

IE5600 – Applied Programming for Industrial Systems

AY 2023/24 Semester 1

Individual Assignment 1 - PyPivot

Objectives

At the completion of the individual assignment, you should:

1. Develop a better understanding of how to apply the computational problem solving process to a moderately complex problem.
2. Implement imperative and structured programming paradigms in Python.
3. Apply appropriate control flow statements to implement algorithm.
4. Apply appropriate multi-dimensional data structures to implement algorithm.

General Requirements

You are required to develop a Python program known as **PyPivot** to manage data in a two-dimensional dataset consisting of column header representing attribute name and row header representing record identifier. The program should then allow user to visualise the data in a pivot table. A pivot table is a table of grouped values that aggregates the individual data items in the original dataset within one or more discrete categories. Typical aggregate functions include count, sum, average, minimum and maximum.

	A	B	C	D	E	F
1	S/N	Name	Gender	Age	Employment	Salary
2	1	Albert	Male	21	Employee	4800
3	2	Bob	Male	21	Employee	5000
4	3	Charles	Male	22	Employee	6000
5	4	Derrick	Male	23	Self-Employed	10000
6	5	Fred	Male	23	Unemployed	1200
7	6	George	Male	25	Self-Employed	12000
8	7	Hubert	Male	26	Unemployed	1000
9	8	Issac	Male	28	Employee	3000
10	9	John	Male	30	Employee	2500
11	10	Kerry	Male	30	Self-Employed	8500
12	11	Linda	Female	21	Self-Employed	9000
13	12	Mindy	Female	21	Self-Employed	8500
14	13	Nicole	Female	22	Unemployed	2000
15	14	Oprah	Female	22	Employee	6500
16	15	Penny	Female	23	Employee	7500
17	16	Queenie	Female	25	Employee	7000
18	17	Ruby	Female	25	Employee	4000
19	18	Stacy	Female	27	Employee	3500
20	19	Tiffany	Female	27	Self-Employed	5000
21	20	Ursula	Female	27	Unemployed	1500

	Row Labels	Count of S/N
3		
4	Female	10
5	21	2
6	22	2
7	23	1
8	25	2
9	27	3
10	Male	10
11	21	2
12	22	1
13	23	2
14	25	1
15	26	1
16	28	1
17	30	2
18	Grand Total	20

Figure 1 – Spreadsheet data and corresponding pivot table in Microsoft Excel.

Figure 1 shows a typical pivot table in Microsoft Excel. As shown in Figure 1, categorical attributes may be placed in a pivot table as columns, rows or a mixture of both. If there are more than one aggregated data items, they can be placed in the pivot table column-wise or row-wise as shown in Figure 2.

Column Labels		Employee		Self-Employed		Unemployed		Total Count of S/N		Total Average of Age	
Row Labels	Count of S/N	Average of Age	Count of S/N	Average of Age	Count of S/N	Average of Age	Count of S/N	Average of Age			
Female	5	24.4	3	23	2	24.5	10	24			
Male	5	24.4	3	26	2	24.5	10	24.9			
Grand Total	10	24.4	6	24.5	4	24.5	20	24.45			

Column Labels		Employee		Self-Employed		Unemployed		Grand Total	
Row Labels	Count of S/N	Average of Age	Count of S/N	Average of Age	Count of S/N	Average of Age	Count of S/N	Average of Age	
Female	5	24.4	3	23	2	24.5	10	24	
Male	5	24.4	3	26	2	24.5	10	24.9	
Total Count of S/N	10		6		4		20		
Total Average of Age	24.4		24.5		24.5		24.45		

Figure 2 – Pivot table showing two aggregated data items column-wise (top) and row-wise (bottom).

Use Cases

Implement the following use cases for PyPivot without the use of any Python modules. In other words, your program should **NOT** contain any `import` statement unless you are importing your own user-defined module(s).

S/N	Use Case	Use Case Description/Business Rules
1	New Dataset (0 mark)	<ul style="list-style-type: none"> Initialise PyPivot with a new empty dataset containing zero column and zero row. If there is an existing dataset (at least one column and/or one row, prompt user to confirm the reinitialisation.
2	Load Test Dataset (0 mark)	<ul style="list-style-type: none"> Auto load the dataset shown in Figure 1. This dataset can be downloaded from Canvas.
3	Add Column (1 mark)	<ul style="list-style-type: none"> Add a new column to the dataset with an alphanumeric attribute name. If the dataset contains at least one row, prompt user to input a default value for the new column. The program should use a one-based index number as the default row identifier. It is not necessary to designate a column as the row identifier. <p>Sample Input:</p> <ul style="list-style-type: none"> Add Name Add Gender Add Age Add Employment Add Salary

S/N	Use Case	Use Case Description/Business Rules
4	Delete Column (1 mark)	<ul style="list-style-type: none">• Delete an existing column in the dataset using its attribute name.• If the dataset contains at least one row, all existing data for the affected column should be deleted. <p>Sample Input:</p> <ul style="list-style-type: none">• Delete Gender
5	Add Row (1 mark)	<ul style="list-style-type: none">• Add a new row to the dataset.• There should be at least one column in the dataset. Otherwise, print out an error message.• Prompt user to input data value for each existing column/attribute.• It is not necessary to perform input data validation. <p>Sample Input:</p> <ul style="list-style-type: none">• Name – Albert• Gender – Male• Age – 21• Employment – Employee• Salary – 4800
6	Delete Row (1 mark)	<ul style="list-style-type: none">• Delete an existing row in the dataset using its row identifier, i.e., the one-based index number.• If the row does not exist, print out an error message. <p>Sample Input:</p> <ul style="list-style-type: none">• Delete 1

S/N	Use Case	Use Case Description/Business Rules
7	View Pivot Table Fields (1 mark)	<ul style="list-style-type: none"> View the list of current pivot table field(s) added by the user. By default, a new dataset does not have any pivot table field. There are three types of pivot table field that can be added by user: <ul style="list-style-type: none"> Columns – Categorical attributes designated as the column header. Rows – Categorical attributes designated as the row header. Values – Data attributes that are aggregated from the original data values. The following aggregation functions should be supported: <ul style="list-style-type: none"> Count Sum Average Minimum Maximum <p>Sample Output:</p> <p>Columns:</p> <ul style="list-style-type: none"> Employment <p>Rows:</p> <ul style="list-style-type: none"> Gender <p>Values:</p> <ul style="list-style-type: none"> Salary – Average To indicate column-wise or row-wise display if there are more than one value fields. <p>See Figure A.1 in Appendix.</p>
8	Add Pivot Table Field (1 mark)	<ul style="list-style-type: none"> Add a pivot table field by prompting user to input attribute name and field type. For value field, prompt user to select the required aggregation function. If there are more than one value fields, prompt user to select column-wise or row-wise display. If the attribute has already been added as a pivot table field, print out an error message. <p>Sample Input:</p> <ul style="list-style-type: none"> Gender Row
9	Delete Pivot Table Field (1 mark)	<ul style="list-style-type: none"> Delete an existing pivot table field by prompting user to input attribute name. If the attribute has not been added as a pivot table field, print out an error message.

S/N	Use Case	Use Case Description/Business Rules
10	View Pivot Table (2 marks)	<ul style="list-style-type: none"> Generate the pivot table according to the current dataset and based on the current pivot table fields definition. That is, each time the user views the pivot table, it should be refreshed according to the latest data and fields. The program should be able to handle all scenarios including one or more column fields and row fields as well as one or more value fields of different aggregation. If a value field cannot be calculated for a particular cell, the cell should be shown as a blank cell. <p>See Figure A.2 in Appendix.</p>
11	View Pivot Table with Grouped Summary (1 mark)	<ul style="list-style-type: none"> Use Case 10 with the generation of row summary, column summary and overall summary for each value field. <p>See Figure A.2 in Appendix:</p> <ul style="list-style-type: none"> In this pivot table, the value field is sum of salary. Accordingly, the pivot table should show the total sum of salary for the row field(s), column field(s) and grand total.

Deliverable Submission

The assignment deliverable to be submitted to the Canvas Assignments tool are to be placed in a single zip archive file with the following folders structure:

- **source** subfolder containing:
 - All Python source files that constitute your program.
 - The main source file containing the program entry point should be named as `pypivot.py`, i.e., your program should be runnable with the command `python pypivot.py`

Upload this zip archive file to the designated Canvas Assignment: Assignments > Individual Assignment 1.

Your deliverables must be submitted latest by **Sunday, 22 October 2023, 11:59 pm**. No submission will be accepted for assessment after this date/time and you will be awarded **0** marks.

-- End of Assignment Specification --

Appendix

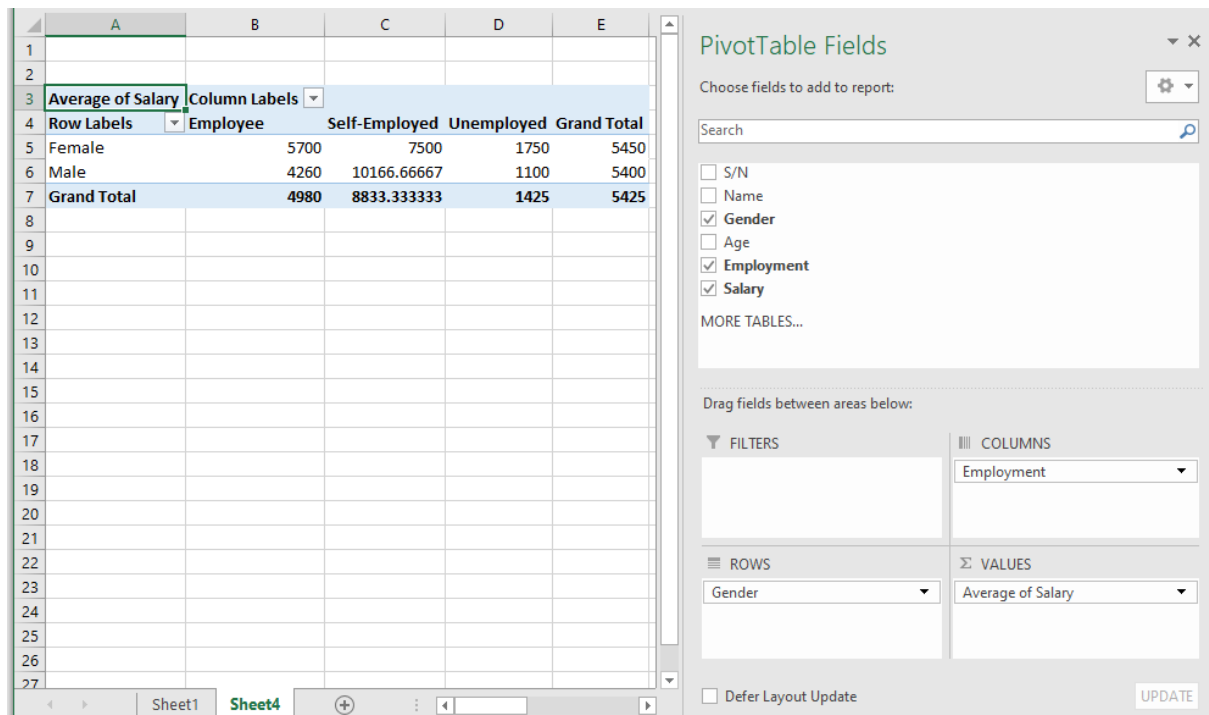


Figure A.1

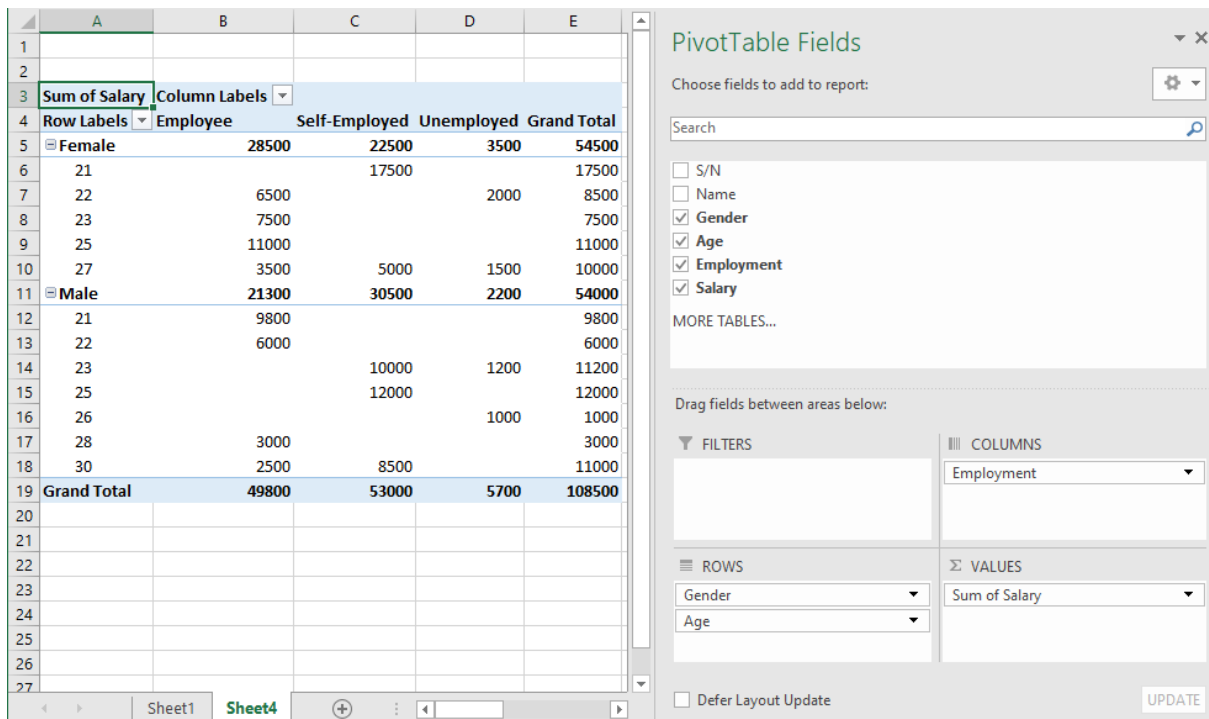


Figure A.2