OWNER: Kyril Fionov
GITHUB: https://github.com/KyrilFionov
PROJECT: Excel File Transformation.
PROBLEM: My business partners share forecast as an unstructured Excel file, which makes it difficult to scale up further data processing.
SOLUTION: Use Python to transform the input file from business partners into a structured table to further load it into online DataBase.

INPUT

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W			X	Y
2		Week start date	Week	Year	KPI 1			KPI 2			KPI 3			Q3G					Q4G					KPI 4				
3					Q3G	Q4G	Delta	Q3G	Q4G	Delta	Q3G	Q4G	Delta	Overlay 1	Overlay 2	Overlay 3	Overlay 4	Total	Overlay 1	Overlay 2	Overlay 3	Overlay 4	Total	Q3G	BL	Adj	Q4G	Delta
4	1	6/19/2022	25	2022	1.554	1.554	0.000	24.3%	24.3%	0	24.6%	24.6%	0	0.054	0.023	0.004	-0.004	0.077	0.054	0.023	0.004	-0.004	0.077	1.487	1.410	1.487	**1.487**	0.000
5	2	6/26/2022	26	2022	1.567	1.567	0.000	24.7%	24.7%	0	22.9%	22.9%	0	0.054	0.023	0.004	-0.004	0.077	0.054	0.023	0.004	-0.004	0.077	1.496	1.419	1.496	**1.496**	0.000
6	3	7/3/2022	27	2022	1.566	1.566	0.000	25.1%	25.1%	0	24.7%	24.7%	0	0.054	0.023	0.004	-0.004	0.077	0.054	0.023	0.004	-0.004	0.077	1.502	1.425	1.502	**1.502**	0.000
7	4	7/10/2022	28	2022	1.574	1.574	0.000	25.3%	25.3%	0	22.2%	22.2%	0	0.054	0.023	0.004	-0.004	0.077	0.054	0.023	0.004	-0.004	0.077	1.500	1.423	1.500	**1.500**	0.000
8	5	7/17/2022	29	2022	1.563	1.563	0.000	24.9%	24.9%	0	22.1%	22.1%	0	0.054	0.023	0.004	-0.004	0.077	0.054	0.023	0.004	-0.004	0.077	1.489	1.412	1.489	**1.489**	0.000
9	6	7/24/2022	30	2022	1.567	1.567	0.000	24.7%	24.7%	0	21.8%	21.8%	0	0.054	0.023	0.004	-0.004	0.077	0.054	0.023	0.004	-0.004	0.077	1.487	1.410	1.487	**1.487**	0.000
10	7	7/31/2022	31	2022	1.559	1.559	0.000	23.8%	23.8%	0	24.3%	24.3%	0	0.054	0.023	0.004	-0.004	0.077	0.054	0.023	0.004	-0.004	0.077	1.487	1.409	1.487	**1.487**	0.000
11	8	8/7/2022	32	2022	1.559	1.559	0.000	23.8%	23.2%	-66	25.6%	25.6%	0	0.054	0.023	0.004	-0.004	0.077	0.054	0.023	0.004	-0.004	0.077	1.478	1.401	1.476	**1.476**	-0.002
12	9	8/14/2022	33	2022	1.557	1.557	0.000	23.9%	24.2%	36	23.9%	23.9%	0	0.054	0.023	0.004	-0.004	0.077	0.054	0.023	0.004	-0.004	0.077	1.478	1.401	1.479	**1.479**	0.001
13	10	8/21/2022	34	2022	1.562	1.562	0.000	24.3%	24.9%	60	23.7%	23.7%	0	0.054	0.023	0.004	-0.004	0.077	0.054	0.023	0.004	-0.004	0.077	1.480	1.403	1.482	**1.482**	0.002
14	1	8/28/2022	35	2022	1.577	1.577	0.000	23.6%	25.9%	223	25.0%	25.0%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.492	1.415	1.478	**1.478**	-0.014
15	2	9/4/2022	36	2022	1.570	1.570	0.000	23.6%	25.5%	195	27.6%	27.6%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.487	1.410	1.472	**1.472**	-0.015
16	3	9/11/2022	37	2022	1.565	1.565	0.000	23.6%	23.9%	24	25.2%	25.2%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.482	1.404	1.462	**1.462**	-0.019
17	4	9/18/2022	38	2022	1.565	1.565	0.000	23.6%	23.5%	-6	25.2%	25.2%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.479	1.402	1.459	**1.459**	-0.020
18	5	9/25/2022	39	2022	1.570	1.570	0.000	23.6%	22.6%	-101	26.6%	26.6%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.477	1.400	1.455	**1.455**	-0.023
19	6	10/2/2022	40	2022	1.569	1.569	0.000	23.7%	21.6%	-211	28.5%	28.5%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.470	1.392	1.444	**1.444**	-0.025
20	7	10/9/2022	41	2022	1.566	1.591	0.025	24.2%	22.7%	-150	27.2%	27.2%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.473	1.395	1.458	**1.458**	-0.015
21	8	10/16/2022	42	2022	1.581	1.581	0.000	24.3%	22.2%	-207	29.0%	29.0%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.474	1.396	1.449	**1.449**	-0.025
22	9	10/23/2022	43	2022	1.580	1.580	0.000	24.3%	22.3%	-197	21.1%	21.1%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.483	1.406	1.458	**1.458**	-0.025
23	10	10/30/2022	44	2022	1.603	1.603	0.000	24.5%	23.2%	-124	23.5%	23.5%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.492	1.414	1.468	**1.468**	-0.023
24	11	11/6/2022	45	2022	1.643	1.643	0.000	24.6%	24.1%	-56	30.4%	30.4%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.510	1.433	1.489	**1.489**	-0.021
25	12	11/13/2022	46	2022	1.677	1.677	0.000	24.7%	24.5%	-20	31.9%	31.9%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.525	1.447	1.504	**1.504**	-0.021
26	13	11/20/2022	47	2022	1.736	1.736	0.000	25.3%	23.8%	-144	28.0%	28.0%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.560	1.482	1.536	**1.536**	-0.024
27	14	11/27/2022	48	2022	1.739	1.739	0.000	25.3%	21.3%	-400	31.5%	31.5%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.582	1.505	1.551	**1.551**	-0.031
28	15	12/4/2022	49	2022	1.748	1.748	0.000	24.9%	21.2%	-372	35.4%	35.4%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.575	1.497	1.545	**1.545**	-0.030
29	16	12/11/2022	50	2022	1.723	1.723	0.000	25.4%	19.3%	-605	34.7%	34.7%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.553	1.476	1.517	**1.517**	-0.036
30	17	12/18/2022	51	2022	1.616	1.616	0.000	24.5%	18.9%	-558	32.2%	32.2%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.497	1.420	1.463	**1.463**	-0.034
31	18	12/25/2022	52	2022	1.463	1.463	0.000	24.6%	19.8%	-477	32.2%	32.2%	0	0.054	0.023	0.004	-0.004	0.077	0.034	0.023	0.004	-0.004	0.057	1.420	1.343	1.388	**1.388**	-0.032
32																												
33																												
34																												

Region 1 | Region 2 | Region 3 | Region 4 | Region 5 | Model

PYTHON CODE

```
import pandas as pd
import re

# sheet_name = None opens all sheets of the input file
xlsx  = pd.read_excel(r'...\input_file.xlsx',sheet_name=None)

# I remove 'Model' sheet since it is not required in the analysis
xlsx.pop('Model')

# I create a variable that specifies a plan scenario I'm interested in
PLAN = 'Q4G'

regions_data = []

for region, data in xlsx.items():
    data.iloc[0:2] = data.iloc[0:2].fillna(method='ffill', axis=1)
    data = data.fillna('')
    data.columns = data.iloc[0:2].apply(lambda x: ''.join([y for y in x if y]), axis=0)
    data = data.iloc[2:]
    plan_data = (data
                 .assign(Region=region
                         ,Plan=PLAN)
                 .rename(columns=lambda data: re.sub(PLAN,'',data))
                 )
    regions_data.append(plan_data)

total_data = pd.concat(regions_data)
```

```
total_data = total_data[[
    'Year'
    ,'Week'
    ,'Region'
    ,'Plan'
    ,'KPI 1'
    ,'KPI 2'
    ,'KPI 3'
    ,'KPI 4'
    ,'Overlay 1'
    ,'Overlay 2'
            ]]
# Pivoting values to later load data in the online DataBase.
# Pivoting data allows me adding more KPIs to the online table in future without having to
change the structure of the table
pivot_data = pd.melt(total_data,
                id_vars=['Year','Week','Region','Plan'],
                value_vars=['KPI 1','KPI 2','KPI 3','KPI 4','Overlay 1','Overlay 2'],
                value_name='value',
                var_name='metric'
                )
pivot_data.to_excel(r'...\output_file.xlsx', sheet_name='plan_data', index=False)
```

OUTPUT

	A	B	C	D	E	F
1	Year	Week	Region	Plan	metric	value
2	2022	25	Region 1	Q4G	KPI 1	1.55369
3	2022	26	Region 1	Q4G	KPI 1	1.566664
4	2022	27	Region 1	Q4G	KPI 1	1.566025
5	2022	28	Region 1	Q4G	KPI 1	1.574206
6	2022	29	Region 1	Q4G	KPI 1	1.563002
7	2022	30	Region 1	Q4G	KPI 1	1.566696
8	2022	31	Region 1	Q4G	KPI 1	1.559175
9	2022	32	Region 1	Q4G	KPI 1	1.558737
10	2022	33	Region 1	Q4G	KPI 1	1.556556
11	2022	34	Region 1	Q4G	KPI 1	1.562351
12	2022	35	Region 1	Q4G	KPI 1	1.577095
13	2022	36	Region 1	Q4G	KPI 1	1.570479
14	2022	37	Region 1	Q4G	KPI 1	1.565391
15	2022	38	Region 1	Q4G	KPI 1	1.564604
16	2022	39	Region 1	Q4G	KPI 1	1.570489
17	2022	40	Region 1	Q4G	KPI 1	1.568718
18	2022	41	Region 1	Q4G	KPI 1	1.591292
19	2022	42	Region 1	Q4G	KPI 1	1.580637
20	2022	43	Region 1	Q4G	KPI 1	1.580024
21	2022	44	Region 1	Q4G	KPI 1	1.603242
22	2022	45	Region 1	Q4G	KPI 1	1.642645
23	2022	46	Region 1	Q4G	KPI 1	1.677206
24	2022	47	Region 1	Q4G	KPI 1	1.735721
25	2022	48	Region 1	Q4G	KPI 1	1.739097
26	2022	49	Region 1	Q4G	KPI 1	1.747722
27	2022	50	Region 1	Q4G	KPI 1	1.723121
28	2022	51	Region 1	Q4G	KPI 1	1.616408
29	2022	52	Region 1	Q4G	KPI 1	1.462562
30	2022	25	Region 2	Q4G	KPI 1	1.7304
31	2022	26	Region 2	Q4G	KPI 1	1.738612
32	2022	27	Region 2	Q4G	KPI 1	1.741736
33	2022	28	Region 2	Q4G	KPI 1	1.774244
34	2022	29	Region 2	Q4G	KPI 1	1.727062
35	2022	30	Region 2	Q4G	KPI 1	1.744965
36	2022	31	Region 2	Q4G	KPI 1	1.756588
37	2022	32	Region 2	Q4G	KPI 1	1.752193
38	2022	33	Region 2	Q4G	KPI 1	1.745612

plan_data