Deakin University
School of IT
Unit Team: SIG 742

Trimester 2, 2023
Assessment Task 1 (Full Mark: 100)

SIG 742: Modern Data Science

Academic Integrity All assignments will be checked for plagiarism, and any academic misconduct will be reported to the unit chair and university. Current usage of generative AI tools for SIG742 assessment should follow guideline **II**.

Instructions

Assignment Questions

There are total 2 parts in this assessment task:

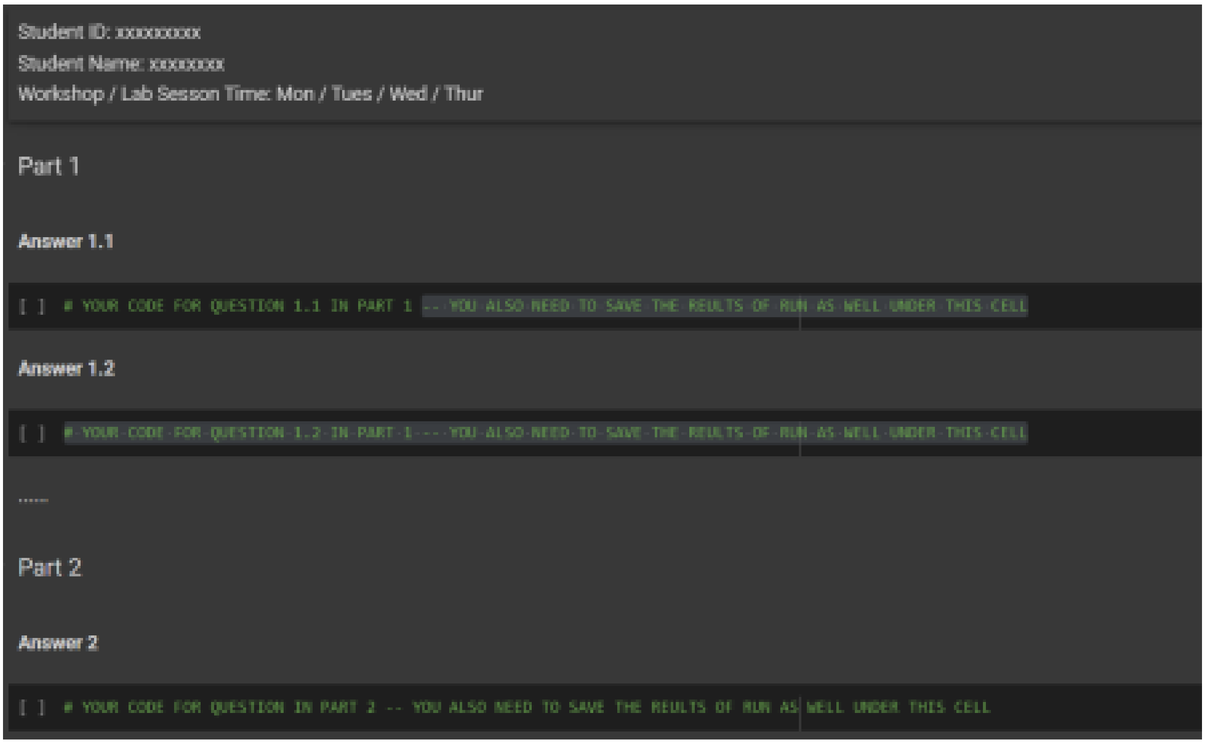
Part 1 The first part will focus on the basic Python programming skills which includes the data types, the control flow, the function and Class, the modules and library from **M02.**

Part 2 The second part is for those who are aiming to achieve 'High Distinction' (HD) for this assessment task, and it will focus on more advanced Python programming skills for data science on particular scenario. This part will require the knowledge covered in **M02** and also some part of **M03**, in particular, numpy and beyond.

What to Submit?

SIG742Task1.ipynb The completed notebook with all the run-able code for all requirements (part 1, also part2 if you choose to work on).

In general, you need to complete, save the results of running, download/export the notebook as a local file, and submit your notebook from Python platform such as Google Colab. You need to clearly list the answer for each question, and the expected format from your notebook will be like in Figure 1.



Student ID: xxxxxxxxx
Student Name: xxxxxxxx
Workshop / Lab Sesson Time: Mon / Tues / Wed / Thur

Part 1

Answer 1.1

```
[ ]  # YOUR CODE FOR QUESTION 1.1 IN PART 1 -- YOU ALSO NEED TO SAVE THE REULTS OF RUN AS WELL UNDER THIS CELL
```

Answer 1.2

```
[ ]  # YOUR CODE FOR QUESTION 1.2 IN PART 1 --- YOU ALSO NEED TO SAVE THE REULTS OF RUN AS WELL UNDER THIS CELL
```

......

Part 2

Answer 2

```
[ ]  # YOUR CODE FOR QUESTION IN PART 2 -- YOU ALSO NEED TO SAVE THE REULTS OF RUN AS WELL UNDER THIS CELL
```

Figure 1: Notebook Format

SIG742Task1Part2.avi (Only if you choose to work on *Part 2* of this assessment task) A video demonstration between 5 and 10 minutes, and the file format can be other common video formats, such as ‘MKV’, ‘WMV’, ‘MOV’ etc.

If you are aiming to achieve ‘High Distinction’ (HD) and choose to work on *Part 2* of this assessment task, one important submission is a short video in which *You* orally present the solutions that you provide in the notebook and illustrate the running of code line by line.

SIG742Task1Part2report.pdf (Only if you choose to work on *Part 2* of this assessment task) A short essay report with no more than 500 words, and the file format can be other common ones, such as ‘.doc’, ‘.docx’, or ‘.md’ etc.

If you are aiming at ‘High Distinction’ (HD) and choose to work on *Part 2* of this assessment task, another important submission is a short essay in which *You* will explain the code you wrote for the solution, any of the resource or material which has been helpful to resolve the problem, and illustrate if there is any alternative way to solve the problem. In addition, you need to follow the common essay writing structure which could be found in https://www.deakin.edu.au/students/study-support/study-resources/academic-skills/essay-writing

Part I

Python Programming

There are 8 questions in this part for total 80 marks, and each question is for 10 marks.

You are required to use Google Colab to finish all the coding in the *code block cell*, and provide sufficient coding comments, and also save the result of running as well.

Question 1.1

ages = [5,31,43,48,50,41,7,11,15,39,80,82,32,2,8,6,25,36,27,61,31] and the length of the list is 21

- find the average value of age (don’t use numpy and do not import other module or library, do not directly use math,statistics module functions except the function sum);
- find the standard deviation from the age list (don’t use numpy and do not import other module or library, do not directly use math, statistics module functions except the function sum)

Hints: You could firstly create a function for calculating the mean, then create the second function for calculate the sum of square deviations of sequence data, then create the thrid one for standard deviation. Check for the steps on how to calculate the standard deviation from https://www.mathsisfun.com/data/standard-deviation-formulas.html

Run your defined function(s) to calculate the average value of age and standard deviation of the age from given list

Question 1.2

Writing the code or function to achieve below requirements:

- You are given the heads of two sorted linked lists list1 = [1,2,4] and list2 = [1,3,4]. Merge the two lists in a one sorted list which is [1,1,2,3,4,4]. (don’t use numpy or import other module / library);
- You are given the heads of two sorted linked lists list1 = [1,2,4] and list2 = [1,3,4]. Merge the two lists in a one sorted list without duplicated elements which is [1,2,3,4]. (don’t use numpy or import other module / library);

Question 1.3

You are required to design the code (by writing code or function) could achieve the mobile number searching mechanism which could allow you to:

- for Four mobile numbers in a list [’10009091003’,’10008293312’,’10007838282’, ‘130001002’]
- print “Yes” when last digit in mobile number is not 3
- print “No” when last digit in mobile number is 3
- print “Not Valid” when length of the mobile number is small than 11 digits

(Optional: You could import re to use the regex function)

Question 1.4

You are required to use For statement to print out below * mark in 5 lines (must use control flow such as For statement).

```
*
**
***
****
*****
```

Example only (you could use others like While)

For i **in range** (5) :

#continue to write your code in below#

Please do not directly print it out, you must finish the question with control flow syntax.

Question 1.5

Given list A = [1,2,3,[4,5,['x1',6,[1,2,3,'sig742']]]],
writing a function to print 1,2,3,4,5,x1,6,1,2,3,sig742
(must define a function to run, the input for function is list A, run the function and print the results line by line or in one line)

Question 1.6

Write the code or function to achieve:
Sort the dictionary list=[{'name': a,' age': 55}, {'name': b,' age': 30}, {'name': c,' age': 50}]
with the value of age (from smallest to largest, you could use default sorted function).

Question 1.7

Define two functions to sum your student ID (all numeric digits in the ID) and then find out if the result is an odd number or even number there are two functions for you to implement (you must create the below two functions):

- one is sum_digits,
- another one is check_studentid.

your code will have two functions as below:

```
def sum_digits ( studentid ):
        #write your code here

        return sum_digits_result
def check_studentid (sum_digits_result) :

        #write your code here
```

Run your functions to find out the result and the result must illustrate whether the result from function check_studentid is odd or even number (this is critical to be used for Question 2).

Question 1.8

Given an array of integers [1, 2, 6, 8], return indices of the two numbers such that they add up to 10.You may assume that each input would have exactly one solution.

Explanation
Given nums = [1, 2, 6, 8], target = 10,
Because nums[1] + nums[3] = 2 + 8 = 10,
return [1, 3].

Your code should have a defined function which could take the array nums= [1, 2, 6, 8] as input and run the function to return a list [1,3] (only [1,3])

Part II

Python Foundations for Data Science

This part is for students who are aiming at 'High Distinction' (HD) for this assessment task. Part II has been designed for a particular scenario based data science problems. There are 2 version in this part for 20 marks: 10 marks for coding, 5 marks for video presentation SIG742Task1Part2.avi (as in 'What to Submit') and 5 marks for the essay SIG742Task1Part2report.pdf (as in 'What to Submit'). You only need to work on one version in this part based on the result in Question 1.7 (details in Part II), and working on the wrong one will result in zero for this question.

For your question, you are required to use Google Colab to finish all the coding in the *code block cell*, provide sufficient coding comments, and also save the result of running.

Which question for you?

If your result in Question 1.7 for the function check_studentid returned as odd, you will need to work on the *version I*; if your result in Question 1.7 for the function check_studentid returned as even, you need to work on the *version II*. You will also need to print/save the result of code running for Question 1.7, and you need to have the correct answer for Question 1.7 to start the Question 2.

- If your answer in Question 1.7 is wrong, Eg: the result should be odd but your result in Question 1.7 is even and you choose even – Zero mark for Question 2;
- If your answer in Question 1.7 is wrong, Eg: the result should be odd but your result in Question 1.7 is even and you choose odd – Zero mark for Question 2;
- If your code in Question 1.7 is not running functionally, Eg: the code in Question 1.7 has error – Zero mark for Question 2;
- If your answer in Question 1.7 is correct, Eg: the result should be odd and your result in Question 1.7 is odd but you choose even – Zero mark for Question 2;
- If you did not have any code for Question 1.7 – Zero mark for Question 2;
- Only if your code in Question 1.7 is running functionally and returns correct answer, your Question 2 will be assessed accordingly.

Question 2 - Version-I

Question 2 Whether the three numbers could form a triangle?

Whether the three numbers could form the triangle? You need to write code (function) with 3 numeric inputs and then check whether the given three numbers could form the triangle.

Question 2.1

You are given three sets of list and each list contains three numbers, you will need to write a code with functionality of inputting the three numbers from each list and find out the answers for below (three) questions. *list*1 = [1, 3, 5], *list*2 = [12, 35, 37], *list*3 = [2, 2, 2.8]

- Define a function to check whether the given three numbers could form the a triangle?
- Define a function to check whether the given three numbers could form the right

  triangle?
- Write the code or function to check whether the given three numbers could form the isosceles triangle?

One possible solution you may try to first define a function to check whether the three input numbers could meet the triangle rule (The sum of the length of the two sides of a triangleis greater than the length of the third side). Then you need to check whether the length of the side could meet the condition of right triangle and isosceles triangle.

Run your code with each list by inputting the numbers and return the results to answer above three questions (you could run your code three times and save the results).

Question 2.2

The problem for Question 2.2 is how to calculate the area of the triangle. When you find out the list(s) in Question 2.1 which could form the triangle, could you also please calculate their area (round to integer)? There is one method to calculate the area of given shape - Heron's formula as below:

$$A = \sqrt{s(s - a)(s - b)(s - c)} \; (1)$$

where $s = \frac{a+b+c}{2}$

You are required to define the function find_area(a,b,c) for this problem, and you will need to run the find_area() and print the results for the list(s) in Question 2.1 which could form the triangle.

Question 2 - Version-II

Question 2 The Monty Hall problem – Do you want to switch the door?

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, **who knows what's behind the doors**, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice? Scenario is as in Figure 2:

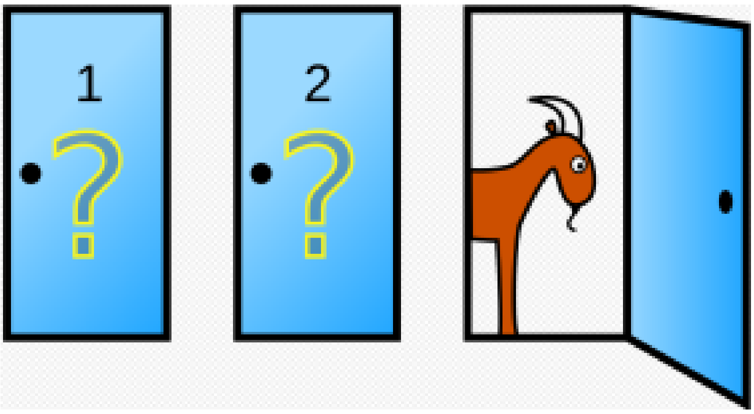


Figure 2: In search of a new car, the player picks a door, say 1. The game host then opens one of the other doors, say 3, to reveal a goat and offers to let the player switch from door 1 to door 2.

Question 2.1

You will need to write a code to find out what is the probability to get the car if you switch the door?

One possible solution You first want to define a variable with random choice to indicate which door has the car. Then you need to also define a variable with random choice to indicate the choice that challenger has selected. Then host will need to pick the door that is not same as the challenger's choice and also did not has the car. Then you need to write the condition to judge whether the other door that challenger did not pick has the car or not. Then repeating the above for *n* times for calculate the probability.

Guideline for the Application of Generative Artificial Intelligence in SIG742 Assignments

This section delineates the expectations for SIG742 students regarding the conscientious and ethical employment of generative Artificial Intelligence (AI) tools in their assignments.

Introduction

Generative AI, with its capacity to spontaneously produce content like audio, code, written text, images, and videos, harbors enormous transformative potential for society and economy. Despite its vast opportunities, there are significant risks and challenges associated with its use.

In academia, generative AI can be a valuable tool, aiding in tasks such as text summarization and refinement for improved readability. However, it also introduces potential pitfalls including IT security issues, intellectual integrity and property protection concerns, and potential breaches of confidentiality. When data is input into generative AI tools, it enters the public domain, accessible by undefined third parties, thereby opening up debates about the rightful authorship of the generated content.

As generative AI technologies and their applications continue to evolve rapidly, they will undoubtedly introduce new facets and concerns to consider.

Directives for Using Generative AI in SIG742 Assignments

AI tools, such as ChatGPT, can generate a broad range of content, from essays and code to poems, based on user input. While these tools serve as a valuable platform for learning and exploration, they should not supersede original work and critical thought. Therefore, SIG742 students are urged to be prudent when employing generative AI tools in their assignments. When incorporating AI tools in SIG742 assessments, please adhere to these guidelines:

Reference AI-generated content Acknowledge and cite the AI tool as the source if you utilize or alter any AI-generated content in your assessment. Use the Harvard referencing style for proper citation format. Neglecting to accurately cite AI-generated content may lead to plagiarism allegations and academic misconduct penalties.

Ensure content appropriateness [Refrain from employing AI tools to generate content that is irrelevant, inappropriate, offensive, or harmful to yourself or others. Uphold the intellectual property rights of others by avoiding the use of AI tools to generate content that infringes on copyrights.

Exercise responsible and ethical use of AI tools Misusing or abusing AI tools in ways that contradict the university’s policies, values, or codes of conduct is unacceptable.

Refrain from using AI tools for coding tasks Coding tasks are intended to assess your programming skills and knowledge, which cannot be evaluated through AI-generated code. Using AI tools to generate code for coding tasks will result in a zero score for that portion of the assessment and may also lead to academic misconduct penalties.

By employing AI tools in SIG742 assessments, you consent to comply with these rules and accept the repercussions of any violation. Deakin University strictly enforces its policy against plagiarism or collusion. The Deakin Academic Integrity Committee will review and make decisions on allegations of academic integrity breaches involving AI tools or other sources. Students could face severe consequences, including exclusion from the unit or the university in extreme cases.

For any questions or concerns about using AI tools for university assessments, please reach out to the teaching team for clarification.