

TASK 1

Steps i have taken to clean the data :

1. Initially I have loaded the data into powerbi and removed empty values by clicking dropdown button

Query Editor: Untitled - Power Query Editor

Query: = Table.PromoteHeaders(Source, [PromoteAllScalars=true])

Columns: Marital_Status, Income, Kidhome, Teenhome, Dt_Customer

Properties: Name: marketing_campaign

Applied Steps: Source, Promoted Headers

Index	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer
43	together	21994	0	1	24-12-2012
44	ingle		0	0	05-11-2013
45	married	7500	1	0	02-10-2013
46	ingle	79941	0	0	28-06-2014
47	married	7500	0	0	09-11-2012
48	together	41728	1	0	24-05-2013
49	ingle		2	1	01-01-2014
50	married	72550	1	1	08-11-2012
51	widow	65486	0	1	12-05-2014
52	married	79143	0	0	11-08-2012
53	ingle	35790	1	0	02-01-2013
54	ivorced	82582	0	0	07-06-2014
55	married	66373	1	1	12-06-2013
56	together	82384	0	0	19-11-2012
57	ivorced	70287	0	0	02-04-2013
58	ingle	27938	1	0	28-04-2014
59	ingle		1	0	17-06-2013
60	married	55954	0	1	03-03-2014

Query Editor: Untitled - Power Query Editor

Query: = Table.PromoteHeaders(Source, [PromoteAllScalars=true])

Columns: Marital_Status, Income, Kidhome, Teenhome, Dt_Customer

Properties: Name: marketing_campaign

Applied Steps: Source, Promoted Headers

Remove Empty dialog box:

- (Select All)
- (blank)
- 101970
- 102160
- 10245
- 102692
- 10404
- 105471
- 10979
- 12571

List may be incomplete. Load more

OK Cancel

Table with 8 columns: Marital_Status, Income, Kidhome, Teenhome, Dt_Customer, and others. The formula bar shows: `= Table.SelectRows("#Promoted Headers", each ([Income] <> ""))`

	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer
1	single	58138	0	0	04-09-2012
2	single	46344	1	1	08-03-2014
3	together	71613	0	0	21-08-2013
4	together	26646	1	0	10-02-2014
5	married	58293	1	0	19-01-2014
6	together	62513	0	1	09-09-2013
7	divorced	55635	0	1	13-11-2012
8	married	33454	1	0	08-05-2013
9	together	30351	1	0	06-06-2013
10	together	5648	1	1	13-03-2014
11	married	7500	0	0	13-11-2012
12	divorced	63033	0	0	15-11-2013
13	divorced	59354	1	1	15-11-2013
14	married	17323	0	0	10-10-2012
15	single	82800	0	0	24-11-2012
16	married	41850	1	1	24-12-2012
17	together	37760	0	0	31-08-2012
18	married	76995	0	1	28-03-2013

2. Next i removed duplicates from the table by clicking at the table icon

Table with 8 columns: Marital_Status, Income, Kidhome, Teenhome, Dt_Customer, and others. The formula bar shows: `= Table.SelectRows("#Promoted Headers", each ([Income] <> ""))`

Context menu options visible:

- Copy Entire Table
- Use First Row as Headers
- Add Custom Column...
- Add Column From Examples...
- Invoke Custom Function...
- Add Conditional Column...
- Add Index Column...
- Choose Columns...
- Keep Top Rows...
- Keep Bottom Rows...
- Keep Range of Rows...
- Keep Duplicates
- Keep Errors
- Remove Top Rows...
- Remove Bottom Rows...
- Remove Alternate Rows...
- Remove Duplicates
- Remove Errors
- Merge Queries...
- Append Queries...

3. Next I changed date format to dd-mm-yyyy

Table with 5 columns: Teenhome, Dt_Customer, Recency, and others. The formula bar shows: `= Table.Distinct("#Filtered Rows")`

Context menu options visible:

- Decimal Number
- Fixed decimal number
- Whole Number
- Percentage
- Date/Time
- Date
- Time
- Date/Time/Timezone
- Duration
- Text
- True/False
- Binary

Dt_Customer	APC Recency	APC MntWines	APC MntFruit
Valid 100%	Valid 100%	Valid 100%	Valid 100%
Error 0%	Error 0%	Error 0%	Error 0%
Empty 0%	Empty 0%	Empty 0%	Empty 0%
04-09-2012	58	635	88
08-03-2014	38	11	1
21-08-2013	26	426	49
10-02-2014	26	11	4
19-01-2014	94	173	43
09-09-2013	16	520	42
13-11-2012	34	235	65
08-05-2013	32	76	10
06-06-2013	19	14	0
13-03-2014	68	28	0
13-11-2012	59	6	16
15-11-2013	82	194	61
15-11-2013	53	233	2

PROPERTIES
Name
marketing_campaign
All Properties

APPLIED STEPS
Source
Promoted Headers
Filtered Rows
Removed Duplicates
X Changed Type

- After that I renamed columns and checked and fixed data types every column in the dataset