A PYTHON PROJECT

ON

***PANDA AND OPENPYXL***

COMPLETED BY TEAM TORNADO

GUIDE: MR. SHOBHIT NIGGAM

**Submitted By:**

Dolly Bagaria

Devvrat Vaidya

Ishant Tiwari

Pramodh Narayan L

# 

# PROBLEM STATEMENT

To design a python application which will be able to open up 2 Excel files. These files will both be data tables laid out in a similar format as the below mentioned examples.

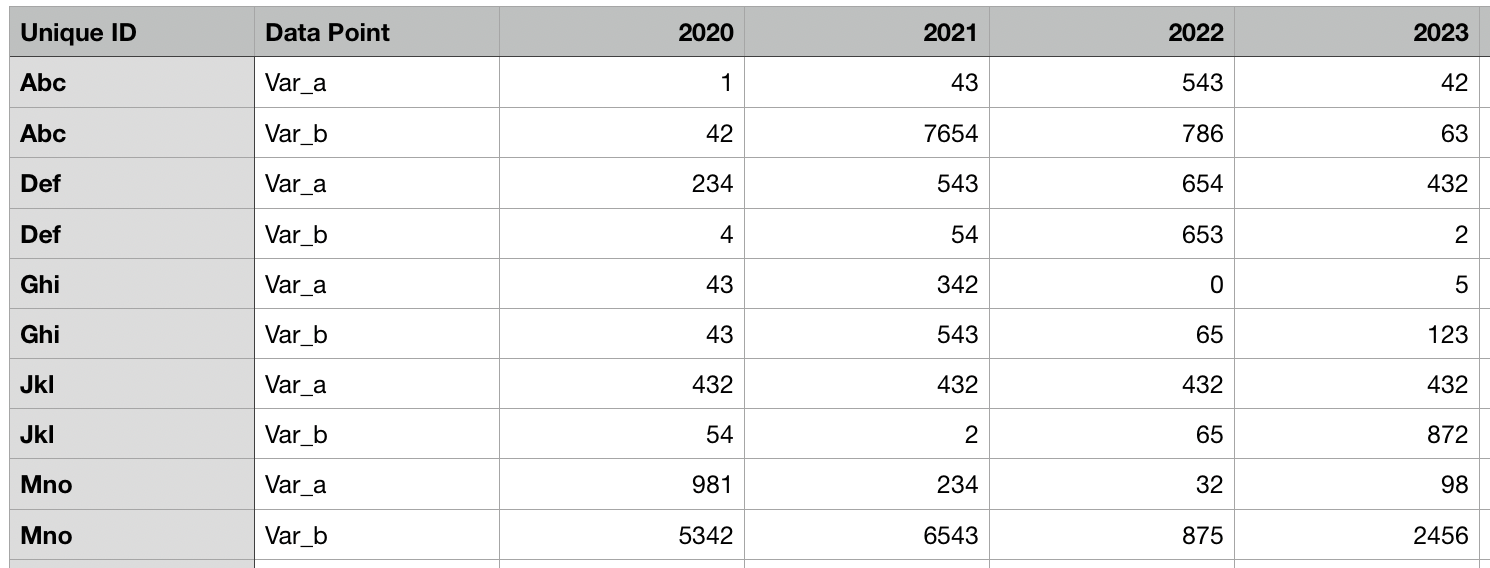


Fig 1 : Table 1 for comparison

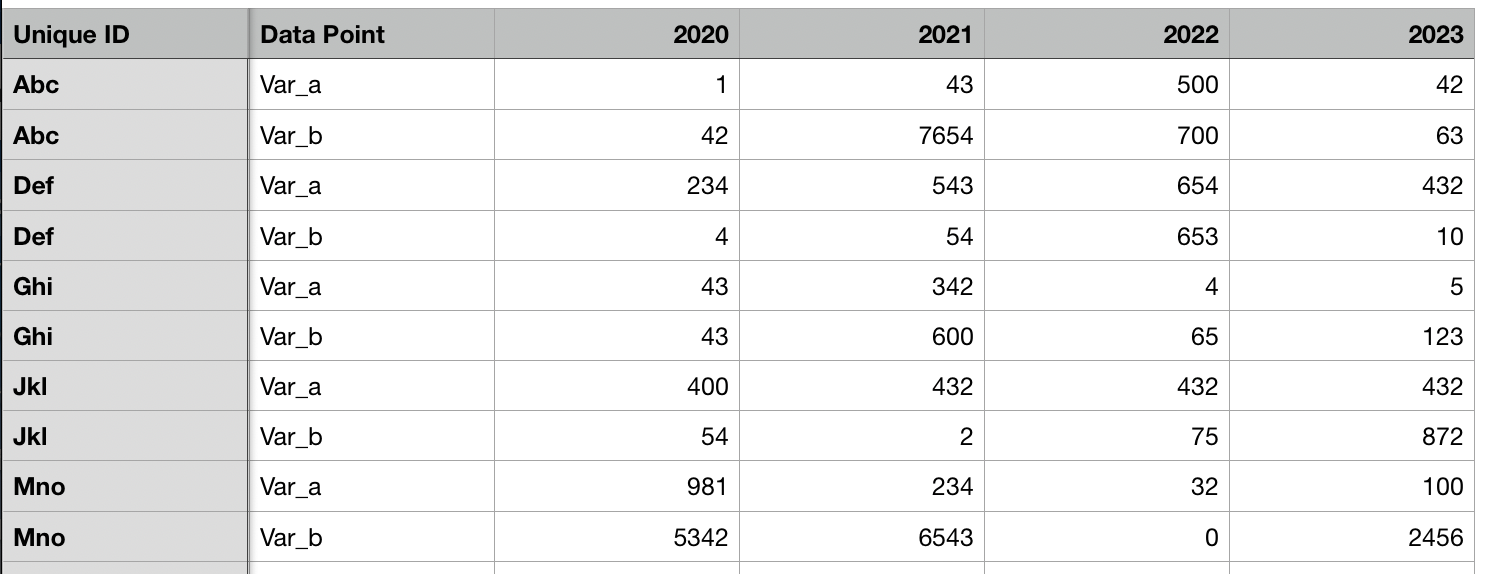


Fig 2 : Table 2 for comparison

It should to be able to tell for each ID what Data Point what has changed with respect to each other. We will be highlighting all the updated rows and columns using a special formatting (in this case the color changes to red).

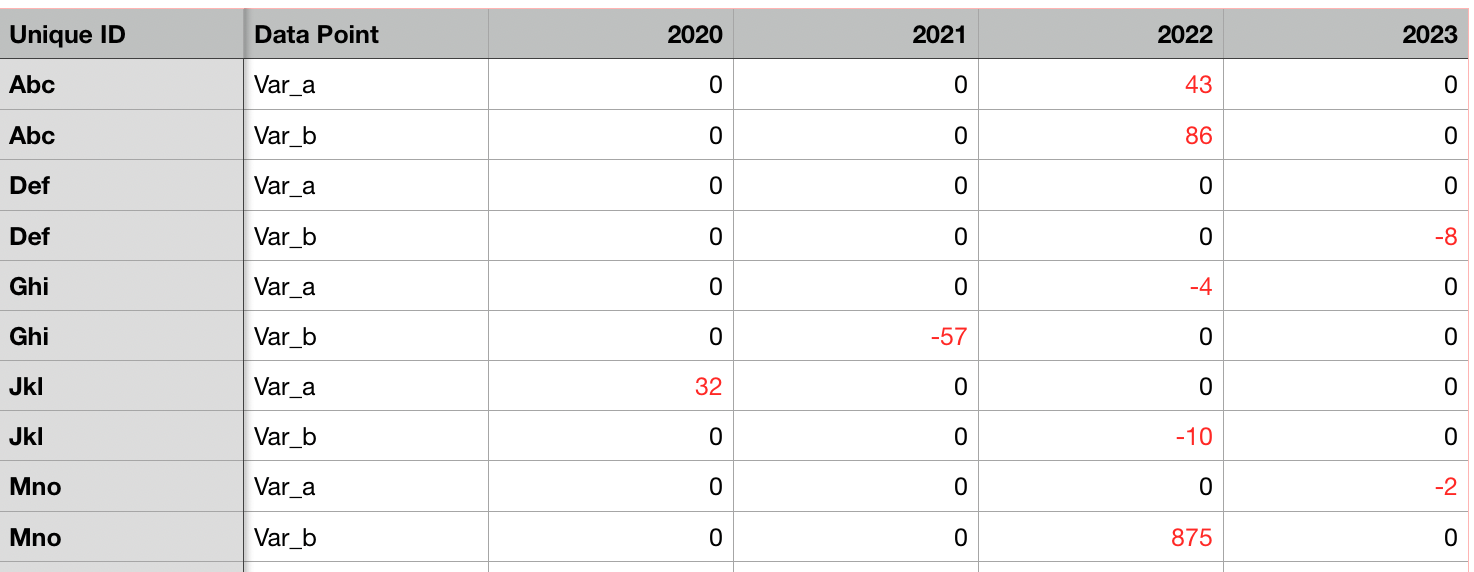


Fig 3 : Comparison Table

# TECHNOLOGIES USED

The project uses **Python 3** as the coding platform and uses two libraries i.e

1. **Pandas -**  an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming
2. **OpenPYXL -**  is a Python library to read/write Excel 2010 xlsx/xlsm/xltx/xltm files. It was born from lack of existing library to read/write natively from Python

# SCOPE OF THE PROJECT

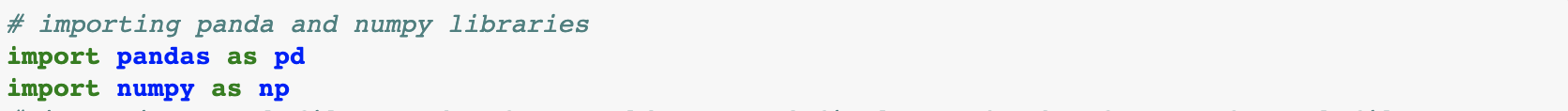
The project will revolve around comparing two excel files and highlighting out the differences to compare values in multiple scenarios, and making a real world application out of it.

## **SOURCE CODE**

**IMPLEMENTATION USING PANDAS**

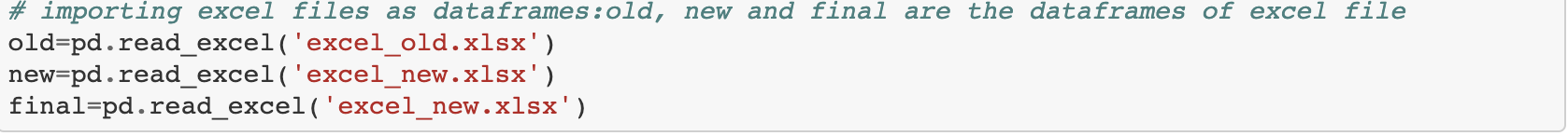
1. **Import**

Importing essential libraries pandas and numpy

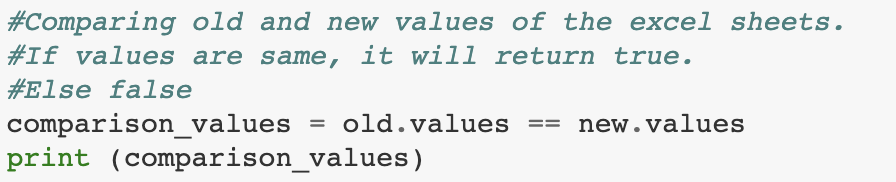


1. **Assigning the excel sheets as inputs in python**

Assigning old and new excel sheets to old and new respectively

****

1. **A boolean matrix is generated using the comparison in-built function between the two generated DataFrames**



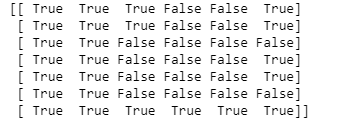
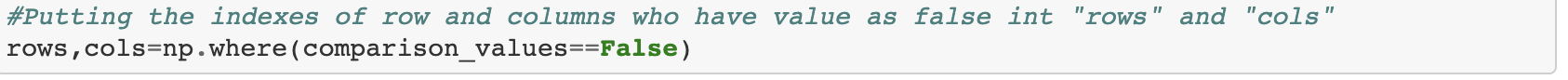


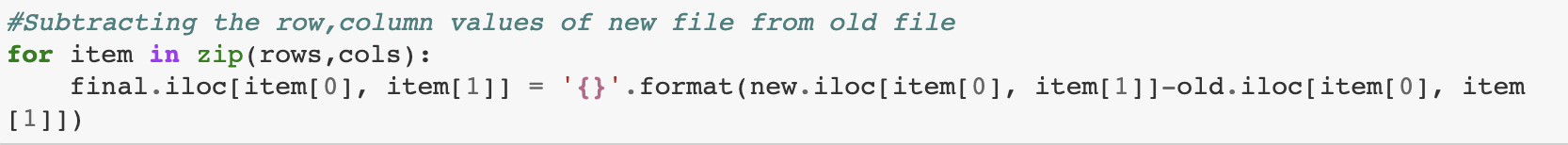
Fig 4 : Boolean Matrix

1. **Looking for rows and columns with ‘False’ values**

****

1. **After subtracting the false column value**

“loc” is used to fetch the row of a dataframe



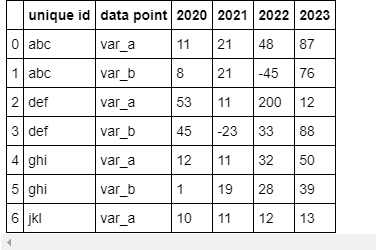
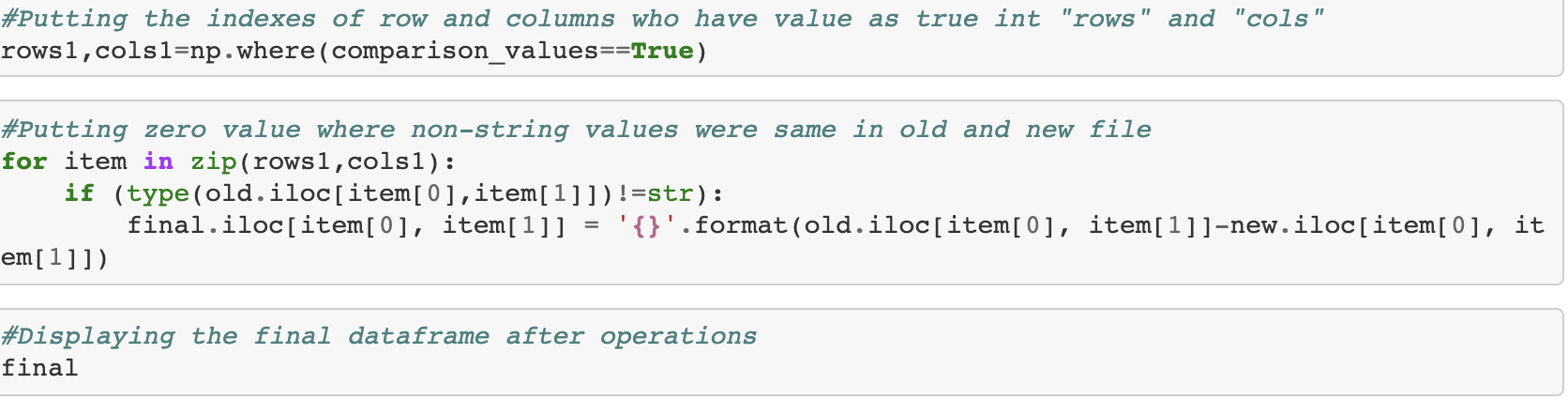


Fig 5: modified excel sheet

1. **Looking for ‘True’ values**



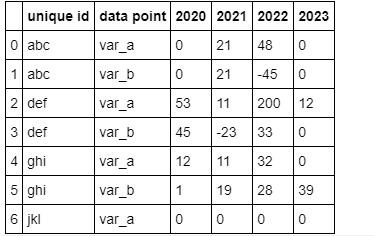
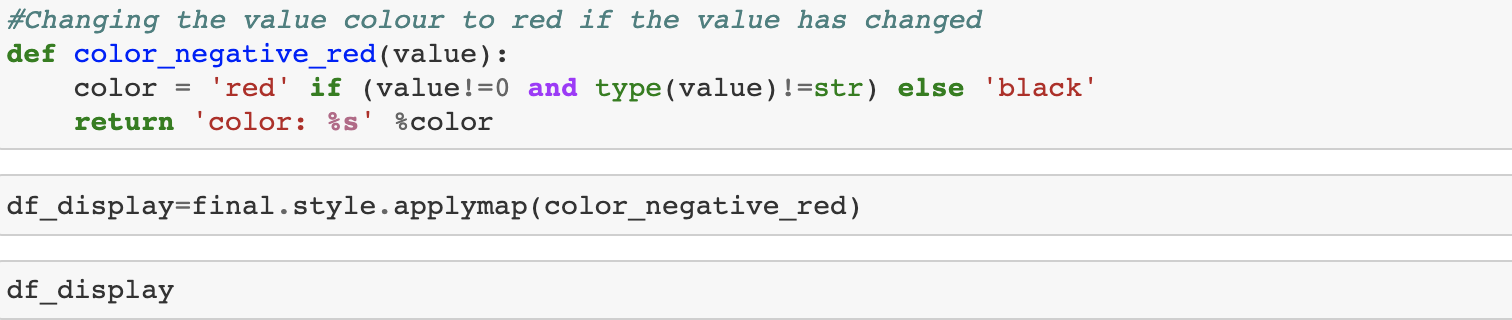
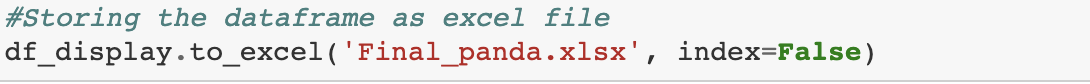


Fig 6: Modified excel sheet

1. **Colouring the values**





## 

Fig 7: Modified excel sheet

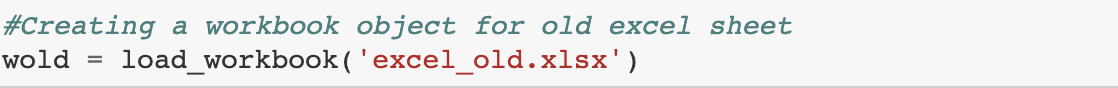
**IMPLEMENTATION USING OPENPYXL**

1. **Importing pyxl libraries**

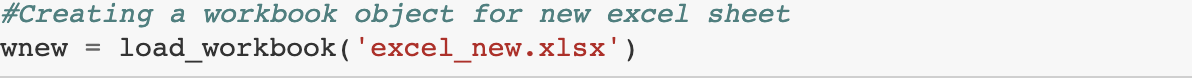
Importing the necessary libraries



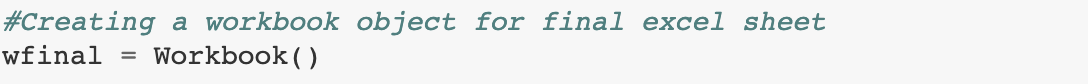
1. **Creating a workbook object from old excel sheet**



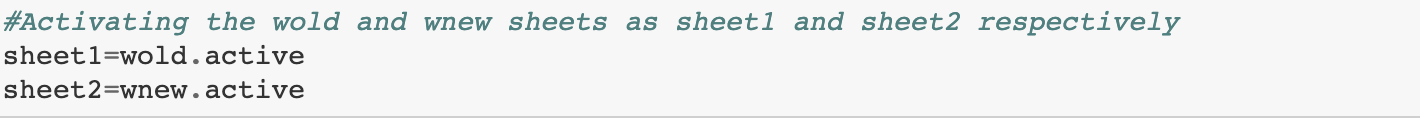
1. **Creating a workbook object from new excel sheet**



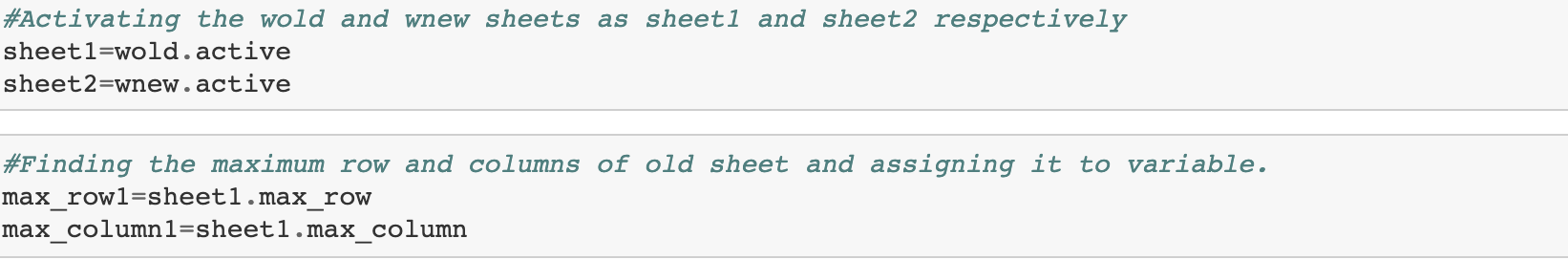
1. **Creating a workbook object for final excel sheet**



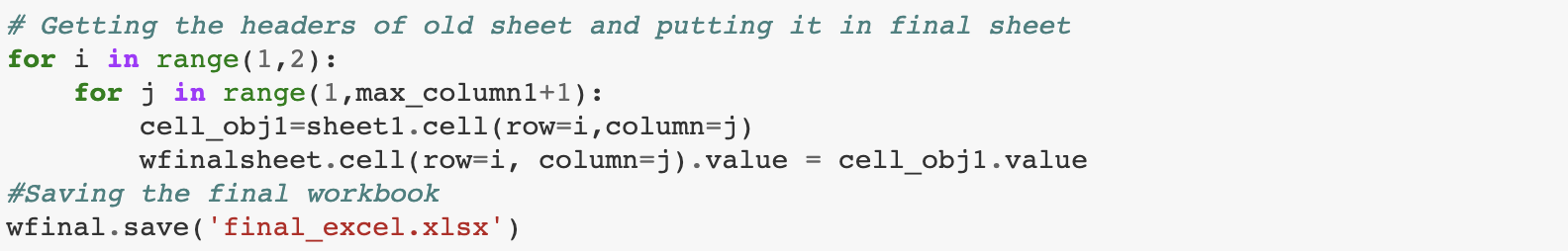
1. **Activating the worksheet and giving the detail as sheet1**



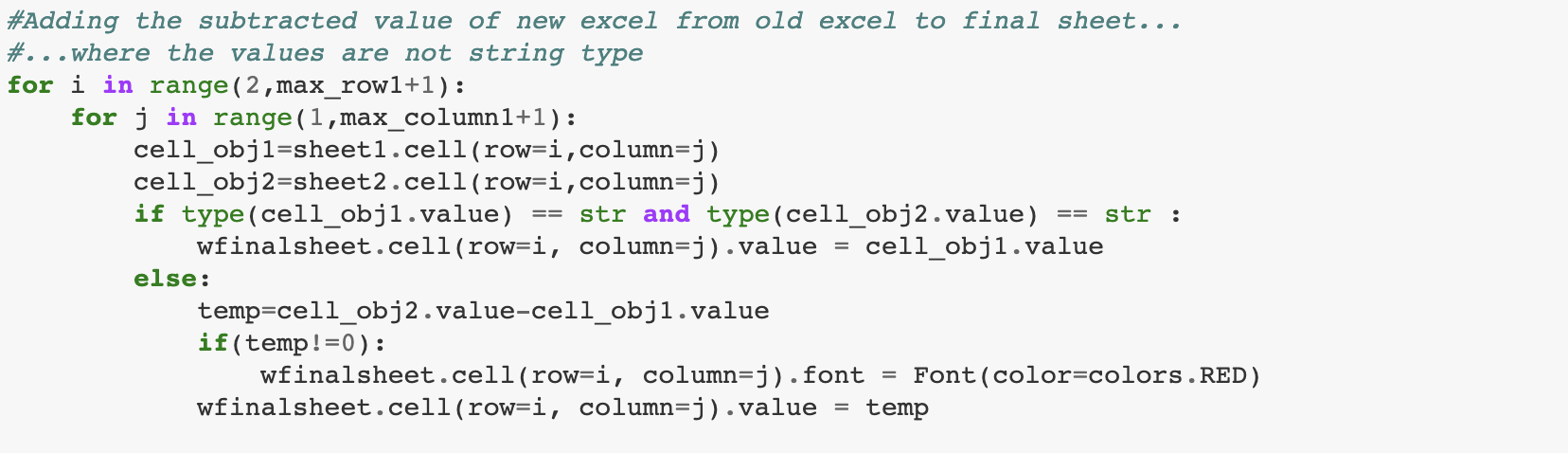
**f) Finding the maximum rows and columns of old sheets**



1. **Getting the headers of old sheet and putting it in final sheet and saving it**



1. **When the values are non-string, adding the values of new excel from old excel to final sheet**



1. **Saving a file**

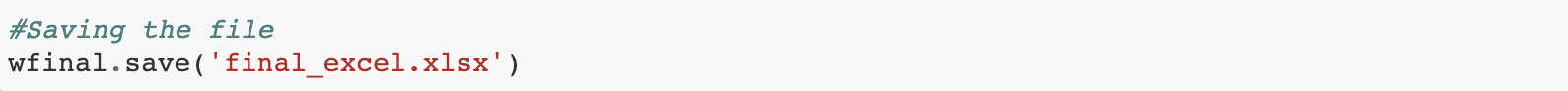
****

Fig 9 : excel\_file\_old.xlsx

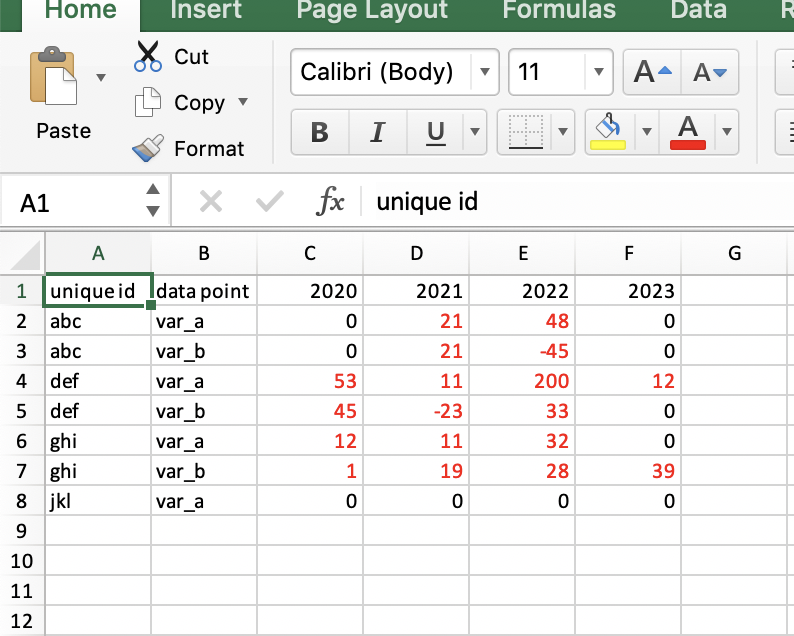


Fig 10 : excel\_file\_diff.xlsx

# Result

The files have been compared using execution of two technologies and a new file has been generated in all instances.