GENESIS- Advanced Python Programming Summary Report

LTTS
GLOBAL
ENGINEERING
ACADEMY

L&T Technology Services

Details

Ver. Rel. No.	Release Date	Prepared. By	Reviewed By	To be Approved	Remarks/Revision Details
1.0	22-03-2020	Abhishek Guria			

Contents

PROBLEM STATEMENT:-

To write program in python to extract data from sheets by giving one or more input and validate them and then to put the whole data of all the sheets of the same in Master sheet. Which will store them.

DEFINING THE IDEA FOR PROJECT :

Customized code is used to read and write data. An excel sheet made by hand containing 5 sheets and one master sheet. Here we search for the details of the person associated with the specific name and registration number on all 5 sheets. Once the data has been downloaded from the all sheets it will be printed on the master sheet. All its usage is used to read better search and write file. The code makes studying easier in the field of data science where there is a lot of data that needs to be extracted.

DETAILED REQUIREMENTS:-

HIGH LEVEL REQUIREMENTS:

ID	Requirements	Explanation	Status
HLR_01	Excel sheets	To get or read Data in different excel sheets for populating master sheet	Implemented
HLR_02	Validating input	Read give input Data from sheet	Implemented
HLR_03	Writing Data	Writing data to Master sheet	Implemented

LOW LEVEL REQUIREMENTS:

ID	Requirements	Description	Status
LLR_01	To Search Data from different sheets by key	Need to search data corresponding to a particular keyword in multiple Excel sheets	Implemented
LLR_02	Read Data from all different sheets	Read Data from different sheets of every single Excel file	Implemented
LLR_03	To Write Data into a master sheet as per the user input after validation	Writing data to a single master sheet as per the user input, after reading all the sheets by automation using python pandas library	Implemented
LLR_04	Print and save Data in assigned sheet	Reading the data as well as writing data after retrieving data in backend via automation and print it if required To show whether data successfully executed	Implemented

SWOAT ANALYSIS:-

Strength: - High Accuracy - Reading data - Writing data	Weakness: - Sometimes overwriting of data occurs - Difficulty in reading huge data in a single sheet
Opportunities: - Has a great advantage in the field of Data Science. - Advantageous in merging multiple sheets	Threats: - Data merging to a single sheet can lead to security issues

4 WH1'S :-

Who : Basically used in research field and data science fields where large data needs manipulation and extraction.

What : Xlsx file or CSV file with python code to read and write data

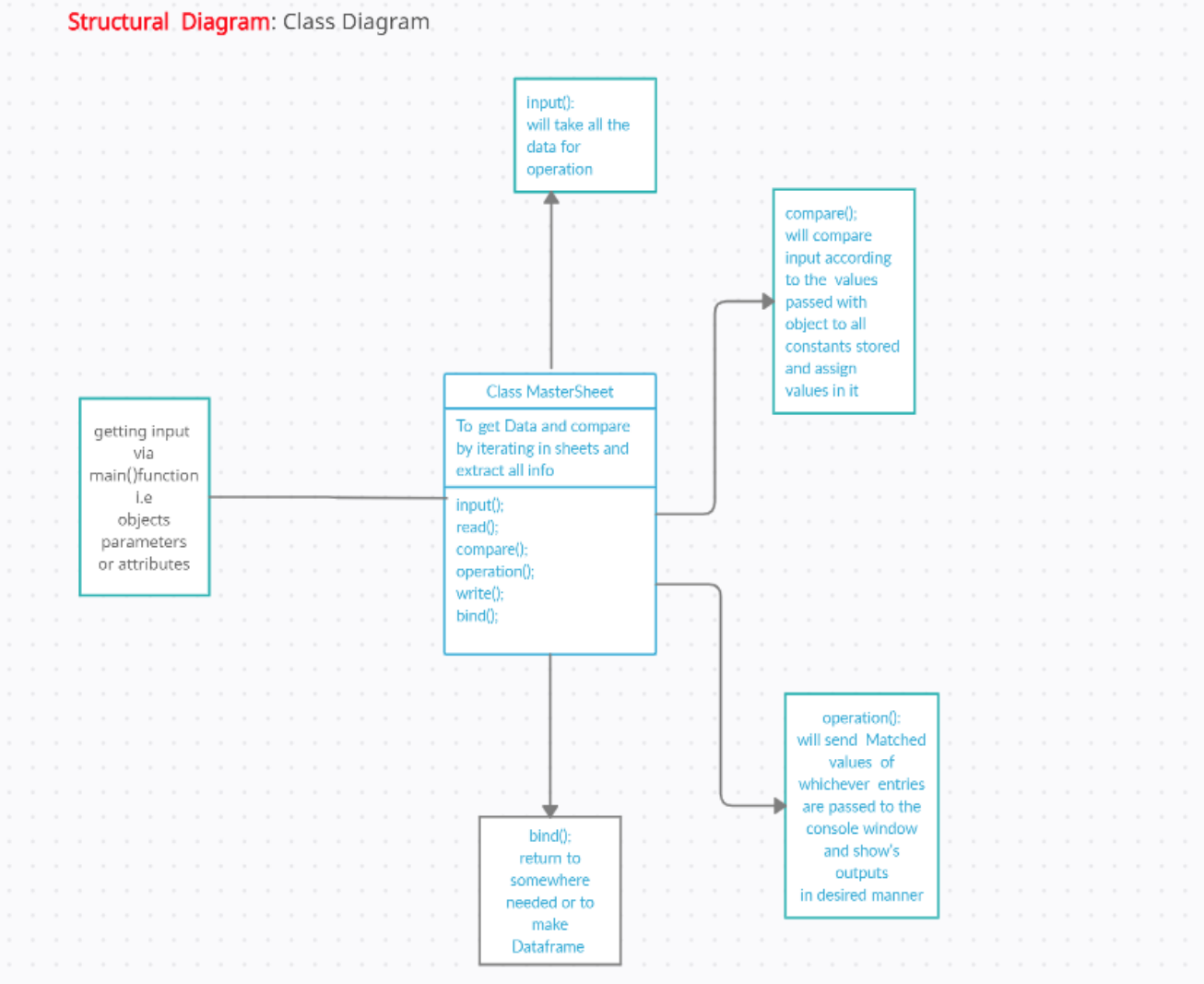
When : Used when large data needs to manipulated and extracted.

Where : In Research and technical fields

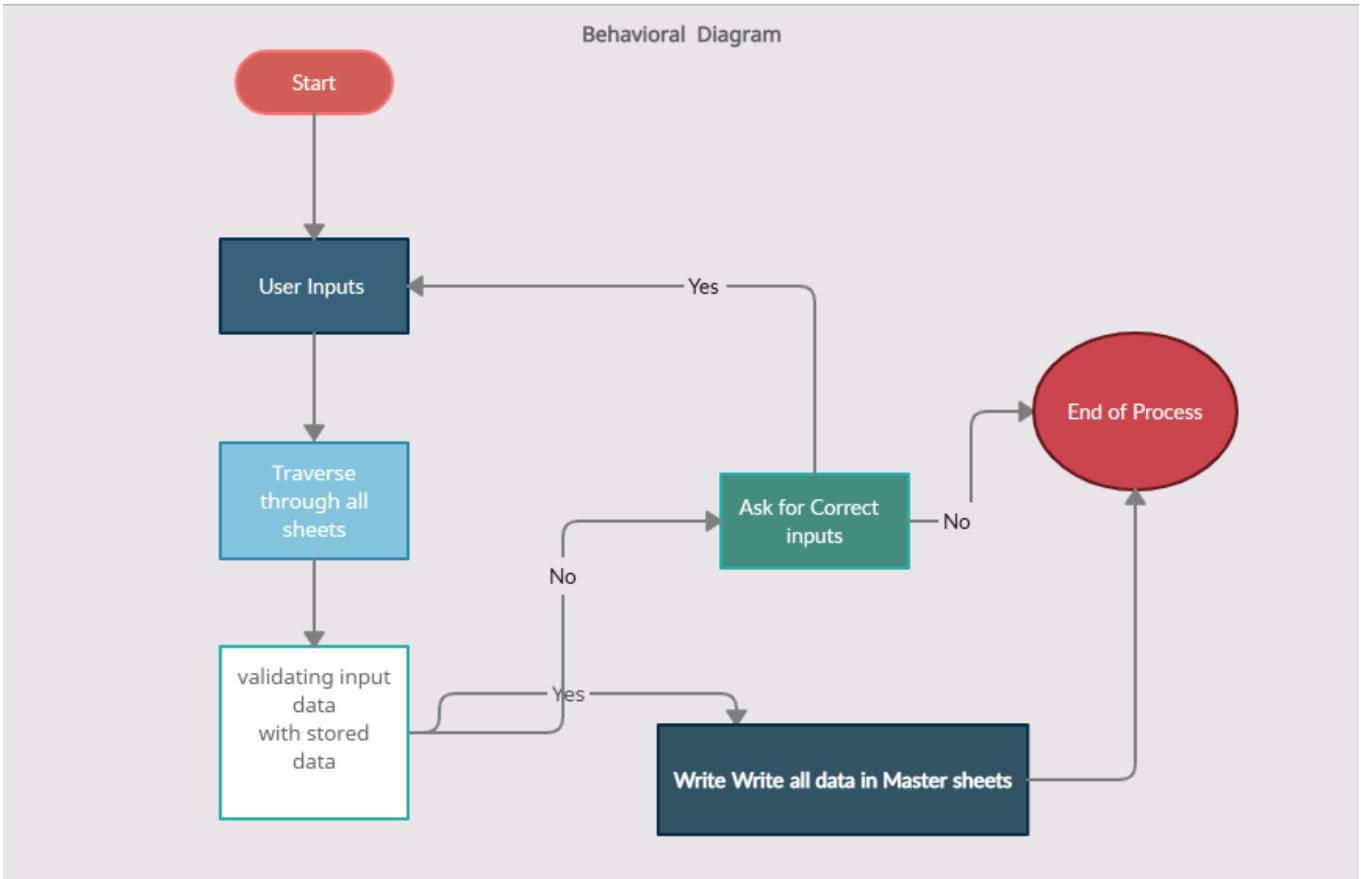
How : User friendly and easily accessible.

DESIGN:-

STRUCTRAL DIAGRAM



BEHAVIORAL DIAGRAM



SOURCE CODE:-

```
import pandas as pd
import openpyxl as op
from openpyxl import Workbook
from openpyxl.utils.dataframe import dataframe_to_rows


def read_path_of_files():
    numberof_sheets = int(input
                          ("Enter Number of sheets To "
                           "read And extract data: "))
    print("1. Enter the Paths of ", numberof_sheets, " Excel sheets")
    for i in range(1, numberof_sheets + 1):
        print('Path of sheet', i, ':')
        filepath_ofSheet.append(input())


def read_files():
    for fpath in (filepath_ofSheet):
        workbok1 = op.load_workbook(fpath)
        dff.append(pd.read_excel(fpath))
        namesheet = workbok1.sheetnames
        print(namesheet)
        allsheetnames.append(namesheet[0])
        wb_read.append(op.load_workbook(fpath))


def write_files_in_one():
    with pd.ExcelWriter('auto5sheets1.xlsx') as writer:
        for (df, a) in zip(dff, allsheetnames):
            df.to_excel(writer, sheet_name=a, index=False)


def header_in_new_sheets():
    MasterSheet = Workbook()
    wb_read = op.load_workbook("auto5sheets1.xlsx")
    wsheet = MasterSheet.active
    wsheet.title = 'output'
    # sheets_read = wb_read.sheetnames
    for sheet in wb_read.sheetnames:
        rs = wb_read[sheet]
        Maxrow = rs.max_row
        Maxcol = rs.max_column
        if sheet == allsheetnames[0]:
```

```
            for j in range(1, Maxcol + 1):
                value = rs.cell(row=1, column=j).value
                wsheet.cell(row=1, column=j).value = value
        else:
            print("You are in now else block")
            maxColmaster = wsheet.max_column
            for j in range(7, Maxcol + 1):
                value = rs.cell(row=1, column=j).value
                print(value)
                maxColmaster = maxColmaster + 1
                wsheet.cell(row=1, column=maxColmaster).value = value
    MasterSheet.save('Openpyxl.xlsx')

    wb_read = op.load_workbook("auto5sheets1.xlsx")
    count = int(input("Enter how Many Data you want to read"))
    for i in range(1, count + 1):
        print('Enter Details for the ', i, ' Data')
        FirstName = str(input('Enter First name '))
        Email = str(input('Enter email Id'))
        PS_No = int(input('Enter PS Number '))
        Datatoload = []
        for sheet in wb_read.sheetnames:
            rs = wb_read[sheet]
            Maxrow = rs.max_row
            Maxcol = rs.max_column
            for i in range(2, Maxrow + 1):
                if rs.cell(row=i, column=1).value == \
                        PS_No and rs.cell(row=i, column=2).value == \
                        FirstName and rs.cell(
                        row=i, column=6).value == Email:
                    if sheet == allsheetnames[0]:
                        for j in range(1, Maxcol + 1):
                            Datatoload.append(rs.cell(row=i, column=j).value)
                    else:
                        for j in range(7, Maxcol + 1):
                            Datatoload.append(rs.cell(row=i, column=j).value)
                    df = pd.DataFrame(Datatoload)
                    df = df.T
        for r in dataframe_to_rows(df, index=False, header=False):
            wsheet.append(r)
        MasterSheet.save('Openpyxl.xlsx')

if __name__ == '__main__':
    filepath_ofSheet = []
    allsheetnames = []
```

```
    dff = []
    wb_read = []
    read_path_of_files()
    read_files()
    write_files_in_one()
    header_in_new_sheets()

# D:\Python\python_project\a\quiz_1_grades.xlsx
# D:\Python\python_project\b\quiz_2_grades.xlsx
# D:\Python\python_project\c\quiz_3_grades.xlsx
# D:\Python\python_project\d\quiz_4_grades.xlsx
# D:\Python\python_project\e\quiz_5_grades.xlsx


# 99003700  Richard Bennett Male    22  richard.bennett@univ.edu
# 99003701  Timothy Parker  Female  25  timothy.parker@univ.edu
# 99003704  Michael Taylor  Male    20  michael.taylor@univ.edu
```

CODE QUALITY:-

PEP8 online

Check your code for PEP8 requirements

All right Save Share

Your code

```
1  import pandas as pd
2  import openpyxl as op
3  from openpyxl import Workbook
4  from openpyxl.utils.dataframe import dataframe_to_rows
5
6
7  def read_path_of_files():
8      numberof_sheets = int(input
9                            ("Enter Number of sheets To "
10                            "read And extract data: "))
11     print("1. Enter the Paths of ", numberof_sheets, " Excel sheets")
12     for i in range(1, numberof_sheets + 1):
13         print('Path of sheet', i, ':')
14         filepath_ofSheet.append(input())
15
```

Check again

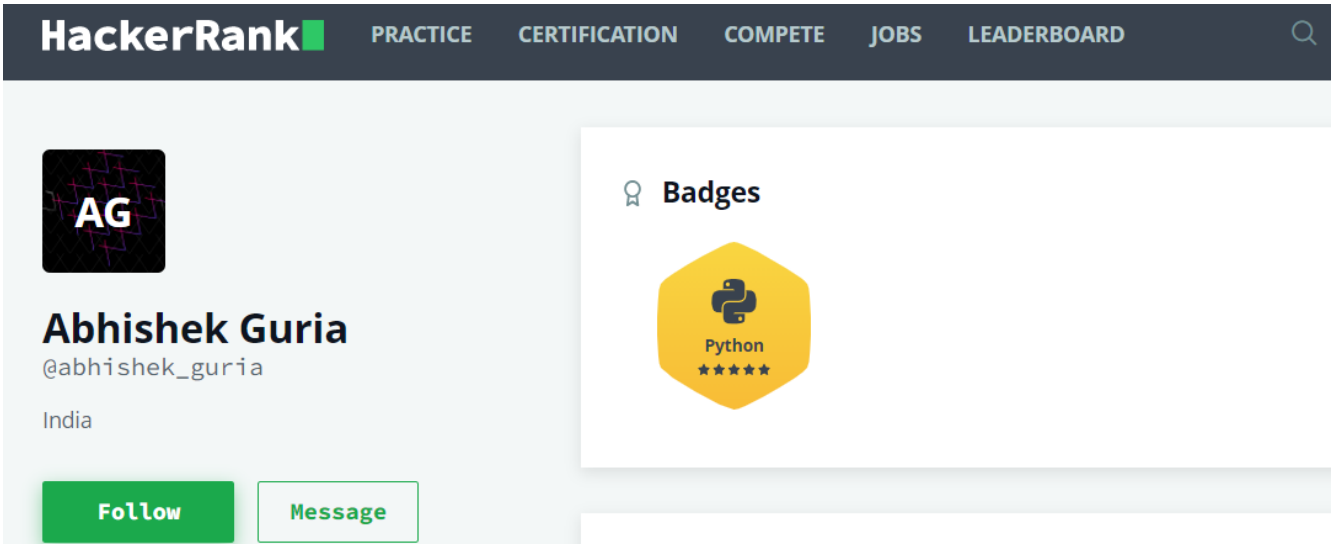
Figure 1 Code Quality

GITHUB LINK: https://github.com/99003726/ Miniproject_Python.git

LEARNING CERTIFICATES:-

1. SOLOLEARN PYTHON CORE COURSE

2. HACKERANK GOLD BADGE



REFRENCES:-

- JAVATPOINT
- GEEKS FOR GEEKS
- PANDAS DOCUMENTATION
- OPENPYXL DOCUMENTATION
- WIKIPEDIA
- HACKERANK
- SOLOLEARN