

Steps to be taken before uploading the harmonized file into the DDE platform:

1. The file should always be in xlsx format
2. The file consists of only 1 sheet (No hidden sheets also allowed)
3. The first column of the file should be the time granular dimension with data type 'Date' and date format is recommended to be in 'YYYY-MM-DD'
4. The dataset should be sorted by the most granular dimension (subject) column followed by time granular column
5. A measure value cannot be represented as blank or '-'.
6. A cell cannot have a formula. All data should be saved as a value
7. Column names should be unique and cannot repeat
8. Column names should not start with numeric value
9. Column names and all dimension values cannot contain spaces and any special characters except underscore (" \_ ")
10. Column names cannot be blank
11. Column names can have a maximum of 50 characters in length
12. Column names and all dimension values should not have two consecutive underscores (" \_\_ ") or should not end with an underscore (" \_ ")
13. Data harmonization is not a part of system driven process

Refer the example below:

	A	B	C	D	E	F	G	H	I	J
1										
2	Time Granularity	Subject								
3	WeekEnding	Region_Storetype	SubBrand	Storetype	Region	Dependant Variable	Measure 1	Measure 2	Measure 3	Measure 4
4	03/10/2019	Region North_1	Energy Drink	1	Region North	718	41	0	2213	0
5	03/10/2019	Region North_2	Energy Drink	2	Region North	3238	41	1	2213	1
6	03/10/2019	Region South_2	Energy Drink	2	Region South	2404	23	1	2213	1
7	03/10/2019	Region South_1	Energy Drink	1	Region South	442	23	0	2213	0
8	03/17/2019	Region South_1	Energy Drink	1	Region South	414	42	0	8038	0
9	03/17/2019	Region North_1	Energy Drink	1	Region North	691	77	0	8038	0
10	03/17/2019	Region North_2	Energy Drink	2	Region North	3433	77	1	8038	1
11	03/17/2019	Region South_2	Energy Drink	2	Region South	2338	42	1	8038	1
12	03/24/2019	Region North_1	Energy Drink	1	Region North	743	104	0	7975	0
13	03/24/2019	Region South_1	Energy Drink	1	Region South	508	44	0	7975	0
14	03/24/2019	Region South_2	Energy Drink	2	Region South	2451	44	1	7975	1
15	03/24/2019	Region North_2	Energy Drink	2	Region North	3577	104	1	7975	1
16	03/31/2019	Region South_1	Energy Drink	1	Region South	450	27	0	8193	0
17	03/31/2019	Region North_2	Energy Drink	2	Region North	4366	59	1	8193	1
18	03/31/2019	Region South_2	Energy Drink	2	Region South	2847	27	1	8193	1
19	03/31/2019	Region North_1	Energy Drink	1	Region North	690	59	0	8193	0

In the above snippet, Time granular column 'WeekEnding' is in 'MM/DD/YYYY' and the sorting is based on time granular column 'WeekEnding'. We need to change the date format to 'YYYY-MM-DD' and sorting as below

	A	B	C	D	E	F	G	H	I	J
1	Format requirements - Time granular column in yyyy-mm-dd, Sort by Subject and then by Time Granularity									
2	Time Granularity	Subject								
3	WeekEnding	Region_Storetype	SubBrand	Storetype	Region	Dependant Variable	Measure 1	Measure 2	Measure 3	Measure 4
4	2019-03-10	Region North_1								0
5	2019-03-17	Region North_1								0
6	2019-03-24	Region North_1								0
7	2019-03-31	Region North_1								0
8	2019-03-10	Region North_2								1
9	2019-03-17	Region North_2								1
10	2019-03-24	Region North_2								1
11	2019-03-31	Region North_2								1
12	2019-03-10	Region South_1								0
13	2019-03-17	Region South_1								0
14	2019-03-24	Region South_1								0
15	2019-03-31	Region South_1								0
16	2019-03-10	Region South_2								1
17	2019-03-17	Region South_2	Energy Drink	2	Region South	2338	42	1	8038	1
18	2019-03-24	Region South_2	Energy Drink	2	Region South	2451	44	1	7975	1
19	2019-03-31	Region South_2	Energy Drink	2	Region South	2847	27	1	8193	1

?

×

+

2

Add Level

×

Delete Level

Copy Level

▲

▼

Options...

☒ My data has headers

Column

Sort by

Sort On

Order

Region\_Storetype

Values

A to Z

Then by

WeekEnding

Values

Oldest to Newest

OK

Cancel

After the changes, dataset format would be as below:

	A	B	C	D	E	F	G	H	I	J
1	Format requirements - Time granular column in yyyy-mm-dd, Sort by Subject and then by Time Granularity									
2	Time Granularity	Subject								
3	WeekEnding	Region_Storetype	SubBrand	Storetype	Region	Dependant Variable	Measure 1	Measure 2	Measure 3	Measure 4
4	2019-03-10	Region North_1	Energy Drink	1	Region North	718	41	0	2213	0
5	2019-03-17	Region North_1	Energy Drink	1	Region North	691	77	0	8038	0
6	2019-03-24	Region North_1	Energy Drink	1	Region North	743	104	0	7975	0
7	2019-03-31	Region North_1	Energy Drink	1	Region North	690	59	0	8193	0
8	2019-03-10	Region North_2	Energy Drink	2	Region North	3238	41	1	2213	1
9	2019-03-17	Region North_2	Energy Drink	2	Region North	3433	77	1	8038	1
10	2019-03-24	Region North_2	Energy Drink	2	Region North	3577	104	1	7975	1
11	2019-03-31	Region North_2	Energy Drink	2	Region North	4366	59	1	8193	1
12	2019-03-10	Region South_1	Energy Drink	1	Region South	442	23	0	2213	0
13	2019-03-17	Region South_1	Energy Drink	1	Region South	414	42	0	8038	0
14	2019-03-24	Region South_1	Energy Drink	1	Region South	508	44	0	7975	0
15	2019-03-31	Region South_1	Energy Drink	1	Region South	450	27	0	8193	0
16	2019-03-10	Region South_2	Energy Drink	2	Region South	2404	23	1	2213	1
17	2019-03-17	Region South_2	Energy Drink	2	Region South	2338	42	1	8038	1
18	2019-03-24	Region South_2	Energy Drink	2	Region South	2451	44	1	7975	1
19	2019-03-31	Region South_2	Energy Drink	2	Region South	2847	27	1	8193	1