SINGAPORE UNIVERSITY OF SOCIAL SCIENCES

ANL252
Python for Data Analytics

Group-based Assignment

JANUARY 2023 Presentation

GROUP-BASED ASSIGNMENT

This assignment is worth 20% of the final mark for ANL252 Python for Data Analytics.

The cut-off date for this assignment is **19 February 2023, 2355hrs**.

This is a group-based assignment. You should form a group of **4 members** from your seminar group. Each group is required to upload a single report via your respective seminar group site in Canvas. Please elect a group leader. The responsibility of the group leader is to upload the report on behalf of the group. Those submitting individually will be given a 10 marks deduction.

It is important for each group member to contribute substantially to the final submitted work. All group members are equally responsible for the entire submitted assignment. If you feel that the work distribution is inequitable to either yourself or your group mates, please highlight this to your instructor as soon as possible. Your instructor will then investigate and decide on any action that needs to be taken. It is not necessary for all group members to be awarded the same mark.

Up to 25 marks of penalties will be imposed for inappropriate or poor paraphrasing. For serious cases, they will be investigated by the examination department. More information on effective paraphrasing strategies can be found on https://academicguides.waldenu.edu/writingcenter/evidence/paraphrase/effective.

If your course involves programming, you are urged to read the following articles as well:
https://wiki.cs.astate.edu/index.php/Plagiarism_in_a_Programming_Context

https://www.turnitin.com/blog/plagiarism-and-programming-how-to-code-without-plagiarizing-2

Note to Students:

You are to include the following particulars in your submission: Course Code, Title of the GBA, SUSS PI No., Your Name, and Submission Date.

The submitted report must be in **word document** format.

Part 1(a) is to be answered using MS Excel.

Parts 1(b) to 1(c) are to be answered using Python and any related libraries.

The **complete Python codes**, corresponding to each part of the question, are to be provided with clear indication, and expressed **in text format** with the correct indentation(s), in the **appendix** of the report. **Screenshots of the codes are not permitted and will get a 20-mark deduction.**

The charts produced are to be included as **images** in the word document.

Question 1

The dataset used in this assignment comprises employee information of an organization. There are 26 variables and 1250 records in the dataset, and its data dictionary is depicted in Appendix 1.

(a) Analyse the given dataset, and provide ***four*** (4) charts **and** their corresponding tables, using MS Excel. Describe the insights and highlight any interesting observations.

The charts and tables are to be provided as part of the answer in the main report.

(40 marks)

(b) Read the same given raw dataset using Python. Create the **same *four*** (4) charts **and** their corresponding tables produced in Part (a), using Python.

The output of charts and tables is to be provided as part of the answer in the main report, while the code is to be provided in the appendix of the report.

(30 marks)

(c) Develop an interactive user input, which provides the following number options for the user to choose from, when the code/program is executed.

1. Current staff strength
2. Current staff annual renumeration
3. Eligible long service award recipients
4. Exit

- Selecting option 1 will provide information of the current total staff strength, and a breakdown of staff strength by business unit and gender.
- Selecting option 2 will provide information of the current total annual staff renumeration expenditure, and the current median monthly salary.
- Selecting option 3 will output information of the employees (EmpID), job role and age, ordered by descending age, with 35 or more years of service.
- Selecting option 4 will allow the user to exit from the program.
- The code/program should continue to run (i.e., allow user to query/prompt user for option selection) so long as option 4 has not been selected.

The output from selecting each option is to be provided as part of the answer in the main report, while the code is to be provided in the appendix of the report.

(30 marks)

APPENDIX 1 – DATA DICTIONARY

Variable	Description
EmpID	Employee identifier
BusinessUnit	Business Unit/Department
Education	Education level (1: Others, 2: High School, 3: Tertiary, 4: Postgraduate)
FieldOfStudy	Field of academic study
Gender	Gender
Age	Age in years
ProximityToHome	Distance between home and work location in km
Travel	Frequency of work travel
JobGrade	Job grade (1-5, where 5 is highest)
JobRole	Job role
MaritalStatus	Marital status
MonthlyPay	Monthly pay
PreviousEmployers	Number of previous employers worked for
SalaryIncrement	Last salary increment in percentage
PerformanceRating	Last appraisal performance rating (1-5, where 5 is highest)
WorkingExperience	Number of years of working experience
TrainingAttended	Number of trainings attended in the previous year
WorkLifeBalance	Employee rating of work life balance (1-4, where 4 is highest)
WithCompany	Number of years working for the organization
CurrentRole	Number of years working in the current role
LastPromoted	Number of years since last promotion
CurrentManager	Number of years reporting to the current manager
SatisfactionCompany	Satisfaction rating with respect to the company (1-4, where 4 is highest)
SatisfactionRole	Satisfaction rating with respect to the job role (1-4, where 4 is highest)
SatisfactionCoworkers	Satisfaction rating with respect to fellow co-workers (1-4, where 4 is highest)
LeftCompany	Whether the employee has left the company

---- END OF ASSIGNMENT ----