Python Pandas Excel value comparison

ddas.tech/python-pandas-excel-value-comparison/

January 3, 2023

Python Pandas Excel value comparison In this post we will compare two excel workbooks that has same number of records but the value columns contain different values. We will check how the value has changed in each cell and report the delta in each cell.

Lets begin by creating the two data frames from two different data sets with same number of records and primary keys and different values in some cells.

```
import pandas as pd
import numpy as np

df1 = pd.read_excel('SampleData/SalesData.xlsx')
df2 = pd.read_excel('SampleData/SalesData_1.xlsx')

print("DataFrame 1")
display(df1)
print("DataFrame 2")
display(df2)
```

```
DataFrame 1
```

	Region	Country	Product	WK_1	WK_2	WK_3	Wk_4
0	America	USA	Laptop	1241	1160	1929	1174
1	America	USA	Phone	1098	1092	1089	1819
2	America	Canada	Laptop	1441	1099	1950	1394
3	America	Canada	Phone	1990	1057	1656	1060
4	Europe	Belgium	Laptop	1084	1116	1002	1566
5	Europe	Belgium	Phone	1574	1958	1793	1213
6	Europe	Finland	Laptop	1325	1374	1300	1579
7	Europe	Finland	Phone	1347	1736	1782	1921

DataFrame 2

	Region	Country	Product	WK_1	WK_2	WK_3	Wk_4
0	America	USA	Laptop	1206	1471	1294	1258
1	America	USA	Phone	1273	1065	1324	1306
2	America	Canada	Laptop	1233	1295	1047	1258
3	America	Canada	Phone	1413	1030	1325	1238
4	Europe	Belgium	Laptop	1485	1108	1105	1139
5	Europe	Belgium	Phone	1077	1364	1324	1076
6	Europe	Finland	Laptop	1037	1046	1148	1334
7	Europe	Finland	Phone	1273	1291	1458	1201

We have the two data frames printed and can observer that the values are different across the weeks.

```
df1.fillna('',inplace= True)
df2.fillna('',inplace= True)
comparison_result = df1

if df1.equals(df2) == False:
    print("Not same, so lets find the delta")
    comparison_values = df1.values == df2.values
    print(comparison_values)
    rows,cols = np.where(comparison_values == False)
    for item in zip(rows,cols):
        if df1.iloc[item[0],item[1]] is not None and df2.iloc[item[0],item[1]] is not
None:
            comparison_result.iloc[item[0],item[1]] = '{} ->
{}'.format(df1.iloc[item[0],item[1]],df2.iloc[item[0],item[1]])

    display(comparison_result)

# On printing the comparison_values we observe that the week values are different
between df1 and df2
[[ True  True  True False False False False]
 [ True  True  True False False False False]
 [ True  True  True False False False False]
 [ True  True  True False False False False]
 [ True  True  True False False False False]
 [ True  True  True False False False False]
 [ True  True  True False False False False]
 [ True  True  True False False False False]]
Finally we print the comparison result. eg in Row 0 Week 1 the value has changed from
1241 to 1206
```

Below is the comparison result for the data frame differences.

	Region	Country	Product	WK_1	WK_2	WK_3	Wk_4
0	America	USA	Laptop	1241 -> 1206	1160 -> 1471	1929 -> 1294	1174 -> 1258
1	America	USA	Phone	1098 -> 1273	1092 -> 1065	1089 -> 1324	1819 -> 1306
2	America	Canada	Laptop	1441 -> 1233	1099 -> 1295	1950 -> 1047	1394 -> 1258
3	America	Canada	Phone	1990 -> 1413	1057 -> 1030	1656 -> 1325	1060 -> 1238
4	Europe	Belgium	Laptop	1084 -> 1485	1116 -> 1108	1002 -> 1105	1566 -> 1139
5	Europe	Belgium	Phone	1574 -> 1077	1958 -> 1364	1793 -> 1324	1213 -> 1076
6	Europe	Finland	Laptop	1325 -> 1037	1374 -> 1046	1300 -> 1148	1579 -> 1334
7	Europe	Finland	Phone	1347 -> 1273	1736 -> 1291	1782 -> 1458	1921 -> 1201

Comparison Result