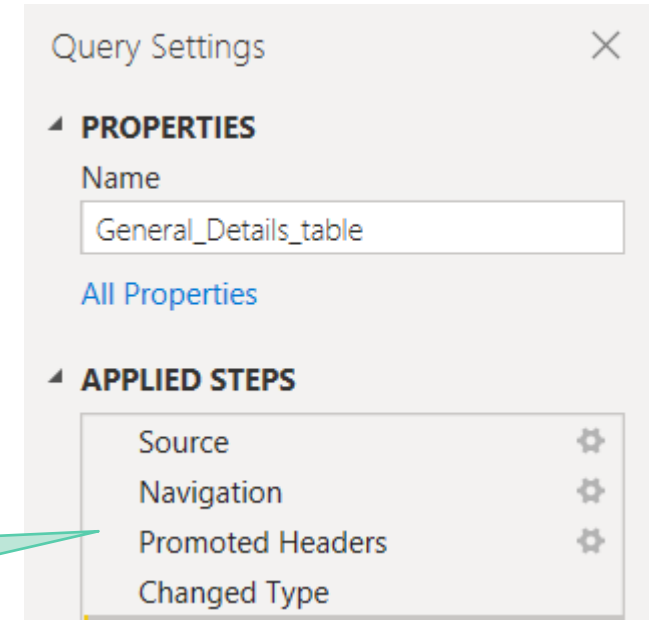
Patient_id	hypertension	heart_disease	avg_glucose_level	bmi	smoking_status	stroke
63884	0	0	162.96	39.4	never smoked	0
67855	0	0	95.04	42.4	never smoked	0
25774	0	0	85.37	33	never smoked	0
24447	0	0	82.67	22.5	never smoked	0
48588	0	0	109.82	23.7	never smoked	0
70336	0	0	60.84	24.5	never smoked	0
45801	0	0	97.49	26.9	never smoked	0
36275	0	0	206.72	26.7	never smoked	0
11577	0	0	214.45	31.2	never smoked	0
36811	0	0	94.09	30.9	never smoked	0
58261	0	0	141.24	28.5	never smoked	0
28526	0	0	203.04	33.6	never smoked	0
7282	0	0	81.84	25.1	never smoked	0
1686	0	0	71.89	27.6	never smoked	0
59368	0	0	243.5	26.1	never smoked	0
18051	0	0	91.61	25.2	never smoked	0
40840	0	0	138.16	19.4	never smoked	0
10449	0	0	75.23	29	never smoked	0
61837	0	0	58.95	24.6	never smoked	0
9487	0	0	99.92	25.8	never smoked	0
49713	0	0	116.23	26.1	never smoked	0
28102	0	0	66.3	27.2	never smoked	0
62608	0	0	136.8	37.3	never smoked	0
40670	0	0	96.57	34.1	never smoked	0
4630	0	0	66.42	23.6	never smoked	0

Stroke Case study

**Step-3: Transform the data;
Prepare it for analysis**

Power Query Editor

- Start with General Details table data transformation
- Click on Transform data, it will open the power query editor
- Check the query settings, Some of the transformations are already applied on the data



Query Settings

Query Name: General_Details_table

APPLIED STEPS

- Source
- Navigation
- Promoted Headers
- Changed Type
- Filtered Rows
- Replaced Value

Column1	Column2	Column3	Column4	Column5	Column6
id	gender	age	ever_married	work_type	Residence_type
9046	Male	67	Yes	WT -Private	Urban
51676	Female	61	Yes	WT -Self-employed	Rural
31112	Male	80	Yes	WT -Private	Rural
60182	Female	49	Yes	WT -Private	Urban
1665	Female	79	Yes	WT -Self-employed	Rural
56669	Male	81	Yes	WT -Private	Urban
53882	Male	74	Yes	WT -Private	Rural
10434	Female	69	No	WT -Private	Urban
27419	Female	59	Yes	WT -Private	Rural
60491	Female	78	Yes	WT -Private	Urban

You can check the effect of a particular transformation step

Query Settings

Query Name: General_Details_table

APPLIED STEPS

- Source
- Navigation
- Promoted Headers
- Changed Type
- Filtered Rows
- Replaced Value

id	gender	age	ever_married	work_type	Residence_type
9046	Male	67	Yes	WT -Private	Urban
51676	Female	61	Yes	WT -Self-employed	Rural
31112	Male	80	Yes	WT -Private	Rural
60182	Female	49	Yes	WT -Private	Urban
1665	Female	79	Yes	WT -Self-employed	Rural
56669	Male	81	Yes	WT -Private	Urban
53882	Male	74	Yes	WT -Private	Rural
10434	Female	69	No	WT -Private	Urban
27419	Female	59	Yes	WT -Private	Rural
60491	Female	78	Yes	WT -Private	Urban
12109	Female	81	Yes	WT -Private	Rural

Gender variable

- 1% null values in gender variable
- Can be replaced with “NA”

fx = Table.TransformColumnTypes(#"Promoted Headers",{{"id", type text}, {"gender", type text}, {"age", type number}, {"ever_married", type

	id	gender	age	ever_married	work_type	Residence_type
1			67	Yes	WT -Private	Urban
2			61	Yes	WT -Self-employed	Rural
3			80	Yes	WT -Private	Rural
4			49	Yes	WT -Private	Urban
5			79	Yes	WT -Self-employed	Rural
6			81	Yes	WT -Private	Urban
7	53882	Male	74	Yes	WT -Private	Rural
8	10434	Female	69	No	WT -Private	Urban
9	27419	Female	59	Yes	WT -Private	Rural
10	60491	Female	78	Yes	WT -Private	Urban
11	12109	Female	81	Yes	WT -Private	Rural
12	12095	Female	61	Yes	WT -Govt_job	Rural
13	12175	Female	54	Yes	WT -Private	Urban
14	8213	Male	78	Yes	WT -Private	Urban
15	5317	Female	79	Yes	WT -Private	Urban
16	58202	Female	50	Yes	WT -Self-employed	Rural
17	56112	Male	64	Yes	WT -Private	Urban
18	34120	Male	75	Yes	WT -Private	Urban
19	27458	null	60	No	WT -Private	Urban

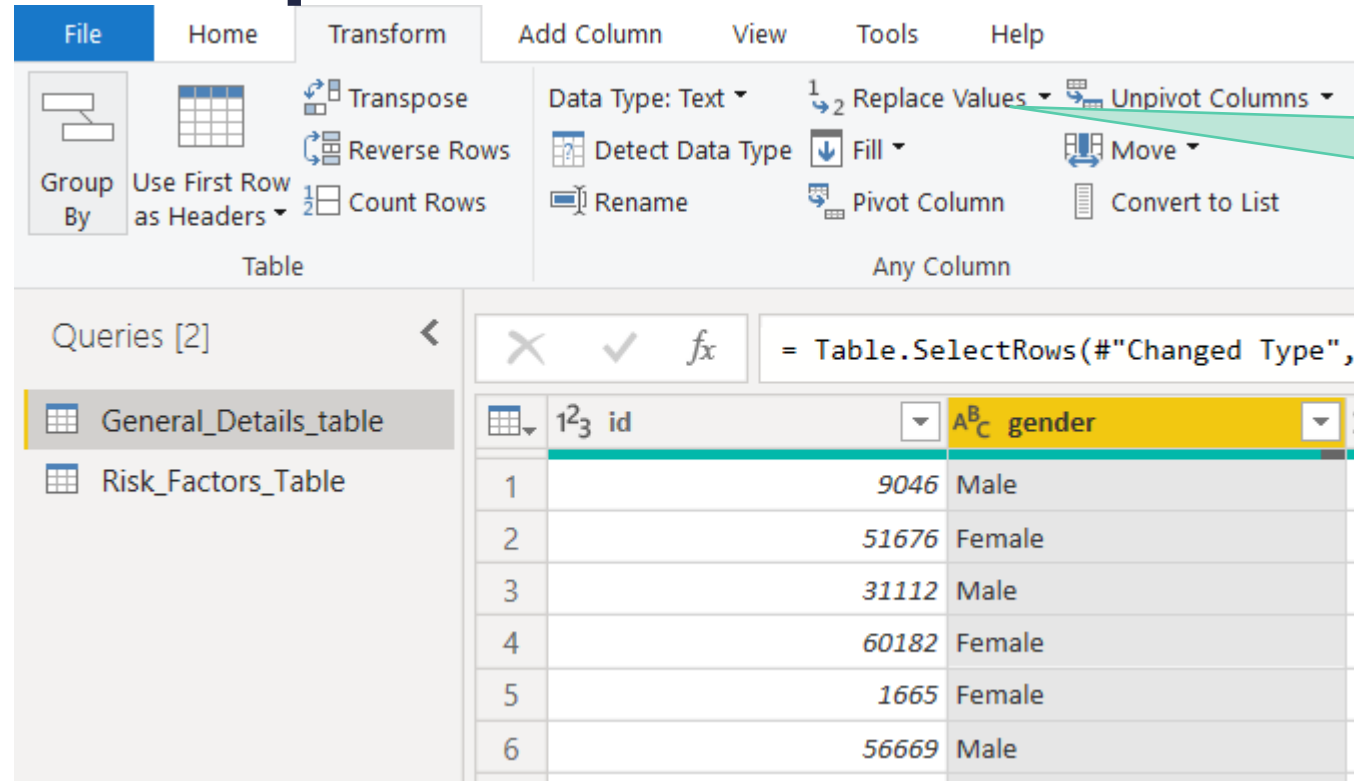
gender

989 (99%) 0 (0%) 11 (1%)

Valid Error Empty

Remove Empty

Replace null values

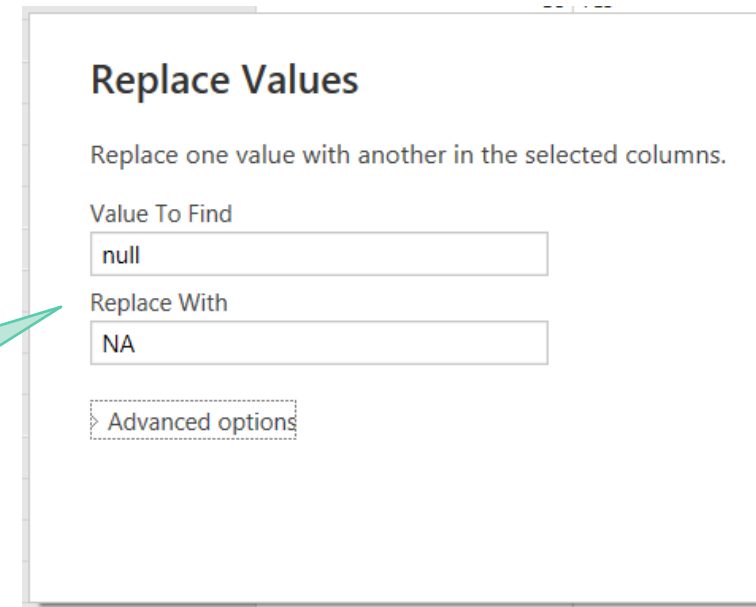


The screenshot shows the Power BI Desktop interface. The 'Transform' ribbon is active, displaying various data transformation options. Below the ribbon, the 'Queries' pane on the left lists 'General_Details_table' and 'Risk_Factors_Table'. The main area shows a data table with columns 'id' and 'gender'.

	id	gender
1	9046	Male
2	51676	Female
3	31112	Male
4	60182	Female
5	1665	Female
6	56669	Male

Menu options >> Transform >> Replace Values

Value to find >> null
Value to replace it with >> NA



The 'Replace Values' dialog box is shown, allowing users to replace one value with another in the selected columns. The 'Value To Find' field is set to 'null' and the 'Replace With' field is set to 'NA'. There is an 'Advanced options' link at the bottom.

Replace Values

Replace one value with another in the selected columns.

Value To Find
null

Replace With
NA

[Advanced options](#)

ABC gender		1.2 age
gender		
989 (99%)	0 (0%)	11 (1%)
Valid	Error	Empty
Remove Empty		
53882	Male	
10434	Female	
27419	Female	
60491	Female	
12109	Female	
12095	Female	
12175	Female	
8213	Male	
5317	Female	
58202	Female	
56112	Male	
34120	Male	
27458		null
25226		null
70630		null
13861		null
68794		null
64778	Male	



ABC gender		1.2 age
gender		
1000 (100%)	0 (0%)	0 (0%)
Valid	Error	Empty
53882	Male	
10434	Female	
27419	Female	
60491	Female	
12109	Female	
12095	Female	
12175	Female	
8213	Male	
5317	Female	
58202	Female	
56112	Male	
34120	Male	
27458	NA	
25226	NA	
70630	NA	
13861	NA	
68794	NA	
64778	Male	
4219	Male	
70822	Male	



Age - Replace null values with mean

- Age is a numeric column we can replace null values with mean.
- But, how to mention the value here?
- We need to use a trick

her in the selected columns.

nan

Replace With

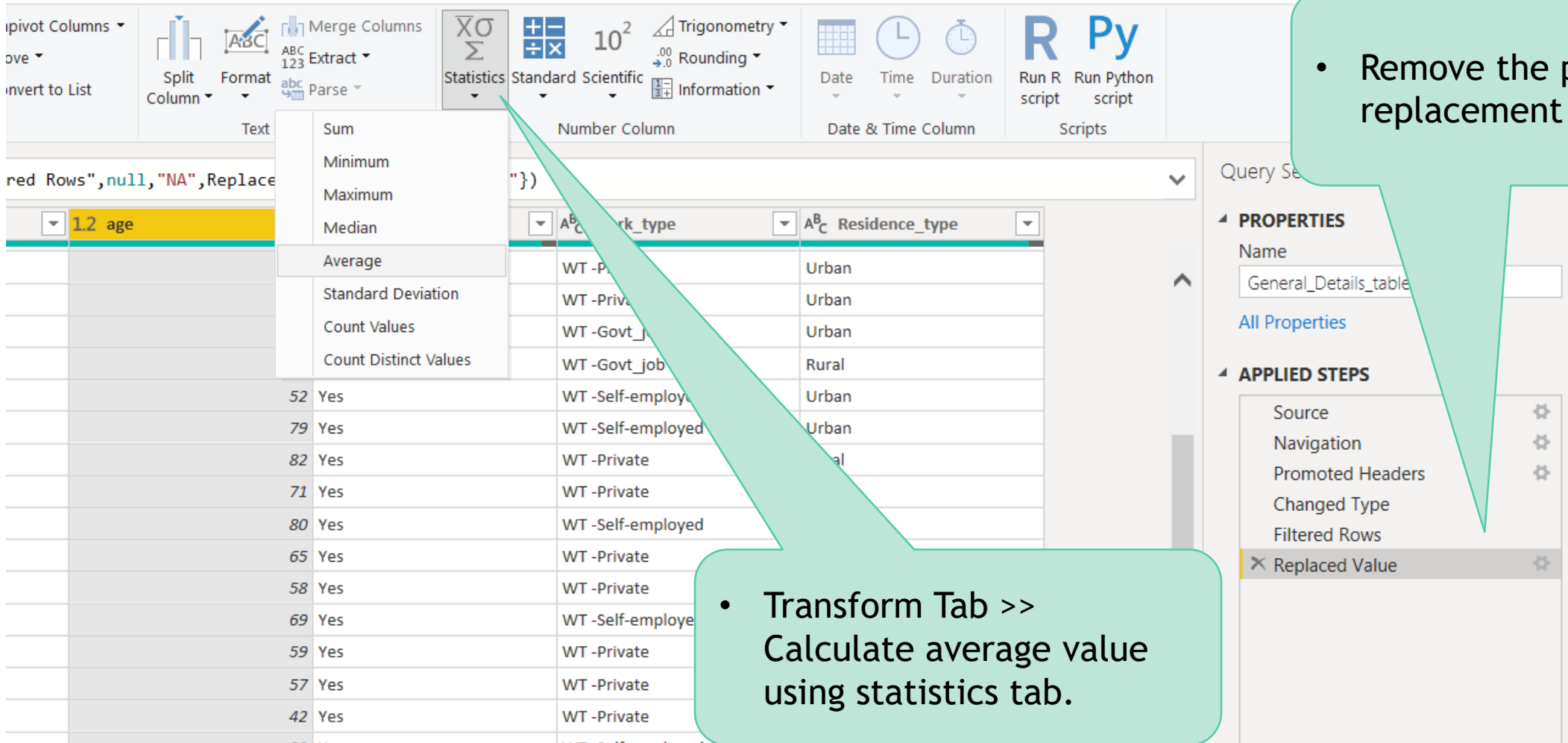
Trick - Replace null values with mean

First replace by “0” get the formula

```
= Table.ReplaceValue("#Replaced Value",null,0,Replacer.ReplaceValue,{"age"}))
```

- This “0” need to be replaced by average value
- Now let calculate the average value of age column and get the formula for it

Trick - Replace null values with mean

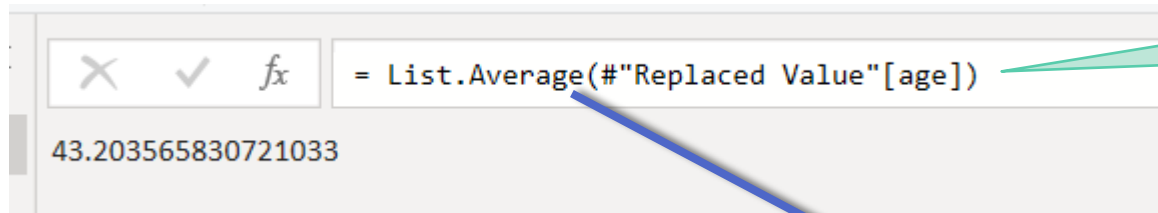


The screenshot shows the DV Analytics interface. The 'Statistics' tab is selected in the top toolbar, and its dropdown menu is open, highlighting the 'Average' option. A callout bubble points to this menu with the text: 'Transform Tab >> Calculate average value using statistics tab.'

In the background, a data table is visible with columns 'age', 'work_type', and 'Residence_type'. The 'age' column has a yellow header row with the value '1.2 age'. The 'work_type' column contains values like 'WT-Private', 'WT-Self-employed', and 'WT-Govt_job'. The 'Residence_type' column contains 'Urban' and 'Rural'.

On the right side, the 'APPLIED STEPS' panel is visible, showing a list of steps: Source, Navigation, Promoted Headers, Changed Type, Filtered Rows, and Replaced Value. A callout bubble points to the 'Replaced Value' step with the text: 'Remove the previous "0" replacement step'.

Trick - Replace null values with mean

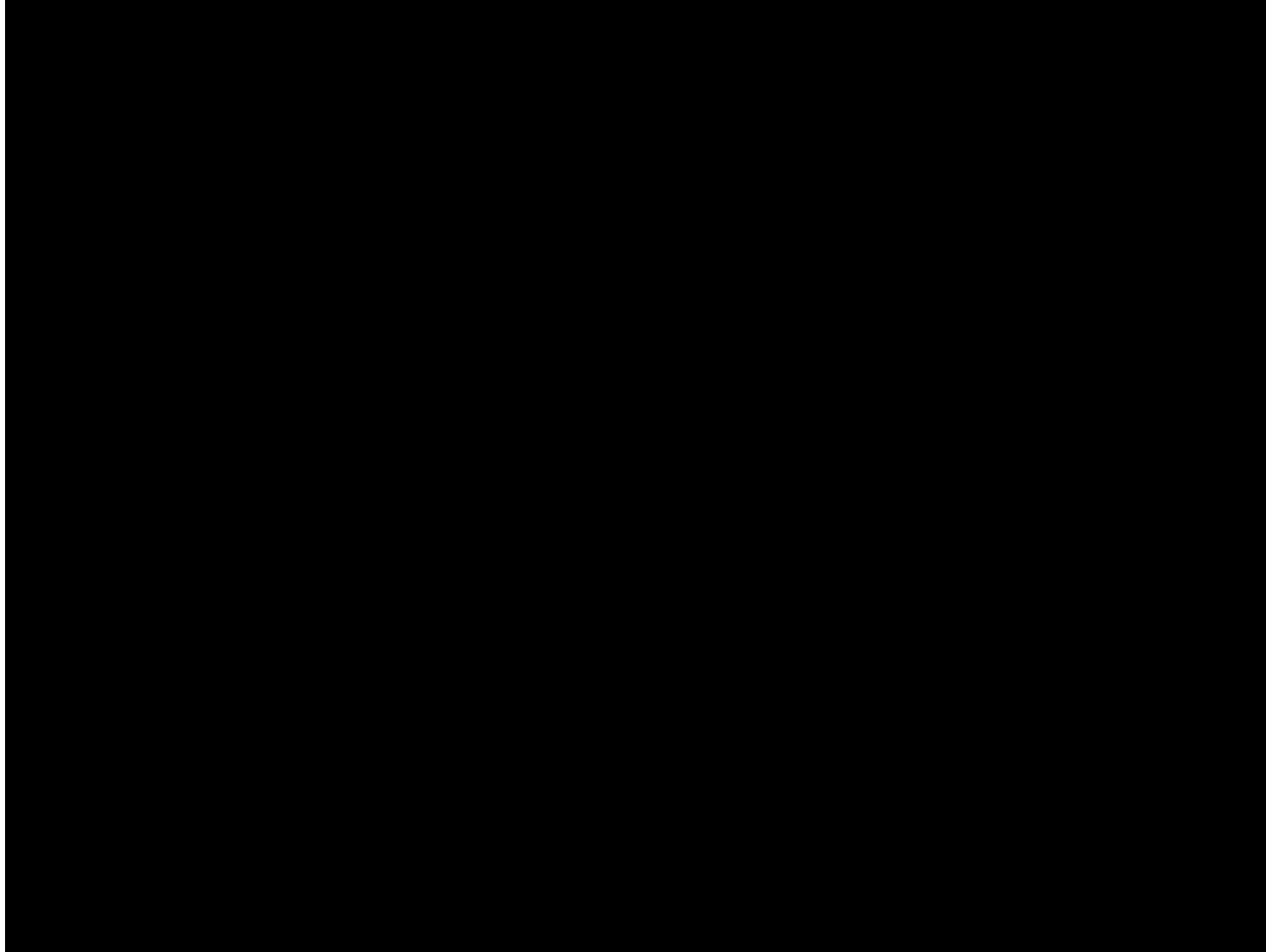


- Take this formula and replace it in the place of "0"

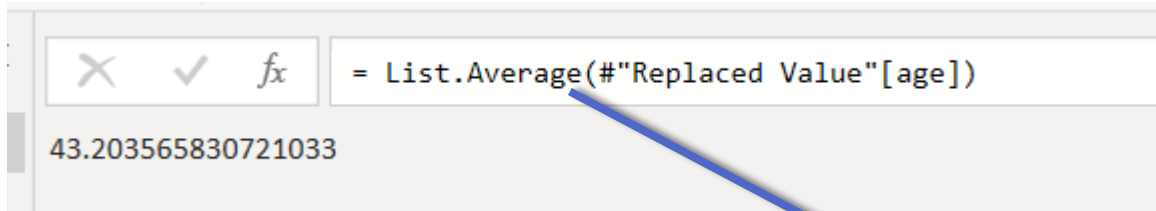
`= Table.ReplaceValue(#"Replaced Value",null,0,Replacer.ReplaceValue,{"age"})`

`= Table.ReplaceValue(#"Replaced Value",null,List.Average(#"Replaced Value"[age]),Replacer.ReplaceValue,{"age"})`

Video- Replace null values with mean



Power Query M Formula Language



- This code is known as the Power Query M Formula Language



`= Table.ReplaceValue(#"Replaced Value",null,0,Replacer.ReplaceValue,{"age"})`

The diagram shows a blue arrow pointing from the `List.Average` formula in the top screenshot to the `0` parameter in the middle formula. Another blue arrow points from the `0` parameter in the middle formula to the `List.Average` formula in the bottom screenshot.

`= Table.ReplaceValue(#"Replaced Value",null,List.Average(#"Replaced Value"[age]),Replacer.ReplaceValue,{"age"})`

M- language

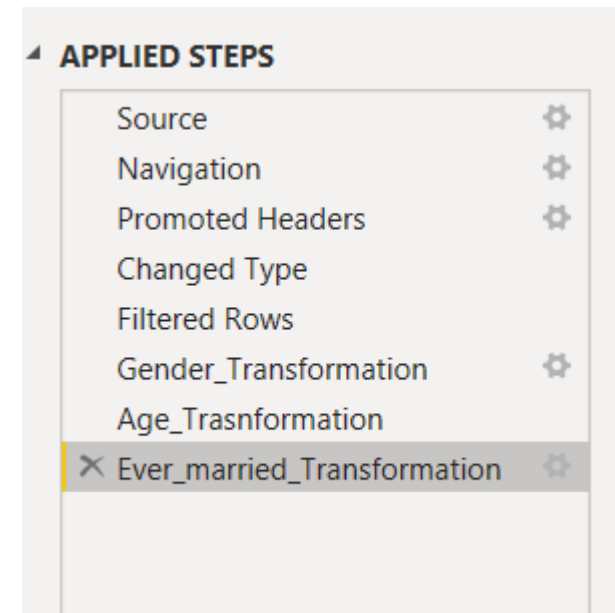
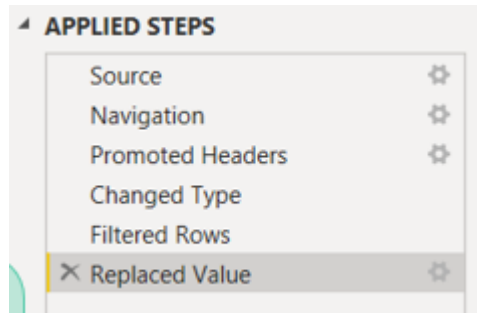
- The M stands for data Mash-up, as power query is all about connecting to various different data sources and “Mashing” them up.
- M code is the language behind the scenes of power query.
- It's a functional, case sensitive language
- When you create a data transformation in the power query editor UI, Power BI is writing the corresponding M code for the query.
- M code comes with a very large library of predefined functions available and you can also create your own.

Transform – “ever married”

- Replace missing values with “NA”

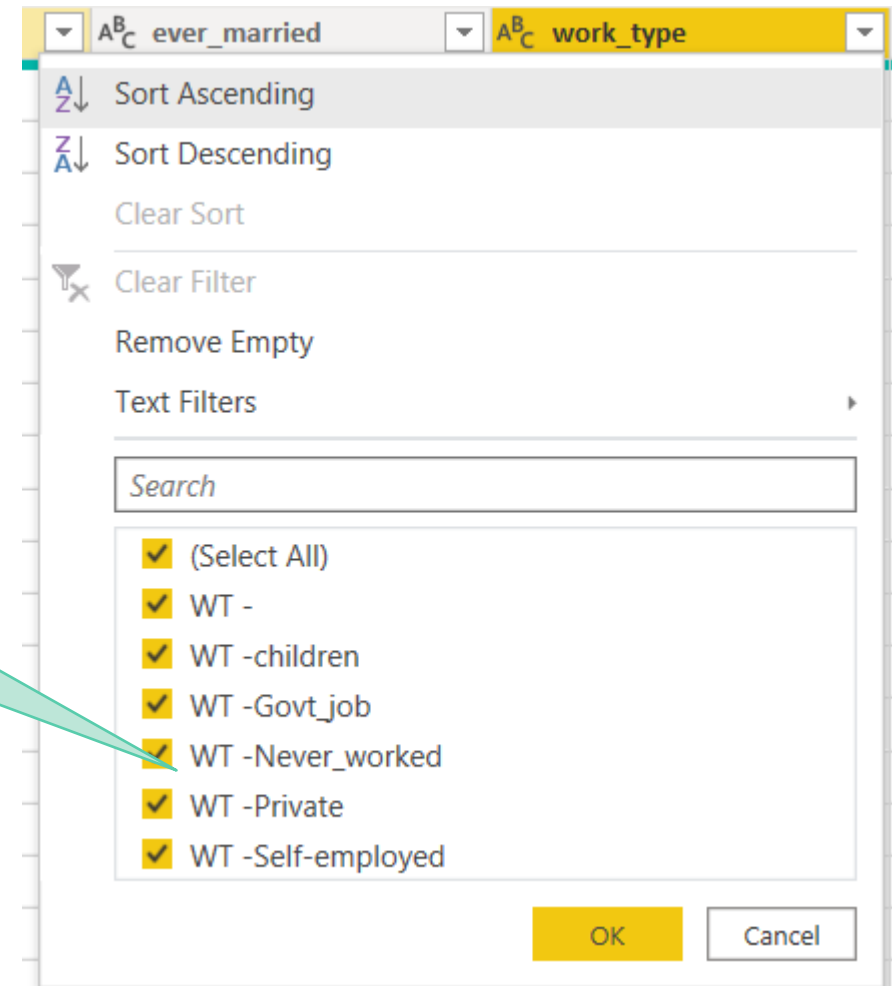
Before you proceed

- Rename each transformation step

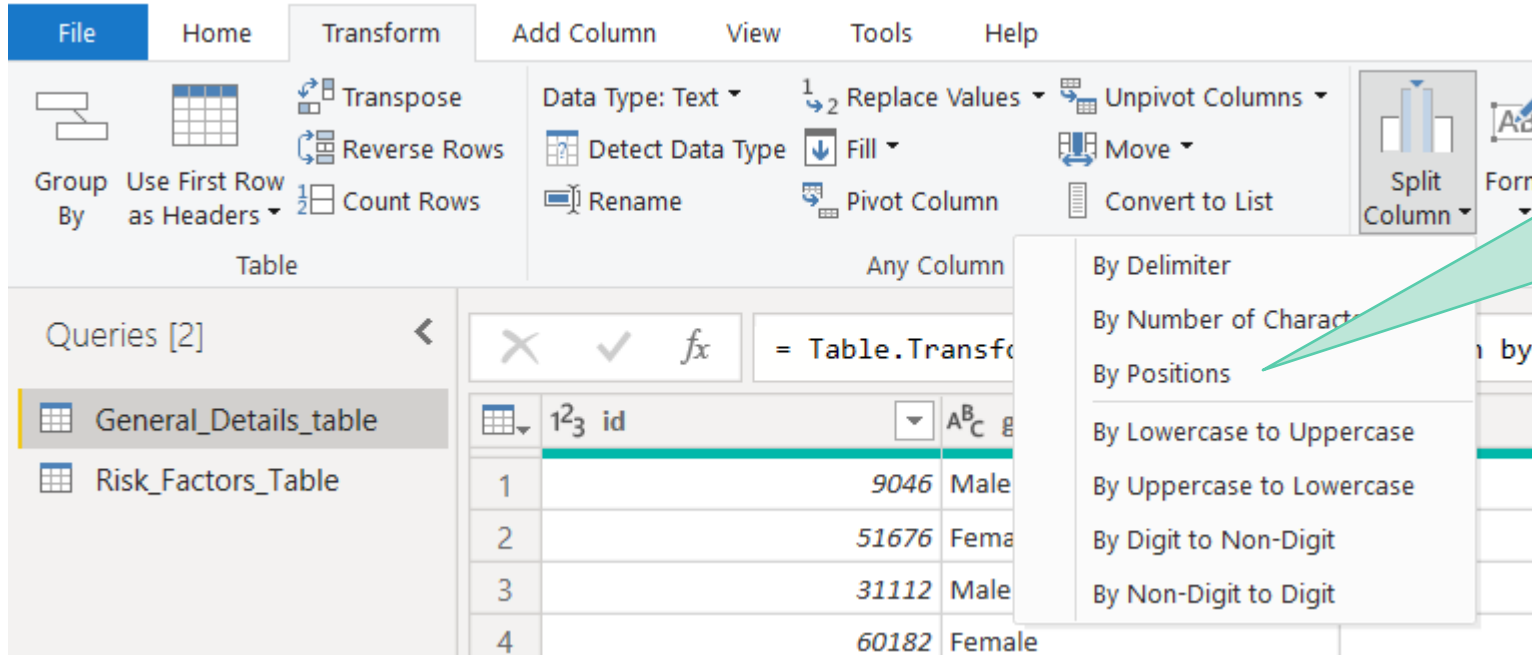


Transform – work_type

- Remove WT from work_type column values
- There are a couple of ways to do this
- Create a duplicate column to apply different methods



Transform – work_type - Using split column

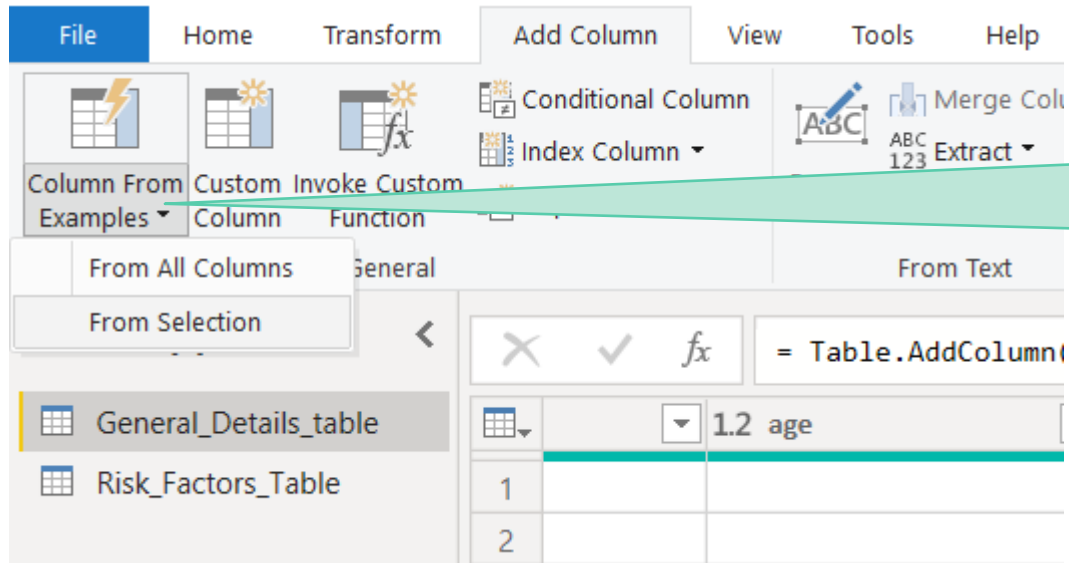


The screenshot shows the Power Query Editor interface. The 'Transform' tab is selected, and the 'Split Column' button is highlighted. The 'Split Column' dropdown menu is open, showing options: 'By Delimiter', 'By Number of Characters', 'By Positions', 'By Lowercase to Uppercase', 'By Uppercase to Lowercase', 'By Digit to Non-Digit', and 'By Non-Digit to Digit'. The 'By Positions' option is selected. The background shows a table with columns 'id', 'work_type', and 'gender'.

	id	work_type	gender
1	9046	Male	
2	51676	Female	
3	31112	Male	
4	60182	Female	

- Transform >> Split Column >> By Position >> 0 - 4

Transform – work_type - Using Column from Examples



- Select the column >> Add Column >> Column From Examples >> From Selection

Transform – work_type - Using Column from Examples

?

OK Cancel

work_type - Copy

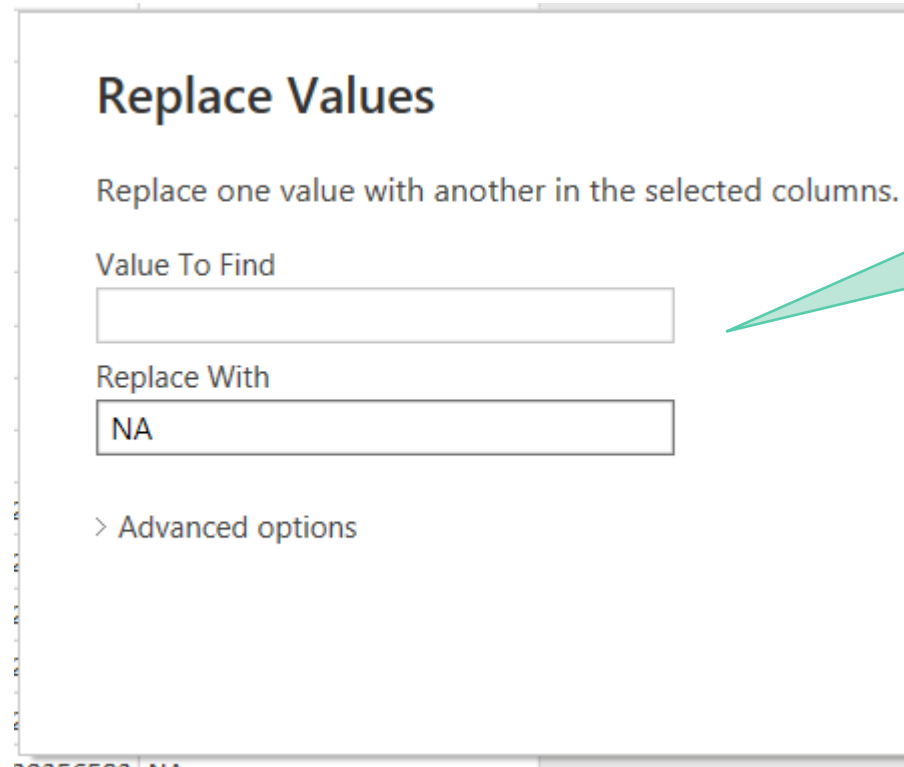
☒

	Text After Delimiter
WT -Private	Private
WT -Self-employed	Self-employed
WT -Private	Private
WT -Private	Private
WT -Self-employed	Self-employed
WT -Private	Private
WT -Private	Private
WT -Private	Private
WT -Private	Private
WT -Private	Private
WT -Private	Private
WT -Govt_job	Govt_job
WT -Private	Private
WT -Private	Private
WT -Private	Private
WT -Self-employed	Self-employed
WT -Private	Private
WT -Private	Private
WT -Private	Private

- Enter a sample value. Rest of the values will be filled automatically.

Transform – work_type- Further Cleaning

- Replace <<space>> values with “NA”



Replace Values

Replace one value with another in the selected columns.

Value To Find

Replace With

> Advanced options

- We have to replace the blank spaces

Transform – Residence_type

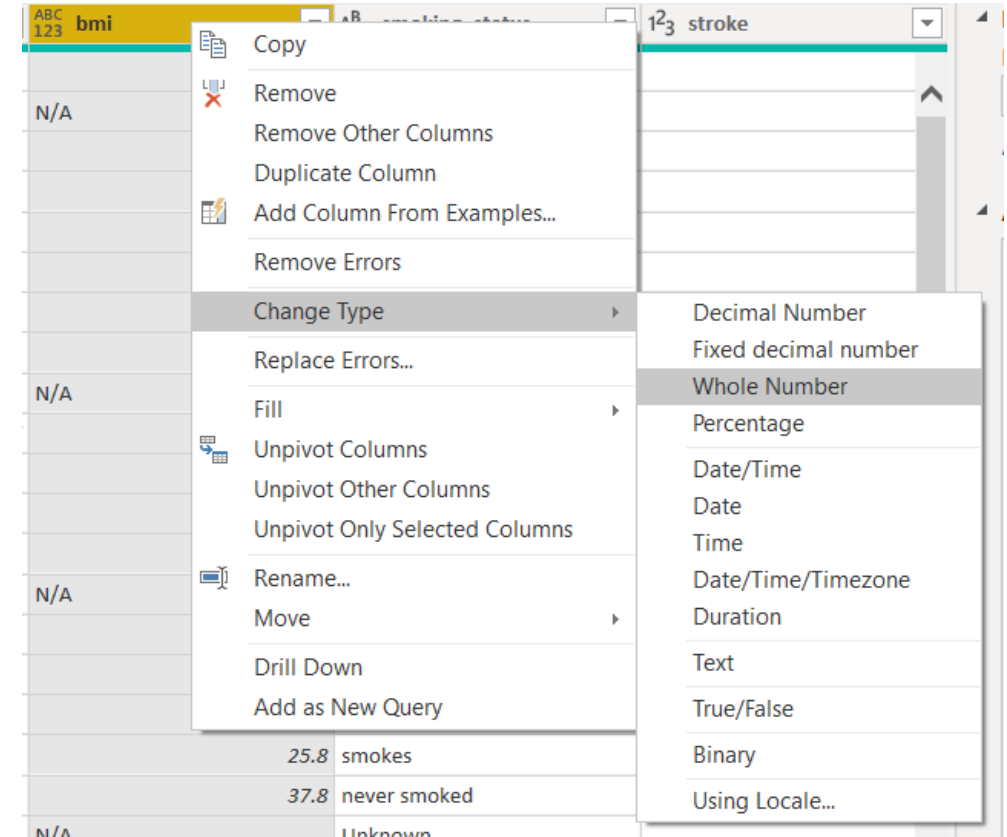
- Replace null values with “NA”

Remaining cleaning

- Finally delete the un-necessary columns related to work type.

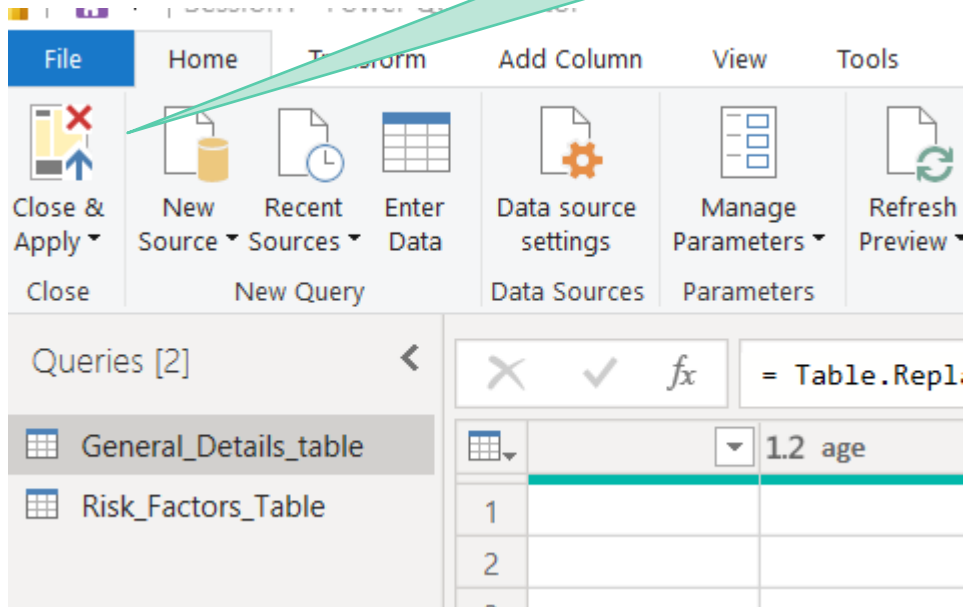
Transforming the Risk factors table

1. Change the bmi column type to numeric
2. Transform >> Replace Values >> Replace Errors >> null
3. Calculate the average >> Take the formula >> Replace with zero and then include the formula



Important step before you proceed...

- You have to click on Close and Apply
- Without this, if you close the query tab, then the steps will not be applied.



Next Step –
**Step4: Model the data and
create derived columns**
